Integrated Master in Biomedical Engineering

Faculty of Sciences and Technology
University of Coimbra

Search Acceleration Methods for Large Retinal Image Databases

Diana Craveiro Mourão

2013

Integrated Master in Biomedical Engineering

Faculty of Sciences and Technology
University of Coimbra

Search Acceleration Methods for Large Retinal Image Databases

Dissertação apresentada à Universidade de Coimbra para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Engenharia Biomédica, realizada sob a orientação do Engenheiro Carlos Manta Oliveira, Critical Health S.A. e do Professor Doutor Luís Alberto da Silva Cruz, Departamento de Engenharia Electrotécnica e de Computadores da Faculdade de Ciências e Tecnologia da Universidade de Coimbra.

Diana Craveiro Mourão

2013

Contents

List of Figures

List of Tables

viii

List of Abbreviations

Abbreviation	Meaning
AMD	Age-Related Macular Degeneration
BoW	Bag-of-Words
DR	Diabetic Retinopathy
FD	Fractal Dimension
KNN	K-Nearest Neighbor
SIFT	Scale Invariant Feature Transform
SURF	Speeded Up Robust Features
SVM	Support Vector Machine
VT	Vocabulary Tree
WEF	Wavelet Energy Feature

Nomenclature

Symbol	Meaning
p	Image partitioning degree
N_r	Number of image subregions with at least one nonzero pixel (Fractal Dimension)
D	Fractal Dimension value
$F(a,b)$	Continuous Wavelet Transform
$\psi_{a,b}$	Mother Wavelet
A_J	Approximation Wavelet Coefficients of the last decomposition level
H_i	Horizontal Wavelet Coefficients of decomposition level i
V_i	Vertical Wavelet Coefficients of decomposition level i
D_i	Diagonal Wavelet Coefficients of decomposition level i
E_i^h	Horizontal Wavelet Energy of decomposition level i
E_i^v	Vertical Wavelet Energy of decomposition level i
E_i^d	Diagonal Wavelet Energy of decomposition level i
$L(x,y,\sigma)$	Image Scale Space
$G(x,y,\sigma)$	Gaussian Function
$D(x,y,\sigma)$	Difference-of-Gaussian Function
H	Hessian Matrix
$m(x,y)$	Image Magnitude Gradient

$\theta(x, y)$	Image Orientation Gradient
k	Number of children nodes
L	Number levels of the Vocabulary Tree
$w(v)$	Weight of node v
N	Total number of images
N_v	Number of images in node v
q	Query Image
d	Database Image
s	Scale at which the SIFT/SUR keypoint was detected
θ	Orientation assigned to the SIFT/SURF keypoint

Acknowledgments

Apart from the efforts of the author, the success of this work depends largely on the motivation and guidance of many others.

I would like to acknowledge with much appreciation the crucial role of Professor Luís Cruz from the Department of Electrical and Computer Engineering at the University of Coimbra, for the guidance throughout the development of this work, as well as for all the patience it required. His suggestions had an important role in the way this dissertation was conducted.

I am also grateful for the opportunity of developing my thesis at such a successful company as Critical Health S.A. which enabled me to not only work in an excellent environment but also meet interesting and competent people. For this I would like to thank all my colleagues from Critical Health S.A., especially to my supervisor Carlos Manta Oliveira for always appreciating the value of my work and motivating me to improve it.

To my parents, Jaime and Isabel Mourão, I cannot say thank you enough for their tremendous support in so many aspects of my life, mainly in this last phase of my studies. They certainly are my role model. A special thanks goes to my sister, Cátia Mourão for sharing such happy moments with me. She does not know how important they were.

Finally I would like to express my deepest gratitude to all my friends who, in one way or another, where present in this important period of my life, namely to André Silva whose unconditional love and support were essential to successfully finish this work. I will be forever grateful to you.

Abstract

Retinal pathologies monitoring may be performed by comparing retinal images of the same patient and eye captured at different times. To perform this comparison, after the acquisition of a new retinal image, it is necessary to locate other images stored in the database that belong to the same patient and eye. Although stored images may be easily labeled (using metadata) to facilitate the database search, errors may occur and relevant information such as patient information may not be correctly retrieved. To overcome this problem, images must be identified solely based on visual features. To perform retina-based person identification several techniques, like image registration, have been developed in the past. Unfortunately the sequential search based on the comparison of the newly retinal image with each image in the database is a costly procedure. The aim of this work is the research and development of faster image search methods based on the selection of a subset of the image database, which is then subject to a verification step during which a retina-based identification method accurately verifies which images from this subset belong in fact to the same patient and eye from which the query image was captured. Naturally, to be efficient, the algorithm must select a subset of images which should be as small as possible while ensuring that the overall identification performance is as high as possible.

