# FEATURE EXTRACTION AND DESCRIPTION FOR RETINAL FUNDUS IMAGE REGISTRATION

ROZIANA BINTI RAMLI

FACULTY OF COMPUTER SCIENCE AND
INFORMATION TECHNOLOGY
UNIVERSITY OF MALAYA
KUALA LUMPUR

2020

# FEATURE EXTRACTION AND DESCRIPTION FOR RETINAL FUNDUS IMAGE REGISTRATION

**ROZIANA BINTI RAMLI**

**THESIS SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY**

**FACULTY OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY
UNIVERSITY OF MALAYA
KUALA LUMPUR**

**2020**

**UNIVERSITY OF MALAYA**

**ORIGINAL LITERARY WORK DECLARATION**

Name of Candidate: Roziana Binti Ramli

Matric No: WHA140056

Name of Degree: Doctor of Philosophy

Title of Project Paper/Research Report/Dissertation/Thesis ("this Work"):

Feature Extraction and Description for Retinal Fundus Image Registration

Field of Study: Image Processing (COMPUTER SCIENCE)

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This Work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date:

Subscribed and solemnly declared before,

Witness's Signature Date:

Name:

Designation:

## ABSTRACT

Retinal fundus image registration (RIR) is performed to align two or more fundus images. A general framework of a feature-based RIR technique comprises of pre-processing, feature extraction, feature descriptor, matching and estimating geometrical transformation. The RIR is mainly performed for super-resolution, image mosaicking and longitudinal study applications to assist diagnosis and monitoring retinal diseases. Registering image pair from these applications involve a combination of challenges such as overlapping area and rotation between images. The challenges of the overlapping area and rotation can be addressed at feature extraction and feature descriptor stages of the feature-based RIR technique, respectively. To address the overlapping area, a reliable and repeatable anatomical information such as retinal vessels is required. However, finding the feature points on retinal vessels can be challenging due to noises with similar structure representation as the vessels. For rotation, a distinctive descriptor is necessary to characterise the feature points on retinal vessels that are lack of textural information and exhibit repetitive patterns in the local patches of fundus image. Therefore, this study proposed new feature extraction and feature descriptor methods for feature-based RIR technique to address these issues. The proposed feature extraction method extracts the feature points on retinal vessels by considering the characteristics of the retinal vessels and noises. The proposed feature descriptor method characterises the feature points with statistical properties obtained from the surrounding region of the feature points. The proposed work is tested on five public datasets, namely, CHASE_DB1, DRIVE, HRF, STARE and FIRE. Aspects of the evaluation include evaluating the extraction accuracy of the proposed feature extraction method and the registration accuracy of the proposed feature-based RIR technique. Experimental results show that the proposed feature extraction method attained the highest overall extraction accuracy (86.021%) and outperformed the existing feature extraction methods; Harris corner (41.613%), SIFT (16.164%), SURF (18.929%), Ghassabi's (28.280%) and D-Saddle (20.509%). The registration success rate of the proposed feature-based RIR technique (67.164%) is also outperformed the existing feature-based RIR techniques; Harris-PIIFD (3.731%), GDB-ICP (27.612%), Ghassabi's-SIFT (12.687%), H-M 16 (16.418%), H-M 17 (19.403%) and D-Saddle-HOG (11.940%). The influence of the overlapping area and rotation on the proposed feature-based RIR technique are significant but the weakest among the evaluated feature-based RIR techniques.

Keywords: Feature-based image registration, fundus image, feature extraction, feature descriptor.

# ABSTRAK

Pendaftaran imej retina (RIR) dilakukan untuk menyelaraskan dua atau lebih imej fundus. Rangka umum teknik RIR berasaskan ciri terdiri daripada pra-pemprosesan, pengekstrakan ciri, deskriptor ciri, padanan dan penganggaran transformasi geometri. RIR dilaksanakan terutamanya bagi aplikasi resolusi super, imej mosaik dan kajian membujur untuk membantu diagnosis dan pemantauan penyakit retina. Mendaftar pasangan imej dari aplikasi-aplikasi ini melibatkan gabungan beberapa cabaran seperti kawasan pertindihan dan putaran antara imej. Cabaran kawasan pertindihan dan putaran dapat ditangani pada peringkat pengekstrakan ciri dan descriptor bagi teknik RIR berasaskan ciri. Untuk menangani kawasan pertindihan, maklumat anatomi yang boleh dipercayai dan berulang seperti saluran darah retina diperlukan. Walau bagaimanapun, mencari titik ciri yang terletak pada saluran darah retina adalah sukar kerana pencemaran yang mempunyai struktur yang hampir sama dengan saluran darah retina. Bagi putaran, deskriptor tersendiri diperlukan untuk mencirikan titik ciri pada saluran darah retina yang kekurangan maklumat tekstural dan mempamerkan corak berulang dalam petak tempatan imej fundus. Oleh itu, kajian ini mencadangkan kaedah pengekstrakan ciri dan deskriptor ciri yang baharu bagi teknik RIR berasaskan ciri untuk menangani isu-isu ini. Kaedah pengekstrakan ciri yang dicadangkan mengekstrak titik ciri pada saluran darah retina dengan mempertimbangkan ciri-ciri pencemaran dan saluran darah retina yang diperhatikan pada profil intensiti. Kaedah deskriptor ciri yang dicadangkan mencirikan titik ciri dengan sifat statistik yang diperoleh dari kawasan sekitar titik ciri. Kerja yang dicadangkan diuji pada lima dataset awam, iaitu, CHASE_DB1, DRIVE, HRF, STARE and FIRE. Aspek penilaian termasuk menilai ketepatan pengekstrakan bagi kaedah pengekstrakan ciri yang dicadangkan dan ketepatan pendaftaran teknik RIR berasaskan ciri yang dicadangkan. Keputusan eksperimen menunjukkan bahawa kaedah pengekstrakan ciri yang dicadangkan mencapai ketepatan pengekstrakan keseluruhan yang tertinggi (86.021%) dan mengatasi kaedah pengekstrakan ciri terdahulu; Harris (41.613%), SIFT (16.164%), SURF (18.929%), Ghassabi's (28.280%) and D-Saddle (20.509%). Kadar kejayaan pendaftaran teknik RIR berasaskan ciri yang dicadangkan (67.164%) juga mengatasi teknik RIR berasaskan ciri yang terdahulu; Harris-PIIFD (3.731%), GDB-ICP (27.612%), Ghassabi-SIFT (12.687%), H-M 16 (16.418%), H-M 17 (19.403%) dan D-Saddle-HOG (11.940%). Pengaruh kawasan penindihan dan putaran pada teknik RIR berasaskan ciri yang dicadangkan adalah signifikan tetapi paling lemah di antara teknik RIR berasaskan ciri yang dinilai.

Kata kunci: Pendaftaran imej berasaskan ciri, imej fundus, pengekstrakan ciri, deskriptor ciri.

## ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisors Assoc. Prof. Dr. Mohd. Yamani Idna Idris, Dr. Khairunnisa Hasikin and Dr. Noor Khairiah A. Karim for all their help and advice throughout my Ph.D. study. Most important of all, I want to express my deepest gratitude towards my family for their love and support throughout this process. I am also thankful to the people whom I did not mention, but have a valuable contribution to this research. This research was supported by the Postgraduate Research Grant under Grant PG039-2015B and MyBrain15 scheme of the Ministry of Higher Education, Malaysia

## TABLE OF CONTENTS

## LIST OF FIGURES

## LIST OF TABLES

## LIST OF ABBREVIATIONS

| | | |
|---|---|---|
| ALOHA | : | An efficient binary descriptor based on Haar features |
| ANOVA | : | Analysis of variance |
| BBF | : | Best-bin-first |
| BRIEF | : | Binary robust independent elementary features |
| BRISK | : | Binary robust invariant scalable keypoints |
| DoG | : | Difference of Gaussian |
| DRIVE | : | Digital Retinal Images for Vessel Extraction |
| Fast-NNS | : | Fast approximate nearest-neighbour search |
| FIRE | : | Fundus Image Registration |
| FOV | : | Field of view |
| FVM | : | Frangi's vesselness measure |
| GDB-ICP | : | Generalized dual-bootstrap iterative closest point |
| GMM | : | Gaussian mixture model |
| GTM | : | Graph transformation matching |
| HOG | : | Histogram of oriented gradients |
| HRF | : | High-Resolution Fundus Image Database |
| *ii*DoG | : | Illumination invariant difference of Gaussian |
| LNNS | : | Lowe's nearest-neighbour search |
| LoSPA | : | Low-dimensional step pattern analysis |
| MSAC | : | M-estimator SAmple and Consensus |
| *n*DoG | : | Normalised difference of Gaussian |
| NNS | : | Nearest-neighbour search |
| OCT | : | Optical coherence tomography |
| PIIFD | : | Partial intensity invariant feature descriptor |
| RANSAC | : | RANdom SAmple Consensus |
| RGB | : | Red, green and blue composites |
| RIR | : | Retinal fundus image registration |
| ROI | : | Region of interest |
| SIFT | : | Scale-invariant feature transform |
| SLO | : | Scanning laser ophthalmoscope |
| SSD | : | Sum of squared differences |
| STARE | : | Structured Analysis of the Retina |

| | | |
|---|---|---|
| SURF | : | Speeded up robust features |
| TPS | : | Thin-plate spline |
| TRE | : | Target registration error |
| UR-SIFT | : | Uniform robust scale-invariant feature transform |

## LIST OF APPENDICES

# CHAPTER 1: INTRODUCTION

## 1.1 INTRODUCTION

Ocular and systemic diseases manifest in the retina in various form of abnormalities such as retinal detachment, cupping of the optic disc, cotton wool spots, narrowed arteries and thickened veins (Abràmoff, Garvin, & Sonka, 2010). These abnormalities can be documented and observed through retinal fundus imaging. Fundus imaging is widely employed in documenting retinal abnormalities because it is safe, feasible and cost-effective.

There are three main applications of retinal fundus image registration (RIR), namely, super-resolution, image mosaicking and longitudinal study applications. The super-resolution application combines multiple fundus images to increase the density of the spatial sampling. This can resolve the blur edges of the retinal vessels caused by the patient movements or improper imaging setup. The image mosaicking application aligns multiple fundus images to generate an image with a wider view of the retina. The wide view image of the retina can be used to view a full extent of the retinal disease in one big picture during diagnosis (Bontala, Sivaswamy, & Pappuru, 2012; D. M. Brown & Ciardella, 2005) and during preparation of eye laser treatment for diabetic retinopathy (B. H. Lee, Xu, Gopalakrishnan, Ong, Li, Wong, & Lim, 2015). The longitudinal study application combines multiple fundus images that are acquired at different screening sessions. This application is essential in monitoring the progression of the retinal diseases which usually undergoes a long degeneration process such as glaucoma and age-related macular degeneration (Adal, van Etten, Martinez, van Vliet, & Vermeer, 2015).

A general approach in performing image registration is to establish correspondences between a pair of images. These images are referred as fixed and moving images. Then, the established correspondences are used to estimate geometrical transformation. The geometrical transformation is composed of operations such as scaling, rotation and translation to align the moving image to the orientation of the fixed image. Existing RIR techniques can be grouped as area-based and feature-based. These groups are according to the type of correspondence utilised in the image registration process.

The area-based RIR technique finds correspondences of intensity pattern between fixed and moving images. These correspondences are established using similarity metrics such as mutual information (Legg, Rosin, Marshall, & Morgan, 2015) and phase correlation (Kolar, Sikula, & Base, 2010). The process of establishing correspondences is repeated while optimising the geometrical transformation until an optimum registration or a maximum number of the iteration is achieved. However, the area-based RIR technique considers the intensity pattern from all part of the retina in fundus image. This cause the intensity from the non-overlapping area to mislead the similarity metric to establish incorrect correspondences. Accordingly, the estimated geometrical transformation from the incorrect correspondences will result in inaccurate registration.

Generally, the feature-based RIR technique is more robust in registering fundus images compared to area-based RIR technique. This is because of the feature-based RIR technique only considers significant local information between images during the registration process. Therefore, the feature-based RIR technique is chosen in this study.

A general framework of the feature-based RIR technique comprises of five main stages. These stages are pre-processing, feature extraction, feature descriptor, matching and estimating geometrical transformation. The pre-processing stage involves the process of converting input images to the desired colour space or improving the quality of the input images. The feature extraction stage extracts local information such as feature points from local patches throughout fixed and moving images. The feature points can be extracted based on anatomical information or invariant feature. The examples of the anatomical related information are retinal vessels and vessel bifurcations whereas invariant feature includes scale-invariant feature transform (SIFT) (Lowe, 2004) and speeded up robust features (SURF) (Bay, Ess, Tuytelaars, & Van Gool, 2008; Bay, Tuytelaars, & Van Gool, 2006). Then, feature descriptor is computed for each feature point to describe their surrounding region. The computed feature descriptor will be used to find the corresponding feature points or matches between the images. Finally, the geometrical transformation between images is estimated according to the established matches.

Registering image pair from super-resolution, image mosaicking and longitudinal study applications involve a combination of several challenges, namely, overlapping area, rotation and image quality between images. Among these challenges, addressing overlapping area and rotation between images will be our main focused in this thesis.

a) Overlapping area – The overlapping area is an intersection region between fixed and moving images. A small overlapping area limits the amount of common region available between images, which can be insufficient to estimate an accurate geometrical transformation. The example of the overlapping area between fixed and moving images is visualised in Figure 1.1.
b) Rotation – The rotation in fundus image is introduced to access part of the retina or due to involuntary movement by the patient. The rotation alters the orientation of the common region between images. This alteration can be challenging for feature-based RIR technique to establish correspondences. The example of the rotation between fixed and moving images is depicted in Figure 1.2, where the moving image is rotated at the angle of 30° anti-clockwise.
c) Image quality – The variation of the image quality between images can be caused by inconsistent imaging setup, uneven absorption of light due to the spherical shape of

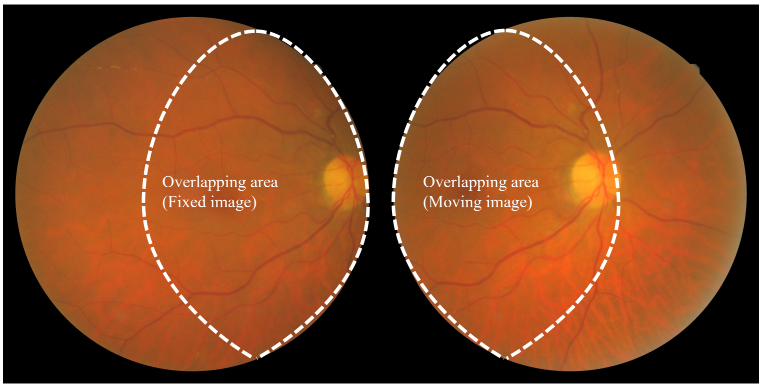



**Figure 1.1: Example of overlapping area between fixed and moving images.**

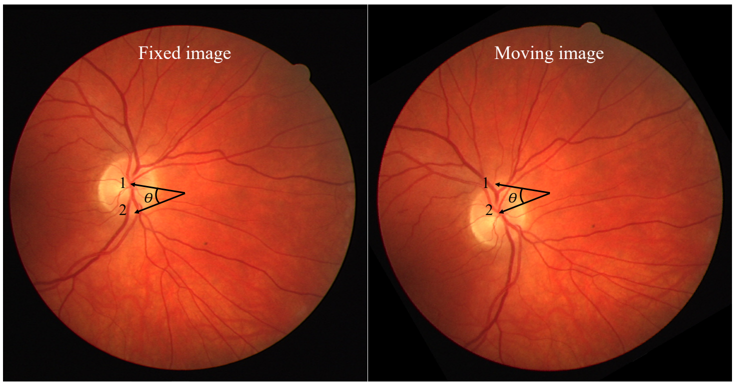



**Figure 1.2: Example of rotation where a moving image is rotated at the angle of 30° anti-clockwise.**

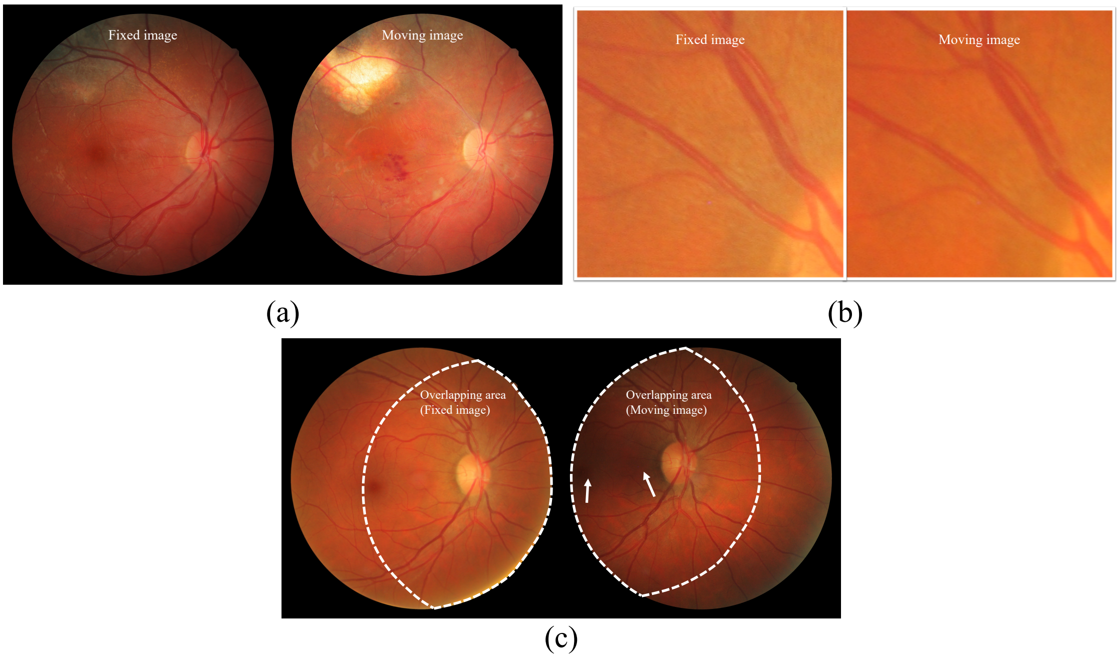



**Figure 1.3: Examples of image quality: (a) intensity difference (b) difference of structure similarity (c) difference of intensity distribution where the white arrows pointing the non-uniform intensity region obscuring the visibility of the retinal vessels.**

the retina or the progression of the diseases. The variation of the image quality as depicted in Figure 1.3, can affect the image pair in many ways. For example, a significant difference in intensity and structure similarity between images reduce the similarity level of the overlapping area. Other than that, the non-uniform intensity distribution can potentially obscure the visibility of anatomical information in the overlapping area. Thus, a high difference of the intensity distribution between images can limit the amount of the common region in the overlapping area, which yields the same challenge as small overlapping area.

## 1.2 RESEARCH PROBLEMS

The challenges of the overlapping area and rotation can be addressed at different stages of the feature-based RIR technique. The overlapping area can be addressed at the feature extraction stage, whereas rotation can be addressed at the feature descriptor stage. Therefore, the feature-based RIR technique requires a reliable feature extraction and feature descriptor methods to address the mentioned challenges.

A reliable feature extraction method extracts feature points that are distributed throughout the image and repeatable between images. In fundus image, anatomical related information of retinal vessels can be found throughout the fundus image and repeatable between images (Deng, Tian, Zheng, Zhang, Dai, & Xu, 2010; S. K. Saha, Xiao, Frost, & Kanagasingam, 2016). Furthermore, the position of the retinal vessels is generally static over a short period of time even in a diseased eye (Xiao, Vignarajan, Lock, Frost, Tay-Kearney, & Kanagasingam, 2012). Therefore, in this study, the feature points are extracted on retinal vessels to ensure sufficient matches can be established to estimate an accurate geometrical transformation. There are two major concerns of existing feature extraction methods in extracting feature points on retinal vessels as follows:

**RP1:** The lack of existing studies examining the characteristics of the retinal vessels and noises in local patches of fundus image.

**RP2:** Most of the existing feature extraction methods are unable to accurately extract feature points on retinal vessels of various sizes and without a proper selection module to accurately distinguish between retinal vessels and noises.

For a reliable feature descriptor method, the feature points are characterised in a way that allows it to be distinguishable from other feature points but comparable between corresponding feature points. The feature descriptor in the existing feature-based RIR techniques are mainly based on gradient direction computed from grids within a local square patch. The rotation invariance is achieved by rotating the square patch according to the dominant orientation of the local patch. The feature descriptor that based on gradient direction is well known for its rotation invariance in object and scene images. However, it can be indistinctive when describing feature points on structure with lack of textural information (Hinterstoisser, Cagniart, Ilic, Sturm, Navab, Fua, & Lepetit, 2012) and repetitive patterns (Fang, Yu, Ma, & An, 2019; Kushnir & Shimshoni, 2014). The lack of textural information and repetitive patterns are commonly observed in the local patches of fundus image (Abràmoff et al., 2010; Deng et al., 2010). Particularly, when the local patches contain bifurcation, low contrast or narrowed vessels. Therefore, the feature descriptor based on gradient direction can be unsuitable to describe feature points on retinal vessels. This concern can be summarised as follows:

**RP3:** Most of the existing feature descriptor methods characterise the feature points based on gradient direction, which can be indistinctive in local patches with retinal vessels as the patches lack of textural information and exhibit repetitive patterns.

## 1.3 RESEARCH QUESTIONS

Given the research problems presented in Section 1.2, the research questions constituted in this thesis are:

**RQ1:** What is the general characteristic of the retinal vessels in local patches of fundus image?

**RQ2:** What are the unique characteristics of the retinal vessels in local patches that set it apart from the noises?

**RQ3:** How to extract feature points from various sizes of retinal vessels?

**RQ4:** How to extract feature points that only located on retinal vessels and exclude those on noises?

**RQ5:** How to distinctively characterise the feature points to establish matches between images?

**RQ6:** How the proposed feature extraction method improves the performance of other feature extraction methods from existing feature-based RIR techniques in extracting feature points on retinal vessels?

**RQ7:** How the proposed feature-based RIR technique improves the performance of the existing feature-based RIR techniques in registering fundus images?

## 1.4 RESEARCH OBJECTIVES

The primary aim of this research is to propose a retinal image registration technique for fundus image. The primary aim is achieved with the following objectives:

**RO1:** To investigate the general and unique characteristics of the retinal vessels in local patches of fundus image.

**RO2:** To propose a feature extraction method based on the characteristics of the retinal vessels.

**RO3:** To propose a feature descriptor method that characterises the feature points based on distinctive information.

The objectives RO1 and RO2 are presented in Chapter 4: Feature Extraction, whereas objective RO3 is presented in Chapter 5: Feature Descriptor. The objectives RO1, RO2 and RO3 can be mapped to the research problems (Section 1.2) and research questions (Section 1.3) as provided in Table 1.1.

## 1.5 RESEARCH SCOPES AND LIMITATIONS

The scopes and limitations of this research are outlined as follows:

1) This research focuses on the feature-based RIR technique.
2) The modality of the retinal images in the datasets used during the experiments is limited to colour fundus images.
3) The target registration applications for the proposed feature-based RIR technique are limited to super-resolution, image mosaicking and longitudinal study applications.

## 1.6 RESEARCH CONTRIBUTIONS

The first main contribution of this thesis is to propose a novel feature extraction method for feature-based RIR technique. The existing feature extraction methods are mainly without a proper feature selection module to distinguish between retinal vessels and noises. Contrarily, the proposed feature extraction method is composed of feature detection and feature selection modules that consider the characteristics of the retinal vessels and noises. This allows the extraction of the feature points on retinal vessels and excludes those on noises with similar structure representation as retinal vessels.

The second main contribution is to propose a novel feature descriptor method for feature-based RIR technique. The existing works mainly utilised gradient direction surrounding the square region of the feature points as feature descriptor. In contrast, the proposed feature descriptor method describes the circular region of the feature points based on statistical properties to distinctively characterise the feature points on retinal vessels even in the presence of rotation between images.

The proposed feature extraction method and the proposed feature descriptor method are evaluated in registering fundus images from super-resolution, image mosaicking and longitudinal study applications. These registration applications are mainly performed in clinical settings. Thus, demonstrates the capability of the proposed work in a real-world application.

## 1.7 THESIS ORGANIZATION

This thesis is divided into six chapters. Chapter 2 gives a brief background related to retinal fundus image and introduces the existing feature-based RIR techniques. Chapter 3 explains the general framework of the proposed feature-based RIR technique. The proposed feature extraction method and its evaluation are presented in Chapter 4. Chapter 5 describes the proposed feature descriptor method. Also, the performance of the proposed feature-based RIR technique is evaluated and discussed in this chapter. Finally, Chapter 6 summarises and concludes the research findings.

**Table 1.1: Mapping between research problems (Section 1.2), research questions (Section 1.3) and research objectives (Section 1.4).**

| RESEARCH PROBLEMS | RESEARCH QUESTIONS | RESEARCH OBJECTIVES |
|---|---|---|
| CHAPTER 4: FEATURE EXTRACTION | | |
| **RP1**: The lack of existing studies examining the characteristics of the retinal vessels and noises in local patches of fundus image. | **RQ1**: What is the general characteristic of the retinal vessels in local patches of fundus image? | **RO1**: To investigate the general and unique characteristics of the retinal vessels in local patches of fundus image.<br>(Section 4.3 Characteristics of Retinal Vessels and Noises) |
| | **RQ2**: What are the unique characteristics of the retinal vessels in local patches that set it apart from the noises? | |
| **RP2**: Most of the existing feature extraction methods are unable to accurately extract feature points on retinal vessels of various sizes and without a proper selection module to accurately distinguish between retinal vessels and noises. | **RQ3**: How to extract feature points from various sizes of retinal vessels? | **RO2**: To propose a feature extraction method based on the characteristics of the retinal vessels.<br>(Section 4.4 Proposed Feature Extraction) |
| | **RQ4**: How to extract feature points that only located on retinal vessels and exclude those on noises? | |
| | **RQ6**: How the proposed feature extraction method improves the performance of other feature extraction methods from existing feature-based RIR | |

| RESEARCH PROBLEMS | RESEARCH QUESTIONS | RESEARCH OBJECTIVES |
|---|---|---|
| | techniques in extracting feature points on retinal vessels? | |
| CHAPTER 5: FEATURE DESCRIPTOR | | |
| **RP3**: Most of the existing feature descriptor methods characterise the feature points based on gradient direction, which can be indistinctive in local patches with retinal vessels as the patches lack of textural information and exhibit repetitive patterns. | **RQ5**: How to distinctively characterise the feature points to establish matches between images? | **RO3**: To propose a feature descriptor method that characterises the feature points based on distinctive information.<br>(Section 5.3 Proposed Feature Descriptor) |
| | **RQ7**: How the proposed feature-based RIR technique improves the performance of the existing feature-based RIR techniques in registering fundus images? | |

# CHAPTER 2: LITERATURE REVIEW

## 2.1 INTRODUCTION

This chapter presents the essential aspects of retinal fundus image registration (RIR). First, the modalities of retinal imaging that regularly used during the screening session are briefly introduced in Section 2.2. Then, the type of the retinal vessels found in fundus image and its appearance in pathology retina are explained in Section 2.3. Finally, the related works are highlighted in Section 2.4 according to the stages in the feature-based RIR technique.

## 2.2 RETINAL IMAGING

Fundus imaging and optical coherence tomography (OCT) are the most commonly used modalities in retinal imaging for the detection and management of ocular and systemic diseases (Abràmoff et al., 2010). These modalities are briefly introduced in the following sub-sections.

### 2.2.1 Fundus Imaging

Fundus image is acquired using an ophthalmoscope or a fundus camera to depict the back of an eye in 2-dimensional. There are three primary structures of the retina visualised in fundus image, namely, optic disc, retinal vessels and macula as shown in Figure 2.1. The optic disc is circular in shape and appears as clearer and brighter than the surrounding region. The network of the retinal vessels is distributed throughout the fundus image and converged at the optic disc. The retinal vessels are diverse in terms of size and contrast. The macula is a dark and an oval-shaped region located near the centre of the retina where the vessels are scarce.

The specifications of the fundus camera, such as resolution and field of view (FOV) varies according to the requirement of the tasks or diagnosis. For example, a high-resolution image is required to perform vascular measurements (Pauli, Gangaputra,

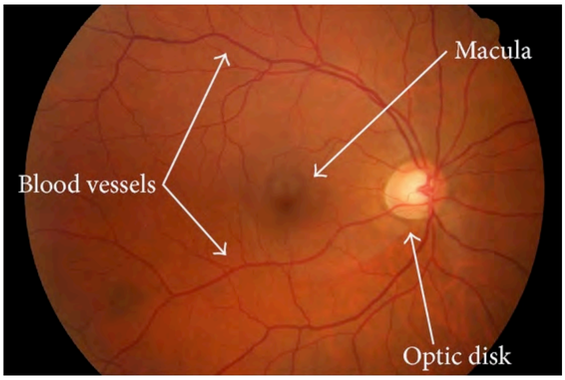



**Figure 2.1: The primary structures of the retina visualised in fundus image (Budai, Bock, Maier, Hornegger, & Michelson, 2013).**

Hubbard, Thayer, Chandler, Peng, Narkar, Ferrier, & Danis, 2012). The ranges of FOV or also known as external angle view of the fundus camera are often reported from 35° to 60° in the literature for the diagnosis of various eye conditions (Bernardes, Serranho, & Lobo, 2011). The screening program for disease documentation and clinical studies frequently utilise fundus image due to the sensitivity and specificity of the modality. Furthermore, fundus imaging is safe, feasible and cost-effective, which make it a realistic screening approach. Despite these advantages, fundus imaging presents three main issues describe below.

First, acquiring a clear fundus image with crisp edges of anatomical feature relies on cooperation from the patient and a proper imaging setup by the ophthalmic photographer (Ryan, Sadda, Hinton, Schachat, Wilkinson, & Wiedemann, 2012). Second, fundus image has a limited FOV of the retina. A fundus image with a smaller FOV covers a smaller part of the retina with minimal distortion. Thus, multiple images are required to be captured at different viewing angles to fully visualise the retina. In opposite, a fundus image with a larger FOV covers a larger part of the retina, but at the cost of distorting the image due to the spherical shape of the retina (DeHoog & Schwiegerling, 2009). Third, the fundus images acquired at different screening sessions often result in a slight variation of the viewpoint on the retina (Noyel, Thomas, Bhakta, Crowder, Owens, & Boyle, 2017). This


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

**Figure 2.2: Types of fundus image (a) colour fundus image (b) fluorescein angiography (Katherine Hu, 2017) (c) single-wavelength SLO (Bennett, 2019) and (d) true colour SLO (Optomap, 2015).**

variation can be challenging for the ophthalmologist to examine and compare the fundus images in monitoring the progression of retinal diseases or treatments. Therefore, fundus image registration is employed to assist the ophthalmologist in managing these issues.

There are several types of fundus images such as colour fundus imaging, fluorescein angiography and scanning laser ophthalmoscope (SLO) as depicted in Figure 2.2. The colour fundus image consists of a full-colour image in red, green and blue composites (see Figure 2.2(a)). The image is captured by illuminating the back of the retina with white light through cornea, pupil and lens (Abràmoff et al., 2010). The colour fundus imaging has been employed in the majority of the screening program for detecting cataracts, glaucoma, diabetic retinopathy and age-related macular degeneration (Khouri, Szirth, Salti, & Fechtner, 2007; Pirbhai, Sheidow, & Hooper, 2005; Yogesan, Constable, Eikelboom, & van Saarloos, 1998).

The fluorescein angiography captures the fluorescence emitted by a contrast agent when the retina is illuminated with blue light (see Figure 2.2(b)) (Johnson, Fu, McDonald, Jumper, Ai, Cunningham, & Lujan, 2013). The emission of the fluorescence allows the observation of the pathological changes such as vessel occlusion, leakage, diabetic retinopathy or tumours. However, the fluorescein angiography is an invasive technique where the contrast agent is administered into the systemic circulation. Administering the

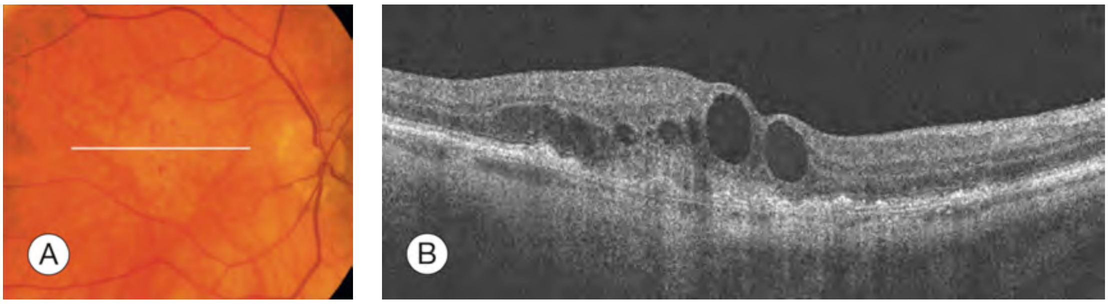


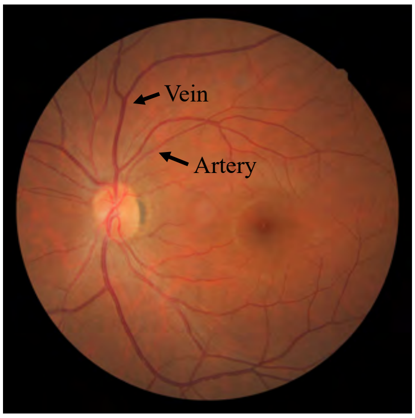

**Figure 2.3: Example of (A) Colour fundus image (B) OCT image of cross-sectional region on foveal indicated by the horizontal white line in (A). These images are obtained from (de Amorim Garcia Filho, Yehoshua, Gregori, Puliafito, & Rosenfeld, 2013).**

contrast agent into the systemic circulation can result in adverse effects such as nausea, vomiting and allergic reaction.

The SLO uses a confocal imaging technique to acquire a monochromatic image (see Figure 2.2(c)) or a true colour image of the retina (see Figure 2.2(d)). The monochromatic image is produced when a single-wavelength laser is used while the true colour image is produced when the images acquired using blue, green, and red lasers are combined (Fleckenstein, Schmitz-Valckenberg, & Holz, 2013). The SLO is capable of visualising retinal structures at a fined detailed and high contrast. The SLO has been used in diagnosing glaucoma and macular degeneration.

### 2.2.2 Optical Coherence Tomography (OCT)

OCT is based on low-coherent light interferometry to produce a cross-sectional image of the targeted area on the retina. The cross-sectional image visualises distinctive layers of the targeted area as shown in Figure 2.3. OCT generally have better precision than fundus image in diagnosing diabetic macular edema, age-related macular degeneration and glaucoma (Walsh, Wildey, Lara, Ouyang, & Sadda, 2010). However, OCT has a limited field of view, requires a highly trained ophthalmic photographers to operate the equipment, a long acquisition time and the cost of the equipment is substantially higher than a colour fundus camera (de Amorim Garcia Filho et al., 2013; Podoleanu, 2012;



(a)

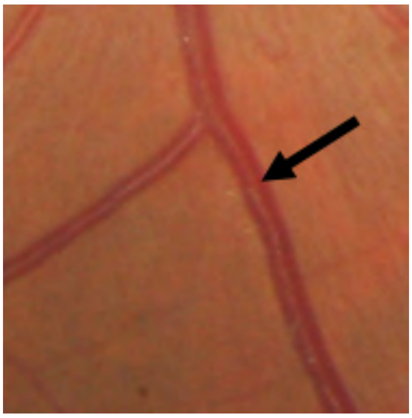

(b)

**Figure 2.4: Example of (a) artery and vein, (b) central light reflex.**

Spaide, Fujimoto, Waheed, Sadda, & Staurenghi, 2018; Wei, Jia, Tan, Potsaid, Liu, Choi, Fujimoto, & Huang, 2013). For these reasons, the utilisation of the OCT in the clinical setting remains limited.

## 2.3 RETINAL VESSELS

The retinal vessels can be divided into arteries and veins as shown in Figure 2.4(a). The arteries are brighter in colour compared to veins because the blood rich in oxygen is transported through the arteries. In opposite, the veins are darker because the blood transported is lacked in oxygen. Furthermore, the veins are larger and central light reflex is more visible in the veins than the arteries. The central light reflex is the bright region or "silver wire" appearance at the centre part of the veins as shown in Figure 2.4(b).

In the pathological retina, the abnormalities of the retinal vessels in fundus image can be observed in terms of thickened (increased in width), narrowed (decreased in width) and tortuosity (twist and turns) (MacGillivray, Trucco, Cameron, Dhillon, Houston, & van Beek, 2014; Rousso & Sowka, 2017). These abnormalities are vital diagnostic information to identify and manage systemic diseases during clinical examination. The thickened and tortuous retinal veins typically indicate the vein occlusion where the blood flow within the vein is obstructed. The blood flow obstruction can build up the pressure within the vein and potentially lead to leakage. This condition can be caused by diabetes, hypertension, cardiovascular disease, hyperviscosity states, arteriosclerosis, collagen vascular disease and sickle cell disease. The thickened and non-tortuous retinal veins can be observed when the blood flow to the retinal veins are decreased due to atherosclerosis, giant cell arteritis and fibrovascular dysplasia. Retinal arteriole macroaneurysms caused by systemic hypertension present the retinal arteries as thickened, whereas the tortuous retinal arteries may indicate the systemic vascular diseases. The narrowed appearance of the retinal arteries can occur due to atherosclerosis and hypertension.

## 2.4 RELATED WORKS

As previously introduced in Section 1.1, the feature-based RIR technique is generally comprised of five main stages; pre-processing, feature extraction, feature descriptor, matching and estimating geometrical transformation. In this section, the existing feature-based RIR techniques that best relate to our work are summarised in Table 2.1, Table 2.2, Table 2.3 and Table 2.4 according to their main contributions either in feature extraction, feature descriptor, matching and geometrical transformation stages, respectively.

### 2.4.1 Feature Extraction

The feature extraction in the existing feature-based RIR techniques mainly based on scale-invariant feature transform (SIFT) detector (Lowe, 2004), a popular feature extraction method in the computer vision field. The initial extraction process of SIFT detector involves finding extrema within local patches throughout the hierarchical difference of Gaussian (DoG) scale space. A hierarchical scale space is a representation of an image over a large range of scales in a “pyramid-like” approach, which allows feature points to be found on structure of various sizes. Therefore, the utilisation of the hierarchical DoG scale space in SIFT detector allows the feature points to be found on retinal vessels of various sizes.

