Skin image mosaicing with topological inference and global adjustment

Khuram Faraz

UNIVERSITÉ DE LORRAINE

École doctorale IAEM Lorraine

Mosaïque d'Images Cutanées avec Inférence Topologique et Ajustement Global

THÈSE

présentée et soutenue publiquement le 14 décembre 2017

pour l'obtention du

Doctorat de l'Université de Lorraine

(Spécialité Automatique, Traitement du Signal et des Images, Génie Informatique)

par

Khuram Faraz

Composition du jury

Rapporteurs :	Ernest HIRSCH	PU, Université de Strasbourg, ICube, Strasbourg, France
	Sylvie TREUILLET	MCF, Université d'Orléans, Orléans, France
Examinateurs :	Frédéric CHAMPAGNAT	CR, ONERA, DTIM, Palaiseau, France
	Sylvain GIOUX	PU, Université de Strasbourg, ICube, Strasbourg, France
Invités :	Marine AMOUROUX	PhD, Université de Lorraine, CRAN, Nancy, France
Directeur de thèse :	Walter BLONDEL	PU, Université de Lorraine, CRAN, Nancy, France
Co-Directeur de thèse :	Christian DAUL	PU, Université de Lorraine, CRAN, Nancy, France

Centre de Recherche en Automatique (CRAN)
UMR 7039 Université de Lorraine — CNRS
avenue de la forêt de Haye 54516 Vandœuvre-lès-Nancy
Tél : +33 (0)3 83 59 59 59 Fax : +33 (0)3 83 59 56 44

Acknowledgements

PhD grant for this work was co-funded by the European Regional Development Funds (FEDER) and the Conseil Régional de Lorraine (Regional Council of Lorraine) in the framework of project InnovaTICs-Dépendance.

I thank Mr. Ernest Hirshch and Ms. Sylvie Treuillet for evaluating my dissertation, which gave me the opportunity to defend my thesis. I thank them also for being part of this defense along with Mr. Frédéric Champagnat and Mr. Sylvain Gioux. I appreciate the efforts these four put into following the presentation of my work and thank them for suggesting some alternative approaches and for pointing out some shortfalls in my work. I also thank Mr. Walter Blondel and Mr. Christian Daul for their supervision of my work and timely proofreading of my dissertation and other related communications. I am thankful to Ms. Marine Amouroux for her support in the data acquisitions aspect of this work. Moreover, I am grateful to Ms. Carole Courrier for facilitating my displacements for the communication of my work and to Ms. Christine Pierson for her professional advice and active assistance pertaining to several administrative aspects.

Contents

List of Figures

List of Tables

Résumé Général

1 Contexte Médical

Les travaux de cette thèse s'inscrivent dans le cadre d'un projet de R&D collaborative "InnovaTICs" (financement FEDER et Région Lorraine) dans le domaine de la télédermatologie. L'objectif est de fournir aux dermatologues un outil permettant une télé-expertise plus efficace des affections cutanées. La télédermatologie facilite les consultations dermatologiques pour les patients situés dans des régions éloignées des centres urbains ou ayant une mobilité réduite. Bien que la télédermatologie soit déjà utilisée et son avantage économique ait été démontré [Mas+09], son usage fait face à certaines limitations : les dispositifs d'acquisition généralement utilisés, tels que les smartphones et les caméras portables, fournissent des images sans rendu colorimétrique standardisé et avec une résolution qui dépend de la distance de prise de vue (d'autant plus réduite que la zone de tissu à imager est grande). Puisque les dermatologues ont besoin non seulement d'observer la zone complète de la peau affectée, mais aussi les zones périphériques saines afin d'effectuer un diagnostic approfondi, l'objectif de ce projet est de surmonter ces limitations en construisant des images panoramiques hautement résolues de surfaces étendues de peau avec une méthode adaptée de mosaïquage d'images de séquences vidéos acquises à l'aide d'un dispositif innovant incluant la maîtrise de la colorimétrie. [Amo+15]

2 Cadre Scientifique

Le mosaïquage d'image à partir d'un ensemble d'images de la même scène ou à partir des images issues d'une séquence vidéo consiste à placer ces images dans un même repère géométrique de sorte que leurs parties communes se superposent. L'organigramme en Fig. 1 montre les étapes essentielles du mosaïquage. Une description rapide de ces principales étapes est donnée ci-après :

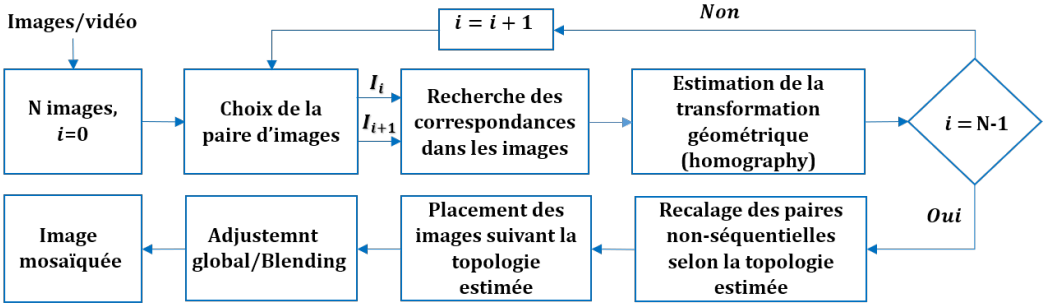


Figure 1: Les principales étapes du mosaïquage.

2.1 Recalage d'images et estimation d'homographie

Comme les vidéo-séquences de la peau sont acquises à proximité de la surface avec un petit champ de vue, les images individuelles peuvent être supposées planes (images visualisant des surfaces quasi-planaires). Ces conditions d'acquisition permettent l'utilisation du mosaïquage 2D, dans lequel les images sont placées dans le plan de la mosaïque en superposant leurs parties communes. Estimer des parties communes revient à déterminer la relation géométrique entre les paires d'images consécutives (I_i, I_{i+1}), c.-à-d. à estimer les paramètres d'une homographie qui donne la meilleure correspondance entre les points homologues. Les images consécutives sont recalées afin de déterminer les paramètres optimaux de la matrice d'homographie $H_{i,i+1}^{est}$, dont les paramètres sont : les facteurs d'échelle (f_x, f_y), la rotation 2D ϕ, les facteurs de cisaillement (s_x, s_y), la translation 2D (t_x, t_y) et le changement de perspective (h_1, h_2). Les coordonnées des pixels de l'image source I_{i+1} transformées avec cette matrice doivent être idéalement déplacés en coordonnées des pixels homologues dans l'image cible I_i. $H_{i,i+1}^{est}$ projette un point $\mathbf{x}_{i+1}$, en coordonnées homogènes, dans I_{i+1} sur son point homologue $\mathbf{x}_i$ dans I_i :

$$\mathbf{x}_i = H_{i,i+1}^{est}\mathbf{x}_{i+1} \tag{1}$$

2.2 Utilisation de la topologie

Pour construire la mosaïque, les images d'une séquence sont placées dans un repère commun à l'aide des transformations géométriques les situant par rapport à une image de référence. Un certain recouvrement entre les images est nécessaire pour pouvoir calculer la relation homographique entre elles. Pour les images n'ayant pas ce recouvrement avec l'image de référence, les homographies des paires d'images amenant vers ces images sont enchainées pour pouvoir calculer leurs placements par rapport à l'image de référence. Cependant, pour de longues séquences, l'enchainement le long de la trajectoire d'acquisition peut engendrer une accumulation considérable de petites erreurs. Des acquisitions couvrant une grande surface peuvent contenir des trajectoires enchevêtrées (trajectoires qui se coupent par exemple). Lorsque la topologie des images est considérée, cette erreur peut être minimisée en choisissant des trajectoires alternatives à travers des croisements de la trajectoire séquentielle afin de réduire le nombre des homographies à enchainer pour atteindre les images éloignées de l'image de référence. Dans [FSV01] et [MFM04], la topologie des chemins des centres des images a été considérée pour créer des mosaïques étendues. En outre, un mosaïquage ne considérant que des trajectoires le long des images séquentielles suppose une continuité dans la séquence. Or, les variations dans l'éclairage ou un flou suite à un mouvement rapide de la caméra peuvent engendrer un échec dans le recalage, ce qui interrompt le mosaïquage. Dans ce cas de figure, le calcul de la topologie de la trajectoire des images permet de trouver des chemins alternatifs pour les images se trouvant en aval du point d'interruption.

2.3 Construction d'une mosaïque visuellement cohérente

Alors que le choix des trajectoires les plus directes peut réduire l'accumulation des erreurs, la précision limitée des méthodes de recalage fait qu'il y persiste des désalignements dans la mosaïque finale. En supposant que les erreurs de recalages ne sont pas significatives, après le calcul des homographies séquentielles, le désalignement dans la mosaïque peut être suffisamment corrigé en utilisant diverses techniques d'ajustement global. A cette fin, des auteurs [MFM04; Wei+12b], se sont servi d'une grille prédéfinie sur le plan de la mosaïque. Chaque point de grille est par la suite projeté selon deux homographies, chacune enchainée le long d'un chemin différent qui lie

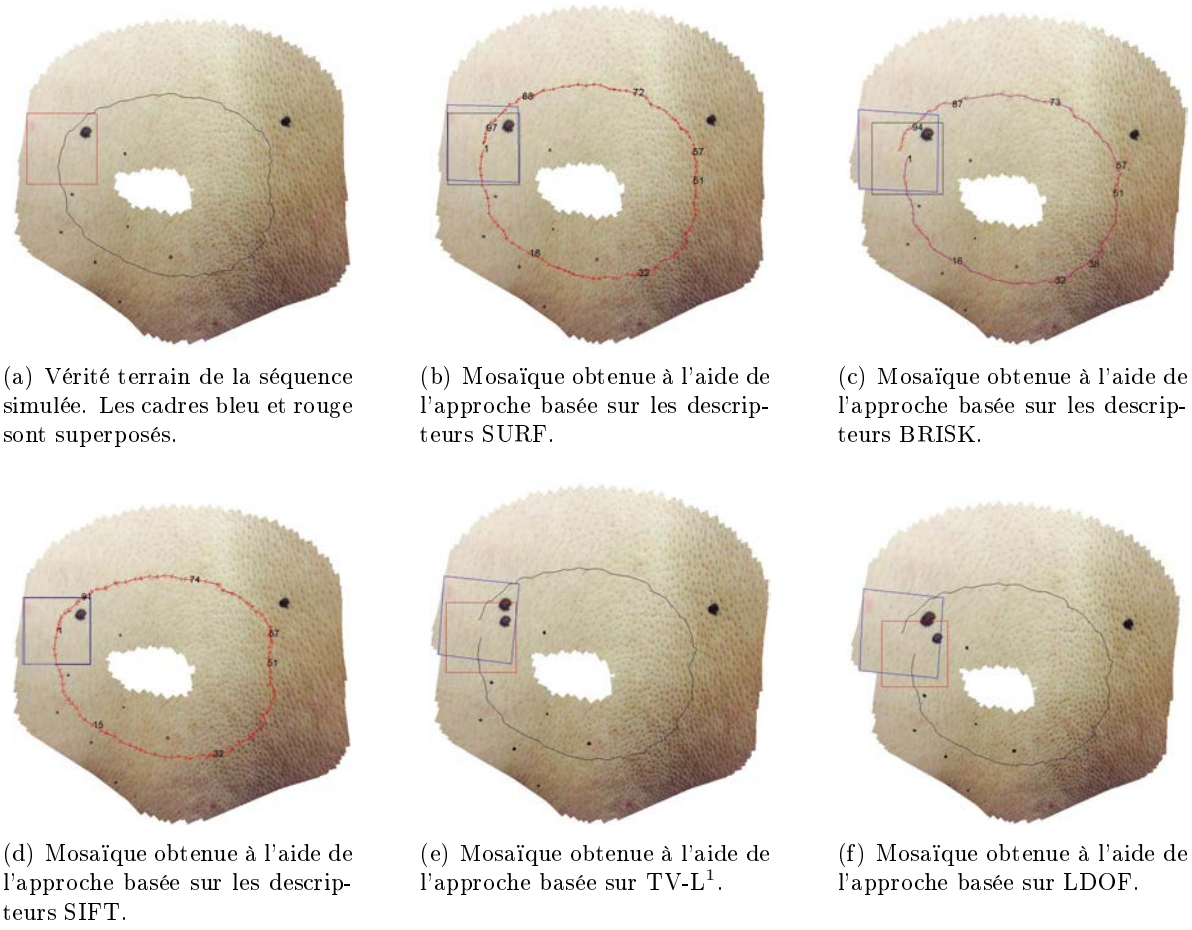
l'image de référence à l'image sur laquelle ce point se situe. La différence entre les positions de ces deux projections, pour l'ensemble des points de grille, est minimisée itérativement. Cependant, cette solution reste coûteuse en termes des calculs. Miranda et al. [ML+08] ont proposé une approche plus rapide pour obtenir un alignement correct de la première et de la dernière image, dans le cas d'une trajectoire d'acquisition en boucle fermée. Leur approche consistait à ajuster les homographies séquentielles d'une manière contrôlée de sorte que la différence entre l'homographie directe entre la première et la dernière image et l'homographie concaténée le long de la boucle soit minimisée. Cette approche permet un ajustement rapide, mais risque d'affecter de manière significative la précision de recalage des paires d'images consécutives, car, outre les homographies, aucune information iconique n'est considérée dans l'ajustement. Dans [BL03], Brown et Lowe ont minimisé, pour chaque image, la somme des distances quadratiques entre les positions des descripteurs SIFT dans cette image et les positions représentant la projection, dans la même image, des descripteurs homologues détectés dans les autres images. Pour cette minimisation, ils font varier les paramètres de caméra à l'aide d'une routine d'optimisation à laquelle les images sont ajoutées une par une (des images avec les meilleures correspondances d'abord). Après l'ajustement global, les discontinuités d'intensité aux bords des images peuvent être gommées à l'aide de différentes approches de blending pour un rendu visuellement cohérent de la mosaïque final [Sze06; Wei+12a]. Le blending consiste à modifier les illuminations des pixels pour qu'il y ait une transition lisse entre les intensités des images qui composent la mosaïque.

3 Contributions

3.1 Choix de la méthode de recalage

La performance de quelques approches de détermination de points homologues a été comparée pour le choix optimal d'une approche de recalage [Far+16]. Des approches appartenant à deux cadres, flot optique et descripteurs des points-clés, ont été considérées. Correlation Flow [DN13] et une autre approche qui calcule le flot optique dans un cadre variationnel total avec une norme L^1 (TV-L^1) [CP11] ont été choisies du domaine de flot optique. SURF (Speeded Up Robust Features, [BTG08]), SIFT (Scale Invariant Feature Transform, [Low04]) et BRISK (Binary Robust Invariant Scalable Keypoints, [LCS11]) sont parmi les approches basées sur les descripteurs des points-clés. LDOF (large displacemnet optical flow, [BM11]), une méthode qui combine le calcul du flot optique avec l'appariement des points-clés, a été choisi également.

Pour avoir la possibilité d'une évaluation quantitative de la cartographie 2D, des séquences vidéos ont été simulées à partir d'une image à haute résolution (3264×2448 pixels) du dos humain. Deux critères de précision ont été utilisés : $\varepsilon_{i,i+1}^{reg}$, l'erreur de recalage entre les images I_i et I_{i+1}, et $\varepsilon_{1,N}^{loop}$, l'erreur de boucle (erreur de mosaïquage entre la dernière et la première image). Ces erreurs représentent la distance euclidienne moyenne entre les deux déplacements des pixels homologues de l'image I_{i+1} dans l'image I_i, un déplacement étant réalisé avec la vraie homographie et l'autre avec l'homographie estimée. La séquence en Fig. 2(a) a été obtenue en simulant numériquement l'acquisition en mouvement circulaire et contient 101 images, extraites de l'image à haute résolution, chacune de dimension 512×512 pixels. Les paramètres d'homographies pour cette séquence sont bornés par ces limites : (f_x, f_y, s_x, s_y)=[0.94, 1.04], ϕ=[0.05, 6.7] degrés, $\sqrt{(t_x^2 + t_y^2)}$=[17, 66] pixels, et (h_1, h_2)=[0, 0.18]$\times 10^{-5}$. Les résultats de mosaïquage, en appliquant différentes méthodes à cette séquence simulée, sont montrés en Figs. 2(b) - 2(f). Bien que SIFT s'avère la méthode la plus précise, compte tenu de son temps de calcul excessif, SURF a été retenu comme la méthode de choix pour le recalage des images cutanées. La Fig. 3 montre



(a) Vérité terrain de la séquence simulée. Les cadres bleu et rouge sont superposés.

(b) Mosaïque obtenue à l'aide de l'approche basée sur les descripteurs SURF.

(c) Mosaïque obtenue à l'aide de l'approche basée sur les descripteurs BRISK.

(d) Mosaïque obtenue à l'aide de l'approche basée sur les descripteurs SIFT.

(e) Mosaïque obtenue à l'aide de l'approche basée sur TV-L[1].

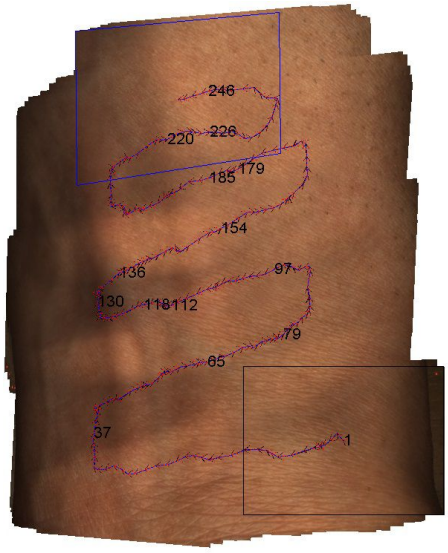
(f) Mosaïque obtenue à l'aide de l'approche basée sur LDOF.

Figure 2: Images mosaïquées à partir d'une séquence simulée numériquement de déplacement d'acquisition d'images d'une surface de peau humaine : a) La vérité terrain (b,c,d,e,f) Mosaïques correspondantes obtenues en utilisant différentes méthodes. L'image de départ de la séquence est encadrée par un carré rouge ou noir, et la dernière image par un quadrangle bleu. Alors que ces deux images coïncident dans l'image de vérité terrain, leur déplacement est une mesure de cohérence globale de la mosaïque dans les autres sous-figures. La trajectoire de séquence, qui forme une boucle fermée dans la vérité terrain, est tracée en noir ou rouge. Cette trajectoire ainsi que la forme du vide circonscrit et les marques noires placées sur l'image sont utiles pour une évaluation visuelle de la mosaïque.

le résultat sur une séquence réelle.

3.2 Mosaïquage en tenant compte de la topologie

Dans l'approche de mosaïquage la plus directe, la relation homographique d'une image avec l'image de référence est calculée en enchainant les homographies des images le long de la trajectoire d'acquisition, comme cela a était fait pour obtenir la mosaïque de Fig. 3(a). Cependant, avec l'enchainement d'un nombre important des homographies, la mosaïque peut subir une déformation significative par l'influence des erreurs accumulées. En plus, l'échec d'un seul recalage empêche le placement des images dans la mosaïque à partir de l'échec de recalage. Pour pallier ces limitations, une topologie plus précise des images de la séquence dans le repère de la mosaïque



(a) Mosaïque obtenue à l'aide des descripteurs SURF

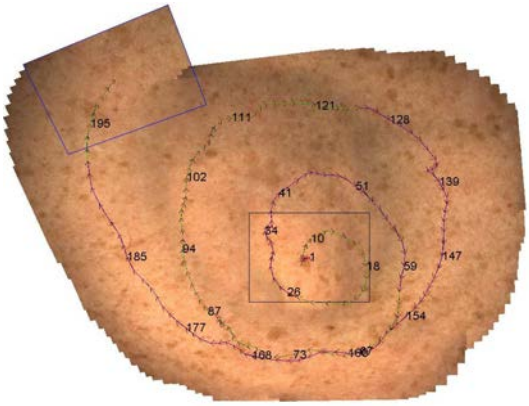
(b) La zone correspondante photographiée avec un smartphone.

Figure 3: a) Mosaïquage d'une séquence, contenant plus de 250 images d'une partie de l'avant-bras antérieur, à l'aide du recalage basé sur SURF. La mosaïque résultante a une taille de 2794×3464 pixels. La trajectoire de la mosaïque est marquée en bleu. L'image initiale est encadrée par un rectangle rouge et des quadrangles verts indiquent des images intermédiaires aux intervalles de 100 images. b) Image de la zone correspondante photographiée avec un smartphone. L'image a une taille de 1406×1372 pixels et représente environ 7×6 cm ^{2} de la surface de la peau.

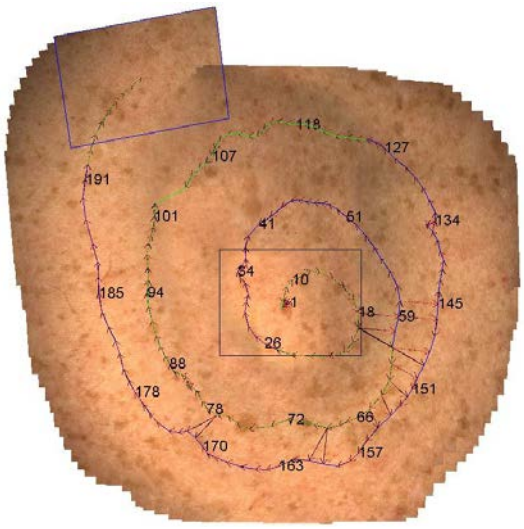
a été cherchée. Dans [VRD13], une acquisition en spirale a été utilisée pour le mosaïquage des vidéos endomicroscopiques. Ce mode d'acquisition a servi de modèle pour la détermination plus précise de la topolgie des images des séquences acquises sur la peau. En plaçant l'image de référence au centre de la spirale, des liens radiaux entre les tours de la spirale peuvent être trouvés soit à l'aide d'une approche itérative [MFM04] soit en cherchant des paires entre les anneaux consécutifs de la spirale avec un large recouvrement à des endroits prédéfinis. En outre, en détectant les paires d'images dont le recalage a échoué (homographie fausse) et en substituant leurs homographies par la dernière homographie correctement calculée, la recherche des chemins alternatifs offre la possibilité de mosaïquer des séquences incluant des homographies "mal calculées". Après avoir trouvé des liens radiaux, la topologie est mise à jour en cherchant les chemins les plus courts, à travers ces liens, entre l'image de référence et les autres images de la séquence à l'aide d'une approche provenant de la théorie des graphes [Dij59]. Cela permet de mosaïquer, tout en limitant l'accumulation des erreurs, des images les plus éloignées dans la trajectoire d'acquisition.

La Fig. 4 montre les résultats obtenus avant et après la mise à jour de la topologie avec les deux approches proposées. Fig. 4(a) montre la mosaïque, obtenue avec le placement des images en enchainant des homographies le long du trajet d'acquisition, d'une séquence acquise sur le dos humain. Cette mosaïque est considérablement déformée à cause d'accumulation des erreurs. Par conséquent, il y a des désalignements et la duplication de la texture. Les mosaïques de Figs. 4(b) et 4(b) ont été obtenues après la mise à jour de la topologie à l'aide des deux approches proposées. Dans les deux cas, l'obtention de plusieurs liens radiaux a permis de trouver les chemins alternatifs plus courts pour atteindre les images de la séquence à partir de

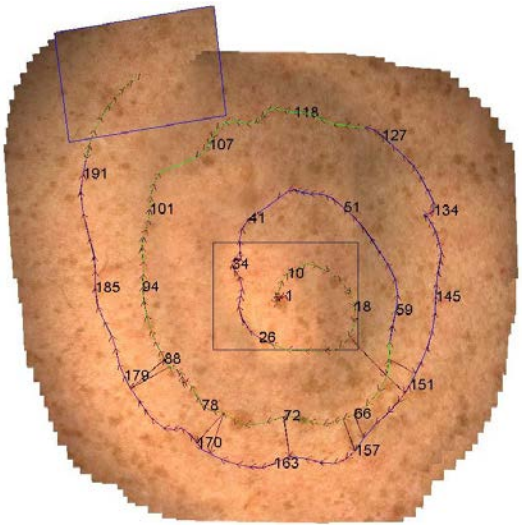
l'image de référence, ce qui a réduit considérablement l'accumulation des erreurs. La cohérence globale de ces deux mosaïques peut être appréciée en les comparant avec l'image acquise avec un smartphone (Fig. 4(d)) de la surface concernée.



(a) Mosaïque (5522 × 4179 pixels) obtenue avec la concaténation des homographies séquentielles. Mosaic obtained using the SURF based registraion scheme by concatenating the homographies over the sequential path.



(b) Mosaïque (4618 × 4630 pixels) obtenue avec la concaténation des homographies le long des trajectoires les plus directes obtenues à l'aide de l'approche itérative.



(c) Mosaïque (4515 × 4544 pixels) obtenue avec la concaténation des homographies le long des trajectoires les plus directes obtenues à l'aide de l'approche basée sur les angles.

(d) Image (2288 × 1906 pixels) acquise avec un smartphone (flash actif) de la zone correspondante.

Figure 4: Illustration, avec le mosaïquage d'une séquence sur le dos humain humaine, de deux approche de mise à jour de topologie. Les quadrangles noir et bleu encadrent, respectivement, la première et la dernière image de la séquence. La trajectoire séquentielle est tracée en bleu et vert. Les traits rouges, avec des flèches indiquant la direction, représentent les trajectoires les plus directes. Les positions, dans la séquence, de quelques images sont illustrées à l'aide du numéro de l'image placé dans la mosaïque à l'endroit du centre de l'image.

3.3 Ajustement Global

Alors que la mise à jour de la topologie améliore considérablement la cohérence de la mosaïque, il peut y exister dans la topologie améliorée de longs intervalles dont la moitié des images sont atteintes d'une direction et l'autre moitié de l'autre. Dans tels cas, la paire d'images se trouvant à la jonction des deux trajectoires peut être désalignée. Une approche rapide d'ajustement global est proposée pour corriger ce genre de désalignements en répartissant les erreurs de recalage d'une manière contrôlée [FBD17]. Le principe de cette approche est illustré à l'aide d'ajustement d'une séquence simulée en boucle fermée. Dans la mosaïque obtenue sans ajustement global en Fig. 3.12(a), la boucle ne ferme pas à cause d'accumulation des erreurs de recalage. L'ajustement global sert à répartir des erreurs de recalage de sorte que la forme finale de la mosaïque, tout en conservant la cohérence visuelle, soit plus conforme à la vraie surface. Les positions des points clés appairés (avec descripteurs BRISK pour cette illustration) ont été exploitées pour une modification contrôlée des homographies séquentielles $\{H^{est}_{i,i+1}\}$ entre les images consécutives I_i et I_{i+1}. Pour chaque paire (I_i, I_{i+1}), les coordonnées des points-clés appairés ($\mathbf{x}^i_D$ et $\mathbf{x}^{i+1}_D$ dans Eq. (2)) sont utilisées pour formuler une énergie dont la minimisation aboutira à un ensemble d'homographies optimisées $\{H^{opt}_{i,i+1}\}$. Pour s'assurer que les modifications des paramètres d'homographie ne provoquent pas de désalignement considérable entre les paires d'images successives, les différences entre les coordonnées des points déplacés, d'une image à l'autre, avec $\{H^{est}_{i,i+1}\}$ et celles des mêmes points déplacés avec $\{H^{opt}_{i,i+1}\}$ sont minimisées dans les deux sens (de I_i vers I_{i+1} et inversement, de I_{i+1} vers I_i, voir Eq. (2) dans laquelle le sens de déplacement est donné par l'ordre des indices i et $i+1$ dans H). Cette minimisation (Eq. (3)) est effectuée sous la contrainte donnée dans Eq. (4). La contrainte imposée assure que l'homographie directe $H^{est}_{1,N}$ entre la première et la dernière image soit égale à l'enchainement des homographies entre ces deux images.

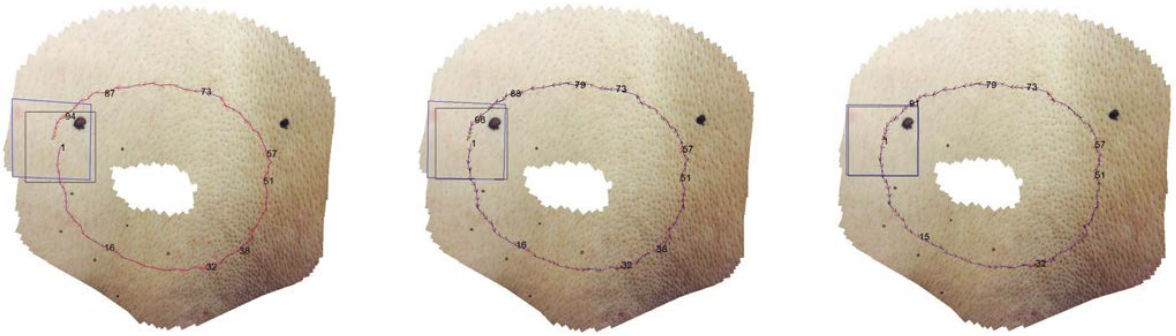
$$E_{erreur} = \underbrace{\sum_{i=1}^{N-1} \left\| H^{est}_{i,i+1}\mathbf{x}^i_D - H^{opt}_{i,i+1}\mathbf{x}^i_D \right\|}_{erreur\ de\ projection\ en\ avant} + \underbrace{\sum_{i=1}^{N-1} \left\| H^{est}_{i+1,i}\mathbf{x}^{i+1}_D - H^{opt}_{i+1,i}\mathbf{x}^{i+1}_D \right\|}_{erreur\ de\ projection\ en\ arrière} \tag{2}$$

$$\{\tilde{H}^{opt}_{1,2}, \tilde{H}^{opt}_{2,3}, ...\tilde{H}^{opt}_{N-1,N}\} = \underset{\{H^{opt}_{1,2}, H^{opt}_{2,3}, ...H^{opt}_{N-1,N}\}}{argmin} (E_{erreur}) \tag{3}$$

$$\text{Sujet à:} \quad \left\| H^{est}_{1,N} - \left(\prod_{j=1}^{N-1} H^{opt}_{j,j+1} \right) \right\| = 0, \tag{4}$$

où N est le nombre des images. Pour accélérer les calculs, au lieu de formuler une contrainte explicite, la contrainte donnée dans Eq. (4) peut être incorporée dans Eq. (2) pour formuler un problème avec une contrainte implicite. Cependant, avec une formulation implicite, la contrainte d'égalité dans Eq. (4) n'est pas strictement respectée. Les résultats avec ces deux formulations sur une séquence simulée (Fig. 2(a)) sont montrés en Fig. 5.

L'applicabilité de cette approche à une séquence acquise est montrée avec l'ajustement d'une boucle de la mosaïque de Fig. 6. La paire d'images (I_{102}, I_{103}) forme une jonction de deux longs chemins (côtés droit et gauche dans la figure 6(a)), chacun utilisé pour accéder aux images à deux côtés de cette paire. Par conséquent, ces deux images sont désalignées, comme l'indique un écart relativement important entre les images I_{102} et I_{103}. Un ajustement avec une contrainte



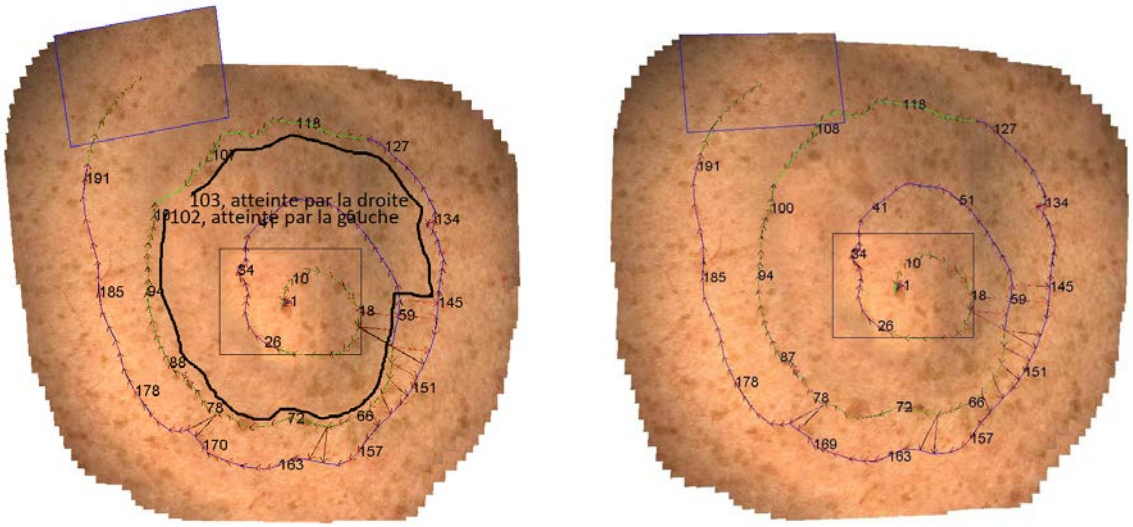
(a) Mosaïque, avant l'ajustement global, obtenue à l'aide des descripteurs BRISK. $\varepsilon^{reg}_{i,i+1}$ moyenne = 0.58 pixel $\varepsilon^{mos}_{1,N}$ 76.6 pixels.

(b) Ajustement global avec une contrainte implicite. Temps de calcul = 8.1 s $\varepsilon^{reg}_{i,i+1}$ moyenne = 0.59 pixel $\varepsilon^{mos}_{1,N}$ = 37.5 pixels.

(c) Ajustement global assujetti à une contrainte explicite. Temps de calcul = 40.0 s, $\varepsilon^{reg}_{i,i+1}$ moyenne = 0.68 pixel $\varepsilon^{mos}_{1,N}$ = 10.8 pixels.

Figure 5: Mosaïque de la séquence simulée montrée en Fig. 2(a), avant et après l'ajustement en utilisant deux modèles d'énergie. La séquence contient 101 images, chacune de taille 512 ×512 pixels (représentant environ 2×2 cm ^{2} de la surface cutanée). Le carré noir représente la première image et le quadrangle bleu la dernière. La ligne rouge indique la trajectoire de mosaïquage.

explicite a été effectué sur la boucle fermée $\{I_{102}, ..., I_{58}, I_{145}, ..., I_{103}\}$ (la boucle concernée est indiquée par la ligne rouge en gras) avec la contrainte que l'homographie accumulée le long de ce chemin soit égale à $H^{est}_{102,103}$. Les images concernées sont à peu près alignées, avec une perturbation minimale dans le reste de la boucle, après l'ajustement.



(a) Mosaïque avant l'ajustement global.

(b) Mosaïque après l'ajustement d'une boucle.

Figure 6: Application de l'ajustement global sur une séquence réelle. La boucle ajustée est formée par le chemin $\{I_{102}, ..., I_{58}, I_{145}, ..., I_{103}\}$. Les erreurs d'homographie ont été redistribuées sur ce chemin sous la contrainte que l'homographie accumulée le long de ce chemin soit égale à $H^{est}_{102,103}$.

General Introduction

Context of this Ph.D. Thesis

This work is part of the collaborative "InnovaTICs" R&D project, that is carried out with funding from "fonds européen de développement régional" (FEDER) and région Lorraine, in the field of teledermatology. This study was performed at CRAN (Centre de Recherche en Automatique de Nancy) laboratory, a joint research unit between the Université de Lorraine and the centre national de la recherche scientifique (CNRS). The objective is to provide the dermatologists with a tool facilitating a teledermatological consultation for patients who have reduced mobility or live in areas far from the urban centers. Although teledermatology is already in use and its economic advantages have been demonstrated [Mas+09], its use faces some limitations: commonly used acquisition devices, such as smartphones and portable cameras, provide images without standardized colorimetric rendering and with a resolution that depends on the acquisition distance, which, with the increase in the affected skin region, needs to be increased for capturing the entire region of interest. Since dermatologists need to observe not only the complete area of the affected skin but also the healthy peripheral areas to make a thorough diagnosis, the objective of this project is to overcome these limitations by constructing highly resolved panoramic images of extended skin surfaces, through a well-adapted image mosaicing approach, from the video sequences. An innovative device developed within the InnovaTICs Dépendance project [Amo+17b] was used for colorimetrically correct rendering of the acquired image sequences.

Scientific Context

Image mosaicing from a set of images of the same scene or from the images composing a video sequence consists in placing these images in a single reference coordinate frame such that their common parts are superimposed. A brief description of the main mosaicing steps given hereafter highlights the scientific problems whose solution is sought in the skin image cartography framework:

The planarity assumption is made in the use of the *image registration approaches*. In dermatology, individual images can be assumed to be planar if the skin video sequences are acquired with a small field of view in close proximity to the surface. These acquisition conditions permit the use of a 2D mosaicing scheme, in which the images are placed in the mosaicing plane by superimposing their common parts. The estimation of the common parts consists in the determination of the geometric relation between the pairs of consecutive images (I_i, I_{i+1}), i.e. estimation of homography parameters which would give the best correspondence between the homologous points in the two images. The consecutive images are registered to determine the optimal parameters of the homography matrix $H_{i,i+1}^{est}$. These parameters are: the scaling factors (f_x, f_y), the $2D$ rotation ϕ, the shear factors (s_x, s_y), the $2D$ translations (t_x, t_y) and the perspective

change factors (h_1, h_2). The coordinates of the pixels of the source image I_{i+1}, transformed with this matrix, must ideally be displaced in the coordinates of the homologous pixels in the target image I_i. $H_{i,i+1}^{est}$ projects $\mathbf{x}_{i+1}$, in homogeneous coordinates, in I_{i+1} onto its homologous point $\mathbf{x}_i$ in I_i. As detailed in this thesis, the robust determination of accurate point pairs is a crucial step in the homographic parameter determination.

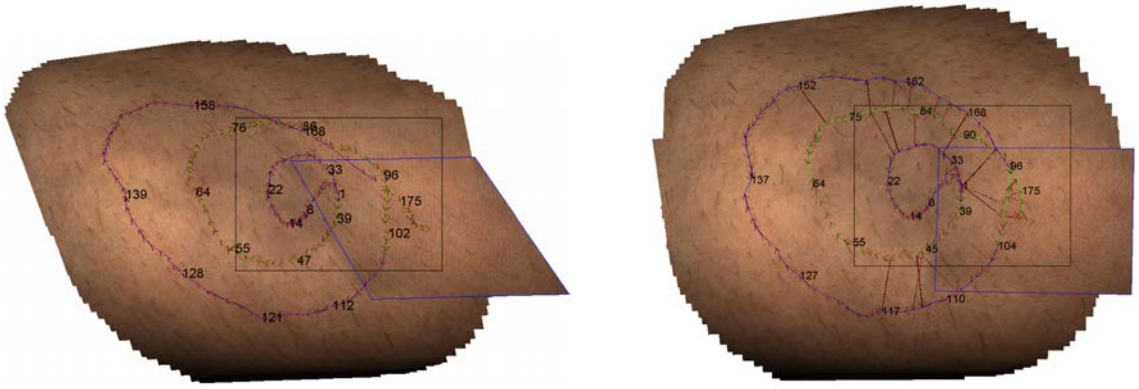
In the mosaicing process starting with the *stitching* step, the warped images of a sequence are placed in a common coordinate system using the geometrical transformations that relate them to a reference image. Some overlap between the images is necessary to calculate the homographic relation between them. The homographies of the overlapping image pairs are concatenated to calculate the placement of the images that do not have a significant overlap with the reference image. However, for long sequences, the concatenation along the acquisition path can generate a considerable accumulation of small errors. Acquisitions covering a large area may contain crossing trajectories. When the *image trajectory topology* is taken into account, this error can be minimized by choosing alternative paths through crossings of the sequential trajectory to reduce the number of homographies to be concatenated to reach images that are far from the reference image. In [FSV01] and [MFM04], the topology of the paths of the image centers was considered to create extended mosaics. Moreover, a mosaic that considers only trajectories along the sequential images assumes continuity in the sequence. However, variations in lighting or blurring due to a rapid movement of the camera may cause a failure in the registration, which interrupts the mosaicing process. In this case, the calculation of the topology of the image trajectory makes it possible to find alternative paths for the images located after this interruption. In this thesis, it is discussed how such image trajectory topology can be adapted to dermatological video sequences.

While the choice of the most direct trajectories can reduce the accumulation of errors, some misalignment may persist in the final mosaic due to the limited accuracy of the registration process and topology based image placement correction. Assuming that the registration errors are not significant after calculating the sequential homographies, the misalignment in the mosaic can be sufficiently corrected by using various *global adjustment* approaches. For this, in [MFM04; Wei+12b], the authors have used a predefined grid on the mosaicing plane. Each grid point was projected according to two homographies, each concatenated along a different path that links the reference image to the image on which that point lies. The difference between the positions of these two projections, for all the grid points, was iteratively minimized. However, this solution remains costly in terms of calculations. Miranda et al. [ML+08] have proposed a faster approach to obtain a correct alignment of the first and last images in the case of a closed-loop acquisition trajectory. Their approach consisted in adjusting the sequential homographies in a controlled manner so that the difference between the direct homography relating the first and the last images and the concatenated homography along the loop was minimized. This approach allows for a fast adjustment but may significantly affect the registration accuracy of the consecutive image pairs because, apart from homographies, no iconic information is considered in the adjustment process. After *the global adjustment*, the intensity discontinuities at the edges of the images can be smoothed using different *blending* approaches to achieve a visually coherent rendering of the final mosaic [Sze06; Wei+12a]. Blending consists in modifying the pixel intensities to obtain a smooth transition between the intensities of the images that make up the whole mosaic. A basic blending approach that proved to be useful for skin image mosaicing was adapted in this thesis.

Main Contributions of the Thesis

The performance of some image registration approaches will be compared for the optimal choice of a registration scheme. Some of such approaches in different frameworks will be considered to find the best compromise between computation time and accuracy. Among these are Chambolle and Pock's [CP11] approach, which calculates the optical flow in a total variational framework, Large Displacement Optical Flow (LDOF) [BM11], an approach that combines the descriptor matching with optical flow, and Brisk [LCS11], SIFT [Low04] and SURF [BTG08], three approaches based on invariant keypoint descriptors. Apart from selecting the best existing point matching approach for homography determination, a scheme for the refinement of matched keypoint correspondence will be presented.

In the most direct mosaicing approach, the homographic relation of an image with the reference image is calculated by concatenating the homographies of the image pairs along the acquisition trajectory. With the concatenation of a large number of homographies, the mosaic can undergo a significant deformation due to error accumulation. In addition, the failure of a single registration prevents the placement of the images that are located after the image pair at which the failure occurs. To overcome these limitations, a dynamic estimation of the sequence topology (i.e. the spatial links between the neighboring images) will be sought. This will allow for the determination of additional paths, through non-sequential image pairs, to reach a given image from the reference image. In [VRD13], a spiral acquisition scheme was used for mosaicing endomicroscopic video sequences. This mode of acquisition will serve as a model for mosaicing the skin surface. By detecting the pairs of images whose registration failed (resulting in a false homography) and considering a spiral acquisition path, this approach potentially offers the possibility of mosaicing sequences that contain some "mismatched" homographies. Radial



(a) Mosaic obtained by finding the "short-circuited" paths.

(b) Mosaic obtained by concatenating the homographies over more direct paths by enforcing the radial links, if possible, every 30 degrees.

Figure 1: Illustration, with the mosaicing of a simulated sequence, of the interest of the most direct trajectories, obtained through the angle-based scheme, to minimize the accumulation of errors. The black and blue quadrangles frame the first and last frame of the sequence, respectively. The sequential trajectory is marked in green and blue, the change of color indicates a change in the spiral branch. The red lines, with arrows indicating the direction, represent the most direct trajectories. The positions in the sequence of a few images are illustrated by the number of the image placed in the mosaic at the center of the image.

links between the spiral rings can be found either by using some existing approaches [Wei+12b; MFM04] or by determining the topology through the angle between the images at consecutive rings and the reference image that is taken at the center of the spiral. This can permit while limiting the accumulation of errors, the placement of the most distant images in the acquisition trajectory onto the mosaicing plane (see illustration in Fig. 1). In addition, a fast global adjustment scheme applicable to the mosaicing of skin image sequences acquired under certain constraints so as to form crossing trajectories will be presented.

Personal Scientific Production

- K. Faraz, W. Blondel, M. Amouroux and C. Daul, "Towards skin image mosaicing." In proceedings of the sixth International Conference on Image Processing Theory, Tools and Applications (IPTA'16), Oulu, Finland, December 2016.
- K. Faraz, W. Blondel and C. Daul, "Global Adjustment for Creating Extended Panoramic Images in Video-dermoscopy.", In proceedings of European Conferences on Biomedical Optics (ECBO'17), Munich, Germany, June 2017.

