Computerized methods for classification and severity assessment of skin pigmentation disorders

Liang, Yunfeng

2018

Liang, Y. (2018). Computerized methods for classification and severity assessment of skin pigmentation disorders. Doctoral thesis, Nanyang Technological University, Singapore.

<http://hdl.handle.net/10356/75905>

<https://doi.org/10.32657/10356/75905>

**COMPUTERIZED METHODS FOR
CLASSIFICATION AND SEVERITY
ASSESSMENT OF SKIN
PIGMENTATION DISORDERS**

LIANG YUNFENG

Interdisciplinary Graduate School

A thesis submitted to the Nanyang Technological University
in partial fulfillment of the requirement for the degree of
Doctor of Philosophy

2018

Acknowledgement

Firstly, I would like to express my sincere gratitude to my supervisor Prof. Lin Zhiping for the continuous support of my Ph.D. study and related research, for his patience, motivation, and guidance throughout my research study.

My deep appreciation also go to my co-supervisors, Prof. Lin Feng and Prof. Thng Tien Guan Steven, and my mentor Prof. Ser Wee, for their great support and their kind care. The insightful comments and encouragement from them helped me widen my research from various perspectives.

My sincere thanks also go to Prof. Sun Lei from Beijing Institute of Technology, who provided me enormous help in developing the algorithms and writing articles. Without his support, it would not be possible to conduct this research smoothly.

I would like to thank the clinical collaborators from National Skin Center of Singapore for their support in the data collection, paper writing, and many other aspects of my research study. I would also like to thank my friends and fellow classmates for their help.

Last but not least, I would like to thank my beloved parents and my wife for the selfless support they have given me throughout my study.

Contents

Acknowledgement	i
Abstract	iv
List of Figures	vi
List of Tables	ix
Abbreviations	x
1 Introduction	1
1.1 Background and Motivation	2
1.2 Scope and Contributions	6
1.3 Thesis Overview	10
2 Literature Review	11
2.1 Computerized Image Classification Methods	12
2.1.1 Local Approaches Based on BoF and SIFT	13
2.1.2 Local Approaches Based on SPM and LLC	15
2.1.3 Sparse and Dense Representation	17
2.1.4 Probabilistic Linear Discriminant Analysis	19
2.2 Melasma and Its Clinical Severity Assessment Methods	24
2.2.1 Melasma	24
2.2.2 mMASI	25
2.3 Thresholding Based Image Segmentation Methods	26
2.3.1 Otsu’s Method	28
2.3.2 Adaptive Otsu Thresholding	28
2.3.3 Sauvola’s Method	29
3 Classification of Skin Pigmentation Disorders	31
3.1 Introduction	32
3.2 Improved PLDA for Large Within-Class Variance Problems	34
3.2.1 PLDA as Individual Matching	35
3.2.2 Voting Based PLDA	40
3.3 Experimental Results	44
3.3.1 Experimental Setup	45

3.3.2	Input Feature Selection	46
3.3.2.1	Color Feature Selection	46
3.3.2.2	Non-color Feature Selection	48
3.3.2.3	Combination of the Input Features	52
3.3.3	Results and Discussions	54
3.4	Summary	58
4	Melasma Severity Assessment via Image Analysis	60
4.1	Introduction	61
4.2	Proposed Melasma Severity Assessment System	62
4.2.1	Image Enhancement	63
4.2.2	Area of Interest Detection	66
4.2.3	Hybrid Thresholding Based Image Segmentation	70
4.2.4	Score Estimation	73
4.3	Experimental Results	77
4.3.1	Evaluation Methods	77
4.3.2	Results and Discussions	79
4.4	Summary	83
5	Optimal Hybrid Thresholding for Melasma Severity Assessment	84
5.1	Introduction	85
5.2	Optimal Hybrid Thresholding Based on Euclidean Distance	88
5.3	Optimal Hybrid Thresholding Based on Generalized Distance Measure	93
5.4	Experiments on Melasma Images	98
5.4.1	Evaluation of OHYTA Based on Euclidean Distance	98
5.4.2	Evaluation of OHYTA Based on Generalized Distance	101
5.5	Summary	103
6	Conclusion	105
6.1	Thesis Summary	106
6.2	Future Work	109
	Author's Publications	112
	Bibliography	114

Abstract

The objective of this thesis is to develop computerized algorithms that can easily and reliably diagnose the non-tumorous skin pigmentation disorders and assess the severity of such conditions based on the facial image photographs of the patients. The topic is critical in the clinical management as the correct diagnosis can avoid improper treatments which may result in adverse effects while the severity assessment determines if the treatment method used is effective or not.

For solving the diagnosis problem based on the image classification techniques, we propose a voting based probabilistic linear discriminant analysis (V-PLDA) method which addresses the large within-class variance problem in the non-tumorous skin pigmentation image data. The proposed V-PLDA method has been validated with a real-world dataset that contains totally 150 samples of five non-tumorous skin pigmentation disorders including freckles, lentigines, Horis nevus, melasma and nevus of Ota. The results show that the proposed V-PLDA method obtains significantly higher classification accuracy than the original PLDA method as well as several state-of-the-art image classification methods.

For the pigmentation severity assessment problem, we focus on a commonly observed non-tumorous skin pigmentation disorder, melasma, by estimating the clinical modified melasma area and severity index (mMASI) using the computerized methods. We first develop a complete system for the assessment, which includes several steps of the image processing, for example, illumination normalization, image segmentation, and darkness estimation, to obtain the estimated severity scores. An image set recorded from 29 patients at two time points is used for validating the proposed algorithm. The results show that the estimated scores using the proposed algorithm have a good correlation with the scores given by the dermatologists.

With the severity assessment algorithm developed, the performance of the algorithm is improved by focusing on the image segmentation process. We propose an optimal hybrid thresholding method by determining the parameter that trades off between the local and the global information according to the image content. The optimal trade-off parameter is obtained by minimizing the distance between the input image and the threshold surface. The same melasma image dataset is used in the experiment and the results show that the proposed optimal hybrid

thresholding algorithm (OHYTA) performs better than several state-of-the-art thresholding based image segmentation methods.

In conclusion, the thesis presents computerized methods for solving the real-world medical problems, namely, the classification of five non-tumorous skin pigmentation disorders and the severity assessment of melasma. Validated with the image data of the patients, it is shown that the proposed methods can obtain satisfactory results that correlate well with those given by the dermatologists. It is also shown that the proposed methods can have better performance than the state-of-the-art algorithms in dealing with the non-tumorous skin pigmentation image set we have at hand. To the best knowledge of the author, this work is the first one that applies computerized methods to solve the non-tumorous skin pigmentation classification and melasma severity assessment problems. The study can serve as a benchmark for the future research works on this topic.

List of Figures

1.1	The process flow of the treatment.	4
2.1	Data distributions of the PLDA model when the within-class variance is small.	20
2.2	Data distributions of the PLDA model when the within-class variance is large. The categorical overlapping is highlighted.	22
2.3	Tied-PLDA model. The categorical overlapping between subclasses is highlighted. The subclass 1 is labeled.	23
2.4	Example images of melasma patients.	24
2.5	The face regions divided in mMASI score system.	25
3.1	The process of clinical diagnosis process in differentiating freckles, lentigines, Hori's nevus, melasma and nevus of Ota.	32
3.2	Four sample images of each skin pigmentation disorder to illustrate the large within-class variance. (a) Freckles. (b) Lentigines. (c) Hori's nevus. (d) Melasma. (e) Nevus of Ota.	33
3.3	Average classification accuracies of the PLDA method and the individual matching method at different levels of the within-class variance. The two vertical lines separate the plot into three parts. . . .	39
3.4	The voting sample selection process. Only the ones having high joint probabilities with the testing sample $\mathbf{x}_t$ are selected (circled ones in the figure).	41
3.5	Average classification accuracies of PLDA and V-PLDA at different subclass sizes.	43
3.6	Average classification accuracy of V-PLDA model using different color spaces.	47
3.7	Average classification accuracy for different selections of each parameter.	51
3.8	Illustration of the real part of the selected Gabor filter bank.	53
3.9	The structure of the fused PLDA model.	54
3.10	(a) The testing sample which is from the class of freckles. (b)-(h) The selected samples in the voting process, where (b)-(d) are the samples from the class of Hori's nevus and (e)-(h) are from the class of freckles.	55

4.1	A set of images taken from a patient with the VISIA Complexion Analysis system: (a) Left view, (b) Frontal view and (c) Right view. Eyes are blocked to protect patient's privacy.	63
4.2	The flow chart of homomorphic filtering. 'DFT' and 'DFT ⁻¹ ' represent the discrete Fourier transform and inverse discrete Fourier transform, respectively. 'H' is the Fourier transform of the high pass filter h	65
4.3	Grayscale images. (a) Lightness image before homomorphic filtering. (b) Lightness image after homomorphic filtering. (c) Input image combine both lightness and color information.	65
4.4	Example binary images B obtained from applying the face region detection with face region represented by white color and non-face region represented by black color. (a) Left view. (b) Front view. (c) Right view.	68
4.5	The process of cutting frontal view image to extract the forehead and chin regions.	69
4.6	The process of cutting side view image to extract the malar region.	70
4.7	The original images (upper images) contain different face regions, forehead (left), left malar (middle) and chin (right), and the corresponding results after AoI detection process (lower images).	70
4.8	One example to show the image thresholding process. (a) Input image with a patch of normal region selected. (b) The output binary image $B_{pigment}$ after the thresholding. (c) The detected boundaries plotted on the input image.	73
4.9	Six different types of quads and the corresponding assigned area. The white color and black color represent on and off state of the pixel, respectively.	75
4.10	Scatter plots showing the relationship between clinical mMASI and estimated eMASI scores. (a) Baseline. (b) Post-treatment. (c) Changes between baseline and post-treatment.	81
4.11	User interface of the melasma assessment software based on the proposed assessment method.	82
5.1	Thresholding segmentations on synthetic images.	86
5.2	Weighted hybrid thresholding segmentation with respect to λ . (a) Original image, (b) ground truth, (c) $\lambda = 0$, (d) $\lambda = 0.2$, (e) $\lambda = 0.4$, (f) $\lambda = 0.6$, (g) $\lambda = 0.8$, (h) $\lambda = 1$	90
5.3	Statistical evaluation of different methods using synthetic data. (a) Precision. (b) Recall. (c) F-measure.	93
5.4	Contour plots of generalized distance measure $E(T, I, \alpha)$ with different α values.	94
5.5	Statistical evaluation of the α value selection using synthetic data. (a) Precision, (b) Recall, (c) F-measure.	96
5.6	Statistical evaluation of the α value selection using synthetic data with different levels of Gaussian noise.	98

5.7	Melasma segmentation results using several thresholding methods. (a) Input image with normal skin region selected. (b)-(e) Segmentation results of mSauvola, mAdOtsu, EHYTA and OHYTA-Eu, respectively. The ellipses on the image highlight the over- or under-segmented regions.	99
5.8	Scatter plots showing the relationship between clinical mMASI scores and estimated oMASI scores using the OHYTA-Eu method. (a) Baseline. (b) Post-treatment. (c) Changes between baseline and post-treatment.	102

List of Tables

2.1	The area score defined in mMASI score system.	26
2.2	The darkness score defined in mMASI score system.	26
3.1	Setup for the subclasses	42
3.2	Results of paired-samples t-test comparing PLDA and V-PLDA with different m values	44
3.3	Parameter selection for Gabor filter bank	50
3.4	Results of paired-samples t-test comparing PLDA and V-PLDA with different input features	55
3.5	Results of the different methods	57
3.6	Confusion matrix for the V-PLDA model (%) with combined HSV and Gabor features as its input	57
4.1	Comparison of clinician mMASI scores and eMASI scores from pro- posed method	80
5.1	The α values that obtain the best performances in terms of mean F-measure at different noise levels.	98
5.2	Comparison of thresholding methods for melasma severity assessment	100
5.3	Results of proposed method with different α values (with the best values highlighted)	103

Abbreviations

AdOtsu	Adaptive Otsu
ANOVA	ANalysis Of VAriance
AoI	Area of Interest
BCC	Basal Cell Carcinoma
HOG	Histogram of Oriented Gradients
IKSVM	Intersection Kernel Support Vector Machine
LBP	Local Binary Patterns
LLC	Locality-constrained Linear Coding
MASI	Melasma Area and Severity Index
OHYTA	Optimal HYbrid Thresholding Algorithm
PLDA	Probabilistic Linear Discriminant Analysis
SC	Sparse Coding
SDR	Sparse and Dense Representation
SPM	Spatial Pyramid Matching
SRC	Sparse Representation-based Classifier
SIFT	Scale Invariant Feature Transform
SNR	Signal to Noise Ratio
VETI	Vitiligo Extent Tensity Index
VQ	Vector Quantization
V-PLDA	Voting based Probabilistic Linear Discriminant Analysis

Chapter 1

Introduction

In this chapter, the background and motivations of this thesis are first described in Section 1.1. In Section 1.2, the scope and contributions of this thesis, where the two problems, non-tumorous skin pigmentation classification and melasma severity assessment are described. Finally, an overview of the thesis is given in Section 1.3.

1.1 Background and Motivation

Skin pigmentation disorders caused by abnormal production and deposition of melanin by special skin cells are generally called melanocytes [1, 2]. With the over or under production of melanin, the skin color would turn darker or lighter, respectively. By affecting patches of skin or the skin of the entire body, skin pigmentation disorders can lead to obvious appearance change and consequently bring huge negative emotional impact on the patients [3–5]. Besides the impact on the appearance, some of the skin pigmentation disorders can be associated with internal diseases [6] or skin cancer [7]. For the purpose of improving the life quality of the patients or early detection of certain diseases [8], it is crucial to have a proper diagnosis and treatment of the skin pigmentation disorders.

To help the dermatologists in the diagnosis of skin pigmentation disorders, different digital image processing algorithms have been developed in the last few decades [9, 10], which have shown good performance in the segmentation and classification of the pigmentation images. However, most of the studies focus on the detection and classification of skin tumors like melanoma [11, 12] and pigmented basal cell carcinoma (BCC) [13, 14]. While such studies are very important as skin tumors are likely to be cancers and threaten the health of the patients [15], more patients are bothered by the non-tumorous skin pigmentation disorders. This is because the non-tumorous conditions like melasma, lentigines, Hori's Nevus, etc., mainly appear on the faces and are regarded as unsightly, causing much emotional stress to sufferers [3].

In contrast to the high demand for a proper treatment of the non-tumorous skin pigmentation disorders, there are only a few studies on applying computerized methods to solve the non-tumorous skin pigmentation related problems [16, 17].

It is important to note that the diagnosis rules for non-tumorous skin pigmentation disorders differ greatly from those of their tumorous counterparts due to the appearance differences in their associated skin images. Tumorous conditions like melanoma, and BCC usually present as a distinct single lesion and a detailed image analysis of the lesion is carried out to look for specifics like size, shape, color, and structure. These prominent features are summarized to form some precise rules, for example, the ABCD rule [11, 18], for skin tumor image classification. In contrast, the features of the non-tumorous skin pigmentation disorders usually affect the whole face and are bilateral. The pigmentation appearance varies a lot in terms of size, color and shape even for the ones in the same class, which makes within-class variance very large. Due to the above-mentioned differences, methods specifically designed for dealing with the tumorous pigmentation disorders may not be suitable for solving the problem of non-tumorous pigmentation disorders. Therefore, to fill the gap in the non-tumorous skin pigmentation image processing, the author carried out the research study presented in this thesis.

During the treatment of the non-tumorous skin pigmentation disorders, there are occasions when the input of the dermatologists is needed. One occasion is in the beginning of the diagnosis process, different types of non-tumorous skin pigmentation disorders need to be classified accurately because different treatment methods need to be adopted for each type of skin pigmentation disorder [19, 20]. Another occasion involving the input of the dermatologists is during the treatment process. Because for certain types of skin pigmentation disorders, such as melasma and vitiligo, the dermatologists need to check the effect of the treatment by assessing the disease severity using predefined scoring systems [21–24]. The process flow of the treatment containing the two occasions is illustrated in Figure 1.1.

It can be observed in Figure 1.1 that the disease is first diagnosed in the classification process. Based on the result of the classification, a proper treatment method is selected. Then, the severity of the disease is assessed at two time points, before

FIGURE 1.1: The process flow of the treatment.

the treatment (baseline) and after the treatment (post-treatment). The severity scores at the two time points are compared to determine if the treatment method is taking positive effect or not. If the condition is improved, the current treatment method would be kept in the next. Otherwise, a new treatment method may be used. In the rest of this thesis, the needs of computerized methods in the classification and severity assessment process are discussed separately: The classification problem is addressed in Chapter 3 after the literature review part in Chapter 2; The severity assessment problem is addressed in Chapter 4 and 5. Chapter 4 builds up a complete system, which is the first one that solves the melasma severity assessment problem. Based on the study in Chapter 4, Chapter 5 improves the melasma severity assessment system by focusing on improving the segmentation method of the system.

In the classification process, many patients seek help from beauty salons and aestheticians instead of dermatologists. This is because non-tumorous skin pigmentation disorders are not fatal and the patients treat them as beauty problems. However, without the professional knowledge of skin pigmentation treatments [25, 26], non-professionals may make a wrong diagnosis [27, 28] and recommend options like lasers and chemical peels which could result in adverse effects on the patients [29]. Therefore, motivated by the need of the patients to have a tool that can easily and reliably diagnose non-tumorous skin pigmentation disorders, the computerized classification methods are developed based on the professional rules provided by dermatologists.

In the severity assessment process, the severity scores are estimated based on the visual observation by the dermatologists. While the information provided by the dermatologists is reliable, there are disadvantages to do the severity assessment process manually [22, 30]. Firstly, the severity scores of the non-tumorous skin pigmentation are based on a visual assessment on the appearance of the pigmentation patches. It is a subjective process and makes the assessment results depend on the experience of the evaluator. Secondly, it is hard for human evaluators to have an accurate assessment of the disease because the scores are grouped in bands. The band form of the scores cannot reflect the minor difference within the same band and enlarge the effects of small changes across bands. Thirdly, the manual assessment can only be done by dermatologists, which is neither time-efficient nor cost-efficient for the patients because the human resources are limited. In contrast to the clinical assessment methods, the computerized methods can well circumvent the above disadvantages by giving consistent, accurate, and fast assessment results. This motivates us to develop the computerized methods to do the severity assessment for the non-tumorous skin pigmentation disorders.

1.2 Scope and Contributions

This thesis focuses on the classification and severity assessment of non-tumorous skin pigmentation disorders. Because the pigmentation patches on the face region are the most appearance and emotion affecting types for the patients, the study only focuses on the processing of skin pigmentation images taken from the face region. We also only focus on processing the image information in the study in the current stage.

In this image classification problem, five common skin pigmentation disorders, namely, freckles, lentigines, Horis nevus, melasma and nevus of Ota are studied. One challenge in classifying the five selected diseases is that the within-class variances of them are very large. The large within-class variance reduces the performance of the classification algorithms. Therefore, to stress this large within-class variance problem, we propose a voting based probabilistic linear discriminant analysis (V-PLDA) based on the probabilistic linear discriminant analysis (PLDA) method [31]. Different from the existing tied-PLDA method [32, 33] that also tackles the large within-class variance problem, the proposed V-PLDA does not require any prior knowledge of the input images. The proposed V-PLDA method is applied to a real-world image set of the five skin pigmentation disorders. The results show that the average classification accuracy of the skin pigmentation image set by the proposed V-PLDA method is significantly higher than all other methods in comparison.

The contributions made in this part of the thesis can be summarized as follows. Firstly, the performance of the PLDA method under the situation of large within-class variance is studied and it is shown that PLDA can be approximated by an individual matching process in such situation. Secondly, a V-PLDA method that tackles the large within-class variance problem is proposed, and the factors that

affect the performance of V-PLDA are studied using simulations. Thirdly, proper features are selected from tests to solve the non-tumorous skin pigmentation classification problem using the proposed V-PLDA method. The work is summarized in a journal paper in [34] and the details are presented in Chapter 3.

As mentioned earlier, with the treatment method determined based on the classification result, the next important procedure is to check the disease severity by observing the appearance of the pigmentation patch before and after the treatment. This is for the purpose of deciding if the treatment method adopted is taking a positive effect or not, so that the strategies for the next step of treatment can be determined. In the study of developing computerized severity assessment method, we only focus on one widely-spread non-tumorous skin pigmentation disorder, melasma [35–38], in the current stage. The proposed method is developed with reference to a well accepted clinical melasma severity score system named modified Melasma Area and Severity Index (mMASI) [21, 39].

There are two stages of the thesis for solving the melasma severity assessment problem. In the first stage, the focus is on developing a scoring system from scratch that imitates the mMASI system using computerized methods. It involves several image processing steps, including preprocessing of the input images so that less noise is contained, segmentation of the pigmentation patches in the preprocessed images, post-processing of the segmentation boundaries and calculation of the severity scores. The results show that the estimated mMASI scores obtained from the proposed system correlate well with the actual mMASI scores given by the dermatologists.

The contributions of the first stage melasma severity assessment are mainly as follows. Firstly, a series of methods for processing the standardized images taken by the VISIA system has been proposed to normalize the illumination, cut the image, and find the skin region. Secondly, we propose an equal hybrid thresholding

method for segmenting the melasma pigmentation region. Thirdly, we propose to utilize the bit quads method to estimate the area score and use a local average color difference method to estimate the darkness score. Overall, based on the above image processing techniques, a complete melasma severity assessment system is developed, which is the first work that deals with such problems. The work is published in a journal paper in [30] and the details are presented in Chapter 4.

After the computerized melasma scoring system is developed, the second stage of the development is improving the system. The focus of the study at this stage is on the image segmentation process in the system. This is because the segmentation is the fundamental and most critical process of the system that determines the accuracy of the estimated scores in the subsequent part [40]. In the developed scoring system, the thresholding method is selected for image segmentation because it is simple and effective in dealing with skin pigmentation images. With the finding that a hybrid of local and global information in the threshold determination process can obtain more accurate segmentation results, the equal hybrid thresholding algorithm (EHYTA) that combines both local and global thresholding methods is used in the severity assessment system in the first stage. The ‘equal’ in the method means that both the local and global information contribute the same amount (with the same trade-off parameter) to the final threshold.

Although the scoring system using EHYTA has obtained satisfactory results, it is found that using the same trade-off parameter for both the local and global information is heuristic and may not be the optimal solution. Therefore, by defining the optimal properties of the threshold surface like the one in [41], the trade-off parameter between the local and global information is optimized by minimizing the Euclidean distance between the hybrid threshold surface and the input image. This method is named as optimal hybrid thresholding algorithm (OHYTA) and is applied in the melasma scoring system. The results show that the system with OHYTA performs better than that with EHYTA and other thresholding based

methods for assessing the severities of the melasma images. The optimal hybrid thresholding algorithm that uses the Euclidean distance in the processing is proposed in the journal paper [42]. To improve the performance, a distance measure that generalizes the Euclidean distance is used in OHYTA and the results are presented in the conference paper [43].