Then, the extrema that are low contrast and located on the edges are rejected to ensure the final feature points are distinctive and repeatable. However, the contrast of the retinal vessels within a fundus image are varied from low to high and inconsistent between images due to the variation of the image quality. Consequently, SIFT detector may mistakenly reject the extrema on retinal vessels.

To overcome the issue above, Ghassabi, Shanbehzadeh, Sedaghat, and Fatemizadeh (2013) proposed the utilisation of uniform robust scale-invariant feature transform (UR-SIFT) (Sedaghat, Mokhtarzade, & Ebadi, 2011) in extracting feature points on retinal vessels. UR-SIFT improves the standard SIFT detector by ensuring the extracted feature points are distributed throughout the hierarchical DoG scale space. The distribution is set in a decreasing manner or reverse from the scale coefficients of the scale space. Specifically, more feature points are extracted in the lower part of the hierarchical DoG scale space where the images are larger and finer. In opposite, fewer feature points are extracted in the upper part of the hierarchical DoG scale space where the images are smaller and coarser. Other than that, UR-SIFT selects feature points according to the strength of the texture surrounding the points. This approach enables UR-SIFT to be more efficient in extracting feature points on retinal vessels compared to the standard SIFT detector. Their work is further improved in (Ghassabi, Shanbehzadeh, Mohammadzadeh, & Ostadzadeh, 2015) by introducing stability score as part of the selection criterion. The stability score incorporates Frangi's vesselness measure (FVM) (Frangi, Niessen, Vincken, & Viergever, 1998), a vessel enhancement filter that suppresses noises in the medical image. By incorporating FVM as part of the selection criterion, Ghassabi et al. (2015) increases the ability of the method to discriminate between noises and retinal vessels.

Extracting feature points on retinal vessels from the underexposed region due to non-uniform intensity distribution is addressed in (Ramli, Idris, Hasikin, & A. Karim, 2017a). In their work, the illumination invariant Difference of Gaussian (*ii*DoG) operator is incorporated in the hierarchical scale space (Vonikakis, Chrysostomou, Kouskouridas, & Gasteratos, 2013). The *ii*DoG operator is composed of normalised difference of Gaussian (*n*DoG) and DoG operators based on piecewise function. The combination of these operators increases the visibility of the underexposed region while left the correctly exposed region unchanged in the hierarchical scale space. This work utilised a similar approach as in SIFT detector to extract extrema from the hierarchical *ii*DoG scale space. Then, a threshold is introduced to discard the extrema on the retinal surface before being selected as feature points. The threshold is set according to the distribution of the pixels value in the local patch of the extrema.

Another approach to improve the extraction of the feature points on retinal vessels is presented in (S. K. Saha et al., 2016; Sajib K. Saha, Xiao, Frost, & Kanagasingam, 2018). The approach involves detecting the feature points based on invariant feature all over the fundus image using Harris corner (Harris & Stephens, 1988), SIFT, speeded up robust features (SURF) (Bay et al., 2008; Bay et al., 2006), binary robust invariant scalable keypoints (BRISK) (Leutenegger, Chli, & Siegwart, 2011) and UR-SIFT methods. Then, the feature points on retinal vessels are identified using the binary mask representing the vasculature map of the retina. The vasculature map is obtained by segmenting the retinal vessels according to (Nguyen, Bhuiyan, Park, & Ramamohanarao, 2013).

Ramli, Idris, Hasikin, A. Karim, Abdul Wahab, Ahmedy, Ahmedy, Kadri, and Arof (2017b) introduced D-Saddle detector to extract feature points on retinal vessels from the low-quality region. The feature points are extracted based on structural information. Other feature points employed in the existing feature-based RIR techniques (Addison

Lee, Cheng, Hai Lee, Ping Ong, Xu, Wing Kee Wong, Liu, Laude, & Han Lim, 2015; Hernandez-Matas, Zabulis, & Argyros, 2017a; S. K. Saha et al., 2016; Sajib K. Saha et al., 2018; Zheng, Daniel, Hunter, Xiao, Gao, Li, Maguire, Brainard, & Gee, 2014) include geometric corner (J. A. Lee, Lee, Xu, Ong, Wong, Liu, & Lim, 2014), vascular-landmark (Can, Stewart, Roysam, & Tanenbaum, 2002), bifurcation and cross-over points. The issues in the existing feature extraction methods highlighted above will be discussed in Section 4.2.

### 2.4.2 Feature Descriptor

The existing feature-based RIR techniques summarised in Table 2.1, Table 2.2, Table 2.3 and Table 2.4 characterise the feature points with gradient direction descriptor and binary descriptor. For the registration of single modality between colour fundus images, the feature descriptor methods such as SIFT, SURF, BRISK, binary robust independent elementary features (BRIEF) (Calonder, Lepetit, Ozuysal, Trzcinski, Strecha, & Fua, 2012), histogram of oriented gradients (HOG) (Dalal & Triggs, 2005) and an efficient binary descriptor based on Haar features (ALOHA) (S. Saha & Démoulin, 2012) are employed without any improvement or modification. Among these methods, SIFT descriptor is highly employed in the existing feature-based RIR techniques.

SIFT descriptor is a local descriptor where the square local patch is aligned to the dominant orientation around the given feature point to achieve rotation invariance. The dominant orientation is determined from the histogram of 36 bins where the angular resolution is 10°. Then, the orientation of the gradient direction for each feature points is computed at the respective Gaussian scale space level to form a descriptor of 128 elements.

The ambiguity of SIFT descriptor can be probable in the presence of nonlinear illumination changes that often observe between multi-modality retinal images.

Accordingly, Tsai, Li, Yang, and Lin (2010) proposed a combination of SIFT descriptor with shape context descriptor to improve the distinctiveness of SIFT descriptor between multi-modality retinal images.

Another issue that can be observed between multi-modality retinal images is the inverse intensity. Therefore, J. Chen, Tian, Lee, Zheng, Smith, and Laine (2010) proposed partial intensity invariant feature descriptor (PIIFD), which inspired by SIFT descriptor. The PIIFD aligns the square local patch to the dominant orientation around the given feature point to achieve rotation invariance. The dominant orientation is computed by averaging squared gradients. Then, the orientation of the gradient direction for each feature point is computed as in SIFT descriptor, but normalised between 0 to $\pi$. This approach allows the corresponding feature points with opposite gradient orientation being characterised with a similar descriptor of 128 elements.

Other than SIFT-based descriptor, Addison Lee et al. (2015) proposed low-dimensional step pattern analysis (LoSPA) to characterise the feature points with patterns of gradient direction in multi-modality retinal images. The LoSPA offers a low dimensional descriptor either with 58 elements or 86 elements. This low dimensional descriptor is capable of speeding up the registration process without compromising the distinctiveness and effectiveness of the descriptor. The issues in the existing feature descriptor methods highlighted above will be discussed in Section 5.2.

### 2.4.3 Matching

The matching process establishes correspondences or matches between feature points in fixed and moving images. The process begins by establishing putative matches where it may include ambiguous matches or outliers. Then, the ambiguous matches or outliers are removed with outlier removal algorithm.

The putative matching method in the existing feature-based RIR techniques as outlined in Table 2.1, Table 2.2, Table 2.3 and Table 2.4 can be grouped into two methods of proximity search namely, minimum distance and nearest-neighbour search (NNS). The minimum distance method establishes the putative matches according to the minimum function of a pairwise distance matrix. The pairwise distance matrix is exhaustively computed between the descriptors in the fixed and moving images using a distance metric such as Euclidean distance (Sajib K. Saha et al., 2018). Multiple distance metrics such as Euclidean with Hamming distances (S. K. Saha et al., 2016) and Euclidean with chi-squared distances (Tsai et al., 2010) are reported to compute the pairwise distance when multiple feature descriptor methods are employed. The approach to compute the pairwise distance matrix using multiple distance metrics are differed between these two studies. Specifically, S. K. Saha et al. (2016) compute a separate pairwise distance matrix for each SIFT and BRIEF descriptors using Euclidean and Hamming distances, respectively. In opposite, Tsai et al. (2010) compute the pairwise distance matrix from the summation of the Euclidean and chi-squared distances for SIFT and shape context descriptors. The advantages of the minimum distance method are it systematically enumerate all possible candidates for the putative matches (from the exhaustive search) and its simplicity as no initialisation of the threshold is required (from selecting match according to minimum function). However, the minimum distance method often results in the putative matches with a high number of outliers.

Therefore, Ghassabi et al. (2013) employed bilateral minimum distance method in their work to reduce the number of outliers in the putative matches. The bilateral minimum distance method considers the minimum function of a pairwise distance matrix that satisfies in both directions for the selection of putative matches. For example, a match is identified when its distances are minimum from fixed to moving images and vice versa.

The putative matching method based on NNS is commonly employed in the existing feature-based RIR techniques. Mainly, Lowe's nearest-neighbour search (LNNS) (Lowe, 2004) is preferred among the existing feature-based RIR techniques (J. Chen et al., 2010; Ghassabi et al., 2015; Hernandez-Matas, Zabulis, & Argyros, 2015, 2016; Hernandez-Matas, Zabulis, Triantafyllou, Anyfanti, & Argyros, 2017b; Ma, Jiang, Liu, & Li, 2017; Ramli et al., 2017a; Ramli et al., 2017b; Tang, Pan, Yang, Yang, Luo, Zhang, & Ong, 2018; Wang, Wang, Chen, & Zhao, 2015; Yang, Stewart, Sofka, & Tsai, 2007). The LNNS establishes the putative matches according to the ratio between the best match (minimum distance) to the second-best match (second minimum distance). If the ratio which yields a value between $[0,1]$ is below the desired threshold, the respective pair of the feature point is accepted as a match. The ratio that is larger than the defined threshold indicates that the match is ambiguous or outliers and will be discarded. This reduces the amount of outliers in the putative matches compared to the minimum distance method. Among the metrics employed to compute the distances are the sum of squared distance (SSD) and Euclidean distance. The number of outliers in the putative matches of LNNS is further reduced in (J. Chen et al., 2010; Ghassabi et al., 2015; Tang et al., 2018; Wang et al., 2015) by implementing the bilateral approach. The bilateral LNNS selects the matches according to the ratio that satisfies the defined threshold in both directions from fixed to moving images and vice versa.

In the existing feature-based RIR techniques, the LNNS is estimated using exhaustive search, best-bin-first (BBF) (Beis & Lowe, 1997) and fast approximate nearest neighbour search with automatic algorithm configuration (Fast-NNS) (Muja & Lowe, 2009). The exhaustive search can provide the exact nearest neighbours but, it can be computationally expensive when a high quantity and high dimensional spaces of the descriptors are involved. Alternatively, the approximation algorithm of BBF and Fast-NNS can speed the process and returns the closest neighbour with high probability.

The NNS approximated by k-d tree (Bentley, 1975) with Euclidean distance is employed in (Addison Lee et al., 2015). The approximation using k-d tree can be fast compared to the exhaustive search if the dimensional spaces of the descriptor remain below than 10 (Lowe, 2004). Other than minimum distance and NNS methods, Zheng et al. (2014) proposed the softassign strategy (Chui & Rangarajan, 2003) and enforcing sparsity properties to perform the matching process.

The outliers removal algorithms based on geometric verification in the existing feature-based RIR techniques (Addison Lee et al., 2015; J. Chen et al., 2010; Ghassabi et al., 2013; Ramli et al., 2017a; Ramli et al., 2017b) eliminates the outliers by examining the geometrical constraint of the matches in affine transformation. Example of the geometric verification based algorithms employed are RANdom SAmple Consensus (RANSAC) (Fischler & Bolles, 1981) and M-estimator SAmple and Consensus (MSAC) (Torr & Zisserman, 2000). The geometric verification based algorithms identify a match as outliers if the distance between the match from the fixed and projected moving images exceeding a defined threshold. In the case of large viewpoint difference between fundus images such as small overlapping area, the affine transformation unable to model an accurate alignment between the images due to the curvature of the eyeball. Consequently, the correct match may yield a large projected distance than the defined threshold and mistakenly eliminated. Therefore, Ramli et al. (2017a) and Ramli et al. (2017b) employed multiple values for the distance threshold to minimise the issue.

To address the limitation of the geometric verification based algorithm in curved object, Wang et al. (2015), Ma et al. (2017) and Tang et al. (2018) proposed the improvements of the robust point matching method. Wang et al. (2015) proposed a single Gaussian robust point matching model. The correct matches are assumed to satisfy a single Gaussian distribution where optimal mapping function is searched in reproduced kernel Hilbert space with a Gaussian radial basis kernel function. Ma et al. (2017) generalize the Gaussian mixture model (GMM) formulation with the descriptors of the putative matches while Tang et al. (2018) proposed the combination of mixture feature and structure preservation.

### 2.4.4 Geometrical Transformation

There are a variety of transformation models employed in the existing feature-based RIR techniques such as similarity, affine, reduced quadratic, second-order polynomial and non-rigid models. The low-order models such as similarity and affine require a minimum of three matches to perform the registration, but these models only emulate a limited number of transformations to register the images. For example, the similarity model includes rotation, translation and scaling, whereas the affine model includes rotation, translation, scaling and shearing.

As the retina is spherical in shape, a higher-order transformation model such as reduced quadratic and second-order polynomial models offer better flexibility to project the curved object. These models require a minimum of six matches to perform the registration. The higher-order transformation models are sensitive in the presence of the outliers, wherein the outliers can drive the model to estimate inaccurate transformation. Therefore, the removal of the outliers is mainly performed in the existing feature-based RIR techniques that utilised a high-order transformation model as described in Section 2.4.3.

Typically, multiple transformation models are employed in the existing feature-based RIR techniques. The works presented in (J. Chen et al., 2010; Ghassabi et al., 2015; Ramli et al., 2017b; Wang et al., 2015) performed the registration using similarity, affine and second-order polynomial models. The model is chosen according to the number of matches available. In opposite, a hierarchy transformation model of generalized dual-bootstrap iterative closest point (GDB-ICP) was proposed in (Yang et al., 2007). GDB-ICP performed the registration by applying three transformation models starting from low-order to higher-order models. These models are similarity, reduced quadratic and second-order polynomial models.

To address the issues of projecting the curved object, thin-plate spline (TPS) and non-rigid models are employed in (Ma et al., 2017; Tang et al., 2018; Wang et al., 2015; Zheng et al., 2014). Hernandez-Matas et al. (2015) proposed a 3-dimensional transformation model, which considers the spherical shape of the retina. This work is further improved in (Hernandez-Matas et al., 2016; Hernandez-Matas et al., 2017b) by adding and improving the initialisation of the pose estimation to their earlier work.

## 2.5 CHAPTER SUMMARY

This chapter described the essential aspects of retinal image registration such as the modalities of the retinal imaging and the primary structures of the retina found in fundus image. The types of the retinal vessel and its abnormalities in pathology retina are also described in Section 2.3. Furthermore, this chapter highlighted the existing feature-based RIR techniques that best relate to our work. These techniques were highlighted according to the stages in the feature-based RIR technique.

**Table 2.1: Existing feature-based retinal image registration techniques with main contribution on feature extraction.**

| Author(s) | Input Image | Pre-processing | Feature Extraction | Feature Descriptor | Matching | Geometrical Transformation |
|---|---|---|---|---|---|---|
| Sajib K. Saha et al. (2018) | Colour fundus | Not reported | Compare performance:<br>- Harris corner, SIFT, SURF, BRISK, UR-SIFT on retinal vessels<br>- Bifurcation with cross-over points | SIFT, SURF, BRISK, BRIEF, ALOHA and PIIFD | - PM: Minimum distance | - Second order polynomial |
| Ramli et al. (2017b) | Colour fundus | Convert input images to grayscale | Proposed D-Saddle detector to extract feature points from the low-quality region | HOG | - PM: Fast-NNS<br>- DM: SSD<br>- OR: MSAC | - Similarity<br>- Affine<br>- Second order polynomial |
| Ramli et al. (2017a) | Colour fundus | Convert input images to grayscale | Proposed iiDoG-SIFT detector with Histogram Threshold to extract feature points on retinal vessels | HOG | - PM: LNNS (Exhaustive)<br>- DM: SSD<br>- OR: MSAC | - Similarity |
| Hernandez-Matas et al. (2017a) | Colour fundus | Not reported | Compare performance:<br>- SIFT<br>- SURF<br>- Harris<br>- Bifurcations | - SIFT<br>- SURF<br>- PIIFD<br>- SIFT | Not reported | - 3D transformation model (considers spherical shape of retina[3)] |
| S. K. Saha et al. (2016) | Colour fundus | Not reported | Proposed two step registration:<br>- SURF points on retinal vessels (use vasculature map)[1]<br>- Bifurcation points[2] | - SURF<br>- BRIEF | - PM: Minimum distance<br>- DM: Euclidean and Hamming | - Affine<br>- Second order polynomial |

| Author(s) | Input Image | Pre-processing | Feature Extraction | Feature Descriptor | Matching | Geometrical Transformation |
|---|---|---|---|---|---|---|
| Ghassabi et al. (2015) | Colour fundus | Not reported | Proposed UR-SIFT detector with FVM to extract feature points on retinal vessels | SIFT | - PM: Bilateral LNNS<br>- DM: Euclidean<br>- OR: Graph transformation matching (GTM) | - Similarity<br>- Affine<br>- Second order polynomial |
| Ghassabi et al. (2013) | - Red-free fundus<br>- Auto-fluorescence<br>- Fluorescein angiographic | Not reported | Proposed using UR-SIFT from (Sedaghat et al., 2011) to extract feature points on retinal vessels | PIIFD | - PM: Bilateral minimum distance<br>- DM: Euclidean<br>- OR: Geometric verification | - Second order polynomial |

PM : Putative matching method
DM : Distance metric
OR : Outliers removal algorithm
[1](Nguyen et al., 2013), [2](L. Chen, Xiang, Chen, & Zhang, 2011), [3](Hernandez-Matas et al., 2017b)

**Table 2.2: Existing feature-based retinal image registration techniques with main contribution on feature descriptor.**

| Author(s) | Input Image | Pre-processing | Feature Extraction | Feature Descriptor | Matching | Geometrical Transformation |
|---|---|---|---|---|---|---|
| Addison Lee et al. (2015) | - Colour fundus<br>- Fluorescein angiographic | Not reported | Geometric corner[1] | Proposed low-dimensional step pattern analysis (LoSPA) | - PM: k-d tree<br>- DM: Euclidean<br>- OR: RANSAC | - Affine |
| J. Chen et al. (2010) | - Red-free fundus<br>- Auto-fluorescence<br>- Infrared | - Convert input images to grayscale<br>- Scale input image to the full 8-bit intensity range<br>- Scale images to a fixed size (1000×1000 pixels) | Harris corner | Proposed PIIFD | - PM: Bilateral LNNS (BBF) OR: Use main orientations of feature point candidates' and geometrical distribution of matches | - Similarity<br>- Affine<br>- Second order polynomial |
| Tsai et al. (2010) | - Red-free fundus<br>- Fluorescein angiographic | Not reported | SIFT | Proposed SIFT with shape context descriptor | - PM: Minimum distance<br>- DM: The sum of Euclidean and chi-squared distances | Edge-Driven DB-ICP Hierarchy model:<br>- Similarity<br>- Reduced quadratic<br>- Second order polynomial |

PM : Putative matching method
DM : Distance metric
OR : Outliers removal algorithm
[1](J. A. Lee et al., 2014)

**Table 2.3: Existing feature-based retinal image registration techniques with main contribution on matching.**

| Author(s) | Input Image | Pre-processing | Feature Extraction | Feature Descriptor | Matching | Geometrical Transformation |
|---|---|---|---|---|---|---|
| Tang et al. (2018) | Multi-modality retinal images | Not reported | SURF | PIIFD | - PM: Bilateral LNNS (BBF)<br>- OR: Proposed the combination of mixture feature and structure preservation | - TPS model |
| Ma et al. (2017) | - Red-free fundus<br>- Fluorescein angiographic | - Equalize the intensity histogram<br>- Denoise with a non-local mean filter | - Edge map<br>- SIFT | SIFT | - PM: LNNS<br>- OR: Proposed feature guided GMM | - Non-rigid model |
| Wang et al. (2015) | - Red-free fundus<br>- Auto-fluorescence | - Obtain green channel of input images<br>- Scale input image to the full 8-bit intensity range | SURF | PIIFD | - PM: Bilateral LNNS (BBF)<br>- OR: Proposed single Gaussian robust point matching model | - Similarity<br>- Affine<br>- Second order polynomial |
| Zheng et al. (2014) | - Colour fundus<br>- Angiogram | Not reported | - Vascular-landmark[1] | - Reinforced self-similarities (SS) descriptor | - Proposed the used of both softassign strategy and enforcing sparsity properties | - TPS model |

PM : Putative matching method
DM : Distance metric
OR : Outliers removal algorithm
[1](Can et al., 2002)

**Table 2.4: Existing feature-based retinal image registration techniques with main contribution on geometrical transformation.**

<table>
<tr><th>Author(s)</th><th>Input Image</th><th>Pre-processing</th><th>Feature Extraction</th><th>Feature Descriptor</th><th>Matching</th><th>Geometrical Transformation</th></tr>
<tr><td>Hernandez-Matas et al. (2017b)</td><td>Colour fundus</td><td>Not reported</td><td>SIFT</td><td>SIFT</td><td>- PM: LNNS<br>- DM: Euclidean</td><td rowspan="2">Add initialization of pose estimation to (Hernandez-Matas et al., 2015), a 3D transformation model that considers spherical shape of retina</td></tr>
<tr><td>Hernandez-Matas et al. (2016)</td><td>Colour fundus</td><td>Not reported</td><td>SIFT</td><td>SIFT</td><td>Not reported</td></tr>
<tr><td>Hernandez-Matas et al. (2015)</td><td>Colour fundus</td><td>Obtain green channel</td><td>SURF</td><td>SURF</td><td>- PM: LNNS<br>- DM: Euclidean</td><td>Proposed 3D transformation model that considers spherical shape of retina</td></tr>
<tr><td>Yang et al. (2007)</td><td>- Red-free fundus<br>- Fluorescein angiographic</td><td>Not reported</td><td>SIFT</td><td>SIFT</td><td>- PM: LNNS</td><td>Proposed GDB-ICP (refine using corner and face points). Hierarchy model:<br>- Similarity<br>- Reduced quadratic<br>- Second order polynomial</td></tr>
</table>

PM : Putative matching method
DM : Distance metric

# CHAPTER 3: OVERVIEW OF METHODOLOGY

## 3.1 INTRODUCTION

This chapter describes the general methodology of the proposed feature-based retinal fundus image registration (RIR) technique in five sections. First, the system requirements in developing the proposed work are described in Section 3.2. Then, the general framework of the proposed feature-based RIR technique is explained in Section 3.3. The overview of the pre-processing, proposed feature extraction method and proposed feature descriptor method are presented in the remaining sections.

## 3.2 SYSTEM REQUIREMENTS

The proposed feature-based RIR technique was developed and implemented in Matlab R2016b with Image Processing toolbox, Computer Vision toolbox and Signal Processing toolbox. The MATLAB R2016b is running on a virtual machine from Google Cloud Engine with specifications of Intel Xeon® E5 2.6GHz (24 vCPUs) and 40GB of RAM.

## 3.3 GENERAL FRAMEWORK

A general framework of the proposed feature-based RIR technique comprises of five main stages as shown in Figure 3.1. These stages are pre-processing, feature extraction, feature descriptor, matching and estimating geometrical transformation. The pre-processing stage converts the input fundus images to the grayscale images. The second stage of the proposed feature extraction method detects and selects feature points on retinal vessels throughout fixed and moving images. The proposed feature descriptor method in the third stage computes the descriptor to describe the surrounding circular region for each feature point. The fourth stage finds the putative matches and removes the outliers between the feature points on fixed and moving images. Finally, the inliers from the putative matches are used to estimate the geometrical transformation between

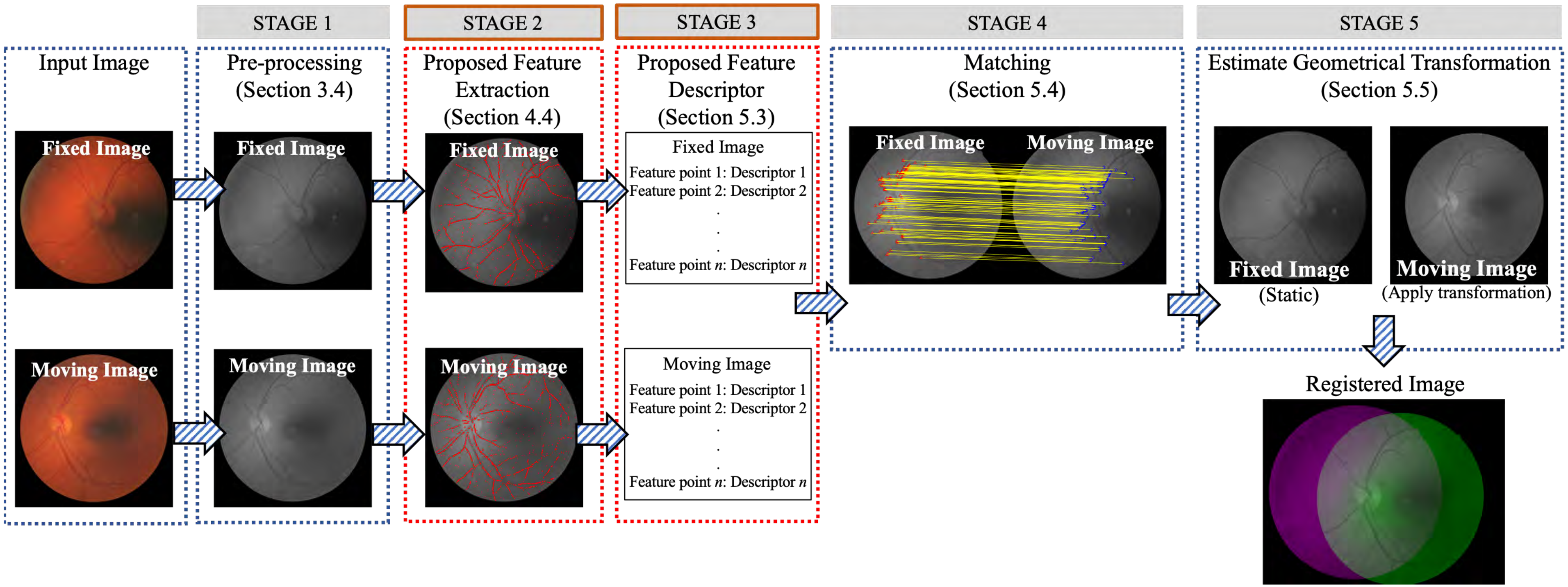



**Figure 3.1: General framework of the proposed feature-based retinal image registration technique. The red boxes highlight the contributions of this thesis.**

the images. The estimated geometrical transformation is applied to the moving image, where the moving image is aligned to the orientation of the fixed image.

## 3.4 PRE-PROCESSING

The proposed feature-based RIR technique processes the input fundus images in grayscale. The conversion of the input images from red, green and blue composites (RGB) to the grayscale is performed through the summation of the weighted R, G, and B components. The weightages for the R, G, B components are according to the calculation of luminance in Recommendation ITU-R BT.601-7 (International Telecommunication Union, 2011) as follows:

$$image_{grayscale} = 0.2989R + 0.587G + 0.1140B \tag{3.1}$$

where, $R$ is the red channel, $G$ is the green channel and $B$ is the blue channel.

## 3.5 FEATURE EXTRACTION

The main aim of the proposed feature extraction method is to extract feature points on retinal vessels from fundus image. The proposed feature extraction method is evaluated in terms of its accuracy to extract feature points on retinal vessels. Its performance is compared with five existing feature extraction methods, namely, Harris corner (Harris & Stephens, 1988), SIFT (Lowe, 2004), SURF (Bay et al., 2008; Bay et al., 2006), Ghassabi's (Ghassabi et al., 2015) and D-Saddle (Ramli et al., 2017b). These feature extraction methods were previously employed in the existing feature-based RIR techniques as highlighted in Section 2.4.

All the feature extraction methods are evaluated on four public datasets containing fundus images with various pathological cases; CHASE_DB1 (*CHASE_DB1 Retinal Image Database*; Fraz, Remagnino, Hoppe, Uyyanonvara, Rudnicka, Owen, & Barman, 2012), DRIVE (*DRIVE: Digital Retinal Images for Vessel Extraction*; Staal, Abràmoff,

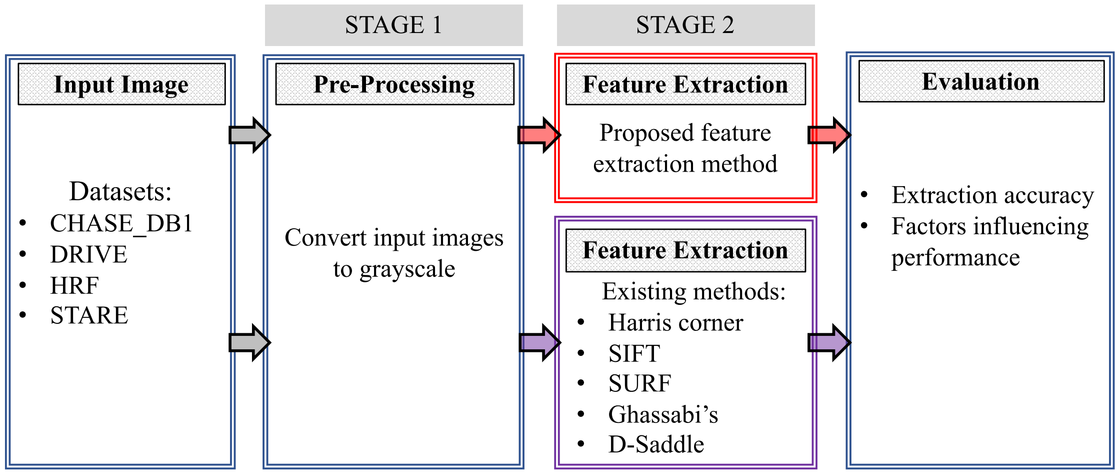



**Figure 3.2: An overview of the work presented in Chapter 4: Feature Extraction.**

Niemeijer, Viergever, & Van Ginneken, 2004), HRF (Budai et al., 2013; *HRF: High-Resolution Fundus Image Database*) and STARE (Hoover, Kouznetsova, & Goldbaum, 2000; *STARE: Structured Analysis of the Retina*). These feature extraction methods are evaluated in terms of extraction accuracy of the feature points on the retinal vessels. Factors influencing the performance of the feature extraction methods are also investigated and discussed. The details of the proposed feature extraction method and its evaluation are presented in Chapter 4: Feature Extraction. The work presented in this chapter can be summarised as shown in Figure 3.2.

## 3.6 FEATURE DESCRIPTOR

The proposed feature descriptor characterises the circular region of the feature points based on distinctive information. For the evaluation, the proposed feature-based RIR technique is evaluated, which includes five main stages of pre-processing, proposed feature extraction as presented in Chapter 4, proposed feature descriptor, matching and estimating geometrical transformation. The performance of the proposed feature-based RIR technique is compared with five existing feature-based RIR techniques; GDB-ICP (Yang et al., 2007), Harris-PIIFD (J. Chen et al., 2010), Ghassabi's-SIFT (Ghassabi et al., 2015), H-M 16 (Hernandez-Matas et al., 2016), H-M 17 (Hernandez-Matas et al., 2017a) and D-Saddle-HOG (Ramli et al., 2017b). All the feature-based RIR techniques are evaluated on Fundus Image Registration (FIRE) dataset (Hernandez-Matas et al., 2017c). Their registration performance is measured in terms of target registration error (TRE). Factors influencing the registration performance for these techniques are also assessed and discussed. The details of the proposed feature descriptor method and the evaluation of the proposed feature-based RIR technique are presented in Chapter 5: Feature Descriptor. The work presented in this chapter can be summarised as shown in Figure 3.3.

## 3.7 CHAPTER SUMMARY

The system requirements in developing and evaluating the proposed feature-based RIR technique were described in this chapter. Then, the general methodology of the proposed feature-based RIR technique was presented. The datasets and benchmark methods that used during the evaluation of the proposed feature extraction method and the proposed feature-based RIR technique were also outlined.

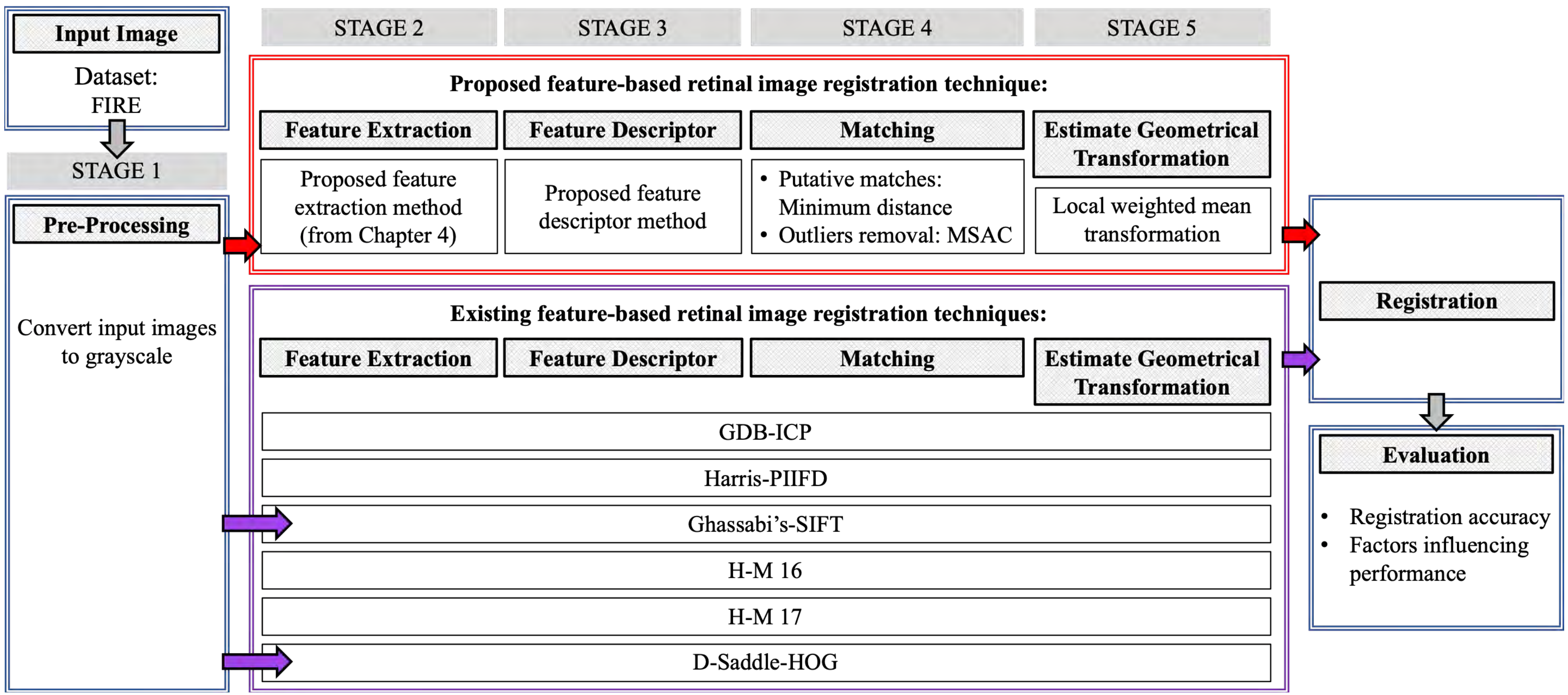



Figure 3.3: An overview of the work presented in Chapter 5: Feature Descriptor.

# CHAPTER 4: FEATURE EXTRACTION

## 4.1 INTRODUCTION

This chapter presents the proposed feature extraction method for Stage 2 of the feature-based retinal fundus image registration (RIR) technique. First, the issues in the existing feature extraction methods are discussed in Section 4.2. Then, Section 4.3 examines the characteristics of the retinal vessels and noises in the local patches of fundus image. This section is crucial in this work, where a fundamental understanding of the retinal vessels and noises characteristics provide an assistant in the development of the proposed feature extraction method. The proposed feature extraction method is presented in Section 4.4, where the feature detection and feature selection modules are described in Section 4.4.1 and Section 4.4.2, respectively. The details of the experimental setup such as existing feature extraction methods and datasets employed in evaluating the proposed feature extraction method are presented in Section 4.5. The experimental results are reported and discussed in Section 4.6. Finally, the presented work is summarised and concluded in Section 4.7 and Section 4.8, respectively. The mathematical symbols and notation used in this chapter can be found in Appendix A.

## 4.2 ISSUES IN EXISTING FEATURE EXTRACTION METHODS

There are several issues that can be outlined from the highlighted feature extraction methods in Section 2.4.1. First, the majority of the feature extraction methods incorporated feature enhancement algorithm such as DoG and *ii*DoG operators in building the hierarchical scale space. The feature enhancement algorithm is incorporated to increase the visibility of the retinal vessels that varies in terms of sizes and contrast. However, at the same time, it also increases the visibility of the noises such as retinal nerve fibre layer, underlying choroidal vessels, microaneurysm, exudates and edge of the optic disc. This makes it more challenging for the feature extraction methods to discriminate between retinal vessels and noises. The proposed feature extraction method avoids this issue by building the hierarchical scale space only from Gaussian smoothing operator without the feature enhancement algorithm as presented in Section 4.4.1.

Second, the feature extraction methods from the existing feature-based RIR techniques are mainly without a proper selection module to select feature points on retinal vessels. A proper feature selection module should consider both retinal vessels and noises characteristics as the noises may appear similarly as the retinal vessels within a local patch. Therefore, the proposed feature extraction method considers both retinal vessels and noises characteristics in the feature selection module to allow robust discrimination between these two. The feature selection module of the proposed feature extraction method is described in Section 4.4.2.