The content of this dissertation is distributed as following:

Chapter 1: Medical Context and Mosaicing Overview. The medical context of the project undertaken in this PhD thesis is presented. A literature review of the teledermatological approaches already in use is given by highlighting their evolution over the last couple of decades along with presenting the advantages they offer and the limitations that they face. After an overview of different aspects of mosaicing pipeline, the main objectives of the PhD thesis are presented here.

Chapter 2: Analysis of Registration Schemes Applicable to Skin Image Mosaicing. After a description of existing works concerning skin image registration, several registration approaches based on descriptor-matching, optical flow methods or a combination thereof are studied, and their performances are compared to select the optimum approach for this project. The results on simulated as well as real sequences are given. Moreover, a contribution concerning the refinement of the point correspondence established through feature-based approaches is presented.

Chapter 3: Skin Image Mosaicing with Topological Inference. The mosaicing of cutaneous surfaces is considered as a whole with a focus on a coherent final rendering of the mosaic that is more conform to the real surface. A mosaicing approach that takes into account the topology of the video sequence acquisition path is described in combination with a global adjustment approach that takes into account the descriptor positions. The effectiveness of the approach is demonstrated through various simulated sequences and its application is demonstrated through mosaicing of some real sequences.

Chapter 1

Medical Context and Mosaicing Pipeline Overview

1.1 Introduction and Project Background

The PhD work presented in this dissertation is part of a collaborative project named "InnovaTICs-Dependance" winner of the "Health Technology" call 2013 from the "Agence de Mobilisation Economique" (AME) of Regional Council of Lorraine (French "Région Lorraine"). It involves 4 partners: one laboratory (CRAN), two enterprises (SEFAM at Villers-Lès-Nancy and SD-Innovation at Frouard, France) and one Regional Hosting and Ressource Center for companies (CAREP, Centre d'Accueil et de Ressources pour Entreprises du Pays Val de Lorraine). Fully co-funded, including the present PhD grant, by the "Région Lorraine" and the European Regional Development Fund (ERDF, or FEDER in French acronym), this project aims at developing medical devices for teleobservance and telediagnosis as well as structuring a network for telemedical follow-up of dependent people. In this framework, the research work of our team in CRAN concerns the development of a multimodality panoramic reflectance imaging system for skin lesions (wounds) tele-expertise including technical innovations such as bimodal video sequence acquisitions (white light and 2 wavelengths light diffusion/absorption), hand-held and communicating facilities, and addressing scientific challenges for robust and precise automatic image mosaicing (subject of the present PhD work). The objective of the project is to provide the dermatologists with a tool for more effective telediagnosis of the skin conditions. Skin lesion tele-expertise makes easier the dermatologists' consultations for the patients located in remote areas or the ones with reduced mobility. Although teledermatology is already in use and its economic advantage has been demonstrated [Mas+09; War+11], the acquisition devices currently used, such as consumer devices (smartphone, digital tablet, digital cameras) and Medical Devices MD (handheld dermoscopes, video-dermoscopes), give limited or even poor efficiency in terms of field of view (MD are intrinsically developed to image small areas $\sim$4 cm^2) and in terms of color rendering, because of uncontrolled illumination and color calibration [Amo+17a]. Chronic wounds may require taking pictures of large skin areas (up to 100 cm^2) without any contact because dermatologists need to have in their sight the complete affected skin zone along with the surrounding area to be able to do a thorough diagnosis of the extended lesions (Fig. 1.1). The goal of this project is to overcome these limitations with a well-adapted mosaicing scheme which increases the field of view of the skin area filmed through a specific medical device, with controlled colorimetry [Amo+15].

(a)

(b)

Figure 1.1: Two examples of extended skin wound lesions.

1.2 Different Settings for Diagnosis of Skin Conditions

Skin, the largest organ of the human body, is susceptible to suffer from various pathoses such as melanoma, ulcer or even simple bruising or lesions. Being entirely exposed to the environment, the skin, on the bright side, accommodates a wide range of non-invasive diagnosis techniques. As Bristow and Bowling in [BB09] presented their concerns, skin cancer accounts for about one-third of all the cancers. Since there are no definitive treatments for cancer, the surveillance for detection at an early stage remains the most suitable approach for its prevention or control. For this to happen, there are two conditions that need to be met widely. Without speaking of the diagnosis methods, first and foremost condition is the awareness, among the general population, of the skin conditions. Since the skin is not perceived as a vital organ, there is a tendency to ignore its conditions. According to [BB09], although more and more people seek diagnosis at early stages upon detection of lesions that might be cancerous, education and age are the key factors that play a role in determining the prevalence of early reporting. People with old age and low education tend to report in later stages. The second condition is the availability and ease of access to specialized medical facilities. The most widely used practice requires patients to present themselves at a clinical establishment where the diagnosis is carried out through direct observation and/or with the help of diagnostic tools by the dermatologists or, in some cases, by the skin health care specialists. With the improving technologies of portable imaging devices and advances in telecommunications, telediagnosis also has found its place in the dermatology domain. Some of the diagnostic techniques, in both these cases, are presented in the next subsection under their respective classifications according to their application setup. Moreover, some diagnostic approaches that are in developmental stages are also discussed.

1.2.1 Dermoscopy in a clinical setup

Dermoscopy consists in examining the skin surface through a magnifier. When a non-polarized light source is used, the skin is examined by bringing the dermoscope in contact with the skin through a liquid medium to avoid surface reflections. The liquid medium can be omitted if a polarized light source is used. This simple device helps the dermatologists to distinguish between benign and malignant tumors/lesions. Compared with the naked eye examination, a dermoscope highly improves the correct classification of skin conditions when examined by a

trained dermatologist [Ves+08]. The application of dermoscopy for early detection of malignant melanoma in the foot is demonstrated in [BB09] and the general importance of dermoscopy in early detection of melanoma is presented in [PHP11]. Being a low-cost device, it may easily be made available to the general practitioners as well, and its generalized use for self-reporting is being encouraged, as discussed in the next subsection.

1.2.2 Dermoscopy/dermatology for telediagnosis

Teledermatology and telemedicine, in general, is the use of telecommunication devices for medical/clinical consultation/expertise at distance. In line with the progress in the digital communication techniques, an increased interest in teledermatology began in the 1990s [PB95], especially to allow patients in rural locations to access a dermatologist. The affected skin areas could be photographed through portable digital cameras. The resulting images could then be transferred through a telephone-modem to the dermatologists, along with the patient information, for further analysis. The concerned dermatologists were trained to diagnose the skin conditions from the photographs. In the 2000s, with further improvement in the digital communication technologies, a larger bandwidth was accessible to the general public. This allowed for video conferences for a telediagnosis. However, although this provided certain advantages over the traditional consultations from the patient's point of view, the economic factors and general cost-effectiveness vs effective-evaluation measures needed to be taken into account. For the patient, it is advantageous to have an instantaneous consultation, but, at the same time, the dermatologist also needs to have a clear schedule. An alternative is the store-and-forward approach, where the data (images, videos, patient information) are transferred to the dermatologist, who can consult them off-line at their convenience.

In 2001, Eedy and Wootton [EW01] published a survey on the use of teledermatological approaches. This survey highlighted the advantages and disadvantages of different telediagnosis approaches, including video-conferencing and off-line image examination, and the general level of satisfaction, both for the health practitioner and the patient. They summarized that the teleconsultations, in general, could be planned as effectively as the conventional consultations, while providing services in the remote areas. The acceptability from the patients made it feasible and the general practitioners had more flexibility in further referrals for specialized consultations. In addition to saving travel costs for the patients, the dermatologists also saved time that they would otherwise have spent visiting the remote areas. Improved equipment, at low costs, for high-quality images, also contributed towards economic benefits. It was also advantageous for the patients who preferred conventional consultations as they had shorter waiting lists. They also highlighted some downsides of the teleconsultations. With the principal focus being on the lesion, this approach reduced the personal patient-doctor interaction. Besides, not all skin conditions could be telediagnosed and there was some uncertainty in the diagnosis. Moreover, a minority, both among the patients and the health practitioners, preferred the traditional consultations. The privacy concern, that could lead to legal liability, was found to be an important concern in telediagnosis. Comparison of video-conferencing and store-and-forward approach showed that the former provided more patient information and facilitated the interaction between multiple health practitioners. However, at the same time, video-conferencing provided lower image quality and was complicated to plan when several dermatologists were involved in the diagnosis. The store-and-forward approach produced superior quality images that not only facilitated the patient follow-up by monitoring the evolution over time of the skin conditions but were also helpful for determining whether some patients needed hospitalization or not. On the other hand, the interaction with the patient as well as among the health practitioners was limited, apart from

requiring careful management so as not to mix-up the data from multiple patients.

In [Tan+10], a study was performed to assess the effectiveness of teledermatology in patient triage. Two hundred patients, with a total of 491 lesions were involved in this study. Images of the lesions were taken by an expert photographer. For acquiring the dermoscopic images, the camera was in contact with the skin through a fluid that reduced the glare. The images obtained were of 1600×1200 pixel in size with 24-bit pixel depth. The patients involved were initially diagnosed, during the direct consultation, by two dermatologists. After a four-week period, the same dermatologists were presented with the previously acquired photographs to make a diagnosis. In the two diagnoses, only 12.3 % lesions were diagnosed differently, with the severity of 12 lesions being under-reported in the telediagnosis. Among the under-reported lesions, only one turned out to be malignant after the histopathology results came in. In this case, solar keratosis was diagnosed instead of what later turned out to be a basal cell carcinoma.

In the late 2000s, smartphones and personal digital assistants equipped with cameras appeared in the consumer market. This, along with the increased bandwidth in the wireless communication, opened the possibilities for mobile teledermatology. Masson et al. [Mas+09], in 2009, reported on the effectiveness of such approach for screening of skin neoplasms and other skin conditions. The resolution of smartphone cameras, at that point, was high enough to enable an effective diagnosis from the images acquired through them. They provided the additional advantage of reaching out the patients in remote areas where internet connection was not available. 91 % of the diagnoses made through this approach corresponded to the ones made during a face-to-face consultation. Telediagnosis of melanocytic skin neoplasms was reported to have 83 % accuracy when compared with histopathologic approaches. Masson et al. [Mas+14], in 2013, published the results of their study performed over 690 patients with the goal of skin cancer prevention. This study sought to perform a triage scheme based on teletransmitted skin images. The dermatologists had to determine, from these images, whether the lesions required further followup, or were to be excised or if an in-person diagnosis was to be made. The patients involved in this study had local consultations with general practitioners who had no special training in dermatology. Based on their assessment, the general practitioner advised for either a regular camera image or an image through polarized light contact dermoscope. A total of 962 dermoscopic and 123 clinical (camera-based) images were acquired for 221 patients whose lesions were evaluated as suspicious. These images were sent to two dermatologists for teleconsultations. They classified 88 % of the dermoscopic images as of excellent quality compared to only 77 % of the clinical images. A few images, in both modalities, were, considered of low quality and no definite analysis could be established in only three cases where the dermoscopic images were of poor quality. All the malignant and benign lesions were correctly classified through teleconsultation, with a diagnostic accuracy of 94 %. Misclassification in the diagnoses was noticeable in the case of dysplastic nevi: one case was incorrectly classified as seborrheic keratosis while another as a basal cell carcinoma (a misclassification like the second case is also reported in Tan [Tan+10], as presented in the previous paragraph).

In [WGN15], a study assessing the reliability of skin neoplasms diagnosis through teledermatology was presented. This study, which involved around two thousand patients having a total of about three thousand lesions, consisted in comparing the in-place diagnoses and telediagnoses by two different dermatologists. Images used were obtained either through a white light camera or through a dermoscope. Contact dermoscopic images were also obtained. Agreement on diagnosis ranged from moderate to almost perfect, with the agreement being slightly better when dermoscopic images were used. Also, contact dermoscopy proved to be a better tool for telediagnosis of pigmented lesions. Nami et al. [Nam+15], in 2015, published a study that detailed the working of a web-based store-and-forward approach. The concerned dermatology website, acces-

sible through a smartphone application, provided a secure storage of patient data and images. The patients involved in this study were the ones who were reporting (at their own initiative or through a general practitioner's referral) a skin disorder for the first time. Images, taken through a smartphone by a student enrolled in a course for general medical practice, involved a zoom on the affected area or a zoomed out view that placed the concerned area at the center. Using face-to-face diagnosis as a reference, 356 out of a total of 391 patients were correctly diagnosed by a teledermatologist.

In [Bö+15], the effectiveness of using smartphone referrals, in comparison with the traditional paper-based referrals by the general practitioners, for triage of skin cancer patients was studied. Two sets of patients, consulted in 20 health care centers, were involved in this study. A total of 816 patients were referred to the dermatological centers through a smartphone application that involved taking images through a compatible dermoscope. For another 746, paper-based referrals were used. It was found that the patients with melanoma and some kinds of carcinoma needing surgical procedure required lesser waiting time and were more manageable when referred through teledermatology application. The teledermatology referrals were more reliable and could save the in-person visit for many patients. Only four teledermoscopic referrals could not be processed due to poor image quality.

The above study by Börve et al. [Bö+15] was performed in Sweden. In a correspondence concerning this study [LJH15], Leitch et al., after presenting some difficulties for implementing such scheme in Scotland, asked for some specifications. Their concerns were specifically about extensive delays from referral to treatment (62 days) and they were interested in innovations that could reduce this delay. Smartphones being commonplace and teledermatology being already used in the UK, technological innovation did not pose a problem since it simply required involving a medical photographer. However, they suggested that the reduction in this delay as reported in [Bö+15] could have been due to the severity of the malignant tumors of the patients in the teledermatology group. Furthermore, they pointed out that the reported shorter delays could simply be due to the longer delivery time of the paper referrals that could have been reduced by using electronic referrals. They looked for clarification on how the teledermatology referrals reduced the waiting time for surgical intervention. Referring to the smartphone application used in the study by Börve et al., they wondered if the reduced number of visits for the patients referred through this system was because this application did not have an inherent option for managing records of multiple lesions for one patient. Consequently, the people with multiple lesions being grouped, by default, in the paper referral group requiring multiple visits (as opposed to the ones referred through the smartphone application). Also, they wondered if the additional benefits in teledermatology were due to the fact that a dermoscopic image provided additional benefit over a clinical image for diagnostic purposes. Moreover, they showed their concerns, in what regards to an opportunity for further education, about the patients with benign tumors who, after a telediagnosis, were seen as not requiring a face-to-face visit.

Responding to the concerns of Leitch et al., the authors of [Bö+15] specified that their approach did not involve an intermediary medical photographer (which would cost extra time). Regarding the potential patient-bias, they agreed that randomization in patient-selection would provide more optimal results, but this, though considered, was not carried out in their study due to increased time and costs. For the reduced waiting time in surgical intervention, they confirmed Leitch et al.'s assumption that teledermoscopic referrals provided a greater insight into the lesion, thus opening a possibility of "pre-booking" a window for the procedure. Regarding the inability of their application to register and transfer multiple images, they confirmed the technical limitations, due to limited WiFi coverage, at the time of the study and agreed with Leitch et al.'s suggestion that this can be circumvented through imaging the most suspicious lesion for the

patients with multiple lesions. To show the impact of improved diagnosis through dermoscopic images, although in their study they did not publish a direct comparison–this not being their objective–they supported their approach through another study [WGN15] (aforementioned) that compares the two modalities. Regarding Leitch et al.'s concern about the patient missing the chance of self-education, they highlighted that the general practitioner received almost immediate feedback from the teledermatologists. This not only improved the general practitioner's diagnostic skills, but, following the former's recommendations, the patient could greatly be informed of the importance of self-care and follow-up procedures. In addition, they pointed out the emerging consumer-oriented applications for self-awareness.

The issue of patient education was raised in the correspondence [LJH15] overviewed above. In 2015, Wu et al. [Wu+15] published a study evaluating the effectiveness of patient-initiated mobile teledermoscopy. They studied the feasibility and effectiveness, along with patient receptivity, of such approaches for followup, over a brief period, of clinically atypical nevi. Out of 34 patients initially involved, 29 completed the study. Images were obtained in two settings. One involved dermoscopic images taken by a dermatologist in an office setting. In the other case, the patients themselves took the images of the concerned regions through a mobile dermoscope attached to a smartphone. The smartphone images were sent, over the internet, to another dermatologist who monitored the evolution of the lesions. After the initial baseline images, follow-up images were taken after 3 to 4 months. The diagnoses by the two dermatologists were compared. To determine the feasibility, the patients' level of ease with acquiring the images was observed. 28 of the 29 patients were able to acquire at least one image that was both in the frame and in focus. The effectiveness was evaluated through concordance in the diagnosis by the two dermatologists. An overall 97 % concordance was found in the 30 image pairs. The lesions for which the discordance occurred, an evolution of the lesion was noticed in the clinic-based examination. However, the teledematologist classified it as involving no-change. This was attributed to the lower image quality, where smaller changes are difficult to observe. For evaluating the patient receptivity, they were surveyed about their confidence in the use of the mobile dermoscope. The patients showed a general confidence in the technology and were willing to pay between $20 and $500 (with a median of $100) for a personal mobile dermoscope. They were generally satisfied with the procedure, particularly due to short waiting time. On the other hand, the principal barrier that could be deterrent in its use was the inability to see the dermatologist and a few patients expressed their concerns about cost and insurance reimbursement. A more exhaustive study of the consumer acceptance of teledermoscopy among the patients with elevated risk of melanoma was presented in [Hor+16]. The majority of the 228 participants, in their fifties or sixties, who were involved in their initial survey found teledermatology of interest. However, about half of them had some misgivings about telediagnosis. Among the patients who participated in the teledermatological process, though the majority indicated that they found it convenient to use, about one fifth were unable to photograph areas of their skin that were difficult to reach. Moreover, about one third required help for submitting the photographs. The authors concluded that teledermatology's acceptance was favorable and that further assessment in this field for early detection of melanoma was warranted.

1.2.3 Skin diagnosis methods in developmental stage

Some approaches for diagnosing skin diseases, though not readily available in clinical settings, have been tested in experimental setups. A few among these are optical coherence tomography [Dal+14], hyperspectral imaging [YNP10], thermal imaging [Rin10] and multiphoton excitation microscopy [MS01]. Two of these approaches are briefly reviewed hereafter.

1.2.3.1 Hyperspectral imaging

Hyperspectral imaging is used for examining the evolution of diabetic foot ulcers in [YNP10]. Hyperspectral imaging consists in obtaining images of the same region with light lying in different wavelengths. Melanin concentration, epidermal thickness, blood concentration in the dermis and dispersion properties of the skin are among the factors affecting the absorption of the incident radiation with a given wavelength λ_j. Hemoglobin absorbs more radiation in the visible light spectrum, whereas the melanin does so in the UV spectrum. This difference in absorption by the chromophores forms the basis of the diagnosis by Hyperspectral imaging. The absorption spectra of oxyhemoglobin and deoxyhemoglobin must be deconvoluted from the spectra of other chromophores to achieve true concentrations of each of the chromophores.

A method to calculate the concentrations of oxyhemoglobin, deoxyhemoglobin and melanin is to use a modified form of the Beer-Lambert law [SF04]:

$$A_{abs}(x, y, \lambda_j) = -\log_{10}[R_d(x, y, \lambda_j)], \tag{1.1}$$

where A_{abs} is the radiation absorbed at coordinates (x, y) at wavelength λ_j and R_d is the normal diffuse reflection. Taking into account the absorption by different chromophores, the total absorption can be written as:

$$\begin{aligned} A_{total}(x, y, \lambda_j) = \varepsilon_{oxy}(\lambda_j) M_{oxy}(x, y) L + \\ \varepsilon_{deoxy}(\lambda_j) M_{deoxy}(x, y) L + \varepsilon_{mel}(\lambda_j) M_{mel}(x, y) L + G, \end{aligned} \tag{1.2}$$

where M is the molar concentration of the chromophore under consideration, ε is the molar absorption coefficient and L is the average distance traveled in the skin by the photon. G is the scattered light which does not reach the sensor and remains independent of λ_j.

If only the effective concentrations of the chromophores are considered, they can be found by minimizing the residue defined in the following equation:

$$r = \sum_{j=1}^{15} (A_{total}(x, y, \lambda_j) - s[OXY(\lambda_j) + DEOXY(\lambda_j) + MEL(\lambda_j)] + G), \tag{1.3}$$

where OXY, $DEOXY$ and MEL represent concentrations of oxyhemoglobin, deoxyhemoglobin and melanin respectively recorded at wavelength λ_j. s is a scale factor determined empirically to be 50 for healthy patients.

Following this analysis, one cannot find the thickness of the epidermis. Yudovsky et al. [YNP10] proposed a method which, from the same reflection data, can simultaneously calculate the concentration of oxygen, the blood volume fraction, the concentration of melanin and the dispersion coefficient of the epidermis and dermis. For this, they have modeled the skin after a plate with parallel planes. Nouvong et al. [Nou+09] also used hyperspectral imaging to study the evolution of diabetic foot ulcers.

1.2.3.2 With polarized light

The light loses its polarization as it disperses into the tissues. Therefore, when polarized light is reflected by the skin, the degree of alteration in the polarization may be associated with the tissue composition. When polarized light is incident on the skin, three types of light emerging from the skin can be distinguished: the one reflected by the air/skin interface (which does not affect the

polarization), the light resulting from diffuse reflection (which results in a random polarization) and the light scattered by the surface tissues (which results in polarization according to the birefringence of the tissues). The light reflected in the third case is important because it can reveal the structure of surface tissues, where skin cancer occurs.

An analytic linear polarizer permits analysis of the light in terms of its source of origin [And91]. When the analyzer is oriented in parallel with the polarization direction, most of the diffuse reflection is filtered and, as a result, the light reflected by the dermal surface can be studied. On the other hand, when the orientation is orthogonal to the polarization filter, it has the effect of filtering out the specular reflection and up to 50 % of the diffuse reflection. Since the light reflected after diffusion is attenuated by blood vessels, hemoglobin and superficial tissues, it permits visualization of the structure of the papillary dermis and the epidermis.

Jacques et al [JRRL02] proposed a modification of the above approach to study the structure of the papillary dermis and superficial reticular dermis in more detail. They avoided capturing the specular reflection by placing the light source and the receptor such that the latter received only that part of the light that resulted from the diffuse reflection or was reflected after dispersion by the surface layers of the skin. The light scattered by the dermal subsurface remained polarized. This light permitted acquisition of the image I_{par} by orienting the analyzer parallel to the orientation of the polarization of the incident light. A second image I_{per} was acquired by orienting the analyzer orthogonal to the incident polarization. The two images were combined algebraically to contain most of the light scattered by the subsurface tissues. The resulting image accentuated the subsurface tissue structures. For the data acquisition, the part of the white light above a threshold of 500 nm wavelength was filtered and then polarized by a linear polarizer. The wavelength was not considered important in this case because the depth of the photographed tissue depends on its birefringence, which depolarizes the light. The collimated light was projected onto the surface at an angle of 15° with respect to the normal. A glass plate was used to apply pressure to whiten the blood vessels in the skin. The linear analyzing polarizer placed in front of the camera allowed to obtain two images according to the method described above. The acquired images were of 600×600 pixels dimensions, with 34×34 microns pixel size.

I_{par} is the sum of the light dispersed by the surface tissues (R_s) and half of the diffused light (R_d). The light received by the sensor is proportional to the incident light (I_0) attenuated by the light absorption by melanin (T_{mel}):

$$I_{par} = I_0 T_{mel}(R_s + 0.5R_d) \tag{1.4}$$

I_{per} contains 50 % of the light resulting from diffuse reflection:

$$I_{per} = 0.5 I_0 T_{mel} R_d \tag{1.5}$$

The polarization ratio (Pol) is defined as:

$$Pol = \frac{I_{par} - I_{per}}{I_{par} + I_{per}} = \frac{R_s}{R_s + R_d} \tag{1.6}$$

This ratio is independent of the intensity of the incident light and the light absorbed by melanin. The image represented by Pol accentuates the surface tissue structures.

Since the time for integration of an image by the sensor is not negligible (1 s), body movement caused by global or local muscle contractions may result in a shift between the two images. This offset results in undesired artifacts. In addition, air-balls caught between the glass plate and the skin can cause blurred areas in the image.

The effectiveness of the optical methods for diagnosis of skin conditions is supported by a vast body of literature. Moreover, the dermatologists are able to make a telediagnosis with high certainty. Some limitations in telediagnosis are posed by low-quality images or improper illumination conditions during image acquisition. A telediagnosis from dermoscopic images was more precise compared to the one from regular camera images, as shown in [WGN15]. Another way of facilitating the telediagnosis, especially for extended lesions in ulcerology–which is not as much treated as cancerology in the extant literature on teledermatology–is by obtaining extended high-resolution panoramic images of the concerned region. The following subsection overviews main steps of the mosaicing process for obtaining panoramic images.

1.3 Mosaicing Overview

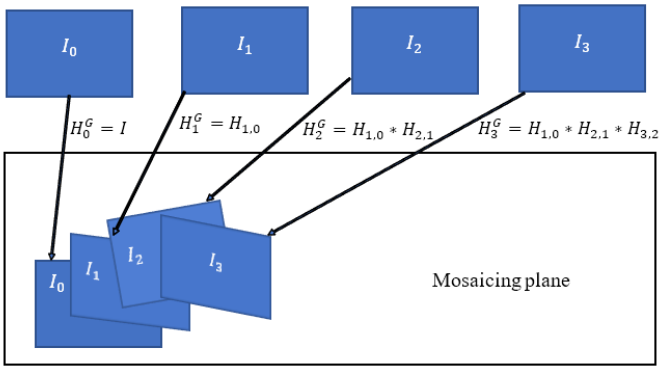


Figure 1.2: An illustration of the mosaicing process.

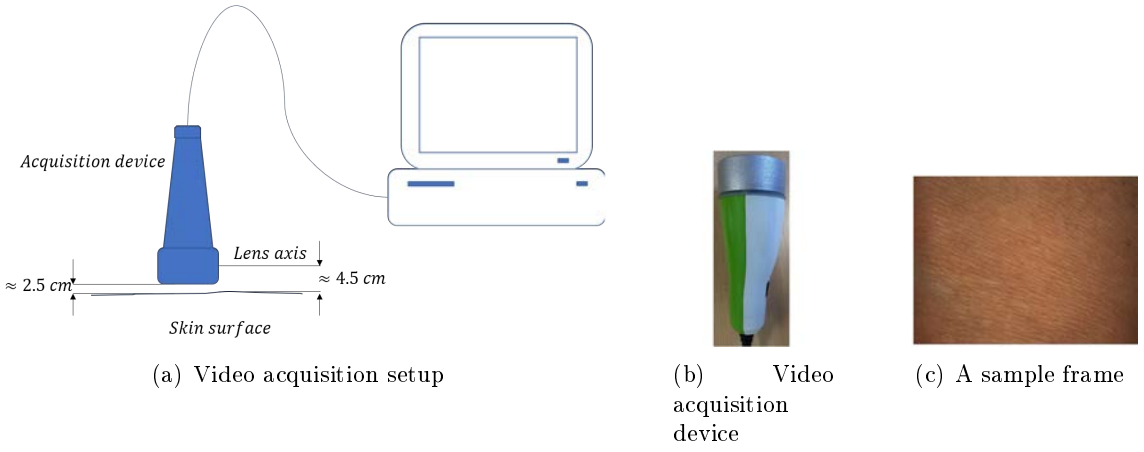
Image mosaicing consists in placing different images of the same scene or the frames of a video sequence on a surface such that their common parts overlap (Fig.1.2). Image mosaicing has been used since the early 1980s to create extended fields of view from aerial images [Jon82]. The interest in creating highly resolved panoramic views by mosaicing the images obtained through hand-held cameras emerged in the mid 1990s [SHK98]. All the mosaicing techniques require alignment of the image pairs. In the early years of mosaicing, dense pixel-based approaches were used for this purpose. At the dawn of the 21st century, the use of specific features saw its increase in the mosaicing process. This allowed not only for a quick alignment of the images, but also permitted "recognizing" the panoramas by matching similar features in a set of non-sequenced [BL03].

As illustrated in Fig. 1.4, the initial stage of mosaicing is the alignment of consecutive image pairs in a set of images or in a video sequence. This involves two key steps. One is choosing an appropriate mathematical model that would provide the corresponding coordinates of homologous points in the two images (image registration). The other is the determination of the geometrical relation that links these two sets of coordinates. This relationship is then used for placing all the images into a common coordinate mosaicing plane. In Fig. 1.4, $H_{i,i-1}$ is a matrix that superimposes the pixels of image I_i on its homologous pixels in I_{i-1} through a geometric transform. An image is chosen as a reference and the positions of the other images with respect to this image are calculated by concatenating the geometric relations over a path leading up to that image. H_i^G is the "global" matrix placing image I_i in the coordinate system of image I_0, whose origin coincides with the mosaic plane origin. H_i^G is computed by concatenating the

homographies of the image pairs over the path from I_0 to I_i:

$$H_i^G = \prod_{j=0}^{i} H_{j,j-1} \tag{1.7}$$

When there is a large number of images to be mosaiced, the global coherency of the image alignment needs to be taken into account because, due to the limited precision of the image registration methods, the error accumulation when placing images I_i into the mosaic produces perceptible distortions or misalignment in the final mosaic. For this reason, the mosaicing process usually involves either a global alignment and/or the estimation of the camera trajectory to find shorter paths between I_i and I_0 by considering the topology of the images. Taking the image topology into account allows for a smart calculation of the path over which the geometric relationship will be concatenated, resulting in a less accumulated error.



(a) Video acquisition setup (b) Video acquisition device (c) A sample frame

Figure 1.3: Video acquisition setup for the present study. Also shown are the acquisition device, developed in the framework of InnovaTICs projet, alongside a sample frame, with 1294×964 pixels dimensions, representing an area of approximately 3×2.25 cm^2 on the anterior forearm.

Given that the skin videos are acquired at close range with a small field of view (as illustrated in Fig. 1.3(a), the surfaces viewed in individual frames can be assumed to be quasi-planar. These acquisition conditions permit the use of a 2D mosaicing scheme, in which images are placed in a planar mosaic (panoramic image) in a way that their common parts overlap. Essential stages of

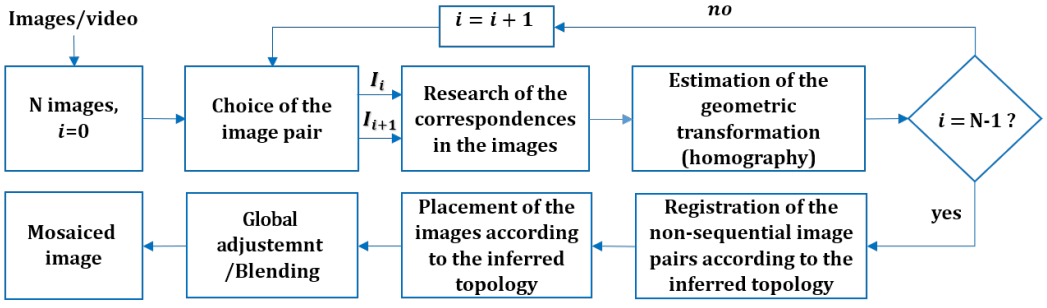


Figure 1.4: Pipeline highlighting the essential mosaicing steps.

the image mosaicing process are shown in Fig. 1.4. Some basic aspects of the mosaicing pipeline will be discussed hereafter.

1.3.1 Image pre-processing

Image pre-processing may be required if the region of interest is only a part of the acquired image or if there are acquisition-related distortions in the images. To extract the required region of interest, the image may simply be cropped. The distortion in the images may occur when they are acquired through a wide lens that has a short focal length f (it is often assumed that for $f > 12$ mm, the radial distortion is negligible). In this case, the outer image regions are displaced away from or towards the center. Since this distortion is associated with the radial displacement from the center, it is referred to as radial or barrel distortion. A simple correction model, that assumes that the distortions are proportional to their distance from the center, is given as [Sze06]:

$$x' = x_0 + x(1 + k_1 r^2 + k_2 r^4) \tag{1.8}$$
$$y' = y_0 + y(1 + k_1 r^2 + k_2 r^4), \tag{1.9}$$

where (x, y) are the image coordinates of a pixel and (x', y') are the coordinates after correction. $(x_0,\ y_0)$ is the projection of the optical center into the image plane and $r = \sqrt{x^2 + y^2}$. Coefficients k_1 and k_2 are the distortion parameters that need to be determined (together with x_0 and y_0) from this pair of polynomial equations. In [ML+04], it was shown that even for strong barrel distortion (i.e. short focal length) two coefficients k_i are sufficient to model the distortion accurately.

1.3.2 Choice of a motion model

A crucial step in image mosaicing is the choice of motion parameters, i.e. selection of the geometric model that determines the transformation between the paired images to be registered. These models vary from simple 2D transformations to perspective transform. They may also consider 3D transformations and projections on a non-planar mosaicing surface. Here, only the linear 2D transform models, that include the perspective transform, are presented.

The simplest motion model involves just an in plane translation, i.e. the coordinates $\mathbf{x} = (x, y)$ are displaced to $\mathbf{x}' = \mathbf{x} + \mathbf{t}$, $\mathbf{t} = [t_x\ t_y]^T$ being the 2D displacement. In homogeneous coordinates, the relationship between the initial position $\mathbf{x}_h = (x, y, 1)$ and the transformed position $\mathbf{x}'_h = (x',\ y',\ 1)$ is given as:

$$\mathbf{x}'_h = \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix} \mathbf{x}_h \tag{1.10}$$

The motion model that considers, in addition to the translation, the 2D in plane rotation ϕ, is known as the Euclidean transform (Euclidean distances are preserved). This transform is also known as the rigid body transform since the shape of the objects in the image remains unmodified. This transform is defined as:

$$\mathbf{x}'_h = \begin{bmatrix} \cos\phi & -\sin\phi & t_x \\ \sin\phi & \cos\phi & t_y \\ 0 & 0 & 1 \end{bmatrix} \mathbf{x}_h \tag{1.11}$$

If the scale factor is added to the above considered transform parameters, the resulting transform, sometimes referred to as similarity transform, preserves the angles between the lines. Similarity transform is given as:

$$\mathbf{x}'_h = \begin{bmatrix} a\cos\phi & -b\sin\phi & t_x \\ b\sin\phi & a\cos\phi & t_y \\ 0 & 0 & 1 \end{bmatrix} \mathbf{x}_h, \tag{1.12}$$

where the coefficients a and b determine the isotropic scale factor.

If asymmetric scaling is used for x and y coordinates, the resulting transform, in which parallel lines remain parallel but shearing appears, is called affine transform and is given as:

$$\mathbf{x}'_h = \begin{bmatrix} f_x\cos\phi & -s_x\sin\phi & t_x \\ s_y\sin\phi & f_y\cos\phi & t_y \\ 0 & 0 & 1 \end{bmatrix} \mathbf{x}_h, \tag{1.13}$$

where f_x and f_y determine the scaling in the x and y directions and s_x and s_y determine the shearing in the corresponding directions. In the above transforms, the homogeneous coordinates only help to do the matrix operations in a more convenient way. They play an effective role in the projective transform, also known as perspective transform or homography, where the perspective transformation is taken into account:

$$\mathbf{x}'_h = \frac{1}{\alpha} \begin{bmatrix} f_x\cos\phi & -s_x\sin\phi & t_x \\ s_y\sin\phi & f_y\cos\phi & t_y \\ h_{3,1} & h_{3,2} & 1 \end{bmatrix} \mathbf{x}_h, \tag{1.14}$$

where the parameters $h_{3,1}$ and $h_{3,2}$ are related to two out-plane rotations and α is the normalizing factor for the homogeneous coordinates.

After the choice of an appropriate geometrical transform model, point correspondences between I_i and I_{i-1} have to be established to computes its parameter values.

1.3.3 Optical flow based point correspondence determination

Optical flow (OF) across two images of the same scene is a vector field that relates the homologous pixels in the two images. In the highly influential work of Horn and Schunk[HS81], the optical flow calculation assumes that the intensity of a pixel does not change over time, i.e. the homologous pixels in the two images have the same value. This intensity constancy constraint is stated as:

$$\|I_1(\mathbf{x}+\mathbf{u_x}) - I_0(\mathbf{x})\|_2 = 0, \forall \mathbf{x}\epsilon\Omega \tag{1.15}$$

with I_1 and I_0 being the source and target images respectively, $\mathbf{x}$ the 2D pixel coordinates, $\mathbf{u_x}$ the 2D optical flow vector and Ω the image plane representing the zone over which the optical flow is calculated.

Using a first order Taylor series approximation, $I_1(\mathbf{x}+\mathbf{u_x})$ can be linearized as:

$$I_1(\mathbf{x}+\mathbf{u_x}) \approx \nabla I_1.\Delta\mathbf{u_x} + I_1(\mathbf{x}+\mathbf{u_x^0}) \tag{1.16}$$

where $\mathbf{u_x}^0$ is an initial guess of the flow vector and $\Delta\mathbf{u_x} = \mathbf{u_x} - \mathbf{u_x}^0$. Using this approximation, the data term for optical flow calculation can be formulated as:

$$\rho_{HS}(\mathbf{u_x}) = I_1(x+u_x) - I_0(x) = \nabla I_1.\Delta\mathbf{u_x} + I_1(\mathbf{x}+\mathbf{u_x}^0) - I_0(\mathbf{x}) \tag{1.17}$$

Since there may exist many potential matches for every pixel, regularization is required. In [HS81], this is achieved through minimizing the variations in optical flow vector gradients. This makes sure that the flow vectors in a neighborhood do not diverge excessively from each other. Together with the regularization term, the energy to be minimized for OF calculation can be written as:

$$E_{HS}(\mathbf{u_x}) = \int_{\Omega} \{ \underbrace{\|\nabla \mathbf{u_x}\|_2}_{regularization\ term} + \lambda \underbrace{\|\rho_{HS}(\mathbf{u_x})\|_2}_{data\ term} \} d\Omega, \tag{1.18}$$

where λ is a weighting coefficient.

1.3.4 Similarity measure based point correspondence

Similarity measures make use of a template (or a sub-image) extracted from one image. This template is then used to find the corresponding template, based on some similarity measure, in the other image. Some of these measures are mutual information, cross-correlation coefficient, Fourier-Melin (or log-polar) transform coefficients, the sum of squared distances and sum of absolute distances. Although these measures can be performed over the entire image plane, due to their large computation times, it is preferable to use sub-images to obtain point or region correspondences. Some of these approaches are discussed below.

Sum of absolute distances SAD between a template f_t of size $m \times n$ in the source image I_t and the template f_w of the same size in the target image I_w is given as:

$$SAD(x, y) = \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} |f_t(i, j) - f_w(i + x, j + y)|, \tag{1.19}$$

where (x, y) indicates the position of the window f_w in image coordinates.

SAD is calculated at different positions by displacing f_w. This metric gives, in fact, a dissimilarity measure: the smaller SAD, most likely similar the corresponding patches. Although an exhaustive search can be performed by centering the window at all the possible pixel locations, the more controlled search can be performed by integrating it into some optimization scheme, like the gradient descent, as used in [Dew78]. Furthermore, simulated annealing can be used to avoid local optima. Vanderbrug and Rosenfeld in [VR77] used further sub-patches of the template to find candidate positions corresponding to these smaller patches. Computation is then reduced by calculating the SAD at these positions only. In [SMA76], it was reported that more reliable results are achieved if image gradients are used instead of the raw pixel intensities.

The centered normalized cross-correlation coefficient between two templates is defined as:

$$CC(x, y) = \frac{\sum_{i=0}^{m-1} \sum_{j=0}^{n-1} g_t(i, j) g_w(i + x, j + y)}{\{\sum_{i=0}^{m-1} \sum_{j=0}^{n-1} g_t^2(i, j) \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} g_w^2(i + x, j + y)\}^2}, \tag{1.20}$$

with $g_t(i, j) = f_t(i, j) - \overline{f_t}$ and $w_t(i, j) = w_t(i, j) - \overline{f_w}$, where $\overline{f_t}$ and $\overline{f_w}$ indicate the average pixel values of the respective windows.

The values of CC vary in the interval [-1, 1], with 1 indicating the maximum similarity. The computation time for CC, in case of an exhaustive research, is larger than that of SAD because it involves several multiplications. However, noting that the numerator in Eq. (1.20) represents a convolution operation, the calculations can be speeded up using fast Fourier transform. Furthermore, the search can be performed in conjunction with the gradient descent methods.

Mutual information is a measure of the certainty that knowing the value of a random variable in one dataset provides about the value of a random variable in another data set. In other words,

it determines the interdependence of the two images. Mutual information consists in calculating the simple and joint probability density functions P_t, P_w and $P_{w,t}$ of the templates to be matched. Let $P_t(a)$ be the probability that a pixel in f_t has an intensity value of a and $P_w(a)$ the one that a pixel in f_w has an intensity value of b. Then $P_{tw}(a,b)$ is the joint probability that the pixels at the same locations in the two superimposed templates have the respective intensity values of a and b. If the two templates have large similarities, their joint probabilities will have higher values, thus minimal joint entropy. On the other hand, when their probabilities decrease as their differences increase, to the point that the templates are completely different, their joint probabilities will simply be $P_t(a)P_w(b)$. Using these notations, the mutual information ($MI(t,w)$) between the two templates f_t and f_w is defined as:

$$MI(t,w) = \sum_{a=0}^{N}\sum_{b=0}^{N} P_{tw}(a,b) log_2 \frac{P_{tw}(a,b)}{P_t(a)P_w(b)}, \tag{1.21}$$

,

where N is the number of possible intensity values a pixel can have.

It can be observed in Eq. (1.21) that the *log* term will be null when there is no dependence between the two templates:

$$log_2 \frac{P_{tw}(a,b)}{P_t(a)P_w(b)} = log_2 \frac{P_t(a)P_w(b)}{P_t(a)P_w(b)} = log_2(1) = 0$$

The peak value of MI will indicate a maximum match. Since its calculation is highly sensitive to noise, it has limited applications in image registration.

Similarity measures discussed so far are easy to use for transforms that contain only translations. Although they can be extended to other homographic parameters, this adds extensive computation overload for calculating these measures at gradual variations in the parameters. While the approaches based on the intensity values (CC and SAD) have their use limited to monomodal image registration, MI can be used to register multimodality images. There also exist similarity measures that are rotation and scale invariant. Template matching in the log-polar domain, for example, is both scale and in-plane rotation invariant. Rigid linear transformation, taking into account the scale change as well, involves four parameters: 2D translations, rotation, and scale factor. Suppose an image $I_t(x,y)$, where x and y correspond to point coordinates on a Cartesian plane, and its transformed version $I_s(x,y)$ on the same plane. Assuming four transformation parameters, the correspondence between the points of I_s and I_t is given by the following relationship:

$$I_t(x,y) = I_s(x^{'}, y^{'}), \tag{1.22}$$

with

$$\begin{bmatrix} x^{'} \\ y^{'} \end{bmatrix} = a \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} t_x \\ t_y \end{bmatrix} \tag{1.23}$$

where ϕ is the angle of rotation, a the scale factor and t_x and t_y are the horizontal and vertical translations. Searching point by point the transformation parameters requires immense calculations. The problem can be divided into two parts by taking advantage of the fact that the magnitude of the Fourier transform is translation independent, the frequency content remaining the same despite spatial displacements. The magnitude of the Fourier transform of Eq. 1.22 is

written as:

$$|FT(I_t(x,y))| = |\mathscr{I}_t(u,v)| = \frac{1}{a^2}\left|\mathscr{I}_s\left(\frac{u\cos\phi + v\sin\phi}{a}, \frac{-u\sin\phi + v\cos\phi}{a}\right)\right|, \tag{1.24}$$

where $\mathscr{I}_t$ and $\mathscr{I}_t$ are Fourier transform images of I_t and I_s respectively, with u and v representing the coordinates on the Fourier plane.

The problem is thus reduced to two parameters: scale and rotation. The relationship between the two images becomes linear when the log-polar coordinates are used. Eq. (1.24) in polar coordinates becomes:

$$|\mathscr{I}_t(r,\theta)| = \frac{1}{a^2}\left|\mathscr{I}_s\left(\frac{r}{a}, \theta + \phi\right)\right|, \tag{1.25}$$

where $r = \sqrt{u^2+v^2}$ and $\theta = \tan^{-1}(\frac{u}{v})$

The relationship is nonlinear in r. By taking the logarithm of the first coordinate, a completely linear relation is obtained:

$$|\mathscr{I}_t(r,\theta)| = \frac{1}{a^2}\left|\mathscr{I}_s\left(\rho - \ln(a), \theta + \phi\right)\right|, \tag{1.26}$$

where $\rho = \ln(r)$

The goal now is to find the horizontal and vertical offsets in the log-polar image frequency domain. For this, the cross-correlation of the images represented by Eq. (1.26) is calculated. The displacement giving maximum correlation gives the sought parameters, which correspond to the rotation and the scale factor in the spatial domain.