The contributions of this stage can be summarized as follows. Firstly, an optimal hybrid segmentation threshold determination criterion is proposed by means of optimization using a distance based deviation between a given image and its segmentation outcome. Secondly, based on the obtained optimal hybrid threshold, an optimal hybrid thresholding segmentation algorithm (OHYTA) is developed. Thirdly, the developed segmentation method is used for melasma severity assessment and is statistically evaluated based on real-world images. The details of this work are presented in Chapter 5.

In summary, the work in this thesis provides different methods that can replace or assist some of the clinical diagnosis process in dealing with non-tumorous pigmentation disorders. It can bring enormous help to both the patients and the clinical organizations. With the help of the proposed method, the patients can get accurate assessment of the disease efficiently. The computerized system also reduces much time and manpower costs of the clinical organizations that are originally spent on the manual assessment process. To the best knowledge of the author, this is the first study that focuses on the diagnosis of non-tumorous skin pigmentation disorders using image analysis methods. It provides a benchmark for the future studies in the related area.

1.3 Thesis Overview

The rest of this thesis is organized as follows:

Chapter 2 reviews the related computerized methods which are used for the image classification and segmentation. The melasma disease and the corresponding clinical severity assessment system, mMASI, are also described.

Chapter 3 presents an improved PLDA method, namely, the V-PLDA method, to classify different non-tumorous skin pigmentation disorders based on the image data. The proposed V-PLDA method tackles the large within-class variance problem in the classification process. It is evaluated with real-world skin pigmentation image data and compared with several state-of-the-art image classification algorithms.

Chapter 4 introduces a melasma computerized severity assessment system, which includes several processing steps like the illumination normalization, image segmentation and score estimation. The system is evaluated on real-world melasma image data with proper evaluation indices used.

Chapter 5 investigates further on the image segmentation method used in Chapter 4. With the properties of the hybrid thresholding method discussed, the optimal hybrid thresholding algorithm is proposed and the performance of the method using different distance measures is discussed. The proposed method is compared with different thresholding methods with the real image data used in Chapter 4.

Chapter 6 concludes the thesis and provides some possible directions of future work.

Chapter 2

Literature Review

In this chapter, the existing diagnosis methods for the non-tumorous skin pigmentation disorders and the related computerized methods for skin pigmentation image classification and segmentation are briefly reviewed. In Section 2.1, both the local and global based image classification approaches are introduced and several representatives of each type are briefly reviewed. The melasma disease and its severity assessment system, namely the mMASI system, are reviewed in Section 2.2 and some thresholding based image segmentation methods are reviewed in Section 2.3.

2.1 Computerized Image Classification

Methods

Over the past few decades, many digital image classification methods have been developed that can be used for skin image classification. They can be broadly classified into local and global approaches [44, 45]. In the local approach, images are described by invariant features of selected local patches. These local patches are distinguished regions marked by region detectors. The features are then extracted from those patches using different feature descriptors, such as, scale invariant feature transform (SIFT) [46–48], local binary patterns (LBP) [49], histogram of oriented gradients (HOG) [50] and so on. Those extracted features are then organized by algorithms like bag-of-features (BoF) [51] or spatial pyramid matching (SPM) [52] before they are fed into different classifiers [53–55]. The local approaches can adapt well to the large within-class variance problem that exist in the image dataset because they can extract the key features of the object using the local patches. Additionally, the nonlinear coding process or the nonlinear classifiers used in the local approaches make the methods able to handle the overlapping between different classes caused by the large within-class variance.

An alternative to the local approach is the global approach, which uses the whole image as an input so that the local region detection process is not required [56–58]. To transform the high dimensional input image into class labels, the global methods project the data onto a suitable lower dimensional space. For example, the Fisher’s linear discriminant analysis (LDA) based methods [56, 59] project the input images onto a space with minimized within-class variance. Another well-known linear discriminant analysis model is the LDA incorporated with probability theory to form the probabilistic linear discriminant analysis (PLDA) model [31]. The PLDA is superior to the LDA because it models the categorical variances in

images more accurately. One problem with some of these global methods, such as LDA and PLDA, is that they can only handle problems with small or moderate within-class variance because the image patterns need to be closely aligned [45]. To tackle the large within-class variance problem in the data, a global method based on the sparse coding (SC) [60] named sparse and dense representation (SDR) [58] is proposed. The SDR method takes both the between- and within-class variances into consideration and therefore can have a more accurate model of the data. One problem with the SDR method is that it is distance based and not robust enough to handle complicated situations. Another global approach that can handle large within-class variance is the tied-PLDA method [32], [33]. In tied-PLDA, the categorical images are further divided into subclasses using a tied analysis. This allows the large within-class variance of the whole class to be replaced by several smaller within-class variances with respect to the divided subclasses.

In this subsection, some representative global and local based image classification methods that can deal with the large within-class variance problem in our non-tumorous skin pigmentation image set are briefly reviewed.

2.1.1 Local Approaches Based on BoF and SIFT

Bag of features (BoF) [51] method is originated from the bag of words (BoW) method [61] that used in the text categorization problems. The BoW method assigns a textual document to one or more categories based on its content. It breaks the text into individual words and counts the frequency of each word used. The frequency information is used as the features for content classification at last. Similar strategies to BoW has been applied to the image classification problems by treating an image as a combination of local patches, which forms the BoF method. In this way, each local image patch can be seen as a visual word and an image

is formed by a bag of those visual words. The frequency of each visual codeword that appears in an image can be used as the features for image classification.

When using the BoF method in the image classification algorithms, the selected local image patches need to be represented by certain region properties, for example, the geometric properties (distance) and orientation information (gradients). These properties are found by using different region descriptors and the scale invariant feature transform (SIFT) [46, 48] descriptor is one of the most popular ones used. The descriptor is designed to be invariant to rotation, the main orientation of the region is obtained by a 36 bin orientation histogram of gradient orientations within a Gaussian weighted circular window. The gradient magnitude (M_{SIFT}) and the local orientation (ϕ_{SIFT}) at a pixel $I(x, y)$ in the image are calculated as follows:

$$M_{SIFT} = \sqrt{(I(x+1, y) - I(x-1, y))^2 + (I(x, y+1) - I(x, y-1))^2} \quad (2.1)$$

$$\phi_{SIFT} = \tan^{-1} \frac{I(x, y+1) - I(x, y-1)}{I(x+1, y) - I(x-1, y)} \quad (2.2)$$

To generate a unique description of each keypoint in the image, a 16×16 window around the point is used. The window is further divided into sixteen 4×4 windows to get the magnitudes and orientations. Then, the orientations obtained are put into an 8 bin histogram of orientations. To stress the influence of the region in the center of the window (around the keypoint), the magnitudes are weighted by a Gaussian window. After the processing, each window would be described by a vector with a length of 128. The SIFT method has achieved satisfactory results when dealing with many image classification problems and further developments of the SIFT descriptor have been made after the original work by Lowe, for example, the dense SIFT [62] and color SIFT [63]. The extracted feature from the BoF and SIFT would be classified using a non-linear classifier. Among different non-linear classifiers, the support vector machine with the histogram intersection kernel

(IKSVM) can achieve satisfactory performance with limited computation cost [64], which is widely used in image classification problems.

2.1.2 Local Approaches Based on SPM and LLC

The BoF method reviewed in the last subsection is successful because it is especially robust against spatial translation of features by disregarding the spatial layout of the features. However, this also brings problems when such information of the object, for example, the shape and location information, are important for the classification. The performance of the BoF method would be not satisfactory under such condition. Therefore, further development of the BoF method has been carried out and the most successful ones are the spatial pyramid matching (SPM) [52] and its improvement, the locality-constrained linear coding (LLC) [65] method.

In the SPM method, the image is partitioned into subregions at different scales. Local features such as SIFT are extracted from these subregions and then combined together to describe the image. The features at different levels of resolutions are weighted differently in the combination: the finer of the scale, the larger of the weight. This is because the larger scale of the image would contain more dissimilar features, which should be penalized. The SPM method is better than the BoF method as it includes the spatial information in the description. However, one problem with the SPM method is that it uses the hard vector quantization (VQ) in the coding process, which can be expressed using the following constrained least square fitting problem:

$$\arg \min_{\mathbf{c}} \sum_{i=1}^N \|\mathbf{x}_i - \mathbf{B}\mathbf{c}_i\|_2^2, \quad s.t. \|\mathbf{c}_i\|_0 = 1, \|\mathbf{c}_i\|_1 = 1, \mathbf{c}_i \geq 0, \forall i \quad (2.3)$$

where $\mathbb{C} = [\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_N]$ is the set that contains all the N codes for the input. $\mathbf{B} = [\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_M]$ is the codebook with M entries. The ℓ_0 -norm constrain $\|\mathbf{c}_i\|_0 = 1$ makes there exists only one non-zero element in the code $\mathbf{c}_i$. The ℓ_1 -norm constrain $\|\mathbf{c}_i\|_1 = 1$ with $\mathbf{c}_i \geq 0$ makes the non-zero element in the code equal to one. Basically, the method would assign the input vector $\mathbf{x}_i$ to its nearest neighbor in the codebook. This hard VQ brings large quantization errors during the coding process, leading to a poor representation and subsequently inaccurate classifications. Besides the coding method, another problem with the SPM method is that, it is in the linear form and need a non-linear classifier to obtain good performance. A non-linear classifier would be more time-consuming than the linear ones.

To improve the hard vector quantization process in the SPM method, the sparse coding (SC) [60] algorithm is used and the method is represented as ScSPM [66]. In ScSPM, the ℓ_0 constraint of the SPM method is relaxed to a sparsity regularization term (ℓ_1 -norm term), making the input $\mathbf{x}_i$ can be represented by more than one entries in the codebook. The process can be expressed as follows:

$$\arg \min_{\mathbb{C}} \sum_{i=1}^N \|\mathbf{x}_i - \mathbf{B}\mathbf{c}_i\|_2^2 + \omega \|\mathbf{c}_i\|_1, \quad (2.4)$$

where ω is the regularization parameter. This sparse coding makes the input vector can be represented by multi salient patterns in the codebook and therefore, reduce the quantization error then the hard VQ used in SPM. The process is nonlinear and can be combined with a linear classier in the classification process to reduce the time consumed.

The ScSPM method is further improved based on the finding that the data coding should not only be sparse but also be local [67], which means the sparse weights should be assigned to the dictionary codes that similar to the input. The coding method is named as locality-constrained linear coding (LLC) and the coding

process can be expressed as follows:

$$\min_{\mathbf{c}} \sum_{i=1}^N \|\mathbf{x}_i - \mathbf{B}\mathbf{c}_i\|_2^2 + \omega \|\mathbf{d}_i \circ \mathbf{c}_i\|_2^2, \quad s.t. \mathbf{1}^T \mathbf{c}_i = 1, \forall i \quad (2.5)$$

where $\circ$ is the element-wise multiplication, and $\mathbf{d}_i$ is the adaptor that weights different code entries according to the distance between the input and the codes. In this way, the coding would be done by considering the local information: The code entries with small distance with the input are more focused. This ensures that an input can be represented by a group of similar codes but not the ones with large differences. With the reasonable coding process, the classification performance can subsequently be improved.

2.1.3 Sparse and Dense Representation

Sparse and dense representation (SDR) [58] is a recently developed global image classification approach, which is based on the subspace dictionary learning. It tackled the limitation of the sparse representation-based classifier (SRC) [57], the lack of within-class variation in the representation, by introducing a dense combination of a common within-class variation dictionary. The corresponding data representation can be expressed as follows:

$$\mathbf{x} = \mathbf{C}\boldsymbol{\alpha} + \mathbf{D}\boldsymbol{\beta} + \mathbf{e} \quad (2.6)$$

where $\mathbf{C}\boldsymbol{\alpha}$ is the class-specific component of the input $\mathbf{x}$ while $\mathbf{D}\boldsymbol{\beta}$ is the non-class-specific component, $\mathbf{e}$ is the residual noise. The matrices $\mathbf{C}$ and $\mathbf{D}$ contain the between- and within-class dictionaries, respectively. To obtain the class information of the input data $\mathbf{x}$, the coefficient vector $\boldsymbol{\alpha}$ is set as a sparse one so that the non-zero terms of the vector can determine the class of the input. This constraint can be realized by minimizing the ℓ_1 -norm of the vector. In contrast to $\boldsymbol{\alpha}$, the

vector β does not provide the class-specific information so that it does not need to be sparse, which can be determined by minimizing its ℓ_2 -norm. The residual $\mathbf{e}$ is determined by minimizing its ℓ_1 -norm to make the representation robust to heavily corrupted input by outlier pixels and occlusions. Therefore, the overall optimization problem for determining the representation information in (2.6) can be expressed as follows:

$$\min_{\alpha, \beta, \mathbf{e}} \|\alpha\|_1 + \omega_1 \|\beta\|_2^2 + \omega_2 \|\mathbf{e}\|_1, \quad s.t. \mathbf{x} = \mathbf{C}\alpha + \mathbf{D}\beta + \mathbf{e} \quad (2.7)$$

where ω_1 and ω_2 are the user-defined parameters. By solving the above minimization problem, the sparse vector α that contains the class information of the input is obtained. The final classification decision is made based on the difference between the representation error of using all the training samples $\mathbf{e}$ and the representation error of using the samples in a class i , $\mathbf{e}_i$. The input sample is classified into the class with the smallest representation error difference:

$$C = \arg_i \|\mathbf{e} - \mathbf{e}_i\|_2 \quad (2.8)$$

where C is the assigned class.

The SDR method can have a good representation of the input data distribution by separating the within-class variance from the between-class variance. In such a way, the identity variable determined by the between-class information becomes more accurate and the method becomes more robust to the noises. One shortcoming of the method is that it is a distance based method, which was shown to be less superior than the probabilistic based models [68]. Therefore, the probabilistic based methods are introduced in the next.

2.1.4 Probabilistic Linear Discriminant Analysis

Probabilistic linear discriminant analysis (PLDA) uses the generative model, which incorporates both within- and between-class variances. In PLDA, the input data are regarded as a linear combination of vectors that associated with the within- and between-class covariances [33]:

$$\mathbf{x}_{ij} = \boldsymbol{\mu} + \mathbf{F}\mathbf{h}_i + \mathbf{G}\mathbf{w}_{ij} + \boldsymbol{\epsilon}_{ij}, \quad (2.9)$$

where $\mathbf{x}_{ij}$ is the j th data sample of the i th class, i and j are in the index sets of $[1, 2, \dots, N_i]$ and $[1, 2, \dots, N_j]$, respectively. $\boldsymbol{\mu}$ is the mean value of the whole training data. In (2.9), $\mathbf{F}$ and $\mathbf{G}$ denote two matrices whose columns are the basis vectors of the between-class covariance and within-class covariance, respectively. $\mathbf{h}_i$ and $\mathbf{w}_{ij}$ are the corresponding weights of the basis vectors. $\mathbf{h}_i$ represents the between-class character and the row elements remain the same for all the samples from the same class i . In contrast, $\mathbf{w}_{ij}$ represents the within-class character and the row elements change with not only the class label i but also the sample label j . $\boldsymbol{\epsilon}$ is a stochastic noise term with a diagonal covariance matrix $\boldsymbol{\Sigma}$. The assembled parameters $\{\boldsymbol{\mu}, \mathbf{F}, \mathbf{G}, \boldsymbol{\Sigma}\}$ are denoted as $\boldsymbol{\theta}$.

From model (2.9) and the assumption of Gaussian distributions of the hidden variables $\mathbf{h}_i$ and $\mathbf{w}_{ij}$ [33]:

$$P(\mathbf{h}_i) = \mathcal{N}_{\mathbf{h}_i}(\mathbf{0}, \mathbf{I}), \quad (2.10)$$

$$P(\mathbf{w}_{ij}) = \mathcal{N}_{\mathbf{w}_{ij}}(\mathbf{0}, \mathbf{I}). \quad (2.11)$$

We can derive the conditional probability as follows [33]:

$$P(\mathbf{x}_{ij}|\mathbf{h}_i, \mathbf{w}_{ij}, \boldsymbol{\theta}) = \mathcal{N}_{\mathbf{x}_{ij}}(\boldsymbol{\mu} + \mathbf{F}\mathbf{h}_i + \mathbf{G}\mathbf{w}_{ij}, \boldsymbol{\Sigma}), \quad (2.12)$$

where $P(\cdot)$ and $P(\cdot|\cdot)$ denote prior probability and conditional probability, respectively; and $\mathcal{N}_{\mathbf{a}}(\mathbf{m}, \mathbf{C})$ denotes a Gaussian distribution in $\mathbf{a}$ with mean $\mathbf{m}$ and covariance $\mathbf{C}$. The data distributions in the case of binary class and a two-dimensional dataset of this model are illustrated in Figure 2.1. In Figure 2.1, the two large ellipses represent the covariances of two different classes. The small ellipses inside the large ones are the distributions of individual samples in each class. The between-class shift ($\boldsymbol{\mu} + \mathbf{F}\mathbf{h}_i$) and within-class shift ($\mathbf{G}\mathbf{w}_{ij}$) are represented by dashed arrows and solid arrows, respectively. $\mathbf{x}_t$ denotes the testing sample which belongs to class 1 in this illustration. It can be observed that the within-class variance is small in the illustration and the two distributions are distinctively distributed.

FIGURE 2.1: Data distributions of the PLDA model when the within-class variance is small.

Based on (2.10) - (2.12), $P(\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{iJ} | \mathbf{h}_i, \mathbf{w}_{ij}, \boldsymbol{\theta})$ is the joint probability of different samples that belong to the same class i . Similarly, the joint probability $P(\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{iJ}, \mathbf{x}_t | \mathbf{h}_i, \mathbf{w}_{ij}, \boldsymbol{\theta})$ considers not only the samples of class i but also the testing sample $\mathbf{x}_t$. Consequently, by considering all of the training samples, the probability of the testing sample $\mathbf{x}_t$ belongs to the i th class can be expressed

as [33]:

$$P_i = P(\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{iJ}, \mathbf{x}_t) \prod_{p=1, p \neq i}^I P(\mathbf{x}_{p1}, \mathbf{x}_{p2}, \dots, \mathbf{x}_{pJ}). \quad (2.13)$$

The classification can be obtained by assigning $\mathbf{x}_t$ to the class i which has the largest probability P_i :

$$C = \arg \max_i P_i, \quad (2.14)$$

where C is the output class of the testing sample $\mathbf{x}_t$. The parameters in $\boldsymbol{\theta}$ can be derived by a Bayesian inference approach, further details can be found in [33, 68, 69].

Here, it is valuable to note that the original PLDA model assumes a small within-class variance as in the situation shown in Figure 2.1, which could not always hold in reality. In the calculation of the joint probability in (2.13), it assumes that all the samples in class i contain discriminant information. However, this assumption is true only when the within-class variance is small: that all the samples in the same class would be closely located to each other implies similar contribution to the joint probability P_i in (2.13). On the contrary, when the within-class variance is large, samples in the same class can have large differences and hence violate the assumption made in the original PLDA.

The impact of the large within-class variance on the performance of PLDA is illustrated in Figure 2.2. In this figure, two important factors impacting on PLDA can be identified. Firstly, with the large within-class variance, the distance between different samples from the same class could be large, for example, for samples $\mathbf{x}_{11}$ and $\mathbf{x}_{12}$. Sample $\mathbf{x}_{12}$ is close to the testing sample $\mathbf{x}_t$ (assuming that $\mathbf{x}_t$ is from class 1) and it suggests that $\mathbf{x}_t$ is likely to be from class 1. On the other hand, sample $\mathbf{x}_{11}$ leads to the opposite conclusion because it has a large distance to $\mathbf{x}_t$. Therefore, when the within-class variance is large, it would be improper to consider all the within-class samples because the large divergence of samples

FIGURE 2.2: Data distributions of the PLDA model when the within-class variance is large. The categorical overlapping is highlighted.

could lead to wrong predictions for a testing sample. Secondly, when the within-class variance is large, it could result in the overlapping between the categorical samples. The subspace overlapping between classes contains samples with small differences between categories. As seen in Figure 2.2, the distance between $\mathbf{x}_{21}$ and $\mathbf{x}_t$ is the smallest among the three training samples. Hence, sample $\mathbf{x}_{21}$ provides misleading information in deciding the class of the testing sample $\mathbf{x}_t$. By utilizing the information of all the training samples together in (2.13), it is clearly seen that the original PLDA does not consider the above mentioned two factors in dealing with the large within-class variance problem.

To tackle the large within-class variance problem, an improved PLDA method named tied-PLDA was proposed [32, 33]. Originally, tied-PLDA was proposed to solve the face recognition problem with large pose changes in the images. By dividing the training samples in each class into subclasses according to the pose angles, the joint probability derived from the tied-PLDA model is given by [33]:

$$P(\mathbf{x}_{ijk}|\mathbf{h}_i, \mathbf{w}_{ijk}, \boldsymbol{\theta}) = \mathcal{N}_{\mathbf{x}_{ijk}}(\boldsymbol{\mu}_k + \mathbf{F}_k \mathbf{h}_i + \mathbf{G}_k \mathbf{w}_{ijk}, \boldsymbol{\Sigma}_k), \quad (2.15)$$

FIGURE 2.3: Tied-PLDA model. The categorical overlapping between subclasses is highlighted. The subclass 1 is labeled.

where k is the subclass label decided by the pose angle. $\mathbf{x}_{ijk}$ denotes the j th sample in the k th pose of class i . $\theta_k = \{\mu_k \mathbf{F}_k \mathbf{G}_k \Sigma_k\}$ is the PLDA parameter of the k th subclass.

The tied-PLDA model is illustrated in Figure 2.3. In this figure, the input data of each class are separated into subclasses. When deciding the class of the testing sample $\mathbf{x}_t$, it only needs to consider the samples in one subclass which represents the testing sample the best. As a result, the training samples far away from the testing sample are not considered. As shown in Figure 2.3, the testing sample $\mathbf{x}_t$ belongs to the subclass 1. Therefore, only the samples in subclass 1 of the two classes will compete to make the final classification decision. In addition to avoiding the misleading samples which are far from the testing sample in participating the classification process, another advantage of tied-PLDA is that it reduces the categorical overlapping. As shown in Figure 2.3, the overlapping between subclasses would only be part of the overlapping between the entire classes. With the reduced overlapping, the two classes could be better separated and the testing sample is more likely to be classified correctly.

FIGURE 2.4: Example images of melasma patients.

2.2 Melasma and Its Clinical Severity

Assessment Methods

As mentioned in the introduction part, severity assessments of the skin pigmentation disorders are the keys to tracking the performance of the treatments. Within the different severity assessment methods that deal with different pigmentation disorders, our study focuses on simulating the modified melasma area and severity index (mMASI) system. The details about the melasma pigmentation disorder and the corresponding assessment method mMASI are reviewed in this section.

2.2.1 Melasma

Melasma is an acquired disorder of hyperpigmentation which commonly affects women of Fitzpatrick skin types IV-VI [70]. The melasma disease commonly appears as diffused brown patches on the face region. Some example images of the melasma patients are shown in Figure 2.4. Melasma is widely spread around the world and is a common presenting complaint in dermatology clinics. It affects more than 5 million people in the U.S.A. [71, 72]. In Southeast Asia, the prevalence

reaches 40% in women and 20% in men [38]. As a chronic and often relapsing condition, it has a significant negative impact on quality of life, even more so than in dermatoses such as acne and rosacea [4]. Melasma is therapeutically challenging and new treatment options are constantly being explored. However, a common obstacle faced is the lack of good and sound outcome measures, which is a critical but difficult-to-standardize aspect in clinical trials.