The major contribution of this work is the application of Vocabulary Trees to the problem of efficient retinal image database search, using different features for retinal image representation. Four different types of features were considered: Fractal Dimension and Wavelet Energy Features (WEF), both used in biometric identification procedures using retinal images, and SIFT and SURF image features. These image descriptors were then organized using Vocabulary Trees to build structures which made the search procedure a lot faster.

This novel approach of retinal image search using Vocabulary Trees, enables the clustering of images accordingly to their similarity, which facilitates the selection of the subset of images most similar to the one being queried. As expected, the best retrieval performance was obtained with image features used in retina-based person identification systems, resulting in the selection of a small subset of images from the database while ensuring that the correct images were retrieved.

The proposed algorithm enables, on average, the use of only 0.042% of the total database in the retina-based person identification method, resulting in a 97.57% overall computation time reduction when compared to the use of the retina-based person identification method in the total database. The developed algorithm achieves a 98.79% accuracy in the retrieval of the correct images.

Keywords: Retinal Image, Retinal Image Recognition, Feature Extraction, Image Registration, Image Texture Analysis, Image Processing, Vocabulary Trees, Search Techniques.

Sumário

A monitorização de patologias da retina pode ser feita através da comparação de imagens retinais do mesmo paciente e mesmo olho capturadas em diferentes períodos de tempo. De modo a fazer esta comparação, após a aquisição de uma nova imagem da retina, é necessário identificar na base de dados as restantes imagens pertencentes a esse mesmo paciente e olho. Embora as imagens armazenadas possam ser facilmente etiquetadas usando metadados de forma a facilitar a pesquisa na base de dados, frequentemente ocorrem erros e informação importante como a identificação do paciente pode não ser obtida correctamente. Para precaver este problema, as imagens devem ser identificadas apenas através de características da própria imagem. Para fazer o reconhecimento de indivíduos baseado na retina existem várias técnicas eficientes que podem ser aplicadas, tal como o registo de imagens. Um aspecto negativo de métodos baseados na pesquisa exaustiva da base de dados de imagens, com a comparação da imagem da retina a pesquisar com todas as imagens armazenadas na base de dados é o elevado custo computacional. O objectivo deste trabalho é a pesquisa e desenvolvimento de métodos que permitam definir um subconjunto de imagens da base de dados, idealmente contendo todas as imagens do mesmo paciente que a imagem da retina em questão, que após passagem por uma segunda fase de identificação na qual é usado um método de reconhecimento de indivíduos com base na retina, permitam a identificação expedita das imagens que pertencem de facto ao mesmo paciente e olho. Para o desenvolvimento de um algoritmo eficiente, o conjunto de imagens devolvido deve ser o mais reduzido possível e o desempenho medido pela taxa de identificação correcta deve ser o mais elevado possível.

A contribuição principal deste trabalho centra-se no uso de Árvores de Vocabulário para pesquisa eficiente em bases de dados de imagens da retina, aliado ao uso de diferentes descritores de características visuais das imagens retinais. Foram considerados quatro tipos diferentes de descritores de imagens: Dimensão Fractal e Energia de Wavelets, ambos usados em estudos de reconhecimento biométrico a partir de imagens da retina e também descritores SIFT e SURF. Os conjuntos de descritores associados às imagens da base de dados foram organizados usando Árvores de Vocabulário por forma a permitirem uma pesquisa rápida da base de dados.

Este procedimento inovador de pesquisa de imagens da retina recorrendo a Árvores de Vocabulário, permite agrupar as imagens de acordo com a semelhança que apresentam entre si, facilitando a selecção de um subconjunto de imagens mais semelhantes à imagem da retina usada na pesquisa. Tal como era esperado, os métodos de extracção de características usados em sistemas de identificação de indivíduos baseados na retina foram os que apresentaram uma maior eficiência, isto é, que permitiram a selecção de um menor subconjunto de imagens da base de dados assegurando que as imagens correctas são identificadas. O algoritmo proposto permite, em média, usar apenas 0.042% da base de dados num método de identificação de indivíduos baseado na retina, representando 97.57% de redução do tempo de cálculo quando comparando com o uso do método de identificação de indivíduos na totalidade da base de dados. O algoritmo desenvolvido permite a identificação correcta das imagens pertencentes ao mesmo paciente em 98.79% dos casos.