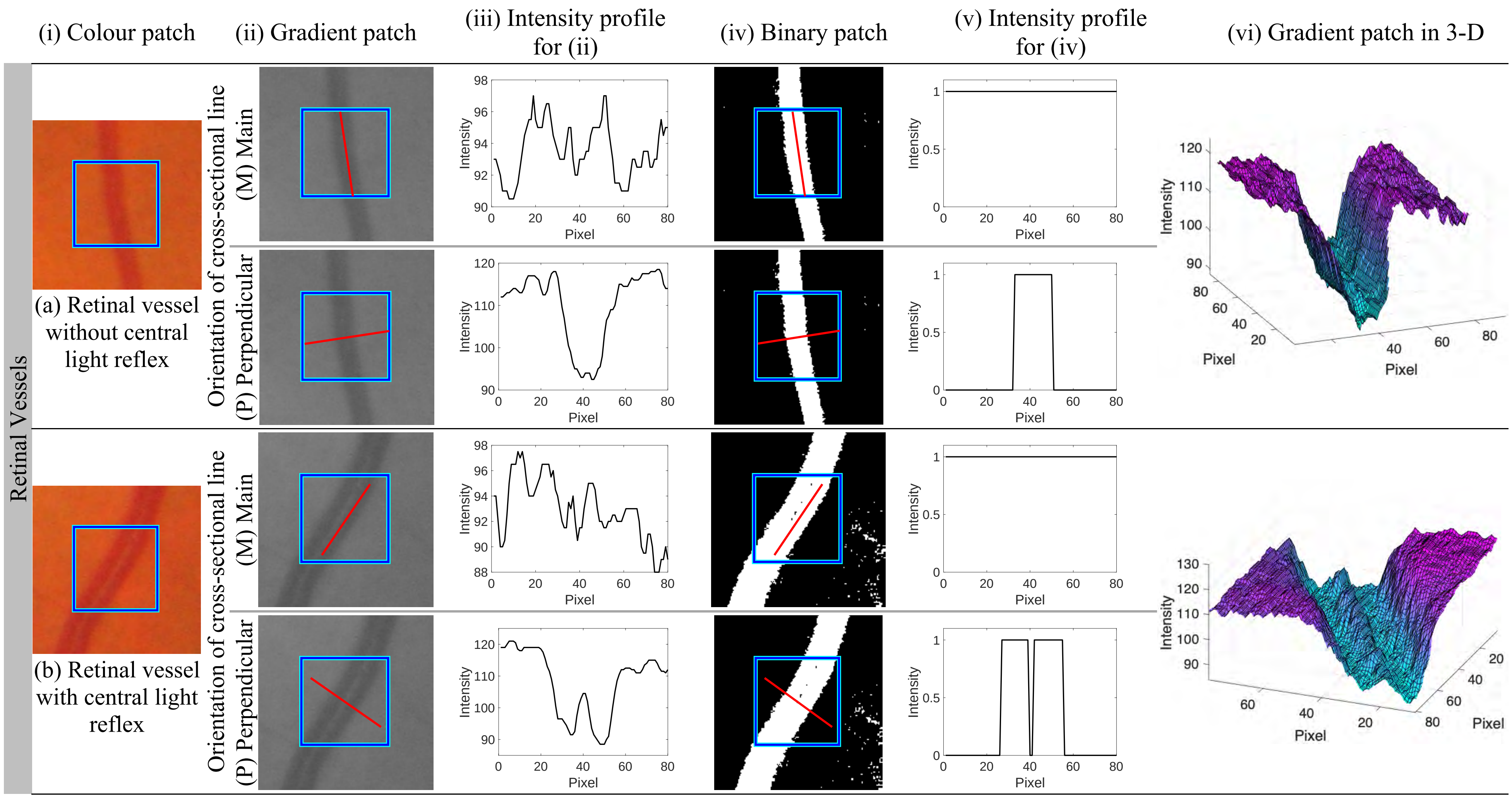
## 4.3 CHARACTERISTICS OF RETINAL VESSELS AND NOISES IN LOCAL PATCHES

This section examines the characteristics of the retinal vessels and noises within the local patches. It is important to understand the similarity and differences of the retinal vessels and noises within the local patches because their appearance can be similar due to the limited information present in the patch. The retinal vessels considered in this examination are the vessels with and without central light reflex whereas the noises considered are retinal nerve fibre layer, underlying choroidal vessels, microaneurysm, exudates and edge of the optic disc. The characteristics of the retinal vessels and noises are examined on the local gradient and binary patches with the size of $80 \times 80$ pixels.

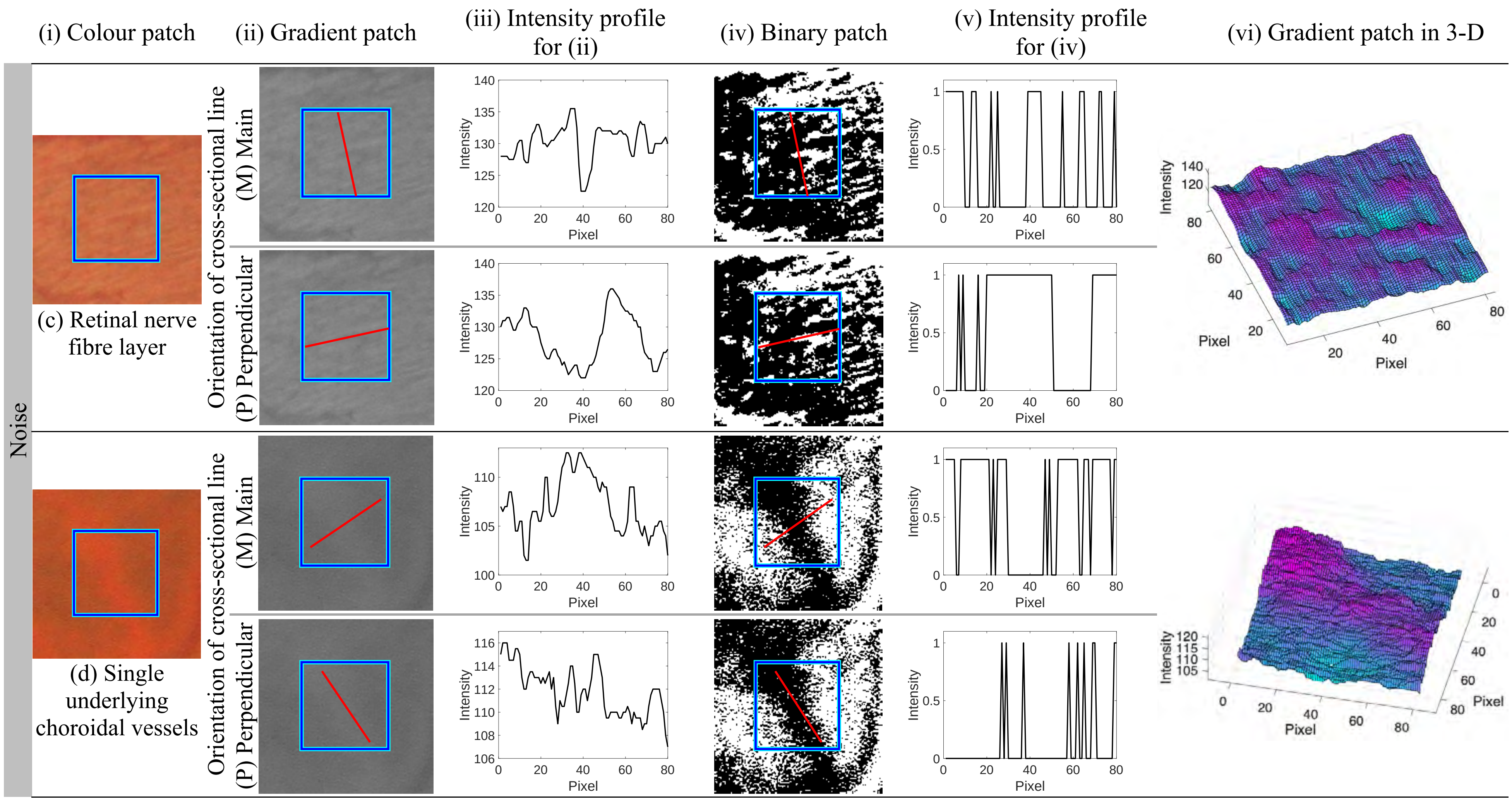
The retinal vessels either with or without central light reflex are generally appeared as continuous curvature structure that visible across the 3-dimensional gradient patch. The continuous curvature structure for the vessel without central light reflex has a U-shape as shown in Figure 4.1(a)(vi) whereas for the vessel with central light reflex has a W-shape as shown in Figure 4.1(b)(vi). The retinal vessels with the variation of size and contrast attain the similar continuous curvature structure but differ in terms of width and depth. Therefore, the information of the curvature structure without specifying its width and depth is considered in the proposed feature extraction method as it can be a reliable characteristic to detect feature points on the retinal vessels of various sizes and contrast. Furthermore, the proposed feature extraction method considers the curvature structure of concave and convex to allow the extraction of the feature points on the valley of the vessels and on the peak of the central light reflex. However, the noises in the 3-dimensional gradient patches as depicted in Figure 4.1(c)(vi) – Figure 4.1(g)(vi) also exhibit curvature structure of concave and convex with varying intensity. For example, single and multiple underlying choroidal vessels appear as continuous curvature structure across the 3-dimensional patch whereas retinal nerve fibre layer, microaneurysm and exudates exhibit a short continuous curvature structure.

The unique characteristics of the retinal vessels and noises are examined through intensity profiles extracted from the gradient and binary patches. The intensity profile is the intensity value of pixels extracted from a cross-sectional line running through the patch. The intensity profile for the retinal vessel has been reported in the literature to identify edges (Bankhead, Scholfield, McGeown, & Curtis, 2012) and estimate width (Araújo, Mendonça, & Campilho, 2018) as it is consistent between fundus images. The intensity profiles shown in columns of Figure 4.1(vi) and Figure 4.1(v) are extracted from the cross-sectional lines running through the centre of the gradient and binary patches, respectively. The cross-sectional lines are positioned along and perpendicular to the main orientation of the patches.

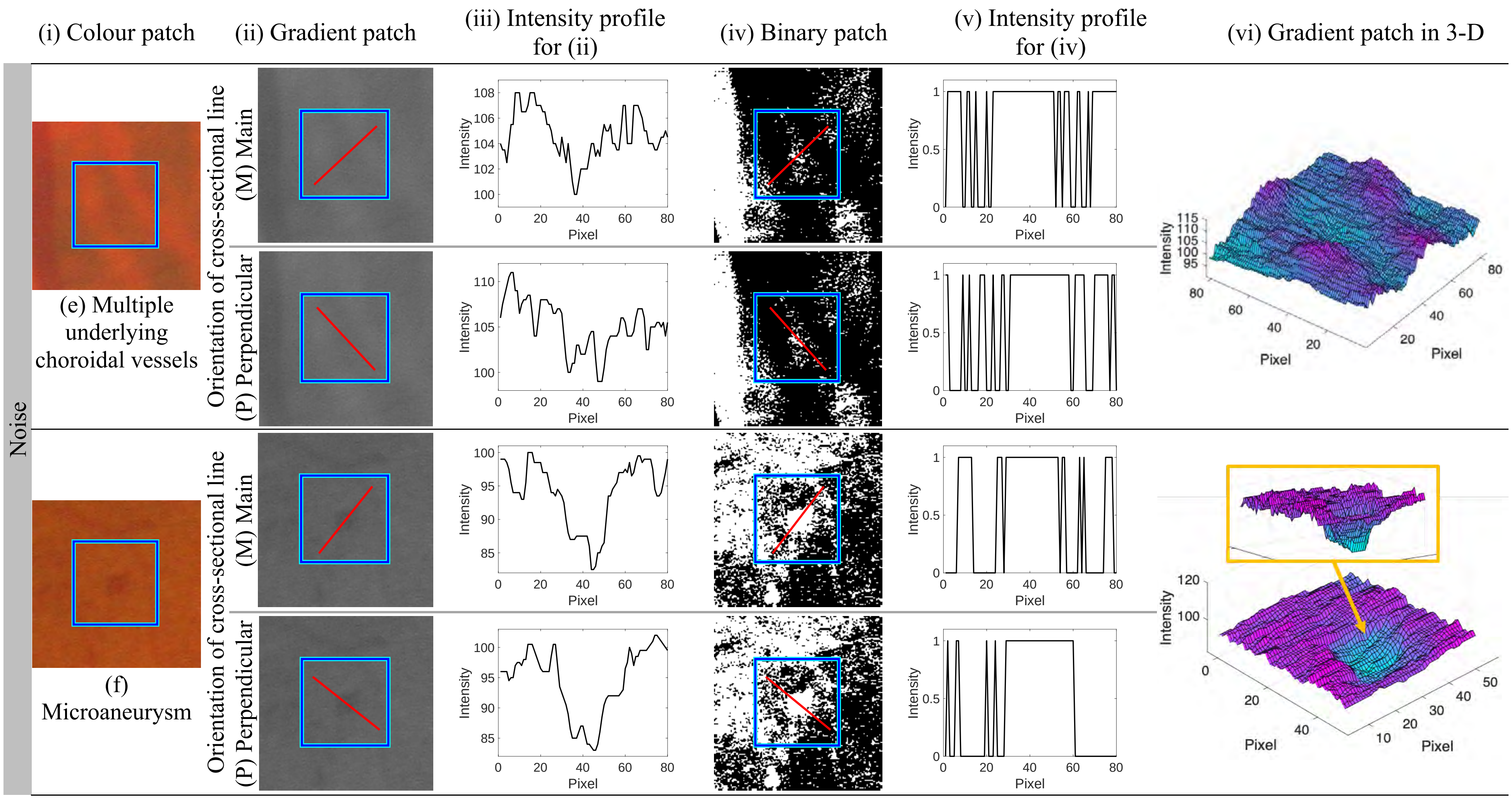
The retinal vessels within the gradient and binary patches appear as a continuous straight line structure. This straight line structure is represented by a long horizontal intensity profile when the cross-sectional line is positioned along the main orientation of



(i) Colour patch
(ii) Gradient patch
(iii) Intensity profile for (ii)
(iv) Binary patch
(v) Intensity profile for (iv)
(vi) Gradient patch in 3-D
Retinal Vessels
(a) Retinal vessel without central light reflex
(b) Retinal vessel with central light reflex
Orientation of cross-sectional line
(M) Main
(P) Perpendicular
Intensity
Pixel



(i) Colour patch
(ii) Gradient patch
(iii) Intensity profile for (ii)
(iv) Binary patch
(v) Intensity profile for (iv)
(vi) Gradient patch in 3-D
Noise
(c) Retinal nerve fibre layer
(d) Single underlying choroidal vessels
Orientation of cross-sectional line
(M) Main
(P) Perpendicular
Intensity
Pixel



(i) Colour patch
(ii) Gradient patch
(iii) Intensity profile for (ii)
(iv) Binary patch
(v) Intensity profile for (iv)
(vi) Gradient patch in 3-D
Noise
(e) Multiple underlying choroidal vessels
(f) Microaneurysm
Orientation of cross-sectional line
(M) Main
(P) Perpendicular
Intensity
Pixel

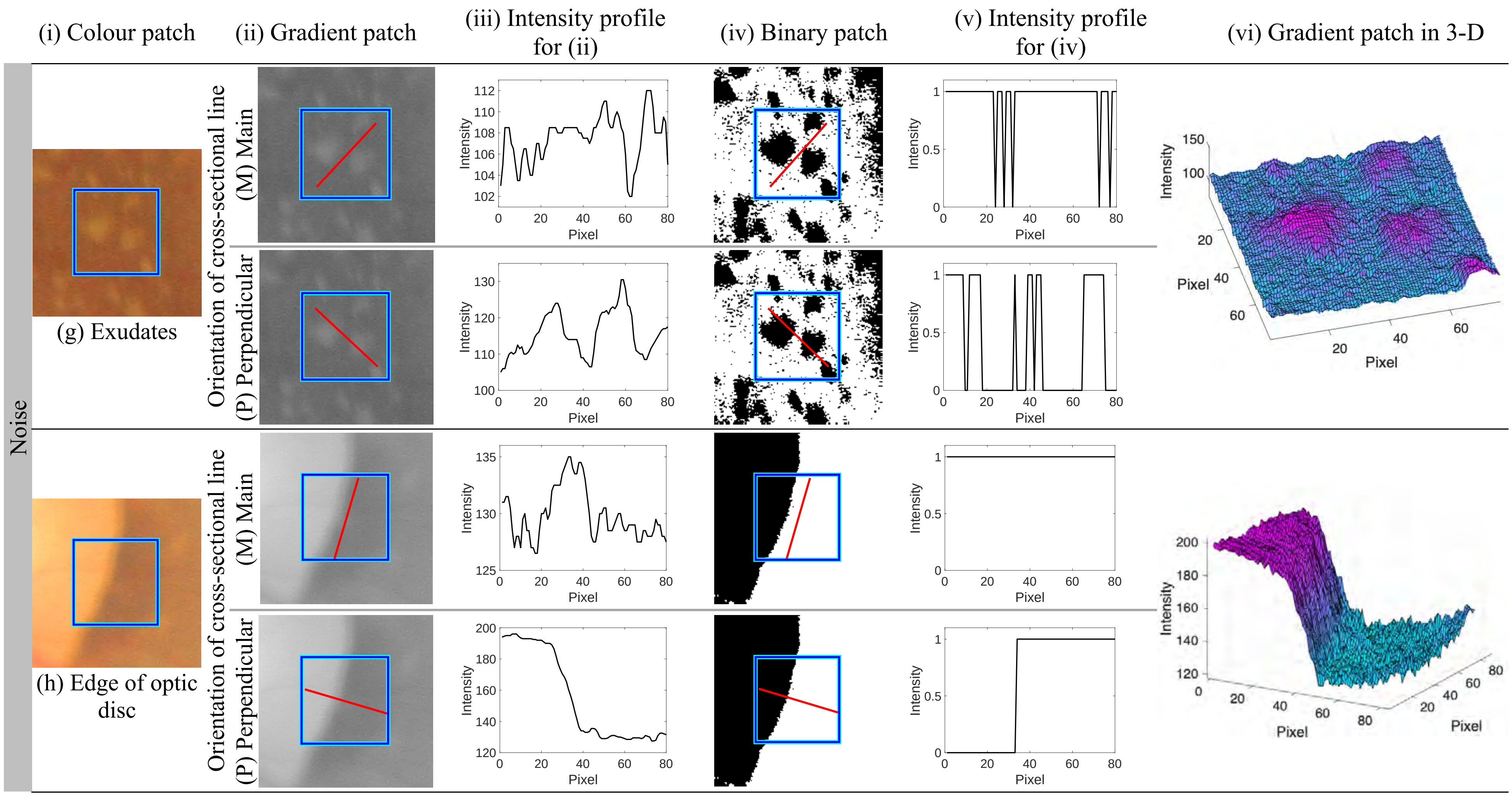



**Figure 4.1: Characteristics of retinal vessels and noises in local patches (blue square). Red lines in (ii) and (iv) are cross-sectional line to extract intensity profiles in (iii) and (v).**

the binary patch as depicted in Figure 4.1(a.M)(v) and Figure 4.1(b.M)(v). The characteristic of a long horizontal intensity profile is consistent between retinal vessels of various sizes and contrast. The similar horizontal intensity profile is also observed in the binary patch with edge of optic disc as depicted in Figure 4.1(h.M)(v). Other noises such as retinal nerve fibre layer, multiple underlying choroidal vessels, microaneurysm and exudates exhibit intensity profile with a shorter horizontal intensity profile. However, in a smaller patch, these horizontal intensity profiles may appear as a long horizontal intensity profile. The intensity profile of the retinal vessels when the cross-sectional line is positioned perpendicular to the main orientation of the binary patch appear as a singular bar for vessel without central light reflex and double bars for vessel with central light reflex. The width of these bars represents the width of the vessel in the patch.

The intensity profile of the retinal vessels extracted perpendicular to the main orientation of the gradient patch typically resembles an inverse Gaussian-like shape, where the width and depth of this shape depict the size of the vessel. In the presence of central light reflex, the bright region at the vessel introduced a higher intensity at the centre of the intensity profile as depicted in Figure 4.1(a.P)(iii). A similar inverse Gaussian-like shape but with a weaker intensity value can also be found in the intensity profile extracted from the gradient patch with retinal nerve fibre layer, microaneurysm and exudates. Discriminating between the noises and retinal vessels based on the intensity value may be inadequate because non-uniform intensity distribution often observed in fundus image. The non-uniform intensity distribution can lower the intensity value of the affected region. This can result in the intensity profile of the retinal vessel from the affected region to have a similar value and shape as the noises from the unaffected region.

The intensity profile of the retinal vessels and noises can be distinguished by examining the position of its valley with the maximum depth. For retinal vessels, the valley with the maximum depth is global minimum on the y-axis. This valley occurs approximately at the centre of the intensity profile as the retinal vessel is positioned at the centre of the patch. Contrarily, the position of the valley with the maximum depth for the noises is inconsistent. Based on the above explanation, we utilise the characteristics of the retinal vessels and noises observed in 3-dimensional patch and intensity profiles as part of the proposed feature extraction method presented in the following section.

## 4.4 PROPOSED FEATURE EXTRACTION

The proposed feature extraction method is composed of feature detection and feature selection modules. The feature detection module considers the general characteristic of the retinal vessels observed in 3-dimensional patch where candidate feature points are detected within the curvature structure. However, the detected candidate feature points are located on retinal vessels as well as noises. The candidate feature points associated with the noises are removed in the feature selection module based on the unique characteristics of the retinal vessels and noises observed in the intensity profiles. Then, the remaining candidate feature points are chosen as final feature points according to the strength of the retinal vessel attributes describing the size and contrast of the vessels within the patch. It should be noted that in the proposed feature extraction method, we do not discriminate between arteries and veins. The term of the retinal vessels used in this thesis refers to either arteries or veins, or both of them. The overview of the steps in the feature detection and feature selection modules is depicted in Figure 4.2.

### 4.4.1 Feature Detection

#### 4.4.1.1 STEP 1: Build hierarchical Gaussian scale space

The initial step of the feature detection module involves building a hierarchical Gaussian scale space as in (Burger & Burge, 2013; Lowe, 2004). The proposed feature extraction method utilised the hierarchical Gaussian scale space ($G$) containing a total of

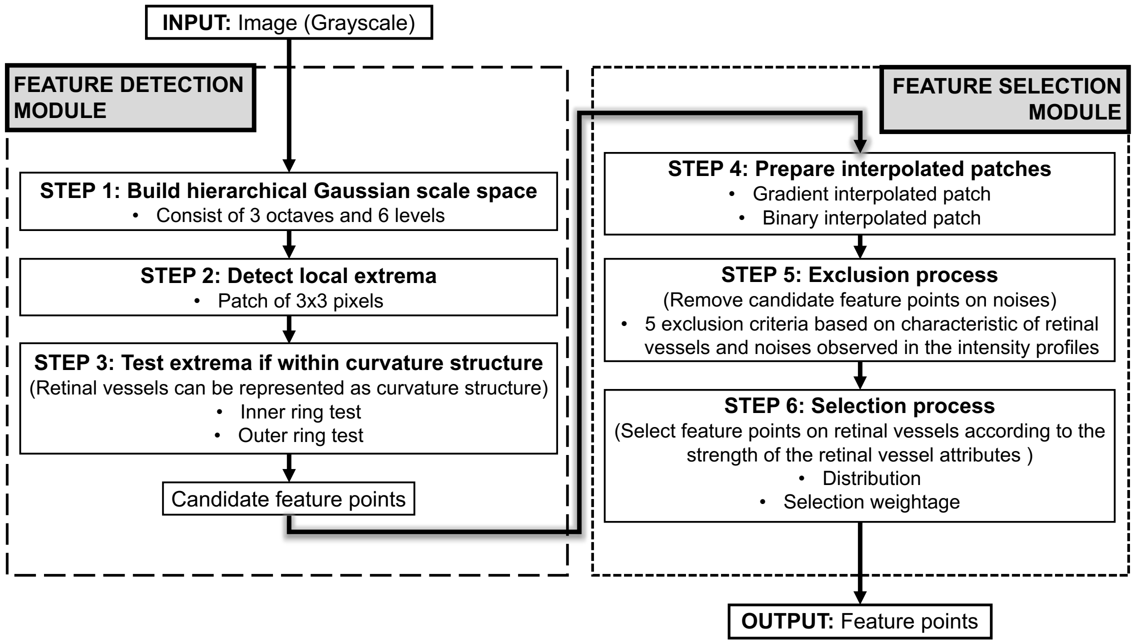



**Figure 4.2: Overview of the steps in the feature detection module and feature selection module.**

three octaves ($P = 3$) with index $p = 0, \cdots, P - 1$ and six levels ($Q = 6$) per octave with index $q = -1, \cdots, Q - 2$. The hierarchical Gaussian scale space allows candidate feature points being detected on various sizes of structures. For example, the candidate feature points are detected on various sizes of retinal vessels at the lower octave of the scale space as the images are larger and finer with detailed information. At the higher octave of the scale space, the candidate feature points are detected on thicker retinal vessels as the images are smaller and coarser with prominent information.

The main procedure in building the hierarchical Gaussian scale space involves creating the initial Gaussian image $G_{p,q}$ at $p = 0$ and $q = -1$. To create the initial Gaussian image $G_{0,-1}$, the input image $I$ is convolved with the initial width of absolute Gaussian kernel $\sigma_{p,q}$ at $p = 0$ and $q = -1$ as follows:

$$G_{0,-1} = I * \sigma_{0,-1} \tag{4.1}$$

The initial width of the absolute Gaussian kernel, $\sigma_{0,-1}$ is denoted by

$$\sigma_{0,-1} = \sigma_0 \cdot 2^{-1/Q-3} \tag{4.2}$$

where, $\sigma_0 = 1.6$ is the base width of the Gaussian kernel and $Q = 6$ is the total level per octave in the scale space. The absolute Gaussian kernel assumes the ideal scenario where the input image $I$ is free from the blur effect. However, in the real-world application, the input image $I$ contains various artefacts. For that reason, the input image $I$ is convolved with relative Gaussian kernel ($\breve{\sigma}_{p,q}$) to build the hierarchical Gaussian scale space. The relative Gaussian kernel assumes that the input image $I$ is pre-filtered with a sampling Gaussian kernel, $\sigma_s \geq 0.5$ (Lowe, 2004). Accordingly, the initial width of the relative Gaussian kernel $\breve{\sigma}_{0,-1}$ can be expressed as

$$\breve{\sigma}_{0,-1} = \sqrt{\sigma_{0,-1}^2 - \sigma_s^2} \tag{4.3}$$

To obtain the Gaussian image of the initial level in the higher octave $G_{p,-1}$ at $p \in [1, \dots, P-1]$, the Gaussian image $G_{p-1,Q-4}$ is down-sampled by half. The subsequent levels $G_{p,q}$ at $p \in [0, \dots, P-1]$ and $q \in [0, \dots, Q-2]$, can be obtained from the convolution between the initial Gaussian image in the respective octave $G_{p,-1}$ with the relative Gaussian kernel of width $\breve{\sigma}_q$ that is given by

$$\breve{\sigma}_q = \sigma_0 \cdot \sqrt{2^{2q/Q-3} - 1} \tag{4.4}$$

Note that, the relative Gaussian kernel is independent of the octave index. Therefore, the similar kernel width at a particular level can be used on another octave as depicted in Figure 4.3. An example of the hierarchical Gaussian scale space and their kernel widths for a retinal image is shown in Figure 4.4.

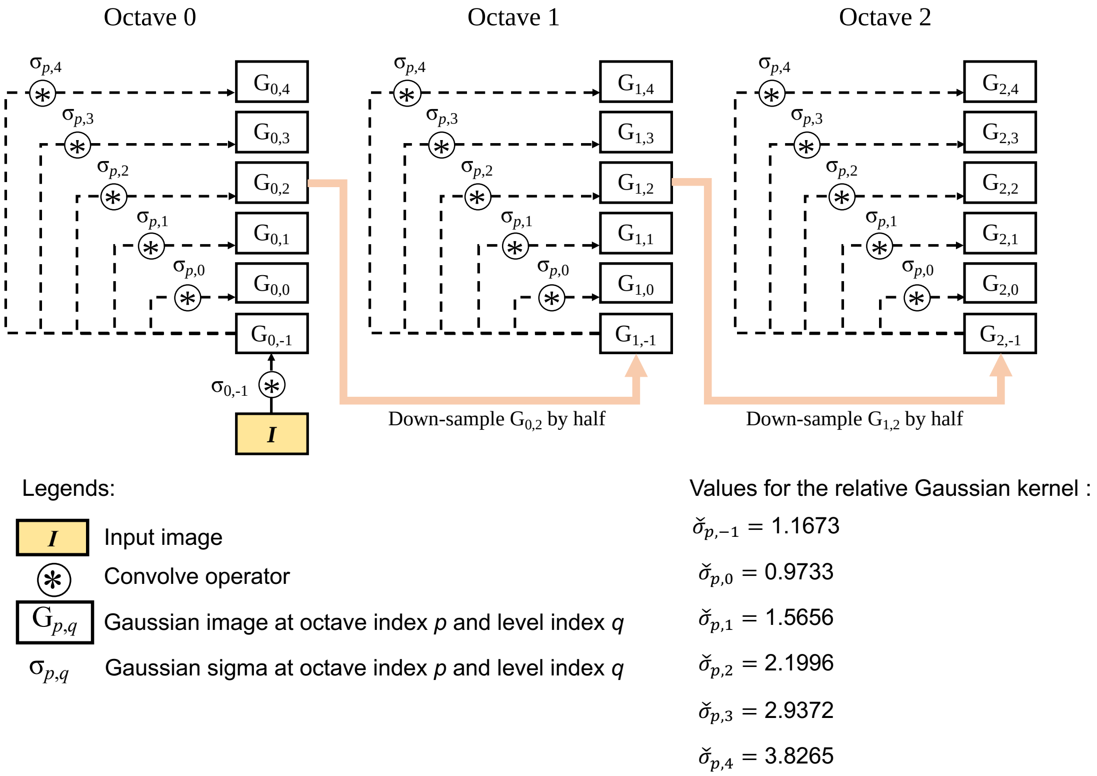



**Figure 4.3: Procedures to build the hierarchical Gaussian scale space.**

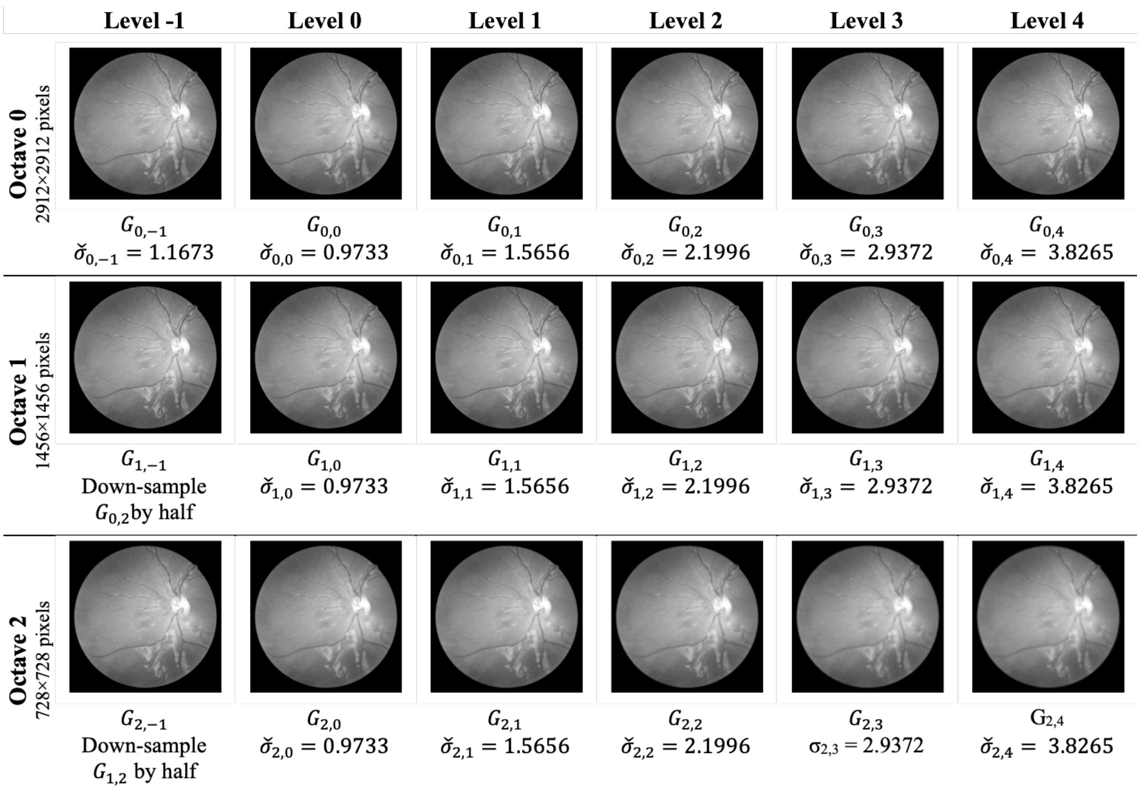



**Figure 4.4: Example of the hierarchical Gaussian scale space and the kernel widths for a retinal image.**

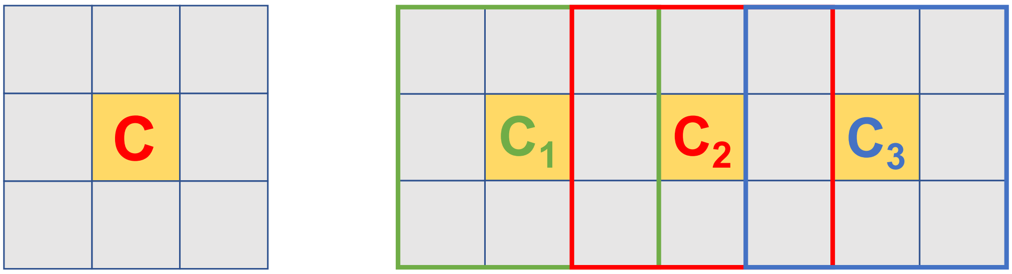
#### 4.4.1.2 STEP 2: Detect local extrema

The process of the feature detection module continues with the detection of extrema throughout the Gaussian images in the hierarchical Gaussian scale space. An extremum is determined from the intensity value of the pixel at the centre of the local patch of $3 \times 3$ pixels as shown in Figure 4.5(a). The centre pixel patch is taken as extremum if its intensity value is maximum or minimum than its eight immediate neighbours in the patch. The patches throughout the image are overlapped by $1/3$ of its size as illustrated in Figure 4.5(b). Only the extrema found within the region of interest (ROI) are considered for further processing. The ROI is identified using a mask image to remove the extrema found on the border of the field of view (FOV).

#### 4.4.1.3 STEP 3: Test extrema if within curvature structure

The retinal vessels in fundus image can be represented as a curvature structure in 3-dimensional as depicted in Figure 4.1(a)(iv) and Figure 4.1(b)(iv). Therefore, the extrema in the scale space are tested if they are within the curvature structure either concave or convex to allow the feature points being extracted on the valley of the vessels and on the peak of the central light reflex. The local curvature structure on an image can be found by performing two tests as reported in (Aldana-Iuit, Mishkin, Chum, & Matas, 2016). These tests are test for inner ring patterns and test for outer ring patterns.

