TU Delft

Delft University of Technology

Computer-Aided Assessment of Longitudinal Fundus Photos for Screening Diabetic Retinopathy

Adal, Kedir

DOI
10.4233/uuid:b1582672-db92-4b79-b331-4e9e307f42b2

Publication date
2019

Document Version
Final published version

Citation (APA)
Adal, K. (2019). *Computer-Aided Assessment of Longitudinal Fundus Photos for Screening Diabetic Retinopathy*. [Dissertation (TU Delft), Delft University of Technology].
https://doi.org/10.4233/uuid:b1582672-db92-4b79-b331-4e9e307f42b2

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.
For technical reasons the number of authors shown on this cover page is limited to a maximum of 10.

Computer-Aided Assessment of Longitudinal Fundus Photos for Screening Diabetic Retinopathy

Computer-Aided Assessment of Longitudinal Fundus Photos for Screening Diabetic Retinopathy

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus Prof.dr.ir. T.H.J.J. van der Hagen,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op dinsdag 12 maart 2019 om 10:00 uur

door

Kedir Mohammed Adal

Master of Science in Vision and Robotics,
Universite de Bourgogne (UB), France,
geboren te Dessie, Ethiopië.

Dit proefschrift is goedgekeurd door de

promotor: prof. dr. ir. L.J. van Vliet
copromotor: dr. ir. K.A. Vermeer

Samenstelling promotiecommissie:

Rector Magnificus,	voorzitter
Prof. dr. ir. L.J. van Vliet,	Technische Universiteit Delft, promotor
Dr. ir. K.A. Vermeer,	Rotterdam Opthmalmic Institute, copromotor
Onafhankelijke leden:	
Dr. F.M.Vos,	Technische Universiteit Delft / Amsterdam Universitair Medische Centra
Prof. dr. F. Meriaudeau,	Universite de Bourgogne
Prof. dr. B. Rieger,	Technische Universiteit Delft
Prof. dr. ir. B.M. ter Haar Romeny,	Technische Universiteit Eindhoven
Dr. M.E.J. van Velthoven,	Het Oogziekenhuis Rotterdam

Financial support: Achmea, Coolsingel Foundation, CZ Fund, Stichting Blindenhulp, SWOO-Flieringa, Foundation Eye Hospital.

Front & Back: Kedir M. Adal

Printed by: Boekendeal.nl

Copyright © 2019 by K.M. Adal.

ISBN 978-94-6384-017-0

An electronic version of this dissertation is available at
http://repository.tudelft.nl/.

To everyone who has supported me along this journey

Contents

1

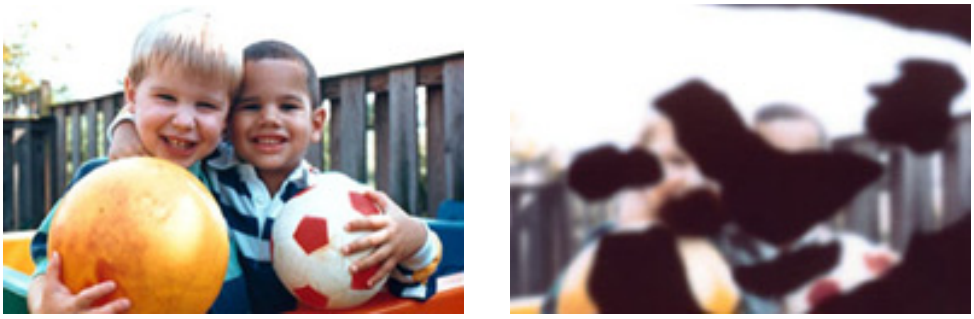
Introduction

1

Diabetic retinopathy (DR) is one of the most common complications of diabetes mellitus that affects retinal blood vessels and can result in vision loss and even blindness if not diagnosed and treated at early stage (see Figure 1.1). DR accounts for 4.8% of the 37 million cases of blindness worldwide [1]. Given the increase in the prevalence of diabetes, there is a greater risk for DR. In 2014, an estimated 422 million people worldwide are reported to have diabetes [2] and in The Netherlands alone, the number of adults in 20-79 age range with diabetes is estimated about 973500 and an additional 367500 undiagnosed cases [3]. The global increase in both diagnosed and undiagnosed diabetic population further exacerbated the risk of DR related eye complications.

Several risk factors, such as the age of onset of diabetes and the duration of diabetes, are associated with the development and progression of DR [2]. In patients who were below 30 years when they developed diabetes, the DR prevalence of 17% and 97.5% was found with a diabetes duration of less than 5 years and 15 or more years, respectively [4]. For patients who developed diabetes after 30 or more years for less than five years and 15 or more years, the DR prevalence varied between 28.8% to 77.8%, respectively [5].

Diabetic patients may not notice the development of DR at the early stages or until it causes vision problems. Thus, regular eye checkup for screening DR is necessary in order to get timely and appropriate treatment measures to prevent vision loss. DR screening can be done by a trained expert through inspection of the retina for pathognomonic abnormalities. However, the global rise in the diabetic population coupled with the required resource for diabetic eye care puts a burden on the accessibility and efficiency of DR screening programs. Computer Aided Diagnosis (CAD) of DR can be instrumental to address this burden by enabling automated analysis and interpretation of digital retinal images for DR related abnormalities. This thesis explores an automated system that analyzes a longitudinal series of retinal images for a regular DR screening.



(a) (b)

Figure 1.1: An example of normal vision (left) and a distorted vision (right) due to advanced DR (image courtesy of: NIH, National Eye Institute).

1.1. Eye anatomy and fundus imaging

The eye is the light-sensing organ of our visual system that enables us to see. The eye consists of several flexible structures that control the amount of light, the path it travels through and the focus location. Light entering the eye through cornea and pupil is focused by a lens onto the retina, the light sensitive tissue covering the inner back-surface of the eye. The retina captures and converts light into a neural signal and subsequently transmits it through optic nerve to the visual cortex for further processing. Figure 1.2 shows a cross section of the eye with some labeled structures.

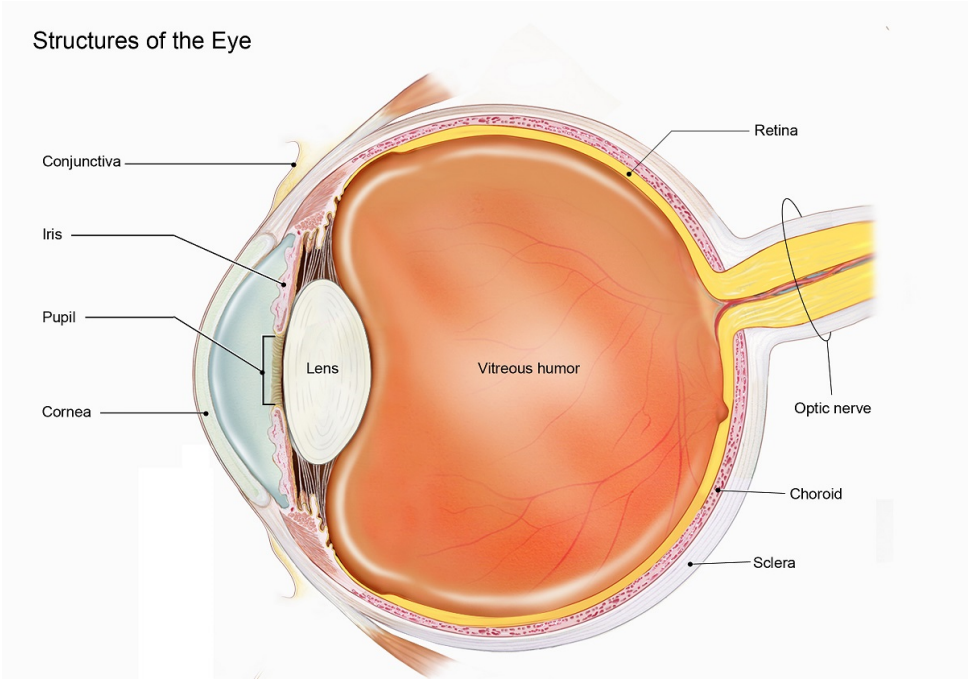


Figure 1.2: Cross sectional view of the human eye model (image courtesy of: NIH, National Eye Institute)

A fundus camera is routinely used in ophthalmology to image the retina. Modern fundus cameras were based on pioneering work by Helmholtz, but it was Gullstrand who perfected the ophthalmoscope in 1910. His study of optical images and of the refraction of light in the eye won him the Nobel Prize in Physiology or Medicine in 1911 [6]. Fundus imaging involves a specialized low power microscope with an attached camera and a noninvasive procedure to capture the interior surface of the eye including the retina, optic disc, macula, and retinal vasculature. In order to facilitate acquisition of a better view of the retinal surface, a mydriatic eye drop can be applied to dilate the pupil and allow more light to enter the eye. Figure 1.3 shows a digital fundus camera and a fundus image of a right eye.

The field of view of fundus cameras is constrained by the small size of the pupil;

1

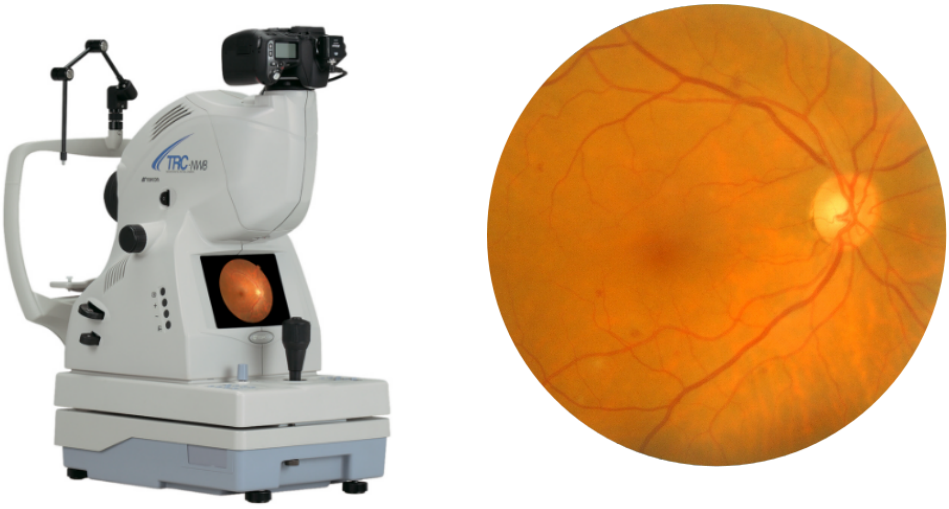


Figure 1.3: Left: A digital fundus camera. Right: fundus image of a right eye.

therefore, conventional approaches to capture a large field of view of the retina involves acquisition of multiple field fundus images covering different retinal regions (see Figure 1.4). Overlapping fundus fields can be registered onto a common coordinate system to create a single mosaic of the retina.

Figure 1.4: Left to right: fundus image set of a right eye consisting of macula, optic nerve, superior, and temporal retinal regions.

1.2. DR signs and screening

At early stages, DR causes microaneurysms, swelling in small blood vessels that may leak blood into the retina causing retinal bleeding (hemorrhage) [7]. Figure 1.5 shows fundus image patches of early DR lesions. As the disease progresses, more retinal blood vessels that nourish the retina start to leak lipids and proteins (exudates). This results in a reduction of nutrient supplies to the retina and stimulates neovascularization, i.e. the growth of new blood vessels. These abnormally grown blood vessels are fragile, thus they may easily leak and cause bleedings.

Since vision problems may not be noticed until more severe DR stages, diabetic patients should undergo periodic eye examination in order to monitor DR progression and receive early treatment measures.

(a) (b) (c)

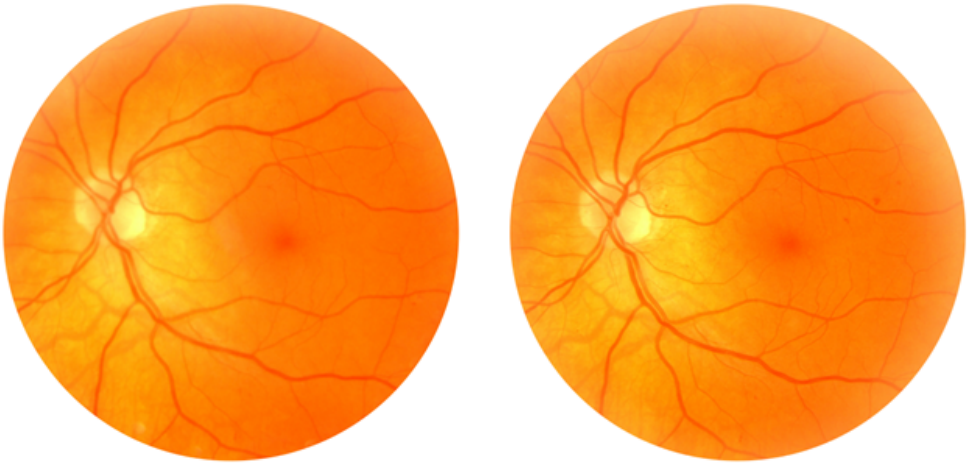
Figure 1.5: Examples of fundus image patches showing DR related retinal lesions.

The current eye care practice for screening DR involves examination of multiple field fundus images for pathognomonic abnormalities by a trained expert. Depending on the observed retinal abnormalities at the time of the examination, diabetic patients are either scheduled for a follow-up examination or referred for immediate intervention. This procedure is time consuming, subjective [8], and does not exploit fundus images from previous screening time points to monitor disease progression. Over the past decades, automated DR screening has attracted many researchers due to its potential to reduce the burden on the available health care. This is especially evident from the increase in the diabetic population worldwide. Unlike manual examination, automated analysis of fundus images enables objective and quantitative assessment of DR and also facilitates the resources needed to provide eye care service for diabetic patients.

1.3. Challenges for screening of longitudinal data

Most of previous research on automated DR screening has focused on analyzing fundus images from a single retinal examination to detect early stage lesions and subsequently identify patients with and without referable DR [9, 10]. These systems achieve a sensitivity comparable to human graders, but with a much lower specificity than human graders. Although such approaches enable to identify DR related retinal abnormalities at the time of examination, it does not give a complete insight into the disease activity since the previous check-up, and thus not applicable for progression assessment.

The development and progression of DR vary between individuals, so longitudinal analysis of fundus images is crucial to understand the disease activity over time. DR is a progressive disease that results in retinal changes due to appearance and disappearance of associated lesions such as microaneurysms. Recent studies suggest that in addition to the number of lesions at the time of examination, the dynamics of these lesions is found to be useful to monitor progression of DR [11, 12]. An automated system for making longitudinal analysis of a series of fundus photos



(a) (b)

Figure 1.6: Fundus image pairs captured a year apart showing a region that changed between the two time points due to appearance of DR lesions. Because of non-uniform illumination, the visibility of some of the lesions and retinal features that are close to the rim is limited in the right image.

for DR can thus be instrumental to assess the disease progression for a proactive and effective screening and intervention planning.

The main goal of this thesis is to develop an automated system for the detection and classification of longitudinal retinal changes to objectively and quantitatively monitor DR progression.

Automated detection of longitudinal retinal changes from a series of fundus images is challenging for several reasons. Firstly, illumination variation between fundus images captured during successive retinal examinations hampers identifying clinically relevant changes. Acquisition of high quality fundus image requires proper adjustment of the settings of the fundus camera parameters such as uniform illumination and focus. However, this is a delicate process and subjective, thus color fundus images often suffer from intra- and inter-visit variation in luminosity and contrast, especially around the rim (see Figure 1.6). This variation poses challenges in matching identical retinal features for aligning overlapping fundus images. In addition, illumination variation combined with the low visibility of DR lesions against the retinal background makes it hard to compare retinal features for the detection of longitudinal retinal change.

Secondly, detecting retinal changes due to small lesions, such as microaneurysms and dot hemorrhages, over time requires very high registration accuracy. In order to correctly register fundus images, the nonlinear spatial deformation caused by the projection of the curved surface of the retina onto a planar image plane needs to be accounted for. This is usually done by estimating image deformation parameters by matching retinal features between overlapping fundus image pairs. The abundance and distribution of retinal features within overlapping regions are cru-

cial to correctly estimate the deformation parameters. Because of the sparseness of retinal features and their uneven distribution in the image, accurate registration of fundus images is challenging.

Third, accurate detection and classification of longitudinal retinal changes requires a robust algorithm to identify clinically relevant changes from those caused by acquisition artefacts and noise. Developing such an algorithm is challenging due to the subtle appearance of early DR lesions, even for expert graders. This also causes variability in interrater agreement for establishing a benchmark dataset for algorithm training. This thesis presents methods to address all of the aforementioned challenges.

The proposed system for automated longitudinal retinal change detection consists of multiple stages: normalization of intra and inter-visit illumination variation, intra and inter-visit image registration, and clinically relevant change detection and classification (see Figure 1.7). Illumination variation is addressed by normalizing the green channel of each color fundus image for luminosity and contrast variation, thereby improving the visibility of retinal features. Then, spatio-temporal retinal changes are detected by a multi-scale image filtering technique. Finally, several local intensity and shape descriptors were extracted from each retinal change location and subsequently used by a support vector machine (SVM) to classify the region as a change due to an early DR related retinal lesion or other type of change. The detected retinal changes are then further explored to investigate the relation between the disease activity and time-to-referral. The developed system can also be used as a clinical tool to assist diabetic eye care experts to analyze and interpret fundus images for longitudinal DR screening, for instance by highlighting DR-related changes since the previous retinal exam.

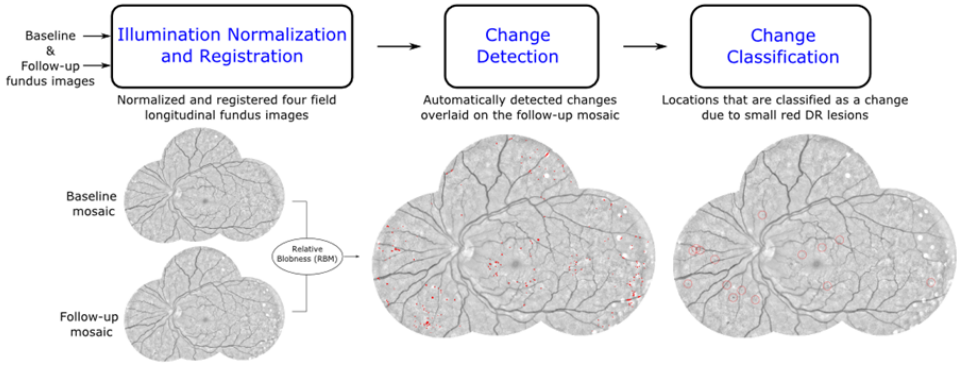


Figure 1.7: An overview of the proposed automated system for the detection and classification of red retinal lesions in longitudinal fundus images.

1.4. Thesis outline

The organization of this thesis is as follows:

Chapter 2 describes a method to normalize illumination and contrast variation

in a series of fundus images, thereby producing uniform luminosity over the entire field-of-view of the images and improving the visibility of smaller retinal structures. In addition, this chapter presents a robust hierarchical coarse-to-fine approach that makes use of the normalized intensity as well as structural information of the retinal vasculature for registering intra- and inter-visit fundus images.

Chapter 3 describes a qualitative approach for accuracy assessment of the proposed registration approach by DR screening experts and performance comparison with two top-ranked state-of-the-art commercial fundus mosaicking programs. The proposed approach facilitates visual inspection of the vasculature alignment in the overlap region between registered intra and inter-visit normalized fundus image series by highlighting possible misalignments. Two expert graders who are involved in a regular DR screening program graded the suitability of the registered fundus images for further use.

Chapter 4 presents an automated method for a quantitative assessment of the registration accuracy of fundus image pairs based on the vasculature in the registered images. The method automatically assesses the registration accuracy of fundus image pairs exploiting the intensity profiles across the vasculature and their difference in the registered images. A new accuracy measure, relative vessel misalignment energy (RVME), which exploits the even and odd signal property of the 1D profile across the vessels in the difference image, is introduced and used to quantify the registration accuracy.

Chapter 5 describes a multi-stage approach for the detection and classification of longitudinal retinal changes due to early DR related retinal lesions such as microaneurysms and dot hemorrhages, from registered fundus mosaics consisting of four-fields. A relative blobness measure (RBM), which is defined as the absolute scale-difference between the extremes of the multiscale blobness responses of fundus images from two time-points, is proposed to detect spatio-temporal retinal changes from longitudinal fundus mosaics. Several intensity and shape descriptors were extracted from each candidate region and subsequently used by a classifier to label the region as a red lesion or a non-red lesion related retinal change.

Chapter 6 presents a retrospective analysis of the red lesion turnover and classification of referable DR in diabetic eyes that have been regularly screened for DR. The red lesion turnover between successive retinal examinations, quantified by the multi-stage approach for the detection and classification of longitudinal retinal changes, was explored as a potential biomarker for predicting referable DR development.

Chapter 7 discusses the technical developments presented in this thesis, potential clinical applications and future research directions for the fully automated CAD system to be used in diabetic eye care.

References

[1] W. H. Organization, *Prevention of blindness from diabetes mellitus: report of a WHO consultation in Geneva, Switzerland, 9-11 November 2005* (World Health Organization, 2006).

[2] World Health Organization, *10 facts about diabetes,* http://www.who.int/features/factfiles/diabetes/en/ (2016), accessed September 15, 2016.

[3] International Diabetes Federation, *Idf diabetes atlas - 8th edition,* http://diabetesatlas.org/IDF_Diabetes_Atlas_8e_interactive_EN/ (2017), accessed June 28, 2018.

[4] R. Klein, B. E. Klein, S. E. Moss, M. D. Davis, and D. L. DeMets, *The wisconsin epidemiologic study of diabetic retinopathy: Ii. prevalence and risk of diabetic retinopathy when age at diagnosis is less than 30 years,* Archives of Ophthalmology **102**, 520 (1984).

[5] R. Klein, B. E. Klein, S. E. Moss, M. D. Davis, and D. L. DeMets, *The wisconsin epidemiologic study of diabetic retinopathy: Iii. prevalence and risk of diabetic retinopathy when age at diagnosis is 30 or more years,* Archives of Ophthalmology **102**, 527 (1984).

[6] A. Gullstrand, *Neue methoden der reflexlosen ophthalmoskopie,* Berichte Deutsche Ophthalmologische Gesellschaft **36**, 326 (1910).

[7] National Eye Institute, *Facts about diabetic eye disease,* https://nei.nih.gov/health/diabetic/retinopathy (2015), accessed September 15, 2016.

[8] M. D. Abràmoff, M. Niemeijer, M. S. Suttorp-Schulten, M. A. Viergever, S. R. Russell, and B. Van Ginneken, *Evaluation of a system for automatic detection of diabetic retinopathy from color fundus photographs in a large population of patients with diabetes,* Diabetes care **31**, 193 (2008).

[9] M. D. Abràmoff, J. C. Folk, D. P. Han, J. D. Walker, D. F. Williams, S. R. Russell, P. Massin, B. Cochener, P. Gain, L. Tang, *et al.*, *Automated analysis of retinal images for detection of referable diabetic retinopathy,* JAMA ophthalmology **131**, 351 (2013).

[10] A. D. Fleming, K. A. Goatman, S. Philip, G. J. Williams, G. J. Prescott, G. S. Scotland, P. McNamee, G. P. Leese, W. N. Wykes, P. F. Sharp, *et al.*, *The role of haemorrhage and exudate detection in automated grading of diabetic retinopathy,* British Journal of Ophthalmology **94**, 706 (2010).

[11] S. Nunes, I. Pires, A. Rosa, L. Duarte, R. Bernardes, and J. Cunha-Vaz, *Microaneurysm turnover is a biomarker for diabetic retinopathy progression to clinically significant macular edema: findings for type 2 diabetics with nonproliferative retinopathy,* Ophthalmologica **223**, 292 (2009).

[12] J. Cunha-Vaz, L. Ribeiro, and C. Lobo, *Phenotypes and biomarkers of diabetic retinopathy,* Progress in retinal and eye research **41**, 90 (2014).

2

A hierarchical coarse-to-fine approach for fundus image registration

This chapter is based on the publication:

K. M. Adal, R. M. Ensing, R. Couvert, P. van Etten, J. P. Martinez, K. A. Vermeer, and L. J. van Vliet, *A hierarchical coarse-to-fine approach for fundus image registration*, in Biomedical Image Registration (Springer, 2014) pp. 93–102.

Abstract

Accurate registration of retinal fundus images is vital in computer aided diagnosis of retinal diseases. This paper presents a robust registration method that makes use of the intensity as well as structural information of the retinal vasculature. In order to correct for illumination variation between images, a normalized-convolution based luminosity and contrast normalization technique is proposed. The normalized images are then aligned based on a vasculature-weighted mean squared difference (MSD) similarity metric. To increase robustness, we designed a multiresolution matching strategy coupled with a hierarchical registration model. The latter employs a deformation model with increasing complexity to estimate the parameters of a global second-order transformation model. The method was applied to combine 400 *fundus images from* 100 *eyes, obtained from an ongoing diabetic retinopathy screening program, into* 100 *mosaics. Accuracy assessment by experienced clinical experts showed that* 89 *(out of* 100*) mosaics were either free of any noticeable misalignment or have a misalignment smaller than the width of the misaligned vessel.*

2.1. Introduction

Registration of retinal fundus images plays a crucial role in computer-aided diagnosis and screening of the human eye for various retinal diseases. Depending on the targeted clinical application, fundus image registration can aid retinal examination in three ways. Firstly, mosaicking creates a larger field-of-view by stitching individual images. Such a mosaic facilitates comprehensive retinal examination at a single glance. Secondly, multimodal registration spatially aligns images from different modalities, thereby fusing complementary information into a single image. Thirdly, longitudinal registration aligns a series of fundus images taken over time. This is especially vital in screening or staging of progressive eye diseases such as age-related macular degeneration (AMD) and diabetic retinopathy [1, 2].

The success of these clinical applications depends on the accuracy of the registration algorithm. Although several fundus image registration algorithms have been proposed in the past decades [3–9], accurate and robust registration of retinal images still remains a challenge. This is mainly due to the sometimes very small image overlap, severe illumination artifacts near the frame boundaries, and the spatial distortion as a result of mapping the curved retinal surface onto the image plane.

Depending on the image information used for matching, existing algorithms can be grouped into intensity-based and feature-based methods. Intensity based methods make use of the similarity between the intensity or RGB values of raw or pre-processed images [3, 4]. Nicola et al. [3] used mutual information as a similarity criterion to estimate the parameters of a global (rigid) affine model. In the study by George et al. [4], the correlation between the binary vasculature masks of segmented fundus image pairs is optimized. These intensity based methods ignore the quadratic and higher order terms of the image distortion.

Feature-based methods [5–9] make use of saliency or landmark points, disregarding most of the structural information embedded in the local correlation of fundus images. In the paper by Ali et al. [5], retinal vessel bifurcations and crossover points are used as landmarks in a hierarchical optimization of a quadratic transformation model. Stewart et al. [6] used vessel bifurcations for initialization of a dual-bootstrap iterative closest point (ICP) algorithm to align the vessel centerlines using a quadratic transformation model. Chanwimaluang et al. [7] used the vasculature tree for initialization and the quadratic model parameters are estimated using the vessel bifurcation and crossover points. In study by Sangyeol et al. [8], a radial distortion correction, estimated using vessel bifurcations, is applied prior to registration in order to correct the distortion caused by the curved to planar surface mapping. Recently, Jian et al. [9] proposed salient feature regions (SFR) as landmark points of fundus images and local features extracted from these points are subsequently matched.

In general, the accuracy and robustness of feature-based methods are highly dependent on the feature detection method, the number of detected features, and their distribution in the image. The latter two conditions are restrictive in registration of fundus images, because vessel branching and crossover points are sparsely and unevenly distributed. Furthermore, this effect gets even worse if the region of

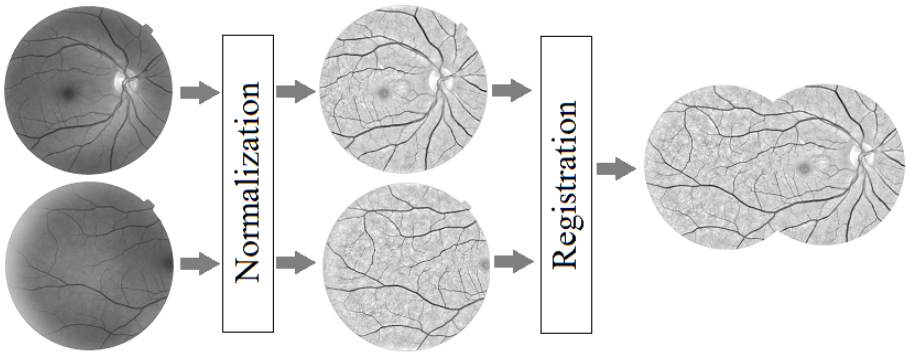


Figure 2.1: Overview of the proposed registration framework. First, the green channels of the fundus images are normalized for luminosity and contrast. Then, a hierarchical coarse-to-fine registration method is applied to produce a mosaic.

overlap between the image pairs becomes smaller.

In this paper, a registration method is proposed that exploits the intensity as well as the structural information of the retinal vasculature. We introduce a novel technique to normalize the green fundus image channel for illumination and contrast variation, thereby improving the visibility of the vasculature and hence the registration accuracy in these regions. The method then aligns retinal vessels based on the normalized images. We designed a multiresolution matching strategy coupled with a hierarchical registration model with a deformation model of increasing complexity for robust optimization of a global second-order transformation model.

2.2. Methods

The proposed method, outlined in figure 2.1, starts by normalizing the image luminosity and contrast, which vary greatly due to illumination conditions. Then the images are spatially aligned by first estimating the lower order transformation model parameters at a coarse resolution level and propagating the results to the next finer resolution level, where higher order model parameters are introduced. To guide the registration by vasculature regions, more weight was assigned to pixels in these regions.