1.3.5 Feature based approaches

The intensity-based methods (optical flow and similarity measure based) seek to establish, through tedious optimization algorithms, a one-to-one correspondence among all the pixels in the two images. The computation time for registration can be reduced significantly if a subset of pixels, corresponding to some uniquely identifiable points or features is used for image matching purposes. Another advantage of the feature based approaches is their ability to register images with a smaller overlap than the one required by the optical flow based methods. However, the success of the feature based methods is contingent upon robust extraction of enough features.

The most basic of the features are geometrical primitives such as lines, corners or points that can be located at the cross sections of lines. The corresponding points provide a simplified way of calculating the geometric relation between the images. If there is a possibility of detecting lines, the points on the corresponding lines can be used for determining the geometric relation. In this subsection, the principle of feature extraction is illustrated through interest point and line detection.

1.3.5.1 Feature detection

Some points in an image can be classified as recognizable if their neighborhood has distinguishable characteristics. One way of finding these points is to detect corners in an image. Such points are detected through a cornerness measure. Among various cornerness measures that exist, one measure makes use of the image gradients [Gos05]. Let $I(x,y)$ represent the intensity of image I at 2D coordinates (x,y) and $I_x(x,y)$ and $I_y(x,y)$ be its gradients in x and y directions

respectively. Using mean gradients $\overline{I_x(x,y)}$ and $\overline{I_y(x,y)}$ within a small window centered at (x,y), the inertia matrix C at this point is defined as:

$$C(x,y) = \begin{bmatrix} \overline{I_x(x,y)}\,\overline{I_x(x,y)} & \overline{I_x(x,y)}\,\overline{I_y(x,y)} \\ \overline{I_y(x,y)}\,\overline{I_x(x,y)} & \overline{I_y(x,y)}\,\overline{I_y(x,y)} \end{bmatrix} \tag{1.27}$$

In [Roh92], $det(C)$, the determinant of the inertia matrix and in [Fö92], the ratio $det(C)/tr(C)$, $tr(C)$ being the trace of the matrix C, were used as the cornerness measures. Tomasi and Kanade [TK92] used eigenvalues of the matrix C to define a cornerness measure. If both eigenvalues are large, this would indicate the presence of a strong corner. Their approach consists in first computing a contour image using a Gaussian filter of variance σ. Then they calculate the inertia matrices, within a circular window of size 3σ, for all the points corresponding to the detected contours. The smaller of the eigenvalues for each inertia matrix is then compared with the smaller eigenvalues of the inertia matrices corresponding to points within a 3×3 neighborhood. If the eigenvalues of the inertia matrix under consideration turns out to be the smallest, it is classified as a corner point. The corner points thus detected are further refined by sorting them, in descending order, with respect to the corresponding smaller eigenvalues. Starting from the corner point at the top of the list, for each corner point in the list, the other corner points within 7σ of this corner point are removed from the list. The distance factor of 7σ is chosen empirically and can be varied. This elimination process leaves the list with the corner points that are well distributed and sufficiently spaced over the image.

Harris and Stephens [HS88] used an improved corner detection based on eigenvalues. For λ_1 and λ_2 representing the eigenvalues, with $\lambda_1 > \lambda_2$, of the inertia matrix, they defined the cornerness measure as:

$$R = \lambda_1\lambda_2 - k(\lambda_1 + \lambda_2)^2 = det(C) - k\ trace^2(C), \tag{1.28}$$

where k is a parameter that determines the detector's sensitivity. The value of k is dependent on the image content.

The trace term in Eq. (1.28) makes it robust against false corner detection at the staircase-like diagonal lines. Moreover, in [HS88], instead of using the simple average of the gradients, they used a weighting Gaussian average. In the cornerness measures approaches that are discussed here, a first-order gradient is used. The second order gradient can also be used to formulate this measure [Bea78; KR82].

In the cornerness measures presented above, the uniqueness of the detected point of interest is not taken into account explicitly. Since the window used for the calculation of the inertia matrix is circular, the detected corner point is, generally, independent of the orientation of the image. One factor that slightly impacts the invariability with respect to the rotation is the use of the 3×3 rectangular window for determining if the detected point of interest is the sharpest in a neighborhood. Another factor that affects the repeatability of the detected points of interest is the image resolution. Due to interpolation/extrapolation of the pixel intensities, the sharpest corner in a neighborhood may vary at different resolutions. This variation in the detected points can be exploited to detect most certain corner points by retaining only those that are detected at various resolutions.

For image matching purposes, it is preferable to prioritize the corner points that are not only unique (i.e. they can be uniquely identifiable in a neighborhood in the image) but are also located in a highly informative neighborhood. The neighborhood with high information content is more likely to give unique points in two images that would correspond to each other. To explicitly take into account these two factors, a uniqueness measure, that ensures that the corner is locally

unique, and an informative measure are taken into account while selecting the image contours over which the interest points would be searched.

A way to quantitatively measure the uniqueness of a corner point (or image contour) is to calculate the cross-correlation (CC) of the circular window centered at this point with windows of the same size centered around the neighboring points [Gos05]. If the normalized CC has a small value for all the neighboring points, this would be indicative of the uniqueness of the corner point. A single value close to 1, the maximum possible value in the normalized CC, would suggest otherwise. For a window $\mathbf{W}(x, y)$ centered at the point coordinates (x, y), the uniqueness measure $U(x, y)$, considering the above criterion, can be formulated as:

$$U(x, y) = 1 - MAX\{CC[\mathbf{W}(x, y), \mathbf{W}(x + x', y + y')]\}, \tag{1.29}$$

where x' and y' represent the window displacement corresponding to an 8-neighborhood, i.e. $x', y' \in \{1, 0, -1\}$ with both values not being simultaneously 0. The cross-correlation between two windows $\mathbf{W_1}$ and $\mathbf{W_2}$ is defined as:

$$CC[\mathbf{W_1}, \mathbf{W_2}] = \frac{\mathbf{W_1}.\mathbf{W_2}}{\|\mathbf{W_1}\| \, \|\mathbf{W_2}\|}, \tag{1.30}$$

where "." represents the dot product.

The information measure makes use of the Shannon entropy. The corner points in a region with a high entropy content contain more information and are more suitable candidates for establishing correct correspondences. The entropy, within a neighborhood defined by a circular window, is calculated as:

$$E = -\sum_{i=0}^{N} p_i log_2(p_i), \tag{1.31}$$

where p_i is the probability that a pixel, within the circular window, would have an intensity i and N is the number of possible intensity values a pixel can have. It is calculated from the histogram $h(i)$ as:

$$p_i = h(i)/T, \tag{1.32}$$

where T is the total number of pixels in the circular window and $h(i)$ is the number of pixels with intensity i.

To make use of the information and uniqueness measures in the selection of the corner points, first, from an initial set of the detected points, only the ones having a high information neighborhood and a uniqueness measure above a threshold are retained. In the retained points, further refinement is performed to keep only the ones well distributed over the image. For this, the selected points are arranged in descending order with respect to the uniqueness measures initially assigned to each of them. The final choice of the interest point is made by selecting, iteratively, the point with the highest uniqueness measure. At each iteration, after having retained the point with the highest uniqueness measure, the uniqueness measures of the rest of the points under consideration are updated by multiplying them with a factor $H = 1 - exp(-d_i^2/D^2)$. Here, d_i is the distance between the i^{th} candidate point and the interest point just selected and D is a parameter that controls the desired spacing between the selected points. After updating the uniqueness measures, the remaining candidates are sorted again and the process is repeated until the desired number of corner points have been retained. The factor H is inversely proportional

to the distance from the selected corner point and, since it rapidly reaches 0, it helps eliminate only the candidates that are in the close neighborhood of the selected corner point.

Instead of using the factor H, the process can be speeded up by eliminating the candidate points that fall within a fixed distance of the selected corner point. Another speed *vs* effectiveness consideration involves the size of the circular window used for calculating the information content and the uniqueness. A smaller window would speed up the process. However, the resulting corner points may lack sufficient uniqueness and information content to achieve successful correspondences. A larger window, on the other hand, although adds a calculation overload, gives more distinguishable and informative points. To make a compromise between these two aspects, a window size 3σ to 5σ pixels, σ being the variance of the Gaussian filter, is generally used for detecting the contours [Gos05].

Apart from the corner points, lines may represent unique features in an image. If an image is known to contain a single line, it may be detected through a simple least square fit. For the images containing multiple lines, Hough Transform [Hou62] is a more suitable and widely used approach. It consists in detecting the lines by presenting all the possible lines passing through the pixels of a binary image considered in Cartesian coordinates into polar coordinates. Let x and y represent the coordinates of a point in the Cartesian plane. The slope-intercept form of the equation of the line passing through this point is given as:

$$y = mx + b, \tag{1.33}$$

with m being the slope of the line and b the y-intercept.

In the mb space, the line presented by Eq. (1.33) is indicated by a single point corresponding to m and b. A line in the mb space, on the other hand, corresponds to a single point in the Cartesian space. Since there is an infinite number of lines that may pass through a single point, a discrete accumulator is used in the mb space to determine the points in the Cartesian space that would correspond to the lines in the mb space. This accumulator, which has the size of $M \times B$, with each of its element representing the lines in mb space within the predefined interval in both m and b values, is initially set to zero. For every combination of m and b values, the corresponding bin in the accumulator is increased by 1 if a corresponding point is detected in the Cartesian space. This approach, however, poses some limitations due to the slope reaching infinity at 90 degree angle. This limitation can be overcome by rotating the image. However, a more appropriate solution is to use the polar coordinates. In polar coordinates, a line in the Cartesian coordinate corresponds to:

$$\rho = (x - x_c)cos(\theta) + (y - y_c)sin(\theta), \tag{1.34}$$

with ρ being the distance from the image center (x_c, y_c) and θ the angle that the line connecting the point (x, y) to the center (x_c, y_c) makes with the abscissa. The $\rho\theta$ space is also referred to as the Hough space. The accumulator in the Hough space contains bins corresponding to the discrete interval of θ and ρ, with the former varying from 0 to 360 degrees and the latter being within the range limited by the image diagonal. Each point in the Cartesian space corresponds to a sine wave in the Hough space. As in mb space, the bins of the accumulator are updated each time a point corresponding to values of a bin is detected in the image plane. In the end, the bins with the highest values are the most likely candidates for the detected lines.

1.3.5.2 Feature matching

Once the features, that are, in some way or another, uniquely identified in a pair of images, the next step in image alignment is to establish their correspondence. In the case of establishing the point correspondences, the problem faced is that points detected in the two images are not necessarily the same, i.e. some points detected in one image may be missing in the other and vice-versa. Besides, due to noise and quantification errors, the detected positions of the corresponding points may be displaced from what they would be according to the actual geometric transformation between the images. An approach for matching point correspondences while simultaneously determining the geometric relationship between the two images is discussed in detail in section 1.3.6. More sophisticated approaches for point matching exploit the neighborhood of the interest points to formulate a point descriptor vector. An extended discussion on the descriptor based approaches follows in Chapter 2.

1.3.6 Estimation of the geometric relationship

The next step, after finding the corresponding points in two overlapping images (I_i, I_{i+1}), is the estimation of the transformation parameters providing the geometric relation between these points. The objective is to find the parameter values of the transformation matrix (homography in our case) $H^{est}_{i,\,i+1}$ such that the pixels of image I_{i+1} transformed with this matrix take the coordinates of their homologous pixels in image I_i. Let $\mathbf{x} = (x,\ y) \in \mathbf{R}^2$ be the pixel coordinates in space Ω of the images. If pixel p_k with coordinates $\mathbf{x}_k = (x_k,\ y_k)$ in image I_i is homologous to pixel q_l with coordinates $\mathbf{x}_l = (x_l,\ y_l)$ in image I_{i+1}, the transformation matrix maps $\mathbf{x}_l$ to $\mathbf{x}_k$. For calculation purposes, the image pixels are mapped using homogeneous coordinates, shown in Eq. (1.35), where the parameters (f_x, f_y), ϕ, (s_x, s_y), (t_x, t_y) and (h_1, h_2) denote the scale factors, in-plane rotation, shearing parameters, 2D translation and perspective changes respectively, whereas α is the perspective scale factor determined by $h_{3,1}$ and $h_{3,2}$.

$$\begin{pmatrix} \alpha\, x_k \\ \alpha\, y_k \\ \alpha \end{pmatrix} = \underbrace{\begin{pmatrix} f_x \cos\phi & -s_x \sin\phi & t_x \\ s_y \sin\phi & f_y \cos\phi & t_y \\ h_{3,1} & h_{3,2} & 1 \end{pmatrix}}_{H^{est}_{i,i+1}} \begin{pmatrix} x_l \\ y_l \\ 1 \end{pmatrix} \tag{1.35}$$

While the correspondence between the homologous pixels is already established in the case of optical flow calculation, such correspondence needs to be established for the interest points detected using the feature based approaches. A basic approach for finding the transformation parameters for this case is through scene coherence [Gos05]. Let $P = \{p_k | k = 1, ..., n_t\}$ represent the detected points in the target image I_i, with $p_k = (x_k, y_k)$ for a total of n_t points, and $Q = \{q_l | l = 1, ..., n_s\}$ represent the detected points in the source image I_{i+1}, with $q_l = (x_l, y_l)$ for a total of n_s points. The objective is to match the homologous points in P and Q. Supposing a projective transformation between the images, the homologous point coordinates $\mathbf{x}$ and $\mathbf{x}'$ are related through the relation:

$$x' = \frac{h_{1,1}x + h_{1,2}y + h_{1,3}}{h_{3,1}x + h_{3,2}y + h_{3,3}} \tag{1.36a}$$

$$y' = \frac{h_{2,1}x + h_{2,2}y + h_{2,3}}{h_{3,1}x + h_{3,2}y + h_{3,3}} \tag{1.36b}$$

Parameters $\{h_{1,1}, h_{1,2},...,h_{3,3}\}$ (corresponding to the elements of the homography matrix in Eq. (1.35)) need to be determined such that every point in P is projected to the coordinates of its homologous point in Q. To solve the set of equations in Eq. (1.36), a minimum of four homologous points are required. Considering the scene coherence, if four points are correctly matched, the rest of the points can be matched by using the geometric relation determined through these points. In the simplest approach, four non-colinear points are selected from each of the detected point sets and the rest of the points are projected using the geometric relation calculated from these points. To find the likelihood that the initial points were a correct match, a match rating has to be calculated. One of these ratings is the direct Hausdorff distance [Gos05; HKR93]. The direct Hausdorff distance measures, for every point p in P, its distance, after geometric transformation, from the closest point q in Q. Then the maximum of all these distances gives a measure that can be used as a match rating:

$$h(P,Q) = \underset{p\in P}{max}\ \underset{q\in Q}{min} \|p-q\| \tag{1.37}$$

A small $h(P,\ Q)$ value would indicate that the closest points are correctly matched, whereas a large value would indicate otherwise. One crucial factor is to take into account the impact of outliers: just one outlier would result in a negative match rating even if all the other points are correctly matched. One way to deal with this situation is to set a threshold and use only those distances that are below this threshold. Another way is to sort the corresponding candidate points in ascending order of their distances and check the matches for a limited number of points. If this number is sufficiently large, indicating that most of the candidate points were correctly matched, the rest of the candidate points can be discarded.

Once the correct correspondences have been found, the initially chosen transform parameters are refined through a least square fit of the homologous points. This is done because, even if the parameters relating the four initial points resulted in a small Hausdorff measure, there may still be some inaccuracy due to noise or quantification errors. A final refinement ensures that the estimated parameters provide the best possible fit for all the points. For the initial choice of the four points, all possible combinations can be tested if the number of points is small (< 10). The combination resulting in the least match rating can then be used as a reference. However, it is not realistic to do so when a large number of interest points are detected. In that case, the simplest approach is to continue randomly selecting the initial points and estimating the transform parameters until the desired accuracy is achieved or a predefined number of trials is reached without success.

If the correspondence of the points in the two images is known, the task is reduced to directly finding the homography. Randomly selected sample points are used in the widely used RANSAC (RANdom SAmple Consensus) approach [FB81; Sze06], which is designed to detect the outliers, i.e. the points that do not fit a given mathematical model. In this approach, k samples are selected from a set of point pairs with initially established correspondences. After determining the geometric relationship between these points, the residue $\mathbf{r}$ in the projected positions of all the other points is calculated as:

$$\mathbf{r} = \sum_{i=1}^{m} \left\| \mathbf{x}_i - \mathbf{x'}_i \right\|, \tag{1.38}$$

where m is the number of points classified as inliers, $\mathbf{x}_i$ is a point in one image and $\mathbf{x'}_i$ is the projection from the other image onto this image of the corresponding point.

After repeating the process S times with randomly selected samples, the geometric relationship giving the least residue is retained. This relationship is further refined, with the least square fit, to achieve the final parameter estimation. The number of minimum trials (S) should be sufficiently large to ensure that the randomly selected samples finally result in a successful parameter estimation. Supposing that the probability of success after S trials is P, and p^k is the probability that the k points in the randomly selected sample provide a valid match, the probability that RANSAC would fail after S trials is:

$$1 - P = (1 - p^k)^S \tag{1.39}$$

From this, the required minimum number of trials is calculated as:

$$S = \frac{log(1 - P)}{log((1 - p^k))} \tag{1.40}$$

In [Ste99], the number of minimum trials to achieve 99 % likelihood of getting successful matches using a different number of sample points (k) is provided (Table 1.1). This indicates that, for the same probability of valid matches of the sample points, the number of minimum required trials increases at a large rate as the number of sample points is increased. This is the reason that a smaller sample is preferred in RANSAC.

Table 1.1: Number of minimum trials required to attain 99 % probability of success for some number of sample points.

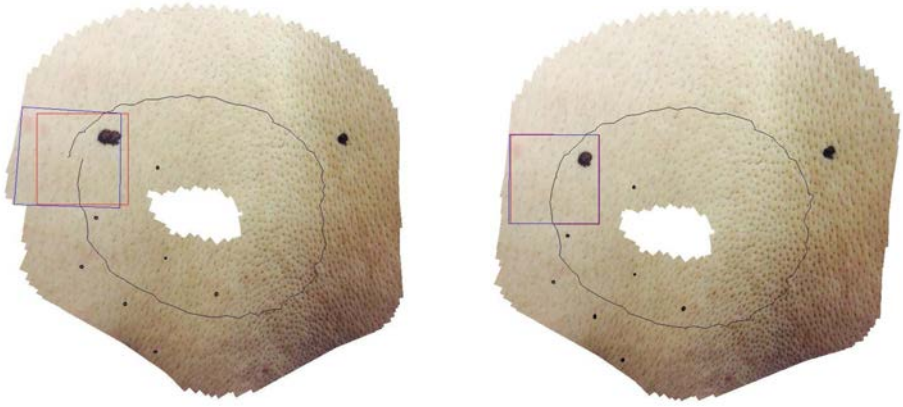
k	p^k	S
3	0.5	35
6	0.5	97
6	0.6	293

In PROSAC (PROgressive SAmple Consensus) [CM05], a more recent version of RANSAC, the initial sample points are randomly selected from a subset containing the points with most confident correspondences.

1.3.7 Global adjustment

Even with a fairly accurate registration, sub-pixel order errors remain and accumulate with homography concatenation when mosaicing a large sequence. Such accumulated errors result in significantly large mosaicing error and consequent distortion in the mosaic. The errors are especially perceptible when the sequence is in a closed loop (see Fig. 1.5) or if there are cross-overs in the acquisition trajectory. Even in the absence of the trajectory crossings, the mosaic may be significantly distorted after several images due to the accumulation of homographic errors. Although perspective transformations are better recovered in 3D registration [KK07], faster computation and simplified acquisition system are the motivations for adapting a 2D registration approach.

Assuming the local registration errors are not significant, misalignment in the initial mosaic can be sufficiently corrected by employing various global adjustment techniques once the sequential homographies have been calculated. In [MFM04; Wei+12b], the authors minimized the displacement, with respect to the points of a predefined grid on the mosaic plane, of the points



(a) Mosaic without global adjustment.

(b) Mosaic of the same sequence after global adjustment.

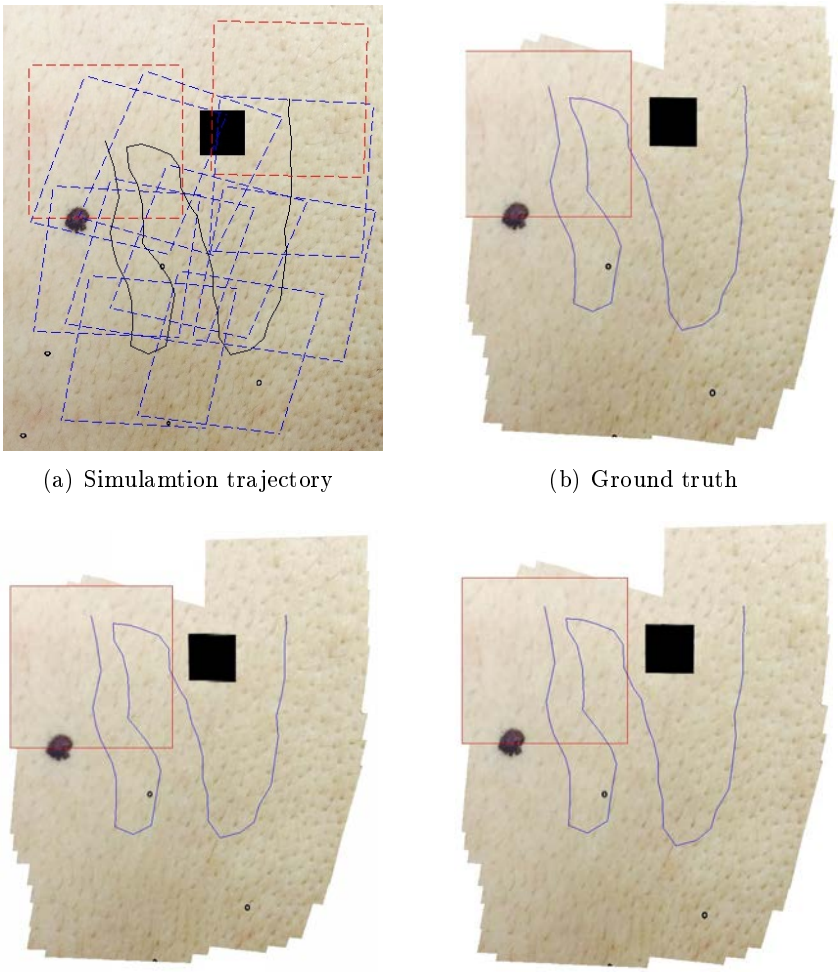
Figure 1.5: Mosaic of a human dorsal region skin, without (left) and with (right) global adjustment. The sequence contains 101 images with 512×512 pixel size each (representing approximately 2×2 cm^2 skin area) each. The red square represents the first image of the loop sequence and blue quadrangle the last one. The black line shows the mosaicing trajectory over image centers.

projected with estimated homographies that could be concatenated over more than one trajectory. However, apart from requiring calculation of additional trajectories over non-consecutive image pairs, this solution is computationally quite expensive. Miranda et al. [ML+08] proposed a faster approach to achieve correct alignment of the first and last images, in case of a closed-loop acquisition trajectory. Their approach consisted in adjusting the sequential homographies in a controlled manner such that the difference between the direct homography relating the first and the last images and the homography concatenated over the loop is minimized. This approach provides a fast adjustment, but at the risk of affecting significantly the registration accuracy between consecutive image pairs because, besides the homographies, no image information is used in the adjustment. In [BL03], Brown and Lowe minimized, for each image, the sum of the quadratic distances between the positions of the SIFT descriptors in this image and the positions representing the projection, in the same image, of the homologous descriptors detected in the other images. For this minimization, the camera parameters were varied using an optimization routine to which images were added one by one (images with the best matches being added first). Fig. (1.5) shows the result results without and with global adjustment of a sequence. The sequence used in this mosaic is simulated such that the first and the last images coincide. The global adjustment helped to close the loop.

1.3.8 Mosaicing with topology inference

Once all the consecutive homographies are determined, all the images belonging to the sequence are projected onto a mosaic plane whose coordinate system is anchored on that of a reference image I_{ref} selected from the sequence. This placement is done by concatenating the homographies $H^{est}_{i-k-1,i-k}$ estimated over the path from the reference image I_{ref} to the image I_i under consideration:

$$H^{est}_{ref \leftarrow i} = \prod_{k=ref}^{k=i-1} H^{est}_{i-k-1,i-k} \tag{1.41}$$



(a) Simulamtion trajectory

(b) Ground truth

(c) Mosaiced using the sequential paths

(d) Mosaiced using "short-circuiting" paths

Figure 1.6: Mosaics of a video sequence simulated from a real skin image. More direct paths reduced the blur and ghost textures. Preservation of the shape of the black square indicates reduction in error accumulation.

Eq. (1.41) helps to estimate the relative positions of all the images in the sequence, provided the registration for all the image pairs has been successful. This allows constructing an initially estimated topology of the images in a video sequence. However, due to the limited precision of the registration methods, this topology is not precise, particularly for larger sequences, where the accumulation of errors distorts the mosaic. So, it is desirable to reduce the number of homographies to be concatenated. This can be achieved in a sequence which contains several crossings in the acquisition trajectory, i.e. the possibility of concatenating the homographies through these crossings should provide some additional non-sequential image pairs with significant overlap. Since the image topology of a non-mosaiced video sequence remains unknown, the initial estimate of the image positions obtained from sequentially registered image pairs is used to infer a more connected topology from which more direct "short-circuiting" paths for the images to be mosaiced can be found. Fig. (1.6) illustrates the improvement resulting from a refined topology.

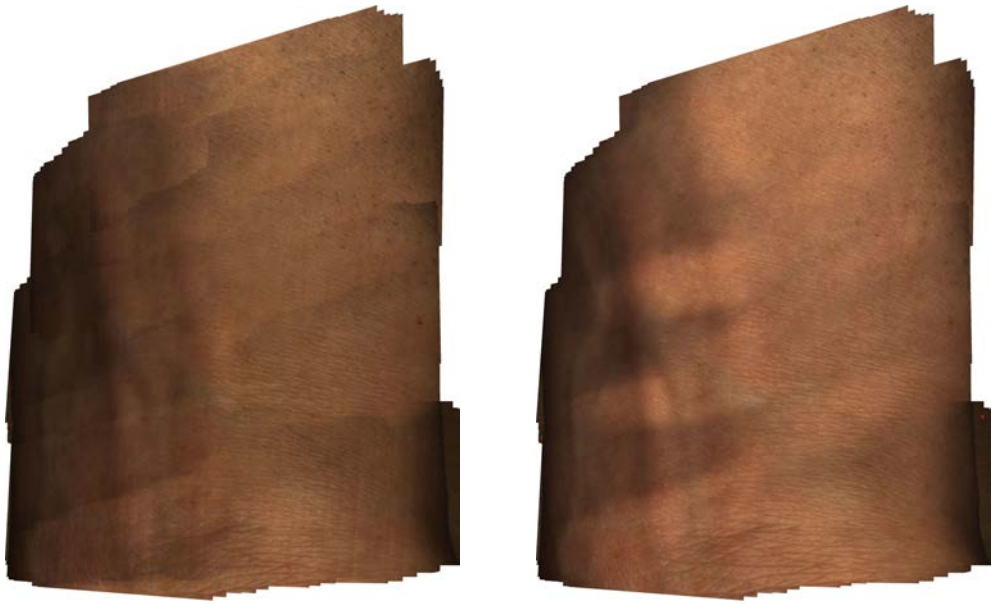
1.3.9 Blending

If the light exposure is uniform over all the images from which the mosaic is created, it would be sufficient just to add them to the mosaicing surface such that their non-overlapping regions are placed together. Since the illumination conditions are susceptible to vary over different frames, a simple juxtaposition of the mutually exclusive parts of the registered images creates visible seams and varying contrasts in the final mosaic. Therefore, it is desirable to implement some blending scheme to achieve a visually attractive and coherent mosaic. The simplest way to achieve this is to take the average of the pixels that overlap:

$$C(\mathbf{x}) = \frac{\sum_k w_k(\mathbf{x})\tilde{I}_k(\mathbf{x})}{\sum_k w_k(\mathbf{x})}, \tag{1.42}$$

where w_k is a binary weight factor that would be 1 if the pixel $\mathbf{x}$ in the warped image $\tilde{I}_k(\mathbf{x})$ belongs to the overlap region under consideration.

A simple average blending, in general, does not produce visually coherent results. Due to exposure differences, visible seams may still be present. A more appropriate approach is to weight the pixels in terms of their distance from the individual image borders, i.e. to assign a lower weight to the pixels near the edges and higher weight to pixels near the center of the images. This approach is called feathering or alpha-blending. The feathering can be performed by using the distance maps of the images. Fig. 1.7(b) shows the result of such blending over the mosaic of a forearm video sequence. Some more sophisticated approaches are gradient based global illumination correction [Lev+06; Wei+12a]. However, they are computationally expensive



(a) Mosaiced by choosing the highest intensity pixels in the overlap regions. This still leaves several sharp seams.

(b) Mosaic after blending. There is a smooth transition between the images and a visually attractive mosaic is obtained.

Figure 1.7: Application of alpha-blending on a mosaic of a forearm sequence.

1.4 Summary and Objectives of the Ph.D. work

In this chapter, a brief overview of the optical techniques used for the diagnosis of skin conditions was presented along with a general introduction of the mosaicing process. The effectiveness of telediagnosis was shown through various existing studies in teledermatology. These studies present a high rate of correct diagnosis of different skin conditions. However, the teledermatologists faced difficulty in analyzing poor quality images. Although images acquired with a dermoscope helped in making a correct diagnosis, the dermatologists may be further facilitated in their diagnosis through extended panoramas of the concerned skin region created through an adapted mosaicing scheme. Various aspects of the mosaicing process were presented and some of these were discussed in some detail. Particularly, different image registration approaches were overviewed and some of the basic keypoint extraction and matching schemes were discussed. The necessity of image topology consideration in image mosaicing and the need for blending was also presented.

(a) Image mosaiced using the SURF based registration

(b) Photographed with a smartphone camera

Figure 1.8: (a): Partial anterior forearm region mosaiced over 250 frames. The resulting mosaic has a size of 2742×3592 pixels. The mosaicing trajectory along with the initial (red rectangle) and every hundredth frame (green quadrangles) are marked on the image. (b): Same anterior forearm, photographed with a smartphone, over the region corresponding to the mosaiced zone. The image has a resolution of 1406×1372 pixels showing approximately 7×6 cm^2 of skin area.

The main objective of the present Ph.D. work was to develop a reliable approach for mosaicing large sequences of skin images. This involved inspecting various stages of the mosaicing pipeline and figuring out the stage that required the most attention. The feasibility of the skin mosaicing was established through mosaicing various simulated sequences using different image registration schemes. For this, not only the precision but also the calculation time was considered with the goal of finding the optimum compromise between registration accuracy and computation time. Since the observation of the pigmentation of the affected skin regions helps the clinical expertise and diagnosis, acquisition under controlled colorimetric and illumination conditions are

desired. This was the driving factor behind the use of a specially designed device [Amo+15], with embedded white LEDs, that is used to acquire high-resolution videos to be used in telediagnosis.

A mosaic of a video sequence (containing 250 frames) of the anterior forearm region acquired along a sinuous trajectory is shown in Fig. 1.8(a). Fig. 1.8(b) shows the corresponding region photographed using a smartphone camera with a resolution of 8 megapixels. The image obtained through mosaicing has a markedly higher resolution, almost quadruple the one attained with the smartphone camera. The visual coherence of the mosaic can be appreciated from the fact that the veins and the texture are well aligned.

After establishing the mosaicability of skin sequences through a survey of various image registration approaches (chapter 2 of this dissertation), the objective was to achieve extended and visually coherent mosaics in a reasonable time. Though a coherent mosaic is obtained in Fig. 1.8(a), this is not always the case for long sequences since, due to the accumulation of registration errors, the mosaic becomes distorted as the number of images grows. In addition, failed registrations in the middle of a sequence interrupt the mosaicing process. After experimenting with circular and zig-zag acquisition schemes, a spiral based approach inspired from [VRD13] is used to overcome these issues. Besides, a global adjustment scheme, based on descriptor locations, is introduced. These various aspects concerning an overall coherent mosaic construction are treated in chapter 3 of this dissertation.

Chapter 2

Analysis of Registration Schemes Applicable to Skin Image Mosaicing

2.1 Introduction

In this chapter, after a brief review of the literature concerning the alignment of human skin images in diverse applications, several advanced registration approaches are discussed. Since image registration is a key aspect of the mosaicing process, the choice of an optimal method, in terms of computation time and precision, is important. The applicability of these approaches on skin images was tested and their performances compared to select an approach that provides the best compromise between computation time and precision. Since the objective is to use these approaches in a mosaicing scheme, their precision and robustness are evaluated over simulated video sequences containing image pairs with varying homographic parameters. After the choice of an optimized approach, results on some real sequences are also given. These results do not incorporate any topology refinement or global adjustment. Unavoidable persistent errors in several results make the case for adapting innovative approaches of topology inference and global adjustment that are presented in the next chapter.

2.2 Existing Studies Using Skin Image Registration

Although there is a vast body of works on mosaicing anatomical surfaces of human organs, like bladder and esophagus, the major part of the existing works on cutaneous images is dedicated to microscopic scale. Works that involve skin image registration at macroscopic scale are very limited in their application since they consider just a few geometric transformation parameters. Some of these works are presented hereafter.

2.2.1 Skin image registration based on microscopic imaging devices

Confocal imaging is used to obtain high-resolution images that are helpful in observing the fiber structure of skin tissue. However, the limited field of view constraints analysis of larger skin surfaces. By building an image mosaic, not only a wide field of view is obtained, but also the signal to noise ratio is improved because the overlapping pixels reduce the average noise. In 2008, Loewke et al. [Loe+08] were interested in the mosaicing of confocal images with the goal of diagnosing cancer at the cellular level. Images were acquired using a portable device that was connected to a data processing module. The device used was able to acquire 5 frames per second

at a resolution of 6 microns with each image containing 200×500 pixels and having an area of 27×85 microns. After correcting geometric distortion inherent to the acquisitions system, the registration was carried out in two steps: first, the optical flow between successive images was calculated and accumulated until a predetermined threshold was reached. Then, the registration of the images at both ends of the calculated motion vectors was refined by minimizing the mutual information of the superimposed regions. For optical flow calculation, the tracking algorithm of Lucas and Kanade [LK81] was used. Final registration was further refined by a gradient descent optimization.

2.2.2 Works involving skin image registration based on macroscopic imaging devices

Holmberg and Lanshammar [HL06], in 2006, sought to establish an accurate and fast method for the analysis of body movements. Until then, this analysis was conducted using markers placed on the body or by using methods based on the topological and colorimetric information of organs. The authors explored the plausibility of using the skin texture images for this purposes by registering square patches of various resolutions extracted from a large leg image. They showed that the texture of the skin alone provides sufficient information for an effective registration. However, the lack of a pronounced texture requires high-resolution images to achieve an accurate registration. They used the mutual information (MI) as a measure of similarity and adapted the simulated annealing algorithm to maximize this criterion. The images were captured with a camera positioned orthogonally at 1.7 m from the subject. Precautions were taken to limit vibrations and shadows. The source images had a resolution of 3000×2000 pixels. Patches of 100×100 pixels were extracted from one frame of the leg image were registered with another frame in the sequence. Four processing parameters (2D translation within ± 20 pixels range, in-plane rotation of up to $10°$, and up to 10 % scale variation) were considered. Three markers attached to the leg were used as an evaluation reference, but they were also used to introduce an initial estimate for the optimization algorithm. The best results were obtained for a starting point within a range of 10 pixels around the point corresponding to the correct registration. Tests were done on different colored skins, without and with hairs. The registration was a success in 72 % of the tests. The presence of hair had no effect on the results. The more reduced the image resolution was, the more accurate initial estimate was required.

In 2009, Noh et al. [NKP09] proposed a method for registering fingerprints in the framework of a biometric study. The scattering of light by the skin blurs veins in the images, which makes their comparison for identification purposes difficult. To overcome this obstacle, their proposed method uses some additional information from the ridges and contours of the fingers. Images are acquired sequentially with white and infrared lights. The transformation parameters (translation and in-plane rotation) are calculated on the white light image and subsequently applied on the infrared image, which is then compared with cataloged infrared images. The outline of the visual image is extracted by applying Sobel filter on the binarized image. The outline image is then projected onto the ordinate. The intensity values in the projection are the accumulated gray-scale values of the corresponding ordinate of the contour image. The accumulated projection of the gray levels results in two salient peaks, one corresponding to the lower contour of the finger and the other to the upper contour. The change in dispersion of the first peak is calculated in relation to the gradual rotation, in both directions, of the original image. At minimum dispersion, the finger is assumed to be well oriented orthogonally. The image aligned in this way is added to the reference catalog. Wrinkles corresponding to interphalangeal joints are considered to find the translation. The crossing point of the first joint and the horizontal axis in

the orthogonal image is defined as a control point. This point, which can be identified using simple mathematical morphology operations, is used to find the 2D translation between the cataloged images and the image under consideration. The thermal image is then transformed according to the three transformation parameters obtained in this way. To facilitate the comparison, the veins in the thermal image are accentuated by applying adaptive histogram equalization followed by application of a line tracking algorithm. The resulting image is dilated morphologically before the comparison. Dilated images of the veins maps are multiplied to quantify the similarity between the compared images. This gives the number of pixels shared by the two images. This method was tested on six subjects with several images of the same subject acquired at separate times. The similarity between the images from the same subject was found to be above 60 % and less than 30 % between different subjects. For large rotations, no significant similarity between the images of the same subjects was established, a limitation attributable to non-uniform exposure to infrared radiation.

While in the previous study bimodality images were used separately to improve the registration, in 2008, Schaefer et al. [Sch+08] proposed a method to produce a composite image by registering thermal and visual images. This was done to facilitate precise identification of the anatomical areas affected by diseases such as scleroderma. The authors put an important emphasis on the preprocessing of the images and computational implementation of their method. Preprocessing of white light images was required to eliminate the background. For this, they have adopted a method proposed by Fleck et al. [FFB96]. This method exploits the observation that the hue of the human skin is bounded by certain values. To use this information, images in RGB space were transformed into the HSB (Hue Saturation Brightness) space. In addition, a texture image T was constructed by subtracting brightness of the image from its median filter smoothed version. Pixels bounded by certain saturation, hue and texture values were classified as belonging to the skin. Background removal was trivial for the thermal image since the ambient temperature was controlled. The gradient descent was chosen as the optimization method for image registration, with mutual information (MI) being used as the similarity measure. B-splines were used for the interpolation of values in the transformed image. The calculation of the MI being the most expensive computational task, a worker-manager approach was used: the image was divided into several patches – each of them being transferred to a worker processor for calculating the Parzen's joint histogram. The initial transformation parameters were also communicated to workers. The managing processor recombined these results to calculate an overall histogram and the MI gradient. New transformation parameters and the step in the direction of the gradient were calculated after each iteration and transferred to the workers for the calculation of a new histogram. The process was repeated until the satisfaction of the stopping criterion.

In 2008, Maglogiannis [Mag03] made use of the Fourier-Melin transform, assuming four transformation parameters (scale factor, in-plane rotation and 2D translation) to register skin images containing lesions. This approach was proposed with the aim to study the lesion evolution over time.

2.3 Optical Flow (OF) Based Approaches

In the approaches dedicated to skin image alignment discussed in section 2.2, very few parameters are taken into account to register images acquired under somewhat constrained acquisition settings. They generally deal with linear transformations limited to 2D translation, in-plane rotation and anisotropic scale factor. However, the geometric transformation is more complex for video sequences acquired through a hand-held device. In this case, a complete homographic

relationship between the images needs to be determined to successfully mosaic a long sequence. In this section, two optical flow based approaches are presented. The point correspondences provided by the OF methods can be used to determine the parameters of a homography.

2.3.1 An intensity based OF approach

An L^2 norm is used in the classical OF calculation scheme of Horn and Schunck (Eq. (1.18)). However, this causes the outliers to have more influence in OF calculation and it also oversmooths the OF discontinuities (which are of concern in scenes involving obstacles and/or moving objects). In several works, an L^1 norm is used instead. In [CP11], in addition to the OF vector gradient constraint for regularization, the authors have introduced an additional illumination constraint w, a scalar value. The formulation of the energy to be minimized is:

$$E_{CP}(\mathbf{u}) = \int_{\Omega} \{\|\nabla \mathbf{u}\|_1 + \|\nabla w\|_1 + \lambda \|\rho_{CP}(\mathbf{u})\|_1\} d\Omega \tag{2.1}$$

with

$$\rho_{CP}(\mathbf{u}) = \nabla I_1 . \Delta \mathbf{u} + I_1(\mathbf{x} + \mathbf{u}^0) - I_0(\mathbf{x}) + \beta w, \tag{2.2}$$

where β is a weight coefficient.

In this study, the optimization was performed over the course of 30 iterations at each scale and repeated twice after warping the source image. The down-sampling factor for the multi-resolution scheme was 0.7 and the value for the coefficient λ was set to be 50.

2.3.2 A correlation-based OF approach

The OF schemes which minimize the difference in intensities of the matched homologous pixels are prone to give false matches if there are huge illumination differences between the two images. This problem can be avoided if the pixels are assigned a value describing their neighborhood pixels. These values are then used, instead of the raw intensity magnitudes, to formulate the energy to be minimized. Such values can be obtained through self-similarity transforms, like localized self-cross-correlation or the sum of squared distances (SSD) with a displaced version of the same image, or neighborhood descriptive transforms, such as census transform. The use of these transforms requires some compromise between achieving robustness against illumination changes and preserving the image information correctly in the energy formulation. In [DN13], a transform named correlation transform (CT) is used to formulate a simple energy that minimizes the SSDs of the CTs corresponding to all the pixels in the images. An interesting property of CT allows for the calculation of the cross-correlation between neighborhoods of the homologous pixels with a simple SSD formulation, a task which otherwise results in a complex energy function. Correlation transform CT_i at pixel $\mathbf{x}$ of an image I_i is defined as:

$$CT_i(\mathbf{x}, \mathbf{s}) = \left(\frac{I_i(\mathbf{x} + \mathbf{s}) - \mu(\mathbf{x})}{\sigma(\mathbf{x})} \right), \mathbf{s} \in \mathfrak{N}_{\mathbf{x}}, \tag{2.3}$$

where $\mathfrak{N}_{\mathbf{x}}$ is a search space centered at pixel $\mathbf{x}$ of image I_i, $\mu(\mathbf{x})$ and $\sigma(\mathbf{x})$ are respectively the mean and variance of intensity values of pixels falling in the search space $\mathfrak{N}_{\mathbf{x}}$ and $\mathbf{s}$ indicates the coordinates of an element in $\mathfrak{N}_{\mathbf{x}}$;

This transform is related to zero normalized cross correlation (ZNCC) between two patches of images I_0 and I_1 by the following relationship:

$$\frac{1}{|\mathfrak{N}_\mathbf{x}|}\sum_{\mathbf{s}\in\mathfrak{N}_\mathbf{x}}\left(\frac{I_1(\mathbf{s}+\mathbf{x})-\mu_1(\mathbf{x})}{\sigma_1(\mathbf{x})}-\frac{I_0(\mathbf{s}+\mathbf{x})-\mu_0(\mathbf{x})}{\sigma_0(\mathbf{x})}\right)^2=2\left(1-ZNCC(I_1,I_0,\mathbf{x})\right), \tag{2.4}$$

with

$$ZNCC(I_1,I_0,\mathbf{x})=\frac{1}{|\mathfrak{N}_\mathbf{x}|}.\frac{\langle I_1-\mu_1(\mathbf{x}),I_0-\mu_0(\mathbf{x})\rangle}{\sigma_1(\mathbf{x}).\sigma_0(\mathbf{x})}, \tag{2.5}$$

where σ_0 and σ_1 are the grey-level variances within the corresponding neighborhoods in images I_0 and I_1 respectively. μ_0 and μ_1 are the mean grey-level values in the same neighborhoods. This property highly simplifies a direct cross correlation calculation between the two patches. The energy formulation in a direct case is:

$$E_{d(ZNCC)}=\sum_{\mathbf{x}\in\Omega}\left\{\frac{1}{|\mathfrak{N}_\mathbf{x}|}\sum_{\mathbf{s}\in\mathfrak{N}_\mathbf{x}}\frac{\left(I_1(\mathbf{s}+\mathbf{u}_\mathbf{x})-\mu_1(\mathbf{x}+\mathbf{u}_\mathbf{x})\right).\left(I_0(\mathbf{s})-\mu_0(\mathbf{x})\right)}{\sigma_1(\mathbf{x}+\mathbf{u}_\mathbf{x}).\sigma_0(\mathbf{x}))}\right\} \tag{2.6}$$

This being quite complex for variational methods, the correlation transform gives the advantage of calculating the cross-correlation indirectly while simplifying the calculations. In this case, data term $E_d(\mathbf{u})$ is formulated to minimize the SSD between CT_1 and CT_0 of the corresponding images I_1 and I_0:

$$E_d(\mathbf{u})=\sum_{\mathbf{x}\in\Omega}\frac{1}{|\mathfrak{N}_\mathbf{x}|}\sum_{\mathbf{s}\in\mathfrak{N}_\mathbf{x}}\left(CT_1(\mathbf{x}+\mathbf{u}_\mathbf{x},\mathbf{s})-CT_0(\mathbf{x},\mathbf{s})\right)^2 \tag{2.7}$$

The smoothness term E_s, reproduced here for a complete presentation of the energy model, is:

$$E_s=\sum_{\mathbf{x}\in\Omega}\sum_{\mathbf{s}\in\mathfrak{N}_\mathbf{x}}bf_{\mathbf{x},\mathbf{s}}\left\|\mathbf{u}_\mathbf{s}-\mathbf{u}_\mathbf{x}\right\|_1, \tag{2.8}$$

where $bf_{\mathbf{x},\mathbf{s}}$ is an anisotropic regularization coefficient, based on bilateral filtering, that determines, based on intensity value comparisons, if the pixels in a neighborhood $\mathbf{s}$ belong to the same object (like discontinuity preservation with an L^1 norm, this sort of filtering finds its utility in OF calculation over 3D scenes and does not concern the registration of planar images).