The inner ring test considers eight pixels surrounding an extremum as shown in Figure 4.6(a). The inner ring pixels denoted by $a_j$ , $j \in [1, \dots, 8]$ are tested for four patterns in the shape of $\times$ and $+$ as depicted in Figure 4.6(b). Each pattern only involves four out of eight inner ring pixels. The patterns are formed when the intensity of two pixels in one direction are brighter than the other two pixels in the orthogonal direction. The four patterns of inner ring test can be expressed as




(a) (b)

**Figure 4.5: (a) The centre pixel, C (yellow square) is taken as an extremum if its intensity value is maximum or minimum than its eight immediate neighbour (grey square). (b) The patches in the image are overlapped by 1/3 of its size.**

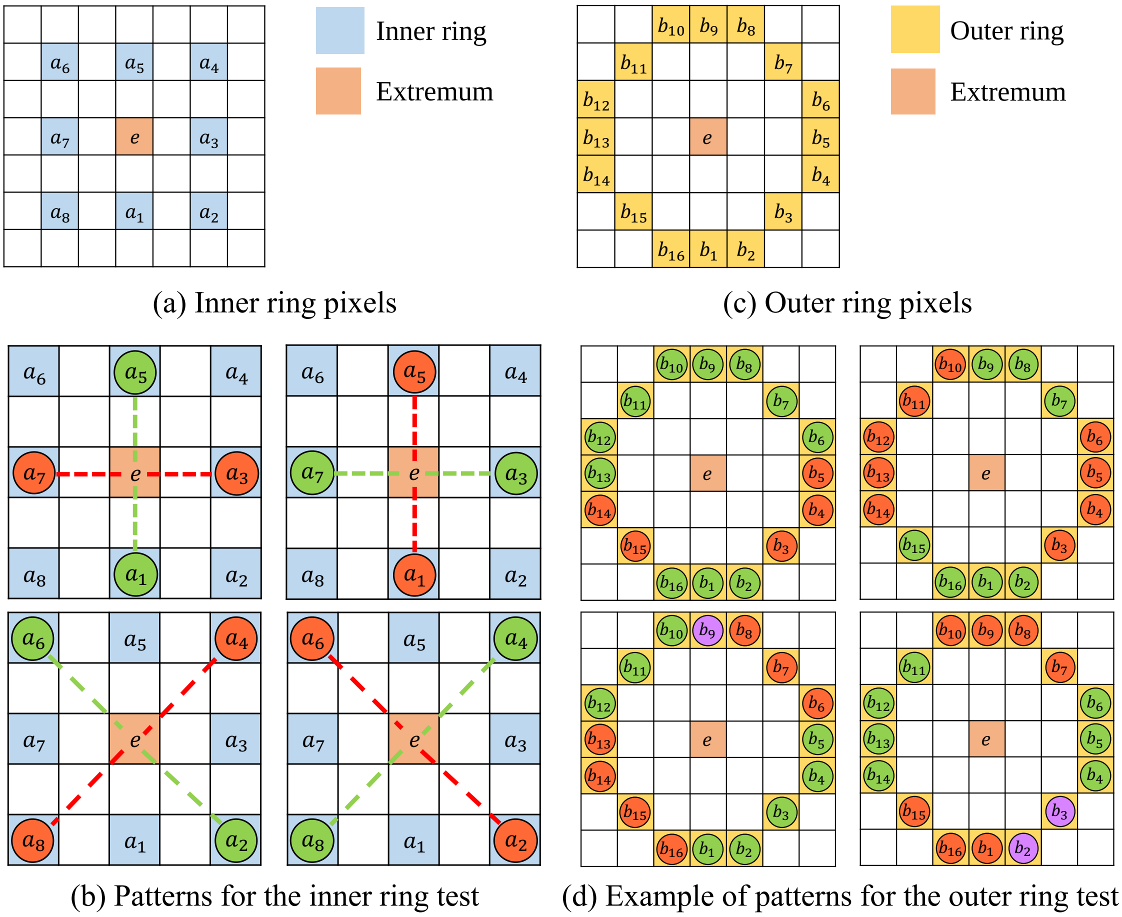



**Figure 4.6: (a) – (b) Inner ring test. (c) – (d) Outer ring test.**

$$
\begin{aligned}
\text{Pattern 1} &: \left(I_{a_1} > I_{a_3}\right) \wedge \left(I_{a_1} > I_{a_7}\right) \wedge \left(I_{a_5} > I_{a_3}\right) \wedge \left(I_{a_5} > I_{a_7}\right) \\
\text{Pattern 2} &: \left(I_{a_3} > I_{a_1}\right) \wedge \left(I_{a_3} > I_{a_5}\right) \wedge \left(I_{a_7} > I_{a_1}\right) \wedge \left(I_{a_7} > I_{a_5}\right) \\
\text{Pattern 3} &: \left(I_{a_2} > I_{a_4}\right) \wedge \left(I_{a_2} > I_{a_8}\right) \wedge \left(I_{a_6} > I_{a_4}\right) \wedge \left(I_{a_6} > I_{a_8}\right) \\
\text{Pattern 4} &: \left(I_{a_4} > I_{a_2}\right) \wedge \left(I_{a_4} > I_{a_6}\right) \wedge \left(I_{a_8} > I_{a_2}\right) \wedge \left(I_{a_8} > I_{a_6}\right)
\end{aligned}
\tag{4.5}
$$

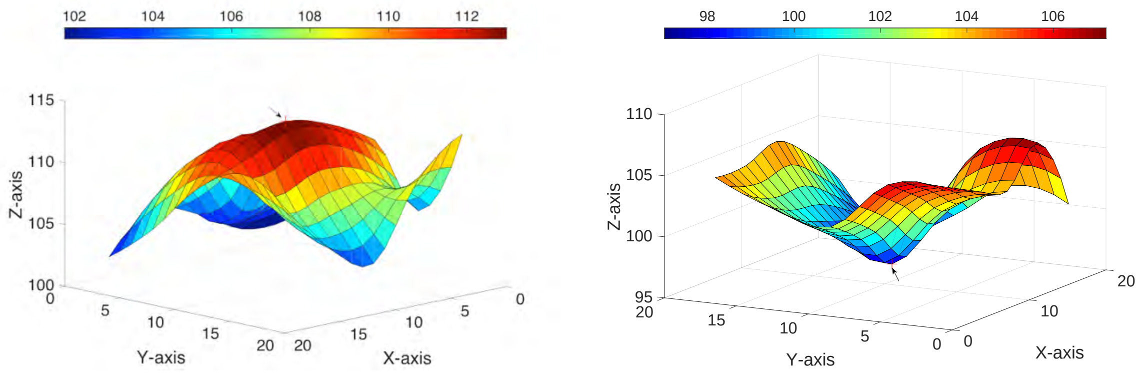
where, $I_{a_j}$ is the intensity of an inner ring pixel $a_j$. The extrema can pass the inner ring test with one or both shapes. The extrema that failed the test are eliminated. Then, the central intensity value $\beta$ is estimated for each extremum that passes the inner ring test. The central intensity value $\beta$ is a median value of four pixels if the extremum passes the inner ring test with one shape and eight pixels if the extremum passes the inner ring test with two shapes.

A circle with a circumference of 16 pixels surrounding an extremum that passes the inner ring test forms the outer ring pixels $b_l, l \in [1, \ldots, 16]$ as shown in Figure 4.6(c). The intensity of an outer ring pixel denoted by $I_{b_l}$ can be divided into three groups, namely, low, middle and high. These groups represent the intensity level of the outer ring pixels according to the central intensity value $\beta$ and offset $\varepsilon$ as follows:

$$\begin{array}{lll} \text{Group } \textit{low} \text{ (red dot)} & : & I_{b_l} < \beta - \varepsilon \\ \text{Group } \textit{middle} \text{ (purple dot)} & : & \beta - \varepsilon \leq I_{b_l} \leq \beta + \varepsilon \\ \text{Group } \textit{high} \text{ (green dot)} & : & I_{b_l} > \beta + \varepsilon \end{array} \tag{4.6}$$

The value of the offset $\varepsilon$ is set to 0.0010 as the intensity values of the pixels are in the range of [0,1] where a black pixel has a value of zero and a white pixel as one (Ramli et al., 2017b). After the outer ring pixels of an extremum are categorised according to the groups defined above, they are tested for the outer ring patterns.

The outer ring patterns consist of consecutive and alternating arcs from group *low* and *high*. The length of the arc can be in between 2 to 8 pixels. The arcs from group *low* and *high* can also be separated by pixels from group *middle* up to 2 pixels. Examples of these patterns are depicted in Figure 4.6(d). The extrema that pass this test are assigned as candidate feature points and included in the feature selection module. The examples of the candidate feature point where the extrema at the top and bottom of the curvature structure are shown in Figure 4.7.




(a) Extremum: Maximum (b) Extremum: Minimum

**Figure 4.7: Examples of the extremum at (a) the top and (b) bottom of the curvature structure found in the scale space.**

### 4.4.2 Feature Selection

The candidate feature points from the feature detection module are extrema found at the top and bottom of the curvature structure. The curvature structure in the gradient patch represents the retinal vessels of various sizes and contrast but it also represents noises such as retinal nerve fibre layer, underlying choroidal vessels, microaneurysm and exudates. Therefore, the feature selection module identifies and selects the candidate points located on retinal vessels and discard the others.

The feature selection module includes two components of exclusion and selection processes. The exclusion process discards the candidate feature points associated with the noises while the selection process selects the candidate feature points on retinal vessels based on the strength of the retinal vessel attributes describing the size and contrast of the vessel. The exclusion and selection processes will require gradient and binary interpolated patches for each candidate feature point as inputs.

#### 4.4.2.1 STEP 4: Prepare interpolated patches

The initial step of the feature selection module is to prepare the gradient and binary interpolated patches. A square patch for each candidate feature point is extracted from the Gaussian image $G_{p,q}$ of the scale space where the candidate feature point is detected. The size of the patch ($s_p \times s_p$) is varied depending on the octave position $p$ of the candidate feature point. This is to ensure that the size of the patch is proportioned to the retinal vessels at the respective octave and allows the selection of the feature points on retinal vessels of various sizes and contrast. The side length ($s_p$) of the patch is an odd number and determined as follows:

$$s_p = s_{initial} - 4(p + 1) \tag{4.7}$$

where, $s_p$ is the side length of the patch at $p \in [0 \ldots P - 1]$, $s_{initial}$ is the initial side length and $p$ is the octave index. The value for the initial side length $s_{initial}$ is determined by observing the width of the thickest vessels on five datasets; CHASE_DB1 (*CHASE_DB1 Retinal Image Database*; Fraz et al., 2012), DRIVE (*DRIVE: Digital Retinal Images for Vessel Extraction*; Staal et al., 2004), HRF (Budai et al., 2013; *HRF: High-Resolution Fundus Image Database*), STARE (Hoover et al., 2000; *STARE: Structured Analysis of the Retina*) and Fundus Image Registration (FIRE) dataset (Hernandez-Matas et al., 2017c). From the observations, the width of the thickest vessels is approximately 25 pixels on fundus images with a resolution larger than 1000×1000 pixels, 12 pixels on fundus images with resolutions in between 1000×1000 pixels to 600×600 pixels and 8 pixels on fundus images with resolutions smaller than 600×600 pixels. Considering the scale or zoom in the clinical practice is usually less than 1.5 (Ghassabi et al., 2015), and to ensure the retinal vessel can be fully captured within the patch, the following values for the initial side length $s_{initial}$ are deduced:

$$s_{initial} \begin{cases} 35 \text{ pixels} & \text{if } G_{0,-1} \quad > 1000 \times 1000 \text{ pixels} \\ 25 \text{ pixels} & \text{if } G_{0,-1} \quad \begin{matrix} \leq 1000 \times 1000 \text{ pixels} \\ > 600 \times 600 \text{ pixels} \end{matrix} \\ 21 \text{ pixels} & \text{if } G_{0,-1} \quad \leq 600 \times 600 \text{ pixels} \end{cases} \tag{4.8}$$

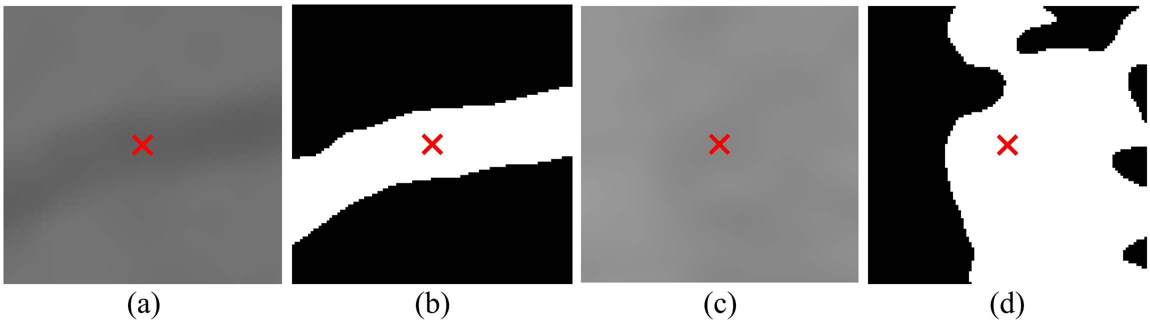



**Figure 4.8: Examples of the gradient and binary interpolated patches extracted from (a) – (b) retinal vessel and (c) – (d) noise. The '×' represents the position of the candidate feature point on the patch.**

The extracted patch is up-sampled using cubic interpolation with the refinement factor of 2 to smooth the region around the edges of the retinal vessels. Then, the gradient interpolated patch is converted to a binary image. Examples of the gradient and binary interpolated patches are shown in Figure 4.8. These patches are used as inputs for the exclusion and selection processes. The details of these processes are explained in the following sub-sections.

#### 4.4.2.2 STEP 5: Exclusion process

The exclusion process discards the candidate feature points associated with the noises based on five exclusion criteria as shown in Figure 4.9. These criteria are the characteristics of the retinal vessels and noises observed on the intensity profile as described in Section 4.3. The examined intensity profiles as shown in Figure 4.1 are extracted from a single cross-sectional line running through the patch, which can be unstable to distinguish between retinal vessels and noises. Therefore, the intensity profiles extracted from multiple cross-sectional lines are considered in the exclusion process to provide a robust representation of the retinal vessels and noises characteristics.

The cross-sectional lines are positioned along or perpendicular to the main orientation of the interpolated patch which is set according to the exclusion criteria. The main orientation is the angle between $x$-axis and major axis of the ellipse on the prominently

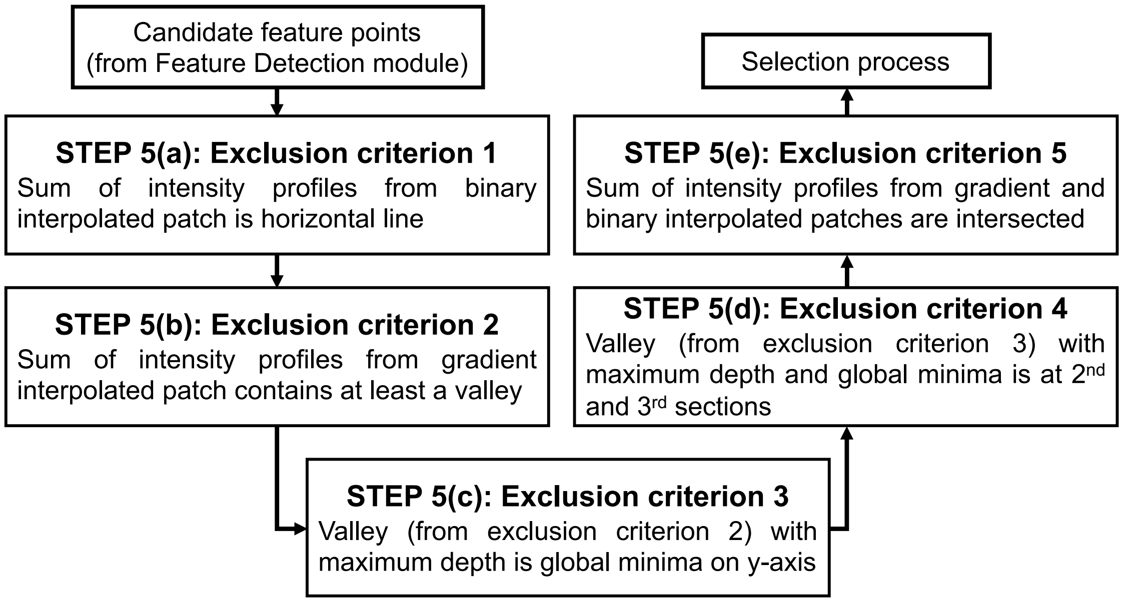



**Figure 4.9: Overview of the exclusion process.**

connected region of the binary interpolated patch. The ellipse has the same normalised second central moments as the binary interpolated patch. These cross-sectional lines are positioned parallel to each other with an identical length. The length of the cross-sectional lines used in the feature selection module is determined in a way that it will not exceed the size of the interpolated patch in any orientation as follows:

$$L_{length} = s_{bin} - (L_{distance} . L_{total}) \tag{4.9}$$

where, $L_{length}$ is the length of the cross-sectional lines, $s_{bin}$ is the side length of the binary interpolated patch on $x$-axis, $L_{distance}$ is the distance between the parallel cross-sectional lines, and $L_{total}$ is the total of the cross-sectional lines.

Then, the extracted intensity profiles are summed to highlight the characteristics of the retinal vessels and noises. The characteristics observed on the sum of the intensity profiles constitute of five criteria to exclude the candidate feature points on noises.

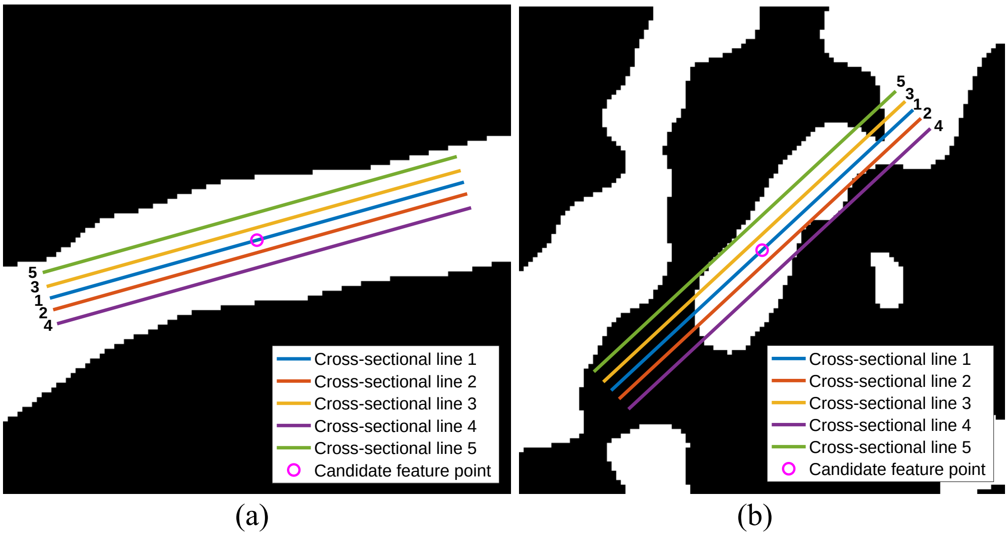



**Figure 4.10: Exclusion criterion 1. Examples of cross-sectional lines in the binary interpolated patch of (a) retinal vessel and (b) noise. The cross-sectional lines are position along the main orientation of the patch.**

(a) ***STEP 5(a): Exclusion criterion 1***

A retinal vessel in a binary interpolated patch forms as a nearly straight and wide connected region going through the candidate feature point as shown in Figure 4.10(a) whereas for the noise the connected region is inconsistent as depicted in Figure 4.10(b). These characteristics can be distinguished by extracting the intensity profiles from five cross-sectional lines ($L_{total} = 5$) with $L_{length}$ length. The distance between the cross-sectional lines is set to three pixels ($L_{distance} = 3$) and positioned along the main orientation of the connected region. This setting is empirically chosen to express the straight connected region of various sizes found in the interpolated binary patch with a retinal vessel. The extracted intensity profiles are then summed and results in a horizontal line for a retinal vessel (see Figure 4.11) and the detection of peaks for a noise (see Figure 4.12). From these characteristics, a candidate feature point with peaks on its sum of intensity profiles is discarded.

(b) ***STEP 5(b): Exclusion criterion 2***

The sum of the intensity profiles extracted from the interpolated gradient patch is examined in STEP 5(b)–(d) of exclusion criterion 2–4 for the resemblance of the inverse

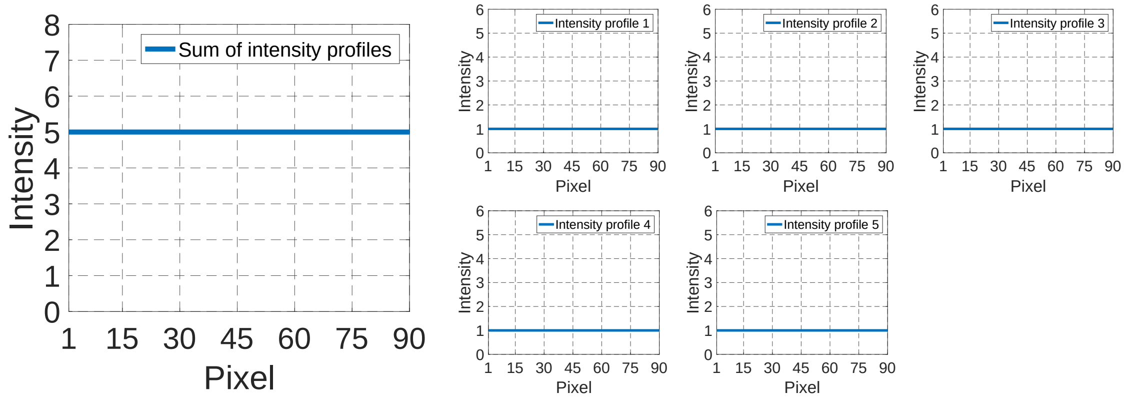



**Figure 4.11: Exclusion criterion 1. Sum of intensity profiles appear as a horizontal line for a retinal vessel. The intensity profiles are extracted from cross-sectional line 1–5 in Figure 4.10(a).**

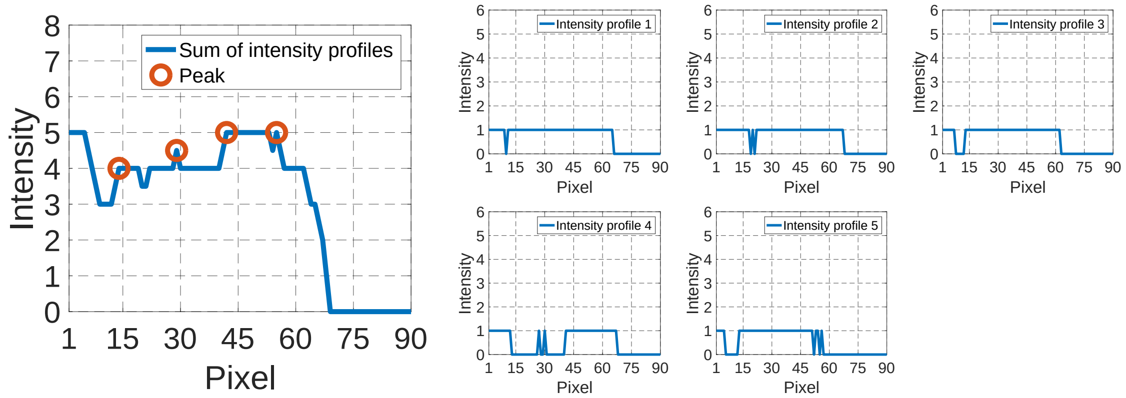
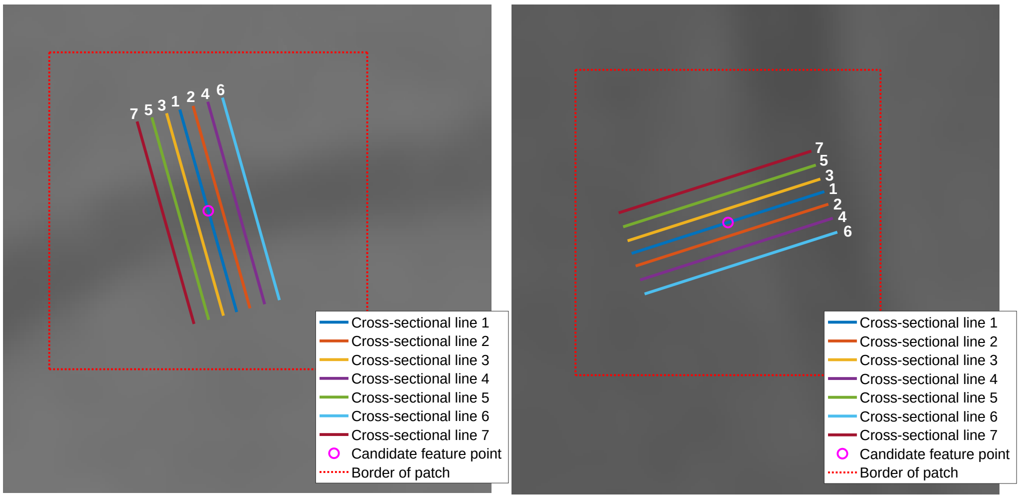



**Figure 4.12: Exclusion criterion 1. Sum of intensity profiles consists of various peaks for a noise. The intensity profiles are extracted from cross-sectional line 1–5 in Figure 4.10(b).**

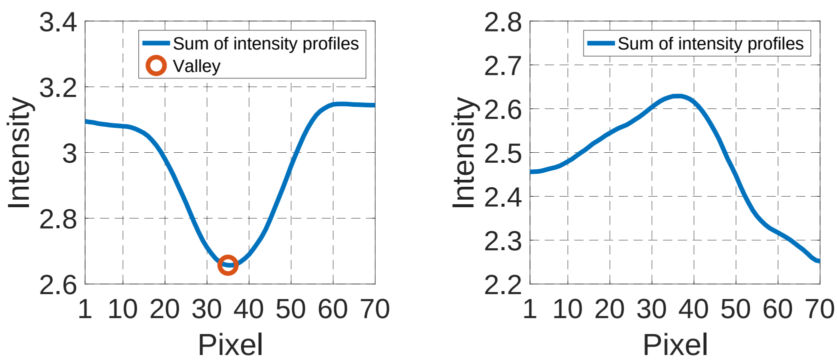
Gaussian-like shape. The inverse Gaussian-like shape of the intensity profile can be a reliable characteristic to distinguish between the retinal vessels and noises. This is because the intensity profile is expected to exhibit the inverse Gaussian-like shape when it is extracted from any part of the vessel in the patch as the curvature structure of the retinal vessel is continuous. Contrarily, the shape of the intensity profile for noises can be inconsistent when it is extracted from various part of the noise in the patch.

Considering the continuous curvature structure of the retinal vessel in the gradient interpolated patch, the sum of the intensity profiles is extracted from seven cross-sectional lines ($L_{total} = 7$) with $L_{length}$ length. These cross-sectional lines are positioned perpendicular to the main orientation of the gradient interpolated patch and separated by




(a) (b)

**Figure 4.13: Exclusion criterion 2. Examples of cross-sectional lines on (a) retinal vessel and (b) noise in gradient interpolated patch. The cross-sectional lines are positioned perpendicular to the main orientation.**




(a) (b)

**Figure 4.14: Exclusion criterion 2. Sum of intensity profiles for (a) retinal vessel and (b) noise from cross-sectional line 1–7 in Figure 4.13.**

five pixels ($L_{distance} = 5$). The cross-sectional lines are expected to intersect the vessel at various part of the vessel as shown in Figure 4.13(a). Therefore, the summation of the extracted intensity profiles will highlight the characteristic of the continuous curvature structure. The sum of the intensity profiles from the defined cross-sectional lines forms an inverse Gaussian-like shape with at least a valley as depicted in Figure 4.14. In opposite, the cross-sectional lines from interpolated patch with a noise such as the edge of the retinal vessel only intersect parts of the vessel as shown in Figure 4.13(b). The sum of the intensity profiles extracted from these cross-sectional lines will have no valley due

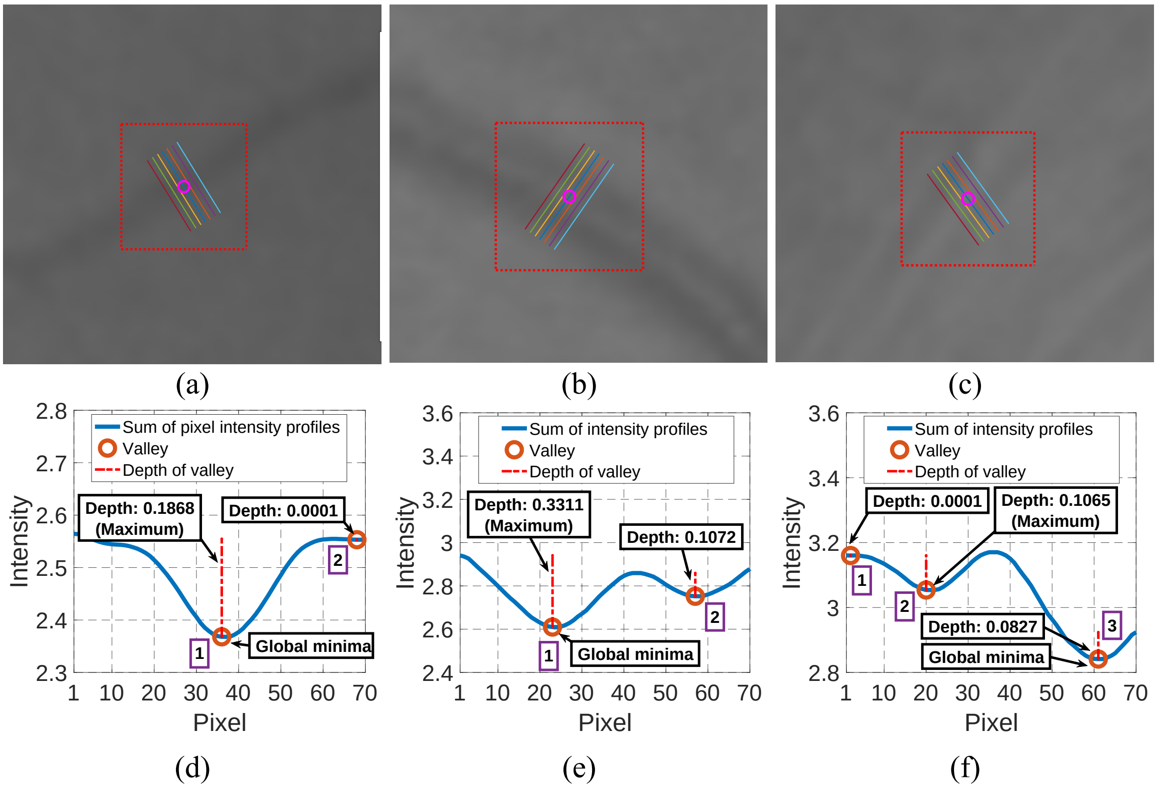
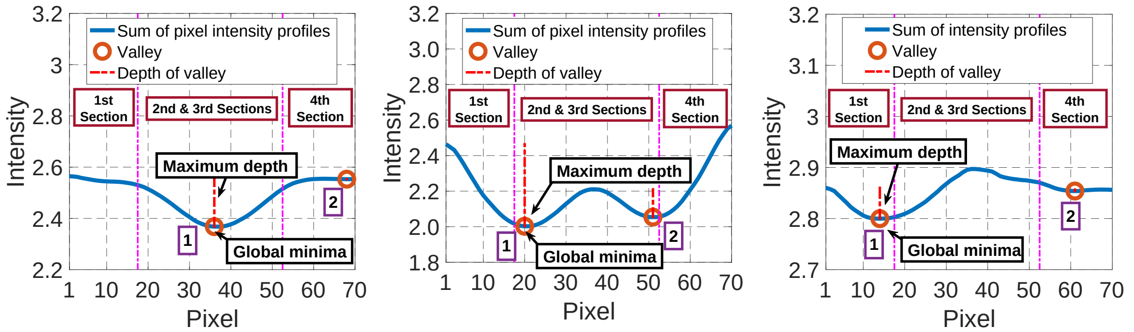



**Figure 4.15: Exclusion criterion 3. Cross-sectional lines on (a) – (b) retinal vessels and (c) noise in the gradient interpolated patch. (d) – (e) Sum of intensity profiles for retinal vessels. Valley 1 is global minimum and has maximum depth. (f) Sum of intensity profiles for a noise. Valley 3 is global minimum but valley 2 has the maximum depth.**

to incomplete information of the retinal vessel. Accordingly, this candidate feature point is discarded from further processing.

(c) ***STEP 5(c): Exclusion criterion 3***

The valleys found on the sum of the intensity profiles from exclusion criterion 2 are further examined for its depth and positioned on $y$-axis. The sum of the intensity profiles with inverse Gaussian-like shape consists a valley where its depth is maximum and at the lowest positioned of $y$-axis or global minimum as shown Figure 4.15(d) – (e). Contrarily, the noise with the similar straight and wide connected region as the retinal vessel in the interpolated patch as shown in Figure 4.15(c) failed to exhibit the same characteristic on the sum of the intensity profiles as depicted in Figure 4.15(f). Based on these characteristics, a candidate feature point is discarded if the valley with the maximum depth is not a global minimum.




(a) (b) (c)

**Figure 4.16: Exclusion criterion 4. Position of the valley with the maximum depth on x-axis. (a) – (b) The valley with the maximum depth is on the 2nd or 3rd section for retinal vessels. (c) The valley with the maximum depth is on the 1st or 4th section for noise.**

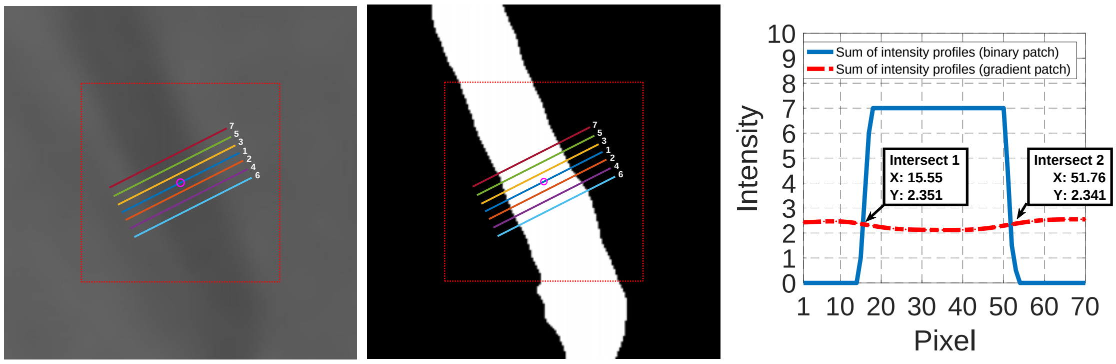
(d) ***STEP 5(d): Exclusion criterion 4***

Next, the valley with maximum depth and global minimum is analysed for their position on the $x$-axis. The sum of the intensity profiles is divided into four sections of equal size. The valley with maximum depth and global minimum is expected to be at the second or third section on the $x$-axis if a candidate feature point located on a retinal vessel as shown in Figure 4.16(a)–(b). For a candidate feature point on noise, the sum of the intensity profiles will have the valley with maximum depth and global minimum either at the first or last section on the $x$-axis as depicted in Figure 4.16(c). This candidate feature point is therefore excluded from further processing.

(e) ***STEP 5(e): Exclusion criterion 5***

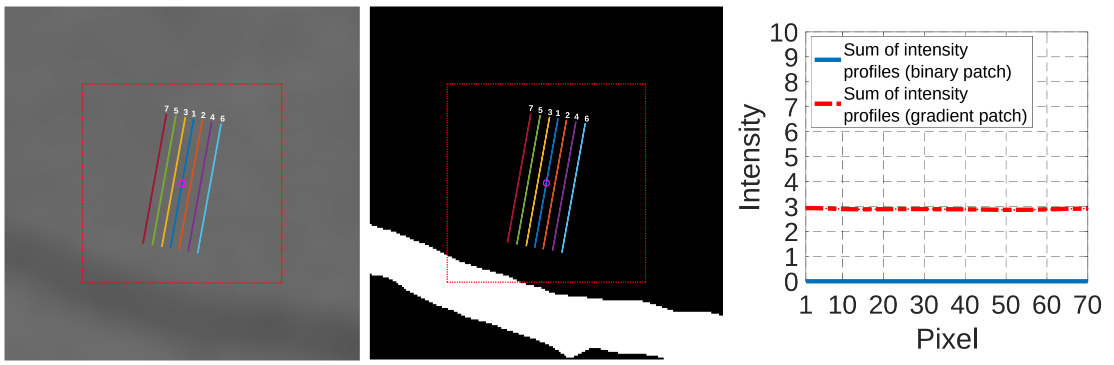
This criterion requires the sum of the intensity profiles to be extracted from both gradient and binary interpolated patches. The cross-sectional lines are set similar to the exclusion criterion 2 where, seven cross-sectional lines ($L_{total} = 7$) with $L_{length}$ length are positioned perpendicular to the main orientation of the interpolated patches. The distance separating the lines is set to five pixels ($L_{distance} = 5$).

Then, the intensity profiles are extracted from the gradient and binary interpolated patches and summed. The sum of the intensity profiles from the gradient and binary




(a) (b) (c)

**Figure 4.17: Exclusion criterion 5. Cross-sectional lines on retinal vessel of (a) gradient and (b) binary interpolated patches. (c) The intersection between the sum of the intensity profiles from binary and gradient interpolated patches.**




(a) (b) (c)

**Figure 4.18: Exclusion criterion 5. Cross-sectional lines on noise of (a) gradient and (b) binary interpolated patch. (c) No intersection can be found between the sum of the intensity profiles from binary and gradient interpolated patches.**

interpolated patches are overlaid on each other to find the intersection between them. The intersection can be observed if a candidate feature point located on a retinal vessel as shown in Figure 4.17(c). Contrarily, the sums of the intensity profiles from the gradient and binary interpolated patches are separated from each other if the candidate feature point is on noise as shown in Figure 4.18(c). Thus, this candidate feature point is discarded from further processing. The settings to extract the intensity profiles and details of the exclusion criteria (STEP 5(a)–STEP 5(e)) are summarized in Table 4.1.

**Table 4.1: Settings and details of STEP5: Exclusion criteria.**

| | | STEP 5(a): Exclusion criterion 1 | STEP 5(b): Exclusion criterion 2 | STEP 5(c): Exclusion criterion 3 | STEP 5(d): Exclusion criterion 4 | STEP 5(e): Exclusion criterion 5 |
|---|---|---|---|---|---|---|
| Settings to extract sum of intensity profiles from interpolated patches | | | | | | |
| Interpolated Patch | | Binary | Gradient | – | – | Binary and gradient |
| Cross-sectional lines | Length | $L_{length}$ | $L_{length}$ | – | – | $L_{length}$ |
| | $L_{total}$ | 5 | 7 | – | – | 7 |
| | $L_{distance}$ | 3 pixels | 5 pixels | – | – | 5 pixels |
| | Orientation | Along main orientation | Perpendicular to main orientation | – | – | Perpendicular to main orientation |
| Details of exclusion criteria | | | | | | |
| Input | | Sum of intensity profiles from binary interpolated patch | Sum of intensity profiles from gradient interpolated patch | Valley from STEP 5(b) | Valley with maximum depth and global minimum from STEP 5(c) | Sums of intensity profiles from binary and gradient interpolated patches |
| Characteristics | Vessel | A horizontal line. Figure 4.11. | Consists at least a valley. Figure 4.14(a). | Valley with maximum depth is global minimum. Figure 4.15(d)–(e). | At $2^{nd}$ or $3^{rd}$ section on $x$-axis. Figure 4.16(a)–(b). | Intersected when overlaid. Figure 4.17(c). |
| | Noise | Consists at least a peak. Figure 4.12. | Without valley. Figure 4.14(b). | Valley with maximum depth is local minimum. Figure 4.15(f). | At 1st or 4th section on $x$-axis. Figure 4.16(c). | Apart from each other when overlaid. Figure 4.18(c). |

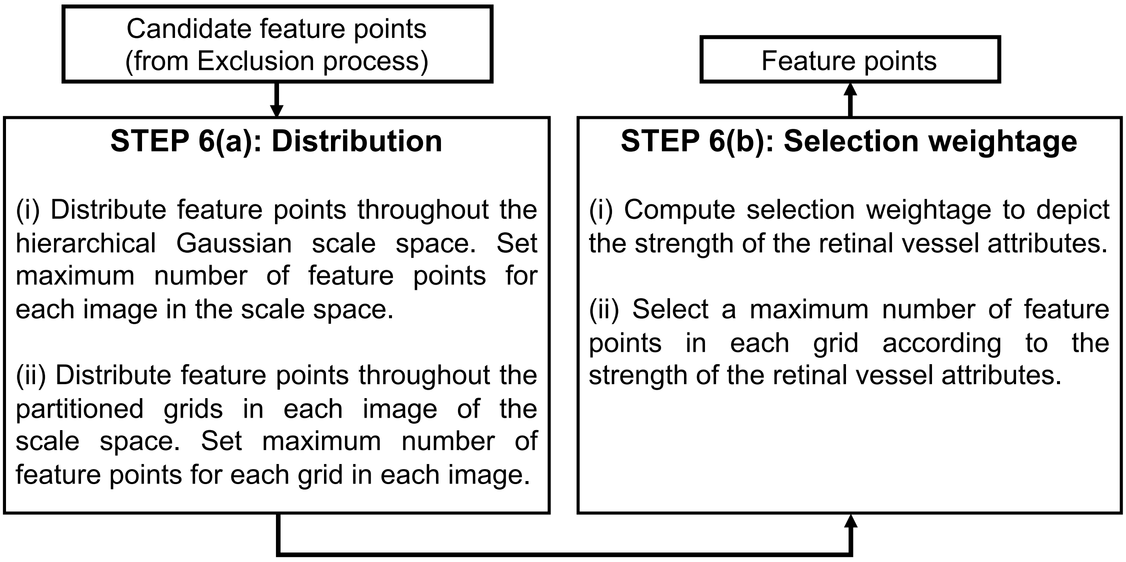



**Figure 4.19: Overview of the selection process.**

#### 4.4.2.3 STEP 6: Selection process

The exclusion process described in Section 4.4.2.2 removes the majority of the candidate feature points detected on noises. However, the remaining candidate feature points may include points detected on noises with a high structural similarity as the retinal vessels in the interpolated patches. Therefore, the selection process is introduced to select the final feature points according to the strength of the retinal vessel attributes and distributed throughout the image. The selection process can be summarised as shown in Figure 4.19.