FIGURE 2.5: The face regions divided in mMASI score system.

2.2.2 mMASI

The Melasma Area and Severity Index (MASI) and modified MASI score (mMASI) are the most frequently used instruments to measure disease severity and treatment response in clinical trials. The MASI is calculated based on the darkness, homogeneity and extent of pigmentation [21]. The MASI has been validated and performed well when compared with the categorical Melasma Severity Scale and mexameter readings. The validated mMASI score, which removes homogeneity as a scoring component, has been shown to be as reliable as MASI scoring [22]. Therefore, in this study, we focus on the simulation of the mMASI system which is illustrated in Figure 2.5.

In Figure 2.5, it can be observed that the face of a patient is divided into 4 regions: forehead, chin, left malar and right malar. The assessor would check the severity of each of the four regions and come to an overall judgement by combining all the information. For each region, two properties of the pigmentation patch are used to describe the severity, namely the area percentage of the pigmentation and the darkness level of the pigmentation. Different scores are defined to describe the severity levels of the two properties as shown in Table 2.1 and Table 2.2, respectively. The final mMASI score is calculated as follows:

$$\text{mMASI} = 0.3A_fD_f + 0.3A_{lm}D_{lm} + 0.3A_{rm}D_{rm} + 0.1A_cD_c \quad (2.16)$$

where A and D denote area values and darkness values of melasma pigment involvement of a patient, respectively. The subscripts denote the four parts of a face image as shown in Figure 2.5, forehead region (f), right malar region (rm), left malar region (lm) and chin region(c).

TABLE 2.1: The area score defined in mMASI score system.

Area Score (A)	0	1	2	3	4	5	6
Percentage	0	1-9%	10-29%	30-49%	50-69%	70-89%	90-100%

TABLE 2.2: The darkness score defined in mMASI score system.

Darkness Score (D)	0	1	2	3	4
Degree	Absent	Silght	Mild	Marked	Maximum

2.3 Thresholding Based Image Segmentation Methods

Computerized image segmentation methods are widely applied in skin pigmentation image processing for the purpose of extracting the pigmentation region

[12, 40, 73, 74]. It is a fundamental and important step of the processing and can determine the accuracy of the subsequent analysis. In general, the segmentation process is based on different properties of the area of interest, for example, the discontinuities between different regions and the homogeneity within a region. Different image segmentation methods have been developed in the literature and some major types include thresholding based methods, edge based methods, region based methods, atlas based methods and artificial intelligence based methods. As the thresholding based methods are focused in this study, the related works are reviewed briefly in this section.

The segmentation model of thresholding methods is simply formulated as:

$$B(x, y) = \begin{cases} 0, & I(x, y) \leq T(x, y) \\ 1, & I(x, y) > T(x, y) \end{cases} \quad (2.17)$$

where $B(x, y)$ is the resultant segmentation label, 0 or 1, for a pixel at location (x, y) , $I(x, y)$ denotes its pixel intensity (or brightness) and $T(x, y)$ denotes a decision threshold for this pixel. Based on this segmentation model, each pixel is sequentially operated and foreground pigmentation regions are finally segmented out from a given image.

The thresholding segmentation model shown in (2.17) is mathematically simple. However, thresholding segmentation performance depends on the setting of thresholds $T(x, y)$ and therefore the determination of these thresholds is critical to segmentation performance. Basically, the methods can be divided into two categories, namely global and local. The global methods would obtain a fixed threshold value based on the information of the whole image while the local methods would change the threshold value adaptively according to the local information. In the next, typical methods for determining the threshold value are presented.

2.3.1 Otsu's Method

Otsu's global thresholding method [75, 76] is one of the most successful image segmentation algorithms. The threshold value is determined by minimizing the total intra-class variance as shown in (2.18).

$$T_O = \arg \min_t \sum_{I(x,y) < t} (I(x,y) - \mu_1)^2 + \sum_{I(x,y) \geq t} (I(x,y) - \mu_2)^2, \quad (2.18)$$

where

$$\mu_1(t) = \frac{1}{N_1} \sum_{I(x,y) < t} I(x,y), \quad \mu_2(t) = \frac{1}{N_2} \sum_{I(x,y) \geq t} I(x,y), \quad (2.19)$$

are the mean of the two pixel sets that divided by the threshold t and N_1 , N_2 are the number of incidents in each set. The problem with Otsu's method and all the other global thresholding methods is that they are not adaptive and cannot satisfactorily solve the problems with background intensity variations.

2.3.2 Adaptive Otsu Thresholding

The adaptive Otsu (AdOtsu) thresholding method [77, 78] is developed based on the Otsu's method. Instead of considering the whole image information, the AdOtsu method tries to select different threshold values based on information in local windows. In this way, the Otsu's method can adaptive to the changes in the local region. The formula of the AdOtsu is as follows:

$$T_{AdO}(x,y) = \Theta \left(\frac{|T_O - T_O(x,y)|}{c} - 1 \right) (T_{O,1-I}(x,y) - T_O(x,y)) + T_O(x,y) \quad (2.20)$$

where T_O is the threshold from the Otsu's method introduced previously, $T_O(x,y)$ is the Otsu's threshold constrained by a local window centered at (x,y) . Subscript $1 - I$ in $T_{O,1-I}$ denotes the inverse of the input image. Θ is a step function

($\Theta(x) = 0$ if $x \leq 0$ and $\Theta(x) = 1$ if $x > 0$) and c is a constant. The method considers the situation where only samples from one class exist in the local window. In the situation where only the background is presented in the local window, the step function would be 1 and the threshold of the inverse image $T_{O,1-I}(x, y)$ would be used. This prevents the large area of background misclassified as foreground. However, there is a problem with this method that it does not consider the situation where the local window only contains the foreground pixels. In such situation, all the foreground pixels would be misclassified as background pixels.

2.3.3 Sauvola's Method

Like the AdOtsu's method, Sauvola's method [79] is also a local thresholding method. It makes use of the regional mean and variance to determine the threshold value of each pixel as follows:

$$T_{Sauvola}(x, y) = \mu(x, y) + [1 + k(\frac{\sigma(x, y)}{R} - 1)], \quad (2.21)$$

where $\mu(x, y)$ is the mean value of the pixels within the local region centered at (x, y) , σ is the standard deviation of the local region, R is the dynamic range of standard deviation and k is a user-determined constant. Sauvola's method can adapt to the change of pixel values and is robust to noises.

Although the local thresholding methods can adapt to different situations well and thus obtain more reasonable segmentation results, they only utilize the local information and thus does not have a global view of the problem. One main problem with the local thresholding methods is that it is hard to select the size of the local region, especially when the size of foreground varies a lot. To solve this problem, the multi-scale methods are proposed in [77]. In the multi-scale image thresholding methods, segmentation results from different local window sizes are

considered. In this way, the foreground objects with different sizes can be properly identified when different window sizes are applied.

Chapter 3

Classification of Skin Pigmentation Disorders

In this chapter, the problem of classifying five selected non-tumorous skin pigmentation disorders using computerized methods is studied. Section 3.1 briefly explains the main challenge in the image classification problem, followed by the proposed V-PLDA method that designed to tackle this challenge in Section 3.2. The details of the experiment on real-world skin pigmentation images are presented in Section 3.3. Finally, a summary of the whole chapter is given in Section 3.4.

3.1 Introduction

This study focuses on the classification of five types of non-tumorous skin pigmentation disorders based on their appearance, which are: freckles, lentigines, Hori's nevus, melasma, and nevus of Ota. These five types are selected because they are the most common types [80] and they can be considered as representatives of appearance diagnosis. The other types of non-tumorous skin pigmentation disorders of the face, such as Lichen planus pigmentosus, acanthosis nigricans, maturation pigmentation, are rare and they usually need more professional investigation, such as biopsy, rather than just using appearance information. Therefore, they are not included in this study.

FIGURE 3.1: The process of clinical diagnosis process in differentiating freckles, lentigines, Hori's nevus, melasma and nevus of Ota.

Each of the five non-tumorous skin pigmentation disorders has its distinct appearance features. The clinical gold rules of diagnosis for these pigmentation disorders are summarized in the flowchart shown in Figure 3.1. Although the decision tree looks concise and easy to follow, the classification of the real-world problems

can be much more complicated because the pigmentation patches do not show their appearance properties consistently. Some more images of the five classes are shown in Figure 3.2. It can be observed that within each class, the pigmentation regions change a lot in terms of sizes, colors, shapes, etc. The human observer can automatically detect and ignore these variations and therefore make an accurate decision. However, in the computerized methods, certain measures need to be taken to tackle such large within-class variance problems to achieve better performance.

FIGURE 3.2: Four sample images of each skin pigmentation disorder to illustrate the large within-class variance. (a) Freckles. (b) Lentigines. (c) Hori's nevus. (d) Melasma. (e) Nevus of Ota.

As reviewed in Section 2.1, different methods that can adapt to the within-class variance have been developed in the literature. The recently developed probabilistic linear discriminant analysis (PLDA) method is particularly successful because

it models the data distribution more accurately. The improved PLDA model, tied-PLDA, is designed to tackle the large within-class variance problem as reviewed. Despite its superior power, the tied-PLDA model requires prior knowledge of the data, for example, the pose angle in the face recognition problem [33], to divide the input data properly in the training process. However, this prior information is not available in the skin pigmentation image classification problem. To deal with the large within-class variance problem without prior knowledge provided, we propose a voting based PLDA (V-PLDA) scheme in the following.

3.2 Improved PLDA for Large Within-Class Variance Problems

In this section, we first show that PLDA can be approximated by an individual matching process when the within-class variance is large. With this approximation, the voting based PLDA (V-PLDA) is introduced subsequently.

As reviewed in Section 2.1.4, all of the samples in class i are assumed to contain discriminant information in the calculation of the joint probability in (2.13). However, this assumption is true only when the within-class variance is small or moderate. In other words, all of the samples in the same class would have similar contribution to the joint probability P_i in (2.13). In contrast, when the within-class variance is large, samples in the same class can have large differences in their contributions to the joint probability and, hence, reduce the performance of the PLDA method. For example, as illustrated in Figure 2.2, assume that the two selected samples, $\mathbf{x}_{11}$ and $\mathbf{x}_{12}$, and the testing sample $\mathbf{x}_t$ are all from class 1. Apparently, sample $\mathbf{x}_{12}$ would have a larger joint probability with $\mathbf{x}_t$ than that $\mathbf{x}_{11}$ with $\mathbf{x}_t$. Therefore, while the information from $\mathbf{x}_{12}$ suggests that $\mathbf{x}_t$ is likely from class 1, the information from $\mathbf{x}_{11}$ tends to suggest the opposite due to its small

joint probability with $\mathbf{x}_t$. By considering both samples $\mathbf{x}_{i1}$ and $\mathbf{x}_{i2}$, the joint probability $P(\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{iJ}, \mathbf{x}_t)$ in (2.13) would be small and unreliable for the decision making. Hence, the original PLDA method requires further improvement in the situation of large within-class variance.

3.2.1 PLDA as Individual Matching

To understand how the original PLDA model works in the large within-class variance situation, the joint probability $P(\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{iJ} | \mathbf{h}_i, \mathbf{w}_{ij}, \boldsymbol{\theta})$ in the PLDA model should be studied first. To derive the detailed expression of the joint probability, the compound form of (2.9) by considering all the samples in class i together is shown in (3.1) [33].

$$\begin{bmatrix} \mathbf{x}_{i1} \\ \mathbf{x}_{i2} \\ \vdots \\ \mathbf{x}_{iJ} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu} \\ \boldsymbol{\mu} \\ \vdots \\ \boldsymbol{\mu} \end{bmatrix} + \begin{bmatrix} \mathbf{F} & \mathbf{G} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{F} & \mathbf{0} & \mathbf{G} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{F} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{G} \end{bmatrix} \begin{bmatrix} \mathbf{h}_i \\ \mathbf{w}_{i1} \\ \mathbf{w}_{i2} \\ \vdots \\ \mathbf{w}_{iJ} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon}_{i1} \\ \boldsymbol{\epsilon}_{i2} \\ \vdots \\ \boldsymbol{\epsilon}_{iJ} \end{bmatrix}, \quad (3.1)$$

which can be expressed compactly as [33]:

$$\mathbf{x}'_i = \boldsymbol{\mu}' + \mathbf{A}\mathbf{y}_i + \boldsymbol{\epsilon}'. \quad (3.2)$$

From (3.2), and by the assumption of Gaussian distribution of the hidden variable $\mathbf{y}_i$ [33]:

$$P(\mathbf{y}_i) = \mathcal{N}_{\mathbf{y}_i}(\mathbf{0}, \mathbf{I}), \quad (3.3)$$

the conditional probability can be derived as [33]:

$$P(\mathbf{x}'_i | \mathbf{y}_i) = \mathcal{N}_{\mathbf{x}'_i}(\boldsymbol{\mu}' + \mathbf{A}\mathbf{y}_i, \boldsymbol{\Sigma}'), \quad (3.4)$$

where $\boldsymbol{\Sigma}' = \text{diag}\{\boldsymbol{\Sigma}, \dots, \boldsymbol{\Sigma}\}$.

By marginalizing over $\mathbf{y}_i$, the marginal distribution is obtained [33]:

$$P(\mathbf{x}'_i) = \mathcal{N}_{\mathbf{x}'_i}(\boldsymbol{\mu}', \mathbf{A}\mathbf{A}^T + \boldsymbol{\Sigma}'). \quad (3.5)$$

Based on (3.1) and (3.2), the combined covariance matrix $\mathbf{A}\mathbf{A}^T + \boldsymbol{\Sigma}'$ is expressed as:

$$\mathbf{A}\mathbf{A}^T + \boldsymbol{\Sigma}' = \begin{bmatrix} \mathbf{D} & \mathbf{F}\mathbf{F}^T & \dots & \mathbf{F}\mathbf{F}^T \\ \mathbf{F}\mathbf{F}^T & \mathbf{D} & \dots & \mathbf{F}\mathbf{F}^T \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{F}\mathbf{F}^T & \mathbf{F}\mathbf{F}^T & \dots & \mathbf{D} \end{bmatrix}, \quad (3.6)$$

where

$$\mathbf{D} = \mathbf{F}\mathbf{F}^T + \mathbf{G}\mathbf{G}^T + \boldsymbol{\Sigma}.$$

The combined covariance matrix (3.6) includes the within-class matrix $\mathbf{G}$ and it provides us with a means to study its impact. When the within-class variance is large, the values of the elements in the within-class matrix $\mathbf{G}$ are large and, hence, the elements of $\mathbf{F}\mathbf{F}^T$ are negligible compared with the corresponding elements in $\mathbf{G}\mathbf{G}^T$. Consequently, the combined matrix $\mathbf{A}\mathbf{A}^T + \boldsymbol{\Sigma}'$ can be approximated by its diagonal elements. This approximation implies that the correlations between different samples in the compound vector $\mathbf{x}'_i = [\mathbf{x}_{i1} \ \mathbf{x}_{i2} \ \dots \ \mathbf{x}_{iJ}]^T$ are small. In this case, the joint probability can be approximated by the product of the individual probabilities:

$$P(\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{iJ}) \approx P(\mathbf{x}_{i1})P(\mathbf{x}_{i2})\dots P(\mathbf{x}_{iJ}). \quad (3.7)$$

Based on (3.7), the joint probability of $\mathbf{x}_t$ and the samples in class i can be checked separately, which gives:

$$P(\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{iJ}, \mathbf{x}_t) \approx \max_j P(\mathbf{x}_{ij}, \mathbf{x}_t) \prod_{q=1, q \neq j}^J P(\mathbf{x}_{iq}). \quad (3.8)$$

By substituting (3.7) and (3.8) into (2.13), the probability of the testing sample $\mathbf{x}_t$ belonging to the i th class is:

$$\begin{aligned} P_i &\approx \max_j P(\mathbf{x}_{ij}, \mathbf{x}_t) \prod_{q=1, q \neq j}^J P(\mathbf{x}_{iq}) \\ &\quad \times \prod_{p=1, p \neq i}^I P(\mathbf{x}_{p1}) P(\mathbf{x}_{p2}) \dots P(\mathbf{x}_{pJ}) \\ &= \max_j P(\mathbf{x}_{ij}, \mathbf{x}_t) \prod_{(p,q) \neq (i,j)} P(\mathbf{x}_{pq}) \\ &= \max_j P_{ij}, \end{aligned} \quad (3.9)$$

where $P_{ij} = P(\mathbf{x}_{ij}, \mathbf{x}_t) \prod_{(p,q) \neq (i,j)} P(\mathbf{x}_{pq})$ is the probability that $\mathbf{x}_t$ is from the same distribution with only the training sample $\mathbf{x}_{ij}$. Therefore, the derived class label C of the testing sample in (2.14) is now obtained approximately as:

$$C = \arg \max_i P_i \approx \arg \max_i \left[\max_j P_{ij} \right]. \quad (3.10)$$

We name the above class assignment process ‘individual matching’ because the training samples are checked individually to obtain the joint probabilities with the testing sample $\mathbf{x}_t$.

It should be noted that the individual matching method can perform well not only in the situation of large within-class variance, but also in the situation of very small within-class variance. This is because when the within-class variance is much smaller than the between-class variance, the elements of $\mathbf{G}\mathbf{G}^T$ are negligible compared with the corresponding elements in $\mathbf{F}\mathbf{F}^T$. This is the other extreme

situation and the covariance matrix in (3.6) can be treated as all the elements have the same value. The input samples are perfectly correlated and the joint probability can be approximated by the individual probabilities:

$$P(\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{iJ}) \approx P(\mathbf{x}_{i1}) \approx P(\mathbf{x}_{i2}) \dots \approx P(\mathbf{x}_{iJ}). \quad (3.11)$$

When the number of training samples in each class is the same. The classification can be done using the same procedure shown in (3.8) and (3.9). The above analysis of the approximation process is further evidenced by the following simulation.

An example with synthetic data is used to show the impact of the within-class variance on the classification accuracy when both the individual matching and PLDA methods are applied. Assume that there exist two classes with fixed class centers of $\boldsymbol{\mu}_1 = \boldsymbol{\mu} + \mathbf{F}\mathbf{h}_1 = [0.8 \ 0.6]^T$ and $\boldsymbol{\mu}_2 = \boldsymbol{\mu} + \mathbf{F}\mathbf{h}_2 = [0.19 \ 0.98]^T$. Under this simple set-up, both matrices $\mathbf{F}$ and $\mathbf{G}$ are reduced to a vector as there is only one basis vector in them. To simulate different levels of the within-class variance, $\mathbf{G}$ is set to have different scales by multiplying the direction vector with a scalar ν :

$$\mathbf{G} = \nu \cdot \frac{\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2}{\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|_2}. \quad (3.12)$$

When the value of ν is larger, the within-class variance will be larger. The basis vector in $\mathbf{G}$ is set to have the direction of $\frac{\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2}{\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|_2}$ so that categorical distributions overlap with each other when the within-class variance is large. The parameter $\mathbf{w}_{ij}$ is randomly generated from a Gaussian distribution defined in (2.11). The covariance of the noise term is set as $\boldsymbol{\Sigma} = \text{diag}\{0.1, 0.1\}$.

In the simulation, 20 samples from each class are generated according to the distributions that have previously been defined. The 10-fold cross-validation procedure [81] is used for the performance evaluation. The data are randomly divided into 10 equal sized subsets, where nine are used for training and the remaining subset

FIGURE 3.3: Average classification accuracies of the PLDA method and the individual matching method at different levels of the within-class variance. The two vertical lines separate the plot into three parts.

is used for testing. The same procedure is then repeated for the next subset until all 10 subsets have been tested, one by one. The average accuracy of the above 10 tests is obtained. To reduce the randomness, the above training and testing procedure is repeated 10 times randomly. The final average classification accuracy of these 10 runs is used for the performance evaluation.

The classification accuracy results are plotted versus the values of ν in Figure 3.3, where the accuracy curves can be roughly divided into three parts. The first part is the region that the within-class variance is extremely small ($\nu < 0.2$) and, hence, the two classes are well separated. This is the situation that (3.11) describes. It can be well approximated by the individual matching method from the analysis. As a result, the individual matching method can achieve similar high level of accuracy to that of PLDA. In the second part, when the within-class variance ($0.2 \leq \nu < 1.5$) increases but is not very large, the accuracy difference between PLDA and the individual matching method is large. The increase of the accuracy difference is due to the error that is caused by approximating PLDA with the individual matching method. In the third part, the classification accuracies of

the two methods converge to about the same level when the within-class variance is very large ($\nu \geq 1.5$). The convergence of the classification accuracies of the two methods in the third part agrees with the previous observation that PLDA can be approximated by the individual matching when the within-class variance is large enough. The statistical analyses of the results by the paired-sample t-test [82] support this argument. The t-test results show that there is a significant difference in the accuracies obtained from the PLDA and the individual matching process in the second part; for example, at $\nu = 0.5$, $t(999) = 6.7827$, $p < 0.001$. In contrast, there is no significant difference in the third part; for example, at $\nu = 2.0$, $t(999) = 0.2442$, $p > 0.05$.

3.2.2 Voting Based PLDA

One problem with using the individual matching method to approximate PLDA is that the sample with the largest joint probability P_{ij} in (3.9) may not always be the desired one. With the large distribution overlapping between classes, it is possible that the sample with the largest joint probability P_{ij} is from a different class than the testing sample. Assume that in Figure 2.2, $\mathbf{x}_{21}$ is the sample that has the largest joint probability with the testing sample $\mathbf{x}_t$ among all of the samples shown but $\mathbf{x}_{21}$ is in a different class (class 2) from $\mathbf{x}_t$. The individual matching would then obtain a wrong classification decision if it uses $\mathbf{x}_{21}$ alone. We propose a voting scheme to deal with this problem and to make the decision process more robust. Instead of considering just a single sample $\mathbf{x}_{ij}$ that has the largest joint probability with $\mathbf{x}_t$, a bag of n samples with the largest joint probabilities with $\mathbf{x}_t$ are used in a voting procedure. The sample selection scheme can be expressed as follows:

$$\mathbb{S} = \{\mathbb{A}(1:n) | \mathbb{A} = \text{sort}(\{P_{ij} : i \in [1, 2, \dots, I]; j \in [1, 2, \dots, J]; \})\}, \quad (3.13)$$

FIGURE 3.4: The voting sample selection process. Only the ones having high joint probabilities with the testing sample $\mathbf{x}_t$ are selected (circled ones in the figure).

where $\text{sort}(\cdot)$ is a function that sorts the elements of the input dataset in a descending order based on their values. Consequently, the set $\mathbb{S}$ contains the n largest P_{ij} values. In the classification process, each of the n elements is associated with a class i . The final classification decision is made based on a voting procedure using these n samples:

$$C = \arg \max_i |\mathbb{S} \cap \mathbb{P}_i|; \mathbb{P}_i = \{P_{i1}, P_{i2}, \dots, P_{iJ}\}, \quad (3.14)$$

where $|\cdot|$ is the set cardinality. If more than one class has the same largest number of votes, then the class that contains the largest joint probability P_{ij} is selected. We term this proposed method as the voting based PLDA, or V-PLDA in short, to reflect the incorporation of the individual matching approximated PLDA with a voting process.