2.2.1. Image Normalization

The main limitations of using the raw intensity values of fundus images for registration are the luminosity and contrast variations caused by non-uniform illumination of the retina during image acquisition. In this work, this intra and inter image variation is compensated for by applying an improved version of Foracchia's luminosity and contrast normalization method [10] to the green channel (I_G) of our RGB fundus images. The method relies on the intensity distribution of the retinal background (excluding vessels, optic disc, and lesions) to estimate local luminosity (L) and contrast (C). To compensate for local variations, the normalized image I_N,

becomes:

$$I_N = \frac{I_G - L}{C}, \tag{2.1}$$

where L and C are respectively the sample mean and standard deviation of the background image in the neighborhood of each pixel. However, since the background image is locally masked by retinal features such as blood vessels, a local signal approximation is required to handle this space-variant reliability map in neighborhood operations. In this paper, a higher-order normalized convolution is used to approximate the luminosity map. It takes into account missing or unreliable data samples and gives a better estimate of linear and quadratic variations in the illumination pattern [11, 12]. This is done by projecting each pixel and its neighbors on a set of basis vectors, chosen from the second-order Taylor expansion of the pixel around the neighbors, to create a new representation [12]. The contribution of each neighbor pixel is controlled by a Gaussian applicability function combined with a confidence measure, which encodes the presence or absence of background pixel values.

In figure 2.2, a typical example of a pair of fundus images from the same eye captured one year apart are shown before and after image normalization. The normalized image pairs (figure 2.2c and 2.2f) appear much more similar than the unprocessed image pairs (figure 2.2a and 2.2d). Moreover, the normalized convolution approach provides a far better contrast of the fine vasculature compared to the method described in [10] (figure 2.2b and 2.2e), especially around the border of the images. This is very crucial in registration of fundus images. As most of the overlap occurs around border regions, the registration accuracy depends on how well the vasculatures in these regions are aligned.

2.2.2. Registration Initialization

Convergence and robustness of image registration requires a good starting point. In this paper, we propose a robust initialization algorithm using overlap-corrected cross-correlation, i.e. standard cross-correlation divided by the number of overlapping pixels from which it is computed (see Eq. 2.2). This allows the cross-correlation to be invariant to the overlap between images. In order to further handle rotation between the image pairs (e.g. due to possible head, eye or camera motion between consecutive image acquisitions), this is done at three rotation angles, $\alpha = 0°, \pm 5°$, and at a very coarse scale, i.e. by blurring with a Gaussian filter of $\sigma = 32$ pixels and downsampling by a factor of $s = 16$.

$$I_{\widehat{CC}}(u, v, \alpha) = \frac{\sum_{x=1}^{M} \sum_{y=1}^{N} I_f(x, y) I_m(x', y')}{\sum_{x=1}^{M} \sum_{y=1}^{N} \Omega_f(x, y) \Omega_m(x', y')}, \tag{2.2}$$

where $I_{\widehat{CC}}$ is the overlap-corrected cross-correlation and I_f and Ω_f (I_m and Ω_m) are the normalized image and field-of-view mask of the fixed (moving) image of size $M \times N$, respectively. $(x', y') = (x \cos\alpha - y \sin\alpha + u, x \sin\alpha + y \cos\alpha + v)$ are the

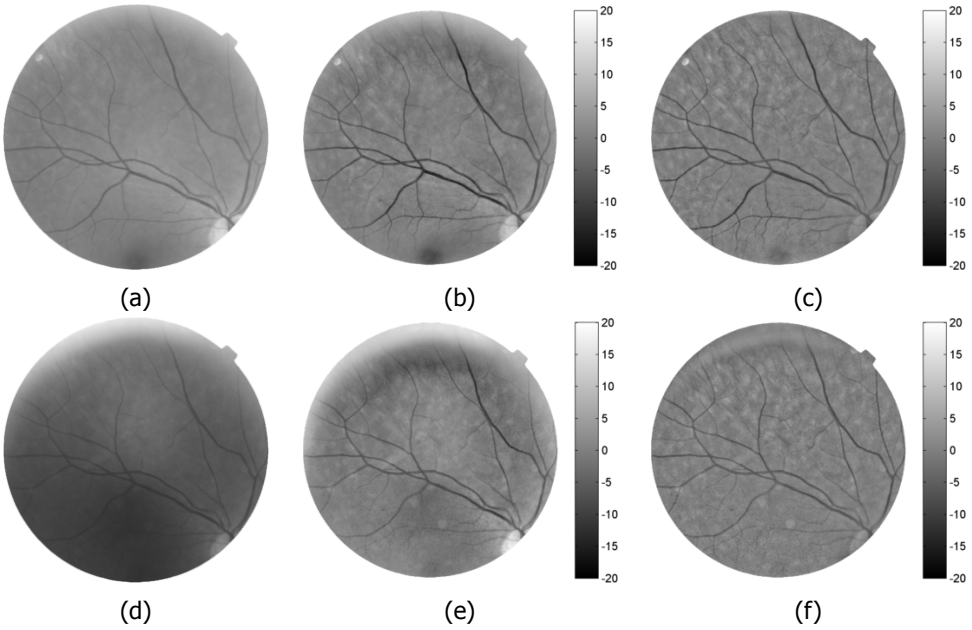


Figure 2.2: An example of illumination normalization on a pair of fundus images. (a) and (d) Green channel of fundus images. (b) and (e) Normalized fundus images using the method described in [10]. (c) and (f) Normalized fundus images using the proposed normalized convolution technique.

rotated and translated pixel coordinates. For each angle, the values of u and v that maximize $I_{\widehat{CC}}$ are tentatively selected. The optimal angle ($\hat{\alpha}$), and the corresponding values for u and v, are then selected by minimizing the mean squared difference (MSD) of $I_f(x, y)$ and $I_m(x', y')$. In our study, since the image pairs are represented at a very coarse scale, the three angles (five degrees apart) are enough to find the starting point for the registration.

2.2.3. Hierarchical Coarse-to-Fine Registration

Since the image pairs are normalized for luminosity and contrast, the MSD can be used as similarity metric. The registration is further guided by the vasculature regions as they provide the main distinctive structures of fundus images, thereby restricting the effect of intensity change in the background region due to factors such as disease progression and artifacts. This is achieved by weighting the contribution of each pixel to the similarity metric using a measure for vesselness $V(x, y) \in [0, 1]$. The vesselness-weighted cost function to minimize is:

$$\varepsilon = \frac{1}{|\Omega|} \sum_{(x,y)\in\Omega} V^2(x, y) \cdot \Big[I_f(x, y) - I_m\big(T(x, y; \Theta)\big) \Big]^2, \tag{2.3}$$

where $T(\cdot)$ is the transformation model parameterized by Θ, I_f and I_m are the normalized values of the fixed (anchor) and moving (floating) image, respectively,

Table 2.1: Transformation model and parameters at each pyramid level of the proposed hierarchical coarse-to-fine registration approach. σ and s are the Gaussian blurring scale and subsampling factor, respectively. The deformation model parameters at each level are optimized using Eqs 2.3 and 2.4. Note that $\hat{\alpha}$ is a fixed angle optimized at the initialization stage (section 2.2.2).

Level	Transformation	Parameters	σ (pixels)	s
1	Translation	$\begin{pmatrix} 0 & 0 & 0 & \cos\hat{\alpha} & -\sin\hat{\alpha} & \theta_1 \\ 0 & 0 & 0 & \sin\hat{\alpha} & \cos\hat{\alpha} & \theta_2 \end{pmatrix}$	16	8
2	Similarity	$\begin{pmatrix} 0 & 0 & 0 & \cos\alpha & -\sin\alpha & \theta_1 \\ 0 & 0 & 0 & \sin\alpha & \cos\alpha & \theta_2 \end{pmatrix}$	8	4
3	Affine	$\begin{pmatrix} 0 & 0 & 0 & \theta_1 & \theta_2 & \theta_3 \\ 0 & 0 & 0 & \theta_4 & \theta_5 & \theta_6 \end{pmatrix}$	4	2
4a	Simplified Quadratic	$\begin{pmatrix} \theta_1 & \theta_1 & 0 & \theta_2 & \theta_3 & \theta_4 \\ \theta_5 & \theta_5 & 0 & \theta_6 & \theta_7 & \theta_8 \end{pmatrix}$	2	2
4b	Quadratic	$\begin{pmatrix} \theta_1 & \theta_2 & \theta_3 & \theta_4 & \theta_5 & \theta_6 \\ \theta_7 & \theta_8 & \theta_9 & \theta_{10} & \theta_{11} & \theta_{12} \end{pmatrix}$	1	1

and Ω is the set of all overlapping pixels in the image pairs. The vesselness maps of both normalized images were computed from the multi-scale ($\sigma \in [1, 9]$ pixels), second-order local image structure [13]. The pixelwise maximum of the two maps was then dilated by a disk structuring element of 25 pixels radius and used as a weight.

As fundus imaging involves mapping the curved retinal surface onto a flat image plane, a transformation model of at least second-order is required to accurately align images. In this work, a global 12 parameter quadratic transformation model is used [5]:

$$T(x, y; \Theta) = \begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \theta_1 & \theta_2 & \theta_3 & \theta_4 & \theta_5 & \theta_6 \\ \theta_7 & \theta_8 & \theta_9 & \theta_{10} & \theta_{11} & \theta_{12} \end{pmatrix} \begin{pmatrix} x^2 & y^2 & xy & x & y & 1 \end{pmatrix}^T, \quad (2.4)$$

where (x', y') are the transformed pixel coordinates and θ_i is an element of the transformation matrix Θ.

In order to improve the robustness in estimating the parameters of the transformation model, a hierarchical multiresolution method is applied. The method employs a four level coarse-to-fine Gaussian pyramid, in which the complexity of the deformation model increases with every step downwards in the pyramid: first translation-only at the top level, second translation and rotation, third an affine transform followed by a simplified quadratic model (4a) and finally a full quadratic model (4b). The simplified quadratic model assumes an isotropic second-order deformation along both x and y dimensions. Each level of the Gaussian pyramid is formed by blurring and downsampling. Table 2.1 summarizes the transformation models, the blurring scale, and subsampling factors.

At each level of the pyramid, the model parameters which minimize the cost function ε, are optimized using Levenberg-Marquardt. In order to take into account

the difference of the magnitude of each parameter's search space, a scaling technique is employed. In addition, the parameters are orthogonalized with respect to each other so as to mitigate intra-parameter correlation. Since the optimization of each level is initialized by the results of the previous level, the risk of getting stuck into a local minimum is greatly reduced. Moreover, the hierarchical coarse-to-fine approach speeds up the convergence of the Levenberg-Marquardt algorithm by providing an appropriate initial estimate of parameters at successive pyramid levels.

2.3. Experiments and Results

2.3.1. Data Description

Data for this study was obtained from an ongoing diabetic retinopathy screening program at the Rotterdam Eye Hospital. 70 diabetes patients who visited the hospital in two consecutive years for diabetic retinopathy screening were included. During each visit, four images of macula-centered, optic nerve-centered, superior, and temporal regions of the retina were acquired from each eye. 400 images from 100 eyes, selected randomly from the first or the second year, were combined into 100 mosaics. At least one eye of each patient was included in this study.

2.3.2. Data Processing

For each eye, the image having the largest overlap with the remaining three images was selected as the fixed image. Then, starting with the fixed image as intermediate result, each of the three images were registered sequentially to the intermediate result in order of decreasing overlap area with the fixed image. The overlap between image pairs was as low as 14%, with an average of 48%. In total, 300 registrations were accomplished to create the 100 mosaics.

After registration, instead of averaging the overlapping area, each mosaic was constructed by overlaying the four individual images on top of each other. This is particularly important to assess the registration accuracy of fine vasculatures as combining by averaging conceals any misalignment or yields spurious blurring in the overlap regions. By changing the order of overlay, each image appeared in the top layer once, resulting in four mosaics. These mosaics were put together to form a mosaic video which was then used for grading.

2.3.3. Fundus Mosaic Grading

Unlike the conventional approach where the centerline error between the aligned vessels is used to quantify the accuracy of alignment, we let clinical experts do the evaluation. Two experienced graders, which are involved in the diabetic retinopathy screening program, independently assessed the accuracy of the normalized mosaic images. Each of the graders evaluated the accuracy of the overall mosaic by assessing how well the vasculatures in the overlap region were aligned and assigned a grade to it. Mosaics were graded based on the region with the worst alignment. The possible grades were:

- **Off**: an image is placed at an incorrect location.

Table 2.2: Evaluation results of 100 mosaics from both graders. Each grader evaluated half of all the data.

	No. of mosaics		
Grade	**Grader 1**	**Grader 2**	**Total**
Off	2	2	4
Not Acceptable	21	3	24
Acceptable	76	37	113
Perfect	6	63	69

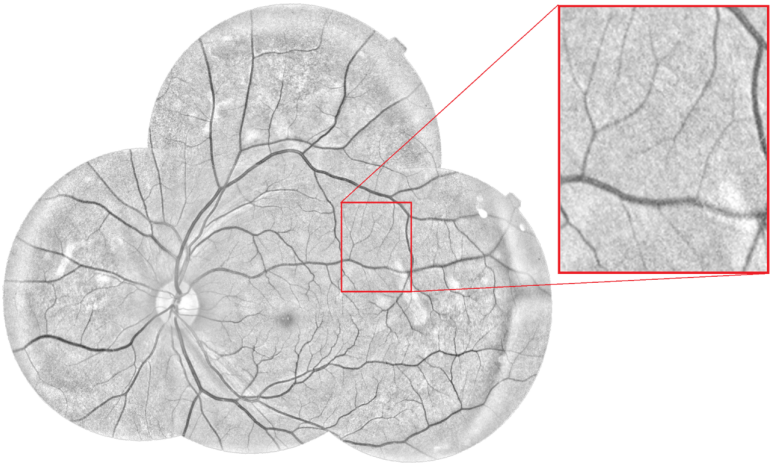


Figure 2.3: A fundus mosaic which was graded as 'perfect'. The zoomed in and overlaid image patch shows part of the mosaic in which three images overlapped.

- **Not Acceptable**: a misalignment larger than the width of a misaligned vessel.
- **Acceptable**: a misalignment smaller than the width of a misaligned vessel.
- **Perfect**: no noticeable misalignment.

It should also be noted that in our evaluation a mosaic is considered as 'not acceptable' even if the misalignment occurs in a very small fraction of the overlap region between two images.

2.3.4. Results

The evaluation results from both graders are summarized in table 2.2. Figure 2.3 shows a mosaic image which was graded as 'perfect'. A mosaic which was graded as 'not acceptable' is shown in figure 2.4. The overlap regions in the mosaics of figure 2.3 and 2.4 are constructed by averaging.

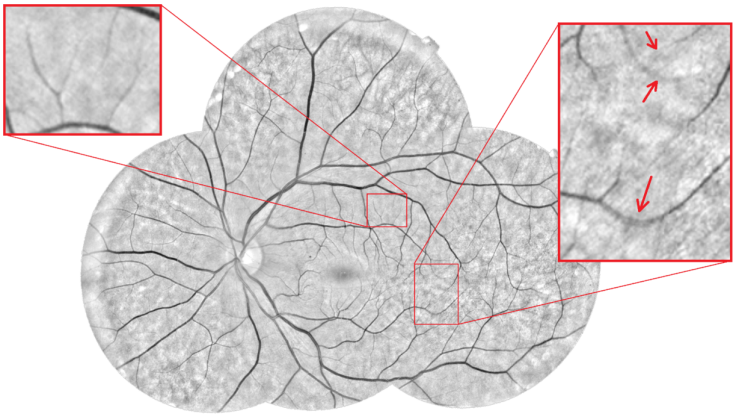


Figure 2.4: A fundus mosaic which was graded as 'not acceptable'. The arrows in the zoomed in and overlaid image patch mark the misaligned micro-vessels, resulting in a blurred or double appearance of the vessels. The image patch on the left shows accurately aligned fine vasculatures.

2.4. Discussion and Conclusion

In this paper, we present a robust hierarchical coarse-to-fine registration method for fundus images. The intensity as well as the structural information of the retinal vasculature are exploited to spatially align the four images. The method registers retinal images after normalization for luminosity and contrast variation within and between images. The alignment is done based on the vasculature-weighted MSD of the normalized images, solving the inherent limitation of feature-based algorithms of being dependent on the number and distribution of features. The robustness benefited greatly from the multiresolution matching strategy. We coupled a hierarchical coarse-to-fine registration with a deformation model of increasing complexity to estimate the parameters of a global second-order spatial transformation model. Careful initialization of each step with the results of the previous scale reduced the risk of getting trapped in a local minimum during the optimization.

Among the 100 mosaics created by the proposed method, 44 mosaics were free of any noticeable misalignment ('perfect' grade) and 45 mosaics received an 'acceptable' grade. Three mosaics were graded as 'off', all due to a failure in the first initialization stage. One of these failures could be attributed to a very poor image quality. Note that none of the 400 images were used to develop the method.

In the remaining eight mosaics, even though the accuracy of the alignment was good in most of the overlap area, a small misalignment of one or two micro-vessels resulted in a 'not acceptable' grade. The misalignments in these mosaics occurred mostly in fine vasculature regions (see figure 2.4). Here, the low signal-to-noise ratio resulted in a weak second-order local structure and, therefore, a low vesselness weight. In these cases, the registration was mainly guided by larger vasculature in regions around it.

References

[1] M. Abràmoff, M. Garvin, and M. Sonka, *Retinal imaging and image analysis,* Biomedical Engineering, IEEE Reviews in **3**, 169 (2010).

[2] L. Zhou, M. S. Rzeszotarski, L. J. Singerman, and J. M. Chokreff, *The detection and quantification of retinopathy using digital angiograms,* Medical Imaging, IEEE Transactions on **13**, 619 (1994).

[3] N. Ritter, R. Owens, J. Cooper, R. H. Eikelboom, and P. P. Van Saarloos, *Registration of stereo and temporal images of the retina,* Medical Imaging, IEEE Transactions on **18**, 404 (1999).

[4] G. K. Matsopoulos, N. A. Mouravliansky, K. K. Delibasis, and K. S. Nikita, *Automatic retinal image registration scheme using global optimization techniques,* Information Technology in Biomedicine, IEEE Transactions on **3**, 47 (1999).

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

[6] C. V. Stewart, C.-L. Tsai, and B. Roysam, *The dual-bootstrap iterative closest point algorithm with application to retinal image registration,* Medical Imaging, IEEE Transactions on **22**, 1379 (2003).

[7] T. Chanwimaluang, G. Fan, and S. R. Fransen, *Hybrid retinal image registration,* Information Technology in Biomedicine, IEEE Transactions on **10**, 129 (2006).

[8] S. Lee, M. D. Abràmoff, and J. M. Reinhardt, *Feature-based pairwise retinal image registration by radial distortion correction,* in *Medical Imaging* (International Society for Optics and Photonics, 2007) pp. 651220–651220.

[9] J. Zheng, J. Tian, K. Deng, X. Dai, X. Zhang, and M. Xu, *Salient feature region: a new method for retinal image registration,* Information Technology in Biomedicine, IEEE Transactions on **15**, 221 (2011).

[10] M. Foracchia, E. Grisan, and A. Ruggeri, *Luminosity and contrast normalization in retinal images,* Medical Image Analysis **9**, 179 (2005).

[11] H. Knutsson and C. F. Westin, *Normalized and differential convolution,* in *Computer Vision and Pattern Recognition, 1993. Proceedings CVPR '93., 1993 IEEE Computer Society Conference on* (1993) pp. 515–523.

[12] C. Wijk, R. Truyen, R. Gelder, L. J. Vliet, and F. Vos, *On normalized convolution to measure curvature features for automatic polyp detection,* in *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2004*, Lecture Notes in Computer Science, Vol. 3216 (Springer Berlin Heidelberg, 2004) pp. 200–208.

[13] A. Frangi, W. Niessen, K. Vincken, and M. Viergever, *Multiscale vessel enhancement filtering,* in *Medical Image Computing and Computer-Assisted Intervention - MICCAI 1998*, Lecture Notes in Computer Science, Vol. 1496 (Springer Berlin Heidelberg, 1998) pp. 130–137.

3

Accuracy assessment of intra- and inter-Visit fundus image registration for diabetic retinopathy screening

This chapter is based on the publication:

K. M. Adal, P. G. van Etten, J. P. Martinez, L. J. van Vliet, and K. A. Vermeer, *Accuracy assessment of intra-and intervisit fundus image registration for diabetic retinopathy screening*, Investigative ophthalmology and visual science 56, 1805 (2015).

3

Abstract

Purpose: *To evaluate the accuracy of a recently developed fundus image registration method (Weighted Vasculature Registration or WEVAR) and to compare it with two top-ranked state-of-the-art commercial fundus mosaicking programs (i2k Retina, DualAlign LLC and Merge Eye Care PACS, formerly named OIS AutoMontage) in the context of diabetic retinopathy (DR) screening.*

Methods: *Fundus images of 70 diabetic patients who visited the Rotterdam Eye Hospital in 2012 and 2013 for a diabetic retinopathy screening program were registered by all three programs. The registration results were used to produce mosaics from fundus photos that were normalized for luminance and contrast to improve the visibility of small details. These mosaics were subsequently evaluated and ranked by two expert graders to assess the registration accuracy.*

Results: *Merge Eye Care PACS had high registration failure rates compared to both WEVAR and i2k Retina ($p = 8 \times 10^{-6}$ and $p = 0.002$, respectively). WEVAR showed significantly higher registration accuracy than i2k Retina in both intra-visit ($p \leq 0.0036$) and inter-visit ($p \leq 0.0002$) mosaics. Fundus mosaics processed by WEVAR were therefore more likely to have a higher score (odds ratio (OR) = 2.5, $p = 10^{-5}$ for intra-visit and $OR = 2.2, p = 0.006$ for inter-visit mosaics). WEVAR was preferred more often by the graders than i2k Retina ($OR = 6.1, p = 7 \times 10^{-6}$).*

Conclusion: *WEVAR produced intra- and inter-visit fundus mosaics with higher registration accuracy than Merge Eye Care PACS and i2k Retina. Merge Eye Care PACS had higher registration failures than the other two programs. Highly accurate registration methods such as WEVAR may potentially be used for more efficient human grading and in computer-aided screening systems for detecting DR progression.*

3.1. Introduction

Diabetic Retinopathy (DR) is one of the most common complications of Diabetes Mellitus (DM) and results in vision loss and even blindness if not diagnosed and treated adequately. An estimated 422 million people worldwide are reported to have diabetes [1], and diabetic retinopathy accounts for 4.8% of the 37 million cases of blindness worldwide [2]. The current practice of DR screening is based on regular examinations of a series of fundus images. A retinal specialist looks for pathognomonic abnormalities. In addition, manual grading is time-consuming, subjective, and limits the efficiency of the available DR screening facilities. Automated registration of fundus images can be instrumental to alleviate this problem and increases the efficiency of DR screening in two ways. Firstly, intra-visit images that capture partially overlapping regions of the same retinal surface can be automatically registered to create a mosaic of the retina, enabling clinicians to do a comprehensive retinal examination at a single glance. Secondly, registration of inter-visit image sets allows longitudinal analysis, facilitating retinal change detection to monitor DR development and progression.

In addition to pre-processing retinal images for more efficient human grading, fundus image registration is often used as part of computer-aided screening systems for detecting DR progression and longitudinal changes [3–6]. Over the last decade, several computer-aided diagnosis (CAD) systems have been developed to analyze digital fundus images for symptoms of diabetic retinopathy [4, 7–17]. The performance of these systems are comparable to expert readers in distinguishing fundus images of a normal retina from those with DR symptoms [4, 10–18]. CAD systems could thus be used in DR screening such that experts only have to evaluate suspicious or difficult cases [16–18]. Moreover, registration of fundus images captured across multiple exams enables CAD systems to identify and analyze retinal surface changes due to disease progression.

Tracking small retinal features, such as microaneurysms, over time needs a very high registration accuracy. This requires a thorough evaluation of image registration methods for DR screening. Evaluation can be done either by expert graders based on visual inspection of the registered image pairs or by objective, automatic computer algorithms that assess the registration accuracy between corresponding landmark points. Due to the sparse distribution of landmark points in the field-of-view and the difficulty to accurately extract and match these points, an objective registration accuracy assessment may be limited to a few regions. On the other hand, visual inspection by expert graders permits qualitative accuracy assessment of the entire field-of-view. Moreover, clinicians are likely to focus on regions of clinical interest, thereby producing a more clinically relevant accuracy assessment.

In this study, the accuracy of a recently developed fundus image registration method (WEVAR) was systematically evaluated by clinical experts in the context of automated diabetic retinopathy screening [19]. The evaluation was performed on intra-visit and inter-visit fundus image sets acquired from diabetic patients who had annual retinal exams for DR. A comparison was made with state-of-the-art commercially available fundus mosaicking programs i2k Retina (DualAlign LLC, Clifton Park, NY) and OIS AutoMontage (OIS, Sacramento, CA). These programs ranked first

and second, respectively, in a recent comparative study that also included IMAGEnet Professional (Topcon, Oakland, NJ; ranked third) [20]. A full evaluation was done for WEVAR and i2k Retina (version 2.1.6), while Merge Eye Care PACS (Version 4.2.0.4221), the successor of OIS AutoMontage, was only partially evaluated due to high registration failure rates.

3.2. Methods

3.2.1. Data description

This retrospective observational study was conducted on fundus images that were captured during annual retinal examinations of diabetic patients who were enrolled in the ongoing DR screening program of the Rotterdam Eye Hospital in The Netherlands. A representative sample of the screening population was gathered by including all patients who were examined in a one week period in June 2013. During this period, a total of 85 patients were screened for diabetic retinopathy. Because repeated examinations were needed for our evaluation, first-time patients and those who were not examined in the year before were excluded. All fundus images were acquired after pupil dilation (one drop of tropicamide 0.5%) using a non-mydriatic digital funds camera (Topcon TRC-NW6S, Tokyo, Japan) with a 45° field-of-view. The fundus images were 2000×1312 pixels in size. Although clinical guidelines suggest two fields per eye for screening purposes [21–23], in this screening program four fields are acquired per examination (See Figure 3.1): images of macula-centered, optic nerve-centered, superior, and temporal regions of the retinal surface were acquired from both eyes.

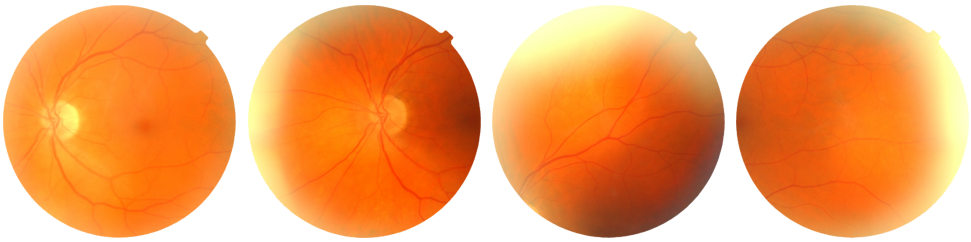


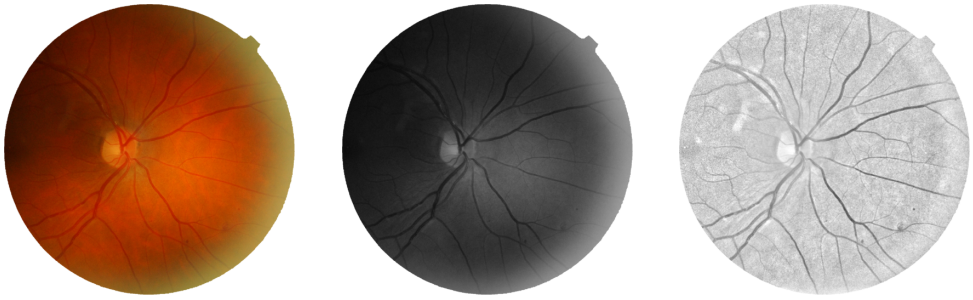
Figure 3.1: An example of a four field fundus image set captured during a retinal examination. From left to right: macula-centered, optic nerve-centered, superior, and temporal fundus images of a left eye.

This study adhered to the applicable code of conduct for the reuse of data in health research [24]. After exporting the fundus images from the clinical image storage system, all data was anonymized prior to further processing.

3.2.2. Fundus Image Normalization

Color fundus images often show highly variable luminosity and contrast due to non-uniform illumination of the retina during acquisition. Because of its higher contrast, the green channel of the digital fundus images (see Figure 3.2a and 3.2b), closely resembling red-free fundus photos, is commonly used in CAD of fundus images. However, the green channel images still show considerable variation in luminosity

and contrast, both within and between images. Foracchia et al. [25] proposed a method to normalize retinal images based on estimates of the local luminosity and contrast from the intensity distribution of the so-called background retina (which excludes features such as vessels, optic disc, and lesions) and subsequently correcting for their variation over the entire retinal image. However, this method does not compensate for all illumination variation, especially around the rim of fundus images. Recently, this limitation was addressed by applying a higher-order normalized convolution, resulting in a considerably larger area with discernible retinal features (see Figure 3.2c) [19].



(a) (b) (c)

Figure 3.2: An example of a fundus image from our data set. (a) Color fundus image. (b) Green channel. (c) Normalized fundus image using the improved normalization method.

The enhanced visibility of retinal features in these normalized images is not only beneficial for further processing by computer algorithms but may also be used by clinicians for a better evaluation of the fundus. The graders who participated in this study preferred the normalized images over the color and green channel fundus images. Therefore, all evaluations in this study were based on normalized image.

3.2.3. Registration Methods for Fundus Image Mosaicking

Fundus image registration is the process of spatially mapping two or more images of the retina into a common coordinate system. The resulting spatial correspondence allows for combining the images into a single mosaic of the retinal surface in order to facilitate comprehensive retinal examination at a single glance [26]. The registered images can also be used in CAD and longitudinal analysis of fundus photos to detect and analyze retinal changes due to disease progression. Because of the spherical shape of the human eye, fundus photography involves a non-linear spatial deformation of the curved retina onto an image plane. Correctly modeling this deformation is central for accurate spatial mapping between fundus images captured from multiple views of the retina [26]. Different attributes of fundus images, such as the raw intensity, the vasculature tree and its bifurcations, may be used to determine the optimal spatial mapping parameters. In this study, two fundus image registration methods were extensively evaluated: WEVAR [19] and i2k Retina, the latter representing the state-of-the-art in fundus image registration methods [20].

The main difference between the two methods lies in the fundus image attributes they use for the registration.

In brief, WEVAR aligns fundus images based on intensity and structural information derived from the retinal vasculature [19]. The method starts by normalizing the green channel of the fundus images for luminosity and contrast. The optimal alignment of the normalized images is then determined using a multi-resolution matching strategy coupled with a deformation model of progressive complexity. For each intra- and inter-visit image set of each eye, the method automatically selects the image having the largest overlap with the other images as the anchor image. Then, the image with the largest overlap to the anchor image is mapped sequentially to the coordinate system of this anchor image. This result becomes the new anchor image and the procedure is repeated for the remaining images until all images have been registered. This yields a set of normalized images which were transformed into a common coordinate system. These outputs are then combined into a mosaic for grading (see Figure 3.3a).