(a) ***STEP 6(a): Distribution***

The distribution of the feature points throughout the fundus image is important to ensure a high registration accuracy (S. K. Saha et al., 2016). There are two procedures involved in distributing feature points. First, the feature points are distributed throughout the hierarchical Gaussian scale space. Then, the second procedure involves distributing the feature points throughout the partitioned grids in each image of the scale space.

- ***STEP 6(a)(i):*** *Distribute feature points throughout the hierarchical Gaussian scale space*

The first procedure of STEP 6(a) is to distribute feature points throughout the hierarchical Gaussian scale space. The feature points can be distributed by setting the maximum number of the feature point for each image in the hierarchical Gaussian scale space. The maximum number is set proportionally inverse to the scale of the Gaussian kernels used when building the scale space. This is to obtain more feature points on the finer images at the lower part of the scale space and fewer feature points on the coarser images at the higher part of the scale space. The maximum number of the feature point ($N_{p,q}$) for an image in the hierarchical Gaussian scale space is computed as described in (Ghassabi et al., 2015; Ghassabi et al., 2013; Sedaghat et al., 2011):

$$N_{p,q} = N_{total} . F_{p,q} \tag{4.10}$$

$$\text{with, } \sum_{p=0}^{P-1} \sum_{q=-1}^{Q-2} F_{p,q} = 1 \tag{4.11}$$

where, $N_{total}$ is the total feature points in the scale space, $F_{p,q}$ is the proportion of the feature points at octave $p$ and level $q$, $p$ is the octave index with $p \in [0 \ldots P-1]$, and $q$ is the level index with $q \in [-1 \ldots Q-2]$.

The proportion of the feature points $F_{p,q}$ is proportionally inverse to the scale coefficient of the Gaussian kernel $\beta_{p,q}$. $F_{p,q}$ can be determined from the proportion in the initial image of the scale space ($F_{0,-1}$). Assuming $F_{0,-1} = f_0$, $F_{p,q}$ can be defined as

$$F_{p,q} = \frac{\beta_{0,-1}}{\beta_{p,q}} f_0 \tag{4.12}$$

The scale coefficient of the Gaussian kernel $\beta_{p,q}$ is denoted by

$$\beta_{p,q} = \mu^{(Q)p+q} \tag{4.13}$$

and the proportion in the initial image of the scale space $f_0$ is given by

$$f_0 = \frac{\mu^{P(Q)-1}}{\sum_{n=1}^{P(Q)} \mu^{n-1}} \tag{4.14}$$

where, the constant factor $\mu$ can be expressed as

$$\mu = 2^{1/Q} \tag{4.15}$$

Substituting Equation (4.13) into Equation (4.12), $F_{p,q}$ can be written as

$$\begin{aligned} F_{p,q} &= \frac{\mu^{-1}}{\mu^{(Q)p+q}} f_0 \\ &= \frac{f_0}{\mu^{(Q)p+q+1}} \end{aligned} \tag{4.16}$$

Accordingly, the proportion of the feature points $F_{p,q}$ is approximately reduced by half from its proportion in the previous octave $F_{p-1,q}$ as shown in Figure 4.20. This proportion is inverse from the scale of the Gaussian kernels that is increased twice from its scale in the previous octave.

In this study, the total feature points in the hierarchical Gaussian scale space ($N_{total}$) is set to 4500 points, which shows empirically to provide a reasonable amount of feature points to perform image registration. If the candidate feature points are detected less than 4500 points, $N_{total}$ is set to 90% of the total candidate feature points detected in the scale space. The maximum number of the feature points $N_{p,q}$ throughout the hierarchical Gaussian scale space with $N_{total} = 4500$ points are depicted in Figure 4.21.

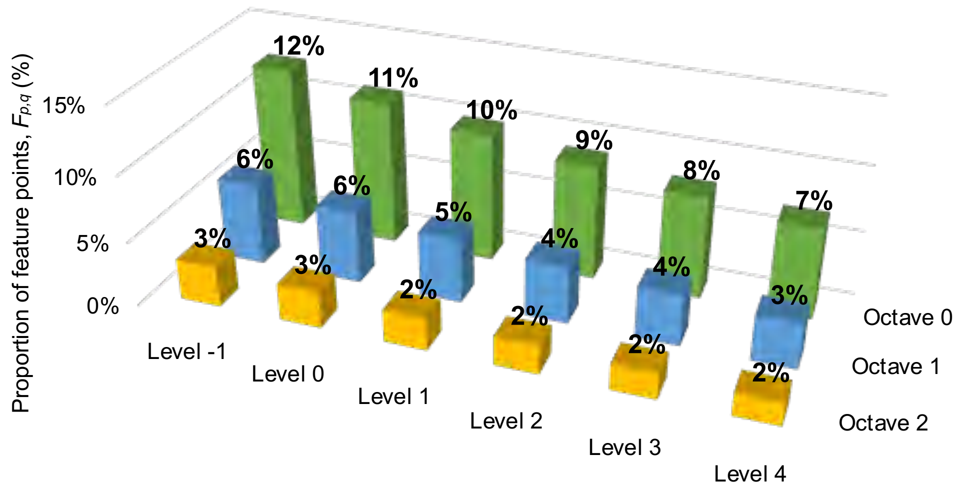



**Figure 4.20: Proportion of feature points $F_{p,q}$ in the hierarchical Gaussian scale space.**

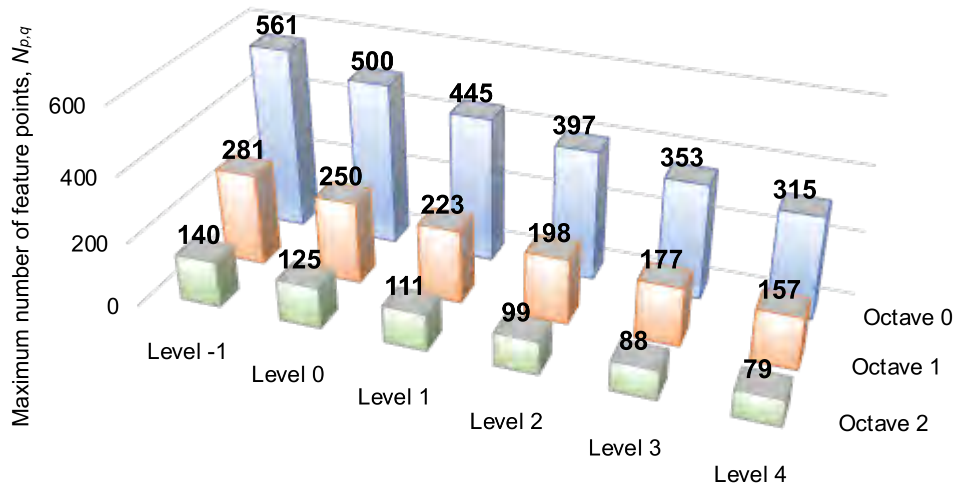
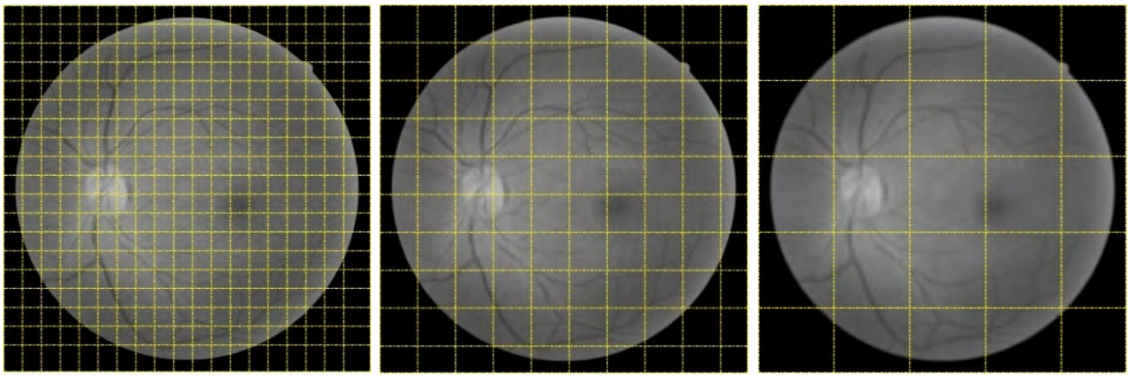



**Figure 4.21: Maximum number of feature points $N_{p,q}$ set in the hierarchical Gaussian scale space with $N_{total}$ = 4500 points.**

- ***STEP 6(a)(ii):*** *Distribute feature points throughout the partitioned grids in each image of the hierarchical Gaussian scale space*

Next, $N_{p,q}$ feature points are distributed throughout the partitioned grids within the Gaussian image $G_{p,q}$. The image $G_{p,q}$ is partitioned into rectangle grids with the size of 150×150 pixels as shown in Figure 4.22. The partitioning process originates at the top left of the image. This cause the grids at the right and bottom sides of the image to be smaller than the defined size. In the case where the grid is smaller than half of the defined grid size, the grid is merged with the previous grid. Otherwise, the grid remains as it is.



$G_{0,2}$ 2912×2912 pixels  $G_{1,2}$ 1456×1456 pixels  $G_{2,2}$ 728×728 pixels

**Figure 4.22: Examples of partitioned grids in the images of the hierarchical Gaussian scale space.**

The maximum number of the feature points in a grid of index $u$ denoted by $N_{p,q,u}$ is computed from the distribution coefficient ($WD_{p,q,u}$) as follows:

$$N_{p,q,u} = WD_{p,q,u}.N_{p,q} \tag{4.17}$$

The distribution coefficient $WD_{p,q,u}$ represents a combination of three factors to describe each grid. The factors considered in this study are entropy ($HG$) (Gonzalez, Woods, & Eddins, 2009), peak deviation nonuniformity ($UN$) (Goerner, Duong, Stafford, & Clarke, 2013) and total candidate feature points detected ($C$).

The first factor is entropy ($HG$) describing the texture of the grayscale image. The grid image that contains high contrast retinal vessels regardless of their sizes will yield a large value of entropy and vice versa. The entropy in a grid of index $u$ denoted by $HG_{p,q,u}$ can be computed as

$$HG_{p,q,u} = -\sum_{\ell}^{\mathcal{L}} \tau_{\ell} log_2 \tau_{\ell} \tag{4.18}$$

where, $\ell$ is the grayscale level of the grid image with $\ell \in [0, \cdots, \mathcal{L}]$ and $\tau_{\ell}$ is the occurrence of the grayscale level $\ell$. However, the grid image that contains low contrast retinal vessels will yield a similar entropy value as the grid image with only noises or retinal surface.

To compensate for this similarity, the second factor of the peak deviation nonuniformity ($UN$) is included in the computation of the distribution coefficient. The peak deviation nonuniformity measures the uniformity of the grayscale level of an image. It is sensitive to the non-uniformities and beneficial in distinguishing between the grid image that contains low contrast vessels and the grid image with only noises. The peak deviation nonuniformity for a grid of index $u$ denoted by $UN_{p,q,u}$ can be defined as

$$UN_{p,q,u} = 100\left(1 - \frac{\max(G_{p,q,u}) - \min(G_{p,q,u})}{\max(G_{p,q,u}) + \min(G_{p,q,u})}\right) \quad (4.19)$$

where, $G_{p,q,u}$ is the grayscale level of the grid image.

In the coarser grid image, the probability of the candidate feature points being detected is low compared to the finer grid image. However, the values of the entropy and peak deviation nonuniformity measured from the coarser and finer grid images only show a minimal difference. To compensate with these factors in the coarser grid image, the total of candidate feature points detected in each grid is considered as the third factor in computing the distribution coefficient ($C$).

The distribution coefficient for a grid of index $u$ in a Gaussian image $G_{p,q}$ can be expressed as the integration of three mentioned factors. The distribution coefficient denoted by $WD_{p,q,u}$ is calculated as

$$WD_{p,q,u} = W_{HG}\frac{HG_{p,q,u}}{\sum_u^U HG_{p,q,u}} + W_{UN}\frac{UN_{p,q,u}}{\sum_u^U UN_{p,q,u}} + W_{CF}\frac{C_{p,q,u}}{\sum_u^U C_{p,q,u}} \quad (4.20)$$

where, $W_{HG}$ is the weight factor for the entropy, $W_{CF}$ is the weight factor for the total of the candidate feature points, $W_{UN}$ is the weight factor for the peak deviation nonuniformity, $p$ is the octave index with $p \in [0, \cdots, P-1]$, $q$ is the level index with $q \in [-1, \cdots, Q-2]$, $u$ is the grid index with $u \in [1, \cdots, U]$ and $U$ is the total grids in a Gaussian image $G_{p,q}$. The values of the weight factors are empirically set to $W_{HG} = 0.3, W_{UN} = 0.3$ and $W_{CF} = 0.4$ to give a distinctive representation in describing the grid image.

(b) ***STEP 6(b): Selection weightage***

- ***STEP 6(b)(i):*** *Compute selection weightage to depict the strength of the retinal vessel attributes*

The selection process is continued by computing selection weightage for each candidate feature point. The selection weightage is composed of three attributes to highlight the characteristics of the retinal vessels in the interpolated patch. The attributes considered are entropy $(HP)$, area of the intersected region $(AV)$ and mean histogram of the gradient direction at the edges of the vessel $(MH)$.

The entropy $(HP)$ computed for each candidate feature point defines the texture of the local gradient interpolated patch. The interpolated patch with a high contrast retinal vessel will yields a high value of entropy and vice versa. The entropy for $\imath$-th candidate feature points denoted by $HP_{p,q,\imath}$ can be determined similar to the Equation (4.18) as follows

$$HP_{p,q,\imath} = -\sum_{\ell}^{\mathcal{L}} \tau_{\ell} log_2 \tau_{\ell} \tag{4.21}$$

where, $\ell$ is the grayscale level of the gradient interpolated patch with $\ell \in [0, \cdots, \mathcal{L}]$, $\tau_{\ell}$ is the occurrence of the grayscale level $\ell$, $p$ is the octave index with $p \in [0, \cdots, P-1]$ and $q$ is the level index with $q \in [-1, \cdots, Q-2]$.

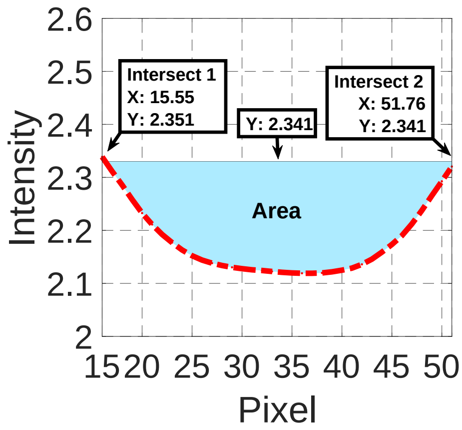



**Figure 4.23: Area of the intersected region.**

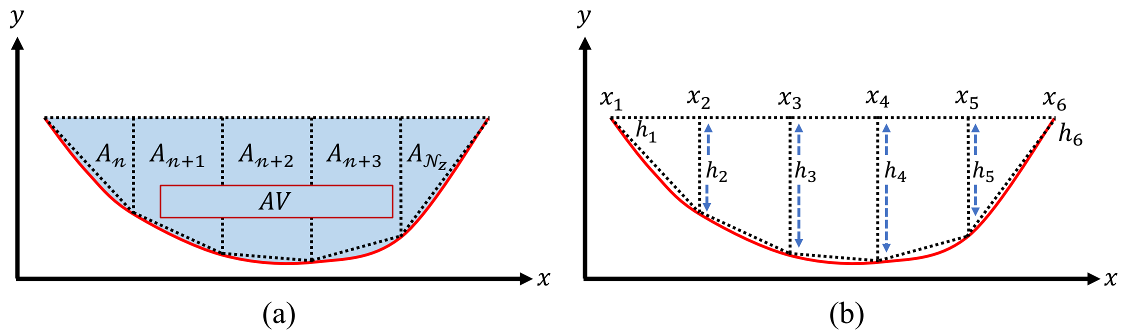



**Figure 4.24: Trapezoids in approximating the area of the intersected region.**

Next, the area of the intersected region is computed as part of the selection weightage. The intersection region is obtained from the intersection between the sum of the intensity profiles from the gradient and binary interpolated patches as described in the exclusion criterion 5. The lowest intersection point on $y$-axis is used as the reference level to measure the area of the intersected region as shown in Figure 4.23. The area of the intersected region highlights the strength of the retinal vessels in terms of size and contrast. For example, the intersected region has a larger area for a thicker and high contrast vessel while a smaller area for a narrower and low contrast vessel.

The area of the intersected region is approximated using trapezoidal rule where the intersected region is partitioned into a total of $\mathcal{N}_z$ trapezoids as shown in Figure 4.24(a). By considering an equal width of trapezoids, the area of the intersected region ($AV$) can be expressed as

$$
\begin{aligned}
AV &= A_n + A_{n+1} + \cdots + A_{\mathcal{N}_z} \\
&= \frac{\Delta x}{2} \sum_{n=1}^{\mathcal{N}_z} (h_n + h_{n+1})
\end{aligned} \tag{4.22}
$$

where, $A_n$ is the area of the trapezoid of index $n$, $\mathcal{N}_z$ is the total trapezoids approximating the intersected region, $\Delta x$ is the width of the trapezoid, $h_n$ and $h_{n+1}$ are the heights of the trapezoid. The width of the trapezoids can be defined by

$$
\Delta x = \frac{x_{\mathcal{N}_z+1} - x_1}{\mathcal{N}_Z} \tag{4.23}
$$

where, the width of each trapezoid is set to $\Delta x = 1$.

The third attribute in the selection weightage is the mean histogram of the gradient orientation at the edges of the retinal vessel. This attribute is estimated using both gradient and binary interpolated patches. Initially, partial derivative is performed on the pixels in the gradient interpolated patch to obtain the vector magnitude of the gradient direction. The central difference approximation is chosen to estimate the partial derivative as it gives a more accurate approximation compared to other techniques such as forward and backward approximations.

The central difference of the gradient interpolated patch is estimated with a unit spaced pixel. Therefore, the horizontal central difference along the $x$-axis ($\Delta\ddot{\beta}_x$) on the gradient interpolated patch can be written as

$$
\Delta\ddot{\beta}_x = 0.5\big(\mathrm{M_G}(x+1, y) - \mathrm{M_G}(x-1, y)\big) \tag{4.24}
$$

with, $x$ and $y$ representing the pixel position on the gradient interpolated patch $\mathrm{M_G}$. The value of $x$ in Equation (4.24) varies between $x = 2$ and $x = s_{inp} - 1$, whereas the value of $y$ in Equation (4.24) and Equation (4.25) varies between $y = 1$ and $y = s_{inp}$. $s_{inp}$ is

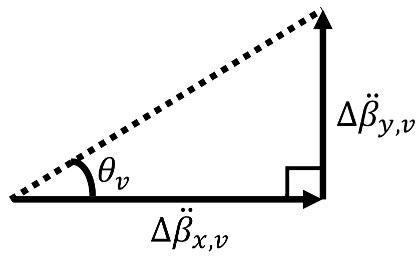



**Figure 4.25: The right triangle to compute gradient orientation for the $v$-th pixel.**

the side length of the interpolated patch. For the pixels that located at the edges of the patch ($x = 1$ and $x = s_{inp}$), $\Delta\ddot{\beta}_x$ is obtained with single sided differences as follows

$$\begin{aligned} \Delta\ddot{\beta}_x(1, y) &= \mathrm{M_G}(2, y) - \mathrm{M_G}(1, y) \\ \Delta\ddot{\beta}_x(s_{inp}, y) &= \mathrm{M_G}(s_{inp}, y) - \mathrm{M_G}(s_{inp} - 1, y) \end{aligned} \tag{4.25}$$

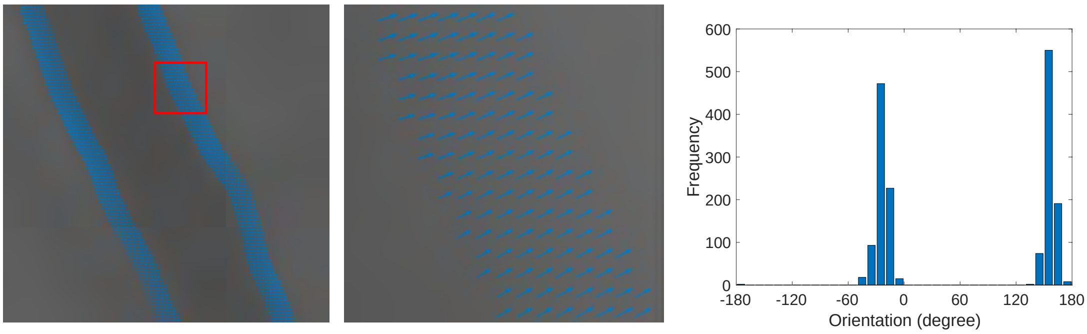
Similar to Equation (4.24) and Equation (4.25), the vertical central difference along the $y$-axis ($\Delta\ddot{\beta}_y$) on the gradient interpolated patch can be defined as in Equation (4.26).

$$\Delta\ddot{\beta}_y = 0.5\left(\mathrm{M_G}(x, y + 1) - \mathrm{M_G}(x, y - 1)\right) \tag{4.26}$$

The value of $y$ in Equation (4.26) varies between $y = 2$ and $y = s_{inp} - 1$ whereas the value of $x$ in Equation (4.26) and Equation (4.27) varies between $x = 1$ and $x = s_{inp}$. For the pixels located at the edges of the patch ($y = 1$ and $y = s_{inp}$), $\Delta\ddot{\beta}_y$ is obtained with single sided differences as in Equation (4.27).

$$\begin{aligned} \Delta\ddot{\beta}_y(x, 1) &= \mathrm{M_G}(x, 2) - \mathrm{M_G}(x, 1) \\ \Delta\ddot{\beta}_y(x, s_{inp}) &= \mathrm{M_G}(x, s_{inp}) - \mathrm{M_G}(x, s_{inp} - 1) \end{aligned} \tag{4.27}$$

The elements of the resulted $\Delta\ddot{\beta}_x$ and $\Delta\ddot{\beta}_y$ represent the vector magnitude of the gradient direction for the pixels in the patch. The central differences of $\Delta\ddot{\beta}_x$ and $\Delta\ddot{\beta}_y$ for the $v$-th pixel can be visualised as a right triangle as shown in Figure 4.25. Accordingly, the gradient orientation for the $v$-th pixel ($\theta_v$) can be expressed as




(a) (b) (c)

**Figure 4.26: (a) Example of the gradient orientation at the edges of the retinal vessel. (b) Close-up from the red rectangle region. (c) Histogram of 36 bins generated for the gradient orientation in (a).**

$$\theta_v = \tan^{-1}\frac{\Delta\ddot{\beta}_{y,v}}{\Delta\ddot{\beta}_{x,v}} \tag{4.28}$$

The gradient orientation $\theta_v$ is estimated in degree and pointing towards a brighter area as shown in Figure 4.26(b).

Then, the edges of the retinal vessel are identified on the binary interpolated patch. The dilation is performed on the binary interpolated patch to increase the thickness of the edges. Once the pixels on the edges of the retinal vessel are identified, the gradient orientation of the pixels are extracted. For a high contrast retinal vessel, the edges are thicker and the gradient orientation at the edges are more uniformed. Contrarily, the edges of the low contrast retinal vessel are thinner and the gradient orientation at the edges are less uniformed. Therefore, the uniformity of the gradient orientation at the edges can be a reliable indicator for the strength of the retinal vessels.

To measure the uniformity of the gradient orientation, a histogram of 36 bins is generated for each candidate feature point. Each of the bin represents a 10° orientation that ranged between -180° to 180° as shown in Figure 4.26(c). The frequency in the histogram signifies the total occurrence of the gradient orientation within the respective bin. Pixels at the edges with high uniformity of the gradient orientation will have a higher frequency and smaller count of bins. Considering this observation, the mean of the non-zero frequency in the histogram ($MH$) is designated as indicator of gradient orientation uniformity. The value of the $MH$ increases as the uniformity of the gradient orientation increases.

The selection weightage denoted by $WS$ is computed for each candidate feature point ($i$) in a Gaussian image $G_{p,q}$. The selection weightage highlights the characteristics of the retinal vessels through three attributes. The combination of these attributes can be expressed as

$$WS_{p,q,i} = W_{HP} \frac{HP_{p,q,i}}{\sum_{i}^{C_{p,q}} HP_{p,q,i}} + W_{AV} \frac{AV_{p,q,i}}{\sum_{i}^{C_{p,q}} AV_{p,q,i}} + W_{MH} \frac{MH_{p,q,i}}{\sum_{i}^{C_{p,q}} MH_{p,q,i}} \tag{4.29}$$

where, $W_{HP}$ is the weight factor for the entropy, $W_{AV}$ is the weight factor for the area of the intersected region, $W_{MH}$ is the weight factor for the mean histogram of the gradient orientation at the vessel edges, $p$ is the octave index with $p \in [0, \cdots, P-1]$, $q$ is the level index with $q \in [-1, \cdots, Q-2]$, $i$ is the candidate feature point index with $i \in [1, \cdots, C_{p,q}]$ and $C_{p,q}$ is the total candidate feature point in a Gaussian image $G_{p,q}$. The values of the weight factors are empirically set to $W_{HP} = 0.3, W_{AV} = 0.4$ and $W_{MH} = 0.3$ to distinctively highlights the strength of the retinal vessel attributes.

- ***STEP 6(b)(ii):*** *Select a maximum number of the feature points in each grid according to the strength of the retinal vessel attributes*

Finally, in each grid of a Gaussian image $G_{p,q}$, a total of $N_{p,q,u}$ candidate feature points with the highest value of the selection weightage $WS_{p,q,i}$ are selected as feature points ($D_{p,q,u}$). The positions of these feature points are then refined to the sub-pixel accuracy at the respective Gaussian image $G_{p,q}$ using a similar approach as in (M. Brown & Lowe, 2002; Lowe, 2004). After that, the refined spatial positions of the feature points in the coordinate system at the respective scale space are converted to the coordinate system of the input image (equivalent to Gaussian image in the initial octave, $G_{0,q}$) as follows:

$$D_n = 2^p . D_{p,q} \tag{4.30}$$

where, $D_n$ is the feature point of index $n$ in the coordinate system of the input image and $D_{n,p,q}$ is the feature point of index $n$ in the coordinate system of the respective octave $p$ and level $q$.

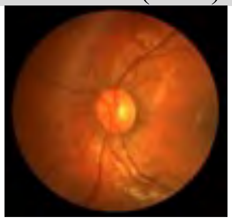
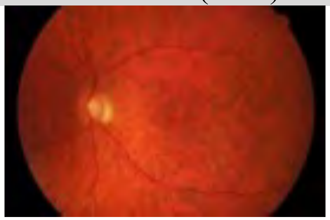
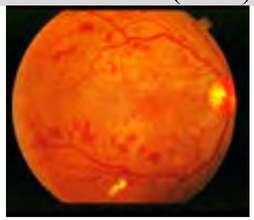
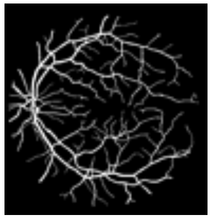
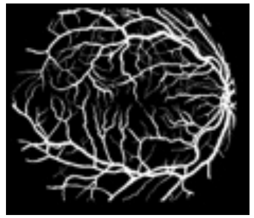
## 4.5 EXPERIMENTAL SETUP

The performance of the proposed feature extraction method is compared with five feature extraction methods that previously used in the existing feature-based RIR techniques. These feature extraction methods are Harris corner (Harris & Stephens, 1988), SIFT (Lowe, 2004), SURF (Bay et al., 2008; Bay et al., 2006), Ghassabi's (Ghassabi et al., 2015) and D-Saddle (Ramli et al., 2017b). Harris corner and SURF are implemented using MATLAB built-in functions, whereas SIFT is implemented using the open-source VLFeat library (Vedaldi & Fulkerson, 2010). The feature extraction methods of Ghassabi's and D-Saddle are developed following the original settings of the studies.

### 4.5.1 Datasets

There are four public datasets employed to evaluate the performance of the feature extraction methods, namely, CHASE_DB1 (*CHASE_DB1 Retinal Image Database*; Fraz et al., 2012), DRIVE (*DRIVE: Digital Retinal Images for Vessel Extraction*; Staal et al., 2004), HRF (Budai et al., 2013; *HRF: High-Resolution Fundus Image Database*) and STARE (Hoover et al., 2000; *STARE: Structured Analysis of the Retina*). These datasets contain fundus images that are affected by various pathological cases as described in Table 4.2. These datasets also provide the ground truth images of the manually segmented

**Table 4.2: Descriptions of datasets used for evaluating extraction accuracy.**

| Descriptions | Datasets | | | |
|---|---|---|---|---|
| | CHASE_DB1 | DRIVE | HRF | STARE |
| Total images | 28 | 40 | 45 | 20 |
| Resolution (pixels) | 999×960 | 564×584 | 3504×2336 | 605×700 |
| Total patients | 14 | 40 | 45 | 20 |
| Age | 9 – 10 years | 25 – 90 years | Not available | Not available |
| Pathological cases | Vessel tortuosity | • 33 images - no sign of diabetic retinopathy<br>• 7 images - signs of mild early diabetic retinopathy | • 15 images of healthy patients<br>• 15 images of diabetic retinopathy<br>• 15 images of glaucomatous | Abnormalities that obscure or the blood vessel appearance such as haemorrhaging etc. |
| Field of view | 30º | 45º | 45º | 35º |
| Year | 2012 | 2004 | 2009 | 2000 |
| Ground truth images | 56 | 60 | 45 | 40 |
| Intensity distribution[1] | 22.6136 | 49.3307 | 34.9433 | 49.5126 |
| Author(s) | Fraz et al. (2012) | Staal et al. (2004) | Budai et al. (2013) | Hoover et al. (2000) |
| Examples (Fundus) |  |  |  |  |
| Examples (Ground truth) |  |  |  |  |

[1]Described by peak deviation nonuniformity intensity (Goerner et al., 2013). Value close to 0 indicates non-uniform intensity distribution in the image.

vessels performed by experts. These ground truth images enable the accuracy measurement of the extracted feature points on retinal vessels.

### 4.5.2 Evaluation Metrics

#### 4.5.2.1 Extraction Accuracy

All feature extraction methods are evaluated in terms of its ability to extract feature points on retinal vessels. This ability is expressed as an extraction accuracy, a ratio between the total of the extracted feature points on retinal vessels to the total of the feature points extracted:

$$Extraction\ accuracy\ (\%) = \frac{total\ feature\ points\ on\ vessels}{total\ extracted\ feature\ points} \times 100\% \tag{4.31}$$

The feature points on retinal vessels are determined by referencing the extracted feature points to the ground truth images of the manually segmented vessels. The extraction accuracy is set to 0% when the total feature points extracted for a particular image is less than 3, due to the minimum requirement to perform a transformation. The statistical test of one-way Analysis of Variance (ANOVA) with Tukey’s post hoc is performed to compare the extraction accuracy between the feature extraction methods.

#### 4.5.2.2 Factors

All the feature extraction methods are further evaluated by investigating factors influencing their extraction accuracy. The factors investigated are changes in image size and intensity distribution throughout the image. The intensity distribution in fundus image can be affected by the spherical shape of the retina, a low light reflection of the macula and the progression of the diseases. The peak deviation nonuniformity (Goerner et al., 2013) as expressed in Equation (4.19) is computed to describe the intensity distribution of the image. A higher value of the peak deviation nonuniformity indicates the intensity distribution of the image is more uniform while a lower value indicates the image is less uniform.

The relation between the factors and extraction accuracy are investigated using Spearman's rank-order correlation. The value of the Spearman's rho ($r_s$) is ranged between –1 to +1. The factor is highly influenced the extraction accuracy or the correlation is strong when the Spearman's rho ($r_s$) is close to ±1. In opposite, the Spearman's rho that closes to 0 indicates the factors has minimal influence or the correlation is weak. The one-way ANOVA and Spearman's rank-order correlation are significant at the 0.05 level identified by a single asterisk or at the 0.01 level identified by two asterisks. These statistical tests are conducted using IBM SPSS Statistics (Version 24) software.

## 4.6 RESULTS & DISCUSSION

### 4.6.1 Extraction Accuracy

The proposed feature extraction method extracted an average of 2482 feature points from each fundus image in the datasets with 2149 of the points are accurately associated with retinal vessels as summarised in Table 4.3. Furthermore, the proposed feature

**Table 4.3: Average of feature points extracted for each dataset.**

| Feature Extraction Methods | Feature points extracted | Datasets | | | | Overall[1] |
|---|---|---|---|---|---|---|
| | | CHASE | DRIVE | HRF | STARE | |
| Harris | All | 986 | 116 | 3445 | 69 | 1419 |
| | Vessel | 180 | 64 | 1550 | 27 | 586 |
| SIFT | All | 2656 | 745 | 36415 | 1432 | 13320 |
| | Vessel | 518 | 140 | 3699 | 272 | 1444 |
| SURF | All | 1860 | 1548 | 5358 | 1741 | 2932 |
| | Vessel | 350 | 251 | 1195 | 281 | 596 |
| Ghassabi’s | All | 4142 | 4071 | 8277 | 3164 | 5373 |
| | Vessel | 955 | 1175 | 2337 | 1039 | 1502 |
| D-Saddle | All | 18846 | 5977 | 83153 | 6892 | 34936 |
| | Vessel | 2775 | 1332 | 17445 | 1547 | 7120 |
| Proposed | All | 2143 | 1349 | 4011 | 1780 | 2482 |
| | Vessel | 1553 | 1214 | 3586 | 1620 | 2149 |

[1]Computed from the summation of all image at once rather than the mean of each dataset to minimize the accumulation of rounding error.

**Table 4.4: Total image in the dataset with at least 3 extracted feature points, due to the minimum requirement to perform a transformation.**

| Feature Extraction Methods | Datasets | | | |
|---|---|---|---|---|
| | CHASE | DRIVE | HRF | STARE |
| **Harris** | 28 | 37 | 45 | 17 |
| **SIFT** | 28 | 40 | 45 | 20 |
| **SURF** | 28 | 40 | 45 | 20 |
| **Ghassabi's** | 28 | 40 | 45 | 20 |
| **D-Saddle** | 28 | 40 | 45 | 20 |
| **Proposed** | 28 | 40 | 45 | 20 |

**Table 4.5: Extraction accuracy (%) of feature points on retinal vessels for each dataset.**

| Feature Extraction Methods | Datasets | | | | Overall[2] |
|---|---|---|---|---|---|
| | CHASE[1] | DRIVE[1] | HRF[1] | STARE[1] | |
| **Harris** | 19.792% | 54.379% | 47.393% | 33.623% | 41.613% |
| **SIFT** | 19.615% | 19.204% | 10.085% | 18.930% | 16.164% |
| **SURF** | 18.888% | 16.390% | 22.310% | 16.462% | 18.929% |
| **Ghassabi's** | 23.340% | 29.106% | 28.604% | 32.813% | 28.280% |
| **D-Saddle** | 14.737% | 22.432% | 21.451% | 22.622% | 20.509% |
| **Proposed** | ***72.462%*** | ***89.763%*** | ***89.378%*** | ***89.963%*** | ***86.021%*** |

[1]Computed from the extraction accuracy of individual image. The extraction accuracy for an image is set to 0% when the feature points are extracted less than 3.

[2]Computed from the summation of all image at once rather than the mean of each dataset to minimize the accumulation of rounding error.

extraction method successfully met the minimum requirement of 3 feature points extracted from each image in the datasets as presented in Table 4.3. The extracted feature points on four datasets constitute for an average extraction accuracy of 86.021%. The highest extraction accuracy of the proposed feature extraction method is observed on STARE dataset (89.963%) and the lowest on CHASE_DB1 dataset (72.462%) as outlined in Table 4.5. Also, the extraction accuracy of the proposed feature extraction method varies approximately 9% between the images in the datasets. This variation is represented by the standard deviation computed from the extraction accuracy of all images as given in Table 4.6 and visualised in Figure 4.27.

**Table 4.6: Descriptive statistics of extraction accuracy (%) for all dataset.**

| Feature Extraction Methods | Total Image | Mean | Std. Dev. | Std. Error | 95% Confidence Interval for Mean | | Min | Max |
|---|---|---|---|---|---|---|---|---|
| | | | | | Lower Bound | Upper Bound | | |
| **Harris** | 133 | 41.613 | 21.317 | 1.848 | 37.956 | 45.269 | 0.000 | 92.857 |
| **SIFT** | 133 | 16.164 | 5.411 | 0.469 | 15.236 | 17.092 | 5.241 | 30.299 |
| **SURF** | 133 | 18.929 | 4.206 | 0.365 | 18.208 | 19.651 | 9.502 | 30.412 |
| **Ghassabi's** | 133 | 28.280 | 5.975 | 0.518 | 27.255 | 29.305 | 17.055 | 44.197 |
| **D-Saddle** | 133 | 20.509 | 4.791 | 0.415 | 19.687 | 21.330 | 12.221 | 31.273 |
| **Proposed** | 133 | 86.021 | 9.199 | 0.798 | 84.443 | 87.599 | 59.677 | 97.842 |

Std. Dev. : Standard deviation.
Std. Error : Standard error.