The voting process of the proposed V-PLDA is illustrated in Figure 3.4. The selected samples in (3.13) are those with largest joint probabilities with the testing sample $\mathbf{x}_t$. In this selection, the samples which are less correlated with $\mathbf{x}_t$ and

may lead to wrong classification results are not considered in the decision making. Another advantage of the voting process is that it makes the decision more robust. The testing sample $\mathbf{x}_t$ would be misclassified using the individual matching when the sample with the largest joint probability is from a different class to the testing sample. This misclassification could happen due to the categorical overlapping caused by the large within-class variance. With the voting process, the result would no longer be determined by just one sample but by several selected samples together. As shown in Figure 3.4, the number of selected samples in class 1 (a total of 4) is larger than that in class 2 (a total of 3). The classification decision is correct via the voting process regardless of which sample has the largest joint probability. The effectiveness of the proposed V-PLDA method is analyzed using another example with synthetic data, which is described in the following.

When the within-class variance is large, the samples in each class may be divided into subclasses where the within-class variance of each subclass is not large. Although we may not have prior knowledge of making this division in some practical applications, it is feasible to do so for synthetic data because it will enable us to understand the impact of the number of samples in each subclass on the performance of the proposed V-PLDA method. Note that within each subclass, the joint probabilities between samples are large because they usually share some common properties. In general, when a sample in a subclass has a large joint probability with the testing sample $\mathbf{x}_t$, then the other samples in the same subclass would also have a large joint probability with $\mathbf{x}_t$. Hence, when there are more samples in each subclass, there will be more samples in the voting process that have similar joint probabilities with the testing sample. This makes the final decision more robust. This observation is illustrated in the following simulation.

TABLE 3.1: Setup for the subclasses

No. of samples in each subclass m	2	5	10
No. of subclasses in each class K	10	4	2

FIGURE 3.5: Average classification accuracies of PLDA and V-PLDA at different subclass sizes.

In the simulation, we use the same set-ups as those in the previous simulation for the individual matching, except that the samples in each class are divided into several subclasses. All of the samples in each subclass share the same within-class character. With the number of samples in each class fixed to 20, the number of samples in each subclass (m) and the number of subclasses (K) can vary in different ways. The three set-ups that are listed in Table 3.1 are tested. In the simulation, the parameter n is selected from a set of values and the one that achieves the best performance is used. The simulation results are plotted in Figure 3.5. It can be observed that the classification accuracy increases with the number of samples in each subclass (m). Comparing PLDA with V-PLDA, the simulation clearly shows that with a fixed number of subclass samples, V-PLDA outperforms PLDA, particularly when the within-class variance is large. To examine the significance of the difference between the accuracies obtained from PLDA and V-PLDA, paired-sample t-tests have been conducted at $\nu = 2.0$ and the results are shown in Table 3.2. It can be observed that the differences in the accuracies for PLDA and V-PLDA at different m values are significant. From this simulation, we can find that

TABLE 3.2: Results of paired-samples t-test comparing PLDA and V-PLDA with different m values

m	PLDA		V-PLDA		t	p
	AC %	SD	AC %	SD		
10	65.27	0.2616	71.93	0.2440	7.482	$1.60e - 13$
5	59.75	0.2666	65.48	0.2346	8.748	$9.08e - 18$
2	54.40	0.2515	55.95	0.2346	2.408	0.016

AC: Average accuracy. SD: Standard Deviation. $N = 1000$.

V-PLDA is effective for classification problems with large within-class variance, particularly when the number of samples in each subclass is relatively large.

As reviewed in Section 2.1, another way of dealing the large within-class variance problem is using the tied analysis. It divides each class into several subclasses based on the similarity between samples. Each subclass contains only the ones having similar appearance. In this way, the large variance of each class is replaced by the small variances of the subclasses and the performance is improved. Although the tied analysis can have a good model of the data distribution, it requires prior knowledge of this subclass division, which is not available for the pigmentation dataset. In contrast, the proposed V-PLDA model is easy to apply and does not need such prior information.

In the next section, the V-PLDA method will be evaluated on a real-world non-tumorous skin pigmentation image dataset.

3.3 Experimental Results

In this section, the problem of classification of non-tumorous skin pigmentation disorders is investigated in details using experiments. The experimental results

of applying PLDA, V-PLDA and several state-of-the-art image classification algorithms on a real-world skin pigmentation image set we have at hand are presented, compared and analyzed.

3.3.1 Experimental Setup

The proposed algorithm is tested on a real-world dataset that consists 30 skin pigmentation images per type. To reduce the impact of external noises (e.g., illumination variations) on the diagnostic accuracy, the data collection process is done under standardized set-ups. A NIKON CMOS sensor digital single-lens reflex camera with 24.2 million pixels is utilized and the output image is saved in the JPEG format. Cosmetic makeup is removed for all of the patients. Flash is used while taking the images to ensure consistent lighting. The distance between the camera and the patient is normally less than 3 meters to record the whole face region, such as the examples shown in Figure 2.4. A patch that contains the whole or part of the pigmentation region is selected from the recorded image and is pre-processed to have the same dimensionality of $100 \times 100 \times 3$ for the subsequent classification. A representative input image of each type is shown in Figure 3.2.

In the experiment, the ground truth of classification is provided by dermatologists. Following the same experimental procedure on synthetic data as described in Section 3.2, the repeated 10-fold cross-validation is used to evaluate each algorithm. The voting parameter n used in the experiment is selected as 7 after comparing the cross-validation results of different settings (the details can be found in Section 3.3.2.1).

3.3.2 Input Feature Selection

With the input image prepared, proper features need to be extracted from them to feed into the V-PLDA model for the classification. The selection of the input feature is based on the clinical golden rules of diagnoses, which are summarized in the flowchart shown in Figure 3.1. The features mentioned in Figure 3.1 can be divided into two types: color features and non-color features. The detailed feature selection process of the two types is described in this subsection.

3.3.2.1 Color Feature Selection

As illustrated in Figure 3.1, color is a critical feature that differentiates the selected skin pigmentation disorders. To have a thorough investigation of the color feature, various color spaces, including RGB, rg (normalized RGB), transformed, opponent, Lab, and HSV are tested. These color spaces would reveal the color information of a pixel from many different perspectives [83]. To compare their performance, these color feature vectors are tested using the V-PLDA method with different voting values. The resulting average classification accuracies of the cross-validation test are plotted in Figure 3.6.

From Figure 3.6, it can be observed that the choice of color space has a great impact on the performance. In comparison, the HSV and Lab color spaces achieve better performance than the others. This may happen because the two spaces separate the intensity from the color information and, hence, match the human perception better [84]. The opponent color space performs somewhat better than the RGB color space because it decorrelates the red, green, and blue information in the color, making the channels contain less redundant information. The rg and the transformed color spaces perform poorly in the experiment. They share the

FIGURE 3.6: Average classification accuracy of V-PLDA model using different color spaces.

similarity of normalizing the input RGB space. This normalization process loses important scale information, so that the classification accuracy is reduced.

Based on the results in Figure 3.6, we select the HSV color space to extract the color information of the image in the rest of the experiment. The detailed color feature extraction process is shown in Algorithm 1. The output color feature vector $\mathbf{x}_{HSV}$ is then fed into the V-PLDA model for the classification.

Using the HSV color information as the input, the PLDA method achieves similar and low accuracy (64.4%) with the individual matching method (62.3%). This indicates that the within-class variance of the pigmentation image data is large, which corresponds to the second region of the Figure 3.3. It can also be observed from Figure 3.6 that the best performance of the algorithm is obtained when the voting number n is around 7. Hence, we fix the voting number as 7 in the rest of the experiment. To check the statistical significance of the results at $n = 7$, a repeated measures analysis of variance has been conducted to compare the effect of color spaces on the classification accuracy. The test shows that there is a

significant effect of color spaces on the classification accuracy for the six color spaces: $F(5, 495) = 33.96, p < 0.001$.

Algorithm 1 Color feature extraction process.

Require: The input color image in RGB color space, $\mathbf{X}_{RGB} \in \mathbb{R}^{p \times q \times 3}$;

Ensure: The color feature vector, $\mathbf{x}_{HSV}$;

```

1: for Each pixel  $(x, y)$  in the image do
2:    $[r \ g \ b] \leftarrow \mathbf{X}_{RGB}(x, y, :)$ ;
3:    $c \leftarrow \min(r, g, b)$ ;  $v \leftarrow \max(r, g, b)$ ;
4:    $s \leftarrow \begin{cases} 0 & \text{if } v = 0 \\ (v - c)/v & \text{else} \end{cases}$ 
5:    $h \leftarrow \begin{cases} 60(g - b)/(v - c) & \text{if } v = r \\ 120 + 60(b - r)/(v - c) & \text{if } v = g \\ 240 + 60(r - g)/(v - c) & \text{if } v = b \end{cases}$ 
6:   if  $h < 0$  then
7:      $h \leftarrow h + 360$ 
8:   end if
9:    $h \leftarrow h/360$ 
10:   $\mathbf{X}_{HSV}(x, y, :) = [h \ s \ v]$ ;
11: end for
12: Reshape  $\mathbf{X}_{HSV}$  to the vector form  $\mathbf{x}_{HSV}$ ;
13: return  $\mathbf{x}_{HSV}$ ;

```

3.3.2.2 Non-color Feature Selection

Besides the color information, the two non-color features mentioned in Figure 3.1, namely the speckled/diffused property and the large/small property of the pigmentation region need to be extracted from the input images. Among the developed image feature representations [85–87], Gabor filter bank [88] is one of the most common methods as its frequency and orientation representations are similar to the mechanism of the human visual system [89]. It has been particularly successful in many computer vision and image processing applications like face recognition [90, 91], fingerprint matching [92], etc. It is found that the Gabor filter bank method is also very powerful in extracting useful features from the skin pigmentation images. Therefore, the detailed Gabor feature extraction process of an input image is explained in this subsection.

In the spatial domain, a two-dimensional Gabor filter contains two parts, a Gaussian kernel function and a complex sinusoidal plane wave:

$$\Psi_{f,\phi}(x, y) = \frac{f^2}{\pi\gamma\eta} \exp^{-f^2[(x'/\gamma)^2 + (y'/\eta)^2]} \exp^{2\sqrt{-1}\pi f x'}, \quad (3.15)$$

with:

$$\begin{aligned} x' &= x \cos \phi + y \sin \phi, \\ y' &= -x \sin \phi + y \cos \phi, \end{aligned}$$

where f is the central frequency of the filter, γ and η are the bandwidths along the major and minor axis of the Gaussian function, respectively. ϕ is the angle between the direction of the sinusoidal wave and the x-axis of the spatial domain. A Gabor filter bank contains multiple Gabor filters with different frequencies (f) and orientations (ϕ) to detect different information in the image. The frequency values are determined by two parameters: the maximum frequency (f_M) and the ratio between two adjacent frequencies (f_r). Therefore, the u th frequency, where $u \in [0, 1, \dots, U - 1]$, is defined as $f_u = f_M / f_r^u$. The orientations are equally separated in the range of $[0 \pi]$, where the v th orientation with $v \in [0, 1, \dots, V - 1]$ can be represented as $\phi_v = v\pi/V$. More details about Gabor filter bank can be found in [88, 93].

Given a Gabor filter bank, the Gabor feature of an input image matrix $\mathbf{X} \in \mathbb{R}^{p \times q \times 3}$ is obtained by first convolving each color channel of $\mathbf{X}$ with different Gabor filters Ψ using symmetric padding. For a Gabor filter with a frequency of f_u and an orientation of ϕ_v , the process can be described as:

$$\mathbf{O}_{f_u, \phi_v}(x, y, z) = \sum_{c=1}^w \sum_{r=1}^w \mathbf{X}(x - c, y - r, z) \Psi_{f_u, \phi_v}(c, r), \quad (3.16)$$

Algorithm 2 Non-color feature extraction using Gabor filter bank.

Require: The input image in RGB color space, $\mathbf{X}_{RGB} \in \mathbb{R}^{p \times q \times 3}$;
The Gabor filter bank $\Psi_{f,\phi}$ with user defined U, V values;
Ensure: The Gabor feature vector of the image, $\mathbf{x}_{Gabor}$;
1: Transform $\mathbf{X}_{RGB}$ to HSV space $\mathbf{X}_{HSV}$ using Algorithm 1;
2: **for** Each frequency $f_u : u \in [0, 1, \dots, U - 1]$ **do**
3: **for** Each orientation $\phi_v : v \in [0, 1, \dots, V - 1]$ **do**
4: Obtain the convolution results $\mathbf{O}_{f_u, \phi_v}$ using (3.16);
5: Calculate the magnitude of the convolution results $\mathbf{M}_{f_u, \phi_v}$ using (3.17);
6: Store $\mathbf{M}_{f_u, \phi_v}$ in the collective output matrix $\mathbf{M}_{f,\phi}$.
7: **end for**
8: **end for**
9: Reshape the collative output $\mathbf{M}_{f,\phi}$ to the vector form $\mathbf{x}_{Gabor}$;
10: **return** $\mathbf{x}_{Gabor}$;

where $\mathbf{O} \in \mathbb{C}^{p \times q \times 3}$. $\Psi \in \mathbb{C}^{w \times w}$ is the Gabor filter with the size of $w \times w$. Subsequently, the magnitude matrix $\mathbf{M}$ of the convolution output $\mathbf{O}$ is calculated:

$$\mathbf{M}_{f_u, \phi_v}(x, y, z) = \sqrt{\text{Re}(\mathbf{O}_{f_u, \phi_v}(x, y, z))^2 + \text{Im}(\mathbf{O}_{f_u, \phi_v}(x, y, z))^2}, \quad (3.17)$$

where $\text{Re}(\cdot)$ and $\text{Im}(\cdot)$ correspond to real and imaginary parts of the input, respectively. Finally, the Gabor feature is obtained by concatenating all of the magnitude matrices $\mathbf{M}_{f_u, \phi_v}$ into the vector form $\mathbf{x}_{Gabor}$, so that it can be fed into the V-PLDA model for classification. The detailed Gabor feature extraction process is summarized in Algorithm 2.

TABLE 3.3: Parameter selection for Gabor filter bank

Parameter	Symbol	Levels	Values
Eta	η	3	0.5, 1.0, 1.5
Gamma	γ	3	0.5, 1.0, 1.5
Max frequency	f_M	2	$1/4, \sqrt{2}/4$
Frequency retio	f_r	2	$\sqrt{2}, 2$
Window size	W	4	15, 25, 35, 45

Gabor filters with different wavelet structures are obtained by changing the kernel size of the Gaussian function and the frequency of the sinusoidal wave, which can both be controlled by parameter f in (3.15). In the experiment, we employed a

Gabor filter bank that contains 40 Gabor filters in five frequency levels ($U = 5$) and eight orientations ($V = 8$). As shown in (3.15), there are several parameters define different properties of the Gabor filter bank: the standard deviations (η and γ), the maximum frequency (f_M) and the ratio between neighboring frequencies (f_r), the window size (W), etc. A change in the values of those parameters would affect the features extracted and subsequently the classification accuracy. Therefore, the effect of the mentioned five parameters $\{\gamma, \eta, W, f_M, f_r\}$ are studied with experiments. The values used in the first stage experiment are listed in Table 3.3 and the result plots of the average accuracy achieved versus different parameters are shown in Figure 3.7.

FIGURE 3.7: Average classification accuracy for different selections of each parameter.

From Figure 3.7, we can observe the following about the parameter selection:

- $f_r = 2$ performs better than $f_r = \sqrt{2}$ at nearly all the selections of other parameters. This is opposite to the situation in [93] that smaller f_r performs better. One reason for the difference is that the images used in [93]

are mostly filled with regular textures, which makes the frequency information constrained within a small range. Therefore, it would be better if the frequency covered by the filter bank is small (small f_r). On the other side, there are no regular textures in the skin pigmentation images. The frequency range should be larger by using a larger ratio.

- In the W plot, the changing trend of the two lines are opposite. This is because the change of the f_r affects the spacial range of the Gaussian: $\sigma_1 = \gamma/f$, where σ_1 is the standard deviation of the Gaussian function. For fixed f_M and γ values, the increase of f_r would lead to a decrease of the smallest frequency value and therefore the increase of the spatial range of the Gaussian. A larger range of Gaussian needs a larger window size to contain. However, the window size cannot be too large to exclude the non-class-related information.
- In the f_M plot, the four points correspond to four different frequency coverage of the filter bank. The one that best cover the frequency range of the dataset would achieve the best performance. Therefore, from the result, the frequency range of $[\sqrt{2}/(4 \times 2^4) \sqrt{2}/4]$ is the best choice.

From the above experiments and analyses, the parameter combination of

$$\{\gamma, \eta, W, f_M, f_r\} = \{1.5, 1.0, 45, \sqrt{2}/4, 2\}$$

is selected in extracting Gabor features for the skin pigmentation image classification problem. The corresponding Gabor filters are shown in Figure 3.8.

3.3.2.3 Combination of the Input Features

With both types of image feature—namely $\mathbf{x}_{HSV}$ and $\mathbf{x}_{Gabor}$ —extracted, it is necessary to combine the two types of information effectively for the subsequent

FIGURE 3.8: Illustration of the real part of the selected Gabor filter bank.

classification process. A common method of feature combination is concatenating them together to form a higher dimensional input vector. However, this method does not work well when the features are not in the same scale. To circumvent the scale difference problem, we propose an alternative to take advantage of the probability framework of the V-PLDA model. Specifically, the information from different features is combined by multiplying the corresponding output probabilities of the V-PLDA model together:

$$P_{ij} = P_{ij}^{HSV} \times P_{ij}^{Gabor}. \quad (3.18)$$

where P_{ij}^{HSV} is the probability derived from (3.9) using the HSV feature as the input. Likewise, P_{ij}^{Gabor} is the probability derived when the Gabor features are the used. To make the final classification decision, the combined probability P_{ij} in (3.18) of all the training samples goes through the same voting process in (3.13) and (3.14).

More outputs of different features can be combined by multiplying the probabilities

together. The model of fusing different types of features together can be illustrated by the chart in Figure 3.9. It needs to be noted that using more features does not necessarily mean achieving higher classification accuracies. In our problem, we only use the color and Gabor features after experiments. They can have a good representation of the clinical rules as shown in Figure 3.1.

FIGURE 3.9: The structure of the fused PLDA model.

3.3.3 Results and Discussions

With the combined features, we are now ready to evaluate the proposed V-PLDA method on the skin pigmentation image set. We start by comparing it with the original PLDA method. The average classification accuracies by PLDA and V-PLDA are listed in Table 3.4 together with the paired-sample t-test results. The average classification accuracies obtained by the proposed V-PLDA method are higher than those obtained by the original PLDA method with the same input features, while the accuracy differences are significant at a level of $p < 0.05$. Table 3.4 also shows that with the selected combined features HSV-Gabor, V-PLDA performs better than with HSV or Gabor alone.

One classification example of using the real image set is shown in Figure 3.10 to illustrate the advantage of the voting process. Within the seven selected voting samples in (b)–(h) have large joint probabilities with the testing sample in (a),

TABLE 3.4: Results of paired-samples t-test comparing PLDA and V-PLDA with different input features

m	PLDA		V-PLDA		t	p
	M %	SD	M %	SD		
HSV	64.40	0.1254	69.27	0.1020	-2.854	0.005
Gabor	69.93	0.1092	73.67	0.1139	-2.433	0.017
HSV-Gabor	70.20	0.1244	76.67	0.1135	-4.082	$0.90e^{-4}$

$N = 200$.

three samples, in (b)–(d), are from a different class to the testing sample. It is important to note that the sample in (b) has the largest joint probability P_{ij} with the testing sample in (a). Hence, the testing sample would have been assigned to the wrong class if we had used the individual matching to approximate the PLDA model. In contrast, the voting process with the seven samples selected leads to the correct classification decision.

FIGURE 3.10: (a) The testing sample which is from the class of freckles. (b)–(h) The selected samples in the voting process, where (b)–(d) are the samples from the class of Hori's nevus and (e)–(h) are from the class of freckles.

To further illustrate the superior performance of the V-PLDA method in dealing with the large within-class variance skin pigmentation problem, the proposed V-PLDA model is compared with several state-of-the-art image classification methods. The methods compared include one from the global approach category (sparse

and dense representation (SDR) [58]) and two from the local approach category (locality-constrained linear coding (LLC) [65], intersection kernel support vector machine (IKSVM) [94]). SDR is similar to PLDA because it also considers both within- and between-class variances to accurately model the given data. The difference is that SDR is a distance-based method rather than a probabilistic method like PLDA. LLC and IKSVM are representative methods that focus on different aspects of the local classification approaches. While LLC focuses on the feature extraction process by proposing an improved feature coding method, IKSVM focuses on the feature classification process by using a non-linear classifier. In comparison, all the three methods have the ability to adapt to the within-class variance in the data. To make sure the power of the above methods compared is fully utilized in the comparison, the parameters in those methods are fine tuned in the experiment. In LLC and IKSVM, the dense SIFT descriptor [62] is used after comparing different local descriptors. The experimental results of the proposed V-PLDA method and the methods compared are shown in Table 3.5.

From Table 3.5, it can be observed that the proposed V-PLDA model achieves the best performance among all four methods in comparison. The average classification accuracy of our skin pigmentation image set by the proposed V-PLDA is higher than other methods in comparison by about 4% or more. An analysis of variance (ANOVA) [82] has shown that the effect of the classification method used on the resulting accuracy is significant, $F(3, 796) = 27.19$, $p < 0.001$. A Tukey post hoc test [82] has revealed that the average classification accuracy is statistically significantly higher by using the proposed V-PLDA method when compared to the other three methods. These results and analysis show that for the skin pigmentation image set that we have at hand, the proposed V-PLDA model possesses superior discriminating power to the methods that were compared.

While the average classification accuracy gives an overview of the performance, a more detailed classification result of V-PLDA in Table 3.5 is shown in the confusion

TABLE 3.5: Results of the different methods

Method	AC %	SD
SDR [58]	67.63	0.1141
LLC [65]	71.47	0.1135
IKSVM [94]	73.30	0.1107
V-PLDA (HSV-Gabor)	77.33	0.0982

matrix in Table 3.6. It can be found that freckles, melasma, and lentigines can be classified more accurately than Hori's nevus and nevus of Ota. For Hori's nevus, over 20% of images are misclassified as freckles. As indicated in Figure 3.1, the two skin diseases are characterized by small speckles and the main difference is in the color. In the real-world situation, due to the large within-class variance, the slate-brown color of Hori's nevus differs only slightly from the brown color in the mild symptom case. Hence, images containing mild symptoms of Hori's nevus might be misclassified as freckles. The same problem exists for nevus of Ota due to the large within-class variance that is caused by the large variation in terms of both color and pigmentation patch size. Consequently, in the case of mild symptoms, the light pigmentation color and small pigmentation patch size of the nevus of Ota image differs only slightly from other types of pigmentation images. These images are more likely to be misclassified.