Palavras-chave: Imagem da Retina, Reconhecimento de Imagem da Retina, Extracção de Características, Registo de Imagens, Análise de Texture de Imagem, Processamento de Imagem, Árvore de Vocabulário, Técnicas de Pesquisa.

Chapter 1

Introduction

1.1 Problem Contextualization

In the developed countries where a considerable large share of the population is an aged population further importance needs to be given to the study of typical diseases of this age group. Eye diseases, in particular those related to aging, have been concerning scientists and professionals in the health industry. More importance has been given to those that may cause vision loss or total blindness such as Diabetic Retinopathy (DR) and Age-Related Macular Degeneration (AMD) [1]. Surveys suggest that over 25 million people worldwide are affected by AMD, which is the leading cause of blindness in people over 55 in the developed countries. Moreover, this number is expected to triple by 2025 [2]. AMD causes the macula cells of the eye to become damaged leading to a gradual loss of the central vision [3]. In addition to AMD prevalence, studies revealed that the number of diabetic people is expected to reach 552 million by 2030, meaning that one person in ten will suffer from this disease [4]. A person with Diabetes is exposed to a variety of diabetic eye diseases including DR, cataracts and glaucoma. DR, the most common diabetic eye disease, is characterized by damage of the retina blood vessels. Cataracts develop at an early age in diabetic patients and cause clouding of the eye's lens. Glaucoma is the increase in the pressure of the fluid inside the eye which can lead to damage in the optic nerve [5].

The early detection of these pathologies is crucial to maintain a good vision and prevent such severe consequences as blindness [6]. When detected at an early stage, these eye diseases may be treatable or their progress may be slowed down. DR and AMD are diagnosable based on perfectly characterized symptoms which may be detectable by visual inspection of the eye fundus [7].

Eye disease screening, especially in the early asymptomatic stage, has been shown to be effective in the prevention of vision loss and reduction of the costs associated with disease progression and treatment [8]. A typical screening process involves the acquisition of retinal images with specific cameras and visual inspection by medical experts. In order to detect anomalies in the eye, different retinal images taken at different time periods are compared in order to identify modifications that may be indicators of an unhealthy eye. However, this manual image comparison process is inefficient since it may be subjective and depends on the medical proficiency, being both expensive and time consuming. These disadvantages conducted

to the development of automated clinical decision support systems focused on detecting modifications in the retina through time [7]. The advances in digital imaging and computing power enabled the use of these new technologies in ophthalmology. Over the past 15 years progresses have been made in the field of automated retinal images analysis systems to be used in screening programs, with better resource use and reduction of diagnostic errors due to observer bias [9]. At the moment some manufacturers already offer automated diagnosis systems which support medical experts in clinical decisions. Critical Health's Retmarker® is one of these retinal image analysis tool which performs automated eye disease diagnosis. Retmarker has three different versions: RetmarkerDR which is a biomarker for DR progression, RetmarkerC which performs automatic detection of retinal changes over time and RetmarkerAMD which is an assisted AMD grading tool [10].

In these clinical decision support systems it is a common practice to compare a newly obtained retinal image with previous acquired ones of that same patient in order to monitor the eye disease progress and evaluate the efficiency of the treatment. Thereby, retinal images need to be stored and easily identified for the proper functioning of the automated diagnosis system. Although database images may be indexed or labeled to easily search for a specific patient, errors often occur, such as misspelled names and typos, and thus the relevant information, i.e. the other images of the same patient, cannot be retrieved. This problem was the main motivation for the development of this thesis. This work is focused on retinal image database search techniques based solely on image features thus avoiding errors due to incorrect database text information. The aforementioned database search consists of the identification of stored retinal images belonging to the same patient and eye as those of a query retinal image.