Together with the two terms, the energy model in [DN13] is:

$$E_{CF}(\mathbf{u})=\lambda E_d(\mathbf{u})+E_s(\mathbf{u}) \tag{2.9}$$

2.4 Feature based approaches

Until the early 2000s, dense OF calculations were preferred over the feature based approaches because the former gave more precise correspondences. This trend began to change with the emergence of more sophisticated feature extraction and description schemes that were invariant to affine transforms. In addition, image features or interest points provide a fast way of processing information in several applications such as object recognition and panorama construction. In [Sze06], a well-known tutorial on mosaicing process, published in 2006, the author states to

have shifted, after being in favor of dense correspondence approaches, towards advocating the feature based approaches for panoramic construction. Although the OF based methods may yield more precise results over short displacements, they require more calculations. Moreover, despite a multiscale approach, they may fail to deal with large displacements (please see results in section 2.6.2 for a comparison, over simulated sequences involving large displacements, of some of the well-known approaches in both categories).

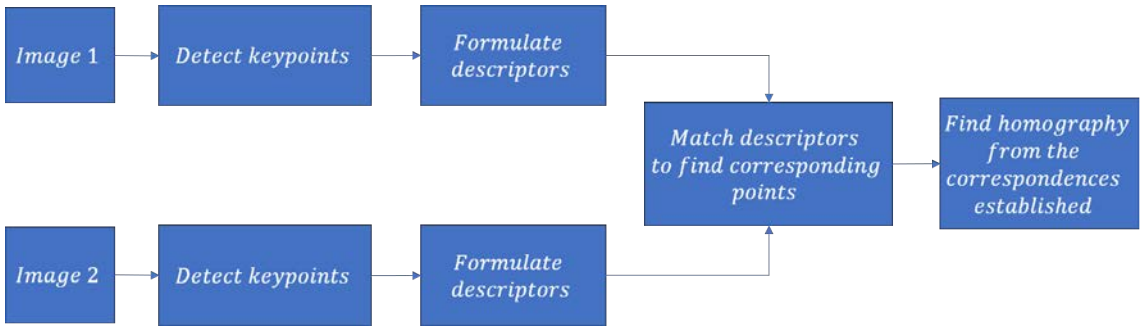


Figure 2.1: Steps involved in feature based registration.

Fig. 2.1 gives a general outline of feature-based image registration approaches. First, locally distinct keypoints are detected on both images. The region surrounding these points is taken into account to form a descriptor since the keypoints alone do not provide sufficient information for making a comparison. The descriptors thus obtained in both images are then compared to find out the best matches. This gives the keypoint correspondences which are then used to calculate homography with RANSAC algorithm. Various methods involved in each of these steps are discussed hereafter.

2.4.1 Keypoint extraction

The detected keypoints should be distinctly identifiable, at least locally, to minimize ambiguity. Some of the main keypoint extraction schemes involving blob or corner point detection are discussed below.

2.4.1.1 Moravec Corner detector

A corner in an image may represent the crossing of two lines or end of a line segment as well as simple isolated points corresponding to the maximum or minimum in a neighborhood. Moravec's corner detection approach [Mor80], one of the earliest detector in this category, uses the sum of squared distances (SSD) between a patch centered at a pixel and a displaced patch of the same size to define a cornerness measure S for each pixel (u, v) in image I:

$$S(x, y) = \sum_{u,v} w(u, v) \left(I(u + x, v + y) - I(u, v)\right)^2, \tag{2.10}$$

where $w(u, v)$ represents a square window (3×3 or 5×5 binary mask) centered at (u, v) to extract a patch and (x, y) is the amount by which the patch is displaced.

S is calculated for each pixel in 8 principle directions for $(x, y) \in \{-1, 0, 1\}$. A map is created that contains the minimum of the 8 conernerness values for each pixel. A large measure would indicate a strong corner presence. Non-maximal suppression is applied on this map to select final corners. This suppression consists in selecting the corner points with maximum cornerness

measure in a local neighborhood. Sharper corners can be selected by removing the corners below a predefined threshold.

2.4.1.2 Harris Corner detector

Moravec Corner detector detects only those corners that are present along horizontal and vertical axes or the four principle diagonals. Harris and Stephens [HS88] overcame this limitation by calculating differentials instead of a direct SSD of displaced patches. In addition, they used Gaussian weighted window. Using the Tailor series expansion of the first order, Eq. (2.10) can be rewritten as:

$$\begin{aligned} S(x,y) &= \sum_{u,v} w(u,v)\left(I(u,v) + I_x(u,v)x + I_y(u,v)y - I(u,v)\right)^2 \\ &= \sum_{u,v} w(u,v)\left(I_x(u,v)x + I_y(u,v)y\right)^2, \end{aligned} \tag{2.11}$$

where $I_x(u,v)$ and $I_y(u,v)$ are derivatives in x and y directions respectively.
Expanding the square in Eq. (2.11) results in:

$$S(x,y) = \sum_{u,v} w(u,v)\left(I_x^2(u,v)x^2 + 2xyI_x(u,v)I_y(u,v) + I_y^2(u,v)y^2\right), \tag{2.12}$$

which can be written in matrix form as:

$$S(x,y) = \begin{bmatrix} x & y \end{bmatrix} H(x,y) \begin{bmatrix} x \\ y \end{bmatrix}, \tag{2.13}$$

with

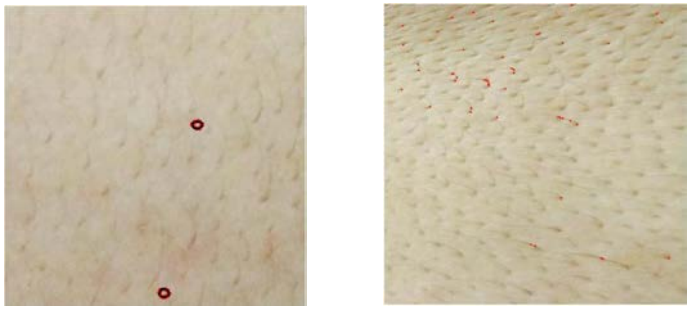
$$H(x,y) = \sum_{u,v} w(u,v) * \begin{bmatrix} I_x(x,y)\, I_x(x,y) & I_x(x,y)\, I_y(x,y) \\ I_y(x,y)\, I_x(x,y) & I_y(x,y)\, I_y(x,y) \end{bmatrix} \tag{2.14}$$

$H(x,y)$ is known as the Harris matrix. Its eigenvalues λ_1 and λ_2 help determine if a corner or an edge is present. If both eigenvalues are close to zero, this would indicate a homogeneous region, if only one of the eigenvalues is small, this would indicate edge presence and both eigenvalues being large would indicate a corner point. The calculation of eigenvalues, however, is expensive since it requires taking a square root. To avoid this, Harris and Stephens proposed a cornerness measure that implicitly depends on the eigenvalues, but which does not require the computation of λ_1 and λ_2:

$$M_c = \lambda_1\lambda_2 - k(\lambda_1 + \lambda_2)^2 = det(H) - k\, trace^2(H), \tag{2.15}$$

where k is a parameter that determines the detector's sensitivity. The value of k is dependent on the image content.

Fig. 2.2 shows the Harris corner points detected on two skin images. It can be noticed that no keypoints on the skin surface were detected when salient marks were present (Fig. 2.2(a)), whereas a few keypoints on the skin surface were detected in an image that did not have such marks (Fig. 2.2(b)).



(a) Skin image without any salient marks

(b) Skin image with a couple of marks

Figure 2.2: Keypoints detected on two skin images using Harris corner detector.

2.4.1.3 Multi-scale Laplacian of Gaussian (LoG) blob detector

A blob in an image represents a region in which the pixel intensities remain more or less constant and which presents contrast with the surrounding regions. To detect blobs of different sizes, the concept of scale-space is used. A scale-space L is constructed by convolving an image with a Gaussian kernel:

$$g(x, y, \sigma) = \frac{1}{2\pi\sigma} e^{-\frac{x^2+y^2}{2\sigma}}, \tag{2.16}$$

where the value of variance σ determines the scale.

The scale-space obtained through convolving with this Gaussian kernel is:

$$L(x, y, \sigma) = g(x, y, \sigma) * I(x, y) \tag{2.17}$$

The Laplacian operator is calculated at a given scale as:

$$\nabla^2 I(x, y) = I_{xx} + I_{yy}, \tag{2.18}$$

where I_{xx} and I_{yy} are second derivatives of image I in x and y directions respectively.

This results in a large response for the blobs of radius $\sigma\sqrt{2}$. However, the detected blobs depend on the size of the Gaussian kernel. For an automatic selection of the scale σ, a scale-normalized LoG is computed as:

$$\nabla^2 L(x, y, \sigma) = \sigma(L_{xx} + L_{yy}), \tag{2.19}$$

The maxima of this operator in scale-space indicate the blob presence.

2.4.1.4 Difference of Gaussian (DoG) blob detector

The heat equation of the scale-space function L is given as:

$$\frac{\partial L}{\partial \sigma} = \sigma \nabla^2 L \tag{2.20}$$

From this equation, Laplacian of Gaussian operator $\nabla^2 L(x, y, \sigma)$ can be approximated in a limiting case of $\frac{\partial L}{\partial \sigma}$ as:

$$\sigma \nabla^2 L = \frac{\partial L}{\partial \sigma} = \frac{L(x, y, k\sigma) - L(x, y, \sigma)}{k\sigma - \sigma} \tag{2.21}$$

Thus, LoG can be approximated by computing the difference of images smoothed by Gaussian operators at two different scales:

$$L(x, y, k\sigma) - L(x, y, \sigma) \approx (1 - k)\sigma^2 \nabla^2 L \tag{2.22}$$

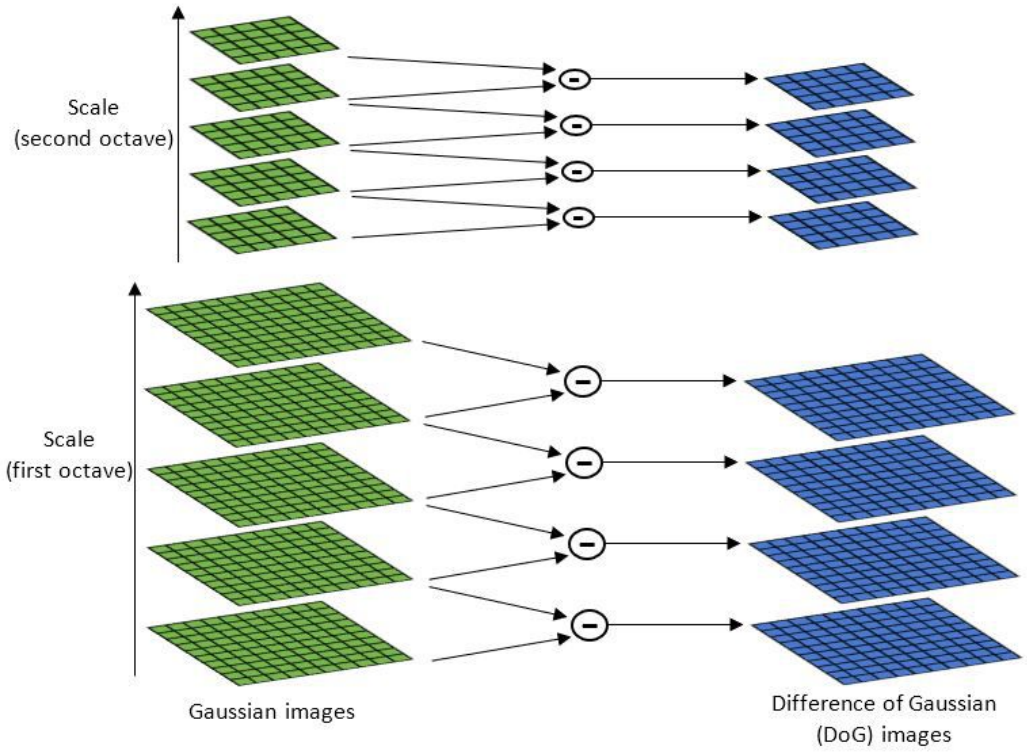


Figure 2.3: Scale-space octave pyramid for SIFT keypoint extraction. The left stack shows the Gaussian convoluted images and the right stack results from taking the difference of the consecutive images on the left.

Bypassing the second derivative calculation provides a faster way of detecting the blobs. This is the approach used by Lowe [Low04] in the keypoint detection stage for SIFT (Scale Invariant Feature Transform) descriptor. Lowe proposed to detect keypoints as the extrema of the Gaussian filtered images across a scale-space pyramid consisting of several octaves. Let σ_0 be the variance of the Gaussian kernel for obtaining the first image in the first octave. The first octave is built by successively convolving the first image with $k\sigma_0$. k is chosen to be $\sqrt{2}$. The second image in the pyramid is obtained by convolution of the first image with $k\sigma_0$, the convolution of this image once again with $k\sigma_0$ results in the third image and so on. A total of $s+3$ images are constructed for each octave with integer s representing the factor by which the image is downsampled for constructing the next octave. For constructing the next octave, the image constructed with $2\sigma_0$ in the previous octave is resampled by a factor of two (by removing every other pixel of the image). The value of s is chosen to be 2 for obtaining 5 scales per octave. Lowe showed through experiments with different number of scales that the detected keypoints become stable (i.e. the same keypoints are detected after several geometric transforms) for $s = 2$ and therefore finer scale gradation is not necessary. Fig. 2.3 illustrates this process. The image stack on the left shows the first two octaves for $s = 2$. The DoG images are obtained by subtraction

of the consecutive images in each octave (shown in the right stack). Once all the DoG images in an octave have been calculated, the local extrema are computed in 26-neighborhood for each pixel as depicted in Fig. 2.4. SIFT is able to detect several features on the skin surface, as is depicted in Fig. 2.5 for an image pair.

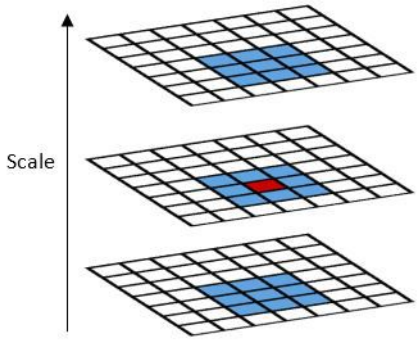


Figure 2.4: A maximum in a given DoG image is classified as a keypoint if it is also a maximum in the 26 neighborhood by including the neighboring DoG images.

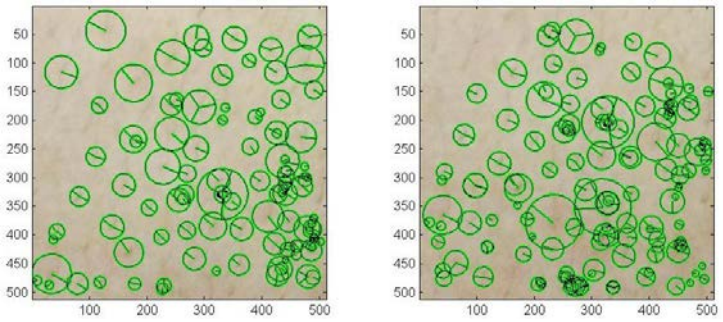


Figure 2.5: Keypoints detected using SIFT. The size of the circles indicates the scale at which the feature was detected and the lines in the circle indicates the dominant gradient orientations, a concept described in section 2.4.2. The two images are of 512×512 pixel size and are related by the homographic parameter values of $f_x, f_y, s_x, s_y = 1.01$, $\phi = -4.61°$, $\sqrt{t_x^2, t_y^2} = 105.65$ pixels and $h_1, h_2 = 3.96 \times 10^{-5}$.

2.4.1.5 Fast Hessian blob detection

Blobs in an image can also be detected through the determinants of the Hessian matrices of an image. An approximation of the Hessian matrix was used in the keypoint extraction scheme for SURF (Speeded Up Robust Features, [BTG08]). Similar to SIFT, which uses DoG approximation of LoG to speed up the keypoint detection process, Bay et al. proposed a fast approach for calculating an approximation of the Hessian of Gaussian through box models of the second derivative of Gaussian filters. Furthermore, they made use of the integral images for faster calculation. In the integral image I_Σ of an image I, the value of the pixel at location (x, y) is calculated by summing up the intensity values of all the pixels contained in the rectangle that

this pixel forms with the image origin:

$$I_\Sigma(x,y) = \sum_{i=0}^{i\leq x}\sum_{j=0}^{j\leq y} I(i,j) \tag{2.23}$$

Fig. 2.6 illustrates the use of the integral image. The sum Σ_{ABCD} of intensities of all the pixels in the rectangle ABCD in image I can be computed by simple additions/subtractions of the I_Σ values at the coordinates of the rectangle corners: $\Sigma_{ABCD} = I_\Sigma(C) - (I_\Sigma(B) + I_\Sigma(D)) - I_\Sigma(A)$. This, in conjunction with box filter, speeds up the calculation of the hessian matrix.

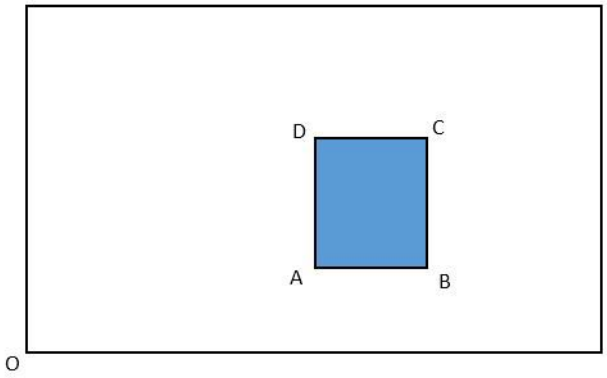
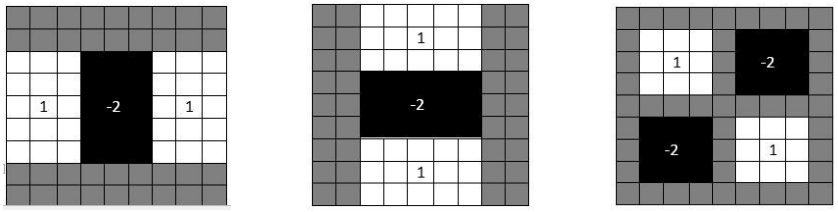


Figure 2.6: The use of inegral image makes it easy to compute the sum of intensities of the pixels bounded by the rectangle ABCD in the intensity (grey-level) image.

The Hessian matrix $\mathscr{H}(x,y,\sigma)$ of the scale-space $L(x,y,\sigma)$, is given as:

$$\mathscr{H}(x,y) = \begin{bmatrix} L_{xx}(x,y) & L_{xy}(x,y) \\ L_{xy}(x,y) & L_{yy}(x,y) \end{bmatrix} \tag{2.24}$$

L_{xx} is achieved by convolving image I with a kernel of the second derivative of Gaussian, i.e. $\frac{\partial^2 g(\sigma)}{\partial x^2}$. In the same fashion, L_{xx}, L_{xy} and L_{yy} are computed through convolution with the respective kernels with the derivatives indicated in the subscripts. The convolution operation is computationally quite expensive. This can be simplified by using box approximations of the filters. Fig. 2.7 shows the box filter approximations of the kernels of the second derivative of Gaussian.



(a) D_{xx} (b) D_{yy} (c) D_{xy}

Figure 2.7: The respective box filters for calculating D_{xx}, D_{yy} and D_{xy}. The grey areas represent a value of 2 in the box filters.

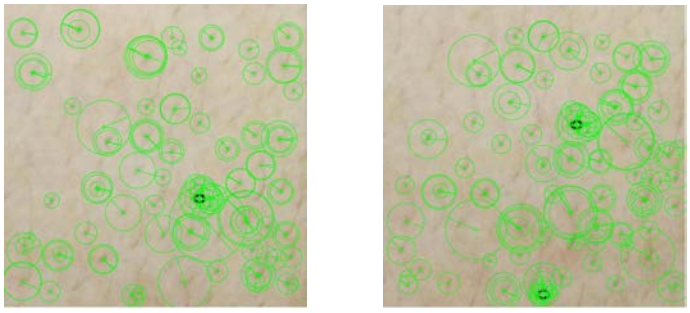
The box filters shown in Fig. 2.7 for the computation of D_{xx}, D_{yy} and D_{xy} are of dimension 9×9 and approximate the second derivative Gaussian filters of $\sigma = 1.2$. For the keypoint

detection, the determinant of the hessian matrix constructed from the box filtered images is calculated as:

$$det(\mathscr{H}_{approx}) = D_{xx}Dyy - (wDxy)^2, \tag{2.25}$$

where w, set at 0.9, is used to compensate for the approximation errors.

For the construction of the scale-space, instead of repeatedly filtering the gradually down-sampled images with the kernel of the same size, the size of the box filter is increased while applying it on the same initial image. The computation time relating to the convolution with box filter of any size remains the same due to the use of the integral images. In SURF too, like in SIFT, several octaves, each containing images filtered with increasing filter size, are built. For the first octave, the box filters used are of 9×9, 15×15, 21×21 and 27×27 dimensions. The smallest possible increase in size is 6 pixels along each axis when one pixel is added at each side of different regions of the box filters while keeping the kernel symmetric and uneven. For the construction of the second octave, the image filtered with the kernel of size 15 (second image in the previous octave) forms the first layer, the next layers are obtained by convolutions with kernels of sizes 27, 39 and 51 (the increment is doubled). The next octave is built in a similar fashion by doubling once again the increment and so on for the next octaves. The keypoints are detected as the maxima, in scale-space, of the determinants of approximated hessian matrices. Fig. 2.8 shows an example of the SURF keypoints detected on a skin image pair.

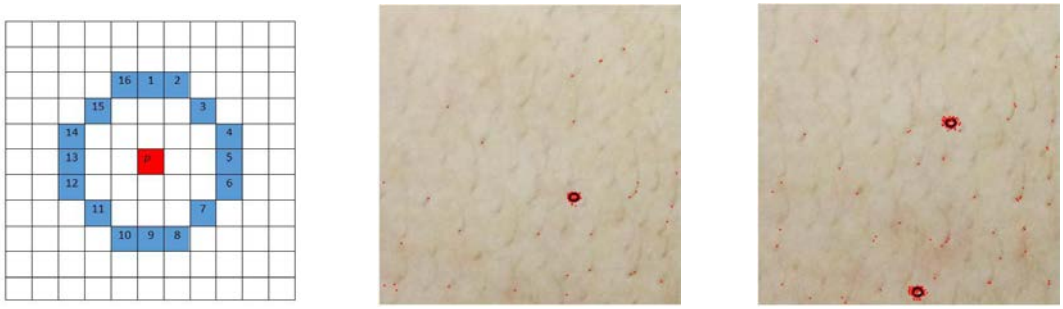


(a) I_1 (b) I_2

Figure 2.8: Keypoints detected on two skin images using the SURF approach. The size of the circles indicates the scale at which the feature was detected and the lines in the circle indicates the dominant gradient orientations, a concept described in section 2.4.2. The two images are of 512×512 pixel size and are related by the homographic parameter values of $f_x, f_y, s_x, s_y = 1.01)$, $\phi = -4.61°$, $\sqrt{t_y^2, t_y^2} = 105.65$ pixels and $h_1, h_2 = 3.96 \times 10^{-5}$.

2.4.1.6 FAST corner detector

FAST (Features from Accelerated Segment Test, [RPD10]) proposes to detect a corner based on a circle of n pixels around the candidate point (see Fig. 2.9(a) with $n = 16$). Through a series of quick tests, it is determined if any k contiguous pixels are brighter or darker than the candidate pixel p in the circle center within a tolerance defined by a threshold t. If this is the case, the candidate pixel is classified as a corner point. Figs. 2.9(b) and 2.9(c) show the keypoints detected on a skin image pair using the FAST corner detector.



(a) illustration of a 16-point circle ($n = 16$) used for corner point determination

(b) I_1

(c) I_2

Figure 2.9: FAST corner search scheme and the keypoints detected on two skin images using this approach with $k = 9$. The two images are of 512×512 pixel size and are related by the homographic parameter values of $f_x, f_y, s_x, s_y = 1.01$, $\phi = -4.61°$, $\sqrt{t_x^2, t_y^2} = 105.65$ pixels and $h_1, h_2 = 3.96 \times 10^{-5}$.

2.4.2 Feature description

The keypoints alone would be quite difficult to match. To match the keypoints unambiguously, it is helpful to take into account the surrounding regions of the keypoints. The interest of feature description is to formulate a mathematical representation of the features that can be conveniently matched using some metrics. The simplest of feature descriptor is to draw boxes of pixels around the detected keypoints. The information in the boxes around the keypoints detected in two images can subsequently be compared using metrics such as the sum of squared distances or cross-correlation. This approach, however, is not invariant to deformations like rotation or scale changes. Invariance against rotation may be achieved by selecting the circular regions around the keypoints. However, the matching results will still be affected by noise and illumination variations. Even though the effect of illumination can be reduced by normalizing the intensity values with respect to the intensity variance within the boxes or the circular patches surrounding the keypoints, this sort of description does not account for complex geometric transformations. Some of the more sophisticated and commonly used feature descriptors are discussed in this subsection.

2.4.2.1 SIFT descriptor

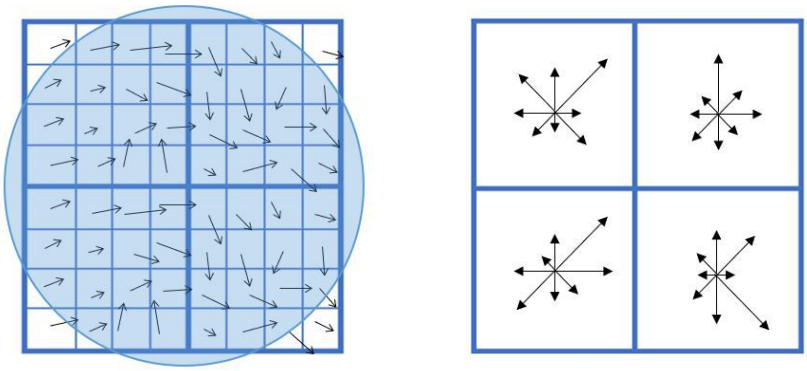
SIFT descriptor [Low04] is based on histograms of oriented gradients (HOG). HOG construction of an image patch involves computing the gradient magnitudes and directions over the pixels contained in that patch. SIFT descriptors have two main elements: the orientation assignment for the feature description vector and the feature description vector itself. To achieve rotation invariance, it is important that the feature descriptors are independent of the viewpoint changes. For assigning the orientation to the descriptor vector, the gradient magnitudes $m(x, y)$ and orientations ($\theta(x, y)$ at all the pixels of the patch surrounding the keypoint are calculated using a 4-connected neighborhood in the scale-space image L as:

$$m(x, y) = \sqrt{(L(x + 1, y) - L(x - 1, y))^2 + (L(x, y + 1) - L(x, y - 1))^2} \tag{2.26}$$

$$\theta(x,y) = tan^{-1}\left(\frac{(L(x,y+1) - L(x,y-1))}{(L(x+1,y) - L(x-1,y))}\right) \tag{2.27}$$

For determining the dominant gradient orientation, a histogram is first constructed from the gradients calculated this way in the patch around the keypoint. The gradient magnitudes are grouped according to the respective gradient angles in 36 bins of 10° intervals. Each gradient magnitude is weighted by $s\sigma$, where s is the scale at which the keypoint was detected and σ is the Gaussian smoothing factor at that scale. For assigning an orientation to the keypoint, the angle associated with the largest bin is selected. Furthermore, to account for noise and other factors that may impact the robust descriptor formulation, additional descriptors for the same keypoint are created with different orientations by selecting the bins whose values are within 80 % of the peak value.

Finally, to construct the descriptor, a 16×16 pixel window centered around the keypoint is used. For the placement of this window, the dominant orientation serves as a reference for achieving rotation invariance. The orientation of the square window is kept the same for all keypoints and the coordinates of the image pixels are rotated relative to the dominant gradient to determine which pixels fall under this window. To see it another way, the window is placed after rotating it by the angle of the dominant gradient and this amount of rotation is then subtracted from the angles of gradients of the pixels falling under this window. An orientation histogram is built for every 4×4 pixel size sub-window with bins spaced at $\pi/4$. The spacing of 4 pixels allows accounting for local variations in the position. The magnitudes of the gradients added to the bins are weighted through a Gaussian with σ half the width of the sampling window so that the gradients closer to the keypoint have more weight. Fig. 2.10 illustrates this process for an 8×8 window. With 8 bins of the histogram for each sub-window, the concatenation of all the histograms for a 16×16 window gives a descriptor of size $16 \times 8 = 128$. To reduce the impact of noise, the values of this vector are thresholded at 0.2. This way, the effect of less salient gradients can be taken into account. The final descriptor is achieved by normalizing the values to unit length.

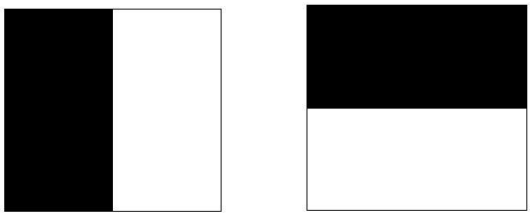


(a) Image gradients (b) Keypoint descriptors

Figure 2.10: SIFT descriptor formulation illustrated for 8×8 window. For each 4×4 sub-region, a histogram of oriented gradients, consisting of 8 bins spaced by $\pi/4$ inerval, is constructed by accumulating the Gaussian weighted gradient magnitudes in the respective bin for each orientation. The blue disc indicates a Gaussian kernal centered at the keypoint location. The gradient vectors are for illustration only and do not represent the actual magnitudes or orientations in a real image.

2.4.2.2 SURF descriptor

The feature description of SURF [BTG08] uses Haar wavelet responses instead of the oriented gradients, used by the SIFT descriptor. Before constructing the descriptor, the dominant orientation is determined. For this, a circular window of radius 6σ centered around the keypoint is used, σ corresponding to the scale at which the keypoint was detected. The Haar wavelet responses are computed in x and y directions using filters of size 4σ (see Fig. 2.11) and are weighted with a Gaussian of size 2σ centered at the keypoint. Due to the use of the integral images, the computation of these responses is very fast. For all the points, the y direction response values are plotted against the x direction responses in a Cartesian space. This space is then divided into regions separated by $\pi/3$ and the values contained in each region are summed. The region containing the largest sum is taken as the dominant direction.



(a) horizontal filter (b) vertical filter

Figure 2.11: Filters for computing Haar wavelet responses. The white region corresponds to a value of 1 and the black region to a value of -1.

For the formulation of the SURF descriptor, a rectangular window of size 16σ, oriented along the dominant direction and centered at the keypoint is used. The area under this window is divided into 16 regularly spaced sub-regions to obtain a 4×4 grid. In the region covered by the window, the Haar wavelet responses dx in the horizontal direction and dy in the vertical direction are calculated at 5×5 regularly spaced sample points. These responses in each sub-region are used to obtain a descriptor vector $v = (\sum d_x, \sum d_y, \sum |d_x| \sum |d_y|)$ for that sub-region. The concatenation of such vectors computed for each of the 16 sub-regions gives 64-dimensional SURF descriptor. If the determination of the dominant direction is omitted, a rotation invariant version called upright SURF (USURF) can be used for applications involving small rotations. Bay et al. have reported USURF to be able to tolerate rotations of up to 15°.

2.4.2.3 Binary descriptors

In contrast with SIFT and SURF, that are gradient-based descriptors, binary descriptors have the advantage of speeding up the descriptor matching process since the binary strings can be compared through fast measures, such as the Hamming distance, which is the number of places at which the values are different in the two strings. The basic concept behind the binary descriptor formulation can be described as involving the following 3 steps:

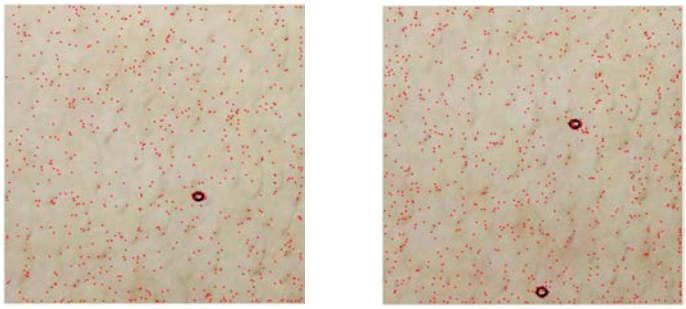
i Choice of a sampling pattern around the keypoint over which a set of point pairs $\{(\mathbf{p_i}, \mathbf{p_j})\}$, with a total of n_p, pairs is selected

ii Dominant orientation determination

iii A descriptor is formed by comparing the intensities (or some other description of the pixel

value) for each point pair. If $I(\mathbf{p_i}) > I(\mathbf{p_j})$, the descriptor vector is assigned a value of 1 for that pair and 0 otherwise.

In BRIEF (Binary Robust Independent Elementary Features), a basic binary descriptor proposed by M. Calonder et. al. in 2010 [Cal+10], the orientation assignment is not considered, and the sample pairs are selected randomly. Sample point pairs can be uniformly distributed or can be more carefully selected through various approaches proposed in [Cal+10]. This may be done, e.g., by using a Gaussian distribution so that the points closer to the keypoint are preferred. In ORB (Oriented Fast and Rotated BRIEF, [Rub+11]), a method for assigning an orientation to BRIEF descriptors is proposed, which involves calculation of the center of inertia of a moment associated with the randomly selected sample pairs. In another binary descriptor, BRISK (Binary Robust Invariant Scalable Keypoints, [LCS11]) rotation invariance is considered. BRISK detects keypoints by adaptation of the FAST keypoint detector to a scale space approach. For the descriptor formulation, BRISK takes the sample points over concentric circles around the keypoint. These sample points are randomly paired and then divided into two categories. The pairs in which the sample points are more distant than a pre-defined threshold (long pairs) are used for orientation assignment by considering the intensity variations along the directions of the vectors connecting the points in each pair. The remaining pairs (short pairs), after rotation in accordance with the orientation assignment, are used to formulate the binary descriptor. Fig. 2.12 shows an example of the BRISK keypoints detected on a skin image pair. A more recent binary descriptor FREAK (Fast REtinA Keypoint, [AOV12]) also uses a sampling pattern of concentric circles. However, the key difference with BRISK's sampling pattern is that the circles have a retinal distribution, i.e. the inner circles have a smaller difference in their radius length than the outer circles. This pattern imitates the receptor distribution on the retina by sampling more points closer to the keypoint and exponentially decreasing the sampling rate as the distance from the keypoint increases. For orientation assignment, instead of using long pairs, FREAK uses a predetermined set of 45 point pairs.



(a) I_1 (b) I_2

Figure 2.12: Keypoints detected on two skin images using the BRISK approach. The two images are of 512×512 pixel size and are related by the homographic parameter values of $f_x, f_y, s_x, s_y = 1.01$, $\phi = -4.61°$, $\sqrt{t_x^2, t_y^2} = 105.65$ pixels and $h_1, h_2 = 3.96 \times 10^{-5}$.

2.4.3 Descriptor matching

Once the descriptors around the keypoints detected in an image pair (I_i, I_{i+1}) to be registered have been formed, they need to be compared so as to find the homologous keypoint pairs. Let

$P = \{p_k | k = 1, ..., n_t\}$ represent the detected points in the image I_i, with $p_k = (x_k, y_k)$ for a total of n_t points, and $Q = \{q_l | l = 1, ..., n_s\}$ represent the detected points in the image I_{i+1}, with $q_l = (x_l, y_l)$ for a total of n_s points. The objective is to match the homologous points in P and Q. In section 1.3.6, an approach based on the direct Hausdorff distance [HKR93] was discussed for determining the homologous points from the scene coherence [Gos05]. The descriptor formulation around the keypoints provides a more direct and accurate way of determining homologous keypoints through feature matching. Let $desc_{p_k}$ and $desc_{q_l}$ represent the feature vectors formed at the keypoints p_k and q_l respectively. The similarity $S_{k,l}$ of two feature vectors can be computed through some metric (such as the Mahalanobis distance or the Euclidean norm):

$$S_{k,l} = \|desc_{p_k} - desc_{q_l}\| \tag{2.28}$$

The descriptor pairs resulting in minimum values of this norm are classified as homologous. Potentially incorrect matches can be removed by discarding the matched descriptor pairs whose similarity measure is below a threshold S_{max}. Lowe [Low04] proposed a test for measuring the quality of the match. In this test, the second-best match for a given feature is considered. Let's suppose the pair $(desc_{p_a}, desc_{q_b})$ results in the best match between a feature vector $desc_{p_a}$ in image I_i with a feature vector in image I_{i+1} and $(desc_{p_a}, desc_{q_c})$ is the second best match of the same feature vector in I_i with another feature vector in I_{i+1}. A significantly high value of $S_{a,b}/S_{a,c}$ would indicate that the detected features correspond to the objects in the image, i.e. the descriptors corresponding to these features are less likely to be an incorrect description of the object. Through several tests, Lowe showed that discarding the matches resulting in $S_{a,b}/S_{a,c}$ value of less than 0.8 significantly improved the number of the correct correspondences established. Bay et al. [BTG08] have exploited the sign of the Laplacian (which corresponds to the trace of the Hessian matrix used in SURF descriptor formulation) to add an additional matching criterion. Since this sign is opposite for dark and light blobs, a match is accepted only if the corresponding features have the same sign.

Different criteria for a quantitative analysis of the keypoint extraction and descriptor formulation approaches can be used. Some of these are "repeatability" of the keypoints (F_c), "matchability" of the feature vectors (M_c) and the ratio of the correct number of matches to the total number of matches(R_c). Repeatability (also called feature score or recall) measures a keypoint extraction approach's ability to detect the same keypoints at different geometric transformations of a given image:

$$F_c = \frac{nf_{i,i+1}}{min(nf_i, nf_{i+1})}, \tag{2.29}$$

where nf_i and nf_{i+1} are the number of keypoints detected in the images I_i and I_{i+1} respectively, the two images having some geometric transform between them, and $nf_{i,i+1}$ is the number of detected keypoints that are common to the two images.

A large F_c value is not very useful for image registration if the corresponding features cannot be matched. Matchability score measures the quality of the features formed at these keypoints:

$$M_c = \frac{nMf_{i,i+1}}{min(nf_i, nf_{i+1})}, \tag{2.30}$$

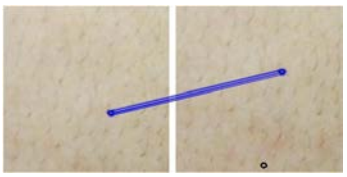
where $nMf_{i,i+1}$ is the number of matched features between the two images I_i and I_{i+1}.

Since there may be some false mathces, R_c measures the precision of the combination of a

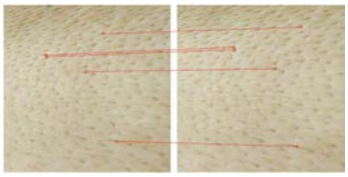
keypoint extraction scheme and descriptor formulation approach:

$$R_c = \frac{nCf_{i,i+1}}{nMf_{i,i+1}}, \tag{2.31}$$

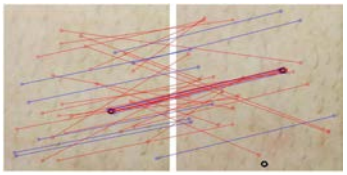
where $nCf_{i,i+1}$ is the number of correctly matched keypoints/feature vectors.



(a) BRISK descriptor matches for Harris corner points detected on (I_1, I_2). $R_c = 12/12$.

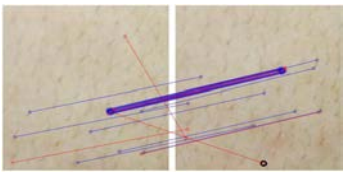


(b) BRISK descriptor matches for Harris corner points detected on (I_{32}, I_{33}). $R_c = 0/5$.

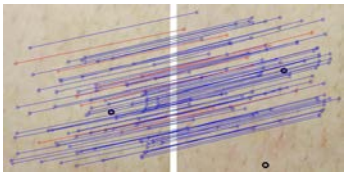


(c) BRISK descriptor matches for BRISK keypoints detected on (I_1, I_2). $R_c = 11/40$.

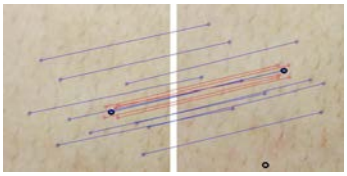
(d) FREAK descriptor matches for BRISK keypoints detected on (I_1, I_2). $R_c = 5/37$.



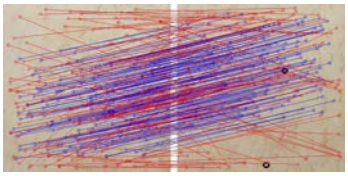
(e) BRISK descriptor matches for FAST keypoints detected on (I_1, I_2). $R_c = 21/30$.



(f) SIFT descriptor matches for SIFT keypoints detected on (I_1, I_2). $R_c = 106/114$.



(g) SURF descriptor matches for SURF keypoints detected on (I_1, I_2) with $0.0001T_S$. $R_c = 13/18$.



(h) SURF descriptor matches for SURF keypoints detected on (I_1, I_2) with $0.00001T_S$. $R_c = 160/290$.

Figure 2.13: Correspondences established by using different point extraction and feature description approaches over the first image pair of seq-II (except Fig. 2.13(b), which uses the pair (I_{32}, I_{33}) to illustrate Harris corners detected in the absence of salient markers). Blue/red lines indicate correct/incorrect matches.