TABLE 3.6: Confusion matrix for the V-PLDA model (%) with combined HSV and Gabor features as its input

		Predicted Class				
		F	L	H	M	O
True Class	F	84.7	0.7	8.0	6.7	0.0
	L	7.0	80.7	8.7	3.7	0.0
	H	22.3	5.0	67.7	4.3	0.7
	M	6.0	5.0	0.3	82.0	6.7
	O	9.0	6.7	13.0	3.0	68.3

F: freckles; L: lentigines; H: Hori's nevus;
M: melasma; O: nevus of Ota.

Although the proposed V-PLDA method has obtained better classification accuracy than the original PLDA method and several state-of-the-art image classification methods for the set of non-tumorous skin pigmentation images at hand, there is still room for improvement. Firstly, the available image set is relatively small and this may affect the performance of the classification methods. Therefore, we plan to acquire more images in the future. Secondly, the current image set was collected under a controlled environment, as presented in Section 3.3.1. To better assist patients and easily and reliably make a diagnosis of non-tumorous skin pigmentation disorders, the algorithm should be further improved to deal with images which are taken under a non-controlled environment. Finally, other features not related to images, such as patients' ages, could be incorporated to improve the classification accuracy of the proposed V-PLDA method.

3.4 Summary

This chapter addresses the first objective of the thesis, i.e., to develop a computerized method that can properly classify different non-tumorous skin pigmentation disorders. Noting that the large within-class variance of the image data can reduce the classification performance, we propose a voting based probabilistic linear discriminant analysis (V-PLDA) method based on the original PLDA model. In contrast to the original PLDA method, which is more suitable for small or moderate within-class variance, the proposed V-PLDA method alleviates the large within-class variance problem by using only those training samples that have high joint probabilities with the testing sample in the classification process. The selected training samples are utilized by a voting process to obtain the final classification decision. In contrast to the tied-PLDA method, the V-PLDA method does not require any prior information of the training samples.

It has been shown that the proposed V-PLDA method outperforms the original PLDA at different within-class variance levels in experiments using synthetic data. The V-PLDA method has also been applied to classify a real-world image set containing five types of non-tumorous skin pigmentation disorders: freckles, lentigines, Hori's nevus, melasma and nevus of Ota. In the experiment, different types of input features are examined and a combination of the HSV color feature and the Gabor texture feature is utilized. The experimental results have shown that the proposed V-PLDA method achieves significantly better performance than the original PLDA method and several state-of-the-art image classification methods when dealing with the given skin image set. We hope that this work will promote more research in helping these patients to diagnose their non-tumorous skin pigmentation problems and to avoid potential improper treatments by non-professionals.

Chapter 4

Melasma Severity Assessment via Image Analysis

In this chapter, the developed image analysis system for the severity assessment of a common skin pigmentation disorder, melasma, is introduced. A brief introduction to the advantages of the computerized melasma assessment to the clinical assessment is given in Section 4.1. The detailed steps of the proposed computerized assessment system is introduced in Section 4.2. In Section 4.3, the experimental results of the proposed system using the real-world melasma image data are presented and discussed. Finally, the chapter is summarized in Section 4.4.

4.1 Introduction

As reviewed in Section 2.2, there are different ways of evaluating the severity of melasma during the clinical treatment and the mMASI score system is one of the most widely accepted methods. While reliable, the mMASI system has its distinct disadvantages:

1. The evaluation can be inconsistent. In practice, evaluators assign mMASI scores with substantial variability. As MASI is based on a subjective visual assessment of the area of involvement and darkness, untrained evaluators may over- or under-estimate the surface area of involvement as well as how dark it is. This results in inter-assessor variability. Assessment of the chin, which makes up 10% of the mMASI score, is particularly difficult [22].
2. The scores cannot reflect the change of severity accurately when the area and darkness components of the score are grouped in bands. This is not ideal because although small changes across bands result in a significant change in scores, significant changes within each band will not be reflected, which results in a false negative error of unchanged score.
3. The score systems are complex and hard to use. The time-consuming nature and inherent lack of intuitive meaning compared with simple measures such as surface area of involvement make the system rarely used outside clinical trials.

The above disadvantages can be well circumvented by using the computerized methods because:

1. In the computerized methods, the evaluation performance depends on a trained mathematical model and the prior knowledge about the training

data. The professional knowledge can be stored in the trained model and the model is fixed for every input data. In this way, the input from the operator is simple and can be done by persons without any professional experience. Another advantage of such a model is that, the outputs are much less affected by the human input and can be more consistent.

2. The minimal change that can be detected by the computerized methods is as small as an image pixel. Different colors are represented by exact digital numbers in the images so that the changes identified can be represented more accurately than that of the band representations used in the clinical methods.
3. Although the development of the computerized methods requires professional knowledge, only small amount of human input is needed in using the developed system. The algorithm can be stored in the computers and easily used by non-professionals.

In considering the above advantages, the study of assessing the melasma severity using computerized methods is carried out. To the author's knowledge, this is the first study on such topic. The details of the developed algorithm are presented in this chapter.

4.2 Proposed Melasma Severity Assessment System

Before introducing the developed melasma severity assessment system, it is necessary to mention how the input image is recorded. In order to make the input images contain fewer illumination problems and also less affected by the changing environment, the VISIA Complexion Analysis system [95] with controlled light

illumination is used to take standardized images. One set of images taken from a patient are shown in Figure 4.1. There are three images for each patient taken from different angles for the purpose of showing different face regions clearly.

FIGURE 4.1: A set of images taken from a patient with the VISIA Complexion Analysis system: (a) Left view, (b) Frontal view and (c) Right view. Eyes are blocked to protect patient's privacy.

The images obtained are fed into the developed computerized system for the severity assessment. Basically, the proposed melasma severity assessment system can be divided into four steps: image enhancement, the area of interest extraction, image segmentation and severity score estimation. The details of each step are introduced in the next.

4.2.1 Image Enhancement

To better localize the skin pigmentation regions in the input images, it is necessary to enhance the input images. We propose a two-step image enhancement procedure for the melasma problem, which contains the procedures of illumination normalization and the combination of color and lightness information. The details are introduced in this subsection.

The illumination normalization process deals with the commonly seen non-uniform illumination problem with the images. In the melasma image data shown in Figure

4.1, it can be clearly observed that the lightness is strong due to the high reflectance of the surface in some parts of the image, while the lightness is weak in the other parts of the image. To tackle such illumination problem, different methods have been developed in the literature, for example, the grey-world algorithm [96], the max-RGB algorithm [97], the retinex algorithm [98], etc. After testing different color constancy methods, the homomorphic filtering method [99, 100] is selected for its good performance.

The homomorphic filtering method is a frequency domain approach that developed based on the illumination-reflectance model [101]. According to the model, it is assumed that the image lightness (L) of each pixel in the image is the product of the illumination of the scene (l) and the reflectance of the object(s) (r):

$$L(x, y) = l(x, y)r(x, y). \quad (4.1)$$

The purpose of the method is to remove the illumination component from the lightness. To achieve that, the two components, l and r , are separated by taking a logarithm transform of (4.1):

$$z(x, y) = \ln(L(x, y)) = \ln(l(x, y)) + \ln(r(x, y)). \quad (4.2)$$

Because the frequency of the illumination is normally lower than that of the reflectance, the illumination component can be removed by convolving the input lightness with a high pass filter h :

$$s(x, y) = h * z(x, y) = h * \ln(l(x, y)) + h * \ln(r(x, y)), \quad (4.3)$$

where ‘*’ indicates the convolution operation between two components. In this case, the convolved result $s(x, y)$ would only contain small amount of high frequency illumination information. The resulting lightness after the homomorphic

FIGURE 4.2: The flow chart of homomorphic filtering. ‘DFT’ and ‘ DFT^{-1} ’ represent the discrete Fourier transform and inverse discrete Fourier transform, respectively. ‘H’ is the Fourier transform of the high pass filter h .

filtering is:

$$L'(x, y) = \exp^{s(x,y)} . \quad (4.4)$$

The whole process of the above algorithm can be expressed using the flow chart shown in Figure 4.2.

One example of the lightness component of an input image and its output from homomorphic filtering is shown in Figure 4.3 (a) and 4.3 (b), respectively. It can be found that the lightness across the face region become more consistent after the filtering.

FIGURE 4.3: Grayscale images. (a) Lightness image before homomorphic filtering. (b) Lightness image after homomorphic filtering. (c) Input image combine both lightness and color information.

Besides the lightness information, color is another important indicator of the skin pigmentation regions, which has already been shown in the last chapter. In the mild cases, the pigmentation color cannot be well separated from the normal skin region, making it hard to make a reasonable segmentation using only the color

dimensions (a and b dimension). Therefore, we propose to combine the lightness with the color dimensions of the image to give an enhanced result for better segmentation. By observing the appearance of the image in each dimension and testing different combinations of these dimensions, we propose to obtain the enhanced input image I using (4.5). One example image I after the enhancement is shown in Figure 4.3 (c). It can be found that the contrast between the pigmentation region and the normal skin region gets larger after the enhancement compared with the corrected lightness image L' in Figure 4.3 (b).

$$I = L' - \sqrt{\frac{1}{2}(a^2 + b^2)} \quad (4.5)$$

4.2.2 Area of Interest Detection

Besides the illumination problem, there are two other problems in using the input images in Figure 4.1 to make the severity assessment based on the mMASI system. Firstly, the input images contain not only the face region of the patients but also the other unrelated regions such as the background of the scene, the hair of the patient. Only the face region contains useful information about the severity of the melasma pigmentation region and therefore it needs to be detected and separated from the non-face regions in the assessment. Secondly, as reviewed in Section 2.2.2, the face region is divided into four parts for the severity assessment in the mMASI score system. However, in the input images shown in Figure 4.1, the whole face region is recorded. To agree with the clinical mMASI assessment, it is necessary to split the input images into the same four parts. To solve the above two problems, a series of image processing techniques are utilized. The whole process is called the area of interest (AoI) detection in the proposed system and the details are introduced in this subsection.

To solve the first problem in the AoI detection, which is the face region detection, we propose to take advantage of the color difference method [102, 103]. Because the non-face regions are normally in the dark color, which is different from the face regions, we propose to calculate the color difference between each pixel and a base point. The calculated color difference values of the face region pixels and non-face region pixels should be at different levels. Then the face region can be differentiated from the non-face region by setting a threshold for the corresponding color difference values.

To calculate the color difference value, the *Lab* color space is used where L stands for lightness, a and b are the color-opponent dimensions. The color difference between a pixel at position (x, y) with the color of $(L(x, y) \ a(x, y) \ b(x, y))$ and a base point with the color of $(L_0 \ a_0 \ b_0)$ can be calculated as:

$$\Delta E(x, y) = \sqrt{(L(x, y) - L_0)^2 + (a(x, y) - a_0)^2 + (b(x, y) - b_0)^2}. \quad (4.6)$$

In the experiment, the base point is set at the origin $(0 \ 0 \ 0)$. With a proper threshold value (T_E) for the color difference (ΔE) selected, the corresponding pixel can be classified as a face pixel or a non-face pixel.

$$B_E(x, y) = \begin{cases} 0, & \Delta E(x, y) \leq T_E \\ 1, & \Delta E(x, y) > T_E \end{cases} \quad (4.7)$$

where $B_E(x, y)$ is the resultant classification label for a pixel at location (x, y) . Label 0 means the corresponding pixel is from the non-face region while label 1 means the pixel is from the face region. The value of T_E is the empirically selected.

The resulting binary images by applying the above color difference method on the input images in 4.1 are shown in Figure 4.4. It can be observed that the background pixels are detected and shown as black color (0). Besides the non-face region, some parts of the face region, such as the eyebrows and nostrils are also

detected at the same time. These regions are also not related to the pigmentation severity and can be excluded in the assessment process.

FIGURE 4.4: Example binary images B obtained from applying the face region detection with face region represented by white color and non-face region represented by black color. (a) Left view. (b) Front view. (c) Right view.

As shown in Figure 4.4, there are certain non-face parts, like the white clip on top of the example images, not detected using one threshold on the color difference value. These regions are finally excluded from the face region after multiple color difference selection processes.

To solve the second problem in the AoI detection, which is splitting the face region into four parts according to the mMASI system, we use images from different angles to extract different parts of the face. To be specific, the frontal view images are used to obtain the forehead and chin regions. To achieve that, the eyes, nose, and mouth of the patient are located in the image by utilizing the information in the binary output B_E from the last step. A brief introduction is as follows.

In the first step, the number of black pixels (with value of zero) in each line of B_E is counted:

$$N_E(x) = \sum_y 1 - B_E \quad (4.8)$$

Because the images taken are standardized and the positions of the eyes, nose and mouth of the patients are roughly in the same image region for different patients.

The line that pass through the eyes should have the maximum number of black pixels in the eyes region:

$$x_{eye} = \max_{x \in X_{eye}} N_E(x) \quad (4.9)$$

where x_{eye} is the horizontal coordinate of the eyes and X_{eye} is the predefined search region. Therefore, the forehead region can be obtained by cropping the image along the horizontal line $x = x_{eye}$ as shown in Figure 4.5.

Using the similar method, the positions of the nostrils and the mouth can be determined. The chin region can be cut out using a polynomial curve that cross the nostrils, the left most and the right most points of lips as shown in Figure 4.5.

FIGURE 4.5: The process of cutting frontal view image to extract the forehead and chin regions.

The same procedure is applied to the side view images to extract the malar regions. One example of processing the left view image is shown in Figure 4.6. The image is cut horizontally along the line that passing the eyes in the first stage to exclude the forehead region. Then a polynomial curve similar to that in the frontal view situation is defined. The image is then cut vertically along the line that passing the highest point of the polynomial curve.

By combining the above two procedures, the overall results of the AoI detection process are illustrated by an example image as shown Figure 4.7. It can be found

FIGURE 4.6: The process of cutting side view image to extract the malar region.

that each processed image would contain one part of the face region. These processed images contain only the face regions involved in the severity assessment and the grayscale image I of those images can be used for the subsequent analysis.

FIGURE 4.7: The original images (upper images) contain different face regions, forehead (left), left malar (middle) and chin (right), and the corresponding results after AoI detection process (lower images).

4.2.3 Hybrid Thresholding Based Image Segmentation

After the image preprocessing steps including image enhancement and area of interest extraction, the next step is to locate the skin pigmentation region in each

input image. Only after the pigmentation region is correctly identified, the area and darkness scores can be calculated. Different image segmentation methods have been proposed in the literature [104], which includes edge-based methods, region-based method, atlas-based method, thresholding, etc. Each type of the method would take advantage of certain properties of the region. In considering that the pigmentation regions do not have regular shapes and sizes, the atlas-based methods cannot work because no templates can be found. Another property of the pigmentation regions is that, the boundaries of the regions are normally diffused and the contrast between the foreground and background regions can be low. In such case, the pigmentation regions cannot be well segmented by the methods based on the edge or region information. After comparing different segmentation methods, it is found that the thresholding based image segmentation methods are suitable for the pigmentation problem. Thresholding is simple and can perform well consistently despite the variations in the pigmentation appearance.

As reviewed in 2.3, the local thresholding methods are more robust to the illumination variation in the image. However, it is problematic to select the window size of the local region in the processing because it requires the selected region contains enough amount of pixels from both classes to obtain an appropriate threshold value. If the pixels are all from one class, the algorithms would normally assume they are from the background, i.e. the normal skin region in our problem. Therefore, in the melasma image segmentation problem, the local thresholding methods can perform well when the melasma pigmentation region is relatively small compared with the local window size. On the other side, when the size of pigmentation region is larger than the local window size, the image would be heavily under-segmented.

Although this window size problem is addressed in different works such as the multi-scale thresholding method [77], these methods still cannot adapt to the situation of an extremely large foreground region, which is common in the melasma

data. On the other side, some research works shown that a combination of the local and global information can achieve better performance [105–107]. Based on this idea, we propose to utilize both the global and local information in our problem to tackle the window size problem. In particular, the global information is obtained by taking advantage of the human knowledge about the input image. This is because it is easy for a human observer to differentiate the melasma and the normal skin regions when there exist severe noises.

To get the global threshold, a small patch of the normal skin region in each input image can be selected to give prior information about the patient as shown in Figure 4.8 (a). Based on this selection, a global threshold t_g can be derived by calculating the mean input image value within the selected patch:

$$T_g = \frac{1}{|\mathbb{S}_n|} \sum_{(i,j) \in \mathbb{S}_n} I(i,j) \quad (4.10)$$

where $\mathbb{S}_n$ is the set containing coordinates of all the pixels in the selected normal skin region and $|\cdot|$ is the set cardinality.

The global threshold T_g and the local threshold T_l contain complementary information about the image: T_g gives the overview of the image and can identify the main part of the pigmentation patch while T_l adapts to the local variations in the image and fine tunes the segmentation result. Therefore, we propose to combine these two to form a hybrid threshold:

$$T_h(x,y) = \frac{1}{2}(T_l(x,y) + T_g), \quad (4.11)$$

where the local and global components are weighted equally in the current setup. I name this method the equal hybrid thresholding algorithm. Follow the style in [108], where the hybrid thresholding algorithm is named as HYTA in short. The short form of the proposed method is EHYTA. The image is then segmented based

on the proposed EHYTA method:

$$B_{pigment}(x, y) = \begin{cases} 0, & I(x, y) \leq T_h(x, y) \\ 1, & I(x, y) > T_h(x, y) \end{cases}. \quad (4.12)$$

One example of the resulting binary image $B_{pigment}$ is shown in Figure 4.8 (b). To have a better observation of the result, the boundaries of the pigmentation region are plotted on the original image as shown in Figure 4.8 (c).

FIGURE 4.8: One example to show the image thresholding process. (a) Input image with a patch of normal region selected. (b) The output binary image $B_{pigment}$ after the thresholding. (c) The detected boundaries plotted on the input image.

4.2.4 Score Estimation

In the previous subsection, the melasma pigmentation region is segmented out from the input image. Based on the segmentation result, the severity scores of the pigmentation region can be estimated. As reviewed in Section 2.2.2, the severity of melasma involves two aspects of the pigmentation in the mMASI system: area of involvement (A) and darkness (D). To agree with the clinical mMASI system, the two parameters are also estimated using the image information in the proposed computerized system. The details about the score estimation are presented in this subsection.

To determine the area of involvement of the pigmentation region, it is necessary to know the area of both the melasma pigmentation region (A_m) and the area of the face region involved (A_{face}). This is because the area score is defined as the area percentage of the pigmentation region to the face involved face region as shown in Table 2.1. To be specific, the area score is estimated as follows:

$$A^e = \frac{A_m}{A_{face}} \times 100\%, \quad (4.13)$$

where A^e denotes the area score calculated using the proposed computerized algorithm to differentiate from the clinical area score A . It can be found that different from the clinical area score, which represents a range of percentage values using one value, the score from proposed method A^e can change smoothly and therefore can be more accurate. This illustrates the advantage of the computerized severity assessment as mentioned previously.

In calculating the area of the melasma A_m and face region A_{face} , the bit quads method [109, 110] is adopted. The bit quad is defined as a 2×2 array of pixel cells. Each of the pixels has a state: on or off. According to the number of on cells and different patterns, the bit quads are classified into 6 types as shown in Figure 4.9. In the area calculation, each pixel is a part of four bit quads and the area of the pixel is the sum of all the four quads.

In determining the darkness of the pigmentation region, two facts need to be taken into consideration. Firstly, the darkness means not only the lightness difference between the pigmentation region and the normal skin region in the image but also the color difference between the two regions. Therefore, the color difference method that utilizes both lightness and color information is utilized to determine the darkness value. Secondly, the darkness contrast between the pigmentation and normal skin regions is a local information: people normally do not compare two regions far apart to determine the darkness level. Therefore, we propose to use the

FIGURE 4.9: Six different types of quads and the corresponding assigned area. The white color and black color represent on and off state of the pixel, respectively.

pixel information constrained by the local window used in the local thresholding method to calculate the darkness. Based on the above two ideas, the darkness score is estimated as follows:

$$D^e = \frac{1}{|\mathbb{S}_m|} \sum_{(x,y) \in \mathbb{S}_m} \sqrt{\Delta L(x,y)^2 + \Delta a(x,y)^2 + \Delta b(x,y)^2}, \quad (4.14)$$

where D^e is the estimated darkness value, $\mathbb{S}_m$ is the set containing coordinates of all detected melasma pixels. The difference in the lightness channel $\Delta L(x,y)$ is defined as:

$$\Delta L(x,y) = L(x,y) - \mu_L(x,y), \quad (4.15)$$

where $\mu_L(x,y)$ is the average lightness value of all the normal skin pixels in the local window centered at (x,y) . The same idea is applied to calculating $\Delta_a(x,y)$ and $\Delta_b(x,y)$ values.

With the melasma area and darkness scores of the four face regions estimated using (4.13) and (4.14), respectively, the overall estimated mMASI score can be calculated according to (2.16). We name this score as eMASI to differentiate from the clinical mMASI.

$$\text{eMASI} = 0.3A_f^e D_f^e + 0.3A_{lm}^e D_{lm}^e + 0.3A_{rm}^e D_{rm}^e + 0.1A_c^e D_c^e, \quad (4.16)$$

where the subscript labels f , lm , rm and c have the same meaning as those in (2.16).

The processing steps of the proposed melasma severity assessment system are summarized as follows:

1. Transform the input image into the *Lab* color space.
2. Apply homomorphic filtering on the lightness component of the image (4.2)-(4.4), combine the color information with the filtered lightness information to form the grayscale input I (4.5).
3. Detect the non-face region using the color difference method (4.6)-(4.7).
4. Divide the input images to get the four parts of the face defined in the mMASI system.
5. Segment out the pigmentation region using hybrid thresholding method (4.11)-(4.12).
6. Calculate the area and darkness score of each region (4.13)-(4.14).
7. Combine all the scores from the four different regions to get the final eMASI score (4.16).

4.3 Experimental Results

For validation, 29 patients with melasma were recruited (26 females and 3 males with a mean age of 48.6 years). All patients had Fitzpatrick skin types III-VI. For each time, three standardized images of the patient like the ones shown in Figure 4.1 are taken. Meanwhile, for each patient, two measurements at different time points, one is before a treatment (baseline) and the other is after the treatment (post-treatment), are taken. Two measurements at different time points are used because that one goal of melasma severity assessment is to evaluate the treatment impact on a patient by a comparison between the before and after treatment measurement scores. The scores from the baseline and post-treatment in mMASI and eMASI are denoted as $mMASI_b$ and $mMASI_p$, $eMASI_b$ and $eMASI_p$, respectively. Since there is no objective assessment rule, in order to obtain a convincing evaluation result, each image is separately examined by three dermatologists and an average score obtained from them is used as the benchmark mMASI score. The eMASI score is obtained by running the proposed severity assessment system once for each set of the input images.