Since it is necessary that the image is identified based on its image features, retinal-based person identification methods were studied. These methods enable the comparison of a pair of retinal images and their identification as belonging or not to the same patient and eye. Several methods used for biometric recognition apply retinal image registration attaining good identification performance [11] [12]. However, the goal of this work is substantially different from a person identification problem using retinal images. While in retina-based person identification the aim is to verify whether two images belong to the same individual, the purpose of this work is the search of stored retinal images belonging to the same patient and eye of an input image without any prior information. Obviously person identification methods could be solely considered if applied between the query image and all the retinal images from the database. However, taking into consideration that this software deals with hundreds of retinal images, solely using this method was found not to be suitable to solve the problem. The solution considered involves the selection of a subset of images from the entire database, as reduced as possible, which groups the most similar retinal images to the one being queried. The similarity between the query retinal image and each one from the retrieved subset of images is appraised by using a retina-based person identification method which enables the identification of the images belonging to the same patient and eye of that of the query image, with less computation time than by applying only the person identification method to the entire database.

For the development of this work, image feature descriptor methods, as well as database search techniques, were studied in detail. As for the image feature descriptors methods, more importance was given to those used in biometric recognition systems using retinal images. The goal is an efficient retinal image representation solely by image feature descriptors. Since the problem consists in the search of images belonging to the same patient of a retinal image being queried, the feature descriptors of retinal images belonging to the same patient and eye must be

similar and the feature descriptors of retinal images from different patients and eye must be very dissimilar. Besides retinal image representation by image feature descriptors, database search techniques were explored. As aforementioned, in a database with thousands of retinal images it is not feasible to perform a comparison between the query and each database image. Ideally retinal image descriptors should be chosen to enable faster selection of images similar to a query image. Even so, usually the time required to search the entire database using naïve approaches is not adequate to an automated system. For this reason, fast image retrieval methods were studied, with emphasis on Vocabulary Trees (VT) which have been successfully used in similar contexts [13] [14]. In VT, at each tree-level the image feature descriptors are divided into clusters accordingly to their similarity. For each cluster a mean centroid descriptor vector is computed. When searching for images belonging to the same patient of an input image, the same type of descriptors are computed from the query image and each one traverse the VT by choosing the closest node at each tree-level (see Chapter 4 – Methodology for more details). The images retrieved are those represented in the selected terminal nodes. The main steps of the solution proposed in this thesis are illustrated in Figure 1. Several VT techniques and image feature computation methods described in the scientific literature were studied and applied in this work. From the research performed no information was found about the use of VTs in large retinal image databases for recognition purposes, which justifies the exploration presented here.
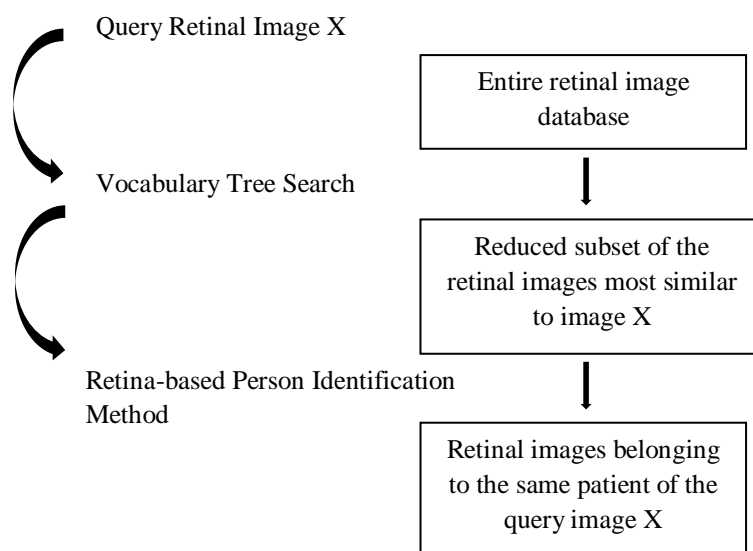
Figure 1: Representation of the goal algorithm for this thesis.

One aspect that may be taken into consideration is the fact that the algorithm developed is meant to be used in an automated diagnosis system, meaning that retinal images probably present symptoms of eye pathologies such as hemorrhages, edemas and other lesions. Therefore, the algorithm must be robust to the presence of these conditions ensuring that image distortions and deviations from the normal do not affect the image retrieval performance. Since this type of automated retinal images analysis systems must be as trustworthy as possible, it was established that in the studies performed throughout this work in nearly 99% of the situations tested, the correct images were to be retrieved.