Fig. 2.13 shows the keypoints matched in the first image pair $(I_1,\ I_2)$ of a simulated sequence seq-II (please refer to section 2.6.1 for details on the simulation) by using various approaches. The images in this pair are of 512×512 pixel size and are related by the homographic parameter values of $f_x, f_y, s_x, s_y = 1.01$, $\phi = -4.61°$, $\sqrt{t_x^2, t_y^2} = 105.65$ pixels and $h_1, h_2 = 3.96 \times 10^{-5}$. For each result shown, the R_C values are given in the respective sub-figure caption. The matches within $0.4S_{max}$ were retained for all the results shown. The blue/red lines indicate the correct/incorrect matches (matches are classified as incorrect if the coordinates of a keypoint in one image mapped into the other image by ground truth homography result in more than one pixel error). Less salient keypoints can be eliminated by thresholding the detected keypoints. This is achieved by discarding the detected keypoints below a fraction of C_{max}, the maximum value of the detected keypoints. In the results shown, Harris, FAST and BRISK keypoints were thresholded respec-

tively at $0.05C_{max}$, $0.1C_{max}$ and $0.1C_{max}$. In a similar fashion, the keypoints detected by SIFT and SURF can be thresholded at a fraction of the maximum T_S of the respective scale-spaces (DoG or Hessian). The keypoints for the results shown in Figs. 2.13(f), 2.13(g) and 2.13(h) were thresholded at $0.001T_S$, $0.0001T_S$ and $0.00001T_S$ respectively. It can be noticed that a lower threshold increases the number of keypoints detected. However, this reduces the R_C value by increasing the incorrect matches. As can be noticed, not all the keypoint detection methods detected the same keypoints. Besides, the ratio of the correctly matched features is different for different methods. The Harris corner detector detected only the most salient points on the black marks in the image pair (I_1, I_2). All of these points were correctly matched by formulating the BRISK descriptors at these points (see Fig. 2.13(b)). Although the Harris corner detector was able to detect some less salient points on the skin surface when the markers were not present (see Fig. 2.13(b)), they were not homologous. A combination of the BRISK keypoints with the BRISK descriptor approach resulted in 11 correct matches out of the 40 matches established. The use of the FREAK descriptor for matching the BRISK keypoints had a lower success rate with $R_C = 5/37$, whereas the use of the BRISK descriptor for matching the keypoints detected with the FAST keypoint detector had a relatively higher success rate with $R_C = 21/30$. The results of the SIFT-based approach are the most successful with $R_C = 106/114$, whereas the SIFT-based approach has results closer to the ones obtained with the FAST/BRISK combination. Given this large variation in the results, the performance of different approaches will be compared in section 2.6 over sequences containing several image pairs to evaluate their consistency in successful skin image registration.

2.5 A combined feature and OF based approach

In [BM11], Brox and Malik have proposed an OF calculation approach, referred to as large displacement optical flow (LDOF), for integrating feature descriptors into a continuous optimization framework. Optimization in the LDOF approach is also carried out in the variational framework. However, in contrast to the TV-L^1 approach, a pseudo L^1-norm function $\psi(q^2) = \sqrt{q^2 + \epsilon^2}$, with $\epsilon = 0.001$ is used to deal with the quadratic penalization problem. LDOF is an extension of the high accuracy OF (HAOF) by Brox et al. [Bro+04]. HAOF energy formulation, that incorporates a gradient constancy assumption to deal with strong illumination variations, is:

$$E_{HAOF}(\mathbf{u}) = \underbrace{\int_\Omega \psi\left(\|I_{i+1}(\mathbf{x}+\mathbf{u}) - I_i(\mathbf{x})\|_2\right) d\Omega}_{E_{int}} + \alpha \underbrace{\int_\Omega \psi\left(\|\nabla I_{i+1}(\mathbf{x}+\mathbf{u}) - \nabla I_i(\mathbf{x})\|_2\right) d\Omega}_{E_{grad}} + \lambda \underbrace{\int_\Omega \psi\left(\|\nabla \mathbf{u}\|_2(\right. d\Omega}_{E_{reg}}, \quad (2.32)$$

with α and γ being the weight coefficients of their respective terms.

While retaining the conventional OF gradient smoothness constraint as the regularization term, LDOF combines a gradient constancy term (E_{grad}) with the intensity constancy term (E_{int}). Two additional interdependent terms, (E_{match} and E_{desc}), for HOG (Hessian of Gaussian) descriptors constancy are added to formulate the LDOF energy [BM11]:

$$E_{LDOF}(\mathbf{u}) = E_{HAOF}(\mathbf{u}) + \beta \underbrace{\int_{\Omega} \delta(\mathbf{x})\rho(\mathbf{x})\psi\left(\|\mathbf{u}(\mathbf{x}) - \mathbf{w}(\mathbf{x})\|_2\right) d\Omega}_{E_{match}} + \underbrace{\int_{\Omega} \delta(\mathbf{x}) \|F_{i+1}(\mathbf{x}+\mathbf{w}) - F_i(\mathbf{x})\|_2 \, d\Omega}_{E_{desc}}, \quad (2.33)$$

where E_{match} minimizes the difference between the optical flow vector $\mathbf{u}$ and the correspondence vector $\mathbf{w}$ obtained by descriptor matching at point $\mathbf{x}$ depending on the value of $\delta(\mathbf{x})$, which is 1 in the presence of a descriptor at point x and 0 otherwise. $\rho(\mathbf{x})$ is a weight assigned through a quantitative measure of the descriptor match success. E_{desc} minimizes the feature correspondence vector $\mathbf{w}$ over the sparse fields (F_i and F_{i+1}) of the feature vectors in the respective image. β is the weight coefficient of the E_{match} term. The solution of Eq. (2.33), as detailed in [BM11], results in a nested set of Euler-Lagrange equations. In this study, the optimization was performed over 10 iterations in the outer and 5 iterations in the inner loop with a down-sampling factor of 0.98 for the multi-resolution OF scheme. The coefficient values used were: $\alpha = 5$, $\beta = 300$ and $\lambda = 30$.

2.6 Comparison of Methods: Results and Analysis

In this section, comparisons of various registration approaches from the frameworks discussed above are presented. The objective was to settle on an approach that provided the best compromise between computation time and accuracy. For these tests, since it is difficult to establish the ground truth on the real video sequences, some displacement sequences were numerically simulated from a high-resolution real skin image. These sequences provide a texture that is representative of the texture captured in the real sequences. Thus, apart from providing a ground truth, they can serve as a reference for evaluating different approaches for skin image registration. The comparative results for some approaches are presented after describing the simulation scheme for these sequences. After settling upon feature-based approaches. SIFT, SURF and BRISK will be further compared on several grounds.

2.6.1 Simulated Sequences

The simulation scheme involves choosing a trajectory shape along which sub-images are extracted from a large high-resolution image. The transformation parameters between two consecutively extracted images can be defined based on certain pre-selected criteria. After selecting a frame size ($P \times Q$) for the sequence to be simulated, the initial image is extracted at the starting point of the trajectory without any transformation. For the rest of the frames, the homographies accumulated up to a frame are used to extract that frame. For the extraction, two approaches can be used: i) the size of the extraction region is fixed, and the coordinates of this region are transformed using the homography of the corresponding frame ii) the size of the frame is fixed, and the coordinates of the frame are back-projected, with the inverse homography, onto the high-resolution image. In both cases, the projected coordinates do not necessarily fall on integer coordinates of the Cartesian grid. So, a scheme needs to be used for assignment of the intensities at the projected coordinates.

Forward warping results in the projection of coordinates of a pixel from the source image to the destination image. The intensity of this pixel then needs to be extrapolated onto the neighboring pixels in the source image. However, apart from the complications of the extrapolation, this is difficult to deal with since, depending on the transformation parameters, the distance ratios between the coordinates of the two images may vary arbitrarily. The backward warping provides a more consistent solution. The back-projection of the coordinates of a pixel in the destination image will either be on the image or outside the image (the high-resolution image is chosen to be large enough and the extraction trajectory is chosen carefully to avoid the outside cases). Now the question is to assign a value, based on the intensity values of the neighboring pixels of the projected coordinate, to the pixel at this coordinate in the destination image. The simplest would be to assign the intensity of the closest neighbor:

$$I_d(p,q) = I_s(round(p'), round(q')), \tag{2.34}$$

where (p', q') are the projected coordinates, in the source image I_s, of a pixel (p, q) in the destination image I_d.

A more sophisticated approach is to interpolate the intensity values from several of the neighboring pixels. A commonly used such approach is bilinear interpolation. This is illustrated through Fig. 2.14. The intensity at location a is found by horizontal interpolation of intensities at coordinates (p_1, q_1) and (p_2, q_1):

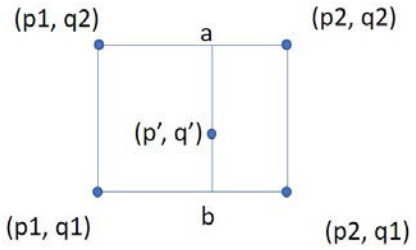


Figure 2.14: Illustration of bilinear interpolation where the intensity at the position (p', q') is interpolated from the values of the 4 closest neighboring pixels.

$$I_s(a) = \frac{p_2 - p'}{p_2 - p_1} I_s(p_1, q_1) + \frac{p' - p_1}{p_2 - p_1} I_s(p_2, q_1) \tag{2.35}$$

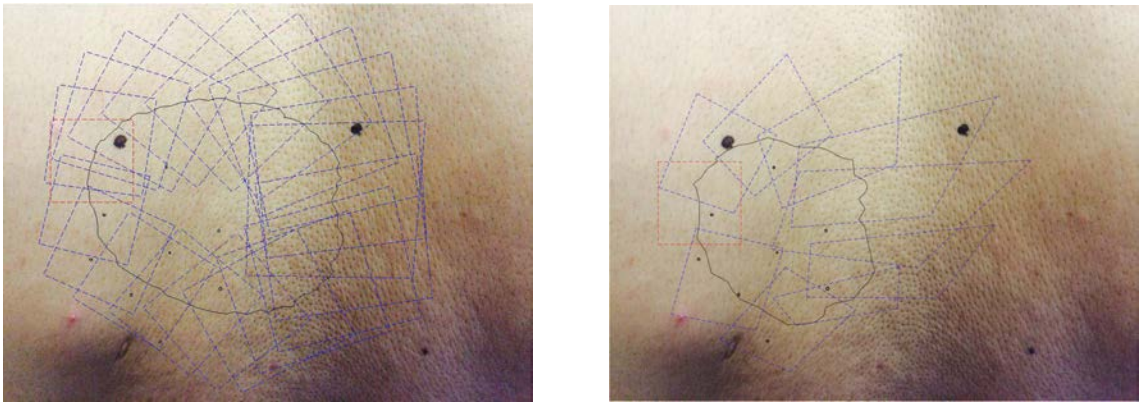
Similar interpolation between intensities at coordinates (p_1, q_2) and (p_2, q_2) gives the intensity at location b

$$I_s(b) = \frac{p_2 - p'}{p_2 - p_1} I_s(p_1, q_2) + \frac{p' - p_1}{p_2 - p_1} I_s(p_2, q_2) \tag{2.36}$$

The intensity at location (p, q) is then allocated by interpolating the intensity at locations a and b in the vertical direction:

$$I_s(p', q') = \frac{q_2 - q'}{q_2 - q_1} I_s(a) + \frac{q' - q_1}{q_2 - p_1} I_s(b) \tag{2.37}$$

Two simulated video sequences (seq-I and seq-II) were obtained by extracting overlapping patches, with a total size of $P \times Q = 512 \times 512$ pixels each, from the high resolution image of Fig. 2.15 which has a size of 3264×2448 pixels and corresponds to a human dorsal area of about



(a) seq-I (b) seq-II

Figure 2.15: Illustration of the seq-I and seq-II simulation path on a high resolution image of 3264×2448 pixel size and corresponding to a human dorsal area of about 20×15 cm^2. The red square represents the first and the last patch. Blue boxes delineate the zones corresponding to every fifth of the rest of the patches extracted along the black trajectory line. The homographic relation between consecutive patches is constrained by the values given in Table 2.1. All patches are mapped onto 512×512 pixels frames using the known (simulated) homographies.

20×15 cm^2. The displacement of the consecutive patches is determined by the homographic parameters within the ranges delimited by the values given in Table 2.1. The parameters, generated randomly with a uniform distribution, are such that the homographies lead to a smooth displacement of the images in the simulated sequence and form a close-looped elliptic trajectory (see the black continuous line in Fig. 2.15). The objective was to obtain sets of known displacements while imitating realistic movements of the acquisition device, namely without irregular variations and sporadic changes in direction. The last image in the sequence is the same as the first one. This helps, through a visual inspection, in finding out the distortion in the mosaic. Somewhat extreme values of the parameters in the seq-II permit to accentuate the performance differences between the compared approaches and help to determine their limits.

Table 2.1: Homography parameter intervals used for computing the displacements between consecutive images of the simulated sequences seq-I (Fig. 2.15(a)) and seq-II (Fig. 2.15(b)). The homographic parameters (s_x, S_y, f_x, f_y, t_x, t_y, $h_{3,1}$, $h_{3,2}$ and θ) are defined in section 1.3.6 for Eq. (1.35).

Seq.	**Value**	$s_x, s_y,$ f_x, f_y	θ **(degrees)**	$\sqrt{t_x^2 + t_y^2}$ **(pixels)**	$h_{3,1}, h_{3,2}$ $(\times \mathbf{10^{-6}})$
I	min	0.94	$\pm$ 0.05	16.94	$\pm$ 0.00
	max	1.04	$\pm$ 6.70	66.12	$\pm$ 7.70
	mean	1.00	$\pm$ 1.81	40.15	$\pm$ 1.80
II	min	0.91	$\pm$ 0.21	38.72	$\pm$ 0.70
	max	1.12	$\pm$ 28.83	137.97	$\pm$ 135.5
	mean	1.00	$\pm$ 5.07	83.74	$\pm$ 37.6

A quantitative evaluation of the registration process can be achieved since the homographic relationships linking the extracted images are known. Two evaluation criteria are used: registration error $\varepsilon_{i,i+1}^{reg}$ and mosaicing error $\varepsilon_{1 \leftarrow N}^{mos}$, with their respective definitions given in Eqs. (2.38) and (2.39).

$\varepsilon^{reg}_{i,i+1}$ is the mean Euclidean distance between corresponding pixels of two images: one warped with the estimated homography $H^{est}_{i,i+1}$ and the other with the true homography $H^{true}_{i,i+1}$. The same metric is used for the $\varepsilon^{mos}_{1\leftarrow N}$ calculation, except that the estimated homography results now from the concatenation of the homographies of all the consecutive image pairs in the sequence. The last and the first images being the same in the sequence, the true homography is the identity matrix in this case.

$$\varepsilon^{reg}_{i,i+1} = \frac{1}{P \times Q} \sum_{\mathbf{c}_{p,q}=(1,1)}^{(P,Q)} \left\| \frac{H^{true}_{i,i+1}\mathbf{c}_{p,q}}{\alpha^{true}_{i,i+1}} - \frac{H^{est}_{i,i+1}\mathbf{c}_{p,q}}{\alpha^{est}_{i,i+1}} \right\| \tag{2.38}$$

$$\varepsilon^{mos}_{1\leftarrow N} = \frac{1}{P \times Q} \sum_{\mathbf{c}_{p,q}=(1,1)}^{(P,Q)} \left\| \frac{H^{true}_{1\leftarrow N}\mathbf{c}_{p,q}}{\alpha^{true}_{1\leftarrow N}} - \frac{H^{est}_{1\leftarrow N}\mathbf{c}_{p,q}}{\alpha^{est}_{1\leftarrow N}} \right\| \tag{2.39}$$

In Eqs. (2.38) and (2.39), $\mathbf{c}_{p,q}$ represents the homogeneous coordinates $(p, q, 1)$ corresponding to a 2D Cartesian grid of size $P \times Q$. The subscripts of the homography matrices H indicate the image indices, with $\leftarrow$ representing the concatenation of homographies. α, with its subscripts indicating the corresponding homography matrix, represents the perspective scale factor of the transformed coordinates.

2.6.2 Comparison of various image registration approaches

At first, registration approaches from very different frameworks (OF, feature-based and the combination of the two) are compared to establish a general trend in speed and precision. This was the subject of a preliminary study [Far+16]. The approaches selected for this purpose are the Chambolle and Pock's TV-L^1 based approach [CP11] and the Drulea and Nedevschi's correlation transform based approach (Correlation Flow [DN13]) in the OF category. For the feature-based approaches, BRISK [LCS11], SIFT [Low04] and SURF [BTG08] are opted for. Large displacement OF (LDOF), which combines the optical flow calculation with the descriptor matching, is also considered. For Correlation Flow, apart from the titular correlation transform, the results for census transform and an intensity transform, which simply selects an intensity patch at a given pixel location, are also evaluated.

Table 2.2 gathers the results obtained using the described methods. Although it might be interesting to compare the pairwise registration accuracy (a comparison which is done later for SIFT and SURF and BRISK), for the mosaicing purposes, consistently successful registration, under varying transformation parameters and image characteristics such as texture and illumination, is equally important. Due to this observation, the success of these approaches is described by grouping the registration results according to certain precision intervals. Sub-half-pixel precision (i.e. $\varepsilon^{reg}_{i,i+1} \leq 0.5$) is considered very accurate for the mosaicing purpose (interval **A**). Errors up to 1 and 2 pixels (intervals **B** and **C**) may be acceptable in limited numbers. An error greater than 2 pixels (interval **D**) is considered as indicative of a failed registration.

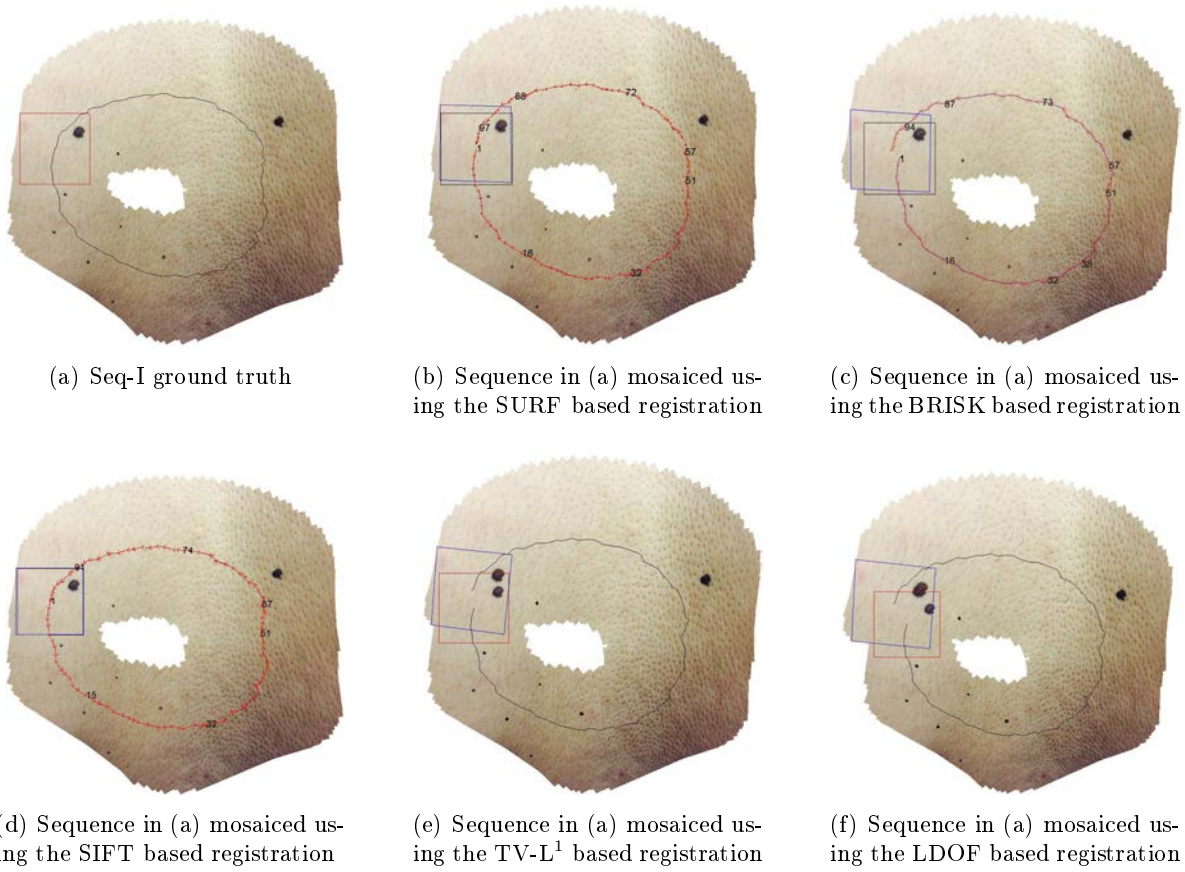
Fig. 2.16 shows seq-I mosaics obtained using various approaches, along with the ground truth mosaic. For this sequence, TV-L^1 and SIFT-based registrations are largely the most successful, with almost all the image pairs registered with less than half pixel error (for both cases, the pairs falling in the interval **B** have errors just slightly above half a pixel ($<$0.6)). LDOF and Correlation Flow (with various transforms) have a considerably lower success rate. For Correlation Flow, it should be noted that this method is designed for extreme illumination variation cases and does so at the expense of accuracy in direct pixel-by-pixel correspondence establishment. Besides, these

Table 2.2: Number of image pairs falling in a given $\varepsilon_{i,i+1}^{reg}$ error range (in pixels) among four intervals **A** = [0.0, 0.5], **B** = (0.5, 1.0], **C** = (1.0, 2.0] and **D** = (2.0, $+\infty$), given for each of the two sequences (seq-I and seq-II) for various registration approaches. The mosaicing error and the registration time per image pair are given as well. Computation times are for an Intel® Xeon® 2.10GHz processor with execution in a MATLAB® environment. All implementation, except LDOF, Correlation Flow and SIFT, which have their computationally expensive routines implemented in C++, are in MATLAB programming language. Registration time averaged over a few pairs is given for descriptor based approaches since it varies depending on the number of keypoints detected.

Seq.	**Method**	**No. of image pairs with $\varepsilon_{i,i+1}^{reg}$ in a given interval**				$\varepsilon_{1\leftarrow N}^{mos}$ **(pixels)**	t (s)	**Result shown in Fig.**
		A	**B**	**C**	**D**			
I	BRISK[LCS11]	50	36	14	0	77	0.13	2.16(c)
	SURF[BTG08]	90	7	2	1	38	1.1	2.16(b)
	SIFT[Low04]	99	1	0	0	7.1	5	2.16(d)
	TV-L1[CP11]	97	3	0	0	117	43	2.16(e)
	LDOF[BM11]	73	22	4	1	356	195	2.16(f)
	(Dr_BC)[DN13]	82	12	5	0	93	287	–
	(Dr_FC)[DN13]	68	27	5	0	100	629	–
	(Dr_IC)[DN13]	49	20	16	15	203	388	–
	(Dr_CT)[DN13]	60	31	5	4	171	286	–
II	BRISK[LCS11]	13	26	9	2	223	0.13	2.17(c)
	SURF[BTG08]	34	12	4	0	25	1.1	2.17(b)
	SIFT[Low04]	49	1	0	0	13	5	2.17(d)
	TV-L1[CP11]	41	6	0	3	686	43	2.17(e)
	LDOF[BM11]	1	18	20	11	356	195	2.17(f)

two approaches are computationally quite expensive (several minutes for registering just a single pair of relatively low-resolution images). TV-L^1 is significantly faster than these two, but the descriptor based approaches provide an even more computational advantage. Among the tested descriptor-based approaches, with only half of the image pairs registered with sub-pixel accuracy, BRISK turns out to be noticeably less successful than SURF and SIFT in registration precision. Although SURF registered 90 pairs with sub-pixel error, a few less precise registrations and one "failed" registration cause some distortion in the mosaic. In terms of computation time, BRISK largely outperforms the other approaches. However, with only 50 % registrations in interval **A**, it is quite less accurate.

The results of Seq-I place SIFT and TV-L^1 as the best candidates for the skin image mosaicing. To further evaluate these approaches, they are tested on Seq-II, which involves larger transformation parameters. Fig. 2.17 shows seq-II mosaics obtained using various image registration schemes. With a visual comparison against the ground truth mosaic, it is clear that SURF and SIFT produce quite coherent and minimally distorted mosaics. The mosaic obtained through TV-L^1 based approach shows distinct distortion, where not only the loop does not close, but also the last image, which has the same dimensions as the first one, is largely distorted. The results for LDOF are shown as well to illustrate another case of a distorted mosaic of the same sequence but with different registration accuracy of the individual pairs. Although for seq-I the results of SIFT and TV-L^1 were comparable, for seq-II, SIFT largely outperforms the latter



(a) Seq-I ground truth

(b) Sequence in (a) mosaiced using the SURF based registration

(c) Sequence in (a) mosaiced using the BRISK based registration

(d) Sequence in (a) mosaiced using the SIFT based registration

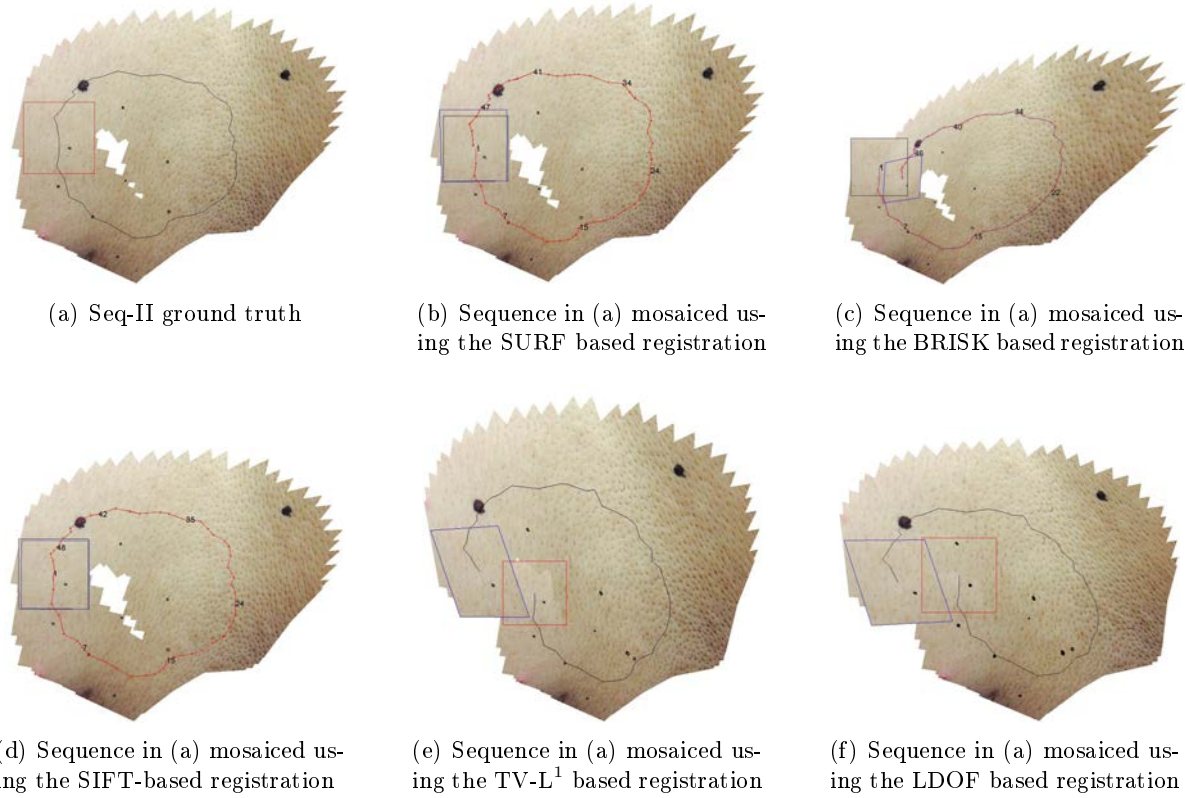
(e) Sequence in (a) mosaiced using the TV-L^1 based registration

(f) Sequence in (a) mosaiced using the LDOF based registration

Figure 2.16: Seq-I ground truth alongside the mosaics obtained using different methods. The initial frame is marked with a red square, with the last frame indicated with a blue quadrangle. While these two frames coincide in the ground truth image, their displacement is a measure of global coherency of the mosaiced image. The sequence trajectory, which forms a closed loop in the ground truth, is traced with a black or red-arrowed line. This trajectory along with the shape of the circumscribed white region as well as the black marks placed on the image are helpful for a visual assessment of the mosaic.

by registering 49 pairs with high accuracy, with just one pair in interval **B**. Although TV-L^1 remains more precise, in general, than SURF (with 41 pairs in interval A for the former against 34 in the same interval for the latter), it has failed registration for 3 pairs. LDOF also results in several failed registrations. Among the descriptor based approaches, SURF is somewhat less precise than SIFT since SURF results in considerably less precise registrations. BRISK is once again less successful than the other two descriptor-based approaches.

In general, it is evident that SIFT results in the most successful, consistent and precise registrations. SURF is less precise, however, given its considerably less execution time, it is not excluded from consideration. Besides, the success of the pairwise registrations is not the only factor influencing the mosaic construction since the eventuality of failed registrations, due to large camera displacements or blur, for example, always exists. A practically useful solution would be to arrive at a compromise between the speed and accuracy. Moreover, once a scheme that takes into account the failed registrations and accumulation of errors over large trajectories has been developed, SIFT and SURF can be used interchangeably, depending on the time constraints for



(a) Seq-II ground truth

(b) Sequence in (a) mosaiced using the SURF based registration

(c) Sequence in (a) mosaiced using the BRISK based registration

(d) Sequence in (a) mosaiced using the SIFT-based registration

(e) Sequence in (a) mosaiced using the TV-L^1 based registration

(f) Sequence in (a) mosaiced using the LDOF based registration

Figure 2.17: Seq-II ground truth alongside the mosaics obtained using different methods. The initial frame is marked with a red square, with the last frame indicated with a blue quadrangle. While these two frames coincide with each other in the ground truth image, their displacement is a measure of global coherency of the mosaiced image. The sequence trajectory, which forms a closed loop in the ground truth, is traced with a black or red-arrowed line. This trajectory along with the shape of the circumscribed white region as well as the black marks placed on the image is helpful for a visual assessment of the mosaic.

a particular mosaicing task.

The mosaicing error criterion used may give a rough estimate of the accuracy of the mosaic if the local registrations are precise in all the consecutive registrations. However, it does not, in general, prove to be a good indicator of the global accuracy, as is evident from the comparison of the SURF based and TV-L^1 based mosaics of seq-I: mosaicing error in the former is smaller even though TV-L^1 based registration is more precise on average. In the results obtained through SIFT, this criterion corresponds well with the visual coherence of the obtained mosaics. In seq-I, this results in a closed-loop and an almost perfect overlap of the first (I_1) and the last (I_N) images ($\varepsilon_{i,N}^{reg}$ being just 7.1 pixels). The overall shape of the obtained mosaic is very close to that of the ground truth mosaic as well. For the mosaic of Seq-II too (with $\varepsilon_{i,i+1}^{reg} = 13$ pixels), SIFT results in an almost closed-loop with a very slight displacement between the first and the last images. The shape of the empty encircled regions does indicate some small distortion in the overall shape of the mosaic. The mosaicing error has even less meaning for the mosaics obtained through TV-L^1 and LDOF for the second sequence. These mosaics contain several registrations with large errors (interval **D**). Besides, large distortions in the mosaics and several duplicate markers, due to failed registrations, can be noticed. The seq-II mosaic obtained through BRISK

illustrates a case where some failed registration can create large distortions in the mosaic, or even interrupt the mosaicing process. This sort of scenarii motivate us to deeply focus on the overall mosaicing process rather than on improving the registration precision and consistency.

2.6.3 BRISK, SURF or SIFT?

Descriptor-based approaches were found to be considerably faster than the optical flow based ones. Although SIFT shows the highest success rate in point matching and precision, it is more closely compared with SURF and BRISK to find out to what extent, with some compromise and adjustments, the latter can be exploited. This interest in this study comes from the observation that SURF is almost five times faster than SIFT. In addition, the computation time for SIFT increases drastically with an increase in the number of detected keypoints. Even though BRISK turned out to be considerably less precise than both SIFT and SURF, it too is retained as a potential choice for mosaicing the real video sequences due to its very low computation time (just 0.13 s/pair). It should be noted that the real sequences used in this study have high-resolution images with 1294×964 pixel size each. Since high-resolution images provide with more discernable features, BRISK is likely to establish more correspondences on the real sequences. The speed *vs* accuracy compromise for making a choice between SURF and SIFT, along with ASIFT (Affine SIFT), was also considered in [Kas+13] for mosaicing of skin images with the objective of monitoring the scar evolution after tumor excision.

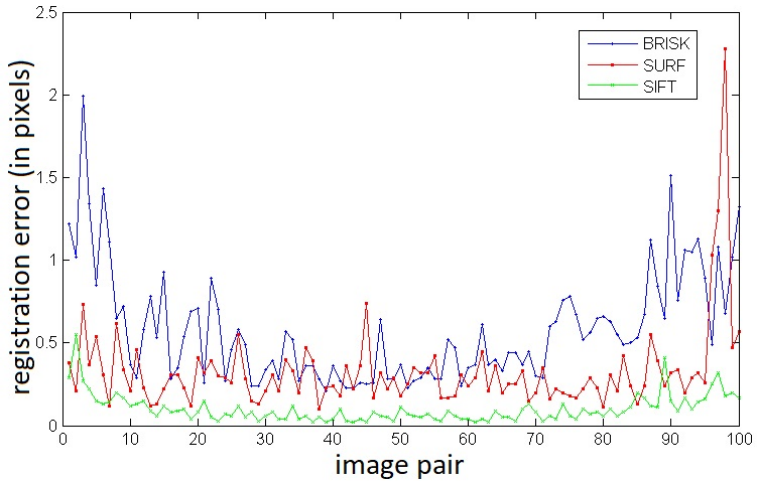
Table 2.3 reproduces the quantitative results, including some additional statistics on the pairwise registration precision, of the three descriptor-based approaches for mosaicing seq-I and seq-II. SIFT has the smallest mean registration error for both sequences, just one-tenth of a pixel for seq-I and one-fifth of a pixel for seq-II. Although the largest error in both cases is more than half a pixel, the majority of the registration errors do not fall far from the mean, as is indicated by very low standard deviations (0.08 and 0.12 respectively) of the errors for the two sequences. For the SURF approach, although a large majority of registrations for image pairs of seq-I led to sub-pixel errors with a mean of 0.33 pixels and a standard deviation of 0.26 pixels, for seq-II the registration failed for 3 image pairs (error in interval **D**). This indicates that SURF can be used over sequences acquired with the slow and smooth motion of camera to avoid, as much as possible, large and abrupt displacements between the images. SIFT appears to be more robust in case of large displacements.

Figs. 2.18(a) and 2.18(b) show, for seq-I and seq-II respectively, plots of the pairwise registration errors $\varepsilon_{i,i+1}^{reg}$ (2.38) for SIFT, SURF and BRISK. In almost all the image pairs, SIFT results are clearly more precise than those of the other two methods. Although BRISK was less

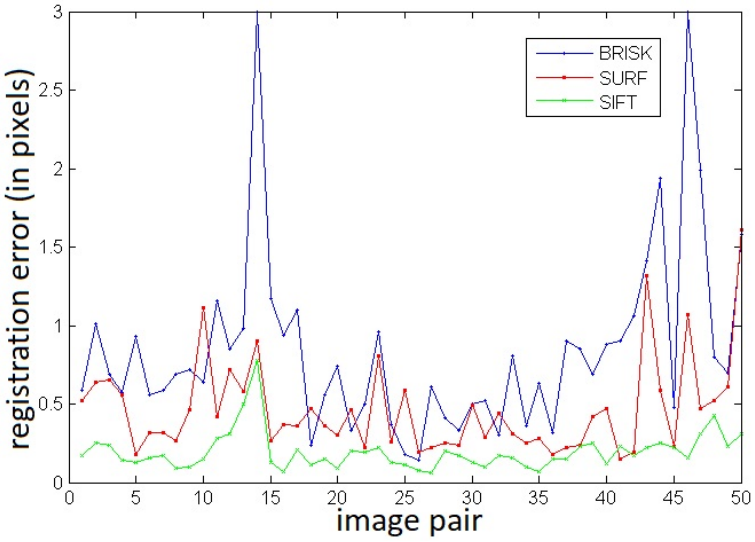
Table 2.3: Comparative results of SIFT, SURF and BRISK

Seq.	**Method**	**No. of image pairs with $\varepsilon_{i,i+1}^{reg}$ in a given interval**				$\varepsilon_{1\leftarrow N}^{mos}$ **(pixels)**	$\varepsilon_{1\leftarrow N}^{reg}$**(pixels)**				t (s)	**Result shown in Fig.**
		A	**B**	**C**	**D**		**min**	**max**	**mean**	σ		
I	BRISK	50	36	14	0	77	0.21	1.99	0.58	0.34	0.13	2.16(c)
	SURF	90	7	2	1	38	0.1	2.28	0.33	0.26	1.1	2.16(b)
	SIFT	99	1	0	0	7.1	0.02	0.55	0.10	0.08	5	2.16(d)
II	BRISK	13	26	9	2	223	0.14	4.15	0.87	0.72	0.13	2.17(c)
	SURF	34	12	4	0	25	0.15	1.61	0.46	0.30	1.1	2.17(b)
	SIFT	49	1	0	0	13.1	0.06	0.78	0.20	0.12	5	2.17(d)

precise, it successfully registered all the image pairs in seq-I. The image pairs at which it had less successful registrations are either in the beginning or at the end of the sequence, corresponding to the left zone of the ground truth mosaic (Fig. 2.16(a)). It can be noticed that this zone has relatively smooth texture, with less prominent skin pores. Pairwise registration precision plot of seq-I shows that BRISK was almost as accurate as SURF between I_{25} and I_{70}, the images corresponding to the high texture zone. This observation provides support for the potential effectiveness of BRISK on sequences containing high-resolution images with small displacement. Even though SURF serves as the method of choice in this study, BRISK's effectiveness and limitations in mosaicing real sequences will be explored in section 3.2.5.



(a) seq-I



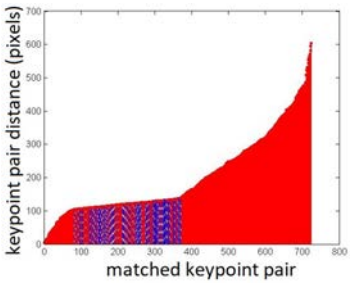
(b) seq-II

Figure 2.18: Comparison of pixel-wise registration errors resulting from using BRISK, SURF and SIFT on seq-I and seq-II.

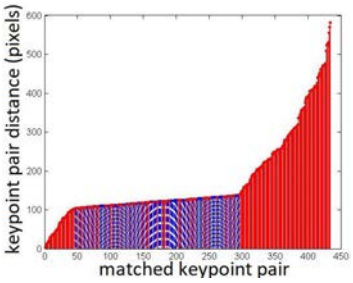
2.6.4 Refinement of the detected correspondences

As was discussed in section 2.4.3, more correspondences can be detected with SURF if keypoints are thresholded at a lower fraction of T_S value (it is recalled that T_S is the maximum of the Hessian filter responses in the scale-space). The correspondences shown in Figs. 2.13(g) and 2.13(h) were obtained with thresholding at $0.0001T_S$ and $0.00001T_S$ respectively. Reducing this threshold significantly increased the number of less salient keypoints detected. At the same time, although more correct matches can be found, lower thresholding also largely increased the number of incorrect matches. SIFT, in general, has a high ratio of correct to incorrect matches as can be observed in Fig. 2.13(f). If the ratio of the correct matches is high, multiple trial approach of the RANSAC algorithm is more likely to calculate a correct homography. However, in the presence of a large number of mismatches, RANSAC may fail to give correct results. It was observed in several experiments that RANSAC's outcome varied for some pairs when it was applied multiple times over the same keypoint matches. Certain RANSAC results succeeded in calculating a close to correct homography while other RANSAC outcomes led to the failure of the homography determination.

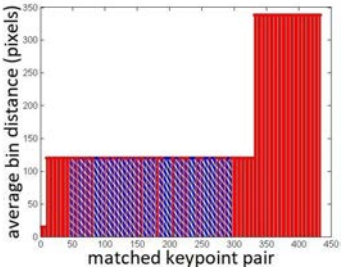
It would be useful to pre-filter the established matches to eliminate potentially false correspondences. In the mosaicing results discussed earlier for the SURF based approach, 30 best matches were retained based on the Euclidean norm $\|desc_{p_k}, desc_{q_k}\|$ of the matched descriptors, with $desc_{p_k}$ and $desc_{q_k}$ representing the feature vectors at the matched keypoints p_k and q_k in the two images. A low value of this norm indicates a strong match between the descriptors. After this selection, the ratio of correct matches may still stay low. Another pre-filtering can be implemented for mosaicing applications by considering the fact that almost all the pixels are displaced in more or less homogeneous manner, i.e. the Euclidean distance $\|p_k, q_k\|$ between



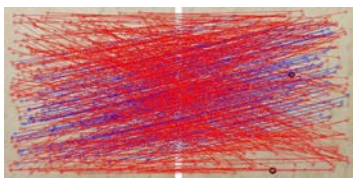
(a) Increasing spatial distance between the initially matched keypoints



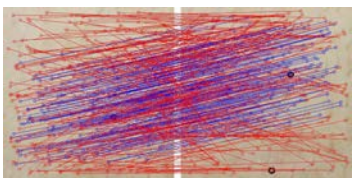
(b) Increasing spatial distance between the matched keypoints thresholded at $0.6S_{max}$



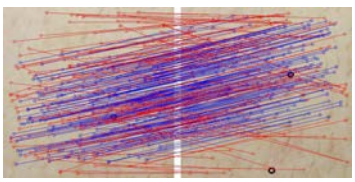
(c) Thresholded keypoint pairs grouped into intervals according to their spatial distance variation



(d) Correspondances of the initially matched keypoints (R_c = 185/724)



(e) Correspondences of the matched keypoints thresholded at $0.6S_{max}$ ($R_c = 179/434$)



(f) Correspondences falling in the largest flat interval of the above plot ($R_c = 179/322$)

Figure 2.19: Selection of best matches from the keypoint correspondences established on the first pair of seq-II using SURF with $0.00001T_S$. The red lines indicate false matches.

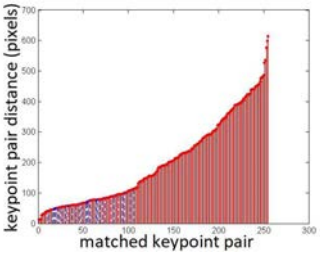
the correctly matched keypoint coordinates does not vary radically. This implies that if the established correspondences are sorted by their distances (i.e. from the smallest Euclidean distance to the largest one between the matched keypoints), the correct matches form roughly a horizontal line segment in a distance-plot (on the x-axis of such a plot, the matched keypoint pairs are ordered according to their increasing distance values, with the y-axis indicating the corresponding distance values). From this observation, the correct matches can be selected by discarding the detected matches at both sides of this roughly flat segment. This is illustrated in Figs. 2.19 and 2.20 for image pairs (I_1, I_2) and (I_{13}, I_{14}) of seq-II.