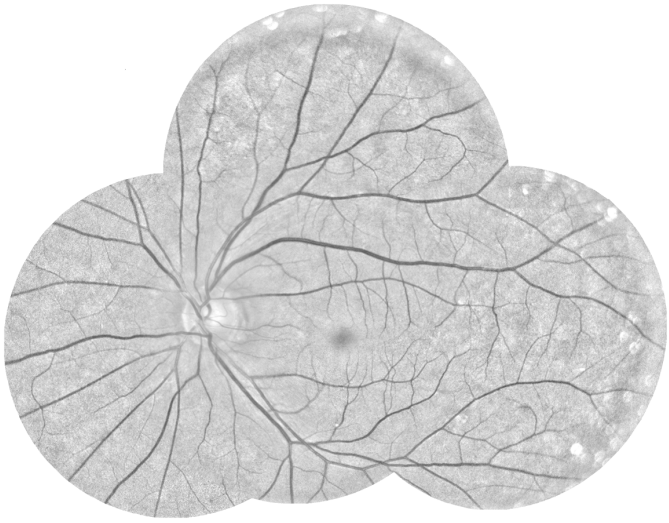
i2k Retina finds similar corresponding regions between a pair of images based on the information extracted from landmark points of the retinal vasculature [27, 28]. The program initializes the alignment by matching features extracted from vessel bifurcations and crossover points in the image pairs. The results are then refined based on vessel centerlines. Hence, the method does not make use of most of the other intensity and structural information within fundus images. Each complete set of color fundus images that needs to be registered were loaded into the i2k Retina program and aligned to one coordinate system using the default program settings. No pre-processing, such as normalization, was performed on the color fundus images before processing by i2k Retina, because the software may have its own internal pre-processing algorithms. To compare the registration produced by both methods, the green channel of the individual color images were normalized for variations in luminosity and contrast as before and then combined into a mosaic using the spatial mapping that was determined during registration (see Figure 3.3b).

3.2.4. Registration Accuracy Assessment

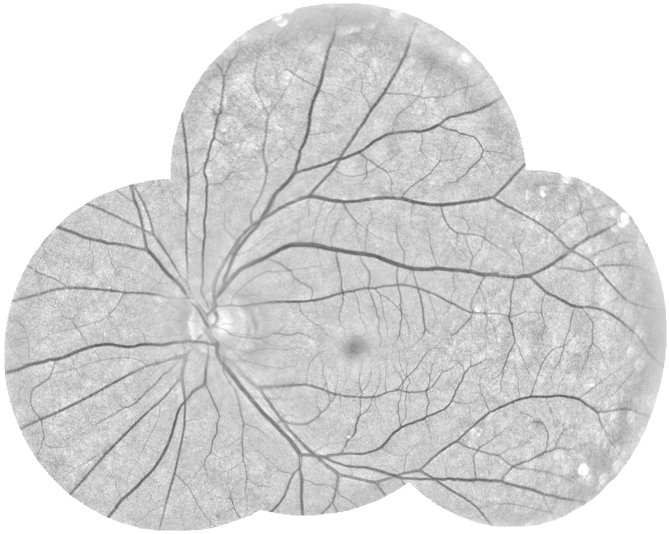
In this study, two experienced graders who are involved in DR care including screening and diagnosis of diabetic retinopathy, independently assessed the registration accuracy of WEVAR and i2k Retina by scoring both intra and inter-visit fundus mosaics. The graders also ranked the mosaics produced by both methods in a side-by-side comparison.

Grading mosaics is a time-consuming task and therefore each grader did not evaluate all data. However, to be able to compare the scores between graders, half of the available data were assessed by both graders. The remaining half were divided equally between the two graders. Note that the two mosaics of each eye by both methods were scored by the same grader. Since the side-by-side comparison was less time-consuming, both graders scored all data.

In the intra-visit evaluation, the accuracy of the fundus mosaics constructed from registered fundus images that were captured during one examination was assessed. Conventionally, when combining multiple fundus images into one mosaic,



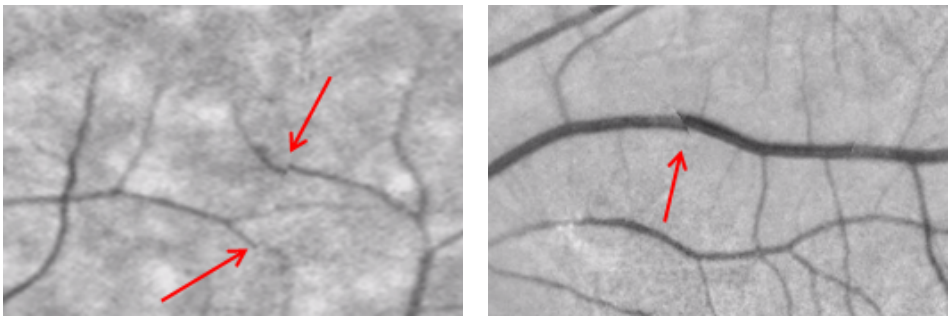
(a)



(b)

Figure 3.3: An example of a correctly registered intra-visit fundus mosaic by WEVAR method (a) and i2k Retina (b).

overlapping areas are averaged. Although averaging or more advanced blending methods produce visually appealing results, it conceals misalignment of retinal features and thereby hinders the quality assessment. In this study intra-visit mosaics were created by stacking the four registered images on top of each other. By chang-



(a) (b)

Figure 3.4: Examples of image patches showing vessel misalignments. The arrows in the image patches mark misalignment locations. (a) Misalignments larger than the width of the misaligned vessels. (b) Misalignment smaller than the width of the misaligned vessel.

ing the order of the images in the stack, each image appeared in the top layer once, resulting in four mosaics that differed in the regions where the images overlapped. These mosaics were put together in a video and played repeatedly for grading (see intra-visit supplementary materials for an example). The graders evaluated each mosaic by visually inspecting the vasculature alignment in the overlapping regions and assigned one of the following grades to it:

- **Off**: at least one image is fully misplaced.
- **Not Acceptable**: misalignment larger than the width of the misaligned vessel (see Figure 3.4a).
- **Acceptable**: misalignment smaller than the width of the misaligned vessel (see Figure 3.4b).
- **Perfect**: no noticeable misalignment.

Graders were instructed to base their score on the region with the worst alignment. Hence a mosaic was graded as 'not acceptable' even if the misalignment occurred only in a small region of the mosaic. i2k Retina sometimes discarded one or more images which could not be registered into a mosaic; these mosaics were given the score 'off '.

In the inter-visit accuracy evaluation, all images were registered to a common coordinate system and a mosaic was produced for each visit. The two mosaics then alternated in a video and played repeatedly for grading (see inter-visit supplementary materials for an example), using the same grading scheme as for the intra-visit evaluation.

In the third evaluation, the registration methods were ranked in a side-by-side comparison for each pair of intra-visit mosaics. The mosaics of both methods were produced from the registered intra-visit fundus images by averaging overlapping areas and each grader ranked all 140 resulting intra-visit mosaic pairs that were

displayed simultaneously on two identical monitors (1920×1080 pixels resolution). The possible grades were 'slightly better 'or 'much better 'for either mosaic, or 'equal 'if both were of the same quality. To avoid bias, the monitor that presented the result of each method was selected randomly for each mosaic pair.

In all three evaluations, the graders were blinded with respect to the method that was used for registration in each mosaic. Moreover, in all accuracy assessments, the mosaics of all eyes from both methods were presented in random order to the graders to avoid any bias.

3.2.5. Data availability

All data that was used in this study is made publicly available through the Rotterdam Ophthalmic Data Repository (http://rod-rep.com). This includes the source data (1120 fundus images), the processed data (1120 normalized fundus images, all intra- and inter-visit mosaic movies and images used for grading) and all grading results (intra- and inter-visit mosaic grading and ranking).

3.2.6. Statistical Analysis

For each evaluation, two types of analyses were performed: First, the grades for both methods were evaluated for each grader separately by conventional nonparametric statistical analyses. Second, a comprehensive statistical model was defined to simultaneously evaluate all grades of both graders.

In the evaluation per grader, the grades assigned to each method were compared. To assess the difference between grades assigned to WEVAR and i2k Retina, a Wilcoxon signed-rank test was applied to the intra- and inter-visit grades. Then, to quantify the preference of a grader for either method, the odds ratio (OR) of the methods was computed from the ranking grades and its significance was tested by Fisher's exact test. The OR for each method was defined as the ratio of the number of cases that the method was preferred over all other cases. To determine the intergrader agreement and consistency, the intraclass correlation coefficient $ICC(3,1)$ was calculated. The ICC values were interpreted as follows: < 0.4 corresponds to poor, $0.4 - 0.75$ was fair to good, and > 0.75 was excellent agreement or consistency [29].

For a comprehensive statistical analysis of each evaluation, proportional odds mixed models were used. Here all mosaics are modelled as random effects whereas the methods and graders (and their interaction) are modelled as fixed effects. The odds ratios resulting from this model were used to quantify the influence of the aforementioned effects on the grade. Such an OR is defined as the ratio of the odds that an image gets a better grade including a certain effect over the same odds excluding that effect. The analysis for the side-by-side method ranking was based on a proportional odds model with only the graders as fixed effects. The results of this model were used to compute the odds ratio for the methods which was then used to determine the preference of one method over the other. The odds of each method was defined as the ratio of the probability that a method is preferred over all other cases.

3.3. Results

During the one week screening period, 85 patients were examined for DR; among these patients, 4 were first-time patients and 11 were not examined the year before, resulting in 70 patients who had consecutive retinal examinations. A total of 1120 fundus images was acquired from 70 patients. At the time of the examination in 2012, the average age of the patients was 63 years ($SD.$ 12 years), 33 (47.1%) were male, and 37 (52.9%) were female. From 70 patients, 140 intra- and 140 inter-visit fundus photo sets were processed by WEVAR and i2k Retina to produce mosaic movies. The 140 intra-visit image sets also were processed by Merge Eye Care PACS; however, the results (described later) did not warrant further evaluation by the expert graders. The mosaic movies from WEVAR and i2k-Retina were independently assessed by two expert graders. Of the mosaic movies from each method, 70 were graded by both graders, the other mosaics were graded by a single grader. The resulting grades are summarized in Tables 3.1 and 3.2. The results showed that WEVAR produced more 'acceptable 'or 'perfect 'mosaics and fewer 'off 'cases than i2k Retina according to both graders. Each grader assigned significantly more often a higher grade to the WEVAR than to i2k Retina in intra-visit (Wilcoxon signed-rank test, $p = 0.0036$ and $p = 0.0006$ for graders 1 and 2, respectively) and inter-visit ($p = 0.0002$ and $p = 0.0001$ for graders 1 and 2, respectively) mosaic evaluations. A partial evaluation of Merge Eye Care PACS revealed that it failed to register one or more images into a mosaic, that is, 'off 'cases, in 19 (of 140) intra-visit image sets. This was significantly higher compared to i2k Retina and WEVAR (McNemar's test, $p = 0.002$ and $p = 8 \times 10^{-6}$, respectively). Therefore, Merge Eye Care PACS was excluded from further evaluation.

Table 3.1: Summary of the grades assigned to the intra-visit mosaics produced by both methods. Each grader evaluated 105 out of 140 available mosaics from each of the methods. 70 mosaics were evaluated by both graders.

			WEVAR			
			Off	Not Acceptable	Acceptable	Perfect
Grader 1	i2k Retina	Off	2	1	1	-
		Not Acceptable	-	11	24	1
		Acceptable	-	8	48	4
		Perfect	-	1	3	1
Grader 2	i2k Retina	Off	2	-	1	3
		Not Acceptable	-	-	3	4
		Acceptable	-	1	19	30
		Perfect	-	2	14	26

In Table 3.3, the ranks assigned to each of the methods in the side-by-side comparison of intra-visit mosaic pairs are summarized. WEVAR produced mosaics which were preferred more often by both graders than i2k Retina. Examples of pairs of mosaics which were compared and ranked are shown in figure 3.5. For

Table 3.2: Summary of the grades assigned to the inter-visit mosaics produced by both methods. Each grader evaluated 105 out of 140 available mosaics from each of the methods. 70 mosaics were evaluated by both graders.

			WEVAR			
			Off	Not Acceptable	Acceptable	Perfect
Grader 1	i2k Retina	Off	1	1	2	-
		Not Acceptable	-	22	21	-
		Acceptable	-	4	50	2
		Perfect	-	0	2	-
Grader 2	i2k Retina	Off	1	-	-	4
		Not Acceptable	-	6	1	7
		Acceptable	-	-	1	3
		Perfect	-	-	1	81

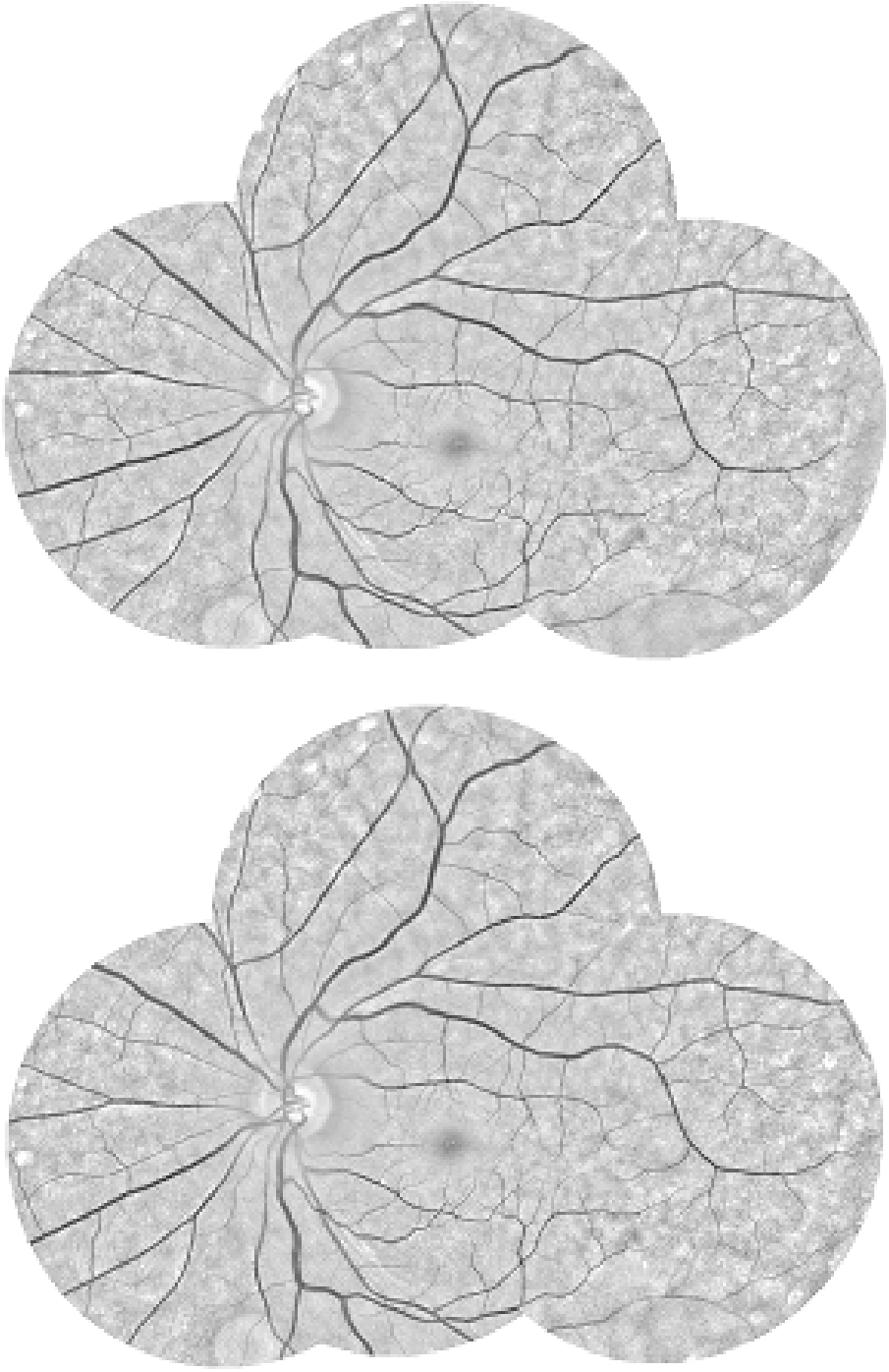
Table 3.3: Summary of the ranks assigned to the methods. Each grader ranked all 280 mosaics produced by the two methods.

	i2k Retina			WEVAR	
	A: Much better	B: Slightly better	C: Equal	D: Slightly better	B: Much better
Grader 1	1	2	120	7	10
Grader 2	1	3	115	13	8

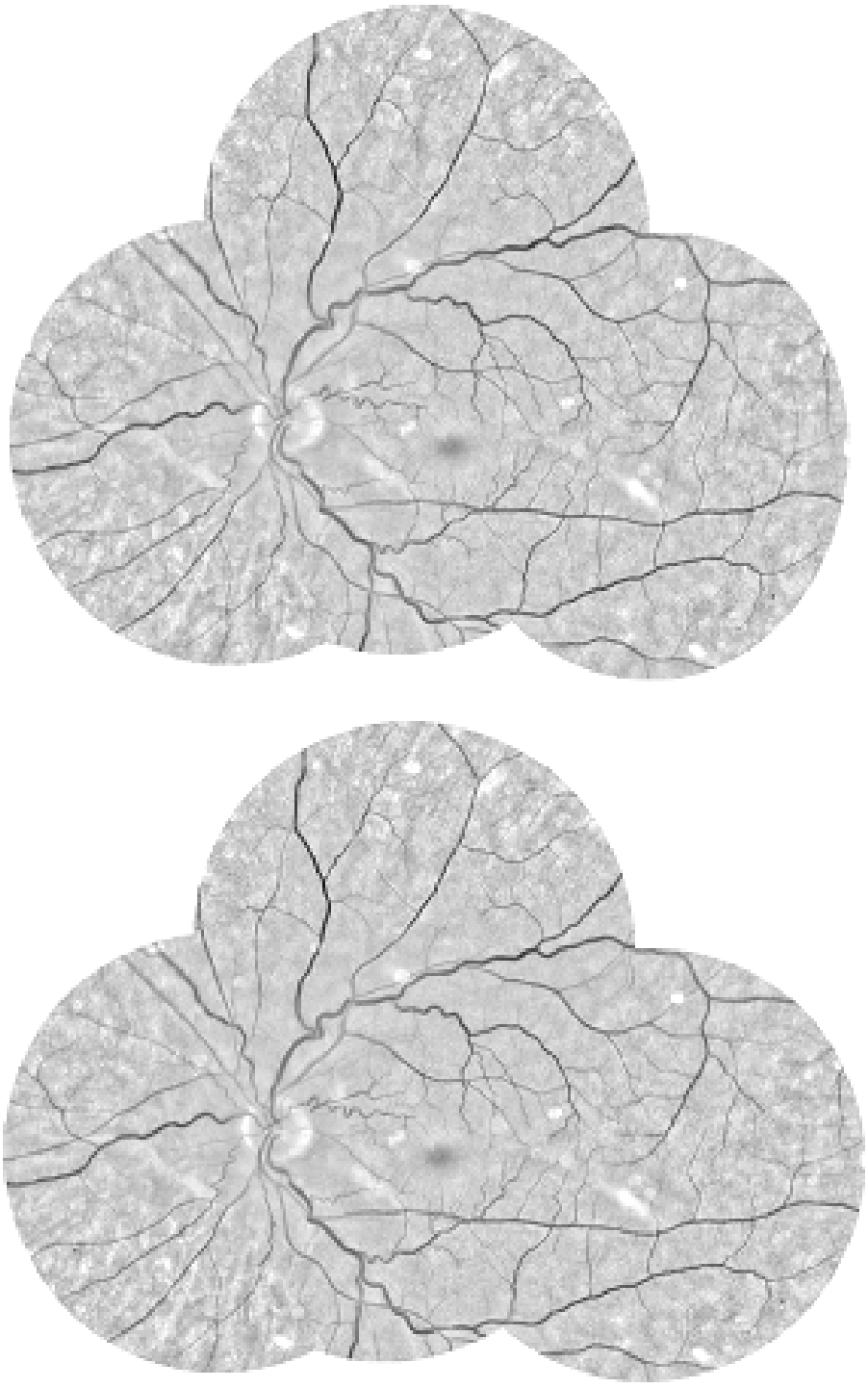
grader 1, the odds of preferring WEVAR, expressed by the ratio of grades D and E over grades A, B and C, was 0.14; the odds of preferring i2k Retina (A and B over C, D and E) was 0.02. The resulting odds ratio was 6.3. For grader 2, the odds ratio was 6.0, showing that a higher rank was assigned significantly more frequently to mosaics produced by WEVAR than to mosaics produced by i2k Retina.

Although the results indicated that WEVAR yields significantly higher accuracy in both intra- and inter-visit registration than i2k Retina, the inter-grader agreement and consistency between the grades assigned by both graders ranged from poor to moderate levels ($ICC(3,1)$ agreement and consistency of 0.52 and 0.65, respectively for the intra-visit grades and 0.35 and 0.71 for inter-visit grades). Thus, the data cannot simply be pooled for analysis by ignoring the grader. Instead, a comprehensive statistical analysis based on the proportional odds mixed model was applied to the grades of both graders altogether.

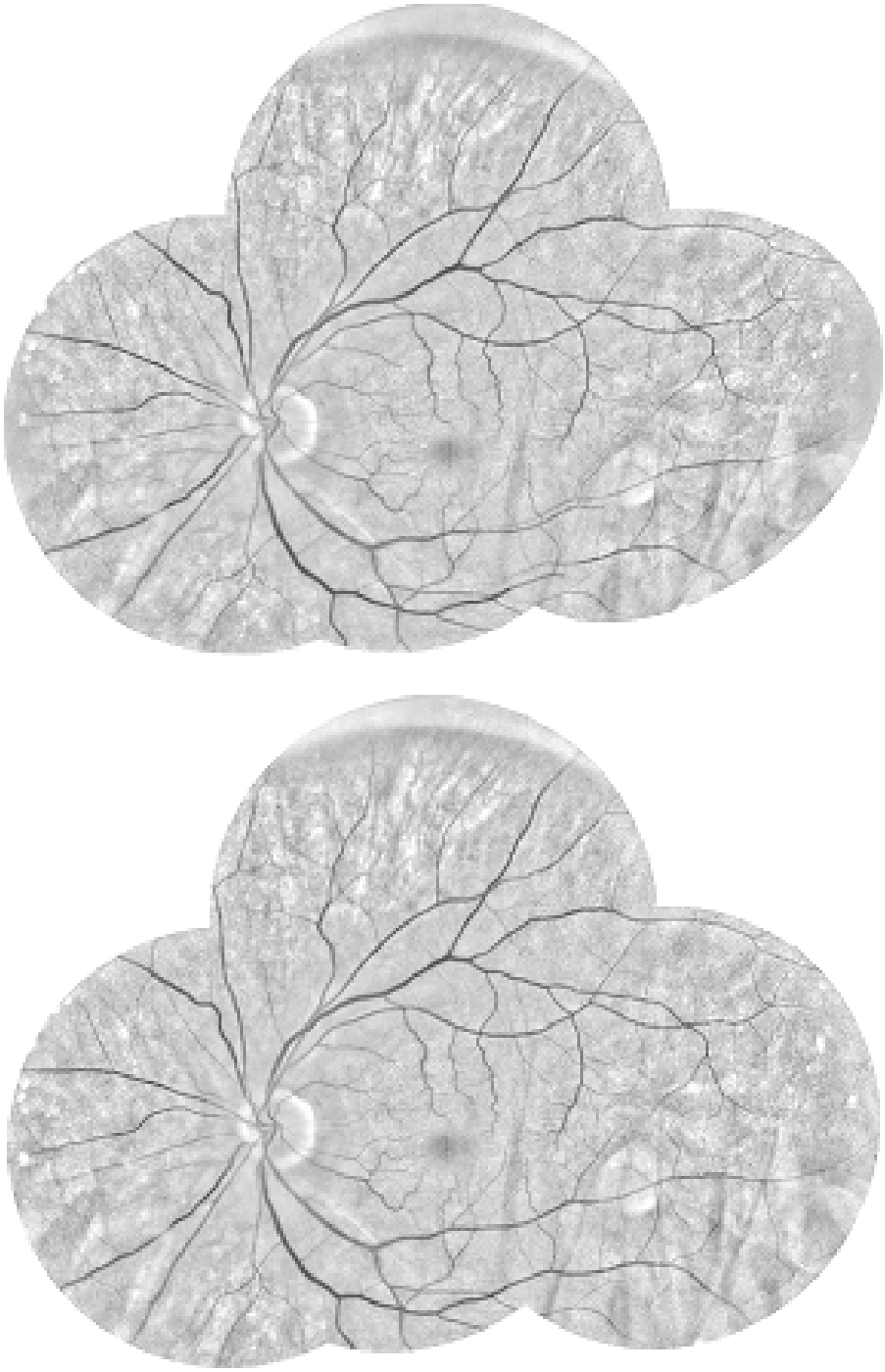
The fitted parameters of a proportional odds mixed model for the intra-visit grades considering the method and grader as fixed effects and the fundus image set as a random effect are summarized in Table 3.4. Interaction between the methods and graders was not included in the final model as its effect on the grades was insignificant ($p = 0.17$). The resulting coefficient of the method indicates that mosaics processed by WEVAR were significantly more likely to receive a higher



(a)



(b)



(c)

Figure 3.5: Examples of mosaic from i2k Retina and WEVAR which were compared and ranked side-by-side. Top: mosaics processed by i2k Retina, Bottom: mosaics processed by WEVAR. The graders were blinded to the identity of the program that produced each mosaic. In (a) the pair of mosaics were ranked as 'equal'. The mosaic by i2k Retina was ranked as 'slightly better' in (b), whereas in (c), the mosaic produced by WEVAR was ranked as 'much better'.

Table 3.4: The estimated coefficients of the proportional odds mixed model fit for the intra-visit grades excluding the effect of the interaction between the methods and graders.

Effects	Estimate	Std. Error	p-Value
method	0.94	0.22	10^{-5}
grader	2.55	0.28	$< 2 \times 10^{-16}$

Table 3.5: The estimated coefficients of the proportional odds mixed model fit for the inter-visit grades.

Effects	Estimate	Std. Error	p-Value
method	0.79	0.29	0.006
grader	5.06	0.53	$< 2 \times 10^{-16}$
method × grader	1.23	0.60	0.037

grade compared to i2k Retina $p = 10^{-5}$. The coefficient of 0.94 corresponds to an odds ratio of $e^{0.94}$. Thus, the odds of receiving a higher score were 2.5 times more likely for WEVAR than for i2k Retina. The model corrected for the fact that the odds of receiving a higher grade from grader 2 were far higher ($OR = 13.0$).

Similarly, Table 3.5 shows the parameters of a proportional odds mixed model fit for the inter-visit grades. In this case, the interaction between the methods and graders was significant. The estimated coefficient associated with the method's effect indicates that for both graders the inter-visit mosaics processed by WEVAR are significantly more likely to receive a higher grade compared to i2k Retina ($p = 0.006$). In addition, grader 2 assigned a higher grade compared to grader 1. The resulting odds ratio of WEVAR vs i2k Retina is $e^{0.79} = 2.2$ and $e^{0.79+1.23} = 7.5$ for grader 1 and 2, respectively.

In order to quantify the difference in accuracy and quality between the intra-visit mosaics of the two methods, the ranks of each method from the side-by-side comparison were also modeled using the proportional odds model. As expected, due to the relative nature of the scores, the grader effect was insignificant ($p = 0.68$). The resulting model parameters are shown in Table 3.6. The odds of preferring WEVAR was $e^{-1.85} = 0.16$, while the odds of preferring i2k Retina was $e^{-3.66} = 0.03$. The odds ratio was therefore 6.1 ($p = 7 \times 10^{-6}$), indicating that the graders were 6.1 times more likely to prefer the results of WEVAR.

3.4. Discussion

This study assessed the accuracy of vessel alignment in mosaics constructed from intra and inter-visit fundus image sets for a recently developed fundus registration method (WEVAR). An extensive accuracy assessment and comparison with two top-ranked state-of-the-art commercial fundus mosaicking programs (i2k Retina and Merge Eye Care PACS) shows that WEVAR yields a significantly higher registration accuracy in both intra-visit ($p \leq 0.0036$) and inter-visit ($p \leq 0.0002$) mosaics. The likelihood of receiving a higher score was 2.5 ($p = 10^{-5}$) and 2.2 ($p = 0.006$) times

Table 3.6: The estimated parameters of the proportional odds mixed model for the ranking grades (see Table 3). The estimates show the log-odds of a grade below the threshold.