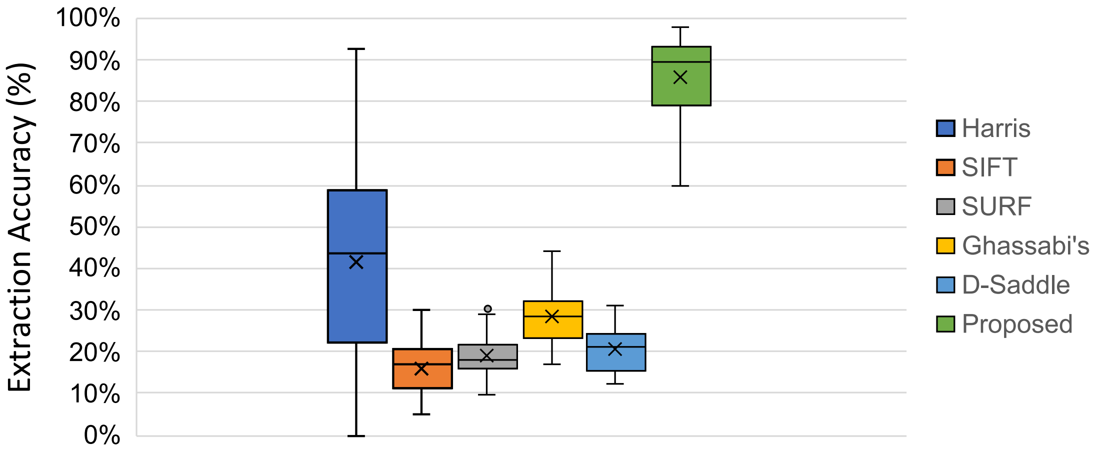



**Figure 4.27: Boxplots of extraction accuracy (%) for all images in CHASE_DB1, DRIVE, HRF and STARE datasets.**

Examples of the extracted feature points with the highest and lowest extraction accuracy in each dataset for the proposed feature extraction method are depicted in Figure 4.28 and Figure 4.29, respectively. From these figures, the proposed feature extraction method mainly extracts the feature points located on retinal vessels. The extracted feature points that associated with noise are observed to be on nerve retinal fibre layer, underlying choroidal vessels, microaneurysm and exudates as shown in Figure 4.29(b), (d), (f) and (h). These feature points are extracted because the noises depict similar characteristics as the retinal vessels within the local patches during the processes in feature detection and feature selection modules.

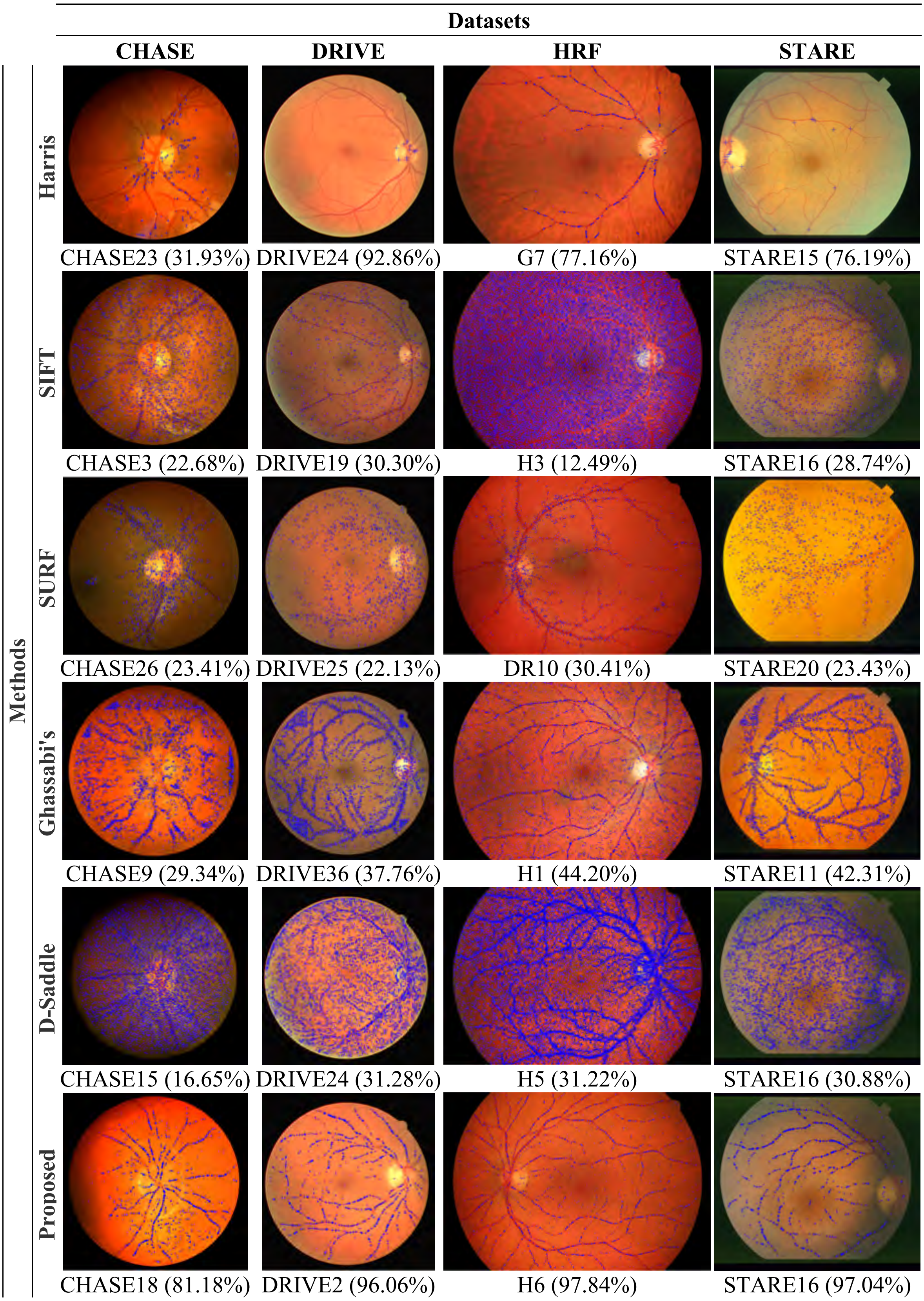



**Figure 4.28: Extracted feature points for the image with the highest extraction accuracy in the datasets.**

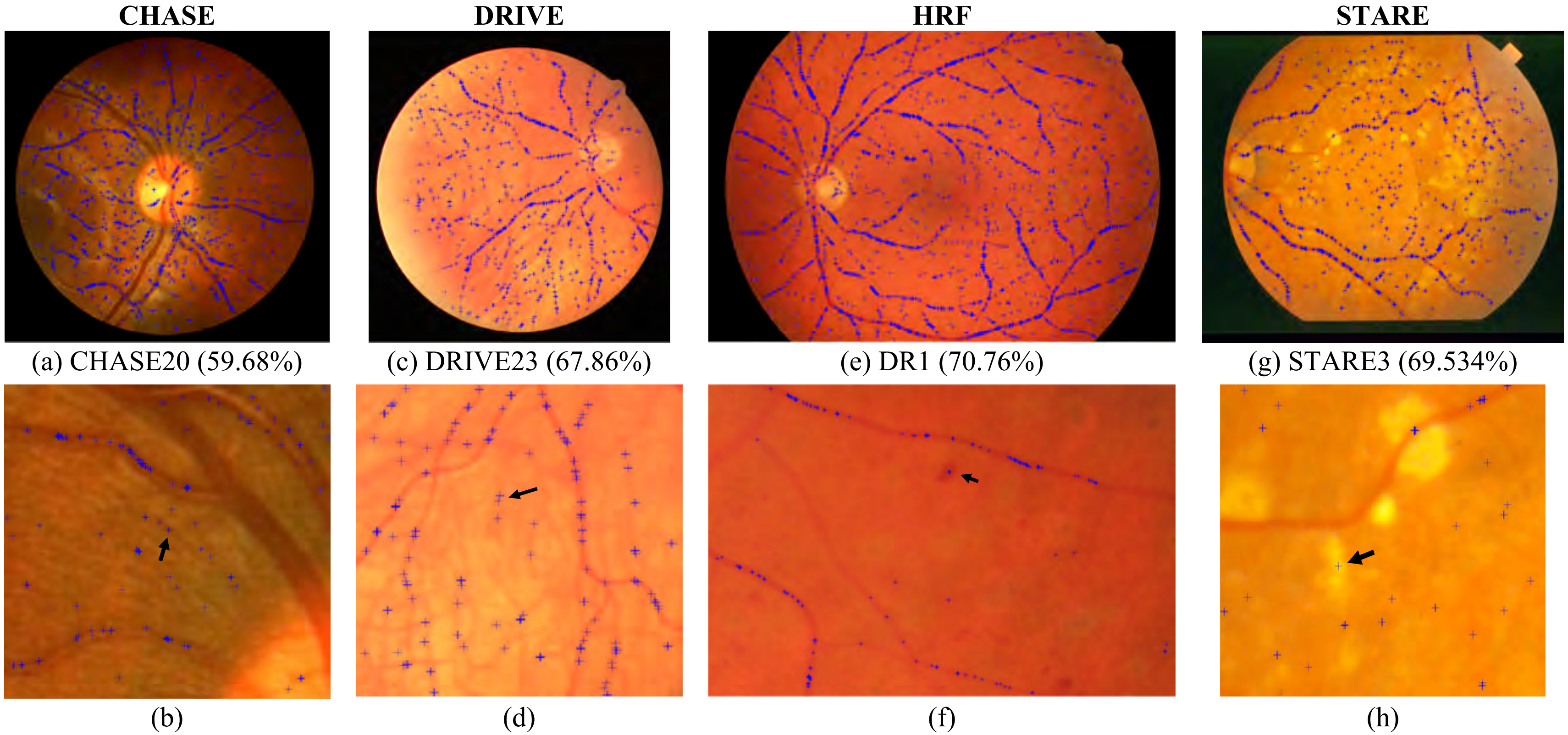



**Figure 4.29: Extracted feature points with the lowest extraction accuracy for the proposed feature extraction method in the datasets. The black arrow point to the extracted feature points on noise: (b) retinal nerve fibre layer, (d) underlying choroidal vessels, (f) microaneurysm and (h) exudates.**

**Table 4.7: Comparisons of extraction accuracy (%) using one-way ANOVA and Tukey's post hoc.**

| Feature Extraction Methods (I) | (J) | Mean Difference (I-J) | Std. Error | $p$ | 95% Confidence Interval Lower Bound | Upper Bound |
|---|---|---|---|---|---|---|
| **Proposed** | **Harris** | 44.408 | 1.271 | <0.001* | 40.777 | 48.039 |
| | **SIFT** | 69.857 | 1.271 | <0.001* | 66.226 | 73.488 |
| | **SURF** | 67.091 | 1.271 | <0.001* | 63.460 | 70.722 |
| | **Ghassabi's** | 57.741 | 1.271 | <0.001* | 54.110 | 61.372 |
| | **D-Saddle** | 65.512 | 1.271 | <0.001* | 61.881 | 69.143 |

* : The mean difference is significant at the 0.05 level.
$p$ : Statistical significance.
Std. Error : Standard error.

Specifically, these noises appear as continuous curvature structure in 3-dimensional, which results in a nearly straight and wide connected region in the binary interpolated patch. Therefore, the sum of the intensity profiles extracted along the connected region in the patch appears as a horizontal line. Furthermore, the sum of the intensity profiles from the gradient interpolated patch for these noises resemble the inverse Gaussian-like shape with unique characteristics that similar to retinal vessels. These unique characteristics described in Section 4.3 are presented as exclusion criteria of STEP 5. In the selection module, these noises also exhibit high similarity as the retinal vessel described by three attributes highlighting the size and contrast of the vessel (see STEP 6(b)). Thus, these noises are mistakenly selected as the final feature points.

Nevertheless, the amount of the feature points on noises mistakenly extracted by the proposed feature extraction method is smaller compared to other methods. This can be observed in the analysis of one-way ANOVA performed between the extraction accuracy of all feature extraction methods. The performed one-way ANOVA analysis shows that the proposed feature extraction method is significantly outperformed the others ($p$ = <0.001*) as presented in Table 4.7. The average extraction accuracy of the proposed feature extraction method has the biggest difference with SIFT (69.857%) and the smallest difference with Harris (44.408%).

The high extraction accuracy of the proposed feature extraction method on four datasets compared to other methods demonstrates its ability to discriminate between retinal vessels and noises. This is because the proposed feature extraction method considers the characteristics that are unique and similar between retinal vessels and noises in extracting the feature points. Contrarily, Ghassabi's and D-Saddle focuses their attention on enhancing the image to increase the visibility of the retinal vessel. However, this enhancement also increases the visibility of the noises in the image. Other than that, they performed the extraction without considering the noises with similar characteristics as the retinal vessels. Consequently, led both of the feature extraction methods to have a low extraction accuracy wherein Ghassabi's attained an average of 28.280% while D-Saddle attained an average of 20.509%. The extracted feature points from these methods that located on noises are observed to be on edge of optic disc, retinal nerve fibre layer, underlying choroidal vessels and macula.

The other feature extraction methods such as Harris, SIFT and SURF have no specific feature selection module for extracting retinal vessels in their scheme. These feature extraction methods were used in the existing feature-based RIR techniques (J. Chen et al., 2010; Hernandez-Matas et al., 2016; Yang et al., 2007) where the authors focused their work on the development of the descriptor and transformation model. Among them, SIFT obtained the lowest average extraction accuracy. SIFT is not associated with any prominent structure during the extraction process, which leads to a high amount of feature points being extracted on the retinal surface. A similar issue is also observed in SURF method. In contrast, Harris corner attained the second highest of the average extraction

**Table 4.8: Correlations between extraction accuracy (%) and factors.**

| Factors | | Extraction Accuracy | | | | | |
|---|---|---|---|---|---|---|---|
| | | **Harris** | **SIFT** | **SURF** | **Ghassabi's** | **D-Saddle** | **Proposed** |
| Image size | $r_s$ | -0.178 | -0.649 | 0.590 | -0.142 | -0.138 | -0.032 |
| | $p$ | 0.041* | <0.001** | <0.001** | 0.104 | 0.114 | 0.712 |
| Intensity distribution | $r_s$ | 0.360 | 0.138 | -0.398 | 0.314 | 0.386 | 0.342 |
| | $p$ | <0.001** | 0.113 | <0.001** | <0.001** | <0.001** | <0.001** |

$r_s$ : Spearman's rho. Value closes to 0 indicates that the correlation is weak, where the factor has minimal influence on the success rate.
$p$ : Statistical significance.
** : Correlation is significant at the 0.01 level (2-tailed).
* : Correlation is significant at the 0.05 level (2-tailed).

accuracy among all the feature extraction methods because the corner structure is considered in their scheme.

### 4.6.2 Factors

Factors such as image size and intensity distribution can affect the performance of the feature extraction method. The Spearman's rank-order correlation is computed between the factors and the extraction accuracy to measure their relationship as presented in Table 4.8. The changes in image size have the weakest influence on the extraction accuracy of the proposed feature extraction method compared to others ($r_s$ = -0.032, $p$ = 0.712). The key to this performance is that the proposed feature extraction method has minimal usage of rigid threshold or rigid variables. In opposite, SIFT is very sensitive to the changes in image size with the strongest correlation among the methods ($r_s$ = -0.649, $p$ = <0.001$^{**}$). The correlation shows that the extraction accuracy of SIFT decreases when the image size is larger.

The extraction accuracy of the proposed feature extraction method is significantly correlated to the intensity distribution in the image ($r_s$ = 0.342, $p$ = <0.001$^{**}$) indicating its performance decreases in the image with less uniform intensity distribution. The proposed feature extraction method is highly dependent on the intensity changes to locate the curvature structure of the retinal vessels in the feature detection module. Furthermore, the proposed feature extraction method does not incorporate any feature enhancement algorithm in building the hierarchical scale space. For these reasons, the performance of the proposed feature extraction method is affected in the presence of the non-uniform intensity distribution in the image. The feature enhancement algorithm such as DoG and *ii*DoG operators can suppress the non-uniform intensity distribution in the image and increases the visibility of the retinal vessels. However, at the cost increasing the visibility of the noises, thus, it is avoided in the proposed feature extraction method. Contrarily, the correlation between SIFT and the intensity distribution is not significant and the weakest among all the methods ($r_s = 0.138$, $p = 0.113$).

## 4.7 SUMMARY

The proposed feature extraction method finds and selects feature points on retinal vessels. The retinal vessel is chosen as a feature because it can be found throughout the fundus image, repeatable between images and reliable in the unhealthy fundus image. The proposed feature extraction method finds the extrema on the curvature structure as candidate feature points by comparing the intensity values within the local patches. The curvature structure in fundus image represents the shape of the retinal vessel in 3-dimensional as well as noises such as nerve retinal fibre layer, underlying choroidal vessels and exudates. Therefore, the feature selection module is introduced to distinguish between the retinal vessels and noises.

The feature selection module is composed of exclusion and selection processes. The exclusion process removes the candidate feature points on noises based on five criteria. These criteria are inferred from the sum of the pixel intensity profiles of the retinal vessels and noises. Then, the selection process is performed on the remaining candidate feature points. The distribution coefficient and selection weightage are calculated for each candidate feature point in the selection process. The selection process is responsible for ensuring the selected feature points are distributed throughout the image and selected according to the strength of the retinal vessel attributes.

The proposed feature extraction method is evaluated in terms of its accuracy in extracting feature points on retinal vessels. There are four public datasets employed in this evaluation, namely, CHASE_DB1, DRIVE, HRF and STARE. These datasets include fundus images of various pathological cases with the ground truth of manually segmented vessels performed by experts. The performance of the proposed feature extraction method is evaluated in these datasets and compared with five feature extraction methods that were used in the existing feature-based RIR techniques. These methods are Harris corner, SIFT, SURF, Ghassabi's and D-Saddle.

The one-way ANOVA analysis performed between the extraction accuracy of all feature extraction methods shows that the proposed feature extraction method is significantly outperformed the others with the highest average extraction accuracy of 86.021% ($p$ = <0.001$^{*}$). Also, the accuracy difference between the proposed feature extraction method and the others ranges from 44.408% to 69.857%. Other than that, factors such as changes in image size and intensity distribution are investigated for their influence on the extraction accuracy. The proposed feature extraction method has shown the weakest correlation with the changes in image size among all the methods ($r_s$ = -0.032, $p$ = 0.712). The weakest influence of the image size on the proposed method is contributed by the minimal usage of rigid threshold or rigid variables. However, the performance of the proposed feature extraction method is affected in the presence of the non-uniform intensity distribution in the image ($r_s$ = 0.342, $p$ = <0.001$^{**}$). The proposed feature extraction method is sensitive to the non-uniform intensity distribution because its highly dependent on the intensity changes in a local patch to locate the curvature structure.

Furthermore, the feature enhancement algorithm such as DoG and *ii*DoG operators is not incorporated in building the hierarchical scale space.

## 4.8 CONCLUSIONS

This chapter presents the proposed feature extraction method for feature-based RIR technique to achieve objectives RO1 and RO2. The proposed feature extraction method is composed of feature detection module and feature selection module, which considers the general and unique characteristics of the retinal vessels as well as noises. The average extraction accuracy of the proposed feature extraction method on four datasets is significantly outperformed Harris corner, SIFT, SURF, Ghassabi's and D-Saddle.

# CHAPTER 5: FEATURE DESCRIPTOR

## 5.1 INTRODUCTION

This chapter presents the proposed feature descriptor method for Stage 3 of the feature-based fundus image registration (RIR) technique. Section 5.2 discusses the issues in the existing feature descriptor methods. Section 5.3 explains the proposed feature descriptor method characterising the circular region of the extracted feature points with statistical properties. The processes of matching (Stage 4) and estimating geometrical transformation (Stage 5) are also explained in Section 5.4 and Section 5.5, respectively. The evaluation presented in Section 5.7 focuses on the assessment of the proposed feature-based RIR technique that combines all five main stages as previously shown in Figure 3.1. Finally, the presented work is summarised and concluded in Section 5.8 and Section 5.9, respectively. The mathematical symbols and notation used in this chapter can be found in Appendix A.

## 5.2 ISSUES IN EXISTING FEATURE DESCRIPTOR METHODS

From the highlighted feature descriptor methods in Section 2.4.2, the new feature descriptor methods are mainly proposed to address various issues in registering multi-modality retinal images as shown in Table 2.2. In our work, we focus on the registration of single modality between colour fundus images. For registering single modality between colour fundus images, Sajib K. Saha et al. (2018), Ramli et al. (2017b), Ramli et al. (2017a), Hernandez-Matas et al. (2017a), Hernandez-Matas et al. (2017b), Hernandez-Matas et al. (2016), S. K. Saha et al. (2016), Hernandez-Matas et al. (2015) and Ghassabi et al. (2015) employed feature descriptor methods such as SIFT (Lowe, 2004), HOG (Dalal & Triggs, 2005), SURF (Bay et al., 2008; Bay et al., 2006), PIIFD (J. Chen et al., 2010), BRISK (Leutenegger et al., 2011), BRIEF (Calonder et al., 2012) and ALOHA (S. Saha & Démoulin, 2012) without any improvement or modifications. The majority of these feature descriptor methods are mainly based on the local gradient direction. The local gradient direction is computed from grids within a square patch that is rotated according to the dominant orientation surrounding the feature point. Rotating the patch allows the assigned descriptor to achieve rotation invariance.

The local patches with retinal vessels especially in the presence of bifurcation, low contrast or narrowed vessels often lack of textural information and exhibit repetitive patterns (Abràmoff et al., 2010; Deng et al., 2010). These properties of the retinal vessels can hinder from the feature descriptor methods based on local gradient direction to distinctively characterise the feature points on retinal vessels (Fang et al., 2019; Hinterstoisser et al., 2012; Kushnir & Shimshoni, 2014). The lack of distinctive descriptor computed to characterise the feature points will cause a high number of mismatches, which can lead to inaccurate estimation of geometrical transformation. For this reason, the feature descriptor method based on the local gradient direction can be insufficient to characterise the feature points on retinal vessels. Therefore, the proposed feature descriptor method considers statistical properties to distinctively characterise the circular region of the extracted feature points.

## 5.3 PROPOSED FEATURE DESCRIPTOR

The proposed feature descriptor method describes the feature points with information from circumferences of circles surrounding the feature points. Example of the circumferences with radiuses of $r_1, r_2, r_3 \dots r_R$ surrounding $n$-th feature point $(D_n)$ is depicted in Figure 5.1. Computing the proposed feature descriptor includes two main steps. First, pixels that lie on circumferences surrounding a feature point are located using the midpoint circle algorithm. Second, the descriptor is computed from the pixels that lie on the circumferences.

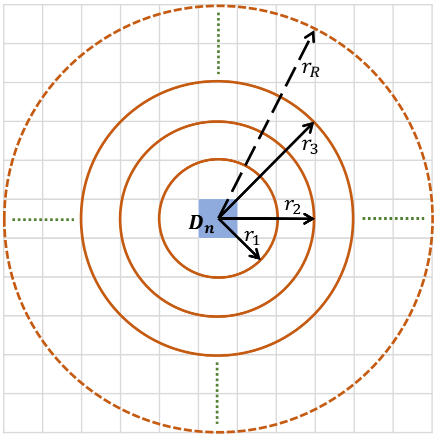



**Figure 5.1: Circumferences surrounding a feature point $D_n$ with radiuses of $r_1$, $r_2$, $r_3$... $r_n$.**

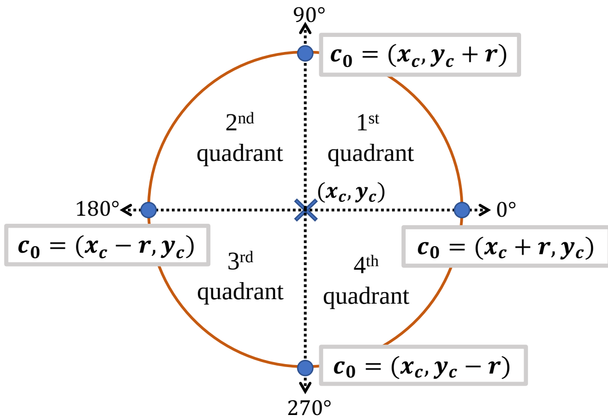



**Figure 5.2: Four possible positions for the initial pixel $c_0$ on the circumference.**

### 5.3.1 Locating Pixels on Circumferences

Generally, the position of a pixel on a circumference for a circle with known radius and centre can be determined based on circle function, $g(x, y)$ as follows

$$g(x, y) = (x - x_c)^2 + (y - y_c)^2 - r^2 \tag{5.1}$$

where, $x_c$ and $y_c$ are the centre coordinate of the circle, $x$ and $y$ are the coordinate of a pixel, $r$ is the radius of the circle. The coordinate of the initial pixel on the circumference, $c_0 = (x_0, y_0)$ can easily be found at the quadrant of the circle. There are four possible initial pixels can be chosen on the circumference as given in Equation (5.2) and visualised in Figure 5.2.

$$\begin{array}{llll} \text{First quadrant} & : & x_0 = x_c + r & , \quad y_0 = y_c \\ \text{Second quadrant} & : & x_0 = x_c & , \quad y_0 = y_c + r \\ \text{Third quadrant} & : & x_0 = x_c - r & , \quad y_0 = y_c \\ \text{Fourth quadrant} & : & x_0 = x_c & , \quad y_0 = y_c - r \end{array} \tag{5.2}$$

Lets assume the initial pixel of the circumference is at $c_0 = (x_c, y_c + r)$. The coordinate for the next pixel $c_1 = (x_1, y_1)$ in the clockwise direction can be chosen either from the pixel at position $(x_0 + 1, y_0)$ or $(x_0 + 1, y_0 - 1)$ as shown in Figure 5.3. To

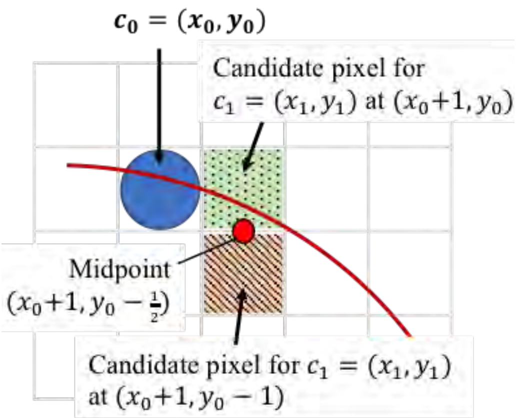



**Figure 5.3: Two possible positions for the next pixel $c_1$.**

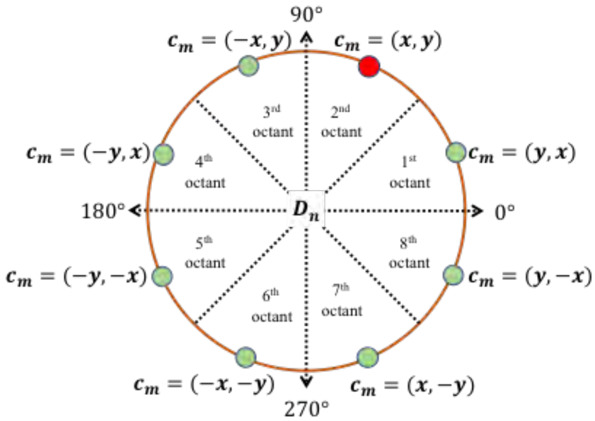



**Figure 5.4: Eight-way symmetry approach.**

choose one of these pixels, the decision parameter denoted by $\varphi$ is computed. The decision parameter is inferred from the midpoint coordinate between these two pixels i.e. $(x_0 + 1, y_0 - \frac{1}{2})$ as given below:

$$\varphi_1 = g\left(x_0 + 1, y_0 - \frac{1}{2}\right) = (x_0 + 1)^2 + (y_0 - \tfrac{1}{2})^2 - r^2 \qquad (5.3)$$

The value of the decision parameter $\varphi$ indicates the position of the midpoint coordinate between the two candidate pixels relative to the circumference. For example, the decision parameter with the value of $\varphi_1 \leq 0$ indicates that the midpoint coordinate is within the circle. Accordingly, the pixel at position $(x_0 + 1, y_0)$ will be chosen as the next pixel $c_1$ as it is closer to the circumference. If the decision parameter $\varphi_1 > 0$, the midpoint coordinate between the two candidate pixels is outside the circle. Thus, the pixel at position $(x_0 + 1, y_0 - 1)$ will be chosen as the next pixel $(x_1, y_1)$ because it is closer to the circumference. The same processes of calculating decision parameter and choosing the next pixel are repeated to find the subsequent pixels on the circumference. These processes can be simplified as follows:

$$c_{m+1} = \begin{cases} (x_m + 1, y_m)\,, & \varphi_m \le 0 \\ (x_m + 1, y_m - 1)\,, & \varphi_m > 0 \end{cases}$$

$$\varphi_{m+1} = \begin{cases} \varphi_m + (2x_m + 3)\,, & \varphi_m \le 0 \\ \varphi_m + 2(x_m - y_m) + 5\,, & \varphi_m > 0 \end{cases} \tag{5.4}$$

where, $c_{m+1}$ is the pixel of index $m+1$ on a circumference and $m$ is the index of the pixel on a circumference.

To speed up the computation, eight-way symmetry approach is used to allow the identification of the pixels on the circumference only from an octant of the circle. The eight-way symmetry approach is performed by dividing the circle into eight octants of identical size. By calculating the position of a pixel $c_m = (x_m, y_m)$ in the second octant (shown by the red circle), the position of the corresponding pixels in other octants (shown by the green circle) can be found as illustrated in Figure 5.4.

### 5.3.2 Computing Feature Descriptor

After the pixels on the circumferences of all radiuses are obtained, their grayscale levels are extracted. Then, the summation, mean and standard deviation are calculated for each circumference. These values are concatenated according to the summation, mean and standard deviation values to form a descriptor as follows:

$$d_n = [sum_1, sum_2 \dots sum_R, mean_1, mean_2 \dots mean_R, std_1, std_2 \dots std_R] \tag{5.5}$$

where, $d_n$ is the feature descriptor representing the $n$-th feature point, $R$ is the total radius, $sum_R$ is the summation of the grayscale levels from circumference with radius $r_R$, $mean_R$ is the mean of the grayscale levels from the circumference with radius $r_R$ and $std_R$ is the standard deviation of the grayscale levels from the circumference with radius $r_R$.

The radiuses to find the circumferences are empirically set from 1 to 55 with an increment of 1. These radiuses are chosen by carefully examining the width of the retinal

| $r_1$ | $r_2$ | $\cdots$ | $r_R$ | $r_1$ | $r_2$ | $\cdots$ | $r_R$ | $r_1$ | $r_2$ | $\cdots$ | $r_R$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $sum_1$ | $sum_2$ | $\cdots$ | $sum_R$ | $mean_1$ | $mean_2$ | $\cdots$ | $mean_R$ | $std_1$ | $std_2$ | $\cdots$ | $std_R$ |

**Figure 5.5: The proposed feature descriptor represents the concatenated information of the summation, mean and standard deviation for each circumference.**

vessels as in Section 4.4.2.2 and assuming the scale or zoom difference is minimal. The concatenated values of summation, mean and standard deviation results in a descriptor with a size of $K = 165$ to represent a feature point as depicted in Figure 5.5.

## 5.4 MATCHING

The matches are obtained by establishing the pairwise distance between the feature descriptors in the fixed and moving images. The distances between the feature descriptors are computed using the sum of squared differences (SSD) defined as follows:

$$SSD(d_{fn}, d_{mn}) = \sum_{k=1}^{K} \left[d_{fn}(k) - d_{mn}(k)\right]^2, \quad k = 1,2,3 \dots K \tag{5.6}$$

where, $SSD(d_{fn}, d_{mn})$ is the sum of squared differences between the feature descriptor $d_{fn}$ and $d_{mn}$, $d_{fn}$ is the feature descriptor representing $fn$-th feature point in the fixed image, $dm_n$ is the feature descriptor representing $mn$-th feature point in the moving image, $k$ is the index of the element in the feature descriptor and $K$ is the size of the feature descriptor.

The estimated SSD between the feature descriptors are organised in a *FN*-by-*MN* correspondence matrix. *FN* is the total feature points extracted from the fixed image and *MN* is the total feature points extracted from the moving image. Each column of the correspondence matrix represents the distances between a feature descriptor $d_{fn}$ in the fixed image to all feature descriptors in moving image, $d_{mn}, d_{mn+1}, d_{mn+2} \dots d_{MN}$. The smallest SSD value in each column represent the putative matches between the feature

| | Feature points in fixed image ($D_{fn}$) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Feature points in moving image ($D_{mn}$) | Feature Point 1 ($D_{f1}$) | … | Feature Point 24 ($D_{f24}$) | Feature Point 25 ($D_{f25}$) | Feature Point 26 ($D_{f26}$) | Feature Point 27 ($D_{f27}$) | Feature Point 28 ($D_{f28}$) | … | Feature Point $FN$ ($D_{FN}$) |
| Feature Point 1 ($D_{m1}$) | | | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | | |
| ⋮ | | | | | | | | | |
| Feature Point 15 ($D_{m15}$) | ............. | | 72748014 | 100559899 | 2902491 | 7482315 | 77449269 | | ................. |
| Feature Point 16 ($D_{m16}$) | ............. | | 62685150 | 86741799 | 2691336 | 6514540 | 60733004 | | ................. |
| Feature Point 17 ($D_{m17}$) | ............. | | 48477816 | 69030706 | 1147950 | 3881051 | 46873974 | | ................. |
| Feature Point 18 ($D_{m18}$) | ............. | | 53856554 | 74791592 | 2228630 | 5933122 | 50231204 | | ................. |
| Feature Point 19 ($D_{m19}$) | ............. | | 81538950 | 108003611 | 4907156 | 11959152 | 78866767 | | ................. |
| ⋮ | | | | | | | | | |
| Feature Point $MN$ ($D_{MN}$) | | | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | | |

**Figure 5.6: Example of the smallest SSD value in a column (highlighted) of the correspondence matrix that represents the putative match.**

points in the fixed image to their correspond feature points in the moving images as shown in Figure 5.6. However, these putative matches contain a high percentage of outliers.

The outliers are eliminated from the putative matches using M-estimator SAmple Consensus (MSAC) algorithm (Torr & Zisserman, 2000). MSAC eliminates the outliers when the square root for the sum of squares distance between the matches exceeded a specified maximum distance. The matches in the moving image are projected based on the nonreflective similarity transformation estimated from two randomly selected putative matches. In this study, the maximum distance is set to values that ranged between 3 and 60 with an increasing step of 3. The random trial is repeated 5000 times while the desired confidence to find the maximum number of inliers is set to 99%.

## 5.5 GEOMETRICAL TRANSFORMATION

The inliers obtained during the matching process are used in estimating geometrical transformation between fixed and moving images. This study utilised local weighted mean transformation to estimate the geometrical transformation. The local weighted mean transformation is inferred from a second-degree polynomial transformation within a radius of influence across the image (Goshtasby, 1988). The radius of influence for an inlier is calculated from the inlier itself to the furthest neighbouring $w$ inliers. The value of $w$ is ranged between 10 and the total inliers with an increasing step of 10. Then, the moving image is transformed to the orientation of the fixed image according to the estimated geometrical transformation.

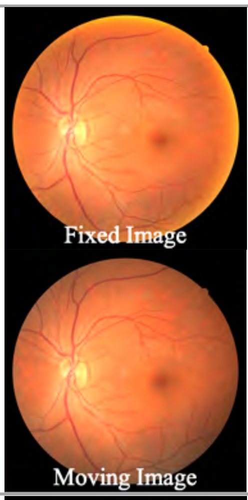
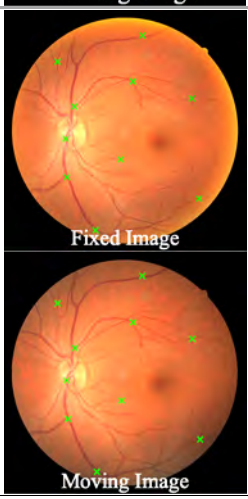
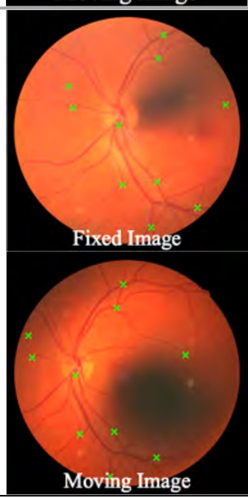
## 5.6 EXPERIMENTAL SETUP

The performance of the proposed feature-based RIR technique is evaluated in this chapter, which includes five main stages as previously depicted in Figure 3.1. These stages are pre-processing (Section 3.4), proposed feature extraction method (Section 4.4), proposed feature descriptor method (Section 5.3), matching (Section 5.4) and estimating geometrical transformation (Section 5.5). The performance of the proposed feature-based RIR technique is compared with existing feature-based RIR techniques, namely, GDB-ICP (Yang et al., 2007), Harris-PIIFD (J. Chen et al., 2010), Ghassabi's-SIFT (Ghassabi et al., 2015), H-M 16 (Hernandez-Matas et al., 2016), H-M 17 (Hernandez-Matas et al., 2017a) and D-Saddle-HOG (Ramli et al., 2017b). The feature-based RIR techniques of Ghassabi's-SIFT and D-Saddle-HOG are developed according to its original settings, whereas the performances for GDB-ICP, Harris-PIIFD, H-M 16 and H-M 17 are obtained from (*FIRE: Fundus Image Registration Dataset*). The dataset and the metrics used in evaluating all the feature-based RIR techniques are explained in the following sub-sections.