4.3.1 Evaluation Methods

With the ground truth mMASI scores given by dermatologists and the estimated eMASI scores calculated by the proposed method, the method can be evaluated by comparing the two set of scores. In considering that only the trends of the scores, but not the detailed values are of the interest in the assessment, four specific evaluation indices are used.

The first evaluation index is the number of errors (N_t) in the changing trend estimation of each patient:

$$N_t = \sum_{i=1}^{29} E_i \quad (4.17)$$

where i is the index of each patient and E_i is the trend match between mMASI and eMASI. $E_i = 0$ means the mMASI and eMASI have the same changing trend while $E_i = 1$ means they have opposite trend.

$$E_i = \begin{cases} 0, & (\text{eMASI}_{p,i} - \text{eMASI}_{b,i})(\text{mMASI}_{p,i} - \text{mMASI}_{b,i}) \geq 0 \\ 1, & (\text{eMASI}_{p,i} - \text{eMASI}_{b,i})(\text{mMASI}_{p,i} - \text{mMASI}_{b,i}) < 0 \end{cases} \quad (4.18)$$

N_t is a crucial parameter because the changing trend directly evaluates the treatment method and therefore it determines the subsequent measure to be taken. The larger the N_t value, the more changing trend errors in the estimated eMASI scores and the worse performance of the method.

The other three statistical evaluation indices are Spearman's rank correlations calculated from baseline (S_b) results, post-treatment (S_p) results and difference (S_d) results. Spearman's correlation is defined as [111]:

$$\rho = 1 - \frac{6 \sum d_i^2}{N_s(N_s^2 - 1)}, \quad (4.19)$$

where N_s is the number of samples in experiments and d_i is the difference between the ranks of mMASI scores and eMASI scores. The rank correlation but not the value correlation is used coincides with the fact that the trend of the scores is the most crucial information in the assessment. To obtain the baseline correlation score S_b , we firstly sort both the baseline doctor generated scores (mMASI_b) and the computer generated scores (eMASI_b) in a descending order over all 29 image sets. The rank of mMASI_b and the rank of eMASI_b are recorded by R_{bm} and R_{bo} ,

respectively. The correlation index can be calculated as follows:

$$S_b = 1 - \frac{6 \sum_{i=1}^{29} (R_{bm,i} - R_{bo,i})^2}{29(29^2 - 1)}. \quad (4.20)$$

S_p is obtained in a similar way as S_b by ranks obtained from the post treatment sores mMASI_p and eMASI_p. S_d is calculated by using the ranks of score changes from the baseline to post-treatment: mMASI_p – mMASI_b and eMASI_p – eMASI_b.

4.3.2 Results and Discussions

An overall comparison between the clinical mMASI and the eMASI scores from the proposed method is shown in Table 4.1. It can be found that the mean clinical mMASI at baseline is 3.64 ± 2.45 , while the mean clinical mMASI post-treatment is 3.16 ± 2.26 . The mean change of clinical mMASI is -0.476 ± 0.87 . In comparison, the eMASI at baseline is 3.54 ± 1.31 . Post-treatment, the mean eMASI is 3.33 ± 1.42 . This corresponds to a mean change in eMASI of -0.21 ± 0.67 .

It can be found that the mean values between the mMASI and eMASI are at about the same level. The changing trends for the averaged mMASI and eMASI score are the same. These imply that the eMASI scores are reasonable estimations of the mMASI scores. The standard deviations of the eMASI scores are relatively smaller than those of mMASI scores, it is reasonable because the proposed method would likely to over-segment the extreme mild cases while under-segment the extremely severe cases. This can be observed from the minimum and maximum values in mMASI and eMASI: the minimum values in eMASI are larger than those from mMASI while the maximum eMASI values are smaller than those from mMASI.

The Spearman's rank correlation between the mMASI and eMASI scores are illustrated using the scatter plots. For the baseline images, Spearman's correlation

TABLE 4.1: Comparison of clinician mMASI scores and eMASI scores from proposed method

	mMASI			eMASI		
	Base	Post	Change	Base	Post	Change
Mean	3.64	3.16	-0.48	3.54	3.33	-0.21
Standard Deviation	2.45	2.26	0.87	1.31	1.42	0.67
Minimum	0.83	0.63	-2.4	1.96	1.62	-1.72
Maximum	9.9	10.8	1.1	8.05	9.5	1.45

coefficient $S_b = 0.735$, shows that automated mMASI correlate well with clinical mMASI ($p < 0.001$) (Figure 4.10 (a)). This is similar for imaging post-treatment ($S_p = 0.608$, $p < 0.001$) (Figure 4.10 (b)). The change in eMASI scores post treatment are also compared with the clinical change in mMASI scores and these also correlate well ($S_d = 0.622$, $p < 0.001$) (Figure 4.10 (c)). Among the 29 testing cases, the number of wrong changing trend inference N_t is 2, which only takes 6.9 percent of the whole data set.

To the author's best knowledge, this is the first study of using the computerized method in the severity assessment of melasma pigmentation. Therefore, no other similar systems in the literature can be compared. This pilot study provides a benchmark for the later developments on the topic of computerized melasma severity assessment.

The developed assessment system mitigates the shortcomings of the clinical assessment method. Firstly, the method is semi-automatic, the only manual step involves selection of skin unaffected by melasma prior to the running of the algorithm. This makes the results much less dependent on the experience of the assessors, making the assessment more consistent. Secondly, the scores obtained changes smoothly and are not grouped in bands as in the clinical assessment methods. It makes the assessment more accurate in detecting all changes, whether large or small, of the pigmentation. Finally, the algorithm is simple and easy to use. A simple executable software is developed based on the proposed assessment system

FIGURE 4.10: Scatter plots showing the relationship between clinical mMASI and estimated eMASI scores. (a) Baseline. (b) Post-treatment. (c) Changes between baseline and post-treatment.

and the user interface is shown in Figure 4.11. It can be easily run on computers with simple setups. There is no need for training and the operator only needs to specify the location of the input files and select the normal skin region using the mouse. The software will generate consistent eMASI scores that can enable clinicians around the globe to standardize melasma severity scoring and outcome measures in an easy and reproducible manner.

While useful, this proposed algorithm has the following weaknesses. Firstly, only a small size of data set is used in the analysis and algorithm development. Besides, most of the patients have mild melasma which reflects the melasma severity of the majority patients seen in the center that the data is collected. The algorithm needs to be validated in a larger prospective cohort, especially those with moderate to severe melasma conditions. Secondly, the input images need to be

FIGURE 4.11: User interface of the melasma assessment software based on the proposed assessment method.

taken under a controlled environment using specialized equipment. This limits the application of the algorithm. Thirdly, although the preprocessing of the input images mitigates the illumination variation, the problem is not totally eliminated. It is still hard to segment the pigmentation region accurately when the illumination problem is severe. For example, the forehead and chin regions contain more illumination problems than those of malar regions, the segmentation of these two regions is less accurate. Besides the illumination problem, the proposed hybrid thresholding method is still not adaptive enough to the dataset. This is because the trade-off parameter is a fixed value, which cannot change according to different situations. This problem is studied and an optimal hybrid thresholding method is developed. More detailed discussion about this trade-off parameter value selection are presented in Chapter 5.

Another thing needs to be noted is, the algorithm assesses the severity of melasma pigmentation regions in the image but it does not differentiate melasma from other types of pigmentation. It needs the input images classified as melasma before they

can be processed. The classification process can be done by either a doctor or a classification algorithm like the one introduced in Chapter 3.

4.4 Summary

In this chapter, the second objective of the thesis, i.e., the severity assessment of skin pigmentation disorder using computerized method is addressed. In particular, we focused on a widely spread disease, melasma, in the study. We have developed and validated a novel algorithm for an automated estimation of the modified melasma area and severity index (mMASI). Different techniques have been utilized in the proposed algorithm to obtain the severity score from the input images. The main ones include, a threshold method based on the color difference to detect the skin region in the image, a method that combines both the lightness and color information to enhance the contrast between the foreground and the background in the melasma image, an equal hybrid thresholding method that segments the melasma pigmentation region in the image, and a darkness estimation method based on the average local color difference. It is shown that the output score, eMASI, correlates closely with the dermatologist-scored mMASI. eMASI has also been shown to be sensitive enough to pick up changes in MASI in response to treatment. The proposed algorithm is the first work that using computerized methods to estimate the mMASI score. It provides a benchmark for the further developments in this topic.

Chapter 5

Optimal Hybrid Thresholding for Melasma Severity Assessment

In this chapter, an optimal hybrid thresholding algorithm (OHYTA) that can optimally determine the trade-off between the global and the local information in the hybrid thresholding method is proposed. The method is applied to solving the melasma severity assessment problem. In Section 5.1, the motivation and the basic principle of the proposed method are briefly explained. Section 5.2 and 5.3 elaborate the OHYTA method with different realizations. Specifically, Section 5.2 introduces OHYTA with Euclidean distance used while Section 5.3 studies OHYTA based on a generalized distance measure. In Section 5.4, the experimental results from the tests on real-world melasma image data are presented and discussed. Finally, the chapter is summarized in Section 5.5.

5.1 Introduction

A computerized melasma severity assessment system has been presented in Chapter 4. It provides a benchmark on the topic of skin pigmentation severity assessment. Despite the developed eMASI scoring representing a promising, novel and reliable method for the assessment of disease severity, there are still many aspects that can be further investigated to improve the performance, such as the illumination normalization, image segmentation, and the score estimation. Within the several aspects listed, the image segmentation process is the most critical step in the assessment because it directly determines the accuracy of the estimated severity scores. In the method, the equal hybrid thresholding algorithm (EHYTA) is used for the segmentation where the trade-off between the global and local information is empirically determined as a fixed value. The fixed trade-off parameter cannot adapt well to the real-world situation due to the large appearance variation of the pigmentation regions. Therefore, in this chapter, we focus on the problem of ‘what is the optimal trade-off between the global and local information in segmenting the skin pigmentation regions?’ and propose an optimal hybrid thresholding method with different distance measures utilized. The proposed method has been validated with both synthetic and real-world skin pigmentation images and has shown its superior performance to the state-of-the-art thresholding based image segmentation methods.

To illustrate the effect of appearance difference on the thresholding methods, a set of synthetic images is created and shown in the first row of Figure 5.1. Specifically, in each image, the foreground is composed of a relatively small disk located at the bottom-left side and a relatively large disk located at the top-right side. Among all these images, the sizes of small disks are the same while the sizes of large disks are different. It is also seen that there are illumination variations in all these images: the left part of the images are brighter than the right part of

FIGURE 5.1: Thresholding segmentations on synthetic images.

them. The goal of using different disks combined with illumination variations is to demonstrate various relationships between the image foreground and background. The second row of Figure 5.1 contains the "ground truth" of segmentations while the following rows contain the segmentation results of different methods. Therefore, these methods can be evaluated by comparing their results with the ground truth images.

The 'mAdOtsu' and 'mSauvola' are the extensions of the adaptive Otsu method and the Sauvola's thresholding method, respectively, which reviewed in Section 2.3. By incorporating with the multi-scale method proposed in [77], the 'mAdOtsu' and

‘mSauvola’ can have a better adaption to the scale change of the object. ‘EHYTA’ is the thresholding method used in the melasma severity assessment system that introduced in Section 4.2. It can be observed from the results that, mSauvola considers only local thresholds and its segmentation output deviates from the ground truth with the increment of the large disk size. That is, if the size of foreground object exceeds the scope of a local region controlled by a window size parameter in mSauvola, the segmentation would fail because local thresholds determined in mSauvola cannot utilize information beyond their local regions. In contrast, the mAdOtsu method considers not only local information but also the global information of the image. In this way, it can have a bigger view of the image, and perform better than mSauvola when the foreground region is large. However, the mAdOtsu method tends to over-segment the image when the foreground size is small. The EHYTA method has more balanced results between the small and large foreground situations by equally combining the local and global thresholds, however, the trade-off between the local and global information is not the optimum, resulting in obvious over-segmentation when the foreground region is large as shown. These synthetic images based experimental results indicate that the local thresholds and the global threshold are required to be properly treated to get a satisfied segmentation result.

In combining the local and global information in the thresholding methods, there are generally two parameters to consider: one is a window size parameter controlling the scope of a local region around each pixel to extract image local information, and the other is a tradeoff parameter compromising local region information and global image information. The window size parameter is application specific where prior knowledge is generally required. Some previous studies about the window size problem can be found in [112] and this problem is not discussed in this work. As mentioned, our focus is on finding an optimal combination of the local region information and the global image information in the thresholding process.

The details of the developed method and the experiments done for validation are presented in this chapter.

5.2 Optimal Hybrid Thresholding Based on Euclidean Distance

In Figure 5.1, it is observed that the EHYTA method works well when the larger disk size is small (for the 2 most left images). Since EHYTA equally combines each local threshold and a global threshold, it implies that a trade-off between each local threshold $T_l(x, y)$ and the global threshold T_g , instead of treating them equally, is necessary to match the various relationships between the image foreground and background. Based on this observation, a generalization of the equally hybrid thresholding method is now proposed. This proposal adds a trade-off parameter $\lambda \in [0, 1]$ to compromise each local threshold $T_l(x, y)$ and the global threshold T_g :

$$T_{wh}(x, y, \lambda) = (1 - \lambda)T_l(x, y) + \lambda T_g \quad (5.1)$$

The global threshold T_g and the local thresholds $T_l(x, y)$ in 5.1 are required to be calculated to obtain the hybrid output $T_{wh}(x, y, \lambda)$. Like the one used in EHYTA, calculation of T_g relies on prior knowledge provided by an application. For example, a patch of image background can be selected to obtain T_g by intensity averaging over the selected patch: $T_g = 1/N_s \sum_{(x,y) \in \mathbb{S}_n} I(x, y)$, where $\mathbb{S}_n$ denotes the set of pixels in the selected patch and N_s denotes the size of this set. For calculation of the local thresholds $T_l(x, y)$, the Sauvola's local threshold determination method is adopted.

With the global threshold T_g and the local thresholds $T_l(x, y)$, hybrid threshold

$T_{wh}(x, y, \lambda)$ is calculated from (5.1) for each given parameter λ . Now, it is interesting to discuss the impact of the parameter λ on the output hybrid threshold. Particularly when $\lambda = 0$, the output of (5.1) is $T_{wh}(x, y, \lambda) = T_l(x, y)$ and this resultant hybrid threshold is identical to the local threshold of each pixel. Consequently, this hybrid threshold considers only the local information around a pixel as the case in Sauvola's method. For the case of $\lambda = 1$, (5.1) becomes $T_{wh}(x, y, \lambda) = T_g$ and this output hybrid threshold is identical to the global threshold T_g . For the general case that $0 < \lambda < 1$, each output $T_{wh}(x, y, \lambda)$ is a weighted sum of the global threshold and one local threshold. Particularly, EHYTA is a special case of equation (5.1) when $\lambda = 0.5$ which corresponds to equal treatment of the local threshold and global threshold. In summary, λ compromises the importance of the local and global thresholds. As a result, both local and global information is incorporated into the hybrid threshold $T_{wh}(x, y, \lambda)$.

To get a better picture of the impact of the parameter λ on segmentation results, experiments are implemented by setting λ at several values and the experimental results are shown in Figure 5.2. From this figure, it is clearly seen that the segmented objects at $\lambda = 0.2$ in the image are comparable to the ground truth (Figure 5.2 (b)). With λ deviating from 0.2, it is observed that the segmented objects are deformed from the ground truth. This figure shows that segmentation performance is changing with respect to λ which implicates a proper setting of λ is necessary.

From the above discussions, we notice that there is a need to find a criterion for the determination of λ . In (5.1), it is noted that the threshold $T_{wh}(x, y, \lambda)$ is a linear combination of $T_l(x, y)$ and T_g . Noting that the final goal of an obtained threshold $T_{wh}(x, y, \lambda)$ is to segment a given image, an optimally selected threshold should lead to a segmentation result that well represents the given image. Therefore, we propose to use a deviation between the segmented outputs and the given image to evaluate the effectiveness of the obtained threshold $T_{wh}(x, y, \lambda)$ and the parameter

FIGURE 5.2: Weighted hybrid thresholding segmentation with respect to λ . (a) Original image, (b) ground truth, (c) $\lambda = 0$, (d) $\lambda = 0.2$, (e) $\lambda = 0.4$, (f) $\lambda = 0.6$, (g) $\lambda = 0.8$, (h) $\lambda = 1$

λ . When the Euclidean distance is adopted, the deviation is formulated as:

$$\begin{aligned}\delta(\lambda) &= \sum_{i=1}^M \sum_{j=1}^N (I(i, j) - T_{wh}(i, j, \lambda))^2 \\ &= \sum_{i=1}^M \sum_{j=1}^N [I(i, j) - (1 - \lambda)T_l(i, j) - \lambda T_g]^2,\end{aligned}\quad (5.2)$$

where (M, N) is the size of a given image. Since $\delta(\lambda)$ is convex with respect to λ , there exists a unique solution to the optimization problem:

$$\lambda_{opt} = \arg \min_{\lambda \in [0, 1]} \delta(\lambda). \quad (5.3)$$

Specifically, λ_{opt} can be obtained by setting the first derivative of $\delta(\lambda)$ to zero and we get:

$$\nabla_{\lambda} \delta(\lambda) = 2 \sum_{i=1}^M \sum_{j=1}^N [I(i, j) - (1 - \lambda)T_l(i, j) - \lambda T_g](T_l(i, j) - T_g) = 0 \quad (5.4)$$

By a straightforward calculation, the optimal parameter λ is obtained:

$$\lambda_{opt} = -\frac{\sum_{i=1}^M \sum_{j=1}^N (I(i, j) - T_l(i, j))(T_l(i, j) - T_g)}{\sum_{i=1}^M \sum_{j=1}^N (T_l(i, j) - T_g)^2}. \quad (5.5)$$

With this λ_{opt} , an optimal weighted hybrid threshold is obtained from (5.1) and denoted by $T_{oh}(x, y, \lambda_{opt})$.

Remark: Although there are thresholding determination methods compromising between local and global information for segmentation, to the best of the author's knowledge, there has been no rule for optimal trade-off parameter determination. The analytic trade-off parameter solution given in (5.5) distinguishes our proposed method from the multi-scale methods [77] as their threshold determinations can hardly be analytically represented.

Algorithm 3 Determination of optimal weighted hybrid threshold $T_{oh}(x, y, \lambda_{opt})$

Require: $I(x, y)$

Ensure: $T_{oh}(x, y, \lambda_{opt})$

Initialization

Select local window size, Set $Neu_s = 0$, $Den_s = 0$

Get T_g

for $i = 1$ to M **do**

for $j = 1$ to N **do**

Get $T_l(i, j)$

$Neu(i, j) = (I(i, j) - T_l(i, j))(T_l(i, j) - T_g)$

$Den(i, j) = (T_l(i, j) - T_g)^2$

$Neu_s = Neu_s + Neu(i, j)$

$Den_s = Den_s + Den(i, j)$

end for

end for

$\lambda_{opt} = \frac{Neu_s}{Den_s}$

return $T_{oh}(x, y, \lambda_{opt}) = (1 - \lambda_{opt})T_l(x, y) + \lambda_{opt}T_g$

The thresholding determination procedure for the optimal hybrid threshold $T_{oh}(x, y, \lambda_{opt})$ is given in Algorithm 3. With the obtained optimal trade-off parameter λ_{opt} and threshold $T_{oh}(x, y, \lambda_{opt})$, an optimal hybrid thresholding algorithm (denoted as OHYTA) is subsequently developed.

In Figure 5.1, synthetic images are used to evaluate several segmentation methods. For comparison, those synthetic images are adopted to evaluate the proposed OHYTA method and the segmentation results are shown in the bottom row of Figure 5.1. By comparing the segmentation results of OHYTA with results of other methods in the figure, it is observed that the proposed OHYTA method achieves visually better segmentation outputs than the compared methods.

To statistically evaluate the segmentation results of our proposed method, we use several statistics for performance evaluation: Precision (Pr), Recall (Re) and F-measure (F). These statistics are defined as:

$$Pr = \frac{TP}{TP + FP}; \quad Re = \frac{TP}{TP + FN}; \quad F = 2 \frac{Pr \times Re}{Pr + Re}$$

where TP, FP, FN represents true positive, false positive and false negative, respectively. TP is the number of correctly detected foreground pixels, FP is the number of background pixels that are wrongly detected as foreground pixels and FN is the number of foreground pixels that are wrongly detected as background pixels.

Evaluation results upon the segmentation outputs based on the synthetic images are shown in Figure 5.3. In this figure, it is observed that the proposed thresholding method attains the most statistically stable and overall better results than the other competing thresholding methods. Specifically, both Precision and F-measure of OHYTA approach the best among the results of all compared methods with the increment of disk size. Although OHYTA takes the second rank in view of Recall scores shown in Figure 5.3(b) where mAdOtsu takes the first rank, it is noted that the Precision and F-measure scores of mAdOtsu varies a lot.

FIGURE 5.3: Statistical evaluation of different methods using synthetic data. (a) Precision. (b) Recall. (c) F-measure.

5.3 Optimal Hybrid Thresholding Based on Generalized Distance Measure

The proposed optimal hybrid thresholding algorithm tries to find the optimal trade-off value by minimizing the distance between the threshold T and the input intensity I . In Section 5.2, the square Euclidean distance is used for the distance measurement for its simplicity. However, it is not well studied what would the performances be when other wide varieties of distance measures are used. Therefore, in order to optimize the performance within a wide range of distance measure

surface, we propose to generalize the measure as follows:

$$E(T, I, \alpha) = (T^\alpha - I^\alpha)^2, \quad (5.6)$$

where α is a positive parameter that controls the model. With a different selection of the α value, the measure would change its focus from region to region. It can be classified into 3 different cases that $\alpha < 1$, $\alpha = 1$, $\alpha > 1$ and the corresponding contour plots are shown in Figure 5.4.

FIGURE 5.4: Contour plots of generalized distance measure $E(T, I, \alpha)$ with different α values.

- $\alpha = 1$: the same with square Euclidean measure where the contours are parallel with $I = T$ line;
- $\alpha < 1$: the contours are dense in the small inputs region while sparse in the large inputs region. One special case of $\alpha = 0.5$ is called the squared-chord distance proposed in [113];
- $\alpha > 1$: the contours are dense in the large inputs region while sparse in the small inputs region.

In Figure 5.4, the spacing between the contours indicates the changing speed of the function. The denser the contours are, the faster changing speed of the distance value has, which indicates the more sensitive of the measure to the input

Algorithm 4 Improved image segmentation method using the generalized distance measure

Input: A gray level image I
Output: The thresholded binary image $B(\alpha)$ of I
Initialization
Initialize $t_g, T_l, \delta_\lambda, \delta_\alpha, \alpha_{min}, \alpha_{max}$
 $\lambda \leftarrow 0, \alpha \leftarrow \alpha_{min}$
while $\alpha < \alpha_{max}$ **do**
 Get $T(\lambda)$ by (5.1)
 while $\lambda < 1$ **do**
 Get $E(\lambda, \alpha) = \sum_{(x,y)} (I(x, y)^\alpha - T(x, y, \lambda)^\alpha)^2$
 $\lambda = \lambda + \delta_\lambda$
 end while
 $\alpha = \alpha + \delta_\alpha$
 Get $\lambda(\alpha)_{opt}$ by (5.7)
end while
return $B(\alpha) = I > T(\lambda(\alpha)_{opt})$

value change. Therefore, in the situation of $\alpha < 1$, the proposed measure is less sensitive to the input change in the high value region. In contrast, the method is less sensitive to the input change in the low value region when $\alpha > 1$. For $\alpha = 1$, the distance contours are parallel with the $I = T$ line which means that the sensitivities of the distance measure to the change of I and T values remain the same in the direction of $I = T$.