Thresholds	Estimate	Std. Error
Between A and B	-4.93	0.71
Between B and C	-3.66	0.38
Between C and D	1.85	0.17
Between D and E	2.68	0.24

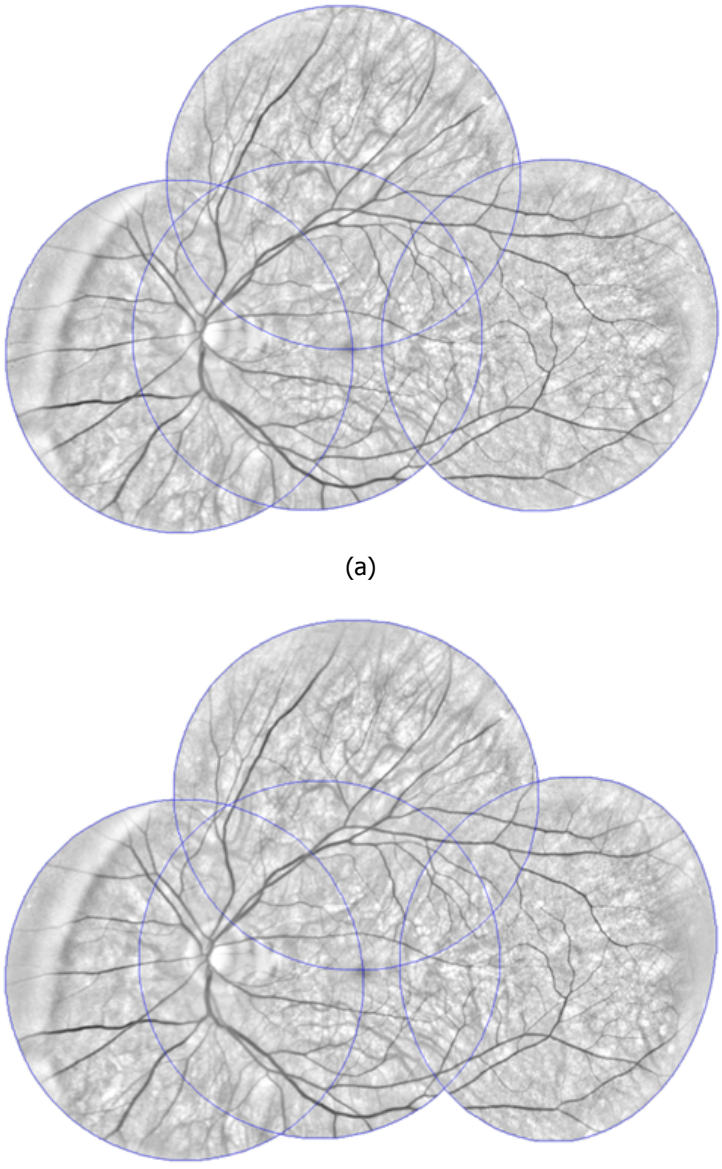
higher for WEVAR than for i2k Retina on intra- and inter-visit mosaics, respectively. Due to a very high registration failure rate, Merge Eye Care PACS was excluded from a full evaluation. Despite the generally higher scores from one grader, the results from both graders show that WEVAR has a significantly higher registration accuracy and significantly fewer failures than i2k Retina. A comprehensive statistical analysis, taking into account the inter-grader variability, also revealed a strong association between the grades assigned to each mosaic and the method used to produce it. In the side-by-side comparison, both graders preferred WEVAR over i2k Retina. The two graders are very experienced in grading DR, but not in assessing registration accuracy. Therefore, the difference between the scores of the two experts might be attributed to differences in evaluation strategies of the mosaics.

The WEVAR program aligns fundus images based on intensity and structural information derived from the retinal vasculature [19]. Therefore, the better the visibility of the retinal vasculature is, the more accurate the registration results become. To this end, the normalized fundus images, created by compensating for the local luminosity and contrast variations, are crucial to enhance the visibility and contrast of especially small retinal structures over the entire field of view and therefore improve the registration accuracy.

Moreover, the enhanced visibility of retinal features in the normalized images may be useful for more sensitive detection of registration errors than color or green channel images. In a clinical setting, the improvement in contrast of retinal structures may enhance the detection of dark red spots or lesions, such as microaneurysms, hemorrhages, and intra-retinal microvascular abnormalities. It will be interesting to further evaluate whether the normalized images can give a higher screening sensitivity without adversely affecting specificity.

The fundus mosaic movies introduced in this study provide a useful way to analyze multiple fundus images of the retina. The intra-visit mosaic movies highlight any possible misalignment between overlapping images, which is useful for human graders to assess the registration accuracy. The inter-visit mosaic movies are also useful in clinical practice, allowing experts to compare a series of fundus images in an efficient manner.

During the side-by-side comparison, differences in image deformation between the mosaics of the two methods were observed. This difference is mainly evident in the registered temporal fundus images (see Figure 3.3 and 3.6). This deformation was observed frequently in the mosaics produced by i2k Retina. This might be due



(a)

(b)

Figure 3.6: An example of an intra-visit fundus mosaics of the same eye produced by using: (a) WEVAR and (b) i2k Retina. Note the difference in deformation between the two mosaics despite the registration of the individual images being correct in both cases.

to the possibly very small number of available features that i2k Retina uses to match the temporal, superior, and macula-centered fundus images that have a (very) small overlap region. Such image deformation cause changes in shape and area of retinal pathology and thus may hinder a correct interpretation. WEVAR on the other hand introduces little deformation, and yet provides higher quality mosaics that allows clinical experts to make accurate diagnosis.

Both intra- and inter-visit fundus mosaics of WEVAR can improve the efficiency of today's DR screening practice. Intra-visit mosaics provide a single large field of the retina for comprehensive analysis and more efficient grading. Inter-visit mosaics can be used to analyze fundus photos of successive retinal exams to monitor DR progression through biomarkers such as the microaneurysm turnover rate. Although our target application is DR screening, these mosaics could also be used in diagnosis and monitoring of other retinal diseases such as age-related macular degeneration.

Current clinical guidelines on referral of DR patients are based on the presence and detection of lesions in fundus images and therefore exclude the dynamics of these lesions. This implies that longitudinal analysis, and therefore accurate registration, of fundus images does not play a significant role in current clinical care. However, recent studies suggest that the progression rate of microaneurysms over time may be a better biomarker for DR progression than the differences between the number of microaneurysms at successive examination [3, 5, 6, 30]. These studies also suggest a correlation between the microaneurysm turnover rate and the likelihood of developing clinically significant macular edema (CSME).

We showed that WEVAR was significantly better in constructing intra- and inter-visit mosaics and obtained a significantly higher registration accuracy than Merge Eye Care PACS and i2k Retina. Merge Eye Care PACS had high registration failures compared to both WEVAR and i2k Retina. A higher registration accuracy is of clinical interest as it is an essential step towards having an automated and reliable objective disease progression measure for progressive eye disease such as DR. An objective progression measure, such as microaneurysm turnover rate, aids clinicians in assessing disease progression for a proactive and effective screening and treatment planning, thereby improving the quality of service provided by eye care centers.

References

[1] W. H. Organization *et al.*, *Global report on diabetes: World health organization,* (2016).

[2] W. H. Organization, *Prevention of blindness from diabetes mellitus: report of a WHO consultation in Geneva, Switzerland, 9-11 November 2005* (World Health Organization, 2006).

[3] K. A. Goatman, M. J. Cree, J. A. Olson, J. V. Forrester, and P. F. Sharp, *Automated measurement of microaneurysm turnover,* Investigative ophthalmology & visual science **44**, 5335 (2003).

[4] H. Narasimha-Iyer, A. Can, B. Roysam, C. V. Stewart, H. L. Tanenbaum, A. Majerovics, and H. Singh, *Robust detection and classification of longitudinal changes in color retinal fundus images for monitoring diabetic retinopathy,* Biomedical Engineering, IEEE Transactions on **53**, 1084 (2006).

[5] R. Bernardes, S. Nunes, I. Pereira, T. Torrent, A. Rosa, D. Coelho, and J. Cunha-Vaz, *Computer-assisted microaneurysm turnover in the early stages of diabetic retinopathy,* Ophthalmologica **223**, 284 (2009).

[6] J. Cunha-Vaz, R. Bernardes, T. Santos, C. Oliveira, C. Lobo, I. Pires, and L. Ribeiro, *Computer-aided detection of diabetic retinopathy progression,* in *Digital Teleretinal Screening* (Springer, 2012) pp. 59–66.

[7] N. Patton, T. M. Aslam, T. MacGillivray, I. J. Deary, B. Dhillon, R. H. Eikelboom, K. Yogesan, and I. J. Constable, *Retinal image analysis: concepts, applications and potential,* Progress in retinal and eye research **25**, 99 (2006).

[8] R. J. Winder, P. J. Morrow, I. N. McRitchie, J. Bailie, and P. M. Hart, *Algorithms for digital image processing in diabetic retinopathy,* Computerized medical imaging and graphics **33**, 608 (2009).

[9] M. Abràmoff, M. Garvin, and M. Sonka, *Retinal imaging and image analysis,* Biomedical Engineering, IEEE Reviews in **3**, 169 (2010).

[10] M. D. Abràmoff, M. Niemeijer, M. S. Suttorp-Schulten, M. A. Viergever, S. R. Russell, and B. Van Ginneken, *Evaluation of a system for automatic detection of diabetic retinopathy from color fundus photographs in a large population of patients with diabetes,* Diabetes care **31**, 193 (2008).

[11] M. D. Abràmoff, J. M. Reinhardt, S. R. Russell, J. C. Folk, V. B. Mahajan, M. Niemeijer, and G. Quellec, *Automated early detection of diabetic retinopathy,* Ophthalmology **117**, 1147 (2010).

[12] S. Philip, A. D. Fleming, K. A. Goatman, S. Fonseca, P. Mcnamee, G. S. Scotland, G. J. Prescott, P. F. Sharp, and J. A. Olson, *The efficacy of automated "disease/no disease" grading for diabetic retinopathy in a systematic screening programme,* British Journal of Ophthalmology **91**, 1512 (2007).

[13] M. D. Abràmoff, J. C. Folk, D. P. Han, J. D. Walker, D. F. Williams, S. R. Russell, P. Massin, B. Cochener, P. Gain, L. Tang, *et al.*, *Automated analysis of retinal images for detection of referable diabetic retinopathy,* JAMA ophthalmology **131**, 351 (2013).

[14] A. D. Fleming, K. A. Goatman, S. Philip, G. J. Williams, G. J. Prescott, G. S. Scotland, P. McNamee, G. P. Leese, W. N. Wykes, P. F. Sharp, *et al.*, *The role of haemorrhage and exudate detection in automated grading of diabetic retinopathy,* British Journal of Ophthalmology **94**, 706 (2010).

[15] C. I. Sánchez, M. Niemeijer, A. V. Dumitrescu, M. S. Suttorp-Schulten, M. D. Abramoff, and B. van Ginneken, *Evaluation of a computer-aided diagnosis system for diabetic retinopathy screening on public data,* Investigative ophthalmology & visual science **52**, 4866 (2011).

[16] G. S. Scotland, P. McNamee, A. D. Fleming, K. A. Goatman, S. Philip, G. J. Prescott, P. F. Sharp, G. J. Williams, W. Wykes, G. P. Leese, *et al.*, *Costs and consequences of automated algorithms versus manual grading for the detection of referable diabetic retinopathy,* British Journal of Ophthalmology , bjo (2009).

[17] G. S. Scotland, P. McNamee, S. Philip, A. D. Fleming, K. A. Goatman, G. J. Prescott, S. Fonseca, P. F. Sharp, and J. Olson, *Cost-effectiveness of implementing automated grading within the national screening programme for diabetic retinopathy in scotland,* British Journal of Ophthalmology **91**, 1518 (2007).

[18] A. D Fleming, S. Philip, K. A Goatman, G. J Prescott, P. F Sharp, and J. A Olson, *The evidence for automated grading in diabetic retinopathy screening,* Current diabetes reviews **7**, 246 (2011).

[19] K. M. Adal, R. M. Ensing, R. Couvert, P. van Etten, J. P. Martinez, K. A. Vermeer, and L. van Vliet, *A hierarchical coarse-to-fine approach for fundus image registration,* in *Biomedical Image Registration* (Springer, 2014) pp. 93–102.

[20] J. Chen, S. Ausayakhun, S. Ausayakhun, C. Jirawison, C. M. Khouri, T. C. Porco, D. Heiden, J. D. Keenan, and T. P. Margolis, *Comparison of autophotomontage software programs in eyes with cmv retinitis,* Investigative Ophthalmology & Visual Science **52**, 9339 (2011).

[21] B. Polak, W. Hartstra, P. Ringens, and R. Scholten, *Richtlijn'diabetische retinopathie: screening, diagnostiek en behandeling'(herziening),* Nederlands Tijdschrift voor Geneeskunde **152**, 2406 (2008).

[22] P. H. Scanlon, R. Malhotra, R. Greenwood, S. Aldington, C. Foy, M. Flatman, and S. Downes, *Comparison of two reference standards in validating two field mydriatic digital photography as a method of screening for diabetic retinopathy,* British journal of ophthalmology **87**, 1258 (2003).

[23] C. Stellingwerf, P. L. Hardus, and J. M. Hooymans, *Two-field photography can identify patients with vision-threatening diabetic retinopathy: a screening approach in the primary care setting,* Diabetes Care **24**, 2086 (2001).

[24] Commissie Regelgeving en Onderzoek, *The code of conduct for the use of data in health research,* `https://www.federa.org/codes-conduct` (2004), accessed October 15, 2014.

[25] M. Foracchia, E. Grisan, and A. Ruggeri, *Luminosity and contrast normalization in retinal images,* Medical Image Analysis **9**, 179 (2005).

[26] A. A. Mahurkar, M. A. Vivino, B. L. Trus, E. M. Kuehl, M. Datiles, and M. I. Kaiser-Kupfer, *Constructing retinal fundus photomontages. a new computer-based method.* Investigative Ophthalmology & Visual Science **37**, 1675 (1996).

[27] A. Can, C. V. Stewart, B. Roysam, and H. L. Tanenbaum, *A feature-based technique for joint, linear estimation of high-order image-to-mosaic transformations: mosaicing the curved human retina,* IEEE Transactions on pattern analysis and machine intelligence **24**, 412 (2002).

[28] C. V. Stewart, C.-L. Tsai, and B. Roysam, *The dual-bootstrap iterative closest point algorithm with application to retinal image registration,* Medical Imaging, IEEE Transactions on **22**, 1379 (2003).

[29] J. L. Fleiss, *Design and analysis of clinical experiments*, Vol. 73 (John Wiley & Sons, 2011).

[30] P. F. Sharp, J. Olson, F. Strachan, J. Hipwell, A. Ludbrook, M. O'Donnell, S. Wallace, K. Goatman, A. Grant, N. Waugh, *et al.*, *The value of digital imaging in diabetic retinopathy.* Health technology assessment (Winchester, England) **7**, 1 (2003).

4

A quadrature filter approach for registration accuracy assessment of fundus images

This chapter is based on the publication:

K. M. Adal, R. Couvert, D.W.J. Meijer, J. P. Martinez, K. A. Vermeer, and L. J. van Vliet, *A quadrature filter approach for registration accuracy assessment of fundus images*, in International Workshop on Ophthalmic Medical Image Analysis of MICCAI, University of Iowa (2014).

4

Abstract

This paper presents a method to automatically assess the accuracy of image registration. It is applicable to images in which vessels are the main landmarks such as fundus images and angiography. The method simultaneously exploits not only the position, but also the intensity profile across the vasculatures. The accuracy measure is defined as the energy of the odd component of the 1D vessel profile in the difference image divided by the total energy of the corresponding vessels in the constituting images. Scale and orientation-selective quadrature filter banks have been employed to analyze the 1D signal profiles. Subsequently, the relative energy measure has been calibrated such that the measure translates to a spatial misalignment in pixels. The method was validated on a fundus image dataset from a diabetic retinopathy screening program at the Rotterdam Eye Hospital. An evaluation showed that the proposed measure assesses the registration accuracy with a bias of -0.1 pixels and a precision (standard deviation) of 0.9 pixels. The small Fourier footprint of the orientation selective quadrature filters makes the method robust against noise.

4.1. Introduction

Registration of medical images can be defined as a spatial mapping between two or more images in order to relate them for diagnosis, screening, or other clinical purposes. The images to be registered may be acquired from different patients and can be of different imaging modalities. However, they can also come from longitudinal data of the same patient to detect pathologies or to quantify disease progression. Depending on the task at hand one chooses a specific registration method. The state-of-the-art medical image registration techniques are presented in recent surveys [1, 2]. In addition to robustness, a key performance measure of these techniques is accuracy.

In medical images where sparsely distributed blood vessels are the main available landmarks, quantifying the registration accuracy is challenging. The most common way to quantify the registration accuracy of such images is to use the vessels' skeleton in the two images and evaluate the distance between them [3]. However, extracting the skeleton is not a trivial task. Some parts may be missing in one of the images and it is also sensitive to noise. Moreover, it does not provide sub-pixel resolution and results in the loss of all the valuable intensity information. Conventional metrics such as intensity difference, cross correlation, and mutual information fail either to handle differences in contrast or defocus between registered image pairs. In addition, these metrics do not provide the spatial registration error in pixels, which is crucial to determine if the registration result meets the required level of accuracy for a certain application.

Other evaluation approaches include visual inspection of the registration result in an overlay mode, comparing the obtained transformation with the "ground truth" transformation, or testing its transitivity [4, 5]. Visual inspection is a very fast way to find large registration errors. As such, it is useful for determining the robustness of an algorithm and to find outliers, but it is not suited for a quantitative assessment of the accuracy for images with small registration errors. Since a ground truth is often not available, one has to simulate a given transformation to use this method for evaluation. Simulated transformations, however, will only cover those deformations that are part of the model and will therefore miss some of the deformations encountered in practical cases. An alternative is to evaluate the transitivity of the algorithm, but this does not guarantee a high registration accuracy because the registration errors may be correlated. Another major limitation common to all of the above approaches is that they do not provide an objective error measure in pixels.

In this paper, we address the problem of assessing the registration accuracy of images in which the vasculature is the main feature. The proposed method uses the vessels as landmarks to quantify the accuracy of the alignment in the direction perpendicular to the vessels. The presence of vessel branches with various orientations ensures a complete accuracy assessment. Thus, the proposed mismatch measure is a quantity related to the physical displacements occurring across the vessels.

A scale and orientation adaptive quadrature filter bank has been used to decompose the 1D profile perpendicular to a vessel in the difference image into an

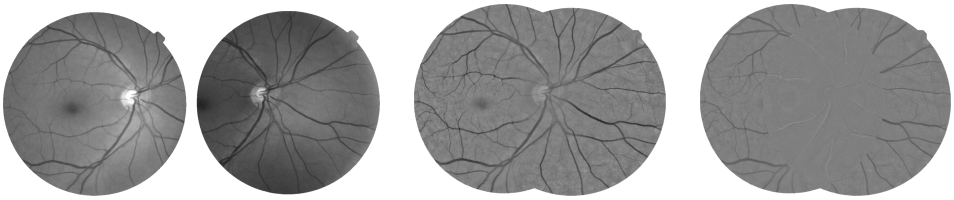


Figure 4.1: Example of registered fundus image pairs. (a) Macula and optic nerve-centered images. (b) Normalized and registered output image. (c) The difference image in the overlap region. Zero difference is depicted in grey.

odd and an even component. The ratio of the energy of the odd component to the total energy of the vessel profile of the two images provides a measure that is directly related to the registration error. This measure is invariant to other disturbing factors due to imaging and illumination artifacts.

We applied our method to assess the accuracy of registered red-free fundus photos acquired for diabetic retinopathy screening. We show that the proposed error measure is strongly related to the spatial registration error in the registered images, thus it can be used as a tool in the longitudinal screening of fundus images for disease progression.

4.2. Material and Method

4.2.1. Material

The proposed algorithm was validated on fundus images obtained from an ongoing diabetic retinopathy screening program at the Rotterdam Eye Hospital. 20 diabetes patients who visited the hospital in two consecutive years for diabetic retinopathy screening were included. During each visit, four images of macula-centered, optic nerve-centered, superior, and temporal regions of the retina were acquired from each eye. For the sake of simplicity, we will use macula and optic nerve-centered images of each patient.

4.2.2. Registration Method

Although the proposed quantitative accuracy assessment method can be applied to evaluate the accuracy of any registration method for images with vessels, we demonstrate it here by applying it to a hierarchical non-rigid fundus image registration approach[6]. The method registers image pairs using intensity as well as structural information of the retinal vasculature after normalization of the green channel for luminosity and contrast variation over the full field of view. The normalized images are registered based on a vasculature-weighted mean square difference (MSD) similarity measure and a multiresolution matching strategy coupled with a hierarchical registration model. Figure 4.1a and 4.1b show an example of individual image pairs and the registered normalized mosaic. Figure 4.1c shows the difference image in the overlap region.

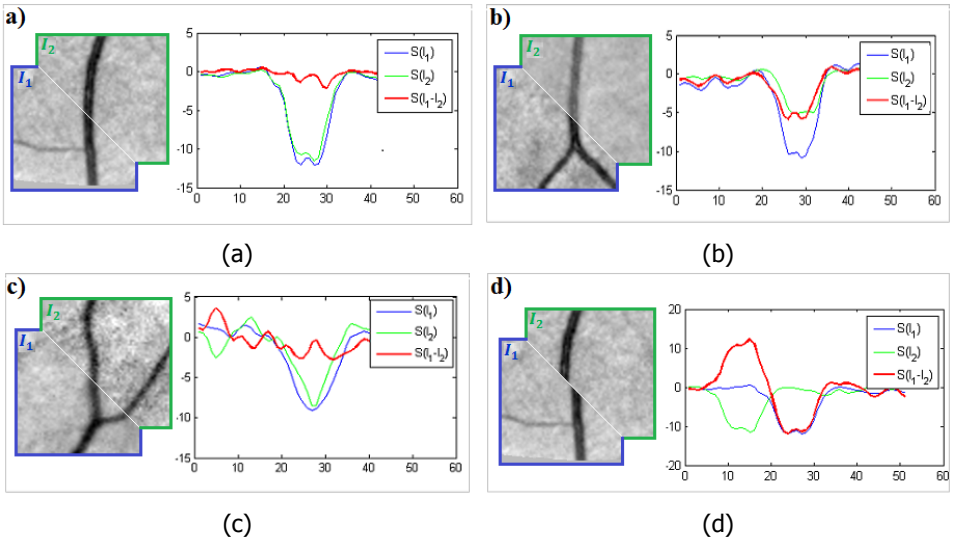


Figure 4.2: Examples of registered image patches with the corresponding vessel profiles and the difference profile. (a-c) Correct alignment. (d) Misalignment.

4.2.3. Registration Accuracy Assessment

When evaluating the registration accuracy of medical images in which the vasculature is the main feature, it is of importance to be able to differentiate vessel misalignments (Fig.4.2d) from other contributions to the MSD such as contrast differences (Fig.4.2b) and (de)focus differences (Fig.4.2c).

In cases where the alignment is correct (Fig.4.2a-4.2c), the profiles of the difference image perpendicular to the aligned vasculatures shows an even signal whereas a misalignment yields an odd (Fig.4.2d) signal. Hence, a new measure called relative vessel misalignment energy (RVME) which exploits this signal property of the difference image is defined to assess the registration accuracy. The RVME measure is expressed as the energy of the odd component of the vessel profile in the difference image divided by the total energy of the two corresponding vessel profiles:

$$RVME = \frac{E\left(S_{odd,\perp}(I_{diff,p})\right)}{E\left(S_{\perp}(I_{1,p_1})\right) + E\left(S_{\perp}(I_{2,p_2})\right)}, \tag{4.1}$$

where $E\left(S_{odd,\perp}(I_{diff,p})\right)$ is the energy of the odd component of the difference image signal profile perpendicular to the vasculatures and centered at a point p. $E\left(S_{\perp}(I_{1,p_1})\right)$ and $E\left(S_{\perp}(I_{2,p_2})\right)$ are the energies of the signal profiles perpendicular to the vasculatures of the registered image pairs I_1 and I_2 centered at p_1 and p_2, respectively.

Quadrature filters

In order to compute the RVME, the odd and even parts of the signal perpendicular to the vessel need to be extracted. A quadrature filter $q(x)$ gives the analytic

representation of a signal that has been filtered by a filter $h(x)$. Such a filter is defined as $q(x) = h(x) + i \cdot \mathcal{H}\big(h(x)\big)$, where $\mathcal{H}$ is the Hilbert transform.

By choosing an even filter $h(x)$, the real and imaginary parts of $q(x)$ allow to differentiate between even and odd profile signals, respectively. Therefore, the analytic representation f of the profile signal S, centered at a given point, becomes:

$$f = S(x) * q(x) = S(x) * \mathrm{Re}\big(q(x)\big) + i \cdot S(x) * \mathrm{Im}\big(q(x)\big) \tag{4.2}$$

If the profile signal is even, then the response to the filter will be real. However, the response to an odd signal profile will be imaginary, which is in agreement with the models in figure 4.2. We can thus use this method to evaluate the RVME.

Orientation space

As the Hilbert transform needs to be applied to the signal in a certain direction, we adopted an orientation space filter bank [7, 8]. The filter bank is composed of rotated versions of an orientation selective quadrature filter. The orientation selectivity and scale selection can be best described in the frequency domain representation of the filter [8]:

$$\Phi(\theta, w) = \underbrace{\exp\left(-\frac{(N\theta)^2}{2\pi^2}\right)}_{\text{angular filter}} \cdot \underbrace{\left(\frac{|w|}{w_c}\right)^{\frac{w_c^2}{b_w^2}} \exp\left(-\frac{w^2 - w_c^2}{2b_w^2}\right)}_{\text{radial filter}}, \tag{4.3}$$

where θ is the orientation angle, w the radial frequency, N the number of filters in the bank which defines the angular resolution, w_c the frequency at which the filter attains its maximum, and b_w the standard deviation of the radial Gaussian. These parameters were set based on the vessels' width and orientation.

The N filtered images are relatively noise-free and the signal in each image represents a specific orientation range, solving the problem of spurious or missing vessel skeleton and simplifying the computation of the RVME. Eq.(1) becomes:

$$RVME = \frac{E\big(\mathrm{Im}(I_{diff} * q_{\hat{\theta}})\big)}{E(I_1 * q_{\hat{\theta}}) + E(I_2 * q_{\hat{\theta}})} \tag{4.4}$$

where $q_{\hat{\theta}}$ is the spatial domain representation of the orientation space filter corresponding to the orientation of the vessel.

Vessel detection and width estimation

In order to match the scale and orientation of the quadrature filters, each vasculature along with its width and orientation must be identified. The vasculature region is first detected from one of the registered images using a multi-scale ($\sigma \in [1, 9]$ pixels) vessel enhancement approach [9] followed by connected component analysis . A vasculature mask is then obtained by keeping objects larger than 2000 pixels, discarding possible noise. This mask is further reduced to a skeleton, and its junctions are removed, leaving only segments of the skeleton along which the registration assessment can be done. It should be noted that even though we used the

skeleton, the assessment can also be done on any point along the vessel segment, solving the requirement of accurate skeleton detection.

Once the skeleton is detected, the width and orientation of the vasculature segments are estimated by exploiting the properties of the local principal curvature computed from the second-order derivatives of the image. Given a scale-normalized Hessian matrix of each pixel $\mathbf{x} = (x, y)$:

$$H(\mathbf{x}; \sigma) = \sigma^2 \begin{bmatrix} I_{xx}(\mathbf{x}; \sigma) & I_{xy}(\mathbf{x}; \sigma) \\ I_{xy}(\mathbf{x}; \sigma) & I_{yy}(\mathbf{x}; \sigma) \end{bmatrix}, \tag{4.5}$$

where each element $I_{..}(\mathbf{x}; \sigma)$ is the convolution of the image at location $\mathbf{x}$ with a second-order Gaussian derivative kernel of scale σ along the specified subscript .

For each vasculature pixel, the eigenvalues λ_1 and λ_2 ($|\lambda_1| \gg |\lambda_2|$) of H correspond to the local (intensity) curvature values across and along a vessel, respectively. The width is approximated by selecting the scale $\hat{\sigma}$ which maximizes the largest principal curvature λ_1:

$$\hat{\sigma} = \arg\max_{\sigma} \lambda_1(\sigma) \tag{4.6}$$

Hence, at each evaluation point, the quadrature filters were tuned to match the vessel width $(2\hat{\sigma})$ and the orientation of the vasculature, determined by the eigenvector corresponding to $\lambda_1(\hat{\sigma})$.

4.3. Experiments and Results

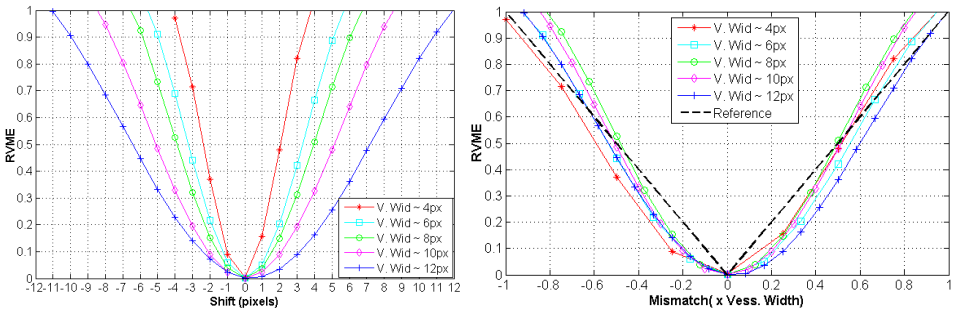
4.3.1. Parameter Optimization

The orientation space filter bank (Eq. 4.3) parameters (N, w_c, b_w) were optimized to match the width and orientation of vessels. The angular resolution N was set to 15 and the optimal values of the remaining parameters are summarized as $w_c = b_w = 0.15\ \hat{\sigma}^{-1}$. These parameters were the same for all fundus data, healthy and diseased.

4.3.2. Evaluation

In order to determine the relation between the RVME measure and the spatial registration accuracy, an evaluation was done on registered image pairs. For each pair, a mismatch was introduced by translating one of the image pairs horizontally (but any other direction would suffice as well) by a known amount before calculating the RVME for vessels running perpendicular to the imposed displacement. The translation was increased by 1 pixel until the vessels were fully misaligned. Figure 4.3a shows the evaluation results stratified by vessel width. Since the orientation selective filters have footprint that runs parallel to vessels, the result of neighboring pixels are correlated, thus evaluation was done at randomly sampled points uniformly distributed across the entire overlap region. Each point on the graph is the average over 50 evaluation points selected from the 20 registered fundus image pairs.

The results show a strong correlation between the RVME measure and the imposed misalignment (registration error). Moreover, figure 4.3b shows the RVME



(a) (b)

Figure 4.3: Registration assessment results. (a) RVME as a function of misalignment in pixels. (b) Correlation between RVME and misalignment as a function of vessel width.

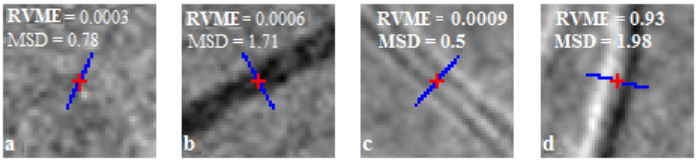


Figure 4.4: Examples of difference image patches. Corresponding MSD (for the entire patch) and RVME values (at the red-cross locations) are shown on each patch.

measure of various vessel widths have an approximately linear correlation, which indicates that it is robust to variation in the vessel width. Hence, given the RVME value and the estimated vessel width, the spatial registration accuracy can be determined in a straightforward manner. This evaluation showed that the proposed measure assesses the registration accuracy with a bias of -0.1 pixels and a precision (standard deviation) of 0.9 pixels.