### 5.6.1 Datasets

The proposed and existing feature-based RIR techniques are evaluated on FIRE: Fundus Image Registration dataset (Hernandez-Matas et al., 2017c). This dataset is the only retinal fundus image registration dataset with ground truth annotation that available publicly. The fundus images were acquired using a Nidek AFC-210 fundus camera. The images are acquired at the resolution of 2912×2912 pixels and 45° field of view (FOV). The FIRE dataset consists of 134 image pairs wherein each pair comprises of fixed and moving images. The ground truth annotations are in the form of corresponding points

**Table 5.1: Descriptions of FIRE dataset.**

| Descriptions | Category $\mathcal{S}$ | Category $\mathcal{P}$ | Category $\mathcal{A}$ |
|---|---|---|---|
| Application | Super-resolution | Image Mosaicking | Longitudinal Study |
| Total image pairs | 71 | 49 | 14 |
| Resolution (pixels) | 2912×2912 | 2912×2912 | 2912×2912 |
| Total patients | | 39 | |
| Age | | 19 – 67 | |
| Pathological case | | Diabetic retinopathy | |
| Anatomical differences between fixed and moving images | No | No | Yes (increased of vessel tortuosity, microaneurysms, cotton-wool, spots) |
| Field of view | | 45º | |
| Year | | 2006 to 2015 | |
| Ground truth | | 10 corresponding points for each image pair | |
| Author(s) | | Hernandez-Matas, Zabulis, Triantafyllou, Anyfanti, Douma, and Argyros (2017c) | |
| Examples (Fundus) |  |  |  |
| Examples (Ground truth) |  |  |  |

between fixed and moving images. A total of 10 corresponding points identified by the experts are provided for each image pair in the dataset. Examples of the image pairs and its ground truth annotation are given in Table 5.1.

The image pairs in FIRE dataset are grouped into three registration applications i.e., super-resolution (Category $\mathcal{S}$), image mosaicking (Category $\mathcal{P}$) and longitudinal study (Category $\mathcal{A}$). All the image pairs are affected by diabetic retinopathy. The presence of the vessel tortuosity, microaneurysms and cotton-wool are also visible on the images. The anatomical changes due to progression or remission of retinopathy can be observed between fixed and moving images in the longitudinal study application. However, the anatomical appearance between fixed and moving images remains unchanged in the super-resolution and image mosaicking applications. Other than that, the image pairs in the image mosaicking application have a smaller overlapping area that ranges between 17% to 89% and a larger range of rotation between 6° to 52° compared to other categories. Detailed properties of the image pairs in FIRE dataset can be found in Table 5.2.

### 5.6.2 Evaluation Metrics

#### 5.6.2.1 Target Registration Error (TRE)

The proposed and existing feature-based RIR techniques are evaluated by measuring the target registration error (TRE). TRE is an average distance measured in pixel from 10 corresponding ground truth annotations between fixed image and transformed moving image. TRE can be computed as expressed below:

$$TRE = \frac{1}{B_{total}} \sum_{B=1}^{B_{total}} \sqrt{(x_{B_{transform}} - x_{B_{fixed}})^2 - (y_{B_{transform}} - y_{B_{fixed}})^2} \tag{5.7}$$

where, $B$ is the index of the corresponding ground truth annotations, $B_{total}$ is the total of the corresponding ground truth annotations, $x_{B_{fixed}}$ and $y_{B_{fixed}}$ are the coordinate of the $B$-th ground truth annotation on the fixed image, $x_{B_{transform}}$ and $y_{B_{transform}}$ are the coordinate of the $B$-th ground truth annotation on the transformed moving image. As described in Section 5.4 and Section 5.5, multiple values are assigned for the maximum distance of MSAC and neighbouring $w$ inliers of local weighted mean transformation to register an image pair. For each combination of these values, TRE is measured and the best (minimum) TRE value is selected as the final result for the respective image pair.

#### 5.6.2.2 Success Rate

The registration of an image pair is considered successful if the TRE value is below than 1 pixel for super-resolution application and 5 pixels for image mosaicking and longitudinal study applications (Matsopoulos, Asvestas, Mouravliansky, & Delibasis, 2004). The registration with TRE larger than these values for the respective application is considered as failed. Then, the success rate is computed from the successful registration given as follows

$$Success\ rate(\%) = \frac{total\ successful\ registration}{total\ image\ pairs} \times 100\% \tag{5.8}$$

The success rate expressed the ability of a feature-based RIR technique to register image pairs and meet the specified TRE requirement for a particular application. The statistical tests of one-way Analysis of Variance (ANOVA) with Tukey's post hoc are performed to compare the registration success rate between the feature-based RIR techniques.

#### 5.6.2.3 Factors

Factors influencing the performance of the feature-based RIR techniques are investigated by establishing the Spearman's rank-order correlation between the factors and success rate. The factors investigated are rotation, overlapping area and image quality. As these factors are not initially provided by the FIRE dataset, the factors are measured as follows.

The rotation for an image pair is measured from the average angle between corresponding ground truth annotations. The overlapping area perceives in percentage is

**Table 5.2: Properties of image pairs in FIRE dataset.**

| Properties/Factors (Average) | Category $\mathcal{S}$ (Super-resolution) | Category $\mathcal{P}$ (Image Mosaicking) | Category $\mathcal{A}$ (Longitudinal Study) |
|---|---|---|---|
| Scale | ≈1 | ≈1 | ≈1 |
| Overlapping Area (%) | 86 – 100 | 17 – 89 | 95 – 100 |
| Rotation (°) | 0° – 12° | 6° – 52° | 1° – 4° |
| Image Quality: | | | |
| • Intensity difference (MSE)[1] | 17 – 1558 | 117 – 1069 | 47 – 1740 |
| • Difference of intensity distribution ($UN_{diff}$)[3] | 0 – 36 | 0 – 23 | 1 – 26 |
| • Structure similarity (SSIM)[2] | 0.779 – 0.940 | 0.834 – 0.925 | 0.823 – 0.918 |

[1]Larger value indicates a higher intensity difference between images in the pair.
[2]Value close to 1 indicates higher structure similarity between images in the pair.
[3]Value close to 0 indicates a similar intensity distribution between images in the pair.

obtained from the overlap area between fixed image and transformed moving image. The moving image is transformed to the orientation of the fixed image using affine transformation inferred from the corresponding ground truth annotations.

Three metrics are employed to perceive the image quality of FIRE dataset; mean squared error (MSE), difference of peak deviation non-uniformity ($UN_{diff}$) and structure similarity index (SSIM). MSE represents the intensity difference to describe if the intensity on one of the images in the pair is brighter than the other. $UN_{diff}$ signifies if the intensity distribution on one of the images in the pair is more uniformed than the other. $UN_{diff}$ is an absolute value and quantified by subtracting UN measurements between fixed and moving images. SSIM describes the similarity of the structure component, which indicates if the vessel edges on one of the images in the pair have a higher level of blurring effect compared to the other. The blurring effect on the vessel edges can be caused by motion or improper focusing.

The measurement of these factors for each category in FIRE dataset are summarised as in Table 5.2. A value of SSIM that close to 1 indicates a high similarity of the structure component between images, whereas a high similarity of intensity and intensity distribution between images are approximated by a smaller value of MSE and $UN_{diff}$. All the factors are measured on grayscale images at the original resolution. Also, it should be noted that the values of the image quality metrics are subjected to image size, conversion of coloured image to grayscale, image filtering and image enhancement. Thus, the values may be varied between studies. The presented measurements of the image quality metrics are obtained from the grayscale images without any image filtering and image enhancement. The conversion to the grayscale image is performed as in Equation (3.1).

Factors influencing the registration success rate of the feature-based RIR techniques are investigated using Spearman's rank-order correlation. The explanation of the Spearman's rho ($r_s$) value can be found in Section 4.5.2.2. The one-way ANOVA and Spearman's rank-order correlation are significant at the 0.05 level identified by a single asterisk or at the 0.01 level identified by two asterisks.

## 5.7 RESULTS & DISCUSSION

This section presents and discusses the registration performance of the proposed and existing feature-based RIR techniques in FIRE dataset. The registration performance of the proposed and existing feature-based RIR techniques were evaluated and measured at the original resolution of the FIRE dataset. The performance of the feature-based RIR techniques in registering image pairs for super-resolution, image mosaicking and longitudinal study applications are compared and discussed in Section 5.7.1. Then, the factors influencing the registration performance are presented in Section 5.7.2.

### 5.7.1 Registration Accuracy

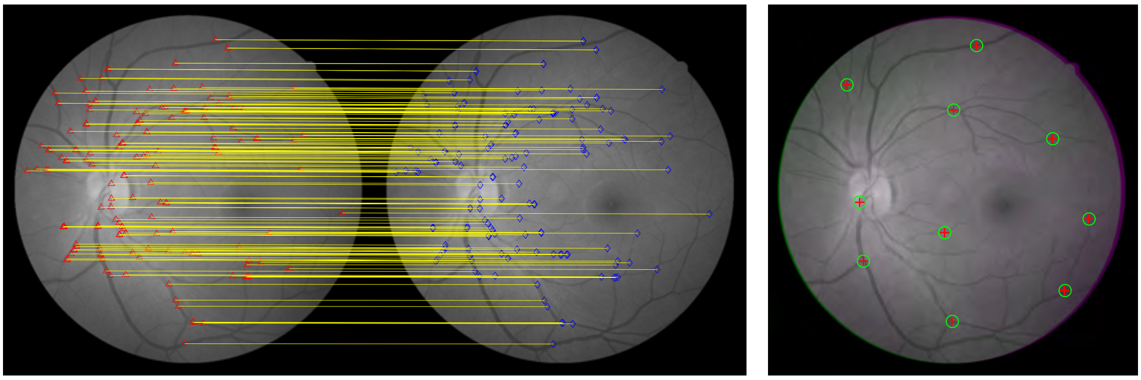
Overall, the proposed feature-based RIR technique successfully registered 67.164% of the image pairs in the FIRE dataset as presented in Table 5.3. By application, the proposed feature-based RIR technique successfully registered 75.510% of the image pairs in the image mosaicking application, 66.197% of the image pairs in the super-resolution application and 42.857% of the image pairs in the longitudinal study application. The

**Table 5.3: Descriptive statistics of the registration accuracy for FIRE dataset. The highest success rate is marked with bold and italic font.**

| Feature-based RIR Techniques | Success rate (%) | | TRE of successful registration (pixels) | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Total Image[1] | Mean | Mean | Std. Dev. | 95% Confidence Interval for Mean | | Min | Max |
| | | | | | Lower | Upper | | |
| **Overall** | | | | | | | | |
| **GDB-ICP** | 37 | 27.612 | 1.988 | 1.268 | 1.566 | 2.411 | 0.486 | 4.952 |
| **Harris-PIIFD** | 5 | 3.731 | 2.573 | 1.613 | 0.571 | 4.576 | 0.785 | 4.244 |
| **Ghassabi's-SIFT** | 17 | 12.687 | 1.529 | 1.352 | 0.834 | 2.225 | 0.665 | 4.917 |
| **H-M 16** | 22 | 16.418 | 1.232 | 0.865 | 0.849 | 1.616 | 0.554 | 3.315 |
| **H-M 17** | 26 | 19.403 | 1.399 | 1.162 | 0.929 | 1.868 | 0.489 | 4.754 |
| **D-Saddle-HOG** | 16 | 11.940 | 2.166 | 1.736 | 1.240 | 3.091 | 0.748 | 4.738 |
| **Proposed** | 90 | ***67.164*** | 1.892 | 1.301 | 1.619 | 2.164 | 0.444 | 4.797 |
| **Category $\mathcal{S}$ (Super-resolution)** | | | | | | | | |
| **GDB-ICP** | 17 | 23.944 | 0.783 | 0.155 | 0.703 | 0.863 | 0.486 | 0.988 |
| **Harris-PIIFD** | 2 | 2.817 | 0.846 | 0.086 | 0.071 | 1.621 | 0.785 | 0.907 |
| **Ghassabi's-SIFT** | 13 | 18.310 | 0.834 | 0.134 | 0.753 | 0.915 | 0.665 | 0.996 |
| **H-M 16** | 18 | 25.352 | 0.839 | 0.118 | 0.780 | 0.897 | 0.554 | 0.995 |
| **H-M 17** | 20 | 28.169 | 0.800 | 0.145 | 0.732 | 0.868 | 0.489 | 0.988 |
| **D-Saddle-HOG** | 10 | 14.085 | 0.895 | 0.090 | 0.831 | 0.960 | 0.748 | 0.999 |
| **Proposed** | 47 | ***66.197*** | 0.818 | 0.143 | 0.776 | 0.860 | 0.444 | 0.998 |
| **Category $\mathcal{P}$ (Image Mosaicking)** | | | | | | | | |
| **GDB-ICP** | 16 | 32.653 | 3.068 | 0.840 | 2.620 | 3.516 | 1.946 | 4.952 |
| **Harris-PIIFD** | 0 | 0 | N/A[2] | N/A[2] | N/A[2] | N/A[2] | N/A[2] | N/A[2] |
| **Ghassabi's-SIFT** | 0 | 0 | N/A[2] | N/A[2] | N/A[2] | N/A[2] | N/A[2] | N/A[2] |
| **H-M 16** | 0 | 0 | N/A[2] | N/A[2] | N/A[2] | N/A[2] | N/A[2] | N/A[2] |
| **H-M 17** | 1 | 2.041 | 3.327 | N/A[2] | N/A[2] | N/A[2] | 3.327 | 3.327 |
| **D-Saddle-HOG** | 2 | 4.082 | 3.532 | 0.636 | -2.186 | 9.250 | 3.082 | 3.982 |
| **Proposed** | 37 | ***75.510*** | 3.067 | 0.965 | 2.745 | 3.388 | 1.094 | 4.797 |
| **Category $\mathcal{A}$ (Longitudinal Study)** | | | | | | | | |
| **GDB-ICP** | 4 | 28.571 | 2.792 | 0.558 | 1.903 | 3.680 | 2.354 | 3.603 |
| **Harris-PIIFD** | 3 | 21.429 | 3.725 | 0.473 | 2.551 | 4.899 | 3.319 | 4.244 |
| **Ghassabi's-SIFT** | 4 | 28.571 | 3.789 | 0.882 | 2.386 | 5.191 | 3.082 | 4.917 |
| **H-M 16** | 4 | 28.571 | 3.004 | 0.214 | 2.664 | 3.343 | 2.857 | 3.315 |
| **H-M 17** | 5 | 35.714 | 3.408 | 0.758 | 2.467 | 4.349 | 2.920 | 4.754 |
| **D-Saddle-HOG** | 4 | 28.571 | 4.658 | 0.070 | 4.546 | 4.769 | 4.583 | 4.738 |
| **Proposed** | 6 | ***42.857*** | 3.060 | 0.725 | 2.299 | 3.821 | 2.352 | 3.927 |

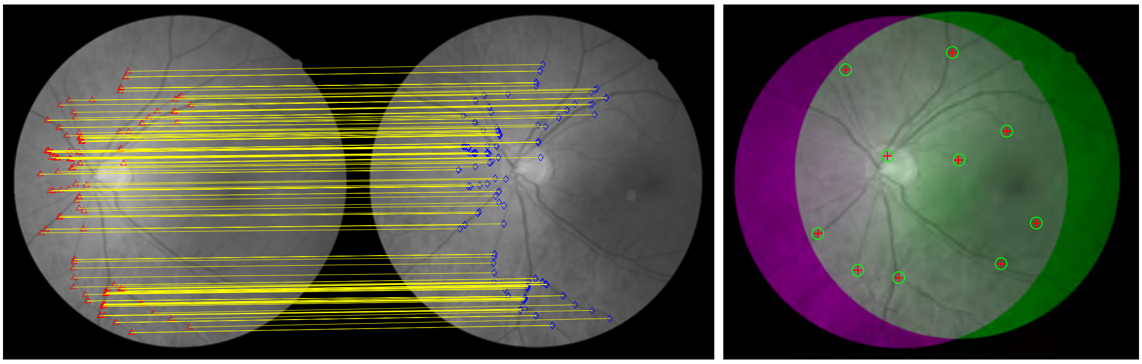
Std. Dev. : Standard deviation.
[1]Total image pairs with successful registration.
[2]Not available. Insufficient number of image pairs with TRE of successful registration available to compute the respective measurement.

example of the successfully registered image pair in each application for the proposed feature-based RIR technique is depicted in Figure 5.7.



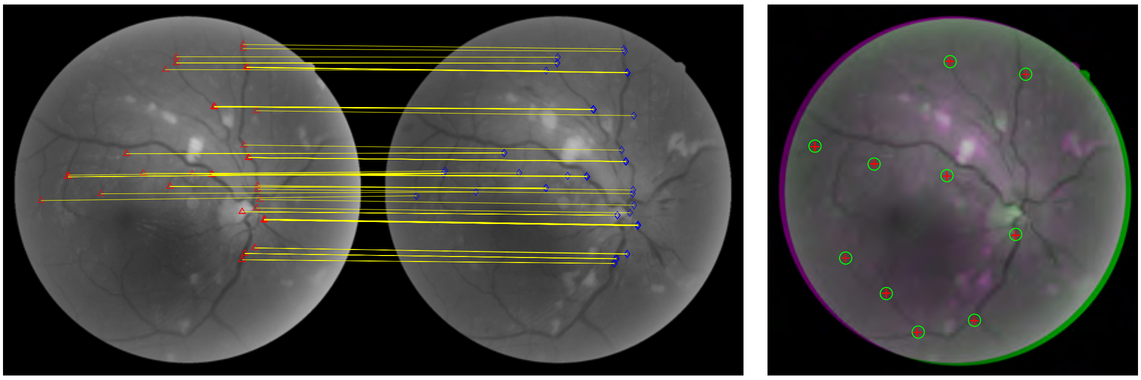
(i) Fixed Image (ii) Moving Image (iii) Registered Image

(a) Super-resolution Application (Pair $\mathcal{S}12$)



(i) Fixed Image (ii) Moving Image (iii) Registered Image

(b) Image Mosaicking Application (Pair $\mathcal{P}7$)



(i) Fixed Image (ii) Moving Image (iii) Registered Image

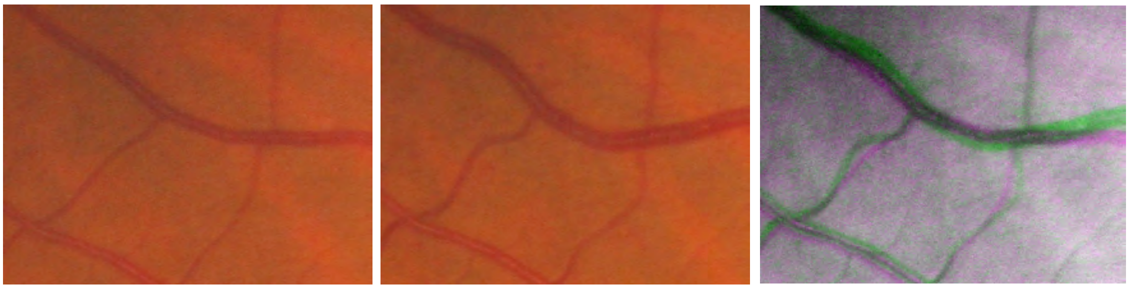
(c) Longitudinal Study Application (Pair $\mathcal{A}4$)

**Figure 5.7: Example of successfully registered image pair in each application for the proposed feature-based technique. The yellow lines indicate the inliers between (i) fixed and (ii) moving images. The green 'o' and red '+' in (iii) registered image indicate the ground truth annotations.**

The proposed feature-based RIR technique attained the lowest success rate in the longitudinal study application. This is because the anatomical appearance differs between images in the pair as the images are taken separated over a long period of time. Particularly, the proposed feature-based RIR technique failed to register image pairs when a substantial difference in vessel tortuosity and thickness are visible between fixed and moving images. The vessel tortuosity is a vascular anomaly that can affect either the small part or entire vascular tree (Ramos, Novo, Rouco, Romeo, Álvarez, & Ortega, 2018). The affected part of the retinal vessel appears as twisted or curved compared to its normal version in which, straight or gently curved. Furthermore, the vessel tortuosity can shift the physical position of the affected vessel on the retina as depicted in Figure 5.8. This condition can cause the correctly established inliers on retinal vessels to have incomparable physical position. The inliers with incomparable physical position can lead to inaccurate estimation of the geometrical transformation particularly, when the local transformation is employed. The local transformation applies a different transformation to a different part of the image and offers a high degree of flexibility to project the curved object such as the retina.

The proposed feature-based RIR technique employed the local weighted mean transformation in its framework. The local weighted mean transformation is a local transformation, which computes the second-degree polynomial transformation for each inlier within a radius of influence from the inlier itself to the furthest neighbouring $w$ inliers. As the distances between the neighbouring $w$ inliers of the proposed feature-based RIR technique are not too close or sparse, the effect of the incomparable physical position can be minimized when the tortuosity difference is small. Therefore, the proposed feature-based RIR technique successfully registered image pairs with small difference in vessel tortuosity but failed when the tortuosity difference is substantial.

Other than vessel tortuosity, a substantial difference of vessel thickness between fixed and moving images also affecting the registration performance of the proposed feature-based RIR technique. The proposed feature descriptor method relies heavily on the structure of the retinal vessel to compute a descriptor. Accordingly, the feature point that lies on a thicker vessel in fixed image is represented by different descriptor than its



(a) Fixed image (b) Moving image (c) Registered image

**Figure 5.8: Difference of vessel tortuosity observed in the image pair of the longitudinal study application.**

corresponding feature point on a thinner vessel in moving image. The lack of similarity between the corresponding descriptors caused a sparse and insufficient amount of the inliers being established. The geometrical transformation estimated from these inliers is inaccurate and lead to failed registration of the image pairs with a substantial difference in the vessel thickness.

The proposed feature-based RIR technique yields a higher success rate in the image mosaicking and super-resolution applications compared to its performance in the longitudinal study application. Mainly because of the anatomical appearance remain unchanged between the images for these two applications as they were acquired during the same examination period. However, registering image pairs from these applications exhibit other forms of challenges.

For example, the super-resolution application that consists of image pairs with a large overlapping area and a small rotation requires a very accurate registration with an error less than 1 pixel. Due to the requirement of a very accurate registration, the proposed feature-based RIR technique failed to register 33.803% of the image pairs from the super-resolution application. The TRE obtained for the failed image pairs are in between 1.034 pixels to 3.866 pixels.

**Table 5.4: Comparisons of success rate (%) between the proposed feature-based RIR technique and others using one-way ANOVA and Tukey's post hoc.**

| Feature-based RIR Techniques (I) | (J) | Mean Difference (I-J) | Std. Error | $p$ | 95% Confidence Interval Lower Bound | Upper Bound |
|---|---|---|---|---|---|---|
| **Overall** | | | | | | |
| **Proposed** | **GDB-ICP** | 39.552 | 4.555 | <0.001* | 26.090 | 53.010 |
| | **Harris-PIIFD** | 63.433 | 4.555 | <0.001* | 49.970 | 76.890 |
| | **Ghassabi's-SIFT** | 54.478 | 4.555 | <0.001* | 41.020 | 67.940 |
| | **H-M 16** | 50.746 | 4.555 | <0.001* | 37.290 | 64.210 |
| | **H-M 17** | 47.761 | 4.555 | <0.001* | 34.300 | 61.220 |
| | **D-Saddle-HOG** | 55.224 | 4.555 | <0.001* | 41.760 | 68.680 |
| **Category $\mathcal{S}$ (Super-resolution)** | | | | | | |
| **Proposed** | **GDB-ICP** | 42.254 | 6.687 | <0.001* | 22.46 | 62.05 |
| | **Harris-PIIFD** | 63.380 | 6.687 | <0.001* | 43.58 | 83.18 |
| | **Ghassabi's-SIFT** | 47.887 | 6.687 | <0.001* | 28.09 | 67.68 |
| | **H-M 16** | 40.845 | 6.687 | <0.001* | 21.05 | 60.64 |
| | **H-M 17** | 38.028 | 6.687 | <0.001* | 18.23 | 57.82 |
| | **D-Saddle-HOG** | 52.113 | 6.687 | <0.001* | 32.32 | 71.91 |
| **Category $\mathcal{P}$ (Image Mosaicking)** | | | | | | |
| **Proposed** | **GDB-ICP** | 42.857 | 5.255 | <0.001* | 27.269 | 58.446 |
| | **Harris-PIIFD** | 75.510 | 5.255 | <0.001* | 59.922 | 91.099 |
| | **Ghassabi's-SIFT** | 75.510 | 5.255 | <0.001* | 59.922 | 91.099 |
| | **H-M 16** | 75.510 | 5.255 | <0.001* | 59.922 | 91.099 |
| | **H-M 17** | 73.469 | 5.255 | <0.001* | 57.881 | 89.058 |
| | **D-Saddle-HOG** | 71.429 | 5.255 | <0.001* | 55.840 | 87.017 |
| **Category $\mathcal{A}$ (Longitudinal Study)** | | | | | | |
| **Proposed** | **GDB-ICP** | 14.286 | 17.908 | 0.985 | -39.71 | 68.28 |
| | **Harris-PIIFD** | 21.429 | 17.908 | 0.894 | -32.57 | 75.43 |
| | **Ghassabi's-SIFT** | 14.286 | 17.908 | 0.985 | -39.71 | 68.28 |
| | **H-M 16** | 14.286 | 17.908 | 0.985 | -39.71 | 68.28 |
| | **H-M 17** | 7.143 | 17.908 | 1.000 | -46.86 | 61.14 |
| | **D-Saddle-HOG** | 14.286 | 17.908 | 0.985 | -39.71 | 68.28 |

* : The mean difference is significant at the 0.05 level.
$p$ : Statistical significance.
Std. Error : Standard error.

Contrarily, registering image pairs from the image mosaicking application is the most challenging among the applications due to the combination of small overlapping area and a large rotation. Despite the challenge, the proposed feature-based RIR technique successfully register 75.510% of the image pairs in the image mosaicking application. The proposed feature-based RIR technique failed to register 24.490% of the image pairs with the failed TRE ranged between 5.213 pixels to 273.91 pixels.

The proposed feature-based RIR technique is further evaluated by comparing its performance with six existing feature-based RIR techniques using one-way ANOVA analysis as presented in Table 5.4. The analysis shows that the overall success rate of the proposed feature-based RIR technique is significantly outperformed the others ($p$ = $<0.001^{*}$). Among the existing feature-based RIR techniques, GDB-ICP obtained the highest overall success rate of 27.612% while Harris-PIIFD attained the lowest overall success rate with only 3.731%.

Image pairs from the image mosaicking application are the most challenging to register for Harris-PIIFD, Ghassabi's-SIFT, H-M 16, H-M 17 and D-Saddle-HOG. All of them obtained the lowest success rate in this application compared to their performance in the super-resolution and longitudinal study applications. Furthermore, Harris-PIIFD, Ghassabi's-SIFT and H-M 16 failed to register any of the image pairs from the image mosaicking application whereas H-M 17 and D-Saddle-HOG only obtained success rate of 2.041% and 4.082%, respectively. However, GDB-ICP exhibits a contrast performance where its success rate of 32.653% in image mosaicking application is the highest compared to its performance in other applications.

H-M 17 outperformed the other existing feature-based RIR techniques in the super-resolution and longitudinal study applications with success rate of 28.169% and 35.714%, respectively. In the longitudinal study application, all feature-based RIR techniques faced a similar challenge to register image pairs with a difference in vessel tortuosity and vessel thickness. However, the existing feature-based RIR techniques failed to register these image pairs including those with a small difference in vessel tortuosity and vessel thickness. In contrast, the proposed feature-based RIR technique only failed to register image pairs with a substantial difference of vessel tortuosity and thickness.

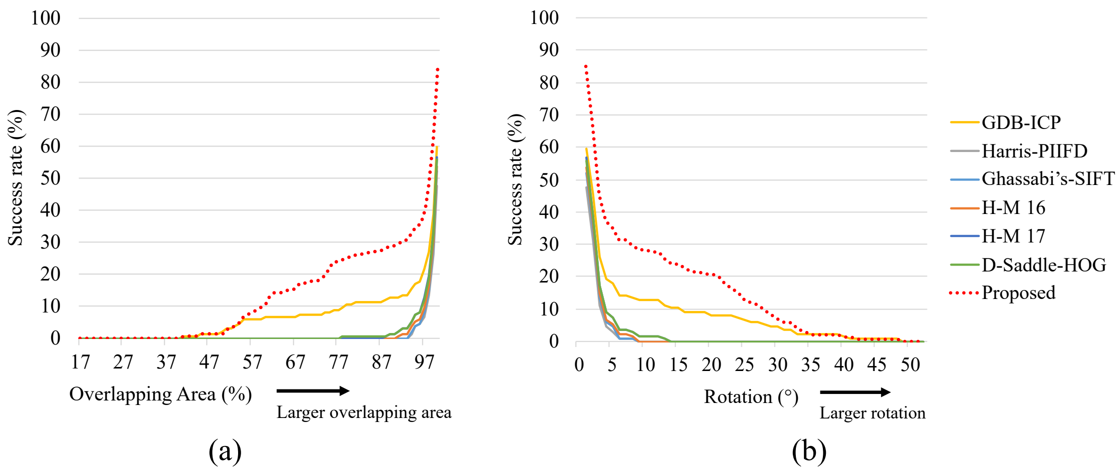



**Figure 5.9: Relations between success rate and (a) overlapping area (b) rotation.**

### 5.7.2 Factors

There are three factors considered in this section; overlapping area, rotation and image quality between fixed and moving images. The influences of the factors on the success rate are investigated by examining their relations as illustrated in Figure 5.9 and computing the Spearman's rank-order correlation as outlined in Table 5.5. The Spearman's rho ($r_s$) and statistical significance ($p$) are discussed and compared between the feature-based RIR techniques. For this evaluation, the successful registration is set at TRE less than 5 pixels despite their registration application and the success rate of all the image pairs in the FIRE dataset are considered in the analysis.

#### 5.7.2.1 Overlapping Area

Registering an image pair with a small overlapping area can be challenging due to a limited intersected region available between images and prior knowledge of the overlapping area is unavailable. To successfully register a retinal image pair with a small overlapping area, it is crucial to extract feature points located on retinal vessels and distributed throughout the image. This will increase the chance of the inliers being established between images and then accurately estimate the geometrical transformation.

**Table 5.5: Correlations between success rate (%) and factors. The weakest correlation for each factor is marked with bold and italic font.**

| Factors | | | Success Rate | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | GDB-ICP | Harris-PIIFD | Ghassabi's-SIFT | H-M 16 | H-M 17 | D-Saddle-HOG | Proposed |
| Overlapping Area (%) | | $r_s$ | 0.443 | 0.732 | 0.795 | 0.785 | 0.773 | 0.769 | ***0.286*** |
| | | $p$ | <0.001$^{**}$ | <0.001$^{**}$ | <0.001$^{**}$ | <0.001$^{**}$ | <0.001$^{**}$ | <0.001$^{**}$ | ***0.001***$^{**}$ |
| Rotation (°) | | $r_s$ | -0.380 | -0.723 | -0.766 | -0.763 | -0.765 | -0.745 | ***-0.261*** |
| | | $p$ | <0.001$^{**}$ | <0.001$^{**}$ | <0.001$^{**}$ | <0.001$^{**}$ | <0.001$^{**}$ | <0.001$^{**}$ | ***0.002***$^{**}$ |
| Image quality[1] | Intensity difference (MSE) | $r_s$ | ***-0.117*** | -0.244 | -0.187 | -0.197 | -0.257 | -0.235 | -0.261 |
| | | $p$ | ***0.177*** | 0.004$^{**}$ | 0.031$^{*}$ | 0.022$^{*}$ | 0.003$^{**}$ | 0.006$^{**}$ | 0.002$^{**}$ |
| | Difference of intensity distribution ($UN_{diff}$) | $r_s$ | 0.172 | 0.178 | 0.206 | 0.208 | 0.210 | 0.199 | ***0.011*** |
| | | $p$ | 0.047$^{*}$ | 0.039$^{*}$ | 0.017$^{*}$ | 0.016$^{*}$ | 0.015$^{*}$ | 0.021$^{*}$ | ***0.896*** |
| | Structure similarity (SSIM) | $r_s$ | -0.103 | -0.018 | -0.085 | -0.057 | ***-0.017*** | -0.027 | 0.074 |
| | | $p$ | 0.238 | 0.837 | 0.331 | 0.512 | ***0.849*** | 0.760 | 0.396 |

$r_s$ : Spearman's rho. Value closes to 0 indicates that the correlation is weak, where the factor has minimal influence on the success rate.
$p$ : Statistical significance.
** : Correlation is significant at the 0.01 level (2-tailed).
* : Correlation is significant at the 0.05 level (2-tailed).
[1]Image quality between fixed and moving images.

The overlapping area between fixed and moving images in the FIRE dataset ranges from 17% to 100%. Generally, the success rate of all feature-based RIR techniques are positive and significantly influenced by the presence of the overlapping area ($r_s > 0$, $p = <0.001^{**}$). The positive correlation specifies that the success rate increases with the increment of the overlapping area.

As presented in Section 4.6, the proposed feature extraction method has demonstrated the ability to accurately extract feature points on retinal vessels of various datasets. This ability explains the weakest correlation between the success rate of the proposed feature-based RIR technique and the overlapping area compared to the existing feature-based RIR techniques ($r_s = 0.286$, $p = 0.001^{**}$). Additionally, the success rate of the proposed feature-based RIR technique decreases at a slower rate than the existing feature-based RIR techniques as the overlapping area becoming smaller as shown in Figure 5.9(a). The registration ability of the proposed feature-based RIR technique is limited when the overlapping area is smaller than 41%.

A similar limitation is also observed on the performance of GDB-ICP but with a more sensitive relation with the overlapping area ($r_s = 0.443$, $p = <0.001^{**}$). GDB-ICP utilises SIFT detector as feature extraction in their scheme. The matches established from the feature points initiate the iterative registration process. The registration process expands surrounding the initial matches by mapping the “corner” or “face” points. However, the established initial matches can be incorrect as the SIFT feature points are extracted from various part of the retina. The incorrect initial matches can falsely initiate the iterative registration process which results in inaccurate estimation of the geometrical transformation. Nevertheless, the success rate of GDB-ICP declines at a slower rate as the overlapping area becoming smaller compared to the other existing feature-based RIR techniques. This performance is contributed by their unique iterative registration process.

The other existing feature-based RIR techniques of Harris-PIIFD, Ghassabi's-SIFT, H-M16, H-M 17 and D-Saddle-HOG performed similarly when the overlapping area becoming smaller given by close Spearman's rho values that ranged between $r_s = 0.732$ and $r_s = 0.795$, ($p = <0.001^{**}$). These feature-based RIR techniques are very sensitive in the presence of the small overlapping area when failed to register image pairs that are approximately smaller than 87%.

#### 5.7.2.2 Rotation

The correlation computed between success rate and rotation express two capabilities of the feature descriptor method. First is the capability to assign distinctive information representing the corresponded feature points. A descriptor with distinctive information allows correct matches or inliers to be established and results in an accurate estimation of the geometrical transformation. This capability is studied in this section by examining the success rate when the rotation between images is minimal ($\leq 1°$). Second is the capability to assign similar descriptor to the corresponded feature points even in the presence of the rotation. This capability is presented by examining the success rate when the angle of the rotation varies between images.

The feature descriptor methods in the existing feature-based RIR techniques are generally based on the gradient direction computed from the grids within a square patch. These feature descriptor methods are HOG, SIFT and PIIFD. Both HOG and PIIFD compute the descriptor from a square patch with a fixed size of 16-by-16 pixels and 40-by-40 pixels, respectively. In contrast, SIFT employed a varying size of the square patch to compute the descriptor. The size of the square patch, which ranges between 8-by-8 pixels to 14-by-14 pixels is varied according to the octave position of the feature point.

For the first evaluation of the feature descriptor capability to assign distinctive information, the existing feature-based RIR techniques achieved a much lower success rate than the proposed feature-based RIR technique when the rotation is minimal as depicted in Figure 5.9(b). The success rates of the existing feature-based RIR techniques are led by GDB-ICP (59.701%) and followed by H-M 17 (56.716%), D-Saddle-HOG (55.970%), H-M 16 (53.731%) and Harris-PIIFD (47.761%). These performances demonstrate that the descriptor based on the gradient direction information is insufficient and indistinguishable to represent corresponded feature points on fundus images. Particularly, when the descriptor is computed for the feature points on the noises with a similar characteristic as the vessels, which can lead to a high amount of the incorrect matches.

Obtaining distinctive information from the retinal vessels in fundus image to represent the corresponded feature points can be challenging as the retinal vessels are lack of textural information and comprise of repetitive patterns (Abràmoff et al., 2010; Deng et al., 2010). For those reasons, we concatenate statistical properties of summation, mean and standard deviation derived from the pixels on circumferences surrounding the feature point. The concatenated statistical properties are calculated from circumferences with radiuses ranged between 1 to 55 pixels, which helps to distinguish the feature points on retinal vessels with various thickness. These statistical properties are chosen as they are repeatable and distinguishable between corresponded feature points. The proposed feature-based RIR technique achieved the highest success rate (85.075%) among all when the rotation is minimal. The highest success rate proves that the proposed feature descriptor method assigns highly distinctive information to represent the corresponded feature points.

In the second evaluation, the capability of the feature descriptor in assigning similar descriptor to corresponded feature points in the presence of various rotation is studied. The rotation between fixed and moving images in FIRE dataset ranges between 0° to 52°.

A negative and significant correlation between the success rate and the rotation is observed for all feature-based RIR techniques as presented in Table 5.5. The negative correlation implies that the success rate decreases as the rotation is larger.

The rotation has the weakest impact on the proposed feature-based RIR technique represented by the smallest Spearman's rho compared to others ($r_s$ = -0.261, $p$ = 0.002$^{**}$). This performance is contributed by the ability of the proposed feature descriptor method to assign similar descriptor representing the corresponded feature points in the presence of rotation between images. The proposed feature-based RIR technique outperformed the others up to the rotation angle of 33° as depicted in Figure 5.9(b). Beyond the angle of 33°, the proposed feature-based RIR technique attained a similar performance as GDB-ICP. Both the proposed feature-based RIR technique and GDB-ICP capable of registering image pairs with rotation between 0° to 48° but failed to register when the rotation is larger than 49°. However, GDB-ICP is more sensitive ($r_s$ = -0.380, $p$ = <0.001$^{**}$) to the presence of the rotation indicated by a steeper trendline in Figure 5.9(b) compared to the proposed feature-based RIR technique. The other existing feature-based RIR techniques performed similarly in the presence of the rotation with close Spearman's rho values that ranged between $r_s$ = -0.723 and $r_s$ = -0.766.

#### 5.7.2.3 Image Quality

Components of image quality investigated in this study are intensity difference, difference of intensity distribution and similarity of structure component. These components are perceived by mean squared error (MSE), difference of peak deviation non-uniformity ($UN_{diff}$) and structure similarity index (SSIM), respectively. Details measurement of these components on FIRE dataset is summarised in Table 5.2.

First, Spearman's correlation is computed between the success rate and MSE to investigate the influence of the intensity difference on the registration performance.