Based on the proposed distance measurement, the OHYTA segmentation method can be expressed as:

$$\lambda(\alpha)_{opt} = \arg \min_{\lambda} \sum_{(x,y)} (T(x, y, \lambda)^\alpha - I(x, y)^\alpha)^2. \quad (5.7)$$

In the above proposed improved segmentation method, the optimal trade-off parameter value λ_{opt} would depend on the value of α . Because the optimization problem (5.7) cannot be analytically solved except for $\alpha = 1$, we use the grid search method to select the λ_{opt} value of each image. To show the impact of different α value and find the optimal value for the melasma problem, the results of different α are obtained. The whole process of proposed image segmentation with generalized

FIGURE 5.5: Statistical evaluation of the α value selection using synthetic data. (a) Precision, (b) Recall, (c) F-measure.

distance measure for different modal parameter α is summarized in Algorithm 4.

In the proposed algorithm, δ_λ and δ_α are the step sizes of λ and α , respectively. α_{min} and α_{max} are the user defined lower and upper bounds of α . For each image to be segmented, an optimal trade-off parameter $\lambda(\alpha)_{opt}$ is selected using (5.7).

The α value would have different effects on the images with different sizes of foreground regions. When the foreground region is small, the major part of the image would consist high valued background pixels. Therefore, to make the large background region identified as a whole, the distance measure should not sensitive to the change in the high value region, which corresponds to a small α value. The same applies to the situation of a large foreground region, where a large α value can obtain a better segmentation result. This can be illustrated by applying the

proposed method on the synthetic images as shown in Figure 5.1. Evaluation results upon the segmentation outputs of using different α values are shown in Figure 5.5. It can be found that a smaller α value tends to obtain better performance in terms of F-measure when the foreground region (disk size) is small. On the contrary, the large foreground region situation requires a large α value to obtain satisfactory results.

Besides the effect on foreground-background ratios of the images, another important function of the α value in the proposed distance measure is the adaptability to the noise. In the situation that the input image contains random noise, the segmentation result would be biased from the ground truth. It is easy to understand that, for the same level of noise, the region with low image intensity would be more affected. This is because the change brought by noises would take higher ratios to the low intensities than those to the high intensities. This difference between the low intensity and high intensity regions can be addressed by the proposed distance measure. As shown in Figure 5.4, a larger α value can be taken if we want to decrease the influence of the intensity change in the low value region to the final result. Therefore, with the increment of noise level, a larger α value need to be used to obtain more reasonable segmentation results. To justify this, experiments are done on the synthetic images containing different levels of Gaussian noise. The results are shown in Figure 5.6.

In the experiment, the averaged F-measure of the five different synthetic images is used to have an overall assessment of the method. It can be observed from Figure 5.6 that, when the noise level is low, i.e., the signal to noise ratio (SNR) is high, a smaller α value can obtain better performance in the tested range and vice versa. A more detailed study is done to find the optimal α value at different levels of noise in the current experimental setup. The results are shown in Table 5.1. These results support our previous analyses on the α value selection in the situation of noises exist. The method is tested using real-world image data in the next.

FIGURE 5.6: Statistical evaluation of the α value selection using synthetic data with different levels of Gaussian noise.

TABLE 5.1: The α values that obtain the best performances in terms of mean F-measure at different noise levels.

SNR(dB)	60	55	50	45	40
α	0.8	0.9	1.3	1.4	1.5

5.4 Experiments on Melasma Images

The proposed optimal hybrid thresholding algorithm (OHYTA) with different distance measures used are evaluated on the real-world melasma images. The same melasma image data set used in Chapter 4 is utilized. Besides the image segmentation process, the other processing steps like the illumination normalization of the input images and the score estimation of the segmented regions are kept the same with those introduced in Chapter 4.

5.4.1 Evaluation of OHYTA Based on Euclidean Distance

Different thresholding based segmentation methods including multi-scale Sauvola (mSauvola), multi-scale adaptive Otsu (mAdOtsu) and equal hybrid thresholding

FIGURE 5.7: Melasma segmentation results using several thresholding methods. (a) Input image with normal skin region selected. (b)-(e) Segmentation results of mSauvola, mAdOtsu, EHYTA and OHYTA-Eu, respectively. The ellipses on the image highlight the over- or under-segmented regions.

algorithm (EHYTA). For the given input image shown in Figure 5.7 (a), its segmentation outputs by using these competing methods are shown in Figure 5.7, where the highlighted yellow-block region in (a) is selected as normal skin region to obtain the global threshold T_g .

It is seen that the input image shows in Figure 5.7 (a) is in a severe melasma condition because there is one large melasma area together with many small lesion patches. Compared with the input image, there are under-segmented parts within the large melasma area where there are many discontinuities in both the segmentation outputs of mSauvola and mAdOtsu (regions highlighted by cyan ellipses in Figure 5.7 (b) and Figure 5.7 (c)). Under-segmentation in these two methods is in line with the fact that when the foreground size is larger than the maximum local window size, these two local thresholding segmentation methods

fail to extract desired thresholds. Consequently, the foreground pixels could be mis-segmented in such cases. Moreover, mAdOtsu has a problem in dealing with background objects with illumination variations. This problem can be observed in Figure 5.7 (c) where there are many background pixels around the face boundaries mis-segmented as foreground pixels (the region in the yellow ellipse in Figure 5.7 (c)). On the contrary, the segmentation output of EHYTA is shown in Figure 5.7 (d) and it is observed that this output is likely to be over-segmented by segmenting normal skin pixels as melasma patterns (the region in the yellow ellipse in Figure 5.7 (d)). As a comparison, the output by using the proposed OHYTA-Eu (see Figure 5.7 (e)) seems reasonable since it is clearly seen that this output compromises between the under-segmented case and the over-segmented case.

Despite the above comparisons based on the segmentation results, the same method for evaluating the accuracy of the melasma severity assessment results is used. As introduced in Section 4.3.1, the number of trend error and correlation scores of the baseline, post-treatment, and the difference are calculated. Moreover, we use $\tau = \frac{S_b + S_p + S_d}{N_t}$ as an overall performance index, which takes into account all the 4 indexes, S_b , S_p , S_d and N_t . The statistical evaluation results are shown in Table 5.2.

TABLE 5.2: Comparison of thresholding methods for melasma severity assessment

	N_t	S_b	S_p	S_d	τ
mSauvola	7	0.743	0.719	0.603	0.295
mAdOtsu	8	0.656	0.589	0.594	0.230
EHYTA	2	0.735	0.608	0.622	0.983
OHYTA	2	0.869	0.651	0.667	1.093

Boldface indicates the best score among the evaluated methods.

From Table 5.2, it can be observed that the performance of mAdOtsu is the worst. This result is in line with our previous experimental results on synthetic images

since mAdOtsu cannot adapt to illumination variations. mSauvola achieves relatively higher correlation scores. However, its number of the error of changes is 7 which is large among the compared results and is unacceptable. An error of change detection means a wrong evaluation of a therapy treatment. One reason for the large change detection error number may be that mSauvola tends to underestimate the severe cases. EHYTA is a particular case of the weighted hybrid method as mentioned earlier. The proposed OHYTA method changes the weight adaptively when dealing with different inputs. Hence, the OHYTA method has a more accurate thresholding surface than the EHYTA method and thus obtains better performance. Particularly, OHYTA attains the best τ -score among all the competing methods. To give a better picture of the statistical evaluation results, scatter plots of the baseline scores, post-treatment scores, and change scores between the benchmark mMASI and the computerized scores obtained from OHYTA-Eu are shown in Figure 5.8 (a), Figure 5.8 (b) and Figure 5.8 (c), respectively.

5.4.2 Evaluation of OHYTA Based on Generalized Distance

To further evaluate the proposed OHYTA method in the situation where the generalized distance measure introduced in Section 5.3 is applied. The experimental results of using different α values are listed in Table 5.3. It can be shown that either a very large or very small α value would result in poor accuracy. The correlation indices would be smaller and more importantly, the number of trend change error (N_t) increases. This is because in such cases, the sensitivity of the distance measure varies dramatically from low to high values and thus would not uniformly reflect the change of severity.

Another finding is that the best performance is not obtained at the situation that $\alpha = 1$. In this experiment, the optimal solution is achieved at $\alpha = 1.2$. This

FIGURE 5.8: Scatter plots showing the relationship between clinical mMASI scores and estimated oMASI scores using the OHYTA-Eu method. (a) Baseline. (b) Post-treatment. (c) Changes between baseline and post-treatment.

is because when $\alpha = 1$, the sensitivity of the distance remains the same in the direction of $I = T$. This works perfectly well if the input image is noise free. However, in the real situation, the images would be contaminated by a certain amount of noise due to the illumination variation and the image capture process. The optimal solution would be changed in such case.

For the same level of noise, small T and I values are affected more in comparison with large T and I values because smaller image values lead to smaller signal to noise ratio. To balance this effect and obtain better performance, the distance measure should be made more insensitive to noises when T and I values are small, which corresponds to a larger α value. This explains why the optimal α value should be larger than 1. Nevertheless, the increase of α cannot be unbounded

TABLE 5.3: Results of proposed method with different α values (with the best values highlighted)

α	S_b	S_p	S_d	N_t	τ
0.5	0.8177	0.6269	0.6353	4	0.5200
0.6	0.8194	0.6244	0.6383	4	0.5205
0.7	0.8369	0.6373	0.6324	4	0.5266
0.8	0.8495	0.6427	0.6519	4	0.5360
0.9	0.8633	0.6476	0.6447	2	1.0778
1.0	0.8584	0.6552	0.6578	2	1.0857
1.1	0.8682	0.6520	0.6531	2	1.0867
1.2	0.8636	0.6410	0.6467	1	2.1512
1.3	0.8530	0.6538	0.6215	4	0.5321
1.4	0.8463	0.6520	0.6164	4	0.5287
1.5	0.8497	0.6688	0.6092	4	0.5319

or the effect of the measure would be excessive. This can be observed in Table 5.3 that the best performance is obtained at $\alpha = 1.2$ but not larger value. To conclude, each α value would correspond to a certain level of noise to achieve optimal segmentation. The higher level of noise is, the larger optimal α should be.

It is noted that the proposed melasma severity assessment method would consume more time than the one in Chapter 4. This is because that it needs to calculate or search for the optimal trade-off parameter. However, the difference is not significant as the computation complexity is not very high. Besides, the system is working off-line and there are no strict requirements for the time consumed.

5.5 Summary

In this chapter, the problem of melasma severity assessment using computerized method is further investigated based on the methods presented in Chapter 4. The hybrid thresholding method for the melasma image segmentation is improved by optimally determining the trade-off parameter between local and global information. This optimization is realized by minimizing the distance between the

threshold surface and the input image. As a result, the trade-off parameter is no longer a fixed value but can change with various situations of the input images, leading to a better adaption to the image background change. The optimization is realized firstly based on the square Euclidean distance. A more generalized distance measure is subsequently introduced to achieve better performance. The proposed methods are first evaluated using synthetic images. It is shown that they can achieve more accurate segmentation results compared with other state-of-the-art image thresholding methods. Then, the methods are evaluated using the real-world melasma image data to estimate the clinical mMASI score. The statistical analysis of the results shows that the proposed method can achieve better overall performance than the previously developed mMASI score estimation system.

Chapter 6

Conclusion

The works presented in the thesis are briefly summarized in Section 6.1, which include three parts, (1) the classification of five non-tumorous skin pigmentation disorders, (2) the severity assessment of melasma, and (3) the optimal hybrid thresholding to improve the severity assessment method in (2). To make further improvement, some possible future works are pointed out in Section 6.2.

6.1 Thesis Summary

In this thesis, the study on developing computerized methods for the diagnosis and severity assessment of different non-tumorous skin pigmentation disorders has been presented. This study is motivated by the fact that the skin pigmentation disorders can bring huge negative effect on the quality of life and the demand for a proper diagnosis and treatment of the diseases is steadily increasing. Different from the large number of research studies on the tumorous skin pigmentation disorders, the work presented in this thesis deals with the non-tumorous skin pigmentation disorders. To be specific, the thesis focuses on the classification of five non-tumorous skin pigmentation disorders and the severity assessment of melasma, which have not been studied in the literature. The thesis includes three main parts and the details are summarized as follows:

(1) The thesis has developed an algorithm to classify five non-tumorous skin pigmentation disorders based on the images of the patients.

This part of the thesis aims to provide the patients a tool to diagnose their non-tumorous skin pigmentation problems and avoid potential improper treatments by non-professionals. In particular, five commonly seen non-tumorous skin pigmentation disorders including freckles, lentigines, melasma, Hori's nevus and nevus of Ota are involved in the study. Considering the large within-class variance of the non-tumorous skin pigmentation data degrades the performance of the image classification algorithms, a voting based probabilistic linear discriminant analysis (V-PLDA) method is proposed in Chapter 3. It is shown that in the situation of large within-class variance, the within-class covariance matrix can be approximated by its diagonal elements. Consequently, the PLDA method can be approximated by an individual matching process, which means the testing sample would be classified to the same class with the sample that has the largest joint probability with it.

It is shown that the individual matching approximated PLDA is sensitive to noises and outliers. We propose a voting procedure to incorporate with the individual matching process to make the classification more robust. It is illustrated using the synthetic data that the proposed V-PLDA method can consistently outperform the PLDA under different setup. It is also shown that the subclass distribution of the data is the factor that influences the performance of the proposed V-PLDA method.

The experiments on the real-world non-tumorous skin pigmentation image data are carried out to evaluate the proposed V-PLDA method. Different input features based on the clinical diagnose rules are examined and tuned. A combination of the color and non-color feature is used as input in the classification process. The results have shown that the proposed V-PLDA method can obtain significantly higher classification accuracy than the original PLDA method as well as several state-of-the-art image classification methods. In addition, the proposed V-PLDA method can serve as a general classification method, which has a good adaptability to the large within-class variance. Therefore, the method can be applied to solving other data classification problems.

(2) The thesis has developed a computerized system to assess the severity of melasma based on the images of the patients.

This part of the thesis aims to develop a computerized system that can assess the melasma pigmentation severity based on the standardized images taken from the patient. Similar to the classification problem, the clinical rules serve as the gold standards that guide the development of the computerized algorithm. The clinical melasma severity assessment score from the modified melasma area and severity index (mMASI) is estimated using the proposed algorithm.

To assess the severity of the melasma pigmentation in the image, different image processing techniques are utilized in a row to get the final result, including the

image enhancement to attenuate the effects of noises and increase the contrast between the foreground and background region, the detection of different face regions, the segmentation of the skin pigmentation regions and the estimation of the area and darkness scores of the detected pigmentation region. Chapter 4 presents a complete algorithm with details of those procedures. To be specific, the homomorphic filter is utilized for the illumination normalization. We propose a thresholding method that equally hybrids the local and global information of the image to segment the skin pigmentation. We use the bit quads method to calculate the area of the skin pigmentation and a local color difference method to calculate the pigmentation darkness. The proposed algorithm is evaluated using the real-world data collected in the National Skin Center of Singapore and the results show that there is a good correlation between the clinical ground truth mMASI scores and the computer estimated mMASI scores. The developed system in Chapter 4 is the first computerized algorithm that deals with the melasma severity assessment problem it has been transformed to an executable software for public use.

(3) The thesis has proposed an optimal hybrid thresholding method to improve the melasma severity assessment system.

This part of the thesis is an extension of the melasma severity assessment algorithm by improving the image segmentation method. The focus is on the segmentation process because, among the several steps mentioned previously, the image segmentation process is the most critical step in the algorithm. To be specific, an optimal hybrid thresholding algorithm (OHYTA) is proposed to segment the skin pigmentation regions. Different from the equal hybrid thresholding algorithm (EHYTA) presented in Chapter 4, OHYTA can change the trade-off parameter adaptively according to the different situations of the input image. This adaptability is achieved by minimizing a cost function that defines the property of the threshold surface, namely, the distance between the threshold surface and the

input image. The squared Euclidean distance is first used in the distance calculation and the results of real-world image set show that the OHYTA method with square Euclidean distance performs better than several thresholding based image segmentation methods compared. The method is improved by using a more generalized distance measure in the cost function.

In summary, the thesis has proposed different digital image processing algorithms for solving the real-world skin pigmentation diagnosis problem. Two specific problems have been addressed namely, the classification of the pigmentation images and the severity assessment of the melasma based on the image. To the best knowledge of the author, this is the first study that focuses on the area of non-tumorous skin pigmentation analysis based on computerized methods. The proposed algorithms have shown satisfactory performance when they are tested on the real-world images collected from the patients and the results provide a benchmark for the future developments.

6.2 Future Work

The works presented in this thesis are the first ones on the topic of classification and severity assessment of non-cancerous skin pigmentation disorders. Different aspects of the works can be improved in the future to obtain more accurate and more reliable algorithms.

For the pigmentation classification algorithm introduced in Chapter 3, we can improve the proposed method from different aspects. For the input features of the system, only the appearance features are used in the proposed method in the current phase. Although these features are the most important ones for the classification, there are other types of features can also provide critical information, for example, the age and the gender of the patients. A more reliable system can be

created in the next stage of development by utilizing all these helpful features. For the proposed V-PLDA method, it works on a balanced data set for the training. This is commonly not the case as some of the diseases would contain much more samples than the others. Therefore, to make better use of the available data, the proposed method can be made adaptive to the data ratio in the training set in the next step of development. Another shortcoming of the proposed V-PLDA method is, the model is based on an approximation process, which is not accurate. We can improve the classification performance by combining V-PLDA with the tied-PLDA method and form a multi-stage classification model. The results of the V-PLDA method can provide the prior information for the tied-PLDA model, which may have a better modelling of the data distribution. For the diseases in the study in the presented work, we only focused on five commonly seen skin pigmentation disorders, which is only a part of the big family of skin pigmentation diseases. Therefore, more types of skin pigmentation disorders can be included in the study in the future.

Another possible way to improve the classification performance is to use deep neural networks [114] as the classifier. The deep learning models, especially the convolutional neural networks are specifically suitable for solving the image classification problems. They have become more and more popular in the image classification problems for their superior performance [115, 116]. One disadvantage of the deep learning models is that they usually need large number of input data for the training. This is a critical problem for our pigmentation image analysis because the real-world data collected from the clinical center are quite limited. However, it is still possible to use these models by enlarging the dataset using data augmentation methods [117, 118].

For the melasma severity assessment algorithms introduced in Chapter 4 and Chapter 5, the segmentation results of certain input images are still not very

accurate. Although the proposed hybrid thresholding methods can adapt to different situations, they cannot handle the situation of strong noise very well. For example, in the images contain the forehead or chin region, the illumination problem like reflection and shadows are more severe. The segmentation of those regions is often less accurate. To solve this problem, further studies on noise removal of those regions can be done in the future. Besides, some studies have been done using other types of image segmentation methods, for example, the edge-based methods [39], to solve the problem. Those methods can be incorporated into the current work to make the algorithm more robust. [24].

The proposed melasma severity assessment system is semi-automatic, which requires the operator select a patch of normal skin region. It can be a possible future work to make the selection process automatic so that the whole system is no longer affected by the human input. The developed severity assessment system can also be transferred for solving other types of diseases' severity assessment in the future. For example, vitiligo, which is assessed by a system called Vitiligo Extent Tensity Index (VETI)

For both the classification and severity assessment algorithms, the number of real-world image data used in the experiment is limited. This is also a factor that constrains the validation and further development of the current methods. Therefore, it would be necessary to get more image data of patients for improving the methods presented in this thesis in the future.

Author's Publications

Journal papers

- Y. Liang, L. Sun, W. Ser, F. Lin, E. Y. Tay, E. Y. Gan, S. T. G. Thng, and Z. Lin, "Hybrid threshold optimization between global image and local regions in image segmentation for melasma severity assessment," *Multidimensional Systems and Signal Processing*, vol. 28, no.3, pp.977-994, 2017.
- Y. Liang, L. Sun, W. Ser, F. Lin, S. T. G. Thng, and Z. Lin, "Classification of non-tumorous skin pigmentation disorders using voting based probabilistic linear discriminant analysis," *Computers in Biology and Medicine*, vol. 99, pp.123-132, 2018.
- E. Y. Tay, E. Y. Gan, V. W. D. Tan, Z. Lin, Y. Liang, F. Lin, S. Wee, and S. T. G. Thng, "Pilot study of an automated method to determine Melasma Area and Severity Index." *British Journal of Dermatology*, vol. 172, no.6, pp.1535-1540, 2015.

Conference papers

- Y. Liang, Z. Lin, L. Sun, W. Ser, F. Lin, and S. T. G. Thng, "An improved image segmentation method for melasma severity assessment," in *IEEE International Conference on Digital Signal Processing*. London, UK, Aug. 2017, pp. 1-5.
- Y. Liang, Z. Lin, L. Sun, J. Cao, "Document image binarization via optimized hybrid thresholding," in *IEEE International Symposium on Circuits and Systems*. Baltimore, MD, USA, May 2017, pp. 1-4.

- Y. Liang, Z. Lin, J. Gu, W. Ser, F. Lin, E. Y. Tay, E. Y. Gan, V. W. D. Tan, and S. T. G. Thng, "Melasma image segmentation using extreme learning machine," in *Proceedings of ELM-2014 Volume 2*. Singapore, Dec. 2014, pp. 369-377.
- Y. Liang, Z. Lin, W. Ser, F. Lin, E. Y. Gan, E. Y. Tay, and S. T. G. Thng, "Automated scoring of melasma using computerized digital image analysis of clinical photographs. a pilot study," presented in *International Pigment Cell Conference*. Singapore, Sep. 2014.
- X. Zhang, Y. Liang, D. Lin, Z. Lin, S. T. G. Thng, E. Y. Gan, and E. Y. Tay, "Reaction-diffusion based level set method with local entropy thresholding for melasma image segmentation," in *International Conference on Control, Automation, Robotics and Vision*. Phuket, Thailand, Nov. 2016, pp. 1-5.