In the example shown in figure 4.4, even if the registration of two images is perfect, the difference image at the location of blood vessels may show a significant residual signal. For example, in figure 4.4a noise leads to an MSD of 0.78. In figure 4.4b, the contrast difference between correctly aligned vessels leads to an MSD value as high as in case of a clear misalignment, while the RVME remains low. Figure 4.4c shows an example of a well-aligned vessel suffering from a clear difference in (de)focus, resulting in a significant MSD. In contrast, these examples show a very small value for the proposed RVME measure, indicating a very accurate registration. In case of actual registration errors, such as in figure 4.4d, the RVME is close to 1, corresponding to the expected value for a registration error of about the width of a vessel. To evaluate the robustness of the RVME measure to higher noise levels than the noise available in the normalized images which is $\sigma^2_{noise} = 1$, a Gaussian noise was added to each of the registered images. Evaluation results at 5 randomly selected points show that the RVME measure barely changes with the noise in the registered image pairs (Fig.4.5).

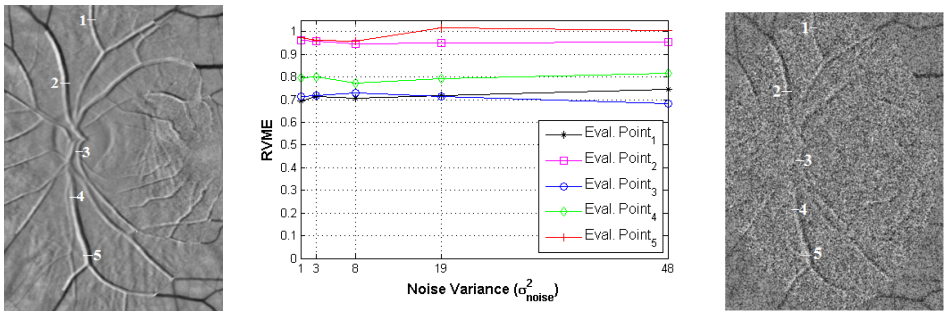


Figure 4.5: Comparison of RVME measures at 5 randomly selected evaluation points on incorrectly registered image pair before (left image, $\sigma^2_{noise} = 1$) and after adding Gaussian noise (right image, $\sigma^2_{noise} = 48$).

4.4. Discussion

In this paper, a new way of quantitatively assessing the registration accuracy of images in which the vasculature provides the main landmarks has been proposed. An accuracy measure (RVME) which exploits the even and odd signal property of the 1D profile across the vessels in the difference image is defined and used to determine the registration accuracy relative to the width of a vessel. The RVME measure is translated to the spatial registration accuracy in pixels by multiplication with the estimated vessel width, enabling an objective and quantitative registration accuracy assessment.

We demonstrated the method by applying it to registered red-free fundus images in order to quantify a misalignment error up to the full width of the widest vessel. Evaluation results showed that the RVME, in contrast with the MSD, does not depend on the intensity variation between registered image pairs, i.e. it is invariant to factors such as contrast, (de)focus, and noise (Fig. 4.4-4.5). It provides an excellent prediction of the imposed displacement (bias of -0.1 pixels and a standard deviation 0.9 pixels) in a controlled experiment.

References

[1] B. Zitova and J. Flusser, *Image registration methods: a survey,* Image and vision computing **21**, 977 (2003).

[2] A. Sotiras, C. Davatzikos, and N. Paragios, *Deformable medical image registration: A survey,* Medical Imaging, IEEE Transactions on **32**, 1153 (2013).

[3] F. Laliberte, L. Gagnon, and Y. Sheng, *Registration and fusion of retinal images-an evaluation study,* Medical Imaging, IEEE Transactions on **22**, 661 (2003).

[4] J. M. Fitzpatrick, *Detecting failure, assessing success,* in *Medical Image Registration* (CRS Press, 2001) pp. 117–139.

[5] G. E. Christensen, X. Geng, J. G. Kuhl, J. Bruss, T. J. Grabowski, I. A. Pirwani, M. W. Vannier, J. S. Allen, and H. Damasio, *Introduction to the non-rigid image registration evaluation project (nirep),* in *WBIR* (Springer, 2006) pp. 128–135.

[6] K. M. Adal, R. M. Ensing, R. Couvert, P. van Etten, J. P. Martinez, K. A. Vermeer, and L. van Vliet, *A hierarchical coarse-to-fine approach for fundus image registration,* in *Biomedical Image Registration* (Springer, 2014) pp. 93–102.

[7] M. van Ginkel, P. Verbeek, and L. J. van Vliet, *Improved orientation selectivity for orientation estimation,* in *Proc. Scan. Conf. on Im. Anal.*, Vol. 1 (1997) pp. 533–537.

[8] F. G. Faas and L. J. van Vliet, *3d-orientation space; filters and sampling,* in *SCIA 2003*, Lecture Notes in Computer Science (Vol. 2749, 2003) pp. 36–42.

[9] A. Frangi, W. Niessen, K. Vincken, and M. Viergever, *Multiscale vessel enhancement filtering,* in *Medical Image Computing and Computer-Assisted Intervention - MICCAI 1998*, Lecture Notes in Computer Science, Vol. 1496 (Springer Berlin Heidelberg, 1998) pp. 130–137.

5

An automated system for the detection and classification of retinal changes due to red lesions in longitudinal fundus images

This chapter is based on the publication:

K. M. Adal, P. G. van Etten, J. P. Martinez, K. W. Rouwen, K. A. Vermeer, and L. J. van Vliet, *An Automated System for the Detection and Classification of Retinal Changes Due to Red Lesions in Longitudinal Fundus Images*, IEEE transactions on bio-medical engineering medical engineering 65, no. 6 (2018): 1382-1390.

5

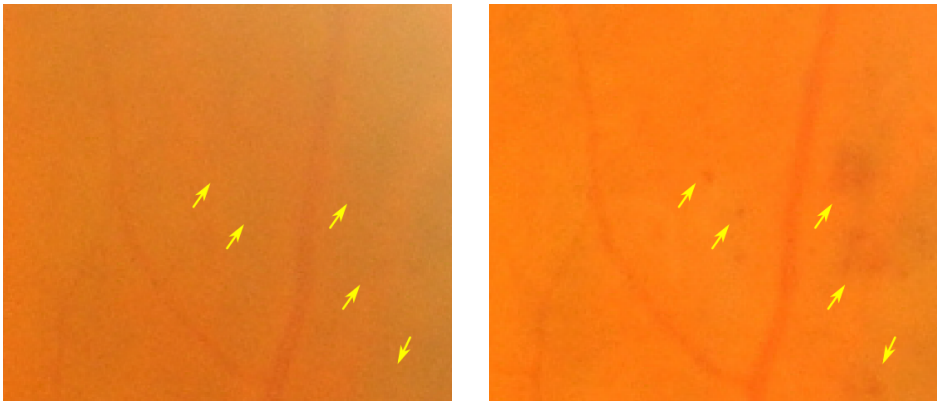
Abstract

People with diabetes mellitus need annual screening to check for the development of diabetic retinopathy (DR). Tracking small retinal changes due to early diabetic retinopathy lesions in longitudinal fundus image sets is challenging due to intra- and inter-visit variability in illumination and image quality, the required high registration accuracy, and the subtle appearance of retinal lesions compared to other retinal features. This paper presents a robust and flexible approach for automated detection of longitudinal retinal changes due to small red lesions by exploiting normalized fundus images that significantly reduce illumination variations and improve the contrast of small retinal features. To detect spatio-temporal retinal changes, the absolute difference between the extremes of the multiscale blobness responses of fundus images from two time-points is proposed as a simple and effective blobness measure. DR related changes are then identified based on several intensity and shape features by a support vector machine classifier. The proposed approach was evaluated in the context of a regular diabetic retinopathy screening program involving subjects ranging from healthy (no retinal lesion) to moderate (with clinically relevant retinal lesions) DR levels. Evaluation shows that the system is able to detect retinal changes due to small red lesions with a sensitivity of 80% *at an average false positive rate of* 1 *and* 2.5 *lesions per eye on small and large fields-of-view of the retina, respectively.*

5.1. Introduction

Diabetic retinopathy is a complication of diabetes mellitus, which progressively damages retinal blood vessels and may result in vision loss and even blindness if not diagnosed and treated adequately. Regular eye examination is necessary for timely detection and treatment of DR at an early stage [1]. The current eye care practice for screening DR involves examination of multiple field fundus images for pathognomonic abnormalities by a trained expert. Depending on the observed retinal abnormalities at the time of the examination, diabetic patients are either scheduled for a follow-up examination or referred to an ophthalmologist for further diagnostic evaluation and possibly treatment. This screening procedure is subjective [2], time consuming and puts a considerable demand on diabetic eye care resources.

Moreover, in addition to examining how far the disease has progressed at the time of examination, the goal of regular DR screening is also to identify patients with a high risk of progression. DR is a progressive disease that results in retinal changes such as the appearance (and the disappearance) of associated lesions such as microaneurysms and hemorrhages (see Fig. 5.1). Recent studies suggest that in addition to the number of lesions at the time of examination, the dynamics of these lesions is useful to monitor progression of DR [3–5]. Automated detection and quantification of retinal changes can thus be an important addition to regular DR screening to objectively assess the disease activity over time for proactively taking appropriate measures. An automated system is also instrumental in patient education, especially in asymptomatic patients. By highlighting and showing DR related retinal changes on a computer display, the patients may have better understanding of their progressing eye condition and the importance of regular checkup and adjustment of their blood sugar level to reduce their risk of developing visual complications.



(a) Baseline image (b) Follow-up image

Figure 5.1: A pair of spatially aligned fundus image patches showing longitudinal retinal change locations (yellow arrows) due to early stage DR lesions between the baseline and follow-up retinal examinations.

Automated detection of longitudinal retinal changes from a series of fundus im-

ages is challenging for several reasons. Firstly, accurate detection and classification of longitudinal retinal changes requires a robust algorithm to discriminate clinically relevant changes from changes caused by illumination variation and noise. Fundus imaging is a process that involves careful manual tuning of fundus camera settings, thus color fundus images often suffer from intra- and inter-image variation in luminosity and contrast (see Fig. 5.2) [6]. Illumination variation coupled with the subtle appearance of early DR lesions renders identifying clinically relevant retinal changes in color fundus images as a very difficult task, even for expert graders.

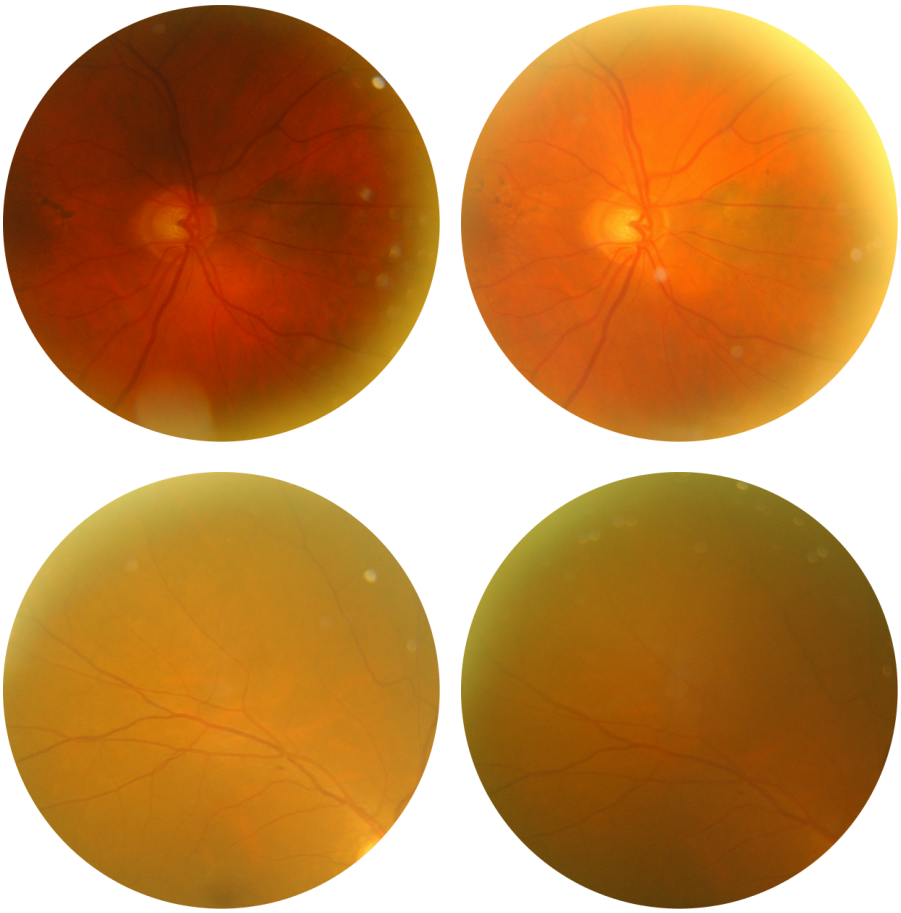


Figure 5.2: Examples of fundus image pairs of the same retina (top and bottom rows) captured one year apart during regular DR screening. Significant variation in illumination and acquisition artefacts can be seen in both image pairs.

Secondly, tracking small retinal features, such as microaneurysms and dot hemorrhages, over time requires very high registration accuracy. In order to correctly register fundus images, the nonlinear spatial deformation caused by the projection of the curved surface of the retina onto a flat imaging plane needs to be accounted

for [7, 8]. This is challenging due to the sparseness of retinal features that can be used for matching and the limited overlapping region between different fundus fields.

Most of the previous studies on computer-aided detection and diagnosis (CAD) systems for DR screening exclusively aimed at analyzing digital fundus images from a single retinal examination [2, 9–22]. A common approach in these studies is to detect early stage retinal abnormalities that are associated with referable DR. Although these CAD systems enable to identify retinal abnormalities for screening DR at the time of examination, they give only limited insight into the activity of the disease since the previous check-up, and thus are not suitable for explicitly monitoring DR.

So far, only a few automated systems have been developed for the detection of longitudinal retinal changes for monitoring DR over time. Examples include systems proposed by Cree et al. [23] and Goatman et al. [24] to detect microaneurysms in longitudinal fluorescein angiogram images. Both systems consist of a method to detect microaneurysms from a region-of-interest centered on the fovea and a registration algorithm to align longitudinal images of the same retina for determining the microaneurysm turnover. Narasimha-Iyer et. al. [25] presented an integrated system for directly detecting and classifying retinal changes from a single-field (macula-centered) color fundus image using a combination of methods for illumination correction, dust removal and segmenting retinal features such as the fovea, optic disc, and blood vessels. The system was later extended to detect changes in vasculature width and appearance/disappearance of lesions [26]. A commercial system also exists for automatically detecting temporal retinal changes due to 'red-dot-like' lesions from a pair of fundus images of the same retina [27, 28].

The main limitation of existing methods for retinal change detection is that they do not address the problems of illumination variation and the space-variant image quality over the entire field-of-view of the retina. Hence, they are not applicable to a large field of the retina, which is required for a comprehensive retinal examination. This is especially a crucial factor when analyzing fundus images in which illumination variation and low image quality affect more than 50% of the field-of-view as shown in Fig. 5.2. This paper addresses these problems and presents a robust and flexible multi-stage approach that is applicable to a wide range of fundus fields for automated detection of longitudinal retinal changes due to microaneurysms and dot hemorrhages (small red lesions).

5.2. Materials and Methods

An overview of the proposed multi-stage approach for automated detection and classification of changes due to small red retinal lesions in longitudinal fundus images is shown in Fig. 5.3. In the first stage, illumination variation is addressed by normalizing the green channel of each color fundus image for luminosity and contrast variation, thereby improving the visibility of retinal features. Then all the baseline and follow-up sets of normalized four-field fundus images are registered into a common coordinate system using a multi-resolution matching strategy coupled to a hierarchical registration model. In the second stage, spatio-temporal

retinal change locations are detected by a novel criterion, blobness measure, based on a multi-scale Laplacian of Gaussian. At the last stage, several local intensity and shape descriptors were extracted from each of the detected change locations and subsequently classified as a change due to a red retinal lesion or no change. Each stage of the proposed approach is described in detail in the following subsections.

5.2.1. Dataset

Data for this study was obtained from a regular DR screening program at the Rotterdam Eye Hospital. Four field (macula-centered, optic nerve-centered, superior, and temporal regions) fundus image sets from 81 diabetic eyes that were acquired for DR screening in 2012 and again in 2013 were used for training (40 eyes) and testing (41 eyes) the proposed approach. All fundus images of 2000×1312 pixels in size were acquired after pupil dilation using a non-mydriatic digital fundus camera (Topcon TRC-NW6S, Tokyo, Japan) with a 45° field of view.

5.2.2. Illumination Normalization and Registration

The green channel of digital fundus images is commonly used in automated fundus image analysis because of its higher contrast between retinal features and the background than the red and blue channels. However, the green channel images show considerable variation in luminosity (brightness) and contrast between retinal structures, both within and between images [6]. In a recent work [29], we have addressed this illumination variation by applying an improved version of the luminosity and contrast normalization technique of Foracchia's et al [6]. The normalization was done by using estimates of the local luminosity and contrast from the intensity distribution of the so-called background retina (i.e., the retina excluding features such as vessels, optic disc, and lesions) and subsequently correcting for their variation over the entire retinal image. In order to take into account missing retinal features, especially around the borders of fundus images, the local luminosity and contrast were estimated based on normalized convolution [29]. This results in a normalized retinal image with a uniform illumination pattern within the entire retinal field and improved visibility of fine retinal details (see Fig. 5.4).

In order to track small retinal features, such as small red lesions, over time, a very high registration accuracy is required. The curved nature of the retinal surface introduces a nonlinear spatial deformation in the process of acquiring fundus images. Therefore, a higher order (quadratic) deformation model is needed for registering fundus images accurately. Over the past decade, several algorithms have been proposed for registration of fundus images [7, 8, 30–34]. In this paper, we used a recently introduced robust fundus image registration method that exploits the normalized intensity as well as the structural information of the retinal vasculature [21]. The method aligns retinal vessels based on a multi-resolution matching strategy coupled to a hierarchical registration model with a deformation model of increasing complexity for robust optimization of a global second-order transformation model. The method was successfully applied to register four-field (macula-centered, optic nerve-centered, superior, and temporal) intra- and inter-visit fundus images that capture different parts of the same retinal surface [35].

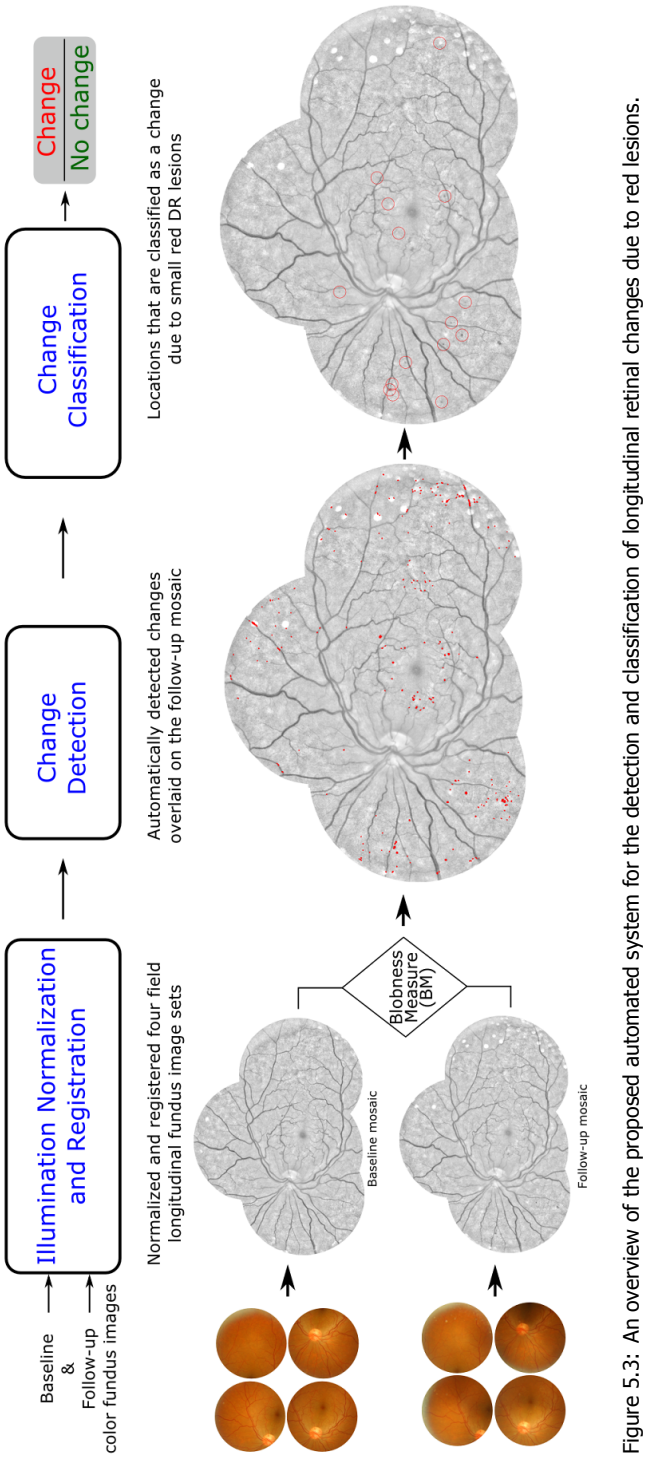


Figure 5.3: An overview of the proposed automated system for the detection and classification of longitudinal retinal changes due to red lesions.

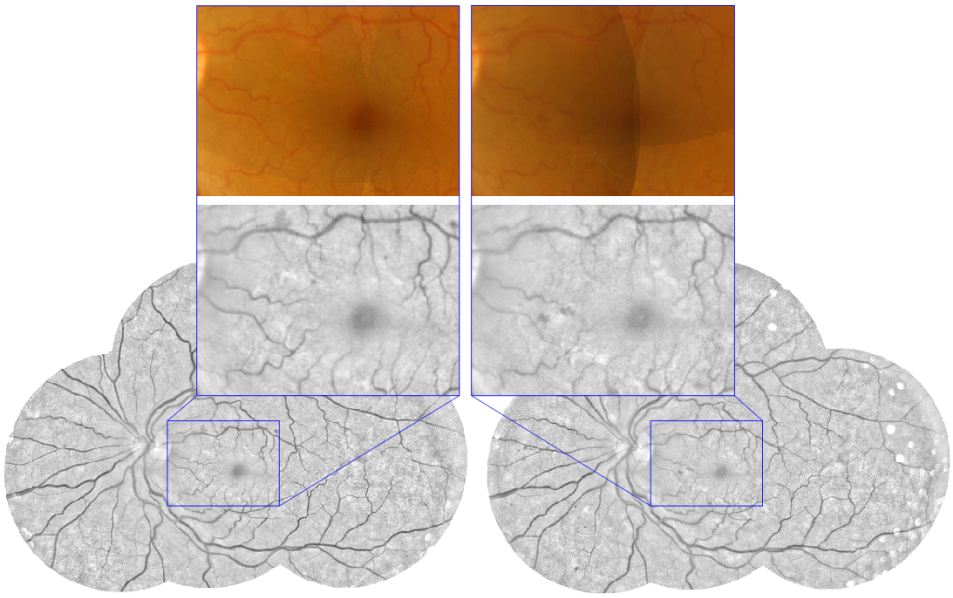


Figure 5.4: Examples of baseline (left) and follow-up (right) normalized fundus mosaics produced by registering four-field fundus images captured from a left eye during regular DR screening. The overlaid color and normalized (enlarged) image patches highlight retinal changes due to DR lesions.

5.2.3. Retinal Change Detection

At early stages, DR is associated with microaneurysms, swellings in small blood vessels that may leak blood into the retina. These lesions commonly appear in color fundus images as small, round dark-red spots (see Fig. 5.4 and 5.5).

Due to the resemblance of these lesions to roundish blobs, the Laplacian of Gaussian (LoG) operator is proposed for detecting them in normalized fundus images. The LoG is sensitive to a certain scale and will therefore provide a maximum response at the scale that matches the size of the object to be detected. In order to take into account the reduction of the LoG response with an increase in σ, the scale-normalized LoG operator is defined as [36]

$$\nabla^2_{norm}(\sigma) = \sigma^2 \nabla^2 G(x, y; \sigma), \tag{5.1}$$

where $\nabla^2 G(x, y; \sigma) = \frac{\partial^2 G(x,y;\sigma)}{\partial x^2} + \frac{\partial^2 G(x,y;\sigma)}{\partial y^2}$ is the LoG operator and $G(x, y; \sigma) = \frac{1}{2\pi\sigma^2} e^{\frac{-(x^2+y^2)}{2\sigma^2}}$ is a 2D Gaussian function of scale σ.

In longitudinal DR screening, the focus is mainly on identifying regions that have changed due to an appearance or disappearance of retinal lesions between DR checkups. Thus, given the two time-point fundus images I_{t_1} and I_{t_2}, spatio-temporal changes are detected by first applying the scale-normalized LoG operator at several scales to each of the two time-point images and then comparing the results. To this end, we propose a blobness measure (BM), which is defined as the

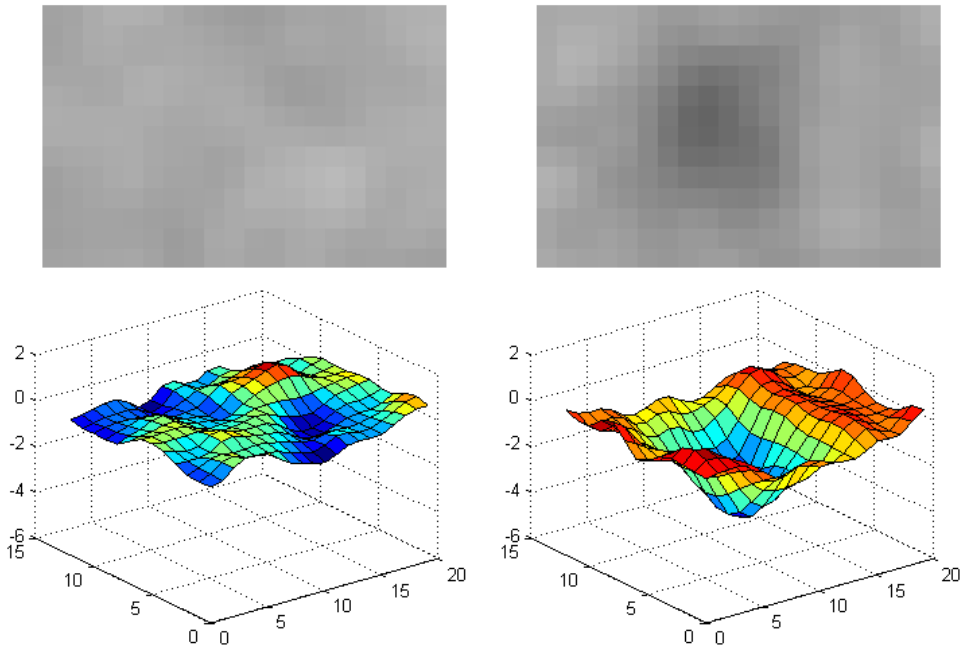


Figure 5.5: A pair of spatially-aligned normalized retinal image patches and their surface topographies before (left) and after (right) a small red DR lesion appears.

absolute difference between the extremes of the multiscale blobness responses of fundus images from two time-points.

$$BM(I_{t_1}, I_{t_2}) = |\max_{\sigma} \nabla^2_{norm}(\sigma) * I_{t_1} - \max_{\sigma} \nabla^2_{norm}(\sigma) * I_{t_2}|. \quad (5.2)$$

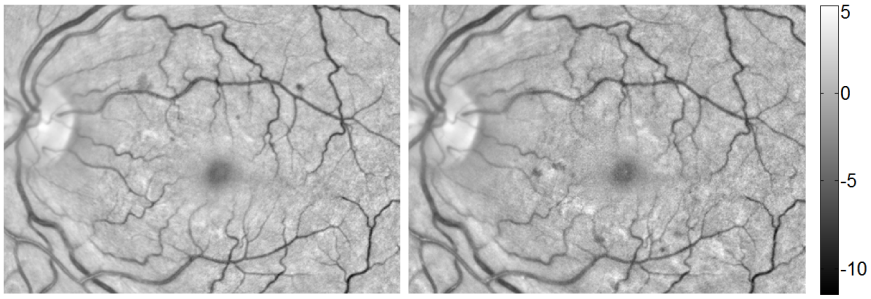
In regions that changed due to an appearance or disappearance of retinal lesions between the two time-points, the BM is expected to be significant (see Fig. 5.6a-5.6c). A candidate change mask (M) is then obtained by thresholding the BM response at θ_{BM} as

$$M = \{(x,y) | BM(I_{t_1}(x,y), I_{t_2}(x,y)) \geq \theta_{BM}\} \quad (5.3)$$

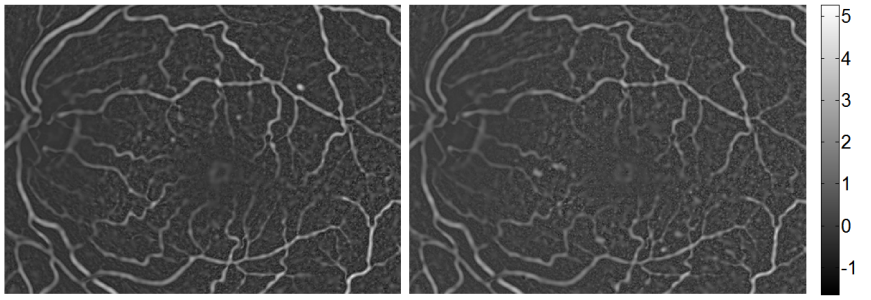
In addition, candidate regions that are smaller than the smallest microaneurysm size (3 pixels or $21\mu m$ in diameter in our dataset) were excluded. Figure 5.6d shows an example change mask extracted from a pair of fundus image patches.