Particularly, to investigate the ability of the feature descriptor method to similarly describe the corresponded regions with a difference in intensity. Among the feature-based RIR techniques, intensity difference has a minimal impact on the success rate of GDB-ICP expressed by the smallest Spearman's rho and insignificant correlation ($r_s$ = -0.117, $p$ = 0.177) as listed in Table 5.5. In contrast, the proposed feature-based RIR technique has a significant and the strongest correlation with intensity difference compared to others ($r_s$ = -0.261, $p$ = 0.002$^{**}$). This is because of the proposed feature descriptor method is highly dependent on the grayscale level to describe the circular region surrounding the feature point. Thus, dissimilar descriptors are assigned to represent the corresponded feature points which lead to incorrect matches. For this reason, we employed MSAC algorithm in our feature-based RIR technique to remove the incorrect matches or outliers. Removing the outliers reduces the sensitivity of the proposed feature-based RIR technique towards the intensity difference.

The second component of the image quality investigated is the difference of the intensity distribution between fixed and moving images perceived by $UN_{diff}$. The non-uniform intensity distribution in fundus image is generally caused by human error or uneven absorption of light due to the spherical retina. The examples of the non-uniform intensity distribution found in fundus image are excessive light exposure near the frame boundary and dark or white spot. The non-uniform intensity distribution can obscure the visibility of the retinal vessels in the affected area. This condition limits the intersected region available between images which present a similar challenge as registering image pair with a small overlapping area.

Therefore, it is crucial for the feature extraction method to extract feature points that are distributed all over the image. A highly distributed feature points will ensure a high number of matches to be established between the unaffected area. The success rate of the proposed feature-based RIR technique has the weakest and insignificant correlation with $UN_{diff}$ among the feature-based RIR techniques ($r_s$ = 0.011, $p$ = 0.896). The correlation signifies the ability of the proposed feature extraction method to extract feature points that are distributed throughout the image. Contrarily, the existing feature-based RIR techniques are significantly affected by $UN_{diff}$ with Spearman's rho values ranged between $r_s$ = 0.172 to $r_s$ = 0.210.

Next, we investigate the correlation between the success rate and the similarity of the structure component. The similarity of the structure component measured by SSIM describes the clarity or sharpness of the vessel edges between fixed and moving images. The similarity of the structure component has a minimal impact on the registration performance of all feature-based RIR techniques represented by the insignificant correlations between the success rate and SSIM. Mainly, because the feature descriptor methods in the feature-based RIR techniques do not rely on the vessel edges to compute the descriptor.

## 5.8 SUMMARY

The proposed feature descriptor method describes the extracted feature points with concatenated values of summation, mean and standard deviation. These statistical properties are calculated from the pixels on circumferences surrounding the feature points. The radiuses to find the circumferences are set from 1 to 55 pixels with an increment of 1. In the evaluation, the proposed feature-based RIR technique is evaluated, which include pre-processing (Section 3.4), proposed feature extraction method (Section 4.4), proposed feature descriptor method (Section 5.3), matching (Section 5.4) and estimating geometrical transformation (Section 5.5). The proposed feature-based RIR technique was tested on FIRE dataset and compared with six existing feature-based RIR techniques, namely, Harris-PIIFD, GDB-ICP, Ghassabi's-SIFT, H-M 16, H-M 17 and D-

Saddle-HOG. The evaluation is conducted in two parts. The first part of the evaluation investigates the performance of the feature-based RIR techniques in registering image pairs from super-resolution, image mosaicking and longitudinal study applications. Then, factors influencing the registration performance of the feature-based RIR techniques were investigated.

In the first part of the evaluation, the proposed feature-based RIR technique obtained the lowest success rate in the longitudinal study application (42.857%) compared to its performance in other applications. The proposed feature-based RIR technique failed to register the image pairs in this application when a substantial difference in vessel tortuosity and thickness are present between images. The vessel tortuosity is a vascular anomaly that causes the vessels to twist or curve over a period of time. Furthermore, the physical position of the affected vessels shifts from its previous position as the tortuosity increased. This causes the established inliers to be on the corresponded vessels with an incomparable position on the retina. As the local transformation is employed in the proposed feature-based RIR technique, the individual transformation applied to these inliers lead to inaccurate registration of the image pair with a substantial difference in vessel tortuosity. Other than that, a substantial difference in vessel thickness between images in the pair can lead to dissimilar descriptor being computed for the corresponded feature points. This is because the proposed feature descriptor method relies heavily on the structure of the retinal vessel to compute a descriptor.

The proposed feature-based RIR technique yielded the highest success rate in the image mosaicking application (75.510%) and followed by the super-resolution application (66.197%) as the anatomical appearance remain unchanged between images. However, registering image pairs from super-resolution application requires a very accurate registration with an error less than 1 pixel whereas registering image pairs from image mosaicking application requires the ability to register image pairs with a combination of small overlapping area and large rotation. The analysis of one-way ANOVA shows that the proposed feature-based RIR technique is significantly outperformed the others with an overall success rate of 67.164% ($p$ = <0.001$^{*}$). Among the existing feature-based RIR techniques, GDB-ICP obtained the highest overall success rate (27.612%) while Harris-PIIFD attained the lowest overall success rate (3.731%).

In the second part of the evaluation, three factors are investigated, namely, overlapping area, rotation and image quality between fixed and moving images. The components of the image quality include intensity difference, difference of intensity distribution and similarity of structure component. The influences of these factors on the success rates are investigated by computing the Spearman's rank-order correlations.

The factor of the overlapping area has the weakest impact on the performance of the proposed feature-based RIR technique indicated by the smallest Spearman's rho value ($r_s$ = -0.286, $p$ = 0.001$^{**}$). This demonstrates the ability of the proposed feature extraction method to extract feature points on retinal vessels and distributed throughout the image. Extracting feature points throughout the image is crucial to ensure sufficient inliers can be established between the overlapping area. However, the registration ability of the proposed feature-based RIR technique is limited when the overlapping area is smaller than 41%. The existing feature-based RIR techniques are more sensitive in the presence of the small overlapping area represented with a larger Spearman's rho values that ranged between $r_s$ = 0.443 and $r_s$ = 0.795 ($p$ = <0.001$^{**}$).

The proposed feature-based RIR technique outperformed the others in registering image pair with various rotation ($r_s$ = -0.261, $p$ = 0.002$^{**}$). Particularly, the proposed feature-based RIR technique capable of registering image pairs with a rotation ranges between 0° to 48°. This evaluation presents the ability of the proposed feature descriptor method to assign distinctive information representing the corresponded feature point and to assign similar descriptor even in the presence of various rotation. The existing feature-based RIR techniques exhibit a more sensitive correlation with rotation perceived by a larger Spearman's rho values between $r_s$ = -0.380 and $r_s$ = -0.766 ($p$ = <0.001$^{**}$).

The proposed feature descriptor method is highly dependent on the grayscale level to describe the circular region surrounding the feature point. Consequently, the proposed feature-based RIR technique is sensitive to the changes of intensity between images ($r_s$ = -0.261, $p$ = 0.002$^{**}$). In contrast, intensity difference between images has minimal impact on the success rate of GDB-ICP as its correlation is not significant with the smallest Spearman's rho among all ($r_s$ = -0.261, $p$ = 0.002$^{**}$).

The non-uniform intensity distribution in fundus image can obscure the visibility of the vessels at the affected area. Registering image pair with a difference of intensity distribution present a similar challenge as registering image pair with a small overlapping area. To register this image pair, the geometrical transformation is estimated between the inliers from the unaffected area. This will require a sufficient number of inliers being established between the unaffected correspondence area. From the computed correlations, the difference of intensity distribution has the weakest impact on the proposed feature-based RIR technique compared to the others with insignificant correlation ($r_s$ = 0.011, $p$ = 0.896). Contrast correlations are observed between the success rates of the existing feature-based RIR techniques and the difference of intensity distribution. Their correlations are significantly correlated to the difference of intensity distribution with Spearman's rho values ranged between $r_s$ = 0.172 to $r_s$ = 0.210.

The similarity of structure component, which describes the clarity or sharpness of the vessel edges between fixed and moving images shows minimal impact on the registration performance of all feature-based RIR techniques. Mainly because the feature descriptor method in the feature-based RIR techniques do not rely on the vessel edges to compute the descriptor.

## 5.9 CONCLUSIONS

This chapter presents the proposed feature descriptor method to achieve objective RO3. The proposed feature-based RIR technique is evaluated in this chapter, which includes pre-processing (Section 3.4), proposed feature extraction method (Section 4.4), proposed feature descriptor method (Section 5.3), matching (Section 5.4) and estimating geometrical transformation (Section 5.5). The overall success rate of the proposed feature-based RIR technique on FIRE datasets is significantly outperformed GDB-ICP, Harris-PIIFD, Ghassabi's-SIFT, H-M 16, H-M 17 and D-Saddle-HOG.

# CHAPTER 6: CONCLUSIONS & FUTURE WORKS

## 6.1 ACHIEVEMENT OF OBJECTIVES

The primary aim of this research is to propose a feature-based retinal image registration (RIR) technique for fundus image. To achieve the primary aim, several objectives are outlined as presented in Section 1.4. The achievements of these objectives are summarised as follows.

The first objective (**RO1**) of this research is to investigate the general and unique characteristics of the retinal vessels in local patches of fundus image. The descriptions of these characteristics are presented in Section 4.3 Characteristics of Retinal Vessels and Noises in Local Patches. In this section, the characteristics of the retinal vessels and noises observed in the gradient and binary patches are examined for their similarity and differences. From our observation on these patches, the retinal vessels are generally appeared as a continuous curvature structure across the 3-dimensional gradient patch. The continuous curvature structure is consistent between the retinal vessels but differed in terms of width and depth depending on the size and contrast of the vessels. Therefore, the curvature structure without specifying its width and depth is employed in the feature detection module (Section 4.4.1.2 STEP 2: Detect local extrema and Section 4.4.1.3 STEP 3: Test extrema if within curvature structure) of the proposed feature extraction method as it can be a reliable characteristic in detecting feature points on the retinal vessels. However, the noises such as single and multiple underlying choroidal vessels, retinal nerve fibre layer, microaneurysm and exudates also exhibit continuous curvature structure in 3-dimensional gradient patch. The retinal vessels and noises can be further distinguished by examining their unique characteristics in the intensity profiles extracted from the 2-dimensional gradient and binary patches. The binary patch with a retinal vessel exhibits a continuous straight line structure that is represented by a long horizontal intensity profile when it is extracted along the straight line structure. The intensity profile extracted from the cross-sectional line intersecting the similar continuous straight line structure in the gradient patch resembles an inverse Gaussian-like shape. The inverse Gaussian-like shape with weaker intensity can also be observed in the intensity profile of the noises from the gradient patch. The intensity profiles of the retinal vessels and noises can be distinguished by examining the position of its valley with the maximum depth on y-axis and x-axis. These unique characteristics observed in the intensity profiles extracted from the 2-dimensional gradient and binary patches are incorporated as part of the feature selection module (Section 4.4.2.2 STEP 5: Exclusion process) of the proposed feature extraction method.

The second objective (**RO2**) of this research is to propose a feature extraction method based on the characteristics of the retinal vessels. The development of the proposed feature extraction method is described in Section 4.4: Proposed Feature Extraction. The proposed feature extraction method is composed of feature detection (Section 4.4.1 Feature Detection) and feature selection (Section 4.4.2 Feature Selection) modules. The feature detection module finds extrema within the curvature structure throughout the hierarchical Gaussian scale space as candidate feature points. The curvature structure is generally observed in 3-dimensional gradient patch with retinal vessels. The utilisation of the hierarchical Gaussian scale space allows the candidate feature points being extracted on various sizes of the retinal vessels. Then, the exclusion and selection processes in the selection module remove the candidate feature points on noises and select the final feature points. The exclusion process removes the candidate feature points on noises according to five criteria. These criteria are the unique characteristics of the retinal vessels and noises observed on the sum of intensity profiles extracted from the gradient and binary patches. The remaining candidate feature points are selected as the final feature points in the selection process according to the strength of the retinal vessel attributes. Furthermore, the selection process ensures that the selected feature points are distributed throughout the image. The performance of the proposed feature extraction method is evaluated on four public datasets. These datasets consist of fundus images with ground truth of the manually segmented vessels performed by experts. The extraction accuracy of the proposed feature extraction method is compared with five feature extraction methods previously used in the existing feature-based RIR techniques, namely, Harris corner, SIFT, SURF, Ghassabi's and D-Saddle. From the experiments, the proposed feature extraction method obtained an overall extraction accuracy of 86.021% on four datasets. This extraction accuracy is significantly outperformed the existing feature extraction methods ($p$ = <0.001$^{*}$). Furthermore, the performance of the proposed feature extraction method is unaffected by the changes of the image size ($r_s$ = -0.032, $p$ = 0.712) but significantly affected by the presence of non-uniform intensity distribution in the image ($r_s$ = 0.342, $p$ = <0.001$^{**}$).

The third objective (**RO3**) of this research is to propose a feature descriptor method that characterises the feature points based on distinctive information. The proposed feature descriptor method describes the extracted feature points with concatenated statistical properties of summation, mean and standard deviation as described in Section 5.3 Proposed Feature Descriptor. These statistical properties are calculated from the pixels on circumferences surrounding the feature points. Then, the performance of the proposed feature-based RIR technique in registering fundus images is evaluated. The proposed feature-based RIR technique comprises of five main stages, namely, pre-processing, proposed feature extraction method (**RO2**), proposed feature descriptor method (**RO3**), matching and estimating geometrical transformation. The proposed feature-based RIR technique is evaluated on FIRE dataset and compared against six existing feature-based RIR techniques; GDB-ICP, Harris-PIIFD, Ghassabi's-SIFT, H-M 16, H-M 17 and D-Saddle-HOG. The first part of the evaluation investigates the performance of the proposed feature-based RIR techniques in super-resolution, image mosaicking and longitudinal study applications. The overall success rate of the proposed feature-based RIR technique is the highest (67.164%) among the feature-based RIR techniques and significantly outperformed the others. The proposed feature-based RIR technique yields the highest success rate in the image mosaicking application (75.510%) followed by super-resolution application (66.197%). The proposed feature-based RIR technique obtained the lowest success rate in longitudinal study application (42.857%) due to changes in the anatomical appearance between images such as the substantial difference in vessel tortuosity and thickness. The second part of the evaluation examines the influences of the overlapping area, rotation and image quality on the success rate. Among these factors, only the overlapping area and rotation will be described here to reflect our objectives RO2 and RO3. The registration performance of the proposed feature-based RIR technique is significantly affected by the presence of the overlapping area ($r_s$ = -0.286, $p$ = 0.001$^{**}$) and rotation ($r_s$ = -0.261, $p$ = 0.002$^{**}$) between images. However, the impact of these factors on the proposed feature-based RIR technique are the weakest compared to the existing feature-based RIR techniques.

## 6.2 CONTRIBUTIONS

There two main contributions of this research to the body of knowledge. The first main contribution of this research is the development of a novel feature extraction method for feature-based RIR technique. The existing feature extraction methods are mainly without a proper feature selection module to accurately extract feature points on retinal vessels. In opposite, the proposed feature extraction method is composed of feature detection and feature selection modules that consider the characteristics of the retinal vessels and noises. This allows the proposed feature extraction method to remove the feature points on noises with similar structure representation as retinal vessels while extracts feature points on the vessels with varying sizes and contrast.

The second main contribution of this research is the development of a novel feature descriptor method for feature-based RIR technique. The existing feature descriptor methods are mainly utilised gradient direction surrounding the square region of the feature points as the descriptor. Contrarily, the proposed feature descriptor method characterised a feature point with concatenated statistical properties surrounding the point. This descriptor is distinctive in characterising the corresponded feature points on retinal vessels with the presence of rotation compared to the traditional feature descriptor method that mainly utilised the information of the gradient direction.

The proposed feature extraction method and the proposed feature descriptor method are evaluated on a total of five public datasets. Four of the datasets evaluate the performance of the proposed feature extraction method and the other one evaluates the performance of the proposed feature-based RIR technique. The proposed feature-based RIR technique was evaluated in registering fundus images from super-resolution, image mosaicking and longitudinal study applications, which mainly performed in clinical settings.

## 6.3 FUTURE WORK

There are several aspects of the proposed feature-based RIR technique that can be improved in the future. First, the proposed feature extraction method is highly dependent on the intensity changes in the local patch to locate the curvature structure. Therefore, improving the approach in locating the curvature structure with less dependency on the intensity changes can increase the performance of the proposed feature extraction method in fundus image with non-uniform intensity distribution. Second, the proposed feature descriptor method is highly susceptible in the presence of the anatomical changes between image. Addressing this issue can improve the usability of the proposed feature-based RIR technique in the longitudinal study application. Third, employing a hierarchy transformation model can further improve the registration performance of the proposed RIR technique. The hierarchy transformation model project the image by iteratively performed the registration from the low-order transformation to the higher-order transformation. This transformation model reduces the error of over transformation that mainly observed when a local transformation is applied to the image pair with minimal distortion such as in super-resolution application. Finally, the proposed feature-based RIR technique can be expanded for other modality of the retinal images.

## REFERENCES


Abràmoff, M. D., Garvin, M. K., & Sonka, M. (2010). Retinal imaging and image analysis. *IEEE reviews in biomedical engineering, 3*, 169-208.

Adal, K. M., van Etten, P. G., Martinez, J. P., van Vliet, L. J., & Vermeer, K. A. (2015). Accuracy Assessment of Intra- and Intervisit Fundus Image Registration for Diabetic Retinopathy ScreeningAccuracy Assessment of Fundus Image Registration. *Investigative Ophthalmology & Visual Science, 56*(3), 1805-1812.

Addison Lee, J., Cheng, J., Hai Lee, B., Ping Ong, E., Xu, G., Wing Kee Wong, D., Liu, J., Laude, A., & Han Lim, T. (2015). *A low-dimensional step pattern analysis algorithm with application to multimodal retinal image registration.* Paper presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

Aldana-Iuit, J., Mishkin, D., Chum, O., & Matas, J. (2016, December). *In the Saddle: Chasing Fast and Repeatable Features.* Paper presented at the 23rd International Conference on Pattern Recognition, Cancun, Mexico.

Araújo, T., Mendonça, A. M., & Campilho, A. (2018). Parametric model fitting-based approach for retinal blood vessel caliber estimation in eye fundus images. *PLOS ONE, 13*(4), e0194702.

Bankhead, P., Scholfield, C. N., McGeown, J. G., & Curtis, T. M. (2012). Fast Retinal Vessel Detection and Measurement Using Wavelets and Edge Location Refinement. *PLOS ONE, 7*(3), e32435.

Bay, H., Ess, A., Tuytelaars, T., & Van Gool, L. (2008). Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding, 110*(3), 346-359.

Bay, H., Tuytelaars, T., & Van Gool, L. (2006, May 7-13). *SURF: Speeded Up Robust Features.* Paper presented at the Computer Vision – ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria.

Beis, J. S., & Lowe, D. G. (1997). *Shape Indexing Using Approximate Nearest-Neighbour Search in High-Dimensional Spaces*. Paper presented at the Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition (CVPR '97).

Bennett, T. J. (2019). Scanning Laser Ophthalmoscopy. *Ophthalmic Photographers' Society.* Retrieved from https://www.opsweb.org/page/SLO

Bentley, J. L. (1975). Multidimensional binary search trees used for associative searching. *Communications of the ACM, 18*(9), 509-517.

Bernardes, R., Serranho, P., & Lobo, C. (2011). Digital Ocular Fundus Imaging: A Review. *Ophthalmologica, 226*(4), 161-181.

Bontala, A., Sivaswamy, J., & Pappuru, R. R. (2012). *Image mosaicing of low quality neonatal retinal images*. Paper presented at the Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on.

Brown, D. M., & Ciardella, A. (2005). Mosaic Fundus Imaging in the Diagnosis of Retinal Diseases. *Investigative Ophthalmology & Visual Science, 46*(13), 2581-2581.

Brown, M., & Lowe, D. G. (2002). *Invariant Features from Interest Point Groups*. Paper presented at the BMVC.

Budai, A., Bock, R., Maier, A., Hornegger, J., & Michelson, G. (2013). Robust vessel segmentation in fundus images. *International Journal of Biomedical Imaging, 2013*.

Burger, W., & Burge, M. J. (2013). SIFT—Scale-Invariant Local Features. In *Principles of Digital Image Processing: Advanced Methods* (Vol. 3, pp. 229 - 295). London: Springer.

Calonder, M., Lepetit, V., Ozuysal, M., Trzcinski, T., Strecha, C., & Fua, P. (2012). BRIEF: Computing a Local Binary Descriptor Very Fast. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 34*(7), 1281-1298.

Can, A., Stewart, C. V., Roysam, B., & Tanenbaum, H. L. (2002). A feature-based, robust, hierarchical algorithm for registering pairs of images of the curved human retina. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 24*(3), 347-364.

*CHASE_DB1 Retinal Image Database*. Retrieved from: https://blogs.kingston.ac.uk/retinal/chasedb1/

*CHASE_DB1 Retinal Image Database*. Retrieved from: https://blogs.kingston.ac.uk/retinal/chasedb1/

Chen, J., Tian, J., Lee, N., Zheng, J., Smith, R. T., & Laine, A. F. (2010). A Partial Intensity Invariant Feature Descriptor for Multimodal Retinal Image Registration. *IEEE Transactions on Biomedical Engineering, 57*(7), 1707-1718.

Chen, L., Xiang, Y., Chen, Y. J., & Zhang, X. L. (2011, 11-14 Sept). *Retinal Image Registration Using Bifurcation Structures*. Paper presented at the 18th IEEE International Conference on Image Processing (ICIP).

Chui, H., & Rangarajan, A. (2003). A new point matching algorithm for non-rigid registration. *Computer Vision and Image Understanding, 89*(2), 114-141.

Dalal, N., & Triggs, B. (2005). *Histograms of oriented gradients for human detection*. Paper presented at the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR).

de Amorim Garcia Filho, C. A., Yehoshua, Z., Gregori, G., Puliafito, C. A., & Rosenfeld, P. J. (2013). Chapter 3 - Optical Coherence Tomography. In S. J. Ryan, S. R. Sadda, D. R. Hinton, A. P. Schachat, S. R. Sadda, C. P. Wilkinson, P. Wiedemann, & A. P. Schachat (Eds.), *Retina (Fifth Edition)* (pp. 82-110). London: W.B. Saunders.

DeHoog, E., & Schwiegerling, J. (2009). Fundus camera systems: a comparative analysis. *Applied Optics, 48*(2), 221-228.

Deng, K., Tian, J., Zheng, J., Zhang, X., Dai, X., & Xu, M. (2010). Retinal Fundus Image Registration via Vascular Structure Graph Matching. *International Journal of Biomedical Imaging, 2010*, 13.

*DRIVE: Digital Retinal Images for Vessel Extraction*. Retrieved from: http://www.isi.uu.nl/Research/Databases/DRIVE/

*DRIVE: Digital Retinal Images for Vessel Extraction*. Retrieved from: http://www.isi.uu.nl/Research/Databases/DRIVE/

Fang, B., Yu, K., Ma, J., & An, P. (2019). EMCM: A Novel Binary Edge-Feature-Based Maximum Clique Framework for Multispectral Image Matching. *Remote Sensing, 11*(24), 3026.

*FIRE: Fundus Image Registration Dataset*. Retrieved from: https://www.ics.forth.gr/cvrl/fire/

Fischler, M. A., & Bolles, R. C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM, 24*(6), 381-395.

Fleckenstein, M., Schmitz-Valckenberg, S., & Holz, F. G. (2013). Chapter 4 - Autofluorescence Imaging. In S. J. Ryan, S. R. Sadda, D. R. Hinton, A. P. Schachat, S. R. Sadda, C. P. Wilkinson, P. Wiedemann, & A. P. Schachat (Eds.), *Retina (Fifth Edition)* (pp. 111-132). London: W.B. Saunders.

Frangi, A. F., Niessen, W. J., Vincken, K. L., & Viergever, M. A. (1998). *Multiscale vessel enhancement filtering.* Paper presented at the International Conference on Medical Image Computing and Computer-Assisted Intervention.

Fraz, M. M., Remagnino, P., Hoppe, A., Uyyanonvara, B., Rudnicka, A. R., Owen, C. G., & Barman, S. A. (2012). An ensemble classification-based approach applied to retinal blood vessel segmentation. *IEEE Transactions on Biomedical Engineering, 59*(9), 2538-2548.

Ghassabi, Z., Shanbehzadeh, J., Mohammadzadeh, A., & Ostadzadeh, S. S. (2015). Colour retinal fundus image registration by selecting stable extremum points in the scale-invariant feature transform detector. *IET Image Processing, 9*(10), 889-900.

Ghassabi, Z., Shanbehzadeh, J., Sedaghat, A., & Fatemizadeh, E. (2013). An efficient approach for robust multimodal retinal image registration based on UR-SIFT features and PIIFD descriptors. *Eurasip Journal on Image and Video Processing*.

Goerner, F. L., Duong, T., Stafford, R. J., & Clarke, G. D. (2013). A comparison of five standard methods for evaluating image intensity uniformity in partially parallel imaging MRI. *Medical Physics, 40*(8), 082302.

Gonzalez, R. C., Woods, R. E., & Eddins, S. L. (2009). Representation and Description. In *Digital Image Processing Using MATLAB* (2 ed.). New Jersey: Prentice Hall.

Goshtasby, A. (1988). Image registration by local approximation methods. *Image and Vision Computing, 6*(4), 255-261.

Harris, C., & Stephens, M. (1988). *A combined corner and edge detector.* Paper presented at the Alvey Vision Conference.

Hernandez-Matas, C., Zabulis, X., & Argyros, A. A. (2015). *Retinal image registration based on keypoint correspondences, spherical eye modeling and camera pose estimation.* Paper presented at the 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC).

Hernandez-Matas, C., Zabulis, X., & Argyros, A. A. (2016). *Retinal image registration through simultaneous camera pose and eye shape estimation.* Paper presented at the IEEE 38th Annual International Conference of the Engineering in Medicine and Biology Society (EMBC), Orlando, Florida.

Hernandez-Matas, C., Zabulis, X., & Argyros, A. A. (2017a). An Experimental Evaluation of the Accuracy of Keypoints-based Retinal Image Registration. *Conf Proc IEEE Eng Med Biol Soc, 2017*, 377-381.

Hernandez-Matas, C., Zabulis, X., Triantafyllou, A., Anyfanti, P., & Argyros, A. A. (2017b). Retinal image registration under the assumption of a spherical eye. *Computerized Medical Imaging and Graphics, 55*, 95-105.

Hernandez-Matas, C., Zabulis, X., Triantafyllou, A., Anyfanti, P., Douma, S., & Argyros, A. A. (2017c). FIRE: Fundus Image Registration dataset. *Journal for Modeling in Ophthalmology, 1*(4), 16-28.

Hinterstoisser, S., Cagniart, C., Ilic, S., Sturm, P., Navab, N., Fua, P., & Lepetit, V. (2012). Gradient Response Maps for Real-Time Detection of Textureless Objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 34*(5), 876-888.

Hoover, A., Kouznetsova, V., & Goldbaum, M. (2000). Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. *IEEE Transactions on Medical Imaging, 19*(3), 203-210.

*HRF: High-Resolution Fundus Image Database*. Retrieved from: https://www5.cs.fau.de/research/data/fundus-images/

*HRF: High-Resolution Fundus Image Database*. Retrieved from: https://www5.cs.fau.de/research/data/fundus-images/

International Telecommunication Union. (2011). *Studio encoding parameters of digital television for standard 4: 3 and wide-screen 16: 9 aspect ratiosRecommendation ITU-R BT.601-7*.

Johnson, R. N., Fu, A. D., McDonald, H. R., Jumper, J. M., Ai, E., Cunningham, E. T., & Lujan, B. J. (2013). Chapter 1 - Fluorescein Angiography: Basic Principles and Interpretation. In S. J. Ryan, S. R. Sadda, D. R. Hinton, A. P. Schachat, S. R.

Sadda, C. P. Wilkinson, P. Wiedemann, & A. P. Schachat (Eds.), *Retina (Fifth Edition)* (pp. 2-50.e51). London: W.B. Saunders.

Katherine Hu, B. (Producer). (2017). Fundus Photography and Fluorescein Angiography of Susac's Syndrome. [Photograph] Retrieved from http://morancore.utah.edu/section-12-retina-and-vitreous/fundus-photography-and-fluorescein-angiography-of-susacs-syndrome/

Khouri, A. S., Szirth, B. C., Salti, H. I., & Fechtner, R. D. (2007). DICOM transmission of simultaneous stereoscopic images of the optic nerve in patients with glaucoma. *Journal of Telemedicine and Telecare, 13*(7), 337-340.

Kolar, R., Sikula, V., & Base, M. (2010). Retinal Image Registration using Phase Correlation. *Analysis of Biomedical Signals and Images*, 244-252.

Kushnir, M., & Shimshoni, I. (2014). Epipolar Geometry Estimation for Urban Scenes with Repetitive Structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 36*(12), 2381-2395.

Lee, B. H., Xu, G., Gopalakrishnan, K., Ong, E. P., Li, R., Wong, D. W. K., & Lim, T. H. (2015). *AEGIS-Augmented Eye Laser Treatment with Region Guidance for Intelligent Surgery*. Paper presented at the 11th Asian Conference on Computer Aided Surgery (ACCAS 2015).

Lee, J. A., Lee, B. H., Xu, G., Ong, E. P., Wong, D. W. K., Liu, J., & Lim, T. H. (2014, 26-30 Aug. 2014). *Geometric corner extraction in retinal fundus images.* Paper presented at the 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society.

Legg, P. A., Rosin, P. L., Marshall, D., & Morgan, J. E. (2015). Feature Neighbourhood Mutual Information for multi-modal image registration: An application to eye fundus imaging. *Pattern Recognition, 48*(6), 1937-1946.

Leutenegger, S., Chli, M., & Siegwart, R. Y. (2011). *BRISK: Binary robust invariant scalable keypoints.* Paper presented at the Computer Vision (ICCV), 2011 IEEE International Conference on.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision, 60*(2), 91-110.

Ma, J., Jiang, J., Liu, C., & Li, Y. (2017). Feature guided Gaussian mixture model with semi-supervised EM and local geometric constraint for retinal image registration. *Information Sciences, 417*, 128-142.

MacGillivray, T. J., Trucco, E., Cameron, J. R., Dhillon, B., Houston, J. G., & van Beek, E. J. R. (2014). Retinal imaging as a source of biomarkers for diagnosis, characterization and prognosis of chronic illness or long-term conditions. *The British journal of radiology, 87*(1040), 20130832-20130832.

Matsopoulos, G. K., Asvestas, P. A., Mouravliansky, N. A., & Delibasis, K. K. (2004). Multimodal registration of retinal images using self organizing maps. *IEEE Transactions on Medical Imaging, 23*(12), 1557-1563.

Muja, M., & Lowe, D. G. (2009, February 5-8). *Fast approximate nearest neighbors with automatic algorithm configuration.* Paper presented at the 4th International Conference on Computer Vision Theory and Applications (VISAPP 2009), Lisboa, Portugal.

Nguyen, U. T. V., Bhuiyan, A., Park, L. A. F., & Ramamohanarao, K. (2013). An effective retinal blood vessel segmentation method using multi-scale line detection. *Pattern Recognition, 46*(3), 703-715.

Noyel, G., Thomas, R., Bhakta, G., Crowder, A., Owens, D., & Boyle, P. (2017). Superimposition of eye fundus images for longitudinal analysis from large public health databases. *Biomedical Physics & Engineering Express, 3*(4), 045015.

Optomap (Producer). (2015). Optomap image of a healthy adult retina. [Photograph] Retrieved from https://en.wikipedia.org/wiki/Scanning_laser_ophthalmoscopy

Pauli, T. W., Gangaputra, S., Hubbard, L. D., Thayer, D. W., Chandler, C. S., Peng, Q., Narkar, A., Ferrier, N. J., & Danis, R. P. (2012). Effect of Image Compression and Resolution on Retinal Vascular CaliberEffect of Image Compression and Resolution. *Investigative Ophthalmology & Visual Science, 53*(9), 5117-5123.

Pirbhai, A., Sheidow, T., & Hooper, P. (2005). Prospective evaluation of digital non-stereo color fundus photography as a screening tool in age-related macular degeneration. *American journal of ophthalmology, 139*(3), 455-461.

Podoleanu, A. G. (2012). Optical Coherence Tomography. *Journal of Microscopy, 247*(3), 209-219.

Ramli, R., Idris, M. Y. I., Hasikin, K., & A. Karim, N. K. (2017a, 31 July - 1 August). *Histogram-based Threshold Selection of Retinal Feature for Image Registration.* Paper presented at the 3rd International Conference on Information Technology & Society (IC-ITS), Penang, Malaysia.

Ramli, R., Idris, M. Y. I., Hasikin, K., A. Karim, N. K., Abdul Wahab, A. W., Ahmedy, I., Ahmedy, F., Kadri, N. A., & Arof, H. (2017b). Feature-Based Retinal Image Registration Using D-Saddle Feature. *Journal of Healthcare Engineering.*

Ramos, L., Novo, J., Rouco, J., Romeo, S., Álvarez, M. D., & Ortega, M. (2018). Retinal vascular tortuosity assessment: inter-intra expert analysis and correlation with computational measurements. *BMC Medical Research Methodology, 18*(1), 144-144.

Rousso, L., & Sowka, J. (2017, 15 Jan). Recognizing abnormal vasculature: a guide to following and educating patients who face this class of sight-threatening diagnoses. *Review of Optometry, 154(1).* Retrieved from https://www.reviewofoptometry.com/article/recognizing-abnormal-vasculature

Ryan, S. J., Sadda, S. V. R., Hinton, D. R., Schachat, A. P., Wilkinson, C. P., & Wiedemann, P. (2012). *Retina* (Fifth ed. Vol. 1 Retinal Imaging and Diagnostics): Elsevier-Saunders.

Saha, S., & Démoulin, V. (2012, 30 Sept.-3 Oct. 2012). *ALOHA: An efficient binary descriptor based on Haar features.* Paper presented at the 2012 19th IEEE International Conference on Image Processing.

Saha, S. K., Xiao, D., Frost, S., & Kanagasingam, Y. (2016). A Two-Step Approach for Longitudinal Registration of Retinal Images. *Journal of Medical Systems, 40*(12).

Saha, S. K., Xiao, D., Frost, S., & Kanagasingam, Y. (2018). Performance Evaluation of State-of-the-Art Local Feature Detectors and Descriptors in the Context of Longitudinal Registration of Retinal Images. *Journal of Medical Systems, 42*(4), 57.

Sedaghat, A., Mokhtarzade, M., & Ebadi, H. (2011). Uniform robust scale-invariant feature matching for optical remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing, 49*(11), 4516-4527.

Spaide, R. F., Fujimoto, J. G., Waheed, N. K., Sadda, S. R., & Staurenghi, G. (2018). Optical Coherence Tomography Angiography. *Progress in Retinal and Eye Research, 64*, 1-55.

Staal, J., Abràmoff, M. D., Niemeijer, M., Viergever, M. A., & Van Ginneken, B. (2004). Ridge-based vessel segmentation in color images of the retina. *IEEE Transactions on Medical Imaging, 23*(4), 501-509.

*STARE: Structured Analysis of the Retina*. Retrieved from: http://cecas.clemson.edu/~ahoover/stare/

*STARE: Structured Analysis of the Retina*. Retrieved from: http://cecas.clemson.edu/~ahoover/stare/

Tang, H., Pan, A., Yang, Y., Yang, K., Luo, Y., Zhang, S., & Ong, S. H. (2018). Retinal Image Registration Based on Robust Non-Rigid Point Matching Method. *Journal of Medical Imaging and Health Informatics, 8*(2), 240-249.

Torr, P. H., & Zisserman, A. (2000). MLESAC: A new robust estimator with application to estimating image geometry. *Computer Vision and Image Understanding, 78*(1), 138-156.

Tsai, C., Li, C., Yang, G., & Lin, K. (2010). The Edge-Driven Dual-Bootstrap Iterative Closest Point Algorithm for Registration of Multimodal Fluorescein Angiogram Sequence. *IEEE Transactions on Medical Imaging, 29*(3), 636-649.

Vedaldi, A., & Fulkerson, B. (2010, 25-29 Oct). *VLFeat: An open and portable library of computer vision algorithms.* Paper presented at the Proceedings of the 18th ACM international conference on Multimedia, Firenze, Italy.

Vonikakis, V., Chrysostomou, D., Kouskouridas, R., & Gasteratos, A. (2013). A biologically inspired scale-space for illumination invariant feature detection. *Measurement Science and Technology, 24*(7), 074024.

Walsh, A., Wildey, R., Lara, C., Ouyang, Y., & Sadda, S. (2010). Detection of Fundus Abnormalities Using 3D-OCT versus Mydriatic Color Fundus Imaging. *Investigative Ophthalmology & Visual Science, 51*(13), 3863-3863.

Wang, G., Wang, Z. C., Chen, Y. F., & Zhao, W. D. (2015). Robust point matching method for multimodal retinal image registration. *Biomedical Signal Processing and Control, 19*, 68-76.

Wei, E., Jia, Y., Tan, O., Potsaid, B., Liu, J. J., Choi, W., Fujimoto, J. G., & Huang, D. (2013). Parafoveal Retinal Vascular Response to Pattern Visual Stimulation Assessed with OCT Angiography. *PLOS ONE, 8*(12), e81343.

Xiao, D., Vignarajan, J., Lock, J., Frost, S., Tay-Kearney, M.-L., & Kanagasingam, Y. (2012). Retinal image registration and comparison for clinical decision support. *The Australasian Medical Journal, 5*(9), 507-512.

Yang, G., Stewart, C. V., Sofka, M., & Tsai, C.-L. (2007). Registration of challenging image pairs: Initialization, estimation, and decision. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 29*(11).

Yogesan, K., Constable, I. J., Eikelboom, R. H., & van Saarloos, P. P. (1998). Tele-ophthalmic screening using digital imaging devices. *Australian and New Zealand Journal of Ophthalmology, 26*, S9-S11.

Zheng, Y., Daniel, E., Hunter, A. A., Xiao, R., Gao, J., Li, H., Maguire, M. G., Brainard, D. H., & Gee, J. C. (2014). Landmark matching based retinal image alignment by enforcing sparsity in correspondence matrix. *Medical Image Analysis, 18*(6), 903-913.