Bibliography

- [1] A. Slominski, D. J. Tobin, S. Shibahara, and J. Wortsman, “Melanin pigmentation in mammalian skin and its hormonal regulation,” *Physiological Reviews*, vol. 84, no. 4, pp. 1155–1228, 2004.
- [2] J. Y. Lin and D. E. Fisher, “Melanocyte biology and skin pigmentation,” *Nature*, vol. 445, no. 7130, pp. 843–850, 2007.
- [3] M. K. A. Basra, R. Fenech, R. M. Gatt, M. S. Salek, and A. Y. Finlay, “The dermatology life quality index 1994–2007: a comprehensive review of validation data and clinical results,” *British Journal of Dermatology*, vol. 159, no. 5, pp. 997–1035, 2008.
- [4] R. Balkrishnan, A. J. McMichael, F. T. Camacho, F. Saltzberg, T. S. Housman, S. Grummer, S. R. Feldman, and M.-M. Chren, “Development and validation of a health-related quality of life instrument for women with melasma,” *British Journal of Dermatology*, vol. 149, no. 3, pp. 572–577, 2003.
- [5] T. Cunliffe. (2013) Hyperpigmentation - of the face and neck. Primary Care Dermatology Society. [Online]. Available: <http://www.pcds.org.uk>
- [6] G.-E. Costin and V. J. Hearing, “Human skin pigmentation: melanocytes modulate skin color in response to stress,” *The FASEB Journal*, vol. 21, no. 4, pp. 976–994, 2007.

- [7] D. Scherer and R. Kumar, "Genetics of pigmentation in skin cancer review," *Mutation Research/Reviews in Mutation Research*, vol. 705, no. 2, pp. 141–153, 2010.
- [8] F. Xie, H. Fan, Y. Li, Z. Jiang, R. Meng, and A. Bovik, "Melanoma classification on dermoscopy images using a neural network ensemble model," *IEEE Transactions on Medical Imaging*, vol. 36, no. 3, pp. 849–858, 2017.
- [9] A. K. Jain, *Fundamentals of digital image processing*. Upper Saddle River, N.J., USA: Prentice-Hall, Inc., 1989.
- [10] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 3rd ed. Upper Saddle River, N.J., USA: Prentice-Hall, Inc., 2006.
- [11] R. Amelard *et al.*, "High-level intuitive features (HLIFs) for intuitive skin lesion description," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 3, pp. 820–831, 2015.
- [12] R. B. Oliveira, E. Mercedes Filho, Z. Ma, J. P. Papa, A. S. Pereira, and J. M. R. Tavares, "Computational methods for the image segmentation of pigmented skin lesions: a review," *Computer Methods and Programs in Biomedicine*, vol. 131, pp. 127–141, 2016.
- [13] K. Shimizu *et al.*, "Four-class classification of skin lesions with task decomposition strategy," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 1, pp. 274–283, 2015.
- [14] S. Y. Chuah, S. I. Tee, W. P. Tan, S. S. J. Lee, S. K. Ng, S. H. Chua, and T. G. S. Thng, "Reflectance confocal microscopy is a useful non-invasive tool in the *in vivo* diagnosis of pigmented basal cell carcinoma in Asians," *Australasian Journal of Dermatology*, vol. 58, no. 2, pp. 130–134, 2017.
- [15] T. L. Diepgen and V. Mahler, "The epidemiology of skin cancer," *British Journal of Dermatology*, vol. 146, no. s61, pp. 1–6, 2002.

- [16] J. Lu, E. Kazmierczak, J. H. Manton, and R. Sinclair, "Automatic segmentation of scaling in 2-d psoriasis skin images," *IEEE Transactions on Medical Imaging*, vol. 32, no. 4, pp. 719–730, 2013.
- [17] H. Nugroho, A. Fadzil, M. Hani, N. Shamsudin, and S. Hussein, "Computerised image analysis of vitiligo lesion: evaluation using manually defined lesion areas," *Skin Research and Technology*, vol. 19, no. 1, pp. 72–77, 2013.
- [18] W. Stolz, A. Riemann *et al.*, "ABCD rule of dermatoscopy-A new practical method for early recognition of malignant-melanoma," *European Journal of Dermatology*, vol. 4, no. 7, pp. 521–527, 1994.
- [19] R. M. Halder and G. M. Richards, "Topical agents used in the management of hyperpigmentation," *Skin Therapy Letter*, vol. 9, no. 6, pp. 1–3, 2004.
- [20] A. M. Rossi and M. I. Perez, "Treatment of hyperpigmentation," *Facial Plastic Surgery Clinics of North America*, vol. 19, no. 2, pp. 313–324, 2011.
- [21] C. K. Kimbrough-Green, C. E. Griffiths, L. J. Finkel, T. A. Hamilton, S. M. Bulengo-Ransby, C. N. Ellis, and J. J. Voorhees, "Topical retinoic acid (tretinoin) for melasma in black patients: a vehicle-controlled clinical trial," *Archives of Dermatology*, vol. 130, no. 6, pp. 727–733, 1994.
- [22] A. G. Pandya, L. S. Hynan, R. Bhore, F. C. Riley, I. L. Guevara, P. Grimes, J. J. Nordlund, M. Rendon, S. Taylor, R. W. Gottschalk *et al.*, "Reliability assessment and validation of the Melasma Area and Severity Index (MASI) and a new modified masi scoring method," *Journal of the American Academy of Dermatology*, vol. 64, no. 1, pp. 78–83, 2011.
- [23] A. Taieb and M. Picardo, "The definition and assessment of vitiligo: a consensus report of the Vitiligo European Task Force," *Pigment Cell & Melanoma Research*, vol. 20, no. 1, pp. 27–35, 2007.

- [24] A. Feily, “Vitiligo Extent Tensity Index (VETI) score: a new definition, assessment and treatment evaluation criteria in vitiligo,” *Dermatology Practical & Conceptual*, vol. 4, no. 4, pp. 81–84, 2014.
- [25] N. A. Vashi and R. Kundu, “Facial hyperpigmentation: Causes and treatment,” *British Journal of Dermatology*, vol. 169, no. s3, pp. 41–56, 2013.
- [26] J.-P. Ortonne and D. L. Bissett, “Latest insights into skin hyperpigmentation,” in *Journal of Investigative Dermatology Symposium Proceedings*, vol. 13, no. 1. Elsevier, 2008, pp. 10–14.
- [27] R. F. Wagner *et al.*, “Residents corner: Diagnoses of skin disease: Dermatologists vs. nondermatologists,” *The Journal of Dermatologic Surgery and Oncology*, vol. 11, no. 5, pp. 476–479, 1985.
- [28] L. Brochez *et al.*, “Diagnostic ability of general practitioners and dermatologists in discriminating pigmented skin lesions,” *Journal of the American Academy of Dermatology*, vol. 44, no. 6, pp. 979–986, 2001.
- [29] T. W. Lin. (2017, Apr.) 5 myths about pigmentation debunked by a dermatologist. THREAD by ZALORA Singapore. [Online]. Available: <https://thread.zalora.sg/>
- [30] E. Y. Tay, E. Y. Gan, V. W. D. Tan, Z. Lin, Y. Liang, F. Lin, S. Wee, and T. G. Thng, “Pilot study of an automated method to determine Melasma Area and Severity Index,” *British Journal of Dermatology*, vol. 172, no. 6, pp. 1535–1540, 2015.
- [31] S. J. D. Prince and J. H. Elder, “Probabilistic linear discriminant analysis for inferences about identity,” in *IEEE International Conference on Computer Vision*. IEEE, 2007, pp. 1–8.
- [32] S. J. D. Prince, J. H. Elder, J. Warrell, and F. M. Felisberti, “Tied factor analysis for face recognition across large pose differences,” *IEEE Transactions*

- on *Pattern Analysis and Machine Intelligence*, vol. 30, no. 6, pp. 970–984, 2008.
- [33] P. Li, Y. Fu, U. Mohammed, J. H. Elder, and S. J. D. Prince, “Probabilistic models for inference about identity,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 1, pp. 144–157, 2012.
- [34] Y. Liang, L. Sun, W. Ser, F. Lin, S. T. G. Thng, Q. Chen, and Z. Lin, “Classification of non-tumorous skin pigmentation disorders using voting based probabilistic linear discriminant analysis,” *Computers in Biology and Medicine*, vol. 99, pp. 123–132, 2018. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0010482518301392>
- [35] P. E. Grimes, “Melasma: Etiologic and therapeutic considerations,” *Archives of Dermatology*, vol. 131, no. 12, pp. 1453–1457, 1995.
- [36] V. M. Sheth and A. G. Pandya, “Melasma: A comprehensive update,” *Journal of the American Academy of Dermatology*, vol. 65, no. 4, pp. 689–714, 2011.
- [37] A. d. Tamega, L. D. B. Miot, C. Bonfietti, T. C. Gige, M. E. A. Marques, and H. A. Miot, “Clinical patterns and epidemiological characteristics of facial melasma in Brazilian women,” *Journal of the European Academy of Dermatology and Venereology*, vol. 27, no. 2, pp. 151–156, 2013.
- [38] A. C. Handel, P. B. Lima, V. M. Tonolli, L. D. B. Miot, and H. A. Miot, “Risk factors for facial melasma in women: A case-control study,” *British Journal of Dermatology*, vol. 171, no. 3, pp. 588–594, 2014.
- [39] X. Zhang, Y. Liang, D. Lin, Z. Lin, S. T. G. Thng, E. Y. Gan, and E. Y. Tay, “Reaction-diffusion based level set method with local entropy thresholding for melasma image segmentation,” in *International Conference on Control, Automation, Robotics and Vision*. IEEE, 2016, pp. 1–5.

- [40] D. L. Pham, C. Xu, and J. L. Prince, “Current methods in medical image segmentation,” *Annual Review of Biomedical Engineering*, vol. 2, no. 1, pp. 315–337, 2000.
- [41] B. N. Saha and N. Ray, “Image thresholding by variational minimax optimization,” *Pattern Recognition*, vol. 42, no. 5, pp. 843–856, 2009.
- [42] Y. Liang, L. Sun, W. Ser, F. Lin, E. Y. Tay, E. Y. Gan, T. G. Thng, and Z. Lin, “Hybrid threshold optimization between global image and local regions in image segmentation for melasma severity assessment,” *Multidimensional Systems and Signal Processing*, vol. 28, no. 3, pp. 977–994, 2017.
- [43] Y. Liang, Z. Lin, L. Sun, W. Ser, F. Lin, and T. G. S. Thng, “An improved image segmentation method for melasma severity assessment,” in *IEEE International Conference on Digital Signal Processing*. IEEE, August 2017.
- [44] P. M. Roth and M. Winter, “Survey of appearance-based methods for object recognition,” Institute of Computer Graphics and Vision, Graz University of Technology, Austria, Tech. Rep. ICG-TR-01/08, 2008.
- [45] K. Grauman and B. Leibe, “Visual object recognition,” *Synthesis Lectures on Artificial Intelligence and Machine Learning*, vol. 4, no. 2, pp. 1–181, 2011.
- [46] D. G. Lowe, “Object recognition from local scale-invariant features,” in *IEEE International Conference on Computer Vision*, vol. 2. IEEE, 1999, pp. 1150–1157.
- [47] M. Brown and D. Lowe, “Invariant features from interest point groups,” in *British Machine Vision Conference*, vol. 4, 2002, pp. 656–665.
- [48] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

- [49] T. Ojala, M. Pietikäinen, and D. Harwood, “A comparative study of texture measures with classification based on featured distributions,” *Pattern Recognition*, vol. 29, no. 1, pp. 51–59, 1996.
- [50] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1. IEEE, 2005, pp. 886–893.
- [51] J. Yang, Y.-G. Jiang, A. G. Hauptmann, and C.-W. Ngo, “Evaluating bag-of-visual-words representations in scene classification,” in *Proceedings of the international Workshop on Workshop on Multimedia information Retrieval*. ACM, 2007, pp. 197–206.
- [52] S. Lazebnik, C. Schmid, and J. Ponce, “Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories,” in *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2. IEEE, 2006, pp. 2169–2178.
- [53] C. Cortes and V. Vapnik, “Support vector machine,” *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [54] J. A. Suykens and J. Vandewalle, “Least squares support vector machine classifiers,” *Neural processing letters*, vol. 9, no. 3, pp. 293–300, 1999.
- [55] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, “Extreme learning machine: theory and applications,” *Neurocomputing*, vol. 70, no. 1, pp. 489–501, 2006.
- [56] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, “Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.

- [57] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, “Robust face recognition via sparse representation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.
- [58] X. Jiang and J. Lai, “Sparse and dense hybrid representation via dictionary decomposition for face recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 5, pp. 1067–1079, 2015.
- [59] X. Wang and X. Tang, “Dual-space linear discriminant analysis for face recognition,” in *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2. IEEE, 2004, pp. 564–569.
- [60] H. Lee, A. Battle, R. Raina, and A. Y. Ng, “Efficient sparse coding algorithms,” in *Advances in Neural Information Processing Systems*, 2007, pp. 801–808.
- [61] Z. S. Harris, “Distributional structure,” *Word*, vol. 10, no. 2-3, pp. 146–162, 1954.
- [62] A. Vedaldi and B. Fulkerson, “VLFeat: An open and portable library of computer vision algorithms,” in *ACM International Conference on Multimedia*. ACM, 2010, pp. 1469–1472.
- [63] A. E. Abdel-Hakim and A. A. Farag, “CSIFT: A SIFT descriptor with color invariant characteristics,” in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2. IEEE, 2006, pp. 1978–1983.
- [64] S. Maji, A. C. Berg, and J. Malik, “Classification using intersection kernel support vector machines is efficient,” in *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2008, pp. 1–8.
- [65] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, “Locality-constrained linear coding for image classification,” in *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2010, pp. 3360–3367.

- [66] J. Yang, K. Yu, Y. Gong, and T. Huang, "Linear spatial pyramid matching using sparse coding for image classification," in *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2009, pp. 1794–1801.
- [67] K. Yu, T. Zhang, and Y. Gong, "Nonlinear learning using local coordinate coding," in *Advances in Neural Information Processing Systems*, 2009, pp. 2223–2231.
- [68] P. Li and S. J. D. Prince, "Probabilistic methods for face registration and recognition," in *Advances in Face Image Analysis: Techniques and Technologies*. Hershey, PA, USA: IGI Global, 2010, pp. 178–197.
- [69] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society*, vol. 39, pp. 1–38, 1977.
- [70] P. E. Grimes, "Melasma: etiologic and therapeutic considerations," *Archives of Dermatology*, vol. 131, no. 12, pp. 1453–1457, 1995.
- [71] V. M. Sheth and A. G. Pandya, "Melasma: a comprehensive update: part II," *Journal of the American Academy of Dermatology*, vol. 65, no. 4, pp. 699–714, 2011.
- [72] A. d. Tamega, L. D. B. Miot, C. Bonfietti, T. C. Gige, M. E. A. Marques, and H. A. Miot, "Clinical patterns and epidemiological characteristics of facial melasma in Brazilian women," *Journal of the European Academy of Dermatology and Venereology*, vol. 27, no. 2, pp. 151–156, 2013.
- [73] G. A. Hance, S. E. Umbaugh, R. H. Moss, and W. V. Stoecker, "Unsupervised color image segmentation: with application to skin tumor borders," *IEEE Engineering in Medicine and Biology Magazine*, vol. 15, no. 1, pp. 104–111, 1996.

- [74] M. Silveira, J. C. Nascimento, J. S. Marques, A. R. Marçal, T. Mendonça, S. Yamauchi, J. Maeda, and J. Rozeira, “Comparison of segmentation methods for melanoma diagnosis in dermoscopy images,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 3, no. 1, pp. 35–45, 2009.
- [75] N. Otsu, “A threshold selection method from gray-level histograms,” *Automatica*, vol. 11, no. 285-296, pp. 23–27, 1975.
- [76] J. Ghaye, C. Succa, D. Demarchi, S. K. Muldur, P. Colpo *et al.*, “Quantitative estimation of biological cell surface receptors by segmenting conventional fluorescence microscopy images,” in *IEEE International Symposium on Circuits and Systems*. IEEE, 2014, pp. 1824–1827.
- [77] R. F. Moghaddam and M. Cheriet, “A multi-scale framework for adaptive binarization of degraded document images,” *Pattern Recognition*, vol. 43, no. 6, pp. 2186–2198, 2010.
- [78] ———, “Adotsu: An adaptive and parameterless generalization of otsu’s method for document image binarization,” *Pattern Recognition*, vol. 45, no. 6, pp. 2419–2431, 2012.
- [79] J. Sauvola and M. Pietikäinen, “Adaptive document image binarization,” *Pattern Recognition*, vol. 33, no. 2, pp. 225–236, 2000.
- [80] S. G. Y. Ho and H. H. L. Chan, “The Asian dermatologic patient,” *American Journal of Clinical Dermatology*, vol. 10, no. 3, pp. 153–168, 2009.
- [81] P. Refaeilzadeh, L. Tang, and H. Liu, “Cross-validation,” in *Encyclopedia of Database Systems*. Springer, 2009, pp. 532–538.
- [82] R. Lowry. (2014) Concepts and applications of inferential statistics. Vassar College. [Online]. Available: <http://vassarstats.net/textbook/>

- [83] W. Y. Lo and S. M. Puchalski, “Digital image processing,” *Veterinary Radiology & Ultrasound*, vol. 49, no. s1, 2008.
- [84] G. Paschos, “Perceptually uniform color spaces for color texture analysis: an empirical evaluation,” *IEEE Transactions on Image Processing*, vol. 10, no. 6, pp. 932–937, 2001.
- [85] X. Xie, “A review of recent advances in surface defect detection using texture analysis techniques,” *Electronic Letters on Computer Vision and Image Analysis*, vol. 7, no. 3, pp. 1–22, 2008.
- [86] T. Chen, X. Wang, S. Chung, D. Metaxas, and L. Axel, “Automated 3D motion tracking using Gabor filter bank, robust point matching, and deformable models,” *IEEE Transactions on Medical Imaging*, vol. 29, no. 1, pp. 1–11, 2010.
- [87] T. Torheim, E. Malinen, K. Kvaal, H. Lyng, U. G. Indahl, E. K. F. Andersen, and C. M. Futsæther, “Classification of dynamic contrast enhanced MR images of cervical cancers using texture analysis and support vector machines,” *IEEE Transactions on Medical Imaging*, vol. 33, no. 8, pp. 1648–1656, 2014.
- [88] L. Shen and L. Bai, “A review on Gabor wavelets for face recognition,” *Pattern Analysis and Applications*, vol. 9, no. 2-3, pp. 273–292, 2006.
- [89] L. Itti, C. Koch, and E. Niebur, “A model of saliency-based visual attention for rapid scene analysis,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.
- [90] M. Yang and L. Zhang, “Gabor feature based sparse representation for face recognition with Gabor occlusion dictionary,” *European Conference on Computer Vision*, pp. 448–461, 2010.

- [91] C. Liu and H. Wechsler, “Gabor feature based classification using the enhanced Fisher linear discriminant model for face recognition,” *IEEE Transactions on Image Processing*, vol. 11, no. 4, pp. 467–476, 2002.
- [92] A. K. Jain, Y. Chen, and M. Demirkus, “Pores and ridges: High-resolution fingerprint matching using level 3 features,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 1, pp. 15–27, 2007.
- [93] F. Bianconi and A. Fernández, “Evaluation of the effects of Gabor filter parameters on texture classification,” *Pattern Recognition*, vol. 40, no. 12, pp. 3325–3335, 2007.
- [94] S. Maji, A. C. Berg, and J. Malik, “Efficient classification for additive kernel SVMs,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 66–77, 2013.
- [95] A. Goldsberry, C. Hanke, and K. Hanke, “Visia system: a possible tool in the cosmetic practice.” *Journal of drugs in dermatology: JDD*, vol. 13, no. 11, pp. 1312–1314, 2014.
- [96] G. Buchsbaum, “A spatial processor model for object colour perception,” *Journal of the Franklin Institute*, vol. 310, no. 1, pp. 1–26, 1980.
- [97] B. Funt and L. Shi, “The rehabilitation of maxrgb,” in *Color and Imaging Conference*, vol. 2010, no. 1. Society for Imaging Science and Technology, 2010, pp. 256–259.
- [98] E. H. Land and J. J. McCann, “Lightness and retinex theory,” *Josa*, vol. 61, no. 1, pp. 1–11, 1971.
- [99] I. Stainvas and D. Lowe, “A generative model for separating illumination and reflectance from images,” *Journal of Machine Learning Research*, vol. 4, no. Dec, pp. 1499–1519, 2003.

- [100] C.-N. Fan and F.-Y. Zhang, "Homomorphic filtering based illumination normalization method for face recognition," *Pattern Recognition Letters*, vol. 32, no. 10, pp. 1468–1479, 2011.
- [101] A. v. Oppenheim, R. Schafer, and T. Stockham, "Nonlinear filtering of multiplied and convolved signals," *IEEE Transactions on Audio and Electroacoustics*, vol. 16, no. 3, pp. 437–466, 1968.
- [102] C. Fernandez-Maloigne, *Advanced color image processing and analysis*. Springer Science & Business Media, 2012, ch. Color Image Segmentation, pp. 219–278.
- [103] G. Xiong, X. Li, J. Gong, H. Chen, and D.-J. Lee, "Color rank and census transforms using perceptual color contrast," in *International Conference on Control Automation Robotics & Vision*. IEEE, 2010, pp. 1225–1230.
- [104] E. Smistad, T. L. Falch, M. Bozorgi, A. C. Elster, and F. Lindseth, "Medical image segmentation on gpus—a comprehensive review," *Medical image analysis*, vol. 20, no. 1, pp. 1–18, 2015.
- [105] E. Gabarra and A. Tabbone, "Combining global and local threshold to binarize document of images," *Pattern Recognition and Image Analysis*, pp. 173–186, 2005.
- [106] E. Espinoza, G. Martinez, J.-G. Frerichs, and T. Scheper, "Cell cluster segmentation based on global and local thresholding for in-situ microscopy," in *IEEE International Symposium on Biomedical Imaging: Nano to Macro*. IEEE, 2006, pp. 542–545.
- [107] E. Zemouri, Y. Chibani, and Y. Brik, "Enhancement of historical document images by combining global and local binarization technique," *International Journal of Information and Electronics Engineering*, vol. 4, no. 1, pp. 1–5, 2014.

- [108] Q. Li, W. Lu, and J. Yang, "A hybrid thresholding algorithm for cloud detection on ground-based color images," *Journal of atmospheric and oceanic technology*, vol. 28, no. 10, pp. 1286–1296, 2011.
- [109] J. E. Adams, "Digital image processing," *Journal of Electronic Imaging*, vol. 16, no. 2, 2007.
- [110] S. B. Gray, "Local properties of binary images in two dimensions," *IEEE Transactions on Computers*, vol. 100, no. 5, pp. 551–561, 1971.
- [111] W. Pirie, "Spearman rank correlation coefficient," *Encyclopedia of Statistical Sciences*, 1988.
- [112] I.-J. Kim, "Multi-window binarization of camera image for document recognition," in *International Workshop on Frontiers in Handwriting Recognition*. IEEE, 2004, pp. 323–327.
- [113] D. G. Gavin, W. W. Oswald, E. R. Wahl, and J. W. Williams, "A statistical approach to evaluating distance metrics and analog assignments for pollen records," *Quaternary Research*, vol. 60, no. 3, pp. 356–367, 2003.
- [114] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, p. 436, 2015.
- [115] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [116] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.
- [117] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, "Return of the devil in the details: Delving deep into convolutional nets," *arXiv preprint arXiv:1405.3531*, 2014.

-
- [118] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.