5.2.4. Red Lesion Classification

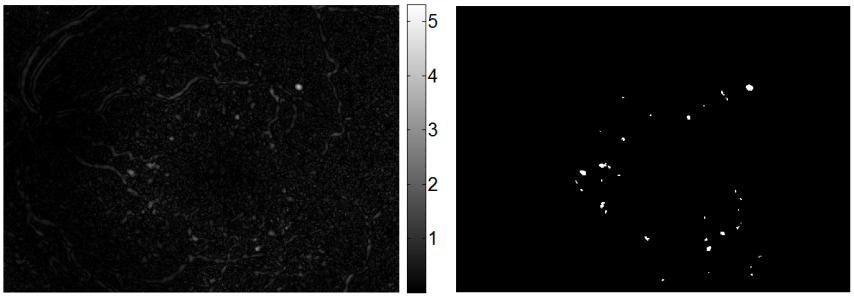
After detecting candidate regions several intensity features, image quality measures, and appearance and shape descriptors were extracted from each candidate region. The complete list of features is



(a) I_{t_1} and I_{t_2}



(b) $\max_\sigma \nabla^2_{norm}(\sigma) * I_{t_1}$ and $\max_\sigma \nabla^2_{norm}(\sigma) * I_{t_2}$



(c) $BM(I_{t_1}, I_{t_2})$ (d) Change mask (M)

Figure 5.6: An example of a pair of spatially-aligned normalized image patches (a), the maximum multi-scale blobness responses (b), and the resulting BM values of each pixel (c). Retinal changes due to appearance or disappearance of lesions appear in the BM image as bright white spots. The white spots in (d) indicate the final candidate change mask locations derived from the BM.

Intensity features (F_{int})

- The red and green channel values of the baseline and follow-up images.
- The normalized intensities of the baseline and follow-up images, I_{t_1} and I_{t_2}.
- The difference in luminosity (L) and contrast (C) between the baseline and follow-up images computed as $\triangle L = |L_{t_1} - L_{t_2}|$ and $\triangle C = \sqrt{C_{t_1}^2 + C_{t_2}^2}$. L and

C are respectively estimated from the instensity distribuition of the retinal background image(excluding vessels, optic disc, and lesions) using the sample mean and standard deviation [29].

- Local power per frequency band of normalized images, which is defined as the extent to which the intensities of retinal regions change rapidly and locally. This is obtained by computing the signal power after band-pass filetring.

A total of 20 intensity features were extracted from each candidate region.

Appearance and shape descriptors (F_{asd})

- The blobness measure (BM) (see eq. 5.2).
- Histogram of oriented gradients (HOG) [37] computed with a cell size of 5×5 pixels and block size of 4×4. This produced 144 features.
- Scale-adapted speeded up robust features (SURF) [38]. 64 SURF features were extracted.
- The local intensity curvatures (eigenvalues) computed from the second-order Gaussian derivatives computed at $\sigma = 2^{i/2}, i \in \{0, 1, 2, 3, 4, 5\}$.

A total of 221 appearance and shape descriptors features were extracted.

Three classifiers, K-nearest neighbor (KNN), random forests (RF), and a support vector machine (SVM) with a radial basis function (RBF) kernel, were independently used to predict the probability that each candidate region is a change due to a red retinal lesion.

5.2.5. Reference Annotation Formation

The reference annotations used for both training and testing the proposed system was gathered from three experts on DR screening (two ophthalmologists and an optometrist). Each of the graders independently annotated the center locations of retinal changes between the baseline and follow-up exam due to small red DR lesions in the fundus mosaics for each eye. The experts were shown both the color and normalized mosaics side-by-side using custom-made software that we developed for lesion annotation. In order to handle inter-grader annotation variability, the reference annotation was defined based on the simultaneous truth and performance level estimation (STAPLE) algorithm [39].

5.3. Experiments and Results

5.3.1. Evaluation metrics

The performance evaluation metrics were the sensitivity (the proportion of correctly detected and classified lesion locations) and average number of false positives per eye. These metrics were computed as

$$Sensitivity = \frac{TP}{TP + FN} \tag{5.4}$$

$$Average\ FPs\ per\ Eye = \frac{FP}{N}, \tag{5.5}$$

where TP is the number of true positives, FN is the number of false negatives, FP is the number of false positives and N is the number of eyes in the test set. A detected location is counted as TP if the distance between its centroid and the closest reference annotation is less than 7 pixels. The evaluation metrics were measured for several threshold levels applied to the (prediction) probability assigned to each of the candidate locations by the classifier and the results are summarized using free-response receiver operating characteristics (FROC) curves.

5.3.2. Parameter Settings

The settings of the two parameters (σ, θ_{BM}) for the retinal change detection algorithm (eq. 5.1-5.3) were optimized based on the training set. The σ values were determined from the relationship between the size (diameter d) of a retinal lesion and the scale at which the lesion response to the scale-normalized LoG operator achieves its maximum. The σ value can be computed as

$$\sigma = \frac{d}{2\sqrt{2}}. \tag{5.6}$$

The estimated diameter of the retinal lesions in our dataset ranges from 3 to 16 pixels ($21\mu m$ to $112\mu m$) and thus scales of $\sigma = 2^{i/2}, i \in \{0, 1, 2, ..., 5\}$ pixels were applied to eq. 5.2. The change detection sensitivity and number of false candidates were optimized by varying a range of values $\theta_{BM} \in \{1.0, 1.1, 1.2, 1.3, 1.4, 1.5\}$. A threshold value of $\theta_{BM} = 1.2$ provided the best compromise between sensitivity (which was set to be at least 96%) and average number of false candidates per eye.

The optimal values for the RBF kernel parameters ($C = 2^{0.5}$, $\gamma = 2^{-9}$) of the SVM classifier and for the number of nearest neighbors ($K = 25$) of the KNN classifier were chosen through 10-fold cross-validation on the training set. A grid-search [40] in combination with cross-validation was used to test various (C, γ) pairs. The RF classifier parameters (the number of trees, the number of randomly selected features for splitting) were set based on the out-of-bag (OOB) data error estimate [41]. During each bootstrap the RF classifier sets aside about one-third of the training samples as OOB data and this data is not used for constructing a tree; therefore, it is used internally as a validation set to estimate the classification error.

5.3.3. Evaluations

We evaluated the proposed approach for the detection and classification of retinal changes due to red lesions from longitudinal retinal mosaics on both a large and a small field of view of the retina on the test set. The evaluation on a large retinal field was done using the four field fundus mosaics, which consists of the macula, optic-disc, superior, and temporal retinal regions. The number of retinal changes (appearances and disappearances) for each eye in the training and testing set is shown in figure 5.7. The total number of retinal changes in the training and testing set were 174 and 164, respectively. The evaluation on a small retinal field was done

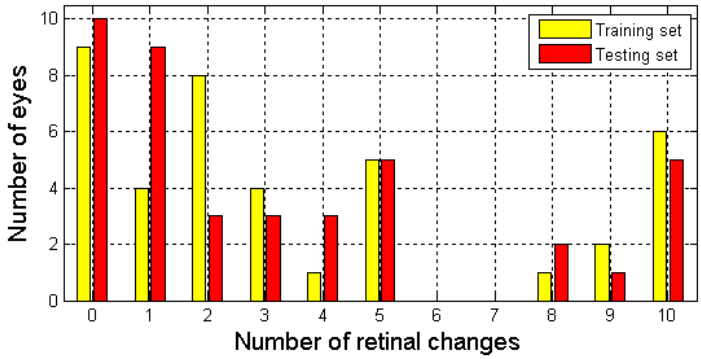


Figure 5.7: Distribution of the number of retinal changes in the training and testing sets.

based on the macula-centered fundus images due to their clinical significance. In addition, the performance of each of the two feature types (intensity vs. appearance and shape) paired with each of the three classifiers (KNN, RF, SVM) was evaluated on the large field fundus mosaics.

5.3.4. Results

Figure 5.8 shows the FROC curves for the systems with various classifier and feature combinations. The blue horizontal line indicates the retinal change detection sensitivity (98%) of the proposed approach on the testing set, which acts as an upper bound on the sensitivity of the whole system. The results show that for all tested classifiers the appearance and shape descriptors produced a much better classification result than the intensity features. Moreover, the performance of the classifiers increased when the combined set of features was used. The SVM classifier performed best among the three classifiers and achieved a sensitivity of 80% at an average false positive rate of 2.5 per eye.

The system that performed best on the large field retinal mosaics (SVM with F_{int}+ F_{asd}) was retrained and used to detect retinal changes from a small field of the retina centered on the macula. The performance of the proposed approach is shown in the FROC curve in figure 5.9. The sensitivity of the retinal change detection algorithm was 97% on macula-centered fundus images. The overall system achieves a sensitivity of 80% at an average false positive rate of 1 per eye.

For each eye in the test set, we visually inspected and analyzed those locations that were detected by our approach but not defined as a DR related change in the reference annotations and thus were counted as false alarms. These locations were found to be in either of the following categories:

- **Dark spots** that resemble tiny red retinal lesions on either the baseline or follow-up mosaics.
- **Retinal vessels** that were affected by illumination artefacts on either the baseline or follow-up mosaics.

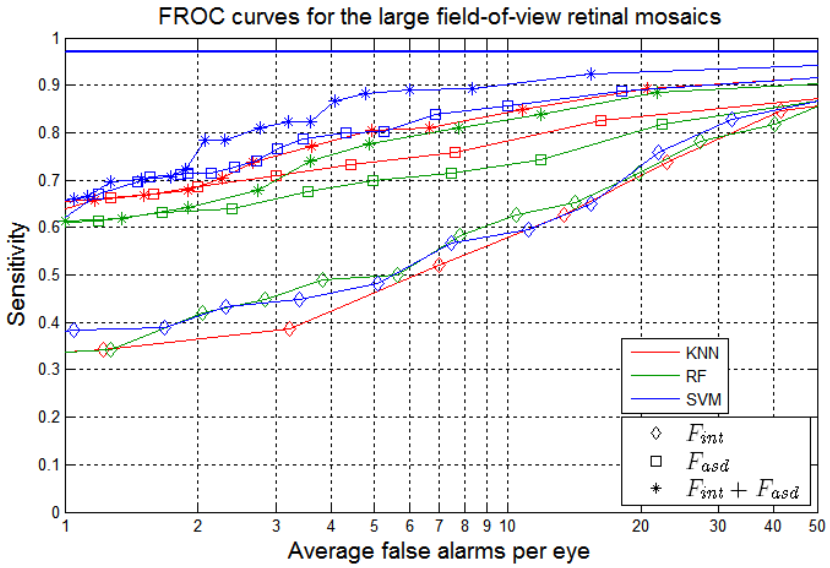


Figure 5.8: FROC curves of the proposed systems for the detection and classification of longitudinal retinal changes due to small red lesion applied to large field-of-view retinal mosaics.

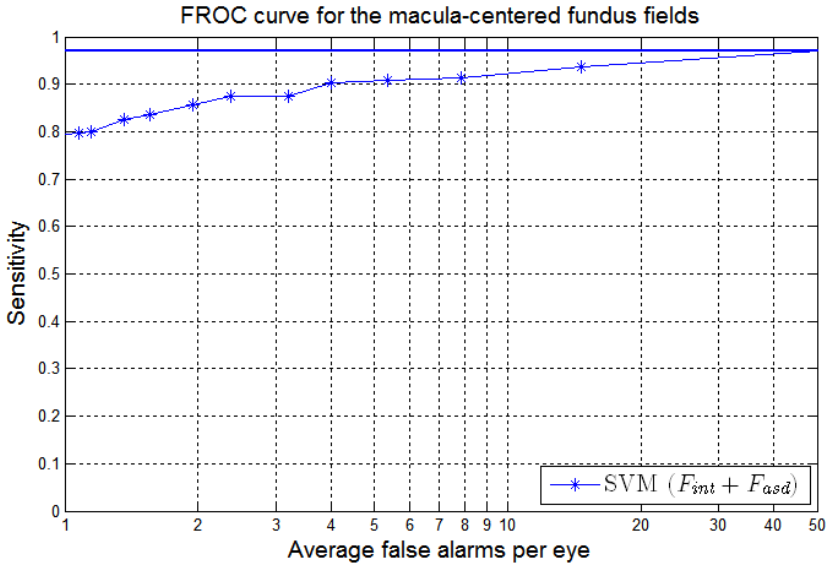


Figure 5.9: FROC curve of the proposed system for the detection and classification of longitudinal retinal changes due to small red lesion applied to macula-centered fundus image pairs.

- **Noisy regions** that have a low signal-to-noise (SNR) ratio.

Figure 5.10 shows the number of false alarms in each of the three categories for

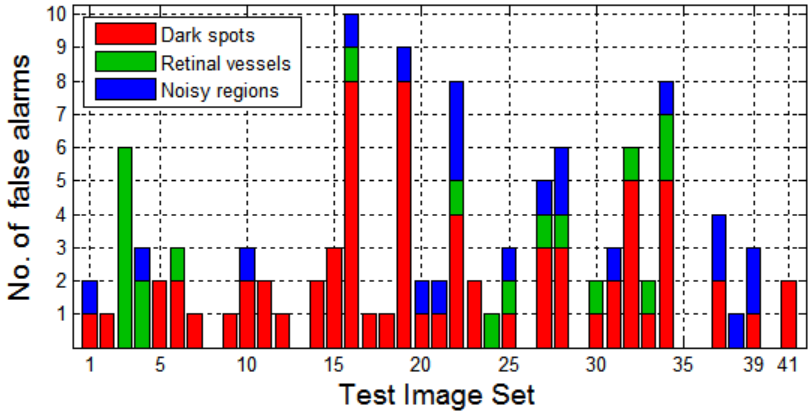


Figure 5.10: Histograms depicting the number of false alarms for each category.

the 41 test eyes at 0.80 sensitivity. Overall, the results show that dark spots that resemble small red lesions caused 63% of the false alarms. The remaining false alarms were detected either on retinal vessels (18%) or in regions with a low SNR (19%). An example from each of the three categories of false alarms are shown in figure 5.11.

5.4. Discussion and Conclusion

In this paper, we have presented a robust and flexible multi-stage approach for tracking retinal changes due to small red DR lesions such as microaneurysms and dot hemorrhages in longitudinal fundus images. The system was applied to both small and large retinal fields of 81 diabetic eyes. Robustness to intra and inter-image illumination variation was achieved by exploiting fundus images that are normalized for luminosity and contrast over the entire field of view. The improvement in the visibility and contrast of especially small retinal features in the normalized fundus images enabled our approach to track subtle retinal changes, including those that are visually difficult to detect on the color fundus images. A simple and effective criterion for blobness (BM) was defined for detecting spatio-temporal retinal change locations from longitudinal normalized fundus images. The BM can also be easily adapted to other related problems for the detection and tracking of small round objects in a series of registered longitudinal images.

The proposed approach was evaluated in the context of a regular diabetic retinopathy screening program involving subjects ranging from healthy (no retinal lesion) to moderate (with clinically relevant retinal lesions) DR levels. Evaluation was done on both a large field-of-view fundus mosaics, which consisted of the macula, optic nerve, temporal, and superior fields, and a small field-of-view of the retina consisting only of the macula-centered fields. The results show that the system was able to detect retinal changes due to small DR lesions with a sensitivity of 80% from large field fundus mosaics and small field fundus images at an average false posi-

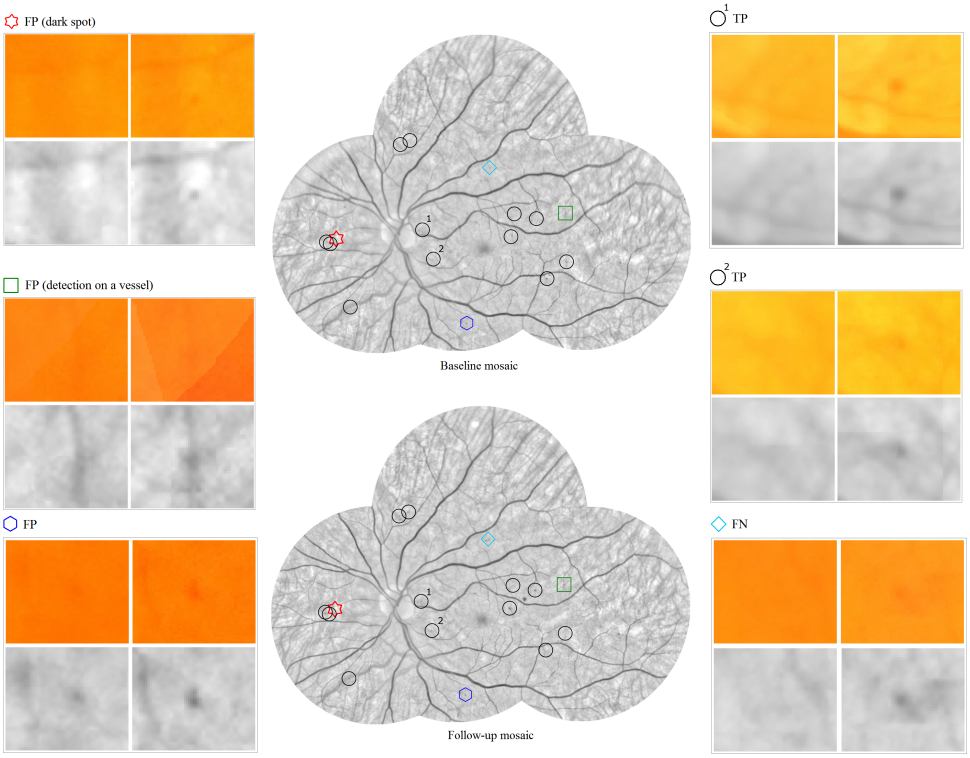


Figure 5.11: An example of the output of the proposed change detection and classification approach at a sensitivity of 80%. The zoomed in color and normalized image patches for the baseline (left) and follow-up (right) indicated by different shapes represent the true positives (TPs), false positives (FPs) and false negative (FN) locations.

tive rate of 2.5 and 1, respectively. In contrast to the small fields, the higher false alarm rate in the large field fundus mosaics is mainly caused by the lower image quality and the presence of significant illumination artefacts, such as white spots (see Fig. 5.4), on the temporal and superior fundus fields.

Visual inspection of the false alarms suggests that most of them were very similar in appearance and shape to small red DR lesions and thus may well be true positives that were erroneously not included in the reference annotation. Indeed, 49% of the detected dark spot locations were small red lesions that were also annotated by one of the graders. It should also be noted that color fundus images are routinely used by eye care experts in DR screening. Therefore, introducing the normalized images during annotation can help the experts see subtle DR related retinal changes, although their annotations might have been biased towards the color images.

The proposed approach could also be applied to determine the red DR lesion count of individual fundus images provided that reference images with known retinal conditions are available. Automated detection of red DR lesions from single time

point images can be very difficult due to the subtle nature of most of the lesions and limited number of lesion pixels. On a publicly available Retinopathy Online Challenge fundus image dataset [42], the top ranking method for automated detection of red lesions from individual images achieved a sensitivity of 53% at an average of 2 false alarms per image [16]. By incorporating reference images and analyzing spatio-temporal change locations, our approach could be applied to detect and determine red lesion count with a higher sensitivity.

Automated detection and quantification of longitudinal retinal changes can be an important addition to regular DR screening. The detected retinal changes can be used for making objective and quantitative analysis of DR progression as well as for more efficient human grading and patient education by highlighting DR related changes since the previous visit.

References

[1] Early Treatment Diabetic Retinopathy Study Research Group, *Early photocoagulation for diabetic retinopathy: ETDRS report number 9,* Ophthalmology **98**, 766 (1991).

[2] M. D. Abràmoff, M. Niemeijer, M. S. Suttorp-Schulten, M. A. Viergever, S. R. Russell, and B. Van Ginneken, *Evaluation of a system for automatic detection of diabetic retinopathy from color fundus photographs in a large population of patients with diabetes,* Diabetes care **31**, 193 (2008).

[3] S. Nunes, I. Pires, A. Rosa, L. Duarte, R. Bernardes, and J. Cunha-Vaz, *Microaneurysm turnover is a biomarker for diabetic retinopathy progression to clinically significant macular edema: findings for type 2 diabetics with nonproliferative retinopathy,* Ophthalmologica **223**, 292 (2009).

[4] C. Haritoglou, M. Kernt, A. Neubauer, J. Gerss, C. M. Oliveira, A. Kampik, and M. Ulbig, *Microaneurysm formation rate as a predictive marker for progression to clinically significant macular edema in nonproliferative diabetic retinopathy,* Retina **34**, 157 (2014).

[5] J. Cunha-Vaz, L. Ribeiro, and C. Lobo, *Phenotypes and biomarkers of diabetic retinopathy,* Progress in retinal and eye research **41**, 90 (2014).

[6] M. Foracchia, E. Grisan, and A. Ruggeri, *Luminosity and contrast normalization in retinal images,* Medical Image Analysis **9**, 179 (2005).

[7] A. A. Mahurkar, M. A. Vivino, B. L. Trus, E. M. Kuehl, M. Datiles, and M. I. Kaiser-Kupfer, *Constructing retinal fundus photomontages. a new computer-based method.* Investigative Ophthalmology & Visual Science **37**, 1675 (1996).

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

[9] M. Niemeijer, B. Van Ginneken, J. Staal, M. S. Suttorp-Schulten, and M. D. Abràmoff, *Automatic detection of red lesions in digital color fundus photographs,* IEEE Transactions on medical imaging **24**, 584 (2005).

[10] T. Walter, P. Massin, A. Erginay, R. Ordonez, C. Jeulin, and J.-C. Klein, *Automatic detection of microaneurysms in color fundus images,* Medical image analysis **11**, 555 (2007).

[11] M. Niemeijer, M. D. Abramoff, and B. van Ginneken, *Information fusion for diabetic retinopathy cad in digital color fundus photographs,* IEEE Transactions on Medical Imaging **28**, 775 (2009).

[12] A. D. Fleming, K. A. Goatman, S. Philip, G. J. Williams, G. J. Prescott, G. S. Scotland, P. McNamee, G. P. Leese, W. N. Wykes, P. F. Sharp, *et al.*, *The role of haemorrhage and exudate detection in automated grading of diabetic retinopathy,* British Journal of Ophthalmology **94**, 706 (2010).

[13] M. Abràmoff, M. Garvin, and M. Sonka, *Retinal imaging and image analysis,* Biomedical Engineering, IEEE Reviews in **3**, 169 (2010).

[14] M. Niemeijer, B. Van Ginneken, M. J. Cree, A. Mizutani, G. Quellec, C. I. Sánchez, B. Zhang, R. Hornero, M. Lamard, C. Muramatsu, *et al.*, *Retinopathy online challenge: automatic detection of microaneurysms in digital color fundus photographs,* IEEE Transactions on medical imaging **29**, 185 (2010).

[15] G. Quellec, S. R. Russell, and M. D. Abràmoff, *Optimal filter framework for automated, instantaneous detection of lesions in retinal images,* IEEE Transactions on medical imaging **30**, 523 (2011).

[16] B. Antal and A. Hajdu, *An ensemble-based system for microaneurysm detection and diabetic retinopathy grading,* IEEE Transactions on biomedical engineering **59**, 1720 (2012).

[17] A. Rocha, T. Carvalho, H. F. Jelinek, S. Goldenstein, and J. Wainer, *Points of interest and visual dictionaries for automatic retinal lesion detection,* IEEE Transactions on biomedical engineering **59**, 2244 (2012).

[18] L. Giancardo, F. Meriaudeau, T. P. Karnowski, Y. Li, S. Garg, K. W. Tobin, and E. Chaum, *Exudate-based diabetic macular edema detection in fundus images using publicly available datasets,* Medical image analysis **16**, 216 (2012).

[19] L. Tang, M. Niemeijer, J. M. Reinhardt, M. K. Garvin, and M. D. Abràmoff, *Splat feature classification with application to retinal hemorrhage detection in fundus images,* IEEE Transactions on Medical Imaging **32**, 364 (2013).

[20] M. D. Abràmoff, J. C. Folk, D. P. Han, J. D. Walker, D. F. Williams, S. R. Russell, P. Massin, B. Cochener, P. Gain, L. Tang, *et al.*, *Automated analysis of retinal images for detection of referable diabetic retinopathy,* JAMA ophthalmology **131**, 351 (2013).

[21] K. M. Adal, D. Sidibé, S. Ali, E. Chaum, T. P. Karnowski, and F. Mériaudeau, *Automated detection of microaneurysms using scale-adapted blob analysis and semi-supervised learning,* Computer methods and programs in biomedicine **114**, 1 (2014).

[22] I. N. Figueiredo, S. Kumar, C. M. Oliveira, J. D. Ramos, and B. Engquist, *Automated lesion detectors in retinal fundus images,* Computers in biology and medicine **66**, 47 (2015).

[23] M. Cree, J. Olson, K. McHardy, P. Sharp, and J. Forrester, *A fully automated comparative microaneurysm digital detection system.* Eye (London, England) **11**, 622 (1996).

[24] K. A. Goatman, M. J. Cree, J. A. Olson, J. V. Forrester, and P. F. Sharp, *Automated measurement of microaneurysm turnover,* Investigative ophthalmology & visual science **44**, 5335 (2003).

[25] H. Narasimha-Iyer, A. Can, B. Roysam, C. V. Stewart, H. L. Tanenbaum, A. Majerovics, and H. Singh, *Robust detection and classification of longitudinal changes in color retinal fundus images for monitoring diabetic retinopathy,* Biomedical Engineering, IEEE Transactions on **53**, 1084 (2006).

[26] H. Narasimha-Iyer, A. Can, B. Roysam, H. L. Tanenbaum, and A. Majerovics, *Integrated analysis of vascular and nonvascular changes from color retinal fundus image sequences,* IEEE Transactions on biomedical engineering **54**, 1436 (2007).

[27] C. M. Oliveira, L. M. Cristóvão, M. L. Ribeiro, and J. Abreu, *Improved automated screening of diabetic retinopathy,* Ophthalmologica **226**, 191 (2011).

[28] J. Cunha-Vaz, R. Bernardes, T. Santos, C. Oliveira, C. Lobo, I. Pires, and L. Ribeiro, *Computer-aided detection of diabetic retinopathy progression,* in *Digital Teleretinal Screening* (Springer, 2012) pp. 59–66.

[29] K. M. Adal, R. M. Ensing, R. Couvert, P. van Etten, J. P. Martinez, K. A. Vermeer, and L. van Vliet, *A hierarchical coarse-to-fine approach for fundus image registration,* in *Biomedical Image Registration* (Springer, 2014) pp. 93–102.

[30] N. Ritter, R. Owens, J. Cooper, R. H. Eikelboom, and P. P. Van Saarloos, *Registration of stereo and temporal images of the retina,* Medical Imaging, IEEE Transactions on **18**, 404 (1999).

[31] G. K. Matsopoulos, N. A. Mouravliansky, K. K. Delibasis, and K. S. Nikita, *Automatic retinal image registration scheme using global optimization techniques,* Information Technology in Biomedicine, IEEE Transactions on **3**, 47 (1999).

[32] C. V. Stewart, C.-L. Tsai, and B. Roysam, *The dual-bootstrap iterative closest point algorithm with application to retinal image registration,* Medical Imaging, IEEE Transactions on **22**, 1379 (2003).

[33] T. Chanwimaluang, G. Fan, and S. R. Fransen, *Hybrid retinal image registration,* Information Technology in Biomedicine, IEEE Transactions on **10**, 129 (2006).

[34] I. N. Figueiredo, S. Moura, J. S. Neves, L. Pinto, S. Kumar, C. M. Oliveira, and J. D. Ramos, *Automated retina identification based on multiscale elastic registration,* Computers in biology and medicine **79**, 130 (2016).

[35] K. M. Adal, P. G. van Etten, J. P. Martinez, L. J. van Vliet, and K. A. Vermeer, *Accuracy assessment of intra-and intervisit fundus image registration for diabetic retinopathy screening,* Investigative Ophthalmology & Visual Science **56**, 1805 (2015).

[36] T. Lindeberg, *Feature detection with automatic scale selection,* International journal of computer vision **30**, 79 (1998).

[37] N. Dalal and B. Triggs, *Histograms of oriented gradients for human detection,* in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, Vol. 1 (IEEE, 2005) pp. 886–893.

[38] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, *Speeded-up robust features (surf),* Computer vision and image understanding **110**, 346 (2008).

[39] S. K. Warfield, K. H. Zou, and W. M. Wells, *Simultaneous truth and performance level estimation (staple): an algorithm for the validation of image segmentation,* IEEE Transactions on medical imaging **23**, 903 (2004).

[40] C.-C. Chang and C.-J. Lin, *LIBSVM: A library for support vector machines,* ACM Transactions on Intelligent Systems and Technology **2**, 27:1 (2011), software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

[41] L. Breiman, *Random forests,* Machine learning **45**, 5 (2001).

[42] M. Niemeijer, B. Van Ginneken, M. J. Cree, A. Mizutani, G. Quellec, C. I. Sánchez, B. Zhang, R. Hornero, M. Lamard, C. Muramatsu, *et al.*, *Retinopathy online challenge: automatic detection of microaneurysms in digital color fundus photographs,* IEEE Transactions on medical imaging **29**, 185 (2010).

6

Investigation of correlation between DR development and longitudinal retinal changes in diabetic eyes

6

Abstract

*Diabetic retinopathy (DR) is a complication of diabetes mellitus, which progressively damages small retinal blood vessels and may result in vision loss and even blindness if not diagnosed and treated timely and adequately. To monitor DR progression, diabetes mellitus patients need to undergo periodic screening. In this study, the correlation between DR development and the retinal surface change over time is investigated. To quantify retinal changes due to formation and disappearance of small round red lesions, a recently developed fully automated longitudinal fundus image analysis system has been applied. The number of detected changes are then analyzed among referable and non-referable DR classes derived from DR levels assigned by trained clainical expert. Evaluation was done on 199 diabetic eyes with a non-referable DR stage at baseline. The data of the yearly examinations of these eyes for four more years were used in this investigation. During this period, 178 eyes (*89.4%*) remained at non-referable DR stage whereas 21 eyes (*10.6%*) progressed to referable DR. The number of detected retinal changes have been analyzed and compared between referable and non-referable diabetic eyes. Evaluation results show a statistically significant association between the number of detected retinal changes and classification of referable DR ($p = 0.008$). The odds ratio for developing referable DR was* 1.15 *($p = 0.018$, 95% CI:* $1.02 - 1.29$*) for each detected change per year. These results suggest that the red lesion activity between successive retinal examinations can be used as a potential biomarker in building a prediction model for future referal.*

6.1. Introduction

Diabetic retinopathy (DR) is a complication of diabetes mellitus (DM), which progressively damages small retinal blood vessels and may result in vision loss and even blindness if not diagnosed and treated timely and adequately. Because of long-standing DM, changes in the capillaries occur leading to thickening of basal membrane, loss of pericytes and death of the endothelium cells, thereby resulting in microaneurysms and capillary non-perfusion [1]. As the disease progresses, more retinal blood vessels that nourish the retina start to close or leak, resulting in formation of exudates (lipids and proteins) or small hemorrhages. This results in a reduction of oxygen supplies to the retina and stimulates the formation of neovascularization, i.e, the growth of new blood vessels that are fragile. The newly formed blood vessels may easily leak and cause bleedings and vitreous hemorrhages. These blood vessels could also become fibrotic over time and lead to tractional retinal detachments.