Algorithm 1 Algorithm for refinement of the matched correspondences

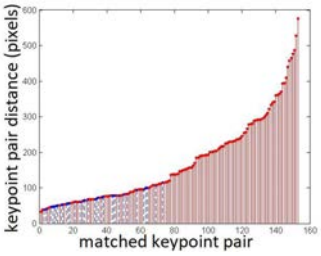
1: $C_m := \{(p_k, q_k) | k = 1, ..., n_m\}$ ▷ all matched keypoints
2: $S_{max} := \underset{(p_i,q_i)\in C_m}{argmax} \|desc_{p_i}, desc_{q_i}\|$;
3: $C_t := \{(p_l, q_l) | l = 1, ..., n_t \wedge \|desc_{p_l}, desc_{p_l}\| < th.S_{max}\}$ ▷ matched keypoints within $th.S_{max}$
4: $C_s := \{(p_c, q_c) | (p_c, q_c) \in C_t \wedge \|p_{c+1}, q_{c+1}\| > \|p_c, q_c\|\}$ ▷ C_t sorted by keypoint coordinate distance
5: $b := 0$ ▷ bin counter
6: $c := 1$ ▷ keypoint counter
7: **while** $c < n_t$ **do**
8: $b := b + 1$
9: $Accu(b) := \{(p_c, q_c)\}$ ▷ initialize accumulator bin b with c^{th} element of C_s
10: $max_b := \|p_c, q_c\|$
11: $min_b := \|p_c, q_c\|$
12: $mean_b := \|p_c, q_c\|$
13: $c := c + 1$
14: **while** $(max_b - mean_b) < sp.mean_b$ & $c < n_t$ **do**
15: $Accu(b) := Accu(b) \cup \{(p_c, q_c)\}$
16: $min_b := \underset{(p_i,q_i)\in Accu(b)}{argmin} \|p_i, q_i\|$;
17: $max_b := \underset{(p_i,q_i)\in Accu(b)}{argmax} \|p_i, q_i\|$;
18: $mean_b := mean\{\|p_i, q_i\| | (p_i, q_i) \in Accu(b)\}$
19: $c := c + 1$
20: b_{max} :=Find_largest_bin_index ($Accu$)
21: $C_r := \{(p_f, q_f) | (p_f, q_f) \in Accu(b_{max})\}$ ▷ refined keypoint matches

Algorithm 1 details the proposed scheme for the refinement of the matched keypoints. Let C_m contain all the matched keypoint pairs (p_i, q_i). Fig. 2.19(a) shows the distance-plot of the all the keypoints initially matched using the SURF based approach for the image pair $(I_1,\ I_2)$. Fig. 2.19(d) illustrates the correspondences on the juxtaposed images. Many false matches can be noticed that result in a low correct to the total number of matches ratio ($R_c = 185/724$). C_t is obtained after an initial refinement by discarding the keypoint pairs whose corresponding descriptor distance is more than $th.S_{max}$, with S_{max} being the maximum of the distances between the matched descriptors and th indicating the fraction of this value used to establish a threshold for discarding the potentially incorrect matches. Fig. 2.19(b) and Fig. 2.19(e) respectively show the distance-plot and correspondences of the sorted matches refined with $th = 0.6$. Although this alone removes most of the incorrect correspondences, a large number of mismatches still remains ($R_c = 179/434$). It can be noticed in the distance-plot that the correct matches (blue lines) form a roughly flat interval. For isolating the keypoints corresponding to this region, the

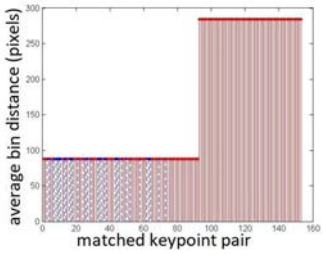
distance-plot is divided into several intervals by grouping the correspondences that fall within a certain threshold of the running average (a related and well-known concept of the cumulative sum has several applications in data analysis). The objective is to group the keypoint pairs in different bins b of a multi-bin accumulator $Accu(b)$ according to their Euclidean distance. Let max_b, min_b and $mean_b$ respectively represent the maximum, minimum and mean of the Euclidean distances between the keypoints contained in the bin b of the accumulator. The first bin is initiated with the first matched keypoint pair. The following correspondences are added one by one to this bin. A new bin is started if the difference between max_b and min_b of this bin surpasses $sp.mean_b$, with sp determining the interval "flatness". A larger sp value permits grouping of keypoints pairs with larger differences in their Euclidean distances. After the end of the process, the bin with the largest number of elements is likely to correspond to the flat interval containing the correct matches. The largest flat region in Fig. 2.19(c) represents the matched keypoint pairs retained using this process with $sp = 1.5$. A large number of incorrect matches were eliminated while retaining all the correct matches ($R_c = 179/322$). The sub-figure just below this sub-figure shows the correspondences over the juxtaposed images. The results of this refinement for the image pairs (I_{13}, I_{14}) are shown in Fig. 2.20. A considerable number of false matches were eliminated while conserving all the correct matches. The pair (I_{13}, I_{14}), which involves large changes in homography parameters, led to a more challenging keypoint refinement task because large homography parameters result in less flat curves of the distance-plot (see Fig. 2.20). For this pairs also, R_c remained low after initial refinement with $0.6S_{max}$ threshold. After the selection of the correspondences falling in the largest interval, this ratio increased significantly.



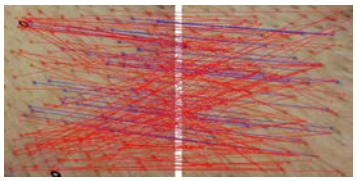
(a) Increasing spatial distance between the initially matched keypoints



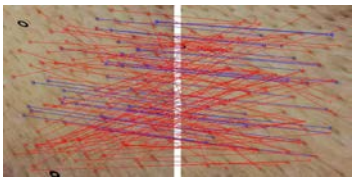
(b) Increasing spatial distance between the matched keypoints thresholded at $0.6S_{max}$



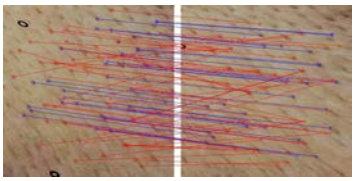
(c) Thresholded keypoint pairs grouped into intervals according to their spatial distance variation



(d) Correspondances of the initially matched keypoints ($R_c = 38/255$)



(e) Correspondences of the matched keypoints thresholded at $0.6S_{max}$ ($R_c = 34/153$)



(f) Correspondences falling in the largest flat interval of the above plot ($R_c = 34/92$)

Figure 2.20: Selection of best matches from the keypoint correspondences established on the pair (I_{13}, I_{14}) of seq-II using SURF with $0.0001T_S$. The red lines indicate false matches.

2.7 Mosaicing of the Real Sequences

Fig. 2.21 shows the mosaic built from a human forearm video sequence using the SURF based image registration. This sequence, covering a part of the anterior forearm region, was acquired along a sinuous trajectory. The mosaic of 250 frames is shown in Fig. 2.21(a), whereas Fig. 2.21(b) shows the corresponding region photographed using a smartphone camera with a resolution of 8 megapixels. The image obtained through mosaicing has a markedly higher resolution, almost quadruple the one attained with the smartphone camera. Although no criterion is available for quantitative evaluation of the result, the visual coherence of the mosaic can be appreciated from the fact that the veins and the texture are well aligned. The resulting mosaic is often distorted on the long sequences due to the accumulation of the registration errors. One such case is shown in Fig. 2.22 for a sequence acquired over the human back. Such cases are the driving factor behind the adaptation of a mosaicing approach that takes into account the acquisition topology and tries to minimize such accumulations by choosing the shorter paths over which the homographies would be accumulated.

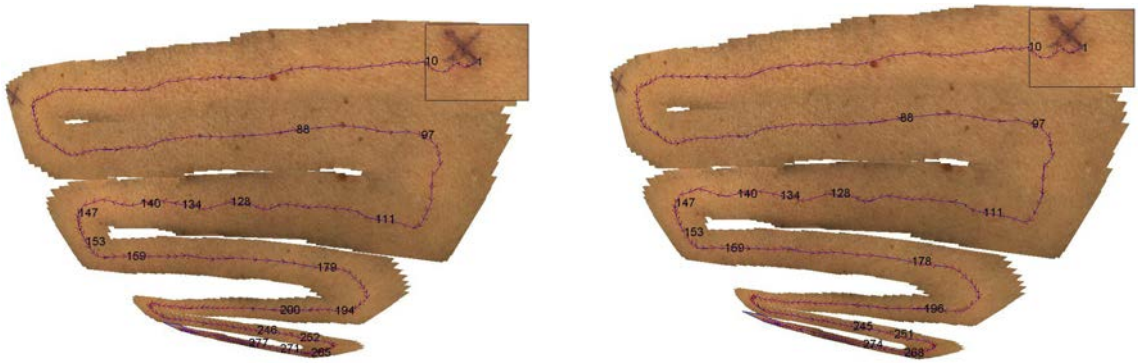
(a) Image mosaiced using the SURF based registration

(b) Photographed with a smartphone camera

Figure 2.21: Real data mosaicing result (a): Partial anterior forearm region mosaiced over 250 frames using the SURF based registration. The resulting mosaic has a size of 2794×3464 pixels. The mosaicing trajectory along with the initial (black rectangle) and final frame (blue quadrangles) are marked on the image. (b): Anterior forearm, photographed with a smartphone, over the region corresponding to the mosaiced zone. The image has a resolution of 1406×1372 pixels showing approximately 7×6 cm^2 of skin area.

2.8 Summary and Discussion

A literature review of the works involving skin image registration for various applications was presented. In these works, the registration was usually performed over images obtained un-



(a) Mosaic obtained using the SURF based image registration

(b) Mosaic obtained using the SIFT based image registration

Figure 2.22: Mosaicing using SURF and SIFT based approaches of a video sequence containing 300 images acquired carefully over human back. The mosaic, despite successful pairwise registrations, is distorted due to the accumulation of errors over a long trajectory.

der very controlled acquisition conditions and the registration approaches used just considered simple (non-homographic) transformation parameters. This chapter discussed some registration approaches with the ability to take into account all parameters of a perspective transformation (a homography is the most complex linear transform between images). In this framework, some optical flow based and feature based methods were compared. For feature-based approaches, an extensive overview of diverse types of keypoint extraction and feature formulation approaches was presented and two of these approaches were analyzed in depth. A comparison of the results over skin image registration for some of these approaches was made. For this comparison, several sequences simulating realistic camera movements were built to enable a quantitative evaluation of the results. Based on an initial comparison, the feature based approaches turned out to best suit the objective of skin image mosaicing. The results of SIFT, SURF and BRISK, the feature based approaches, were further closely compared to find out if the latter two could be used, despite lower precision than SIFT, for skin image mosaicing. The interest in such comparison comes from the observation that SURF and BRISK are considerably faster than SIFT. Nevertheless, SURF and BRISK can be used for obtaining mosaics with acceptable errors under some compromise over precision and accuracy. In efforts to improve the robustness of the feature based methods, an approach for refining the detected feature correspondences was presented to have less uncertainty in homography calculation from these correspondences. In mosaicing of the long real sequence (containing over 100 images covering up to 10×10 cm^2 of skin surface), the need for a global mosaicing scheme by considering the topology of the images in the sequence is observed. Such scheme is the theme of the next chapter.

Chapter 3

Skin Image Mosaicing with Topological Inference

3.1 Introduction

The performance comparison of several image registration schemes applied to skin video sequences performed in chapter 2 led us to retain SURF as the method of choice for the present study. The concatenation of the homographies of consecutive pairs along the acquisition trajectory posed some limitations, as was observed in mosaicing of simulated as well real video sequences. As was noted for the real sequence shown in Fig. 2.22, the accumulation of errors over several image pairs in a long sequence is likely to lead to distortion in the final mosaic. This issue is not necessarily resolved by using a more precise image registration scheme, such as a SIFT-based registration. In addition, if there are some failed registrations in the middle of a sequence, the mosaicing process is interrupted from that point onward. This is due to the fact that the geometric transformation, with respect to the reference image, of the images following that point cannot be determined due to lack of a sequence of homographies leading up to those images. A topological inference scheme was sought to deal with these problems. The topology of the mosaiced sequence refers to the relative positions of the mosaiced images on the mosaicing plane. Prior to the mosaicing, no information about these positions is available. A pairwise registration of all the images along the acquisition path provides spatial information only about the consecutive images. The relative positions of all the images based on this initial topology may not be accurate enough due to error accumulation. The objective is to sequentially refine the estimated image locations from the initial topology in the mosaicing plane. For images resulting in failed registrations, an estimated location is substituted to have a rough estimate of the topology that would later help in finding alternative paths to reach the images situated after the interruption point in the same sequence.

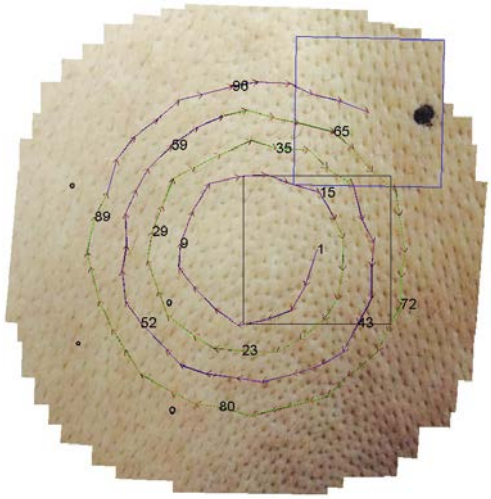
Two schemes for finding these alternative paths for video sequences obtained through a well-adapted acquisition scheme are presented after a brief overview of some existing mosaicing approaches that take topological considerations into account. One of the presented schemes is adapted from an existing method [MFM04]. This adaptation was made over an appropriate camera movement scheme to cover a large area with the possibility of finding more direct paths to a large number of images. Though this approach turns out to work fine to achieve the defined goal, it has certain limitations since it finds the additional paths based on the assumption that the initial topology estimate is not far from the correct topology. This may pose a problem if the non-sequential pairs through which the additional path are found do not have a significant

overlap. This limitation may be overcome through sequentially eliminating such pairs and subsequently re-calculating the additional paths. However, this becomes computationally expensive. Another proposed approach, although not different in its purpose, verifies in advance if a direct path will be possible at a few key locations. This approach provides an improved control over the topology estimation process and does not get stuck in the iterative refinement of the direct paths found. Besides, this provides more flexibility by pre-defining the frequency of the direct links. The failed registrations are detected through a combination of some criteria for avoiding the interruption of the mosaicing process. The images where the registration fails are not mosaiced and the images surrounding them are reached through the alternative paths found through one of the two approaches.

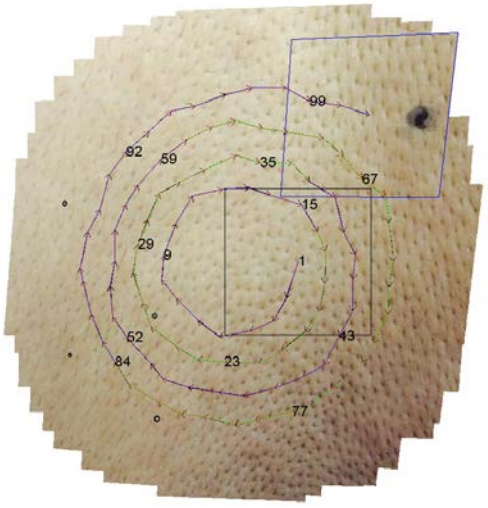
A reduction in the number of homographies to be concatenated to reach a given image in the video sequence largely reduces the error accumulation and the resulting mosaic provides quite a coherent view. However, the neighboring images which are reached through different paths may be misaligned. This misalignment is significant if the length of the accumulated paths for both or one of these neighboring images is large. A global alignment to overcome this issue is sought through controlled variation of the homographies. In the first step, a new and fast approach for the global alignment of a sequence containing a large number of images is developed since the existing bundle adjustment approaches, that take into account all the image pixels, tend to be computationally expensive. The second step is the adaptation of this approach to the multi-path mosaicing strategy that is developed for this study. The first step is successfully achieved and its effectiveness on simulated sequences forming a closed loop is demonstrated. For the second step, which requires simultaneous adjustment over several loops formed through the multi-path approach, preliminary results in limited cases are presented and its prospect is discussed.

3.2 Topology Inference

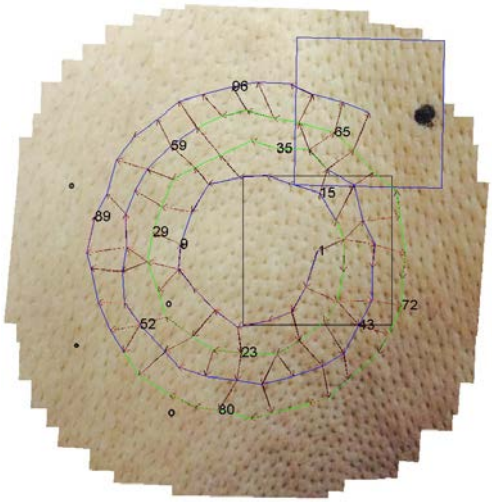
The topology of the video sequence over a surface represents the spatial relationship of the constituent images. No precise geometric relationship between the images is known prior to their registration. The only available information is that the sequential images in the sequences are next to each other. The displacement of the sequential images can be determined by computing homographic relations between them. This provides the simplest topology where the relative positions of the sequential images (i.e. consecutive images in the video sequence) are known. The sequential homographies can then be concatenated to find the coordinates of all the images in the video sequence with respect to a reference image, provided that the sequence is uninterrupted, i.e. the homographic relation between all the consecutive images is successfully calculated. The interest behind determining the topology of the sequence is to find non-sequential overlapping image pairs to find shorter or alternative paths to reach a given image. Some crossings or overlaps in the acquisition trajectory should exist to achieve this objective. Such crossings can exist in an acquisition over a zig-zag pattern or any other pattern that results in sub-loops in the trajectory. In [VRD13], a spiral acquisition was used for mosaicing endodermoscopic hepatic sequences. This acquisition scheme was opted for due to several conveniences that it offers. Although the acquisitions were made through a controlled robotic arm in [VRD13], approximate spiral trajectories can be easily followed manually after little practice. Fig. 3.1(a) shows a simulated sequence (ground truth mosaic) where a spiral trajectory is followed. This sequence is simulated to follow an Archimedean spiral such that partial overlap exists between images in the adjacent spiral rings. A mosaic obtained by concatenating the homographies along the sequential trajectory is shown in Fig. 3.1(b). This results in some noticeable distortion in the



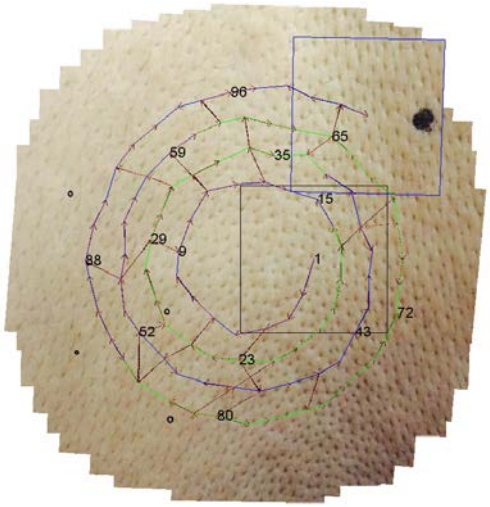
(a) Ground truth mosaic of a sequence simulated from a high-resolution skin image over a spiral trajectory.



(b) Mosaic obtained using the SURF based registration scheme by concatenating the homographies over the sequential path.



(c) Mosaic obtained by concatenating the homographies over paths obtained through an iterative approach discussed in section 3.2.2.



(d) Mosaic obtained by concatenating the homographies over paths obtained using an angle-based approach discussed in section 3.2.3.

Figure 3.1: Mosaicing results obtained for a simulated sequence (a), using two different approaches (c, d), to minimize the accumulation of errors visible in (b). The black and blue quadrangles locate the first and last frames of the sequence, respectively. The sequential trajectory is marked in green or blue (a change in color marks the beginning of a new spiral ring). The dashed red lines, with arrows indicating the path direction, represent the most direct paths. The positions in the sequence of a few images are illustrated by the number of the image placed in the mosaic at the center of the image.

mosaic due to the accumulation of small registration errors. The last image (highlighted through a blue quadrangle), that should have the same dimensions as the first image (black square in the center), is distorted. In addition, some ghost texture can be observed in Fig. 3.1(b). The black mark in the last image notably has a shadow due to misalignment of different images in

which this mark appears. By placing the reference image at the center of the spiral, radial paths towards images located in outer rings of the spiral can be found. This, in turn, reduces the accumulated errors and permits estimation of a topology that results in visually less distorted mosaics. Figs. 3.1(c) and 3.1(d) show refined mosaics obtained after finding more direct paths while considering the non-sequential image pairs (the approaches used for obtaining these mosaics are presented in subsections 3.2.2 and 3.2.3). The resulting mosaics have considerably more visual coherence, with almost no ghost texture. In this case, the distortion in the topology is not considerable because the transformation parameters of the simulated sequence are relatively moderate. The reason behind this choice was to establish the applicability of the used topological inference approaches. More thorough tests performed on real sequences are presented in section 3.2.5. In the real sequences, there are several other factors that need to be taken into account. One of these factors is failed registrations, that interrupt the construction of the initial sequential topology. Another factor, particularly of concern in longer sequences, is crossings or diversions of spiral rings due to the accumulation of registration errors or imperfect spiral trajectory in the manual movement of the acquisition device. This may give a false initial estimate of the overlapping pairs in the neighboring rings. This is the motivation behind sequential update of the topology, i.e. refinement of the topology in several steps, each step seeking to improve upon the previously adjusted topology. After a brief literature review of the approaches considering a topology refinement for mosaicing, two schemes for updating the image topology are presented hereafter.

3.2.1 Some existing works involving topology update

Refinement or update of the image path topology in the mosaic plane requires finding the non-sequential image pairs that have significant overlap. However, when every possible combination of the image pairs is considered, their registration would result in a large computation overload. There are works that use some ways to limit the number of image pairs across which the alternative paths would be considered. In an approach proposed in [MFM04] non-consecutive image pairs are iteratively selected one by one if they satisfy certain criteria based on two measures. One is the overlap measure $\delta_{i,j}$ between images I_i and I_j. The other is a path cost measure γ_{ij}. The overlap measure is calculated as:

$$\delta_{ij} = \frac{max(0, |c_i - c_j| - |d_i - d_j|/2)}{min(d_i, d_j)}, \tag{3.1}$$

where c_i and c_j are the centroids of the respective warped images and d_i and d_j are their side lengths (for square images). If $\delta_{i,j} > 1$, there is no overlap between the images and the images are exactly superimposed when $\delta_{i,j} = 0$.

The path cost γ_{ij} is calculated as:

$$\gamma_{ij} = \frac{\delta_{ij}}{\Delta_{ij}}, \tag{3.2}$$

where Δ_{ij} is the sum of the overlap measures $\delta_{i,j}$ over the previously available shortest path for connecting the image pair (I_i, I_j) which is a candidate of being a part of an alternative path. γ_{ij} ranges between 0 and 1. A high value would indicate that a short path already exists between these images and that, consequently, creating a new link between these images would not have much impact.

Algorithm 2 gives the process used for selecting the non-consecutive image pairs. Set T contains only the sequential image pairs and set S contains all the other possible pairs except the sequential ones. Another set L is initialized with the sequential pairs. After this initialization, the additional pairs are added from set S to set L if they satisfy some criteria and discarded otherwise. In both cases, the considered pairs are removed from set S at each iteration while the discarded pairs appear neither in S nor in L. For the implementation of this algorithm, these sets are transformed into connectivity graphs whose nodes represent the images in the sequence. Weights $\delta_{i,j}$ are assigned to the edges $e_{i,j}$ of the graph, with the indices indicating the image pair. The edges of the graph of set T connect the nodes corresponding to the sequential pairs, whereas the graph of set S is its complement. In the discussion that follows, the set labels will refer to the respective graph names. The graph L is initialized with the edges of graph T. The edges of S with $\delta_{i,j}$ greater than a threshold δ_{max} are removed before the start of the iterations. Also, the edges resulting in γ_{ij} greater than a threshold γ_{max} are removed (since only the sequential paths exist at this point, the $\Delta_{i,j}$ calculated for this purpose are obviously over the sequential paths). During each iteration, the edge with the minimum γ_{ij} is removed from S and added to L. The γ_{ij} values for all the edges in S are updated at the end of each iteration and the ones with a value greater than γ_{max} are eliminated. The process ends when no edges are left in S. The graph L now contains the non-consecutive image pairs which have sufficient overlap and are not too close to each other. The alternative (shortest) paths from the reference image to a given image can be found from this graph by using an appropriate approach for computing a minimum spanning tree (MST), over the connected edges in the graph (such an approach is also used during the computation of γ_{ij} for finding the shortest path connecting the images under consideration).

In [Wei+12b], Marzotto's approach was used with a more precise δ_{ij} measure. This measure corresponds to the smaller of the two values resulting from the ratio of the overlap area A_{ij} of the two images with the complete areas A_i and A_j of the individual images; areas considered being those of the images placed onto the mosaicing plane:

$$\delta_{ij} = min\left(\frac{A_{ij}}{A_i}, \frac{A_{ij}}{A_j}\right) \tag{3.3}$$

Algorithm 2 Iterative algorithm for selecting the non-consecutive overlapping pairs [MFM04]

procedure ITERATIVE_ALGORITHM_FOR_ADDITIONAL_IMAGE_PAIRS_SELECTION($\mathscr{T}$)
▷ $\mathscr{T}$ is the image sequence topology
  $T := \{(i,j)|j = i+1\}$; ▷ sequential pairs with weights $e_{i,j} = \delta_{i,j}$
  $S := \{(i,j)|j > i+1\}$; ▷ all other pairs with weights $e_{i,j} = \delta_{i,j}$ according to $\mathscr{T}$
  for each $(i,j) \in S$ compute $\delta_{i,j}$ and $\gamma_{i,j}$;
  $S := S \setminus \{(i,j)|\delta_{i,j} \geq \delta_{max} \vee \gamma_{i,j} \geq \gamma_{max}\}$;
  $L := T$;
  while $S \neq \emptyset$ **do**
    $e_{k,l} := \underset{(i,j)\in S}{argmin}\gamma_{i,j}$;
    $S := S \setminus \{e_{k,l}\}$;
    $L := L \cup \{e_{k,l}\}$;
    for each $(i,j) \in S$ compute $\gamma_{i,j}$;
    $S := S \setminus \{(i,j)|\gamma_{i,j} \geq \gamma_{max}\}$;

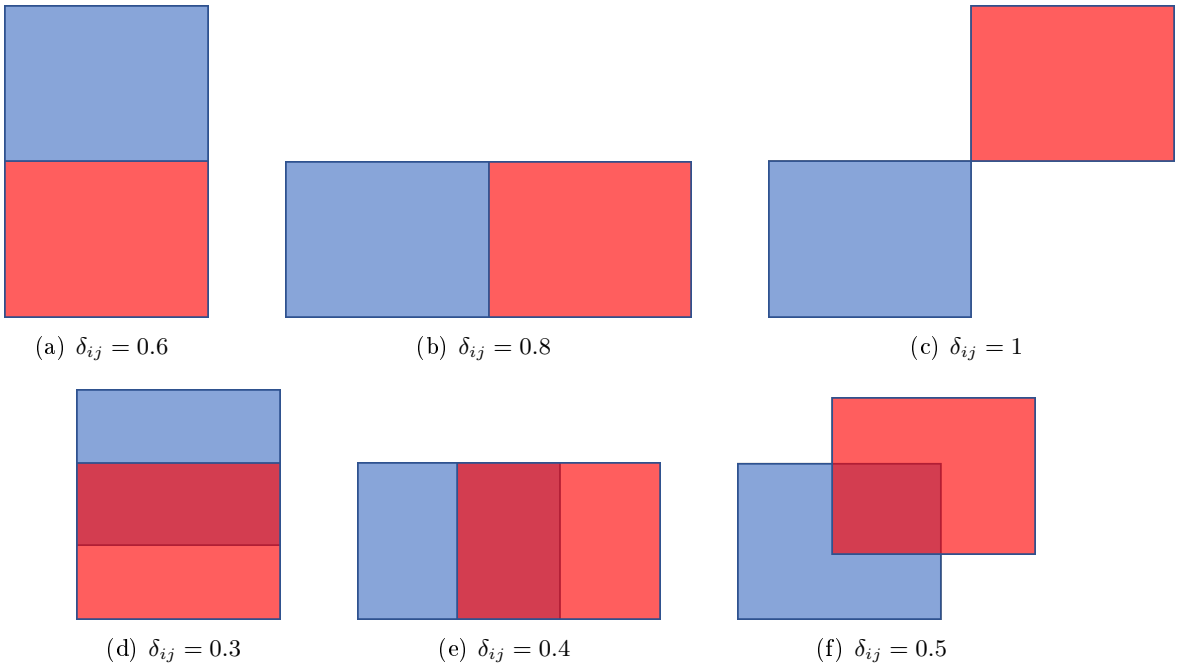
In an earlier work [SHK98], similar considerations as in [MFM04] were used for topology estimation. In this work, $\delta_{i,j}$, defined in Eq.(3.1), was used as an overlap measure and Δ_{ij}, given in Eq.(3.2), was used as a path cost measure. Apart from these measures, a certainty measure ρ_{ij} of the registration success of these pairs is also taken into account. This certainty is calculated by computing the normalized cross-correlation of the image pair after superimposing one of the images on the other with the estimated transformation. Additional non-consecutive image pairs are iteratively added to the connectivity graph if the corresponding δ_{ij}, Δ_{ij} and ρ_{ij} values are above predefined thresholds.

A dynamic topology refinement scheme was used in [EGG13] for mosaicing of underwater image sequences. In this approach, at first, the SIFT descriptors were extracted for all the images in the sequence. Since descriptor matching is computationally expensive, only a subset of randomly selected descriptors was matched among all the images, which gave a rough estimate of potentially overlapping images. An MST was constructed from this estimate and the images along the paths of this MST were registered. In case of failed registrations, the corresponding edges were eliminated, and a new MST computed. In [FSV01], topology refinement was used in combination with a global adjustment scheme for mosaicing underwater images. The estimated camera trajectory was iteratively refined by finding the non-consecutive image pairs with significant overlap and then by redistributing the estimated homographies through a global adjustment to get a better estimate of the topology.

3.2.2 Topology update using iterative scheme (first proposed approach)

In this approach, Marzotto's algorithm is used with some adaptations. Since the height and the width of the images used in the video sequences are not the same, instead of using the image sides (d_i and d_j in Eq. (3.1)), the averages of the two diagonals of the images placed in the mosaic plane are used in the calculation of the overlap measure δ_{ij}. The sides of the image frames used in this study have a ratio of 1/1.34, resulting in a diagonal length of 1.67 units (1 unit being the length of the smaller side). Thus, if the value of δ_{ij} computed in this way is greater than $1/1.67 \approx 0.6$, the overlap is negligible or null. Fig. 3.2 shows δ_{ij} values for an image pair with different alignments. Even though this criterion gives only an approximate overlap measure due to different orientations of the overlapping images and changes in their geometry after warping, it is adequate for detecting image pairs with significant overlap. Furthermore, the elimination of less direct paths (lines 10 and 11 in algorithm 2) is repeated 10 times at each iteration to reduce the complexity of the algorithm, which becomes prohibitively expensive for a large graph with numerous edges for several nodes. Dijkstra's algorithm for minimum spanning tree (MST) [Dij59] is used for the computation of shortest paths.

Agorithm 3 gives the detail of this approach, which sequentially updates the topology $\mathscr{T}$ of the mosaiced images. This update is sequentially performed N_s times (N_s was chosen to be 2 for the results presented in this work). For each update, after selection of the non-consecutive overlapping pairs, MST gives the shortest paths from the reference image to each of the other images in the sequence. However, these paths assume that the initially estimated topology is close to the true topology. In real cases, due to the deviation of the mosaicing trajectory from the acquisition trajectory, δ_{ij} does not necessarily give the actual overlap measure. Moreover, non-consecutive images are difficult to register due to large variations in transformation parameters. The registration of numerous additional image pairs is likely to fail due to these factors. This results in "broken paths" leading to interruption in mosaicing process. The registration of each additional pair is verified through some automatic failure detection criteria (described in subsection 3.2.4). Image pairs with failed registrations are removed from the connectivity graph



(a) $\delta_{ij} = 0.6$ (b) $\delta_{ij} = 0.8$ (c) $\delta_{ij} = 1$

(d) $\delta_{ij} = 0.3$ (e) $\delta_{ij} = 0.4$ (f) $\delta_{ij} = 0.5$

Figure 3.2: Overlap measures for diffent alignements of an image pair considering length to width ratio of 1.34 for each image. The light red and blue rectangles represent the two images. The overlap zone is shown in dark red color.

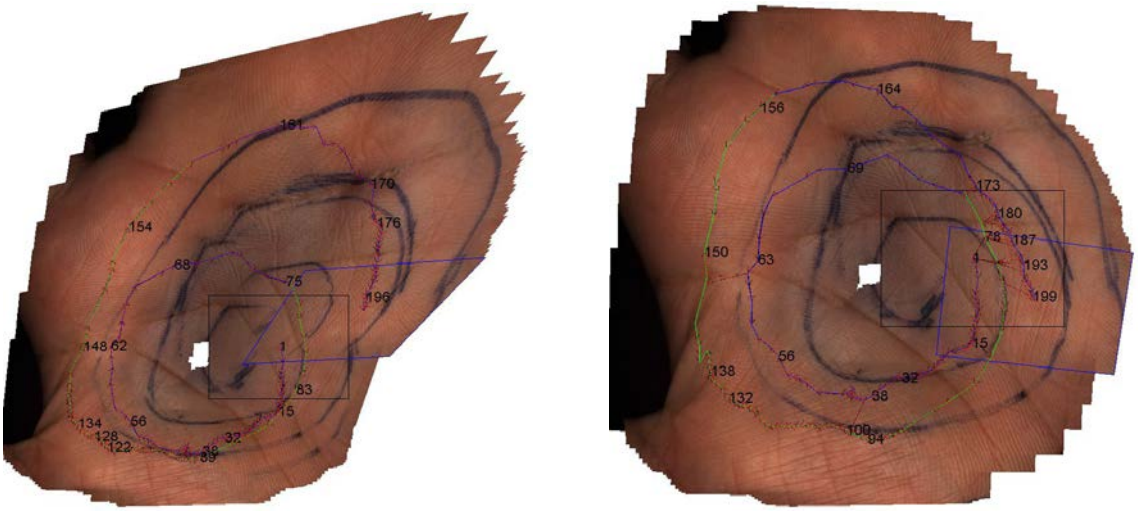
Algorithm 3 Use of iterative algorithm for finding the shortest paths

```
1: 𝒯                                  ▷ initial topology after sequential image pair registration
2: N_s                                ▷ number of sequential topology updates
3: for N_s number of times do
4:     S := Iterative_algorithm_for_additional_image_pairs_selection (𝒯)   ▷ S is the connectivity graph containing additional pairs
5:     Broken_paths:= 1   ▷ a flag for determining if there are any failed registrations in the new paths
6:     while Broken_paths=1 do
7:         Find shortest paths in S by computing an MST
8:         Verify registration of all new pairs in the shortest paths
9:         if Failed registrations are found then
10:            Remove pairs with failed registrations from S
11:            Find shortest paths in S by computing an MST
12:        else
13:            Broken_paths := 0;
14:    update 𝒯 with the shortest paths found by MST
```

and the MST is calculated once again for the remaining pairs. If the resulting paths include some pairs which were not selected in the first calculation, their registration is verified, and a new MST is calculated if some of these pairs fail to register. The process is repeated until no failed registrations remain. The additional pairs with successful registrations permit mosaicing with a topology more conform to the correct topology. Since the adjusted topology may still be

distorted for longer sequences, particularly at the outer spiral rings, it is adjusted once again by repeating the entire process. This is likely to give a still better estimate since the first topology refinement gives a better overlap estimate for the non-consecutive image pairs.

Fig. 3.3 juxtaposes two mosaics of a sequence acquired over human palm before and after the topology update. The manually traced spiral was used to facilitate acquisition over a spiral trajectory (the results without this aid will be presented shortly). In these results, several radial paths were found, which provided the possibility of mosaicing the images in the outer spiral branches with the concatenation of significantly fewer homographies. The mosaic in Fig. 3.3(a) is acquired over a sequential path using the SURF based image registration approach. The mosaic in Fig. 3.3(b) shows the result using more direct paths found through the proposed approach. The thresholds values $\delta_{max} = 0.2$ and $\gamma_{max} = 0.5$ were used for discarding the pairs not satisfying the respective overlap and path cost measures. A significant reduction in the distortion of the mosaic can be noticed after topology adjustment. The alignment of the palm lines was improved. The shape of the manually traced spiral indicates a mosaic with fewer distortions. Although a more relaxed overlap threshold can be used to find more radial paths, this increases the number of edges in the graphs, which in turn results in a large increase in computation time. In addition, this increases the number of additional pairs whose registration needs to be verified.



(a) Mosaic obtained by concatenating the homographies over the sequential path using the SURF based registration scheme. The resulting panorama has an area of 4815×4513 pixels.

(b) Mosaic of 3588×3356 pixel size obtained by concatenating the homographies over paths obtained through iterative approach with $\delta_{max} = 0.2$ and $\gamma_{max} = 0.5$. Updated mosaic was obtained in 9 min 15 s.

Figure 3.3: Mosaicing results obtained for a sequence over human palm by implementing the strategy of most direct trajectories to minimize the accumulation of errors. The black and blue quadrangles locate the first and last frames of the sequence, respectively. The sequential trajectory is marked in green or blue (a change in color marks the beginning of a new spiral ring). The dashed red lines, with arrows indicating the direction, represent the most direct trajectories. The positions in the sequence of a few images are illustrated by the image numbers placed in the mosaic at the center of the images. The area of interest is of roughly 8×8 cm^2 size. The sequence contains 200 images all of which were mosaiced without skipping any.

3.2.3 Topology update by finding selective radial links (second proposed approach)

One limitation of the iterative algorithm is that it supposes that the initial topology is approximately close to the actual acquisition trajectory. However, as the error accumulation increases, the distortion in the topology results in a less correct calculation of δ_{ij}. As a result, additional links may be found between image pairs that do not have sufficient overlap for a successful registration. Thus, it may generate some broken paths. Although this problem was solved in section 3.2.2 by repeatedly calculating an MST after testing the registration of the detected additional pairs and eliminating the pairs with failed registrations, this becomes computationally expensive, especially if there are large distortions in the mosaic.

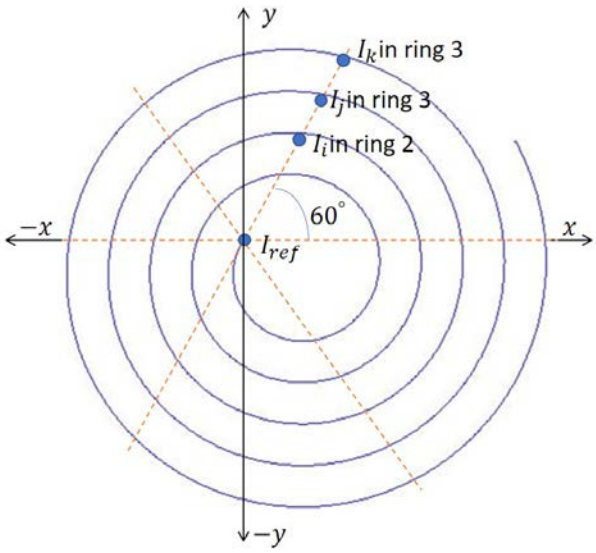


Figure 3.4: Illustration of radial links between different rings of a spiral trajectory.

The spiral trajectory is exploited to formulate a novel approach for finding the alternative paths with less calculation overload. This is done by keeping track of the spiral branches and angles between the axes of a given image and that of the reference image placed at the center of the spiral. Fig. 3.4 illustrates the principle of this approach. The objective is to find images I_i, I_j and I_k in the adjacent spiral links and establish a "radial link" if these images have significant overlap. Radial links at different angles would provide shorter paths to reach a given image from the reference image. Algorithm 4 details this approach. The nodes in graph T contain edges connecting only the sequential images. Graph S is initialized with graph T. Let $ring(i)$ represent the index of the spiral ring on which the image I_i is present. At regular angular intervals, for each ring of the spiral, the images resulting in a successful registration with the images at corresponding angles in the two immediately adjacent rings ($ring(i)$+1 and $ring(i)$+2) are searched to obtain additional image pair sets D_1 and D_2. Since the estimated positions of the images are known after the initial topology estimate, the images at the same angle θ_n, within a tolerance ϕ, in different rings can easily be identified. A tolerance in the search space is used to take into account the distortion in the topology. Due to this tolerance, a large number of candidate pairs are detected at each θ_n. For finding a successful pair (i.e. with successful registration), the pairs with $\delta_{ij} > 2.5$ are first eliminated. The remaining ones are registered one by one, in ascending order of their angle differences. The first pair resulting in a successful

registration (according to the criteria described in the next subsection) is retained as a radial edge for that θ_n and is added to the set S. Once all the possible radial links at all the θ_n have been found and added to graph S, MST calculation over this graph gives the shortest paths for each image from the reference image. Since the registration of the additional pairs has already been verified, further refinement of MST is not required.

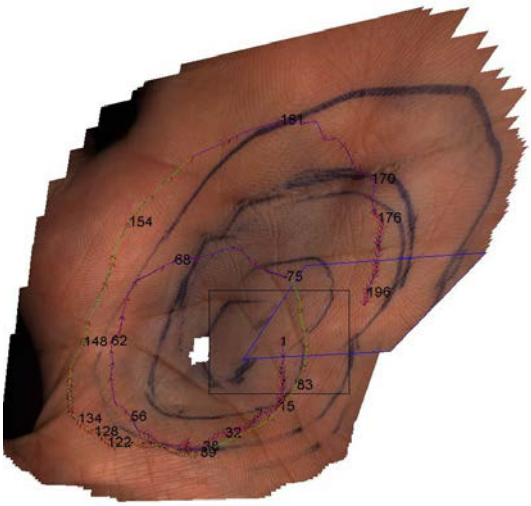
Algorithm 4 Angle-based algorithm for finding the shortest paths

$\Delta\theta = \{\Delta\theta_1,\ \Delta\theta_2,\ \Delta\theta_3\}$ ▷ Search angles with $\Delta\theta_1 > \Delta\theta_2 > \Delta\theta_3$
$\phi := \{\phi_1,\ \phi_2,\ \phi_3\}$; ▷ Search angle tolerance with $\phi_1 > \phi_2 > \phi_3$
$\mathscr{T}$ ▷ initial topology after sequential image pair registration
for $k := 1$ to 3 **do**
  $T := \{(i,j)|j = i+1\}$; ▷ sequential pairs
  $\theta = \{\Delta\theta_k,\ 2\Delta\theta_k,\ 3\Delta\theta_k, ...,\ 2\pi\}$
  $S := T$;
  $R := \emptyset$
  for each spiral ring b **do**
    for each $\theta_n \in \theta$ **do**
      $D_1 := \{(i,j)|angle(i) = \theta_n \pm \phi_k \wedge angle(j) = \theta_n \pm \phi_k \wedge ring(i) = ring(j) + 1\}$;
      $p :=$ Find_image_pair_with_successful_registration(D_1);
      if $p \neq 0$ **then**
        $R := R \cup p$;
      $D_2 := \{(i,j)|angle(i) = \theta_n \pm \phi_k \wedge angle(j) = \theta_n \pm \phi_k \wedge ring(i) = ring(j) + 2\}$;
      $p :=$ Find_image_pair_with_successful_registration(D_2);
      if $p \neq 0$ **then**
        $R := R \cup p$;
  $S := S \cup R$;
  Find shortest paths in S by computing an MST
  Update $\mathscr{T}$ with the shortest paths found

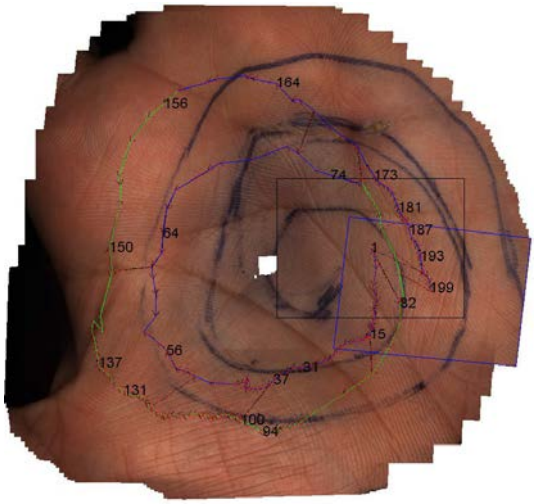
This approach provides more control over the process of finding the additional paths by predefining the locations where they are searched. Contrary to the iterative approach discussed earlier, the determination of the candidate additional pair locations is instantaneous. The computation overload depends on the frequency of such locations and the number of image pairs for which the registration is tested. The frequency is controlled by the intervals of θ_n (angle step or $\Delta\theta$) and the number of pairs to be tested per angle step is controlled by the tolerance ϕ. Although choosing both a small angle step and a large tolerance increases the likelihood of finding the radial links, this may also increase the number of registrations to be tested when the initial topology estimate results in a distorted mosaic. A hierarchic approach is used to decrease the computational complexity. The topology is refined in three stages. For the first stage, fewer additional pairs are searched at 90° intervals ($\Delta\theta_1$) with = 15° tolerance (ϕ_1). Since the distortion is likely to be reduced after the first refinement, a smaller tolerance $\phi_2 = 10°$ is used at the second stage with 60° intervals ($\Delta\theta_2$). The third stage uses even smaller tolerance of $\phi_3 = 5°$ with search locations situated at 30° intervals ($\Delta\theta_3$). These values can be varied depending on the distortion observed in the initially obtained mosaic.

Fig. 3.5 shows the mosaic before and after finding the shortest paths using the angle-based algorithm over the same sequence that was used in the previous section for the iterative approach. Fig. 3.5(a) shows the mosaic obtained by concatenating the homographies over the sequential

path. The mosaic shown in Fig. 3.5(b) results after finding the radial paths with the angle step and angle tolerance $(\Delta\theta, \phi)$ values being $\{(90°, 15°), (60°, 10°), (30°, 5°)\}$ in the three-stage version of the described algorithm. The radial paths found in this way helped with the construction of a coherent and less distorted mosaic. The computation time for updating the mosaic was less than half the time taken by the iterative approach. Fig. 3.6 shows the results of the two topology update approaches for another sequence acquired over the human palm. Both approaches significantly improved the mosaic. The coherence of the mosaic can be appreciated by comparing it with the smartphone image of the mosaiced region. For this sequence, although the angle-based approach was about four times faster than the iterative approach, it failed to find radial links near the end of the outermost spiral ring, resulting in misalignment in the top left region of the mosaic (see Fig. 3.6(c)).