Because the risk of retinal damage due to DR progression over time and the latency between development of DR and early DR symptoms, diabetic patients need to undergo periodic screening. This involves the acquisition of fundus photos covering multiple retinal fields and manual inspection by trained eye care experts to assess the onset and progression of DR. A trained expert reviews the fundus photos and, based on the observed retinal condition, decides either to refer the patient to a retinal specialist for further retinal examination and possibly treatment or to schedule another screening appointment. The current DR screening practice is subjective and time-consuming. This, together with the global rise in the diabetic population puts a significant burden on the accessibility of eye care resources and the efficiency of DR screening programs.

DR is a progressive disease that results in the appearance (and the disappearance) of associated retinal lesions, such as microaneurysms and dot hemorrhages, at an early stage. An objective assessment of retinal changes over time is thus crucial to explore whether red lesion turnover, which is defined as the total number of progressed and regressed red lesions since the previous examination, is predictive for future referral. Automated analysis of longitudinal series of fundus photos plays an important role to objectively assess spatio-temporal retinal change over time. This provides complementary prognostic information in addition to the observed retinal change at the time of the retinal examination. Recent studies suggest that microaneurysm turnover, which is the rate of formation [2, 3] or both the formation and disappearance [4] of microaneurysms, can be a useful biomarker to predict DR progression to clinically significant macular edema. A semi-automatic method was used to analyze the macula centered (field-2 as defined in ETDRS protocol [5]) field of the retina [2, 3].

This study aims to explore whether the red lesion turnover can be used as predictor to identify patients with a high risk of progression to referable DR. This was done by evaluating and comparing the red lesion turnover between two groups of diabetic eyes: referable and non-referable. Both groups were diagnosed with non-referable DR at baseline and they were annually screened for DR for four more years. The fundus photos gathered from all five retinal examinations were auto-

matically analyzed to detect, classify and quantify retinal changes due to early DR lesions between consecutive retinal examinations. To this end, recently developed and published methods for automated fundus image registration [6, 7] and detection and classification of retinal changes in longitudinal four-field fundus images [8] were applied.

6.2. Methods

6.2.1. Data description

The dataset used in this retrospective study was gathered from diabetic patients who were enrolled in the DR screening program of the Rotterdam Eye Hospital in The Netherlands. Four fundus fields consisting of macula-centered, optic nerve-centered, superior, and temporal regions of the retinal surface were gathered from both eyes of 125 diabetic patients that have been screened for DR for 5 consecutive years between 2007 and 2013. Each fundus image was 2000×1312 pixels in size and acquired after pupil dilation (one drop of tropicamide 0.5%) using a non-mydriatic digital funds camera (Topcon TRC-NW6S, Tokyo, Japan) with a 45° field-of-view.

This study adhered to the applicable code of conduct for the reuse of data in health research [9]. After exporting the fundus images from the clinical image storage system, all data was anonymized prior to further processing.

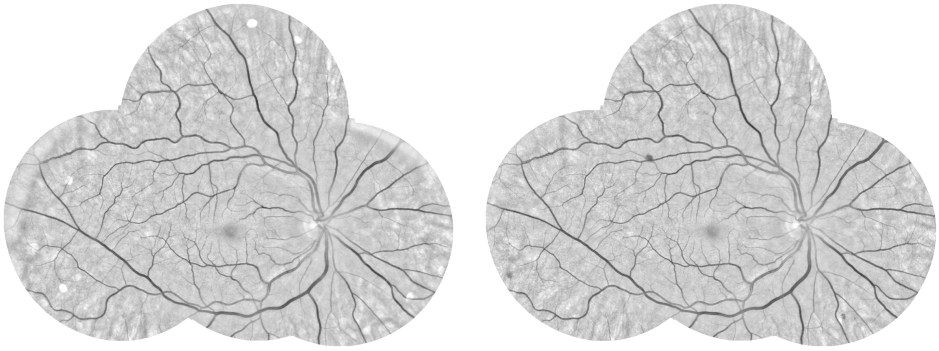
6.2.2. Automated longitudinal fundus image analysis

The longitudinal fundus images were analyzed using a fully automated system. For each eye, all the four field fundus images acquired at five consecutive retinal examination are normalized for illumination variation and registered to a common coordinate system using state-of-the art registration approach [6, 7], resulting in a single large field-of-view mosaic of the retinal surface for each examination. Figure 6.1 shows examples of such longitudinal fundus mosaics for the first two DR examinations of an eye. The number of red lesion related retinal changes between consecutive examinations is extracted using an automated system for the detection and classification of retinal changes in longitudinal fundus images [8].

6.2.3. Longitudinal fundus mosaic grading

For each eye, a trained retinal expert graded each of the mosaics based on the International Clinical Diabetic Retinopathy (ICDR) disease severity scale. The ICDR disease severity scale defines five levels for DR: none, mild, moderate, severe, and proliferative (see Table 6.1) [10, 11]. Both the color and normalized mosaics were used for grading.

The grades in each of the five examinations are translated to non-referable and referable DR. Non-referable DR was defined as no apparent retinopathy or mild NPDR, whereas referable DR was defined as moderate or one of the more severe levels (shaded in red in Table 6.1). The eyes are then categorized into referable and non-referable groups based on whether they were considered referable at any stage during the five year follow-up. Eyes in which non-referable DR was detected in all



(a) Baseline mosaic. (b) Follow-up mosaic.

Figure 6.1: Examples of normalized fundus mosaics of produced from baseline (a) and follow-up examination (b) of an eye.

Table 6.1: The International Clinical Diabetic Retinopathy (ICDR) disease severity scale. Non-referable and referable DR classes are shaded in green and red, respectively.

Disease severity level	**Findings observable upon dilated ophthalmoscopy**
No apparent retinopathy	No abnormalities
Mild non-proliferative DR (NPDR)	Microaneurysms only
Moderate NPDR	More than just microaneurysms but less than Severe NPDR
Severe NPDR	Any of the following: more than 20 intraretinal hemorrhages in each of 4 quadrants, definite venous beading in 2+ quadrants, prominent IRMA in 1+ quadrant
Proliferative DR	One or more of the following: neovascularization, vitreous/preretinal hemorrhage

the five exminations were grouped as non-referable and eyes in which referable DR was detected in any of the five examinations were grouped as referable.

6.2.4. Statistical analysis

The association between the red lesion turnover between consecutive examinations and classification of referable DR was analyzed and compared between referable and non-referable eye groups. The analysis was done in groups of eyes that were non-referable at baseline and later either progressed to referable DR level or remained as non-referable during follow up. The number of detected retinal changes were analyzed in the two groups and the significance of the relationship was tested by Fisher's exact test. In addition, the significance of the red lesion turnover as a predictor for future referal was quantified using odds ratio (OR) derived from logistic regression. A *p*-value below 0.05 is considered to be statistically significant.

6.3. Results

A total of 5000 color fundus images acquired from a total of 250 diabetic eyes screened for DR during a five year period were processed to create longitudinal fundus mosaics. Eyes for which any of the four fields were missing or of poor quality (41 eyes, 16.4%) at any of the five examinations were excluded from the automated processing and expert grading. Out of 209 eyes that were graded by an expert, 10 eyes (4.8%) were already at moderate or severe DR levels (referable stage) at baseline and hence excluded from further analysis. Of the remaining eyes, a total of 178 eyes (89.4%) were graded as no DR or mild NPDR (non-referable stage) in all the follow-up examinations whereas 21 eyes (10.6%) progressed from the non-referable to the referable stage during one of the follow-up examinations.

Figure 6.2 shows the average number of detected retinal changes between successive DR examinations. The result shows that on average the red lesion turnover in referable eyes is higher compared to the non-referable eyes (y-intercept difference of 1.62 changes per year, *p*-value = 0.048). In addition, the red lesion turnover in both groups increases with time, suggesting that the risk of accumlated damage to the retina and hence DR progression increases with the duration of diabetes.

In figure 6.3, the frequencies and cumulative distributions of the number of detected retinal changes at the time of referral in the referable eyes group is shown along with the detected changes between successive retinal examinations in non-referable eyes group. The results show a significant correlation between the number of retinal changes and classification of referable DR (Fisher's exact test, *p*-value = 0.002). It was observed that a few referable eyes (3 out of 21, 14.3%) progressed into moderate DR due to the presence of either non-round or relatively large hemorrhages with no apparent retinal change due to small red lesions in previous DR examinations. Since the automated system was designed to detect changes due to small red lesions, no retinal change was found in those eyes, thereby resulting in the extreme left tail in the referable eye distribution shown in figure 6.3. On the other hand, the right tail in the distribution of non-referable eye group is due to the presence of mild NPDR (non-referable DR) cases in which microaneurysms only are

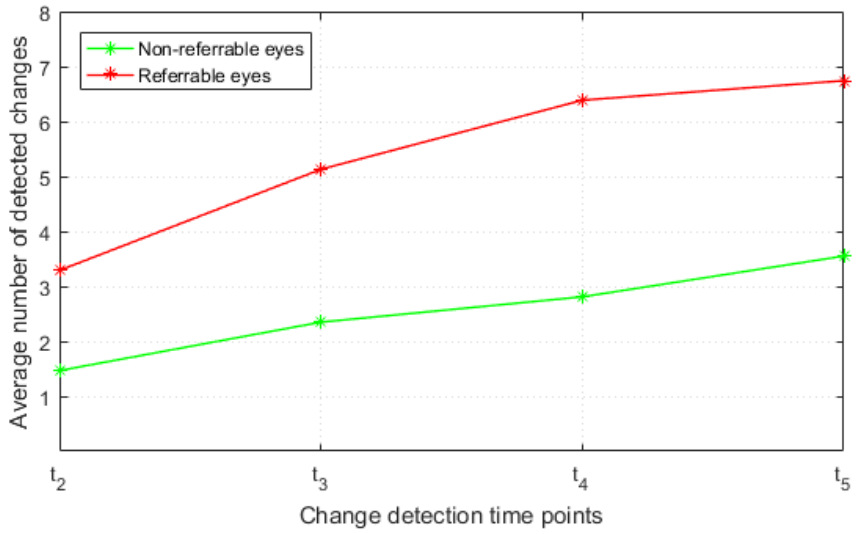


Figure 6.2: Average number of retinal changes between consecutive retinal examinations done at t_{i-1} and t_i.

detected.

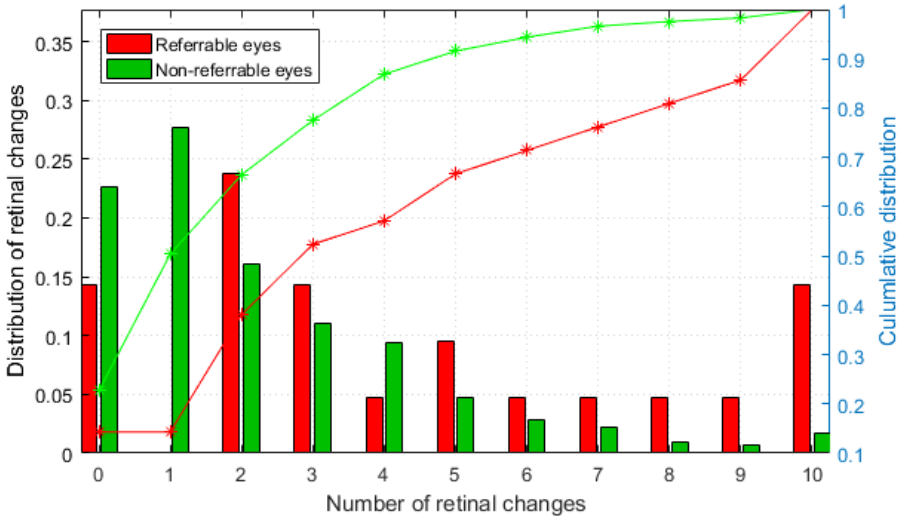


Figure 6.3: Histogram of the detected retinal changes and the corresponding cumulative distribuition up to the referal time point.

In order to analyze the difference in the red lesion turnover between the two diabetic eye groups, the number of retinal changes one year prior to the year of referral

in the referable eyes group is analzyed and compared with the number of detected retinal changes up to and including the fourth examination of the non-referable eyes group. The changes detected between the fourth and fifth examinations were discarded as we cannot warrant that these eyes would not be referred in one year time. Figure 6.4 shows the resulting frequencies and cumulative distribuitions. The results show that the dynamics of red lesions in the retina is strongly associated with the classification of referable DR (Fisher's exact test, *p*-value = 0.008). The trends in both figures 6.3 and 6.4 show that progression to referable DR is often accompanied by an increase in the red lesion turnover. This association enables the use of the red lesion turnover as a biomarker in building a prediction model for future referal.

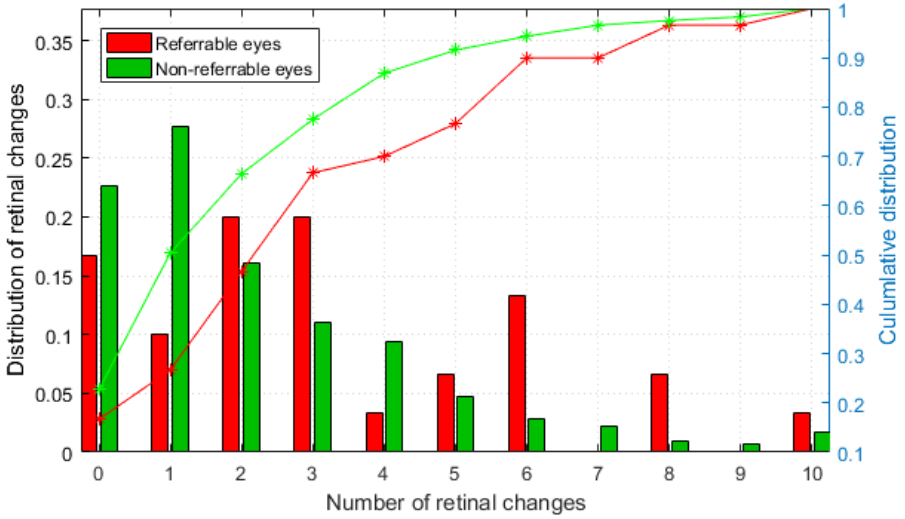


Figure 6.4: Histogram of the detected retinal changes and the corresponding cumulative distribuition prior to referal.

The fitted parameters of the logistic regression model for the prediction of referable DR are summarized in table in 6.2. The result shows that effect of the red lesion turnover on the classification of referable DR was significant (*p*-value = 0.018). Thus, the odds ratio of developing referable DR vs non-referable DR was $e^{0.142} = 1.15$ (95% CI: $1.02 - 1.29$) per unit (lesion per year) increase in red lesion turnover. Given no detected retinal change, the odds of developing referable DR was $e^{-3.262} = 0.04$. It should be noted that although this simple prediction model shows only group level differences and a more rigorous model that includes other DR risk factors needs to be constructed to predict the likelihood of future referral for individual patients.

Table 6.2: The estimated coefficients of the logistic regression model for referable DR prediction.

Coefficients	Estimate	Std. Error	p-Value
Intercept	-3.262	0.268	$< 2 \times 10^{-16}$
Red lesion turnover	0.142	0.060	0.018

6.4. Discussion and Conclusion

In this study, we retrospectively investigated the correlation between retinal changes due to small red retinal lesions and the classification of referable DR in diabetic eyes that have been regularly screened for DR. A recently developed fully automated longitudinal fundus image analysis approach has been applied to quantify the number of retinal changes due to early DR related lesions and analyzed along with the DR levels assigned by a trained expert. A total of 199 diabetic eyes of a non-referable DR stage at baseline were followed for four consecutive years. During this period, 178 eyes (89.4%) remained at non-referable DR stage whereas 21 eyes (10.6%) progressed to referable DR. The number of detected retinal changes between successive retinal examintations show that the red lesion activity and hence the risk of retinal damage increases with the duration of diabetes. Moreover, a statistically significanct association was found between the red lesion turnover and classification of referable DR ($OR = 1.15, p = 0.018$). For each observed change per year, a 15% increase in the risk of developing referable DR was found. These results suggest that the red lesion turnover can be used as a predictor of future risk for referable DR development.

The current DR screening practice involves periodic retinal examination and the grading is based on the presence and detection of lesions at the time of the retinal examintation. Patients who did not develop referable DR are followed up with periodic examinations. By predicting the risk of developing referable DR within a given time based on previously acquired data, an efficient and personalized diabetic eye care can be facilitated, ensuring a reduction in associated health care expenses. Recent studies suggest that the dynamics of retinal change due to microaneurysms may be used as a biomarker to predict diabetic patients that are likely to develop clinically significant macular edema [2–4].

For better understanding of DR activity over time, a longer follow-up period is useful. This study includes a short follow-up period, resulting in a relatively low number of referable cases which might not adequately represent the underlying distribuition of red lesion activities. Previous studies suggested that referable DR prevalence is strongly associated with the duration of diabetes [12, 13]; therefore, extending the follow-up period can be beneficial to get a representative sample. Another limitation of this study is that the automated system was designed to detect only retinal changes due to small round red lesions. Therefore, larger and non-round hemorrhages, exudates, cotton-wool spots and intraretinal microvascular abnormalities (IRMA) which are indirect signs of vascular hyperpermeability and capillary closure or retinal ischemia were not detected and investigated in this study.

In conclusion, this exploratory study showed that the number of retinal changes

due to small red lesions between successive retinal examinations is significantly correlated with the risk of developing referable DR. The red lesion turnover can be exploited in future work as biomarker in constructing a prediction model for DR progression. In addition to the red lesion turnover, the duration of diabetes and age of diabetic patients play a key role in DR progression [12, 14] and hence need to be taken into account.

References

[1] D. G. Cogan, D. Toussaint, and T. Kuwabara, *Retinal vascular patterns: IV. Diabetic retinopathy,* Archives of Ophthalmology **66**, 366 (1961).

[2] S. Nunes, I. Pires, A. Rosa, L. Duarte, R. Bernardes, and J. Cunha-Vaz, *Microaneurysm turnover is a biomarker for diabetic retinopathy progression to clinically significant macular edema: findings for type 2 diabetics with nonproliferative retinopathy,* Ophthalmologica **223**, 292 (2009).

[3] C. Haritoglou, M. Kernt, A. Neubauer, J. Gerss, C. M. Oliveira, A. Kampik, and M. Ulbig, *Microaneurysm formation rate as a predictive marker for progression to clinically significant macular edema in nonproliferative diabetic retinopathy,* Retina **34**, 157 (2014).

[4] M. L. Ribeiro, S. G. Nunes, and J. G. Cunha-Vaz, *Microaneurysm turnover at the macula predicts risk of development of clinically significant macular edema in persons with mild nonproliferative diabetic retinopathy,* Diabetes Care **36**, 1254 (2013).

[5] Early Treatment Diabetic Retinopathy Study Research Group and others, *Grading diabetic retinopathy from stereoscopic color fundus photographs—an extension of the modified airlie house classification: ETDRS report number 10,* Ophthalmology **98**, 786 (1991).

[6] K. M. Adal, R. M. Ensing, R. Couvert, P. van Etten, J. P. Martinez, K. A. Vermeer, and L. van Vliet, *A hierarchical coarse-to-fine approach for fundus image registration,* in *Biomedical Image Registration* (Springer, 2014) pp. 93–102.

[7] K. M. Adal, P. G. van Etten, J. P. Martinez, L. J. van Vliet, and K. A. Vermeer, *Accuracy assessment of intra-and intervisit fundus image registration for diabetic retinopathy screening,* Investigative Ophthalmology & Visual Science **56**, 1805 (2015).

[8] K. M. Adal, J. Martinez, K. Rouwen, K. Vermeer, *et al.*, *An automated system for the detection and classification of retinal changes due to red lesions in longitudinal fundus images.* IEEE Transactions on bio-medical engineering **65**, 1382 (2018).

[9] Commissie Regelgeving en Onderzoek, *The code of conduct for the use of data in health research,* `https://www.federa.org/codes-conduct` (2004), accessed October 15, 2014.

[10] C. Wilkinson, F. L. Ferris III, R. E. Klein, P. P. Lee, C. D. Agardh, M. Davis, D. Dills, A. Kampik, R. Pararajasegaram, J. T. Verdaguer, *et al.*, *Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales,* Ophthalmology **110**, 1677 (2003).

[11] American Academy of Ophthalmology, *International clinical diabetic retinopathy disease severity scale, detailed table,* http://www.icoph.org (2002), accessed August 10, 2017.

[12] R. Klein, B. E. Klein, S. E. Moss, M. D. Davis, and D. L. DeMets, *The Wisconsin Epidemiologic Study of Diabetic Retinopathy: III. Prevalence and risk of diabetic retinopathy when age at diagnosis is 30 or more years,* Archives of Ophthalmology **102**, 527 (1984).

[13] J. W. Yau, S. L. Rogers, R. Kawasaki, E. L. Lamoureux, J. W. Kowalski, T. Bek, S.-J. Chen, J. M. Dekker, A. Fletcher, J. Grauslund, *et al.*, *Global prevalence and major risk factors of diabetic retinopathy,* Diabetes Care **35**, 556 (2012).

[14] Early Treatment Diabetic Retinopathy Study Research Group and others, *Fundus photographic risk factors for progression of diabetic retinopathy: ETDRS report number 12,* Ophthalmology **98**, 823 (1991).

7

Conclusion

Diabetic retinopathy (DR) is a complication of diabetes mellitus, which progressively damages retinal blood vessels and may result in vision loss and even blindness if not diagnosed and treated timely and adequately. Because the risk of retinal damage due to DR progression over time, diabetic patients need to undergo periodic screening. This involves acquisition of fundus photos covering multiple retinal fields and manual inspection by trained eye care experts to assess onset and/or progression of DR. Manual inspection of fundus photos to objectively assess DR development is time consuming and difficult due to, for example, differences in the included retinal surface and deformation and changes in illumination. In this thesis, a fully automated approach for analyzing longitudinal fundus photos for assessing DR progression is presented. The presented approach involves illumination normalization to improve the visibility of retinal features, mosaicking multiple field fundus photos acquired during intra- and inter-visits, and detection and quantification of longitudinal retinal changes to objectively assess DR development. Evaluation done in the context of regular diabetic retinopathy screening on diabetic patients ranging from non-referable DR (no DR or mild NPDR) to referable DR (moderate or severe DR levels) show that the proposed approach can be a useful addition to regular DR screening to effectively and efficiently assess DR development.

7.1. Technical contribuitions

In **Chapter 2**, a robust hierarchical coarse-to-fine approach for registering intra- and inter-visit fundus images was presented. Firstly, illumination variation between fundus images is addressed by using higher-order normalized convolution to estimate the local luminosity and contrast from the intensity distribution of the so-called background retina (which excludes features such as vessels, optic disc, and lesions) and subsequently correcting for their variation over the entire retinal image. Secondly, the normalized intensity as well as the structural information of the retinal vasculature are exploited to spatially align fundus images from multiple fields of the retina based on the vasculature-weighted mean squared difference (MSD) of the normalized images.

To prevent the registration algorithm from getting trapped in a local minimum, we employed a multi-scale approach coupled with a deformation model of progressive complexity to estimate the parameters of a global second-order spatial transformation model parameters. In addition, the initializing of each scale was set based on the results of the previous scale. Extensive evaluation on multiple field fundus images show that the proposed approach can successfully register fundus images pairs which have an overlap region as low as 14%. Moreover, the registration algorithm was robust to illumination variation and uneven distribution of retinal features within the fundus fields.

In **Chapter 3**, a qualitative accuracy assessment was done on fundus mosaics produced by the proposed registration algorithm and compared with two top-ranked state-of-the-art commercial available fundus mosaicking programs: i2k Retina (DualAlign LLC, Clifton Park, NY) and OIS AutoMontage (OIS, Sacramento, CA). Evaluation was done on four-field (macula-centered, optic nerve-centered, superior, and temporal) fundus images of 70 diabetic patients who visited the Rotterdam Eye Hos-

pital in 2012 and 2013 for a diabetic retinopathy screening program. The full evaluation results showed that the proposed approach produced intra- and inter-visit fundus mosaics with higher registration accuracy than i2k Retina in both intra-visit (odds ratio (OR) = 2.5, $p = 10^{-5}$) and inter-visit ($OR = 2.2, p = 0.006$) mosaics. State of the art clinical systems regularly fail at producing mosaics without registration artifacts. We have shown that it is possible to get high quality intra- and inter-visit mosaics in 65% and 49% of the cases, respectively, where the existing system fails.

In **Chapter 4**, an automated method for a quantitative approach for assessing the registration accuracy of fundus image pairs is presented. The method automatically assesses the registration accuracy of fundus image pairs exploiting the intensity profiles across the vasculature and their difference in the registered images. A new accuracy measure, relative vessel misalignment energy (RVME), which exploits the even and odd signal property of the 1D profile across the vessels in the difference image, is introduced. Our evaluation results show that the RVME behaved as expected, as it was invariant to contract, (de)focus, and noise. The RVME measure could be translated to the spatial registration accuracy in pixels, with a bias of -0.1 pixels and a precision (standard deviation) of 0.9 pixels, by multiplication with the estimated vessel width. The RVME measure can thus be used as a robust measure to objectively quantify registration accuracy between retinal vessels as low as 4 pixels in diameter.

In **Chapter 5**, a robust and flexible multi-stage approach is presented for the detection and classification of longitudinal retinal changes due to early DR related retinal lesions such as microaneurysms and dot hemorrhages. To detect spatio-temporal retinal changes, the absolute difference between the extremes of the multiscale blobness responses of fundus images from two time-points is proposed as a simple and effective blobness measure. The detected retinal changes are then classified as red lesion or non-red lesion related based on several intensity and shape features and supervised classifier. Evaluation on image sets acquired during regular DR screening program involving subjects ranging from healthy to moderate DR levels shows that the proposed approach can detect red lesion related retinal changes with a sensitivity of 80% at an average false positive rate of 1 and 2.5 lesions per eye on small and large fields-of-view of the retina, respectively. The results suggest that, by incorporating reference images and analyzing longitudinal retinal changes, the proposed approach can be used to determine red lesion count with a much higher sensitivity compared to state-of-the art methods for red lesion count based on single time-point images. The proposed approach for automated detection of longitudinal retinal changes can thus be an important addition to regular DR screening for an objective and quantitative analysis of DR progression.

In **Appendix A**, a deep convolutional neural network (CNN) has been trained for the detection of spatio-temporal retinal changes from illumination normalized fundus images. Evaluation and comparison with two other CNNs trained separately on color and green channel fundus images shows that the CNN trained on normalized images has a higher accuracy than the other two CNNs in detecting retinal changes. The results suggest that illumination normalization greatly facilitates the

performance of CNNs to quickly and effectively learn retinal change signatures.

7.2. Clinical applications

In **Chapter 6**, the correlation between retinal changes due to small red retinal lesions and the classification of referable DR is explored retrospectively in diabetic eyes that have been screened yearly for DR. To quantify red lesion turnover, the fully automated longitudinal fundus image analysis approach presented in **Chapter 5** has been applied. Evaluation was done on 199 diabetic eyes that have been at non-referable DR stage at baseline and were followed for four consecutive years. The red lesion turnover between successive retinal examinations were analyzed and compared between 178 eyes (89.4%) that remained at non-referable DR stage and 21 eyes (10.6%) that progressed to referable DR. The results show that the development of referable DR is significantly associated with red lesion turnover (odds ratio of 1.15, $p = 0.018$, 95% CI: $1.02 - 1.29$). The red lesion turnover can thus be used as a biomarker for predicting future risk of referable DR development.

Although the methods introduced in the previous chapters play a key role in achieving a fully automated computer aided diagnosis (CAD) system for assessing DR progression, the intermediate results are also useful in clinical practice. Among the key results that have potential for clinical applications are:

- **Illumination normalization:** The enhanced visibility of retinal features in normalized fundus images can simplify fundus image grading by clinical experts. However, whether normalized images can give a higher screening sensitivity without compromising specificity is currently being explored in a joint collaboration between Rotterdam Ophthalmic institute and NIHR Biomedical Research Centre at Moorfields Eye Hospital. Preliminary results showed that normalized fundus images improve the visibility of retinal features and hence facilitates better detection of red DR lesions [1].

- **Fundus mosaicking:** The ETDRS standard for DR detection and classification requires stereoscopic color fundus from 7 standard fields to cover a large field-of-view of the retina [2]. Manual processing of these individual fields to construct a large retinal field is time consuming and resource demanding and hence a limiting factor in clinical practice. The robust automated mosaicking algorithm introduced in this thesis can be useful to construct high quality mosaic of the retina, thereby enabling clinicians to do a comprehensive retinal examination efficiently. The algorithm can also be applied to align retinal images from multiple examinations, thereby facilitating analysis of temporal retinal changes.

- **Automated detection of retinal changes:** The CAD system for the detection and classification of longitudinal retinal changes could also be useful in analyzing retinal images acquired during regular DR screening. The system can help to detect, objectively asses and highlight the retinal changes since the previous examination, allowing eye care experts to get insight about the disease activity over time for a comprehensive DR grading.

7.3. General discussion and future directions

The main goal of this thesis was to develop a fully automated CAD system that can be used to objectively assess DR progression in longitudinal series of fundus images. To this end, several methods are proposed to address challenges such as intra- and inter-visit illumination variation, fundus image registration, and detection and classification of longitudinal retinal changes due to red DR lesions. The CAD system was subsequently used to investigate the association between red lesion turnover and classification of referable DR. The statistically signification association results suggest that the red lesion turnover between successive retinal exams can be used as a biomarker to identify patients with high risk of progression to referable DR.

Areas for future investigation include developing a rigorous risk prediction model for referable DR development. Given the red lesion turnover from previous retinal examinations, predicting the likelihood of developing a referable DR within a given time could be instrumental for health care programs to provide efficient diabetic eye care service. Other DR risk factors such as age, duration of diabetes, and blood sugar level could also be useful in accurately determining the risk and thus need to be considered.

References

[1] K. Adal, K. W. Rouwen, T. Peto, P. G. van Etten, J. P. Martinez, L. J. van Vliet, and K. A. Vermeer, *Detection of lesions and severity grading for diabetic retinopathy in normalized fundus images,* Investigative Ophthalmology & Visual Science **58**, 691 (2017).

[2] Early Treatment Diabetic Retinopathy Study Research Group and others, *Grading diabetic retinopathy from stereoscopic color fundus photographs—an extension of the modified airlie house classification: ETDRS report number 10,* Ophthalmology **98**, 786 (1991).

A

Detection of retinal changes from illumination normalized fundus images using convolutional neural networks

This chapter is based on the publication:

K. M. Adal, P. G. van Etten, J. P. Martinez, K. W. Rouwen, K. A. Vermeer, and L. J. van Vliet, *Detection of Retinal Changes from Illumination Normalized Fundus Images using Convolutional Neural Networks*. In Medical Imaging 2017: Computer-Aided Diagnosis (Vol. 10134, p. 101341N). International Society for Optics and Photonics.

Abstract

Automated detection and quantification of spatio-temporal retinal changes is an important step to objectively assess disease progression and treatment effects for dynamic retinal diseases such as diabetic retinopathy (DR). However, detecting retinal changes caused by early DR lesions such as microaneurysms and dot hemorrhages from longitudinal pairs of fundus images is challenging due to intra and inter-image illumination variation between fundus images. This paper explores a method for automated detection of retinal changes from illumination normalized fundus images using a deep convolutional neural network (CNN), and compares its performance with two other CNNs trained separately on color and green channel fundus images. Illumination variation was addressed by correcting for the variability in the luminosity and contrast estimated from a large scale retinal regions. The CNN models were trained and evaluated on image patches extracted from a registered fundus image set collected from 51 *diabetic eyes that were screened at two different time-points. The results show that using normalized images yield better performance than color and green channel images, suggesting that illumination normalization greatly facilitates CNNs to quickly and correctly learn distinctive local image features of DR related retinal changes.*

A.1. Introduction

Diabetic retinopathy (DR) is a complication of diabetes mellitus, which progressively damages retinal blood vessels and results in vision loss and even blindness if not diagnosed and treated adequately. Regular eye examination is necessary for the detection and treatment of DR at an early stage [1]. The current eye care practice for regular DR screening involves manual examination of multiple fundus photos and is resource demanding, subjective, and does not exploit images from previous retinal exam for progression assessment. Automated detection of longitudinal DR related changes provides an objective measure of retinal abnormalities over time and enables clinicians to objectively assess DR progression.

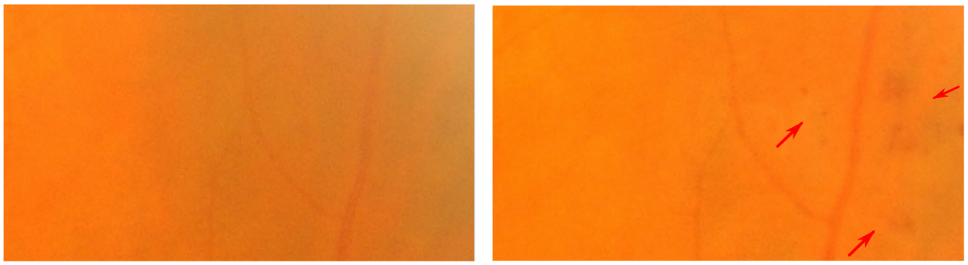
DR progression is accompanied by retinal changes due to appearance and disappearance of lesions such as microaneurysms, hemorrhages, exudates, and cotton wool spots. In addition to the number of these lesions at the time of examination, the dynamics of lesions such as the lesion turnover rate can provide more insight into the disease activities over time [2, 3]. Therefore, automated detection and quantification of longitudinal retinal changes can be an important addition to regular DR screening.

Automated detection of longitudinal retinal changes due to microaneurysms and dot hemorrhages from a series of fundus images is challenging due to intra and inter-image illumination variation between fundus images captured at different retinal checkups (Fig. A.1b). A previous approach for longitudinal fundus image analysis excluded retinal regions around the borders of fundus images due to illumination variation and analyzed the remaining regions to identify change locations [4]. Retinal changes are then detected using hand crafted features extracted from each of the images [4, 5]. Recently, deep convolutional neural networks (CNNs) are shown to be successful in automatically learning local image features for object detection and classification [6, 7]. Deep CNNs can learn local object features in a hierarchical fashion using kernels that are limited to a small neighborhood. In this paper, we explore the performance of a CNN model for the detection of DR related retinal change in pairs of fundus images which are normalized for illumination variation and compare its performance with two other CNNs trained separately on color and green channel fundus images.

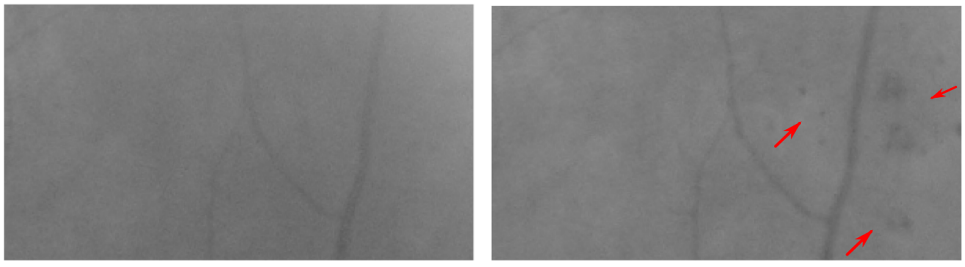
A.2. Methodology

A.2.1. Illumination Normalization and Registration

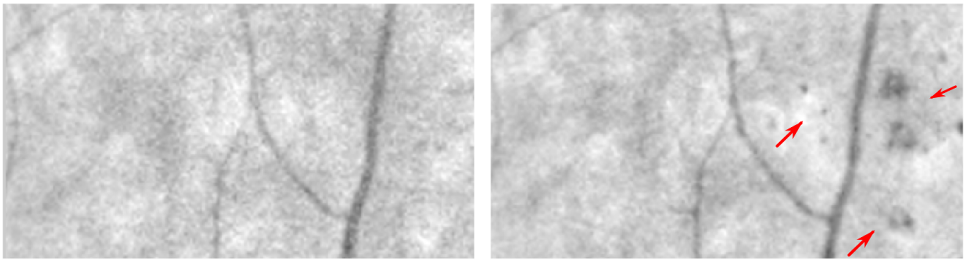
In color fundus images, the red and blue channels suffer from low contrast between the retinal features and the background. Because of its higher contrast, the green channel of the digital fundus images is commonly used in DR screening by eye care experts as well as in automated fundus image analysis. However, the green channel images still show considerable variation in luminosity and contrast, both within and between images. This variability was normalized by estimating the luminosity and contrast from the local intensity distribution of the so-called background retina (i.e., the retina excluding features such as vessels, optic disc, and lesions) and subsequently correcting for their variation over the entire retinal image [8](Fig. A.1c).



(a) Color fundus image patches.



(b) Gree channel fundus image patches.



(c) Normalized fundus image patches.

Figure A.1: Examples of various fundus image patches showing retinal changes due to microaneurysms and hemorrhages (red arrows) between the baseline (left) and follow-up (right) DR checkups.

Then, the intra- and intervisit fundus image series were aligned by a registration method that makes use of the normalized intensity as well as structural information of the retinal vasculature using a multiresolution matching strategy coupled with a hierarchical registration model [8]. The color and green channel fundus images were also registered into the same coordinate system as the normalized images using the final estimate of the transformation model parameters.

A.2.2. Data and Reference Annotation

Data for this study was obtained from the regular DR screening program at the Rotterdam Eye Hospital. Four field (macula-centered, optic nerve-centered, superior, and temporal regions) fundus image sets from 51 diabetic eyes who were examined for DR in 2012 and 2013 were included. For each eye, three expert graders inde-

pendently inspected and annotated the registered color and normalized images for microaneurysm and hemmorrhage related retinal change between the two screening time-points. The experts annotated the center of each changed region. The reference annotation was defined as the union of all the annotations by the three graders. The estimated diameter of the annotated regions ranges from 3 to 16 pixels ($21\mu m$ to $112\mu m$).

A.2.3. Convolutional Neural Network Architecture

The CNN model consists of two convolutional layers and one fully connected layer (Fig. A.2). In both convolutional layers, kernels of size 5×5 were applied to learn local image features. To progressively reduce the size of the spatial representation of the objects in the image and the amount of parameters, the kernels were applied with a stride of 2. A rectified linear unit (ReLU) activation function was employed after each convolutional layer. Then, 32 features maps were generated by the fully connected layer and fed into a softmax classifier to compute the probability that a *change* (or *no change*) has occured between the baseline and follow-up retinal regions. Note that in addition to the CNN shown in Fig. A.2 with a specific configuration of the normalized input images ($32 \times 32 \times 2$), other CNNs consisting of different image types (color or green channel) and configuration (difference images) were also explored.

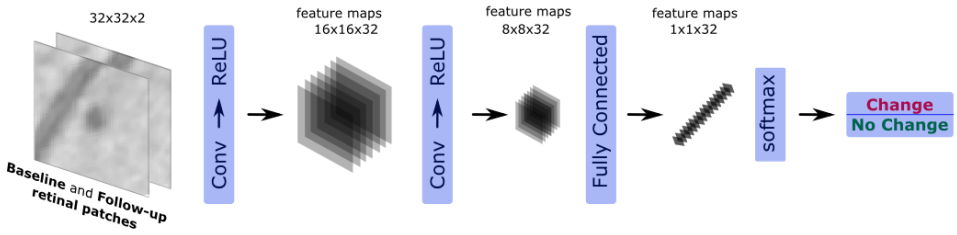


Figure A.2: An overview of the convolutional neural network architecture.

A.3. Experiments and Results

A total of 531 retinal change locations were annotated on the 51 baseline (I_B) and follow-up (I_F) registered image pairs. A positive class, which represents the occurrences of a change, was formed by gathering image patch pairs of 32×32 pixels centered on each annotated change location. A negative class was created by randomly sampling 531 image patches of size 32×32 pixels from locations that were not marked by any of the experts. After splitting all the gathered patches as training (80%) and evaluation (20%) set, the sample size of each set was increased by applying data augmentation techniques such as Gaussian blurring, rotation, and flipping. In total, 7650 training and 1908 evaluation baseline and follow-up image patch pairs were gathered from each of the color, green channel and normalized fundus image types.

For each image type, the CNN model hyperparameters were optimized indepen-

dently using a subset of samples from the training set. We explored two approaches to combine I_B and I_F and feed them into the CNNs. In the first approach, both I_B and I_F were directly fed into the CNN models as separate channels (Fig. A.2). In the second approach, the absolute difference between the two images, $|I_B - I_F|$, was used as input to the CNNs. The performance evaluation metrics were the sensitivity and the specificity, which are computed as

$$Sensitivity = \frac{TP}{TP + FN} \tag{A.1}$$

$$Specificity = \frac{TN}{TN + FP}, \tag{A.2}$$

where TP is the number of true positives, FN is the number of false negatives, TN is the number of true negatives and FP is the number of false positives. The results are summarized using receiver operating characteristics (ROC) curves. All the experiments were done using TensorFlow [9].

The results show that in both approaches the accuracy of the CNNs trained on the normalized images was higher than the accuracy of the CNNs trained on either the color or the green channel images (Fig. A.3). In addition, the results suggest that illumination normalization helps the CNN to quickly learn distinctive local image features of DR related retinal changes and thus converge fast. The ROC curves in Fig. A.4 show that a higher sensitivity and specificity was achieved for the CNN that is trained on the normalized images than on the color or green channel images. For both the normalized and green channel images, a slight increase in performance was observed when directly using the I_B and I_F input image patches as separate channels than combining them as $|I_B - I_F|$. For the color and green changel images, the performance does not reach the same level as for the normalized images even after many iterations. This may be due to the normalization operating on a larger scale (151×151 pixels) than the kernel size of the CNN (5×5 pixels).

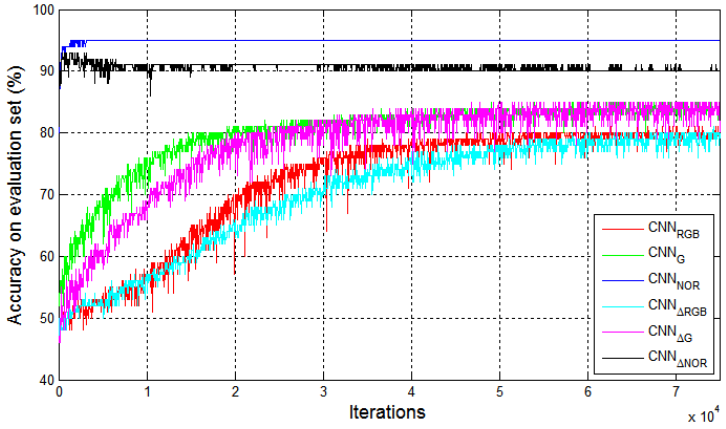


Figure A.3: Accuracy on the evaluation set after each iteration.

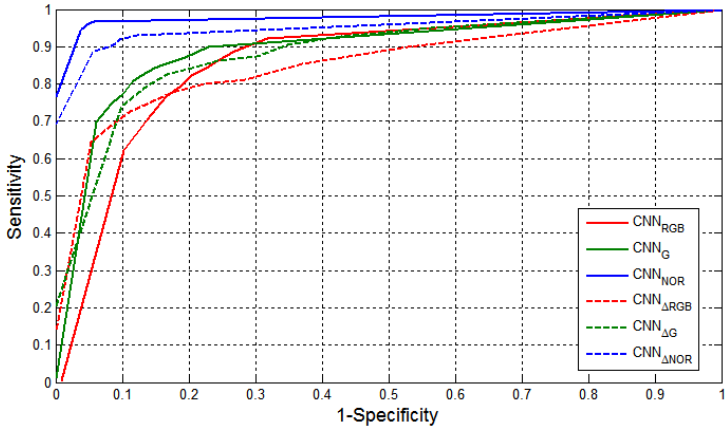


Figure A.4: ROC curves of CNN models trained on color (RGB), green channel (G), and normalized (NOR) fundus images for the detection of retinal changes due to DR lesions. For each image type, the CNNs that are trained directly on I_B and I_F are indicated by CNN_* and those networks trained on $|I_B - I_F|$ are indicated by $\text{CNN}_{\Delta *}$.

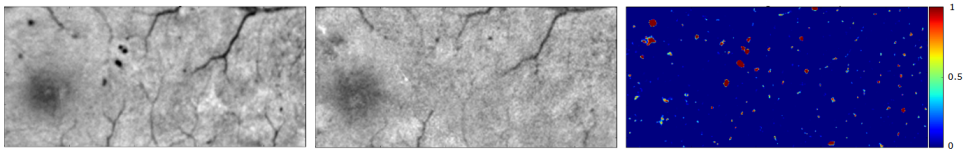


Figure A.5: An example of baseline (left) and follow-up (middle) retinal regions on which a CNN trained on normalized images was applied to produce pixelwise probability map (right) for DR related retinal changes.

A

A.4. Conclusion

In this paper, we presented an approach for the automated detection of longitudinal retinal changes from a series of fundus images. The approach employed a deep CNN trained on normalized fundus images that are corrected for intra- and inter-visit illumination variations, thereby enabling the CNN to correctly learn highly representative local image features of DR related retinal changes. Evaluation showed that the CNN network trained on normalized fundus images outperforms two other CNNs trained separately on color and green-channel images. The detected DR related changes may be used for objective assessment of DR progression as well as for more efficient human grading by highlighting DR related changes since the previous visit (Fig. A.5). Future work includes incorporating contextual information between neighboring pixels using fully CNNs and incorporating postprocessing methods to remove regions that are less likely to be clinically relevant from the pixelwise probability map.

Acknowledgments.
This work is financially supported by CZ Fonds (201400002), Stichting Blindenhulp and Stichting Wetenschappelijk Onderzoek het Oogziekenhuis.

References

[1] Early Treatment Diabetic Retinopathy Study Research Group, *Early photocoagulation for diabetic retinopathy: ETDRS report number 9,* Ophthalmology **98**, 766 (1991).

[2] S. Nunes, I. Pires, A. Rosa, L. Duarte, R. Bernardes, and J. Cunha-Vaz, *Microaneurysm turnover is a biomarker for diabetic retinopathy progression to clinically significant macular edema: findings for type 2 diabetics with nonproliferative retinopathy,* Ophthalmologica **223**, 292 (2009).

[3] J. Cunha-Vaz, L. Ribeiro, and C. Lobo, *Phenotypes and biomarkers of diabetic retinopathy,* Progress in retinal and eye research **41**, 90 (2014).

[4] H. Narasimha-Iyer, A. Can, B. Roysam, C. V. Stewart, H. L. Tanenbaum, A. Majerovics, and H. Singh, *Robust detection and classification of longitudinal changes in color retinal fundus images for monitoring diabetic retinopathy,* Biomedical Engineering, IEEE Transactions on **53**, 1084 (2006).

[5] J. Cunha-Vaz, R. Bernardes, T. Santos, C. Oliveira, C. Lobo, I. Pires, and L. Ribeiro, *Computer-aided detection of diabetic retinopathy progression,* in *Digital Teleretinal Screening* (Springer, 2012) pp. 59–66.

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, *Imagenet classification with deep convolutional neural networks,* in *Advances in neural information processing systems* (2012) pp. 1097–1105.

[7] Y. LeCun, Y. Bengio, and G. Hinton, *Deep learning,* Nature **521**, 436 (2015).

[8] K. M. Adal, R. M. Ensing, R. Couvert, P. van Etten, J. P. Martinez, K. A. Vermeer, and L. van Vliet, *A hierarchical coarse-to-fine approach for fundus image registration,* in *Biomedical Image Registration* (Springer, 2014) pp. 93–102.

[9] M. Abadi *et al.*, *TensorFlow: Large-scale machine learning on heterogeneous systems,* `http://tensorflow.org/` (2015), software available from tensorflow.org.

Summary

Diabetic retinopathy (DR) is a complications of diabetes mellitus, which progressively damages small retinal blood vessels and result in vision loss if not treated and controlled timely. Because of an increase in the risk of vision loss with the duration of diabetes and the latency between DR progression and early symptoms, diabetic patients require periodic screening. The required regular screening by a trained clinician, based on fundus photos, is time consuming, subjective, and resource demanding. Furthermore, the current practice does not scale well with the global rise in the diabetic population. Computer-aided screening offers a solution to this problem. This thesis presents several building blocks for automated analysis of a series of fundus images for DR.

DR screening involves acquisition and inspection of one or more fields of the retinal surface for early DR lesions such as microaneurysms (small round red lesions). Regular screening is aimed at identifying diabetic patient with a high risk of progression to referable DR. The dynamics of retinal change over time provides insight into the disease activity and can be used as a biomarker for DR progression. The fully automated system presented in this thesis enables tracking retinal changes due to small round red lesions in longitudinal series of fundus images, thereby providing an objective measure of the disease activity over time.

The proposed approach spatially aligns illumination normalized intra- and inter-visit fundus images by exploiting the intensity and structural information of the retinal vasculature. Illumination variation was addressed by normalizing luminosity and contrast variations in each of the retinal fields. To increase robustness to local minimum, a multiresolution registration approach coupled with a deformation model of increasing complexity was applied. Qualitative accuracy assessment on the suitability of the registration results for longitudinal analysis showed that the proposed approach can align images with a very high registration accuracy, warranting detection of retinal changes due to especially small retinal lesions (3 pixels or $21\mu m$ in diameter).

A robust and flexible multi-stage framework is proposed to detect and classify retinal changes in registered longitudinal fundus image sets. The framework exploits the enhanced visibility and contrast of small retinal features in the normalized images. To detect spatio-temporal retinal change locations, a simple and effective criterion for blobness measure (BM), which is defined as the absolute difference between the extremes of the multiscale blobness responses of fundus images from two time-points, was proposed. A supervised classifier was trained based on several intensity and shape features extracted from candidate retinal change locations to detect relevant early DR related changes. Evaluation in the context of regular DR screening involving subject with no, mild, and moderate DR levels shows that the proposed framework was able to detect retinal changes due to small DR lesion with

a sensitivity of 80% from both large and small field fundus mosaics with a relatively low false positive rate of 2.5 and 1, respectively.

The red lesion turnover, extracted using the proposed multi-stage framework, over time was explored as a potential predictor for future referral. The red lesion turnover between successive retinal examinations were retrospectively evaluated and compared between eyes with referable and non-referable DR. Evaluation results shows, on average, the red lesion turnover increases with duration of diabetes in both groups of eyes; however, a higher red lesion turnover rate is found in referable eyes compared to non-referable eyes. Furthermore, a statistically significant correlation was found between the red lesion turnover and classification of referable DR, suggesting that the red lesion turnover can be used as a potential biomarker in building a prediction model for future referral.

Overall, this thesis introduced a robust and flexible CAD system for analyzing a longitudinal series of fundus images for an objective assessment of DR progression. Several fully automated methods are presented to address fundus image illumination variation, registration, and retinal change detection. The CAD system was used to explore the red lesion turnover as a potential imaging biomarker for assessing DR progression. Furthermore, the intermediate results of the CAD system, such as normalized fundus images and fundus mosaics, are also shown to be potentially useful to facilitate inspection of fundus images by trained eye care experts.

Samenvatting

Diabetische retinopathie (DR) is een complicatie die optreedt bij diabetes mellitus, waarbij kleine retinale bloedvaten steeds verder beschadigen, hetgeen in verlies van gezichtsvermogen kan uitmonden als adequate behandeling uitblijft. Omdat het risico op verlies van gezichtsvermogen met de duur van diabetes toeneemt en het symptomen in de eerste fase van DR progressie uitblijven, is regelmatige screening van diabetespatiënten nodig. Deze benodigde regelmatige screening door een getrainde clinicus, gebaseerd op fundusfoto's, is tijdrovend, subjectief en legt een groot beslag op beschikbare bronnen. Bovendien is de huidige praktijk niet goed op te schalen met de voortdurende toename van diabetes in de populatie. Computerondersteunde screening biedt een oplossing voor dit probleem. In deze thesis worden verschillende onderdelen voor een automatische analyse op DR van series van fundusfoto's gepresenteerd.

DR screening omvat de acquisitie en inspectie van een of meerdere velden van het oppervlak van het netvlies op vroege DR laesies zoals microaneurysmata (kleine, ronde, rode laesies). De reguliere screening is gericht op het identificeren van diabetespatiënten met een verhoogd risico op progressie tot doorverwijsbare DR. De dynamica van de veranderingen in het netvlies gedurende de tijd geeft inzicht in de ziekteactiviteit en kan gebruikt worden als een biomarker voor progressie van DR. Het volledig automatische systeem dat in deze thesis wordt gepresenteerd maakt het mogelijk om de retinale veranderingen door kleine, ronde, rode laesies te volgen in series van fundusfoto's en biedt daarmee een objectieve maat voor de ziekteactiviteit gedurende de tijd.

De voorgestelde aanpak zorgt eerst voor een spatiele correspondentie tussen intra- en inter-visite fundusfoto's, na normalisatie voor belichting, door gebruik te maken van intensiteits- en structuurinformatie van het retinale vaatstelstel. Variatie in belichting werd aangepakt door te normaliseren voor variaties in helderheid en contrast in elk van de retinale velden. Om robuust te zijn voor lokale minima werd een aanpak gebruikt, gebaseerd op multi-resolutie registratie gekoppeld aan een deformatiemodel van toenemende complexiteit. Een kwalitatieve analyse van de nauwkeurigheid van de toepasbaarheid van de registratieresultaten voor een tijdsanalyse toonde aan dat de voorgestelde aanpak in staat is om beelden met een zeer hoge nauwkeurigheid te registreren, dat de detectie van retinale veranderingen, vooral door kleine retinale laesies (met een diameter van 3 pixels of $21\mu m$), mogelijk maakt.

Een robuust en flexibel meertraps-framework voor de detectie en classificatie van retinale veranderingen in geregistreerde sets van fundusfoto's wordt geïntroduceerd. Dit framework maakt gebruik van de verbeterde zichtbaarheid en contrast van kleine retinale kenmerken in de genormaliseerde beelden. Om locaties met spatio-temporele retinale veranderingen te detecteren wordt een eenvoudige

en effectieve maatstaf voor 'blobness' geïntroduceerd, dat gedefinieerd is door het absolute verschil tussen twee extremen van de multi-schaal blobness respons in fundusfoto's van twee tijdpunten. Een gesuperviseerde classificator werd getraind, op basis van verscheidende intensiteits- en vormkenmerken gedefinieerd op kandidaatlocaties met retinale veranderingen, voor het detecteren van relevante vroege DR-gerelateerde veranderingen. De evaluatie binnen de reguliere DR screening van patienten zonder, met milde, en met matige DR liet zien dat het voorgestelde framework in staat was om retinale veranderingen door kleine DR laesies te detecteren met een sensitiviteit van 80% in zowel kleine als grote fundusvelden met een relatief lage proportie fout positieven van respectievelijk 2.5 en 1.

De turnover (verandering) van rode laesies, bepaald met het beschreven meertrapsframework, gedurende de tijd werd onderzocht als mogelijke voorspeller van toekomstige verwijzing. De turnover van rode laesies tussen opeenvolgende retinale onderzoeken werden retrospectief geëvalueerd en vergeleken tussen ogen met en zonder verwijsbare DR. Deze evaluatie liet zien dat, gemiddeld, de turnover van rode laesies toenam met de duur van diabetes in beiden groepen ogen; een hogere turnover snelheid werd echter gevonden in de verwijsbare ogen vergeleken met de niet-verwijsbare ogen. Bovendien werd een statistisch significante correlatie gevonden tussen de turnover van rode laesies en DR classificatie, wat suggereert dat de turnover van rode laesies gebruikt kan worden als een mogelijke biomarker voor het bouwen van een voorspellend model voor toekomstige verwijzingen.

In deze thesis werd een robuust en flexibel CAD systeem geïntroduceerd voor het analyseren van longitudinale series van fundusfoto's voor een objectieve beoordeling van DR progressie. Verschillende volledig automatische methodes werden gepresenteerd voor het omgaan met variatie van belichting, registratie, en het detecteren van retinale veranderingen. Het CAD systeem werd gebruikt om de turnover van rode laesies als mogelijke beeldvormende biomarker voor het bepalen van DR progressie te onderzoeken. Daarnaast bleken de tussenliggende resultaten van het CAD systeem, zoals genormaliseerde fundusfoto's en mozaïeken daarvan, mogelijk nuttig te zijn bij de inspectie van fundusfoto's door getrainde oogzorgexperts.

Acknowledgements

One of the most rewarding experiences that this PhD journey has brought me was the opportunity to get to know and work with many people. I hereby would like to thank everyone who have made my stay and work in The Netherlands enjoyable and productive.

First and foremost, it is with immense gratitude that I acknowledge my daily supervisor and co-promoter **Dr. Ir. Koen A. Vermeer** and my promotor **Prof. Dr. Ir. Lucas J. van Vliet**, without whom this thesis would not have been possible. I could not ask for more!

Dear Koen: I thank you so much for giving me the opportunity to work on this project and it has been such a great pleasure working with you. Your dedication and continued support have been valuable to the success of my work. Your coaching style and constant challenges helped me improve myself and grow professionally. I thank you so much for always being there for me throughout this PhD period!

Dear Lucas: It has also been such a great honor and pleasure working with you and thank you so much for giving me this opportunity. Our regular meetings were not only where I had productive discussion and new ideas about my research but also where I get a motivation to energize myself and work more. You are a big influence on my career and I thank you for that.

I would also like to say a big thank you to **Dr. Jose P. Martinez** and **Dr. Peter G. van Etten** for providing me all valuable clinical inputs needed for my research. Your commitment and appreciation of my work have been vital during this PhD journey. I appreciate all the after-hours that you have dedicated to support my research. Many thanks to Kenneth for providing clinical inputs and annotations.

Netty, thank you so much for giving me the opportunity to work at The Rotterdam Ophthalmic Institute (ROI) and for helping me out with all the paper works before and after moving to The Netherlands.

Jelena and **Henk**, it has been a pleasure knowing you and working with you in the same research group. Thank you all for all the coffee breaks that we discussed about our frustrations and ways to deal with it. Many thanks to Kari, Susan, Elrozy, Niels, Boy, Gijs, Verena, Mia for all the fun social events and conference trips.

Many thanks to all ROI colleagues Rene, Annemiek, Caroline, Aletta, Mathi, Max, Stijn, Esma, Angela, Aline, Lisette, Juleke, Valentina, Grzegorz, and Eva for making my research at ROI enjoyable and productive.

To all QI group members of TU Delft: Robiel, Babak, Juan, Zhang, Jeroen (2x), Milos, Lennard, Nadia, Mojtaba, Robert, Jelle, Tom, Zhang, Jianfei, Joor, Willem, Lena, Anna, Gyllion, Robert, Ronald and all others that I have not mentioned your name, thank you for making my stay and research work in the group interesting.

I would also like to thank my Ethiopian friends in The Netherlands: Aemro, Tigist, Girma, Sirak, Chalachew, Aklilu, Getachew, Henok, Feven, Meseret, Meles,

Abeleneh, Alhayat, Admassu, Surafel, Konjit, Sara, Bety and others that I have not mentioned your name, for making my stay in The Netherlands enjoyable.

Last but not least, I would like to thank my families back my home country for being a big part of my academic career.

About the author

Kedir M. Adal was born in Dessie, Ethiopia on August 15, 1985. He studied and obtained his Bachelor of Science (BSc) degree in Electrical Engineering from Mekelle Univeristy, Ethiopia in 2006. He pursued a Master of Science (MSc) in the field of Computer Vision and Robotics and obtained his degree in 2012 from Universite de Bourgogne (UB), France. During his MSc study, he gained interest in image analysis and computer-aided diagnosis (CAD) and subsequently did his thesis on automated analysis of fundus images for the detection of early DR lesions. In February 2013, he started his PhD in a collaboration project between TU Delft, Rotterdam Eye Hospital and Rotterdam ophthalmic institute to develope a fully automated CAD system for assessing DR progression in longitudinal fundus images, which resulted in this thesis.