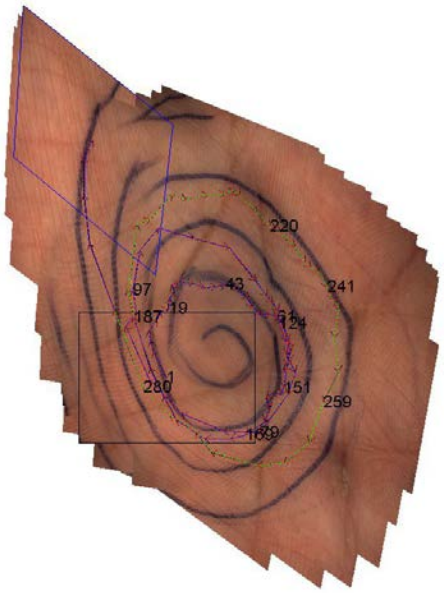


(a) Mosaic obtained by concatenating the homographies over the sequential path using the SURF based registration scheme. The resulting panorama has an area of 4815×4513 pixels.

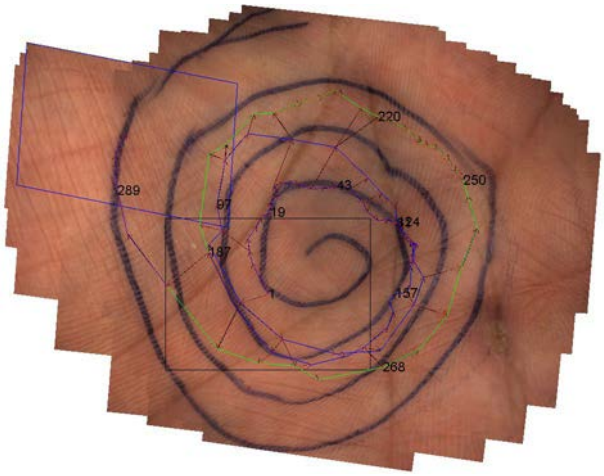


(b) Mosaic obtained by concatenating the homographies over paths obtained using angle-based approach. The resulting mosaic has an area of 3588×3356 pixels. It was obtained with following parameters at the three stages of the algorithm: $(\Delta\theta,\ \phi) = \{(90°,\ 15°),\ (60°,\ 10°),\ (30°,\ 5°)\}$. Updated mosaic was obtained in 4 min 27 s.

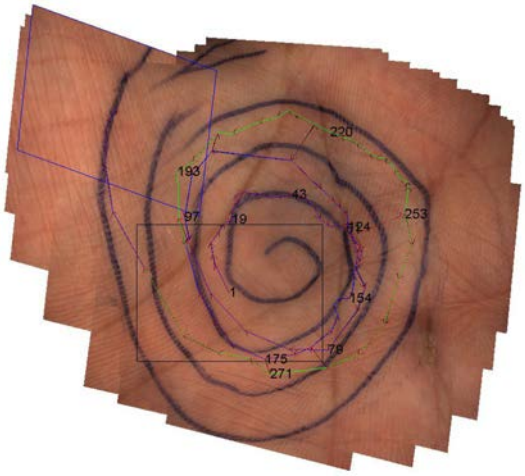
Figure 3.5: Mosaicing results obtained for a sequence over human palm by implementing the strategy of most direct trajectories to minimize the accumulation of errors. The black and blue quadrangles locate the first and last frames of the sequence, respectively. The sequential trajectory is marked in green or blue (a change in color marks the beginning of a new spiral ring). The dashed red lines, with arrows indicating the direction, represent the most direct trajectories. The positions in the sequence of a few images are illustrated by the image numbers placed in the mosaic at the center of the images. The area of interest is of roughly 8×8 cm^2 size. The sequence contains 200 images all of which were mosaiced without skipping any.



(a) Mosaic of 3046×4064 pixel size obtained using the SURF based registration scheme by concatenating the homographies over the sequential path.



(b) Mosaic of 3710×2903 pixel size obtained by concatenating the homographies over paths obtained through iterative approach with $\delta_{max} = 0.2$ and $\gamma_{max} = 0.5$. Updatd mosaic was obtained in 7 min 15 s.



(c) Mosaic obtained by concatenating the homographies over paths obtained using angle-based approach (presented in the next subsection). The resulting mosaic has an area of 3542×3209 pixels. It was obtained with following parameters at the three stages of the algorithm: $(\Delta\theta,\ \phi) = \{(90°,\ 15°),\ (60°,\ 10°),\ (30°,\ 5°)\}$. Updatd mosaic was obtained in 2 min 6 s.

(d) Smartphone image, taken with an active flash, of the corresponding area. The image is of 2848×2392 pixel size.

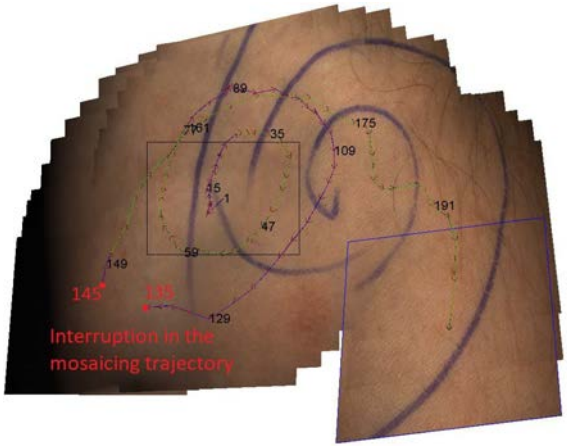
Figure 3.6: Mosaicing results obtained for a sequence over human palm using the most direct trajectories. The black and blue quadrangles respectively represent the first and last frames of the sequence. The sequential trajectory is marked in green or blue (a change in color marks the beginning of a new spiral ring). The dashed red lines, with arrows indicating the direction, represent the most direct trajectories. The positions in the sequence of a few images are illustrated by the image numbers placed in the mosaic at the center of the images. The area of interest is of roughly 9×7 cm^2 size. The sequence contains 300 images every third of which was mosaiced.

3.2.4 Detection of failed registrations

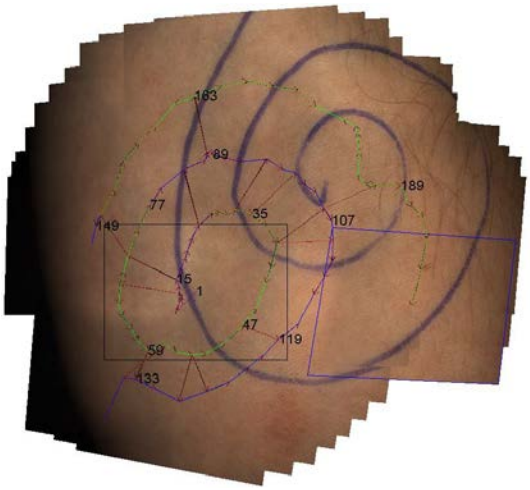
Failed registrations pose a major problem in mosaicing of large sequences. Despite careful camera movement, due to sudden hand movements or illumination changes, for instance, the registration of some image pairs may fail. The mosaicing process is interrupted due to the failed registration of a few consecutive image pairs. In addition, automatic detection of the failed registration is important during the selection of non-consecutive pairs for computation of additional paths in the proposed methods. For determining the success of an image pair registration, two main criteria are used. One is based on the number of inliers ($N_{inliers}$) among the detected point correspondences. The registration is classified as failed if this number is less than 25 (an arbitrarily selected value). The other criterion A_i/A_{i-1}, that supplements the first one, is to compute the ratio of the area of the warped image with the area of the preceding image in the sequential trajectory. If this ratio is not in the range [0.7, 1.4], the registration is considered failed. Two additional conditions are used for detecting large aberrations: if the ratio of the two diagonals ($diag_1/diag_2$) of the image warped using the estimated homography is not in the range [1/5, 5] or if its area A_i differs by a factor of 10 from that of the reference image A_{ref}, the registration is classified as failed. When failed registrations are detected during the calculation of additional paths, the corresponding edges are simply removed from the graph of additional pairs. However, it is more delicate if the failure occurs during the registration of sequential image pairs. To deal with such errors, the node corresponding to the image which failed to be registered is flagged and the corresponding homography is replaced by that of the preceding image pair to have a rough estimate of the topology for continuing the mosaicing process. If the failure does not occur over a large number of consecutive pairs, the sequence may still be mosaiced through alternative paths.

Fig. 3.7 shows an example of the treatment of image registration failures. The lack of continuity in the plotted sequential trajectory in Fig. 3.7(a) indicates failed registrations between images I_{135} and I_{145}. The homographies of the image pairs in this segment were replaced with that of the last successful pair (i.e. (I_{133}, I_{135})). Table 3.1 lists the values of the used failure detection criterion for the image pairs at which the failure was detected. For comparison, these values are also given for a few pairs preceding and following the interruption point. More than 30 inlier points were detected for the image pairs to (I_{133}, I_{135}) and the error criterion values were within the acceptable range. Only 16 inlier points were detected for the image pair (I_{135}, I_{137}). In addition, largely divergent values of the area and diagonal ratios indicate that the registration failed. It should be noted that different applications of RANSAC for this pair give different results, which is likely due to a low R_C value (the ratio of the correct number of matches to the total number of matches). The matched keypoint refinement approach presented in section 2.6.4 has the objective of reducing this variation in RANSAC output. However, this approach might not yield the desired refinement for a very small R_C. Although RANSAC can be applied several times to select the outcome that results in acceptable values of the error detection criteria, it was considered safer to discard image pairs with fewer than 25 inliers since the image registration is likely to be less precise in this case. Image registration failed for further 4 pairs, with only 8 inlier points detected for three of them and 13 for the remaining one (the points classified as inliers by the RANSAC algorithm are not necessarily the correct matches in case of a low R_C value).

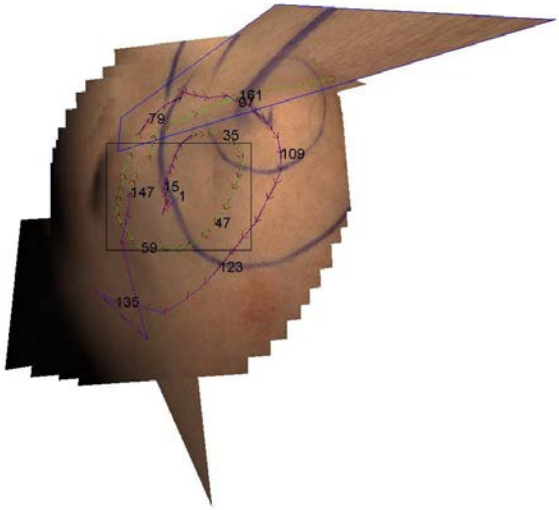
The area and diagonal ratios for the image pairs between I_{137} and I_{145} were calculated by substituting $H^{est}_{133,135}$ for the homographies of the preceding image pairs for which the registration failed. This substitution results in monotonic increase or decrease in the accumulated homography parameters, which causes the are ratios to vary considerably. The result of this effect can



(a) Mosaic obtained using the SURF based registration scheme by concatenating the homographies over the sequential path. The mosaic has a size of 4680×36377 pixels. Th homographies of image pairs with failed registration was replaced with $H^{est}_{133,135}$.



(b) Mosaic obtained by concatenating the homographies over paths obtained using the angle-based approach. The panorama of size 3633×3344 pixels was obtained with the three-stage angle approach with parameters: $(\Delta\theta,\ \phi) = \{(90°,\ 20°),\ (60°,\ 15°),\ (30°,\ 10°)\}$.



(c) Failed mosaic using the SURF based registration scheme by concatenating the homographies over the sequential path without discarding the failed registrations.

(d) Smartphone image, taken with an active flash, of the corresponding area. The image is of 2680×1984 pixel size.

Figure 3.7: Mosaicing results obtained for a sequence acquired over human forearm containing failed registrations. The black and blue quadrangles represent the first and last frames of the sequence respectively. The sequential trajectory is marked in green or blue (a change in color marks the beginning of a new spiral ring). The dashed red lines, with arrows indicating the direction, represent the most direct trajectories. The positions in the sequence of a few images are illustrated by the image numbers placed in the mosaic at the center of the images. The area of interest is of roughly 8×8 cm^2 size. The sequence contains 200 images every other of which was mosaiced.

Table 3.1: Values of image registration failure detection criteria for the image pairs at which the failed registration was detected in the mosaic of Fig. 3.7(a). Values of these criteria for a few image pairs preceding and following the failed registrations are also given.

Image pair	$N_{inliers}$	A_i/A_{i-1}	$diag_1/diag_2$	A_i/A_{ref}	**Registration status**
(I_{129}, I_{131})	59	0.996	1.007	0.995	success
(I_{131}, I_{133})	42	0.938	1.007	0.924	success
(I_{133}, I_{135})	32	1.048	1.013	0.968	success
(I_{135}, I_{137})	16	0.0392	9.371	0.042	failed
(I_{137}, I_{139})	13	0	0.302	0	failed
(I_{139}, I_{141})	8	1.034	1.039	1.088	failed
(I_{141}, I_{143})	8	0.967	1.063	1.055	failed
(I_{143}, I_{145})	8	1.077	1.051	1.213	failed
(I_{145}, I_{147})	28	0.988	1.050	1.145	success
(I_{147}, I_{149})	48	1.004	1.026	1.150	success
(I_{149}, I_{151})	72	1.021	1.009	1.174	success

be noticed in relatively large A_i/A_{ref} value (1.145) for the image pair (I_{145}, I_{147}), for which the registration was a success (please note that A_i/A_{ref} values given for the image pairs for which the registration failed do not represent the expected monotonic increase since the accumulated homography with $H^{est}_{133,135}$ substitution is updated *after* the computation of the error detection criteria). A relatively large acceptable deviation in A_i/A_{ref} is chosen to handle this sort of situation (an alternative could be to adapt a more elaborate model for homography prediction by taking into account homographies of several image pairs preceding the interruption point). The mosaicing over the sequential trajectory was continued by substituting the failed homographies with the last successfully calculated one up to the image I_{145}. 28 inliers, just above the acceptable number, were detected for the image pair I_{145}, I_{147} and it was successfully registered. The number of inliers increased steadily for the next two image pairs. A steady decrease and then an increase in the number of inliers for the image pairs around the interruption point also illustrates that the failed registration in a continuous sequence is unlikely to occur at an isolated image pair. The video acquisition device used in this study was set to capture 30 frames/s. A sudden movement of the hand or the subject may influence the registration of several image pairs in a row.

Fig. 3.7(b) shows the mosaic obtained after updating the topology of the mosaic of Fig. 3.7(a) using the angle-based approach. Several radial links were found, which not only were helpful in mosaicing a sequence which could not have been mosaiced using the sequential trajectories alone but also improved the overall coherence of the mosaic. The coherence of the mosaic can be appreciated by the shape of the spiral traced on the region of interest. Besides, a comparison with the smartphone image (Fig. 3.7(d)) of the mosaiced region shows that the mosaiced surface conforms to the skin surface.

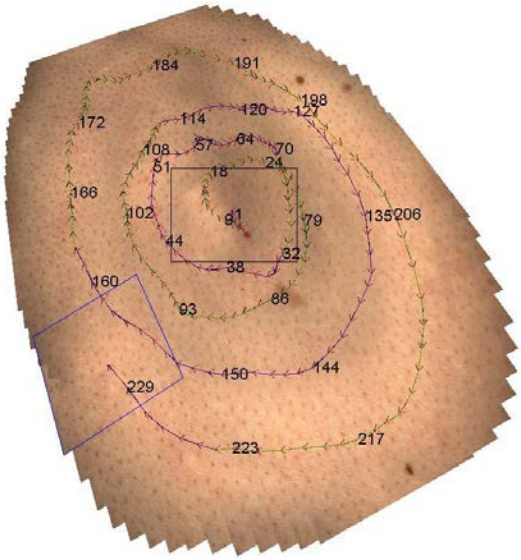
For highlighting the importance of global mosaic correction, Fig. 3.7(c) shows the mosaic obtained without discarding the failed registrations, i.e. the mosaicing process was continued without testing the failed homographies. The mosaicing process effectively fails at the image I_{135}. The images following I_{145} are also not correctly placed, despite successful registration, due to false homography concatenation. The whole mosaicing process was eventually interrupted at image I_{161} when the addition of the following image resulted in a mosaicing plane size too large to be processed by the limited computer RAM.

3.2.5 Results on different human skin surfaces

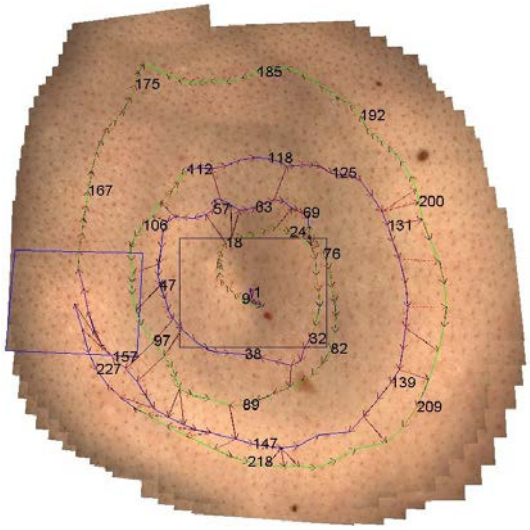
Both approaches for topology refinement were tested on sequences acquired on different body parts of some subjects to demonstrate its applicability on surfaces with different texture and surface shapes. Figs. 3.8 to 3.10 show results using both approaches for mosaicing video sequences acquired over the back of three subjects, along with smartphone images of the corresponding zones. The comparison with the smartphone image demonstrates a better color rendering and higher resolution of the panoramas obtained by mosaicing the sequences acquired through a colorimetrically calibrated device. The mosaic over sequential paths in Fig. 3.8(a) is quite distorted. In addition, there are several ghost textures due to misalignment of non-consecutive images, as can be observed from the duplication of some moles. Both approaches were successful in finding several radial links. The distortion is considerably reduced and no significant ghost textures are left (see Figs. 3.8(b) and 3.8(c)). The surface shape after topology correction corresponds well to the surface shown in the smartphone image. This fact can be observed by comparing the relative mole positions in the smartphone image (ground truth) with the ones in the built mosaics. Better quality of the mosaiced surface can also be appreciated from sharply contrasted moles and finer details of the skin texture. Even though there is a long part of the sequence, between images I_{156} and I_{199}, for which no radial link was found, radial paths surrounding this part minimized the distortion. It should be noted that about half of the images in this section were reached through one path and the rest through another one. Although no considerable misalignment occurs at the junction of the images reached through different paths for this particular segment, such cases may call for a controlled redistribution of homographies over the concerned section so that a misalignment can be avoided. This aspect is addressed in the next section.

Fig. 3.9 shows the mosaicing results on the skin surface of a human back with less visible pores. The mosaic obtained by concatenating the homographies over the sequential path is considerably distorted (see Fig. 3.9(a)). Several moles and freckles are duplicated due to misaligned images. Overall distortion in the mosaic can be noticed by comparing it to the smartphone image of the corresponding zone shown in Fig. 3.9(d). The images in the outer spiral ring show large perspective distortion, resulting in uneven resolution of the mosaiced surface. The mosaics obtained by both the proposed approaches (Figs. 3.9(b) and 3.9(c)) conforms to the smartphone image, as can be noticed from the lack of duplicate texture and by comparing the locations of the corresponding moles/freckles in the smartphone image. The results of Fig. 3.10 are somewhat difficult to interpret since there are no readily noticeable markers on the concerned surface. The effectiveness of the two approaches can, however, be appreciated by comparing their results. Both approaches result in mosaics that have comparable overall shape even though different alternative paths were found by the two approaches.

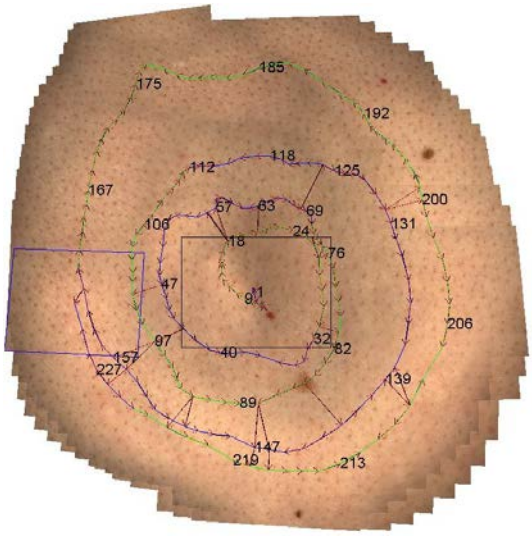
The computation time of both the approaches was relatively high (~ 20 to ~ 40 minutes) for all three sequences. This is explained, in part, by large distortions in the sequential trajectory of the initial mosaic, which results in the selection of a large number of false candidates for non-sequential overlapping image pairs. The registration testing of these pairs is computationally expensive. Sequences of Figs. 3.8 and 3.10 required more time (~ 20 and ~ 40 minutes respectively) because relatively more radial links were found, that consequently required registration of more additional pairs. Unusually high time (~ 40) of the sequence in Fig. 3.10 is also due to a large number of detected SURF keypoints. Around 1000 keypoints were detected for images in this sequence after thresholding at $0.2S_{max}$. For comparison, a few hundred keypoints were detected on the images of the other sequences whose results are presented. An increase in the number of keypoints increases the computation time by requiring formulation and matching of more descriptors.



(a) Mosaic obtained using the SURF based registraion scheme by concatenating the homographies over the sequential path. The resulting mosaic has a size of 5177×5567 pixels.



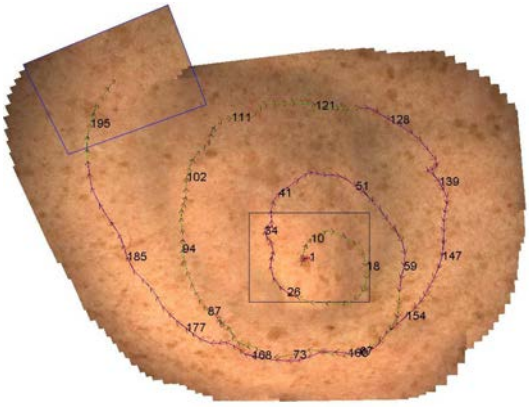
(b) Mosaic obtained by concatenating the homographies over paths obtained through the iterative approach. The panorama with a size of 4449×4501 pixels was obtained for folowing parameter values: $\delta_{max} = 0.2$, $\gamma_{max} = 0.5$. Updatd mosaic was obtained in 17 min.



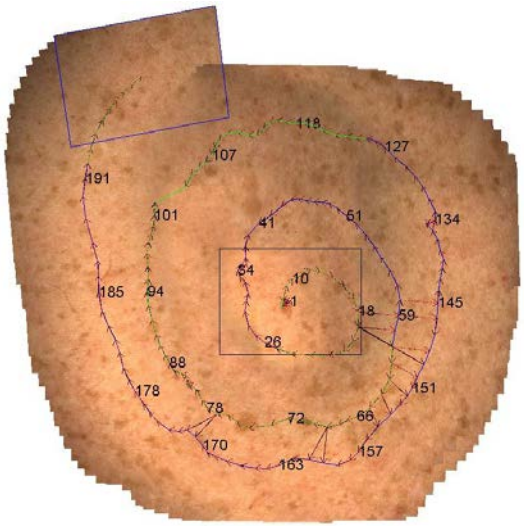
(c) Mosaic obtained by concatenating the homographies over paths obtained using the angle-based approach. The three stage algorithm with parameters $(\Delta\theta,\ \phi) = \{(90°,\ 15°),\ (60°,\ 10°),\ (30°,\ 5°)\}$ led to a mosaic with a size of 4461×4483 pixels. Updatd mosaic was obtained in 21 min.

(d) Smartphone image, taken with an active flash, of the corresponding area. The image size is 2620×2480 pixels.

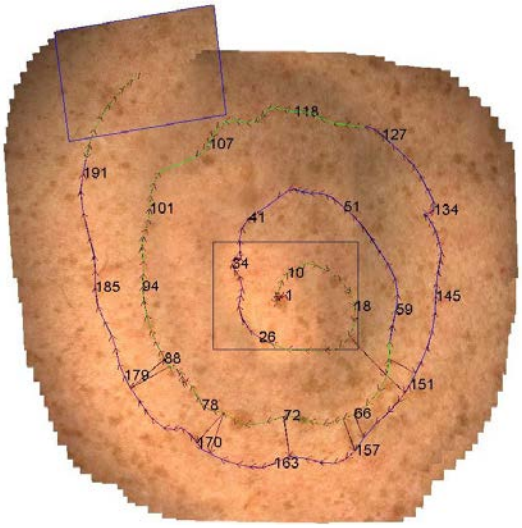
Figure 3.8: Mosaicing results obtained for a sequence over human back using the most direct trajectories. The black and blue quadrangles respectively represent the first and last frames of the sequence. The sequential trajectory is marked in green or blue (a change in color marks the beginning of a new spiral ring). The dashed red lines, with arrows indicating the direction, represent the most direct trajectories. The positions in the sequence of a few images are illustrated by the image numbers placed in the mosaic at the center of the images. The area of interest is of roughly 12×12 cm^2 size. The sequence contains 230 images all of which were mosaiced without skipping any.



(a) Mosaic obtained using the SURF based registraion scheme by concatenating the homographies over the sequential path. The resulting mosaic has a size of 5522×4179 pixels.



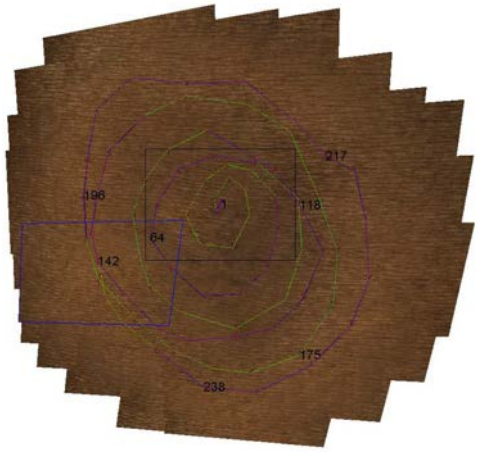
(b) Mosaic obtained by concatenating the homographies over paths obtained through the iterative approach. The panorama with a size of 4618×4630 pixels was obtained for folowing parameter values: $\delta_{max} = 0.2$, $\gamma_{max} = 0.5$. Updatd mosaic was obtained in 12 min.



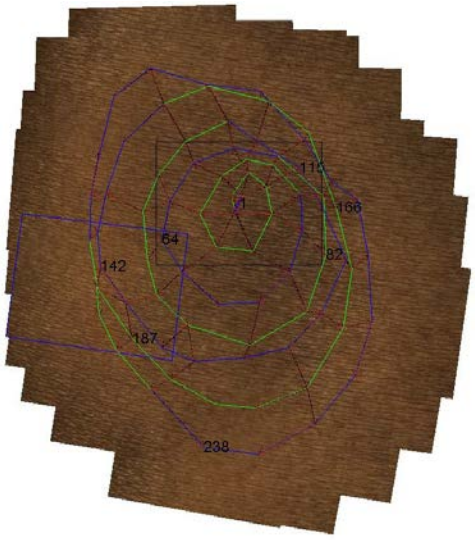
(c) Mosaic obtained by concatenating the homographies over paths obtained using the angle-based approach. The three stage algorithm with parameters $(\Delta\theta,\ \phi) = \{(90°,\ 15°),\ (60°,\ 10°),\ (30°,\ 5°)\}$ led to a mosaic with a size of 4515×4544 pixels. Updatd mosaic was obtained in 11 min.

(d) Smartphone image, taken with an active flash, of the corresponding area. The image size is 2288×1906 pixels.

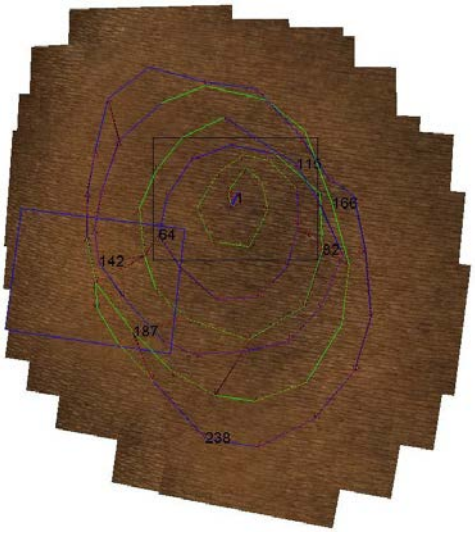
Figure 3.9: Mosaicing results obtained for a sequence over human back using the most direct trajectories. The black and blue quadrangles respectively represent the first and last frames of the sequence. The sequential trajectory is marked in green or blue (a change in color marks the beginning of a new spiral ring). The dashed red lines, with arrows indicating the direction, represent the most direct trajectories. The positions in the sequence of a few images are illustrated by the image numbers placed in the mosaic at the center of the images. The area of interest is of roughly 12×12 cm^2 size. The sequence contains 200 images all of which were mosaiced without skipping any.



(a) Mosaic obtained using the SURF based registraion scheme by concatenating the homographies over the sequential path. The resulting mosaic has a size of 4035×3964 pixels.



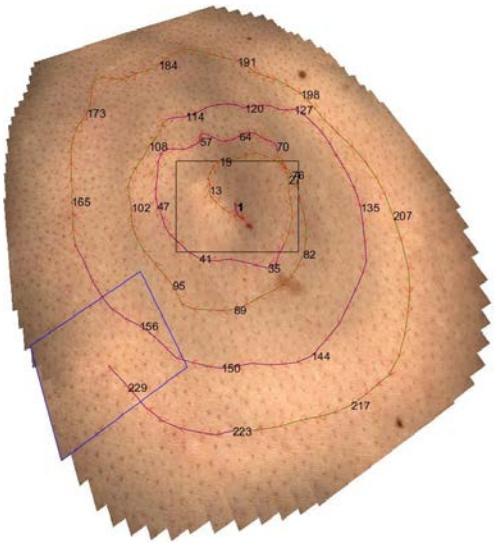
(b) Mosaic obtained by concatenating the homographies over paths obtained through the iterative approach. The panorama with a size of 3977×3771 pixels was obtained for folowing parameter values: $\delta_{max} = 0.2$, $\gamma_{max} = 0.5$. Updatd mosaic was obtained in 44 min.



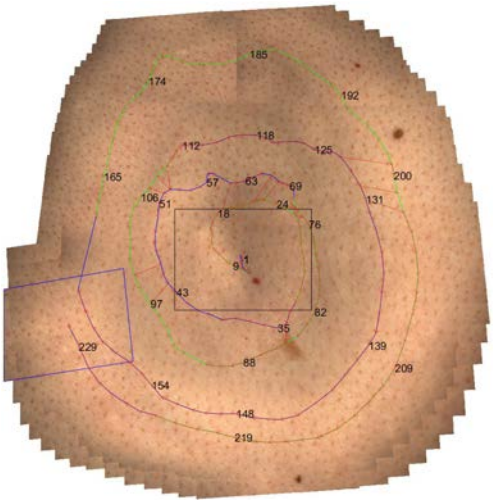
(c) Mosaic obtained by concatenating the homographies over paths obtained using the angle-based approach. The three stage algorithm with parameters $(\Delta\theta,\ \phi) = \{(90°,\ 15°),\ (60°,\ 10°),\ (30°,\ 5°)\}$ led to a mosaic with a size of 3584×4069 pixels. Updatd mosaic was obtained in 38 min.

(d) Smartphone image, taken with an active flash, of the approximate corresponding area. The image size is 3516×4053 pixels.

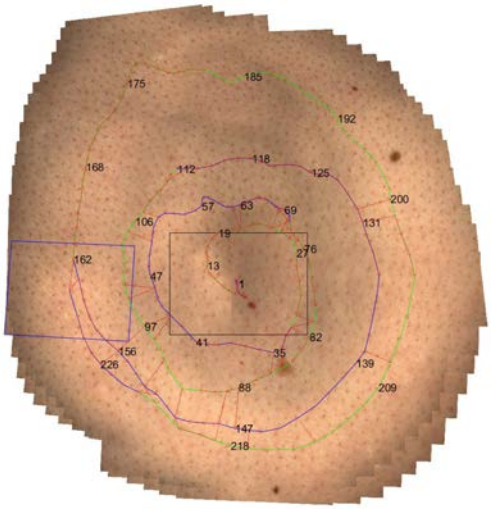
Figure 3.10: Mosaicing results obtained for a sequence over human back (dark color skin) using the most direct trajectories. The black and blue quadrangles respectively represent the first and last frames of the sequence. The sequential trajectory is marked in green or blue (a change in color marks the beginning of a new spiral ring). The dashed red lines, with arrows indicating the direction, represent the most direct trajectories. The positions in the sequence of a few images are illustrated by the image numbers placed in the mosaic at the center of the images. The area of interest is of roughly 10×10 cm^2 size. The sequence contains 250 images, every third of which was mosaiced.



(a) Mosaic obtained using the BRISK based registraion scheme by concatenating the homographies over the sequential path. The resulting mosaic has a size of 5078×5582 pixels.



(b) Mosaic obtained by concatenating the homographies over paths obtained through the iterative approach. The panorama with a size of 4528×4607 pixels was obtained for folowing parameter values: $\delta_{max} = 0.2$, $\gamma_{max} = 0.5$. Updatd mosaic was obtained in 9 min.



(c) Mosaic obtained by concatenating the homographies over paths obtained using the angle-based approach. The three stage algorithm with parameters $(\Delta\theta,\ \phi) = \{(90°,\ 15°),\ (60°,\ 10°),\ (30°,\ 5°)\}$ led to a mosaic with a size of 4471×4632 pixels. Updatd mosaic was obtained in 6 min.

(d) Smartphone image, taken with an active flash, of the corresponding area. The image size is 2620×2480 pixels.

Figure 3.11: Mosaicing results obtained for a sequence over human back using the most direct trajectories. The black and blue quadrangles respectively represent the first and last frames of the sequence. The sequential trajectory is marked in green or blue (a change in color marks the beginning of a new spiral ring). The dashed red lines, with arrows indicating the direction, represent the most direct trajectories. The positions in the sequence of a few images are illustrated by the image numbers placed in the mosaic at the center of the images. The area of interest is of roughly 12×12 cm^2 size. The sequence contains 230 images all of which were mosaiced without skipping any.

The real sequence mosaics and the mosaicing update results discussed so far were all obtained using the SURF based approach. SURF was used under a compromise between speed and accuracy. In the comparisons made in section 2.6.3, it was pointed out that BRISK could potentially detect more correspondences in the high-resolution image pairs of the real sequences. Fig. 3.11 shows the mosaicing and topology update results obtained using the BRISK based approach for a sequence acquired over the human back. This sequence is the same that was used to obtain the SURF-based results shown in Fig. 3.8. The mosaic of Fig. 3.11(a), in which the homographies are concatenated over the sequential path, was obtained by using the BRISK based image registration. A comparison with the mosaic obtained by the SURF based approach (see Fig. 3.8(a)) shows that BRISK was as successful as SURF in registering all the sequential image pairs, with BRISK being over 10 times faster than SURF. The image registration time using SURF varied from 5 s to over 10 s depending on the number of keypoints detected. The registration time using BRISK was less than 0.5 s/pair (for comparison, SIFT-based registration of an image pair in the same sequence is achieved in more than one minute). It is recalled that fast descriptor matching using the Hamming distance contributes to the reduction in BRISK's computation time.

The topology update results indicate that BRISK was less successful than SURF in registering the non-sequential image pairs. The iterative approach found just a few radial links (see Fig. 3.11(b)). Although these radial links largely improved the mosaic, some seams indicate a misalignment in the mosaic. The angle-based approach, on the other hand, was successful in finding several radial links (see Fig. 3.11(c)). However, there are some segments for which no radial link was found, resulting in misalignment, as is indicated by the blur in the bottom left corner of the mosaic. The mosaic update using the BRISK based approach was considerably faster than the SIFT-based one (the computation times are indicated under the respective mosaics). Considering the significantly less computation time of the BRISK based image registration, the time taken for the mosaic update was still higher than one might expect. This is explained partly by a large number of failed registrations of the candidate non-sequential overlapping pairs, which required several MST updates in the iterative approach and led to an almost exhaustive testing of the candidate pairs in the angle-based approach. About half of the computation time results from a less optimized implementation/handling of the topology structure and the overall mosaicing pipeline.

3.3 Global Adjustment

The topology refinement and detection of shorter paths already significantly improve the overall mosaic. However, some misalignment may occur over a large segment of sequentially overlapping images for which no alternative path through non-consecutive image pairs is found. This produces a case where about half the images in a segment are reached through one path and the remaining through another path. Such a situation can be observed in the results shown in Figs. 3.9 and 3.10. The image pair at the junction of two paths may not be well aligned if the accumulated error over one or both paths is significant, resulting in ghost textures. A global adjustment scheme is sought to overcome this problem. Closed-trajectories (loop-shaped paths) formed across the spiral rings can be exploited to redistribute the errors in the estimated homographies in a controlled way such that the images at the junction are aligned without significantly affecting the alignment of the other images over the considered paths.

The proposed global adjustment scheme [FBD17] seeks to adjust the estimated homographies $\{H_{i,i+1}^{est}\}$ between consecutive images I_i and I_{i+1} of a sequence forming a closed loop. Keypoints

detected by a descriptor based approach (SIFT, SURF or BRISK) are exploited for this purpose. For each consecutive image pair (I_i, I_{i+1}), the keypoints classified as inliers by the RANSAC algorithm and given in homogeneous coordinates as vectors $\mathbf{x}_D^i$ and $\mathbf{x}_D^{i+1}$ in Eq. (3.4) are used to formulate an energy whose minimization will result in a set of optimized homographies $\{H_{i,i+1}^{opt}\}$. To ensure that the variations in homographies do not cause considerable misalignment in successive image pairs, the differences between the descriptor coordinates mapped, from one image to the other, with $\{H_{i,i+1}^{est}\}$ and the ones mapped with $\{H_{i,i+1}^{opt}\}$ are minimized in both forward ($I_i \rightarrow I_{i+1}$) and backward ($I_{i+1} \rightarrow I_i$) directions. The minimization of the energy in Eq. (3.5) is performed under the constraint given in Eq. (3.6). The imposed constraint is that the direct homography $H_{1,N}^{est}$ between the first and the last loop images equals the concatenated homographies between these two images:

$$E_{error} = \underbrace{\sum_{i=1}^{N-1} \left\| H_{i,i+1}^{est}\mathbf{x}_D^i - H_{i,i+1}^{opt}\mathbf{x}_D^i \right\|}_{forward\ projection\ error} + \underbrace{\sum_{i=1}^{N-1} \left\| H_{i+1,i}^{est}\mathbf{x}_D^{i+1} - H_{i+1,i}^{opt}\mathbf{x}_D^{i+1} \right\|}_{backward\ projection\ error} \tag{3.4}$$

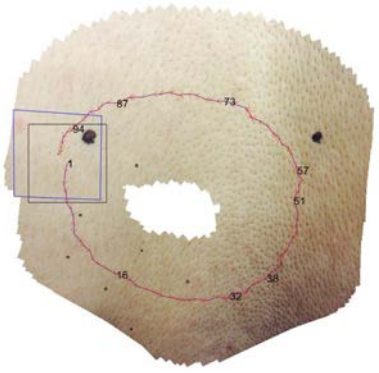
$$\{\tilde{H}_{1,2}^{opt}, \tilde{H}_{2,3}^{opt}, ...\tilde{H}_{N-1,N}^{opt}\} = \underset{\{H_{1,2}^{opt}, H_{2,3}^{opt}, ...H_{N-1,N}^{opt}\}}{argmin} (E_{error}) \tag{3.5}$$

$$\text{Subject to:} \quad \left\| H_{1,N}^{est} - \left(\prod_{i=1}^{N-1} H_{i,i+1}^{opt} \right) \right\| = 0, \tag{3.6}$$

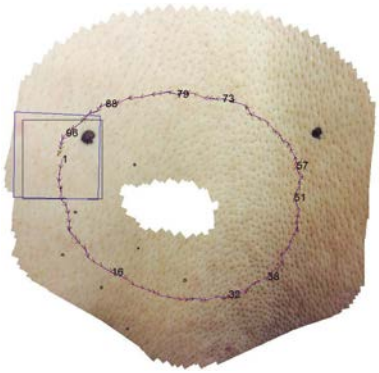
where N is the number of images. This global adjustment approach is referred to as constrained adjustment.

Although a set of predefined homologous coordinates on the pairwise registered images can serve the same purpose as the descriptor positions, it would be with the assumption that the initial homographies $\{H_{i,i+1}^{est}\}$ are precise. Optimization with respect to the descriptor locations provides more robustness because of less disruption in the optimized homographies, which are obtained through the least square fit over the matched descriptors. Moreover, the constraint given in Eq. (3.6) can be directly incorporated into Eq. (3.4), instead of an explicit formulation. This global adjustment variant is referred to as unconstrained adjustment. The results for the two adjustment variants are compared hereafter on different simulated sequences.

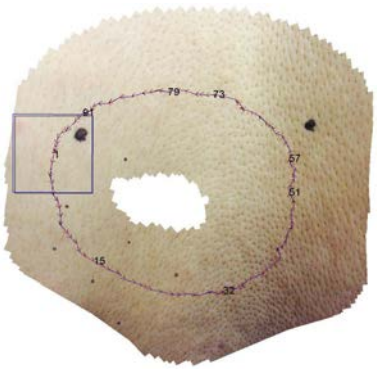
Validation tests were performed on two closed-loop sequences (seq-I and seq-II). These sequences are simulated from a high-resolution image of the human back by extracting overlapping images over a closed trajectory. Each image is extracted from a region corresponding to the mapping with the simulated homography of a 512×512 pixel grid. The simulated homographies are such that they represent realistic manual moving of the camera, i.e. the variations in the homography parameters between the sequential pairs are smooth. Since the objective is to redistribute registration errors over closed paths, these sequences are suitable for preliminary tests. The effectiveness of the proposed method is first demonstrated on seq-I (see Fig. 3.12(d)) mosaiced using BRISK based approach because this mosaic results in a significantly open loop (see Fig. 3.12(a)) and thus provides the opportunity to test the method against large errors. 30 randomly selected matched keypoints were used in the global adjustment process. The matched points resulting in more than 0.5 pixel projection error were ignored so that only precise correspondences influence the homography optimization. The unconstrained and constrained optimization problems were solved using the Levenberg-Marquardt algorithm, available through the function *lsqnonlin*



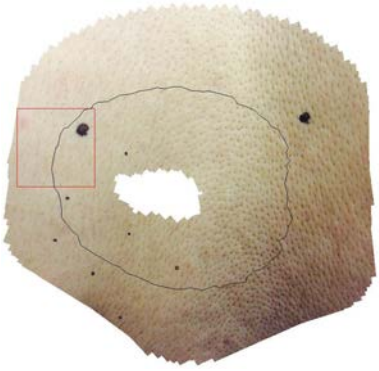
(a) Mosaic obtained without global adjustment and using homologous BRISK descriptors. Mean $\varepsilon_{i,i+1}^{reg} = 0.58$ pixel, $\varepsilon_{1,N}^{mos} = 76.6$ pixels.



(b) Mosaic after unconstrained adjustment. Optimization time = 8.1 s. Mean $\varepsilon_{i,i+1}^{reg} = 0.59$ pixel, $\varepsilon_{1,N}^{mos} = 37.5$ pixels.



(c) Mosaic after constrained adjustment. Optimization time = 40.0 s. Mean $\varepsilon_{i,i+1}^{reg} = 0.68$ pixel, $\varepsilon_{1,N}^{mos} = 10.8$ pixels.



(d) Ground truth mosaic

Figure 3.12: Mosaic of a human dorsal region skin before and after adjustment using two optimization models. The simulated sequence contains 101 images with a 512×512 pixel size (representing approximately 2×2 cm^2 skin area) each. The red square represents the first image and blue quadrangle the last one. The traced line shows the image center trajectory. Thirty homologous keypoints were used for each pair in the optimization process.

in MATLAB®'s optimization toolbox, and the interior-point algorithm, available through the function *fmincon* in the same toolbox, respectively.

Figs. 3.12(b) and 3.12(c) show the results after global adjustment of the sequence given in Fig. 3.12(a) by using the two approaches. Although pairwise registration precision was well-preserved using the unconstrained adjustment, with the convergence achieved in 2000 iterations, the loop failed to close (see Fig. 3.12(b)). With the constrained adjustment, although more iterations (up to 6000) were required for sufficient convergence, the loop was closed with only a small increase in mean registration error, as shown in Fig. 3.12(c). Pairwise registration error comparison of the sequence before and after adjustment is shown in Fig. 3.13. It can be observed that the unconstrained adjustment preserved very well the registration accuracy of consecutive images, as is indicated by almost overlapping plots of the registration errors before and after

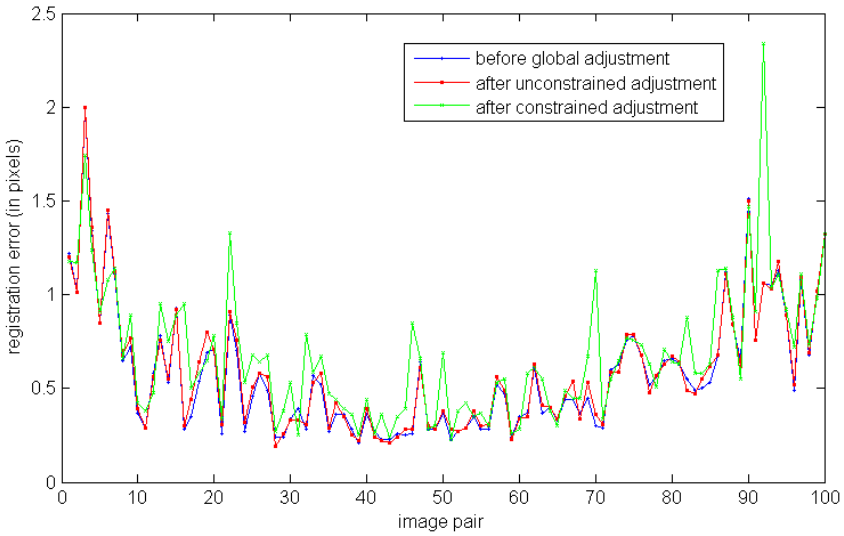
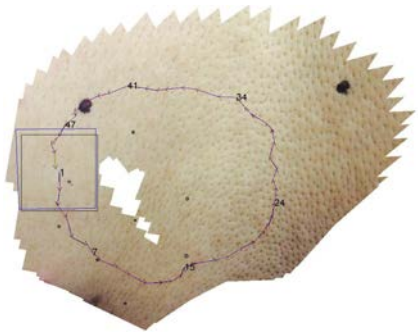


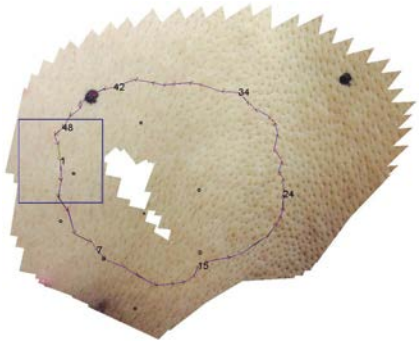
Figure 3.13: Pairwise registration error comparison before and after global adjustment using two models. These curves were obtained for the mosaics given in Fig. 3.12.

adjustment using this approach. A few peaks in the error plot of the sequence adjusted using constrained optimization indicates that a misalignment resulted in the image pairs corresponding to these peaks. Precisely, the aim being to close the loops, such registration error increase in some image pairs is inevitable to reduce the mosaicing errors.

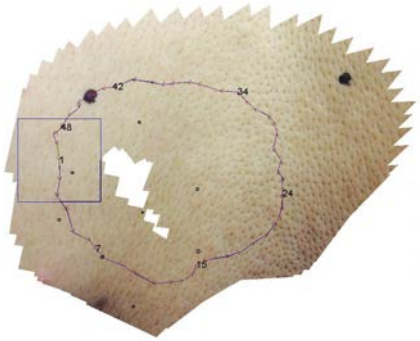
A fixed number of keypoint locations was chosen for the ease of calculation since the number of inliers varies among the registered image pairs. The impact of varying this number is explored by various global adjustment tests on a mosaic of seq-II. Fig. 3.14(a) shows the mosaic of this sequence obtained using the SURF based registration. The selection of the matched keypoints for homography computation was made using the approach presented in section 2.6.4. The resulting keypoints range from under 100 to over 300. The tests were performed using 30, 100 and 200 inlier locations, resulting in adjusted mosaics shown in Figs. 3.14(b), 3.14(c)and 3.14(d). The inliers were selected randomly and if an image pair had fewer inliers than the selected number, the empty places were filled by duplicating some keypoints. It was noticed that as less as 30 keypoint locations were already sufficient to achieve a coherent global adjustment. An increase in the number of selected inliers did not have any significant impact in visually improving the mosaic. Although this increase slightly reduced the errors in the pairwise registration accuracy, the constraint was less precisely fulfilled with 200 inliers, as indicated in Fig. 3.14(d) by the mosaicing error after adjustment, which is slightly larger than that for 30 and 100 inliers (the reader may compare the overlap of the first and last rectangles in Figs. 3.14(b) to 3.14(d) which is the worst in Fig. 3.14(d)). Pairwise precision comparison before and after the adjustment with 100 inlier locations is shown in Fig. 3.15. The plot indicates that the error is homogeneously redistributed without significantly affecting the registration accuracy. Fig. 3.16 shows the global adjustment with constrained optimization of seq-II mosaiced using the SIFT-based approach. In this case, the mosaicing error being small before adjustment, the loop was almost perfectly closed after the adjustment, as is indicated by the decrease of the mosaicing error to just 0.7 pixels. A visually coherent global adjustment in the mosaic is obtained by using the descriptor locations to control the variations in consecutive homographies. In the tests performed on a closed-loop sequence,



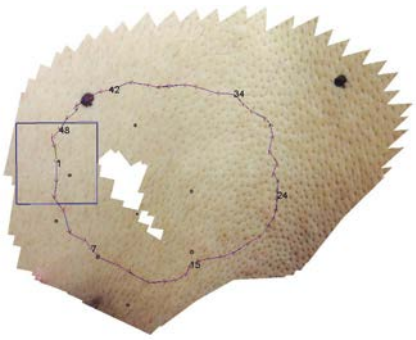
(a) Mosaic obtained without global adjustment and using homologous SURF descriptors. Mean $\varepsilon_{i,i+1}^{reg} = 0.47$ pixel, $\varepsilon_{1,N}^{mos} = 28.6$ pixels.



(b) Mosaic after constrained adjustment using 30 randomly selected keypoint correspondences. Optimization time = 16.6 s. Mean $\varepsilon_{i,i+1}^{reg} = 0.57$ pixel, $\varepsilon_{1,N}^{mos} = 2.8$ pixels.



(c) Mosaic after constrained adjustment using 100 randomly selected keypoint correspondences. Optimization time = 26.2 s. Mean $\varepsilon_{i,i+1}^{reg} = 0.55$ pixel, $\varepsilon_{1,N}^{mos} = 1.7$ pixels.



(d) Mosaic after constrained adjustment using 200 randomly selected keypoint correspondences. Optimization time = 41.9 s. Mean $\varepsilon_{i,i+1}^{reg} = 0.51$ pixel, $\varepsilon_{1,N}^{mos} = 5.7$ pixels.

Figure 3.14: Global adjustment comparison for the different number of keypoint correspondences. The simulated sequence contains 51 images with a 512×512 pixel size (representing approximately 2×2 cm^2 skin area) each. The red square represents the first image and blue quadrangle the last one. The traced line shows the image center trajectory in the mosaic coordinate system.

the loop is almost closed by enforcing the constraint that the concatenated homographies match the direct homography between the first and the last images. This adjustment strategy can be extended to the real sequences, acquired over a spiral trajectory, by identifying the loops formed through the radial links.

Applicability to the real sequence is demonstrated with the adjustment of a loop of the mosaiced sequence of Fig. 3.17(a). The image pair (I_{102}, I_{103}) forms a junction of two long paths (left and right path sides in Fig. 3.17(a)), each used for accessing images at both sides of this pair. Consequently, these two images are misaligned, as indicated by a relatively large gap between the images I_{102} and I_{103}. A constrained adjustment was performed over the closed path $\{I_{102}, ..., I_{58}, I_{145}, ..., I_{103}\}$ (the closed-loop trajectory is indicated with red bold line) with the constraint that the accumulated homography over this path equals $H_{102,103}^{est}$. The concerned

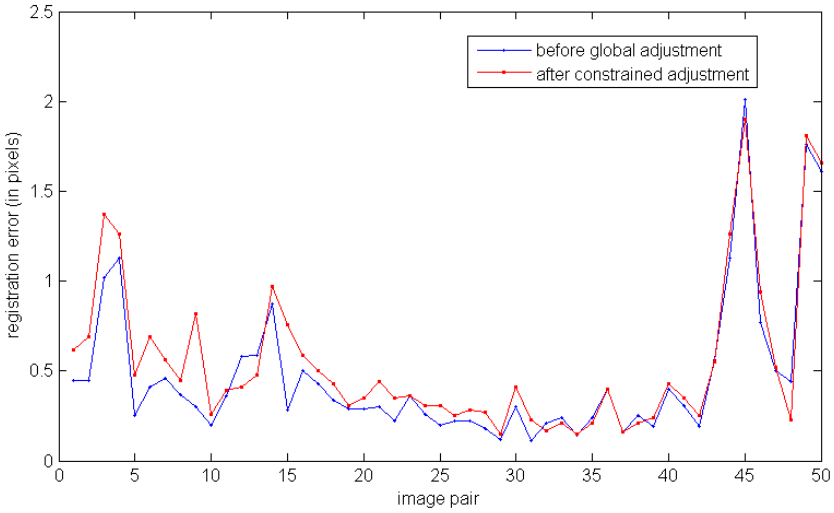
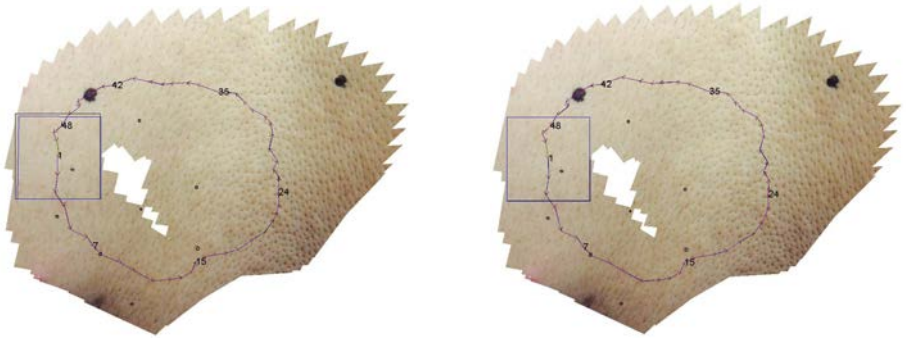


Figure 3.15: Pairwise registration error comparison before and after global adjustment of the mosaic of Fig. 3.14(a) using an explicit constraint formulation with 100 randomly selected SURF keypoint correspondences.

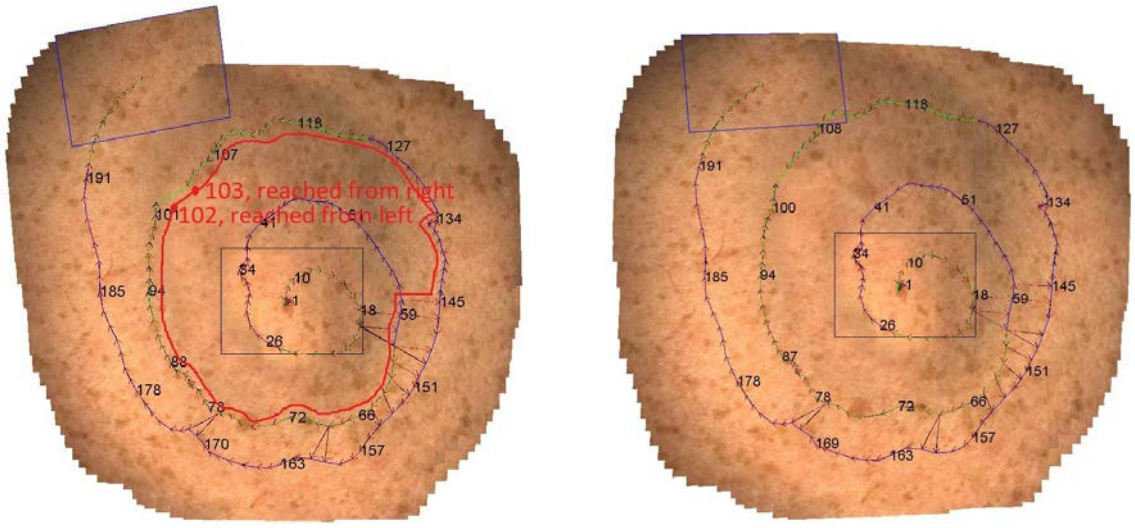


(a) Mosaic obtained without global adjustment and using homologous SIFT descriptors. Mean $\varepsilon_{i,i+1}^{reg} = 0.19$ pixel, $\varepsilon_{1,N}^{mos} = 15.2$ pixels.

(b) Mosaic after constrained adjustment using 100 randomly selected keypoint correspondences. Optimization time = 24.8 s. Mean $\varepsilon_{i,i+1}^{reg} = 0.23$ pixel, $\varepsilon_{1,N}^{mos} = 0.7$ pixel.

Figure 3.16: Global adjustment of a sequence mosaiced using the SIFT-based registration scheme. The simulated sequence contains 51 images with a 512×512 pixel size (representing approximately 2×2 cm^2 skin area) each. The red square represents the first image and blue quadrangle the last one. The traced line shows the image center trajectory in the mosaic.

images are pretty much aligned with minimal disturbance in the rest of the mosaic after the adjustment. A complete adjustment scheme, that would automatically detect the long loops and would perform simultaneous optimization on all the detected loops, remains to be realized. The automatic loop detection in mosaics after topology inference was not implemented yet. However, this first encouraging result indicates that such an automated loop detection and global adjustment algorithm after topology inference can potentially further improve the visual quality



(a) Mosaic before global adjustment

(b) Mosaic after global adjustment of one loop

Figure 3.17: Application of the global adjustment scheme on the real sequence from Fig. 3.9. The adjusted loop is formed over the path $\{I_{102}, ..., I_{58}, I_{145}, ..., I_{103}\}$. The homography errors were redistributed over this path under the constraint that the accumulated homography over this path equals $H^{est}_{102,103}$.

of mosaics.

3.4 Conclusion and Discussion

The application of the proposed mosaicing process to the real sequences was presented in this chapter. It was observed that the error accumulation over long sequences creates significant geometrical distortions in the mosaic. Therefore, an appropriate scheme for mosaicing large sequences was sought. Although the comparison of different registration approaches indicated that a more precise method can improve the overall mosaic, the issue of distortion over long sequences still persists due to the limited precision of the registration approaches. Besides, the failed registrations also posed a problem for mosaicing the images following the registration failure point in a sequence. Several existing works involving the creation of large panoramas propose to solve these issues with image topology refinement. This refinement is based on the initial topology estimate that gives the spatial relationship between the consecutive images in the sequence. Topology is updated after finding the spatial links between non-consecutive pairs and registering some selected pairs with significant overlap. The registration of non-consecutive pairs permits reaching a given image via a shorter path from the reference image, which in turn reduces the error accumulation resulting in a better topology estimate and a mosaic closer to the mosaiced surface. Some trajectory overlaps and/or crossings are required for this type of approaches to be feasible. Acquisition over a spiral trajectory was chosen due to some advantages, which include ease of manual movement of the camera and the possibility of finding several radial paths from the reference image placed at the center of the spiral. Moreover, such a camera trajectory provides the possibility of estimating the position of the images for which the pairwise registration over the acquisition trajectory fails.

Two approaches for finding the alternative and more direct paths were proposed. One approach lies in the adaptation of an existing work that builds a connectivity graph of the images

in the sequence. The connectivity graph, whose nodes represent the images, initially links only the sequential images. Additional links between the non-consecutive image pairs are iteratively added when the corresponding image pairs have significant overlap and whose addition would result in considerably shorter paths for a large number of images. After computation of the connectivity graph from these criteria, the alternative paths through a minimum spanning tree (MST), which gives the shortest path between two nodes in a graph, were found for all the images. The alternative paths calculated in this way may not necessarily provide a continuous path since the registration of some additional images may fail. Therefore, the registration of all the image pairs in the alternative paths was tested using some proposed criteria. When some alternative paths turned out to be discontinuous, MST was recalculated after removing the links corresponding to the failed registrations from the connectivity graph. After successfully finding the continuous paths, the sequence topology was updated and the entire process was repeated to further refine the topology. Although this scheme significantly improved the results, the computation of additional links in an iterative manner becomes computationally expensive as the size of the connectivity graph (both number of nodes and links) increases. Recalculation of MST for finding alternative paths in case of discontinuous paths adds to the computation time if the initial topology estimate is less precise due to spatial distortion of the mosaic. The second approach, that exploits the spiral trajectory for finding the alternative paths is designed to provide more control over the process.

Over a sequential trajectory, the overlapping images across the consecutive spiral links can be predicted from the initial topology estimate. An angle-based approach for finding the alternative paths looked for such images at predetermined positions at the spiral rings by considering the angle formed by that position with the spiral center. Overlapping image pairs across the consecutive rings were searched at regular intervals. Some tolerance in the relative position over the consecutive rings was permitted to account for the distortion in the initial topology. This tolerance results in several potentially overlapping pairs at the considered locations. The registration of the images pairs lying within a defined tolerance limit was tested in their order of appearance. The first successful image pair for each location was added to the connectivity graph, that initially connected the sequential image pairs only. The graph computed in this way provided the certitude that the alternative paths were continuous, thus necessitating the MST calculation just once. Although the registration of the additional pairs is still time-consuming, the identification of the candidate non-consecutive image pairs is instantaneous contrary to the iterative approach. The topology with the angle-based approach was updated in three steps with a gradual increase in angle interval and a decrease in position tolerance at each step. This dynamic refinement was used to minimize the number of additional image registrations that needed to be tested.

Mosaics of several sequences acquired on different skin surfaces of different subjects showed the effectiveness of both approaches. Coherent and high-resolution panoramas of extended surfaces were obtained, and their shape conformed to the mosaiced surface. The possibility of mosaicing a sequence with failed registrations was also demonstrated. The alternative paths helped to take into account the images which could not be reached due to an interruption in the sequential path. A comparison of BRISK and SURF for sequential mosaicing and mosaicing update with refined topology showed that while both approaches were successful in registering the sequential image pairs, SURF was more effective in the registration of non-sequential pairs. Moreover, an approach for fast global adjustment was presented and its effectiveness was demonstrated through simulated sequences. This approach seeks to reduce the misalignment of the images for which shorter alternative paths were not found. Its applicability on the real sequences was demonstrated.

Conclusion and Perspectives

Summary of Work Done

Among the image registration approaches that were tested for skin image registration, some gave consistently successful results over large displacements. Since precision was not the only efficiency criterion, SURF, although not proved to be the best in terms of accuracy, was finally selected as a method of choice for this study due to its considerable lower computation time in comparison to all other tested methods. Besides, the comparison of mosaicing using different approaches showed that the mosaic of longer sequences was likely to be distorted due to error accumulation. Some distortion persisted despite registration with the sub-pixel accuracy of all the image pairs. In addition, the mosaicing process was prone to be interrupted due to a single failed registration. In light of these considerations, rather than focusing on improving the registration accuracy (which might be of more importance in applications involving registration of a couple of or a few images as well as achievement of super-resolution from images of the same scene acquired from the same point of view), more attention was given to the overall mosaicing process. Nevertheless, importance was given to the homography computation accuracy in the image registration step since, even though small variations in the registration of image pairs do not have a significant impact on the overall mosaic, a failed registration considerably distorts the mosaic and can interrupt the entire process. A novel approach for the refinement of the initially established keypoints correspondences using SURF was proposed. The objective behind this refinement was to improve the probability of the correct homography estimation through RANSAC, which randomly selects the keypoint locations detected in the two images (even if RANSAC is designed to statistically reduce the effect of the outliers through multiple trials, probability of the correct computation remains proportional to the ratio of inliers to outliers). Moreover, the number of inlier points detected by RANSAC and the area ratios of two consecutive images warped using the estimated homographies were used for defining the criteria for detecting potentially failed registrations (i.e. the registration is considered as failed either for too few inlier point pairs or for too large difference in the areas, after warping, of two adjacent images). Image pairs involving failed registration were discarded so that the proposed mosaicing scheme, which is designed to handle such cases, may allow the mosaicing process to continue without interruption.

For the mosaicing process, a new strategy for estimating the trajectory topology of the images forming the video sequence was proposed. The objective of topology estimation was to locate non-consecutive image pairs with sufficient overlap for accurate registration. Acquisition along a spiral trajectory was opted for to increase the possibility of having a large number of such overlaps. Since the registration of all the overlapping images is too expensive in terms of computation time, two approaches for selecting the optimally located overlapping images were proposed and implemented. One approach exploited an existing scheme [MFM04] for iteratively selecting the overlapping pairs that would reduce the path length while ignoring the ones that, even though had significant overlap, did not significantly reduced the already calculated paths.

In another approach, non-consecutive image pairs with successful registration were searched at regular intervals at the consecutive spiral rings. In both approaches, the image at the center of the spiral was taken as the reference image and Dijkstra's algorithm for finding the shortest paths [Dij59] was used to find the paths connecting the other images to the reference image. The evaluations performed on the simulated sequences showed the success of both the approaches in finding non-consecutive image pairs with successful registration. The tests performed on real sequences showed the effectiveness of these approaches in improving the coherency of the mosaic. Mosaics obtained from sequences acquired over different parts of the body and over different subjects showed an increased resolution compared to the smartphone images, apart from conserving the colorimetric characteristics of the skin without much illumination variation. The seams caused by illumination variations were not so prominent and were easily suppressed with a basic feathering (blending). Besides, a global adjustment scheme for redistributing the homography errors to obtain a more coherent mosaic was developed by exploiting the keypoints matched by descriptor-based approaches. The effectiveness of this approach was demonstrated on simulated sequences and its applicability on the real sequences was demonstrated.

Perspectives

Some improvements can be made in the topology refinement and the image placement on the mosaicing plane. Although a spiral trajectory increases the number of overlapping images, too many overlaps and crossings of the spiral rings can result in ghost textures. Other acquisition patterns could be explored, for example, a zig-zag trajectory with the reference image placed in the center of the acquisition zone. In addition, instead of stitching all the constituent images of the sequence to form the final mosaic, the placement of a few selected images depending on their mutual overlap could reduce the ghost textures.

The presented global adjustment approach takes into account only homographies of the consecutive image pairs along with their descriptor locations in the overlapping regions. For a potentially more coherent adjustment, it could be interesting to consider several overlapping images simultaneously for the homography refinement, as done in [BL03]. This would require the identification of keypoints common to a set of overlapping images. In [BL03], the homography parameters were adjusted, without any additional constraint, to minimize the distance between the same keypoints detected in different overlapping images. This could be combined with the constraint used in this study [FBD17] by finding the groups of overlapping images but would require further considerations for the formulation and solution of the resulting optimization problem. A better global adjustment scheme may also help with the use of BRISK based registration scheme in the mosaicing process. Although BRISK proved to be as effective as SURF for registering sequential image pairs, it had considerably less success than SURF in registering the non-sequential image pairs. Since BRISK is significantly faster than SURF, the proposed angle-based topology refinement scheme may be used in combination with BRISK to find radial links at shorter intervals without adding much computational overload. Consequently, a better global adjustment scheme combined with BRISK may be helpful in faster panoramic construction without compromising the precision.

An important limitation of 2D mosaicing approaches applied to 3D scenes is that perspective deformation parameters may not be recovered as precisely as required to avoid large perspective distortion due to the accumulation of errors inherent to the concatenation of imperfect homographies over long paths. Besides, the planarity assumption might be less valid for the curved skin surface. For a plane surface, such as back, sequences containing up to 100 images and covering

about 20 cm^2 surface area can be mosaiced generally without distortion. However, for highly curved organs, such as the finger, the mosaic starts to diverge after about 50 images covering about 10 cm^2 surface area. For the organs with intermediary curvatures, such as wrist or ankle, the results are somewhat erratic, i.e. some sequences containing over 100 images are coherently mosaiced and others result in a distorted mosaic. A 3D mosaicing approach, although costly in computation time, might help overcome this limitation, as described in the feasibility study of the internal bladder wall reconstruction [BHDS16]. However, such an active vision-based surface reconstruction and enlargement method also require additional instrumental development. Moreover, consideration of practical aspects coming after clinical trials on real lesions will provide helpful feedback for alternative solutions in the proposed scheme.

Bibliography

[Ali+15] S. Ali, C. Daul, E. Galbrun, M. Amouroux, F. Guillemin, and W. Blondel, "Robust bladder image registration by redefining data-term in total variational approach", in *SPIE*, 2015.

[Amo+15] M. Amouroux, A. Haudrechy, K. Hill, and W. Blondel, "Morpho-functional optical diffusion imaging usable for vascular ulcers diagnosis during telehealth procedures", vol. 9792, Tokyo, Japan, 2015.

[Amo+17a] M. Amouroux, S. L. Cunff, A. Haudrechy, K. Hill, and W. Blondel, "Image quality assessment for teledermatology: From consumer devices to a dedicated medical device", in *Proceedings of Society of Photo-optical Instrumentation Engineers*, vol. 10056, 2017.

[Amo+17b] M. Amouroux, W. Blondel, A. Haudrechy, and K. Hill, "Wireless medical device for acquiring bimodal skin videos with light control", *Patent WO2017198575 (A1)*, filed May 18, 2016, and issued November 23, 2017.

[And91] R. R. Anderson, "Polarized light examination and photography of the skin", *Archives of dermatology*, vol. 127(7), pp. 1000–1005, 1991.

[AOV12] A. Alahi, R. Ortiz, and P. Vandergheynst, "Freak: Fast retina keypoint", 2012.

[BB09] I. R. Bristow and J. Bowling, "Dermoscopy as a technique for the early identification of foot melanoma", *Journal of Foot and Ankle*, vol. 2(14), 2009.

[Bea78] P. R. Beaudet, "Rotationally invariant image operators", in *Proceedings of the 4th International Joint Conference on Pattern Recognition*, 1978, pp. 579–583.

[BHDS16] A. Ben-Hamadou, C. Daul, and C. Soussen, "Construction of extended 3d field of views of the internal bladder wall surface: A proof of concept", *3D Research*, vol. 7(3), pp. 1–23, 2016.

[BL03] M. Brown and D. G. Lowe, "Recognising panoramas", in *Proceedings of International Conference on Computer Vision*, 2003.

[BM11] T. Brox and J. Malik, "Large displacement optical flow", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33(3), pp. 500–513, 2011.

[Bro+04] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, "High accuracy optical flow estimation based on a theory for warping", in *Proceedings of European Conference on Computer Vision*, 2004.

[BTG08] H. Bay, T. Tuytelaars, and L. V. Gool, "Surf: Speeded up robust features", *Computer Vision and Image Understanding*, vol. 110(3), pp. 346–359, 2008.

[Bö+15] A. Börve, J. D. Gyllencreutz, K. Terstappen, B. E. J. Backman, A. Aldenbratt, M. Danielsson, M. Gillstedt, C. Sandberg, and J. Paoli, "Smartphone teledermoscopy referrals: A novel process for improved triage of skin cancer patients", *Acta Dermato-Venereologica*, vol. 95(2), pp. 186–190, 2015.

[Cal+10] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "Brief: Binary robust independent elementary features", in *Proceedings of European Conference on Computer Vision*, 2010.

[CM05] O. Chum and J. Matas, "Matching with prosac – progressive sample consensus", in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005.

[CP11] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging", *Journal of Mathematical Imaging and Vision*, vol. 40(1), pp. 120–145, 2011.

[Dal+14] E. Dalimier, A. Bruhat, K. Grieve, F. Harms, F. Martins, and C. Boccara, "High resolution in-vivo imaging of skin with full field optical coherence tomography", in *Proceedings of Society of Photo-optical Instrumentation Engineers*, vol. 8926, 2014.

[Dew78] A. K. Dewdney, "Analysis of a steepest-descent image-matching algorithm", *Pattern Recognition*, vol. 10, pp. 31–39, 1978.

[Dij59] E. W. Dijkstra, "A note on two problems in connexion with graphs", *Numerische Mathematik*, vol. 1, pp. 269–271, 1959.

[DN13] M. Drulea and S. Nedevschi, "Motion estimation using the correlation transform", *IEEE Transactions on Image Processing*, vol. 22(8), pp. 3260–3270, 2013.

[EGG13] A. Elibol, N. Gracias, and R. Garcia., "Fast topology estimation for image mosaicing using adaptive information thresholding", *Robotics and Autonomous Systems*, vol. 61(2), pp. 125–136, 2013.

[EW01] D. J. Eedy and R. Wootton, "Teledermatology: A review", *British Journal of Dermatology*, vol. 144, pp. 696–707, 2001.

[Far+16] K. Faraz, W. Blondel, M. Amouroux, and C. Daul, "Towards skin image mosaicing", in *Proceedings of the 6th International Conference on Image Processing Theory, Tools and Applications*, Oulu, Finland, 2016.

[FB81] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography", *Communications of the Association for Computing Machinery*, vol. 24(6), pp. 381–395, 1981.

[FBD17] K. Faraz, W. Blondel, and C. Daul, "Global adjustment for creating extended panoramic images in video-dermoscopy", in *Proceedings of European Conferences on Biomedical Optics*, Munich, Germany, 2017.

[FFB96] M. Fleck, D. Forsyth, and C. Bregler, "Finding naked people", in *Proceedings of European Conference on Computer Vision*, vol. 2, 1996, pp. 593–602.

[FSV01] N. Fracias and J. Santos-Victor, "Underwater mosaicing and trajectory reconstruction using global alignment", in *Proceedings of IEEE OCEANS*, Honolulu, Hawaii, 2001.

[Fö92] W. Förstner, "A feature based correspondence algorithm for image matching", *International Archives of Photogrammetry and Remote Sensing*, vol. 26, pp. 150–166, 1992.

[Gos05] A. A. Goshtasby, *2-D and 3-D image registration: for medical, remote sensing, and industrial applications*. Wiley-Interscience, 2005, ISBN: 0471649546.

[HKR93] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge, "Motion estimation using the correlation transform", *IEEE Transactions on pattern analysis and machine intelligence*, vol. 15(9), pp. 850–863, 1993.

[HL06] B. Holmberg and H. Lanshammar., "Possibilities of texture based motion analysis", *Computer Methods and Programs in Biomedicine*, vol. 84(1), pp. 1–10, 2006.

[Hor+16] C. Horsham, L. J. Loescher, D. C. Whiteman, H. P. Soyer, and M. Janda, "Consumer acceptance of patient-performed mobile teledermoscopy for the early detection of melanoma", *British Journal of Dermatology*, vol. 175(6), pp. 1301–1310, 2016.

[Hou62] P. V. C. Hough, "Method and means for recognizing complex patterns", *U. S. Patent 3,069, 654*, 1962.

[HS81] B. K. P. Horn and B. G. Schunck, "Determining optical flow", *Artificial Intelligence*, vol. 17(1-3), pp. 185–203, 1981.

[HS88] C. Harris and M. Stephens, "A combined corner and edge detector", in *Proceedings of Alvey Vision Conference*, vol. 15(50), 1988, pp. 147–151.

[Jon82] A. D. Jones, "Manual of photogrammetry, eds c.c. slama, c. theurer and s.w. hendrikson, american society of photogrammetry, fourth edition.", *Cartography*, vol. 12, no. 4, pp. 258–258, 1982.

[JRRL02] S. L. Jacques, J. C. Ramella-Roman, and K. Lee, "Imaging skin pathology with polarized light", *Journal of Biomedical Optics*, vol. 7(3), pp. 329–340, 2002.

[Kas+13] R. Kassab, S. Treuillet, F. Marzani, C. Pieralli, and J. C. Lapayre, "An optimized algorithm of image stitching in the case of a multi-modal probe for monitoring the evolution of scars", in *Proceedings of Society of Photo-optical Instrumentation Engineers*, vol. 8572, 2013.

[KK07] K. Koser and R. Koch, "Perspectively invariant normal features", in *Proceedings of International Conference on Computer Vision*, 2007, pp. 1–8.

[KR82] L. Kitchen and A. Rosenfeld, "Gray-level corner detection", *Pattern Recognition Letters*, vol. 10, pp. 95–102, 1982.

[LCS11] S. Leutenegger, M. Chli, and R. Y. Siegwart, "Brisk: Binary robust invariant scalable keypoints", in *Proceedings of IEEE International Conference on Computer Vision*, 2011.

[Lev+06] A. Levin, A. Zomet, S. Peleg, and Y. Weiss, "Seamless image stitching in the gradient domain", *IEEE Transactions on Image Processing*, vol. 15(4), pp. 969–977, 2006.

[LJH15] C. Leitch, R. Jones, and S. A. Holme, "Smartphone teledermoscopy referrals: Comment on the paper by börve et al.", *Acta Dermato-Venereologica*, vol. 95(7), pp. 869–871, 2015.

[LK81] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision", in *Proceedings of Imaging Understanding Workshop*, vol. 2, 1981, pp. 121–130.

[Loe+08] K. Loewke, D. Camarillo, W. Piyawattanametha, D. Breeden, and K. Salisbury, "Real-time image mosaicing with a hand-held dual-axes confocal microscope", in *Proceedings of Society of Photo-optical Instrumentation Engineers*, vol. 6851, San Diego, United States, 2008.

[Low04] D. G. Lowe, "Distinctive image features from scale-invariant keypoints", *International Journal of Computer Vision*, vol. 60(2, pp. 91–110, 2004.

[Mag03] I. Maglogiannis, "Automated segmentation and registration of dermatological images.", *Journal of Mathematical Modelling and Algorithms*, vol. 2, pp. 277–294, 2003.

[Mas+09] C. Massone, A. M. Brunasso, T. M. Campbell, and H. P. Soyer, "Mobile teledermoscopy–melanoma diagnosis by one click?", *Seminars in Cutaneous Medicine and Surgery*, vol. 28(3), pp. 203–205, 2009.

[Mas+14] C. Massone, D. Maak, R. Hofmann-Wellenhof, H. P. Soyer, and J. Frühauf, "Teledermatology for skin cancer prevention: An experience on 690 austrian patients", *Journal of the European Academy of Dermatology and Venereology*, vol. 28(8), pp. 1103–1108, 2014.

[MFM04] R. Marzotto, A. Fusiello, and V. Murino, "High resolution video mosaicing with global alignment", in *Proceedings of Conference on Computer Vision and Pattern Recognition*, 2004.

[ML+04] R. Miranda-Luna, W. Blondel, C. Daul, Y. Hernandez-Mier, and D. Wolf, "A simplified method of video-endoscopic image barrel distortion correction based on grey level registration", in *Proceedings of IEEE International Conference on Image Processing*, Singapore, 2004, pp. 3383–3386.

[ML+08] R. Miranda-Luna, C. Daul, W. Blondel, Y. Hernandez-Mier, D. Wolf, and F. Guillemin, "Mosaicing of bladder endoscopic image sequences: Distortion calibration and registration algorithm", *IEEE Transactions on Biomedical Engineering*, vol. 55(2), pp. 541–553, 2008.

[Mor80] H. P. Moravec, "Obstacle avoidance and navigation in the real world by a seeing robot rover", PhD thesis, Stanford University, 1980.

[MS01] B. R. Masters and P. T. C. So, "Confocal microscopy and multi-photon excitation microscopy of human skin in vivo", *Seminars in Cutaneous Medicine and Surgery*, vol. 8(1), 2001.

[Nam+15] N. Nami, C. Massone, P. Rubegni, G. Cevenini, M. Fimiani, and R. Hofmann-Wellenhof, "Concordance and time estimation of store-and-forward mobile teledermatology compared to classical face-to-face consultation", *Journal of the European Academy of Dermatology and Venereology*, vol. 95(1), pp. 35–39, 2015.

[NKP09] S. W. Noh, H. J. Kong, and S. Y. Park, "Registration of finger vein image using skin surface information for authentication", in *Proceedings of Society of Photo-optical Instrumentation Engineers*, 2009.

[Nou+09] A. Nouvong, B. Hoogwerf, E. Mohler, B. Davis, A. Tajaddini, and E. Medenilla, "Evaluation of diabetic foot ulcer healing with hyperspectral imaging of oxyhemoglobin and deoxyhemoglobin", *Diabetes Care*, vol. 32(11), pp. 2056–2061, 2009.

[PB95] D. A. Perednia and N. A. Brown, "Teledermatology: One application of telemedicine", *Bulletin of the Medical Library Association*, vol. 83(1), pp. 42–47, 1995.

[PHP11] A. Plüddemann, C. Heneghan, and C. P. Price, "Dermoscopy for the diagnosis of melanoma: Primary care diagnostic technology update", *British Journal of General Practice*, vol. 61(587), pp. 416–417, 2011.

[Rin10] F. Ring, "Thermal imaging today and its relevance to diabetes", *Journal of Diabetes Science and Technology*, vol. 4(4), 2010.

[Roh92] K. Rohr, "Modeling and identification of characteristic intensity variations", *Image and Vision Computing*, vol. 10, pp. 66–76, 1992.

[RPD10] E. Rosten, R. Porter, and T. Drummond, "Faster and better: A machine learning approach to corner detection", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32(1), pp. 105–119, 2010.

[Rub+11] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "Orb: An efficient alternative to sift or surf", in *IEEE international conference on Computer Vision*, 2011.

[Sch+08] G. Schaefer, R. Tait, K. Howell, A. Hopgood, P. Woo, and J. Harper, "Automated overlay of infrared and visual medical images", *User Centered Design for Medical Visualization*, vol. 8, pp. 174–183, 2008.

[SF04] A. Sassaroli and S. Fantini, "Comment on the modified beer-lambert law for scattering media", *Physics in Medicine and Biology*, vol. 49(14), N255–7, 2004.

[SHK98] H. S. Sawhney, S. Hsu, and R. Kumar, "Robust video mosaicing through topology inference and local to global alignment", in *Proceedings of European Conference on Computer Vision*, 1998.

[SMA76] M. C. Svedlow, D. McGillem, and P. E. Anuta, "Experimental examination of similarity measures and preprocessing methods used for image registration", in *Proceedings of Symposium on Machine Processing of Remotely Sensed Data*, 1976, pp. 9–17.

[Ste99] C. V. Stewart, "Robust parameter estimation in computer vision", *Society for Industrial and Applied Mathematics Reviews*, vol. 41(3), pp. 513–537, 1999.

[Sze06] R. Szeliski, "Image alignment and stitching: A tutorial", *Foundations and Trends in Computer Graphics and Vision*, vol. 2(1), pp. 1–104, 2006.

[Tan+10] E. Tan, A Yung, M. Jameson, A. Oakley, and M. Rademaker, "Successful triage of patients referred to a skin lesion clinic using teledermoscopy (image it trial)", *British Journal of Dermatology*, vol. 162(4), pp. 803–811, 2010.

[TK92] C. Tomasi and T. Kanade, "Shape and motion from image streams under orthography: A factorization method", *International Journal of Computer Vision*, vol. 9, pp. 137–154, 1992.

[Ves+08] M. E. Vestergaard, P. Macaskill, P. E. H. PE, and S. W. Menzies, "Dermoscopy compared with naked eye examination for the diagnosis of primary melanoma: A meta-analysis of studies performed in a clinical setting", *British Journal of Dermatology*, vol. 159(3), pp. 669–676, 2008.

[VR77] G. J. Vanderbrug and A. Rosenfeld, "Two-stage template matching", *IEEE Transactions on Computers*, vol. 26(4), pp. 384–393, 1977.

[VRD13] T. Vercauteren, B. Rosa, and J. Dauguet, "A viterbi approach to topology inference for large scale endomicroscopy video mosaicing", in *Proceedings of Medical Image Computing and Computer-Assisted Intervention*, vol. 16(1), Nagoya, Japan, 2013, pp. 404–411.

[War+11] E. M. Warshaw, Y. J. Hillman, N. L. Greer, E. M. Hagel, R. MacDonald, I. R. Rutks, and T. J. Wilt, "Teledermatology for diagnosis and management of skin conditions: A systematic review", *Journal of the American Academy of Dermatology*, vol. 64(4), pp. 759–772, 2011.

[Wei+12a] T. Weibel, C. Daul, D. Wolf, and R. Rösch, "Contrast-enhancing seam detection and blending using graph cuts", in *Proceedings of 21st International Conference on Pattern Recognition*, Tsukuba-Japan, 2012, pp. 2732–2735.

[Wei+12b] T. Weibel, C. Daul, D. Wolf, R. Rösch, and F. Guillemin, "Graph based construction of textured large field of view mosaics for bladder cancer diagnosis", *Pattern Recognition*, vol. 45(12), pp. 4138–4150, 2012.

[WGN15] E. M. Warshaw, A. A. Gravely, and D. B. Nelson, "Reliability of store and forward teledermatology for skin neoplasms", *Journal of the American Academy of Dermatology*, vol. 72(3), pp. 426–435, 2015.

[Wu+15] X. Wu, S. A. Oliveria, S. Yagerman, L. Chen, J. DeFazio, R. Braun, and A. A. Marghoob, "Feasibility and efficacy of patient-initiated mobile teledermoscopy for short-term monitoring of clinically ctypical nevi", *JAMA Dermatology*, vol. 151(5), pp. 489–496, 2015.

[YNP10] D. Yudovsky, A. Nouvong, and L. Pilon., "Hyperspectral imaging in diabetic foot wound care", *Journal of Diabetes Science and Technology*, vol. 4(5), pp. 1099–1113, 2010.

Résumé

La télédermatologie présente plusieurs avantages par rapport aux consultations traditionnelles en cabinet avec un dermatologue. Elle est particulièrement utile pour faciliter l'accès aux soins dermatologiques pour les patients ayant des problèmes de mobilité ou habitant loin des secteurs géographiques médicalisés. Un schéma de mosaïquage automatique d'images dédié à la création des panoramas étendus des vidéo-séquences de peau est proposé pour surmonter les limitations posées par le champ de vue réduit des images stationnaires acquises par les dispositifs actuellement utilisés. Les vidéo-séquences utilisées à cet effet sont acquises en utilisant un dispositif spécialement conçu pour un rendu colorimétrique contrôlé de la surface de la peau. Après une étude des diverses méthodes de recalage d'images existantes, une approche optimale est proposée, avec un certain compromis entre la précision de recalage et le temps de calcul, pour la superposition des parties communes des images cutanées. En outre, une approche pour affiner la correspondance initiale des points caractéristiques extraits est présentée. L'étude présentée porte principalement sur la construction cohérente d'une mosaïque dans son ensemble. Pour atteindre cet objectif, un schéma de mosaïque capable de générer des panoramas cohérents à partir de vidéo-séquences longues est présenté. Ce schéma estime dynamiquement la topologie de la trajectoire des images dans le plan de mosaïquage. Cela permet de placer les images sur le plan panoramique avec un nombre réduit d'images sur le chemin suivi pour atteindre une image donnée à partir d'une image de référence, ce qui réduit non seulement l'accumulation des erreurs, mais permet également d'éviter les interruptions dans le mosaïquage en excluant les paires d'images dont le recalage ne serait pas réussi. L'approche proposée offre une robustesse vis-à-vis des recalages échoués en trouvant des trajets alternatifs. En outre, un mode d'ajustement global pour améliorer davantage la cohérence de la mosaïque est présenté.

Mots-clés: traitement d'image, recalage d'images, ajustement de mosaïque, image panoramique, télédermatologie, correspondance des points-clés

Abstract

Teledermatology offers several advantages in comparison to the traditional in-place consultations with a dermatologist. It is particularly useful for easing the access to the dermatological care for patients with mobility or travel constraints. A dedicated mosaicing scheme for creating extended panoramas of skin video sequences is proposed to surmount the limitations posed by the small field of view of stationary images acquired by currently used devices. The video sequences used for this purpose are acquired using a specially designed device for a colorimetrically correct rendering of the skin surface. After a study of various image registration approaches, an approach optimally suited to skin image registration with some compromise between registration accuracy and computation time is selected. In addition, an approach for refining the initially detected key-point correspondence is presented. Central focus of this study is on the overall coherent construction of the mosaic. To achieve this objective, a mosaicing scheme capable of generating coherent panoramas from long video sequences is presented. This scheme dynamically estimates the topology of the image trajectory in the panoramic plane to mosaic the images by reducing the number of images over the path used for reaching a given image from a reference image in order to place it on the panoramic plane. A small number of images reduces the accumulated errors, thus improving the visual coherency of the overall mosaic. Besides, the proposed approach offers robustness against failed registrations, which would interrupt the mosaicing process in the absence of the alternative paths. Moreover, a global adjustment scheme for further improving the coherency of the mosaic is presented.

Keywords: image processing, image registration, mosaic adjustment, panoramic image, teledermatology, key-point correspondence