1.2. Document Overview

This document is divided into seven chapters. Chapter 1 presents the problem contextualization and a brief description of the resolution implemented. Chapter 2 provides important information to understand the problem, such as eye anatomy concepts and related common diseases and the use of automated diagnosis systems, which was the motivation for this dissertation. Chapter 3 is a review of published work about image feature descriptors, in particular retinal image feature descriptor methods used in biometric recognition procedures using retinal images, and database search techniques for efficient image retrieval. Chapter 4 lists the methods implemented for the problem resolution. This chapter contains two essential parts: one describing the image feature computation methods and other detailing the procedure used for database image retrieval. In Chapter 5 the results obtained are thoroughly evaluated and discussed in terms of suitability to the problem resolution and the final algorithm is herein described. Chapter 6 consists of a brief summary of the most important conclusions obtained. Suggestions for future work development are indicated in Chapter 7.

One appendix is included in this work, Appendix A, in which flowcharts of the different image feature descriptor computation methods implemented in this dissertation are illustrated.

1.3. Scientific Contributions

-Mourão D., Manta C. M., Cruz L. S., "Retinal Image Recognition using Tree Data Structure Search Method with Fractal Dimension Descriptors" , Accepted for oral presentation at HCIST 2013 - International Conference on Health and Social Care Information Systems and Technologies, Lisboa, Portugal, 2013.

Chapter 2

Retinal Imaging

This chapter provides the essential information to understand the problem which motivated this research, as well as all the main features of the solution proposed. A light introduction to the anatomy of the eye is provided and the most common eye pathologies relevant to this work as well as their characteristic lesions are presented. This information is important to understand and evaluate the proposed method. Besides this information, automated retinal image analysis systems to which the search method developed is target are introduced.

2.1. The Eye

2.1.1. Eye Anatomy

Sight is one of the most important human senses, enabling us to understand the environment around us. The eye, this amazingly complex structure, is constantly taking in the light reflected from objects near us. The light passes through the eyeball converging to a point where a set of cells at the back of the eye convert the light energy into electric impulses which are then transmitted to the brain. Each one of the eye structures contributes in a way to the overall visual process [15].

The eye may be considered as a slightly asymmetrical sphere with an approximate sagittal diameter of 25mm and a transverse diameter of 24 mm. In a cross-sectional view of the eye three different layers would be visible [16] . Figure 2 and 3 illustrate the arrangement of each layer and structure in the eye anatomy.

The external layer is formed by the sclera and the cornea, both making up the supportive structure of the eye. Both cornea and sclera have essentially the same chemical composition, although the cornea is transparent while the sclera appears opaque or translucent. This difference is explained by the fibrils arrangement since the chemical composition is identical [17]. The cornea is the clear part of the eye's protective covering and its main function is to focus the light for photoreception while the sclera mainly protects the eyeball [15] [18]. The sclera is a tough white outer coating of fibrous tissue which covers the eyeball and the muscles that move the eye are attached to it [19].
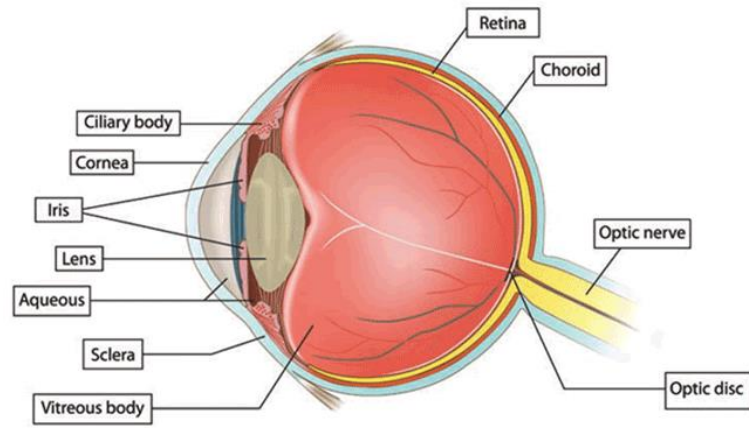
Figure 2: Eye anatomy [18].

The middle layer, also called uvea, contains the main blood supply of the eye and it is constituted by the choroid, the ciliary body and the iris. The outer layer of the uvea is the choroid. It is essentially a layer of blood vessels between the sclera and the retina that nourish the back of the eye. It is connected to the ciliary body in the front of the eye and to the optic nerve at the back of the eye [20]. The ciliary body is a muscular forward continuation of the choroid which main function is the change of the lens shape when the eye is focusing on something [15]. Lens is the structure responsible to focus light rays onto the retina [21]. It is composed of flexible tissue which enables the change in shape of this structure. The lens become more rounded to focus on near objects and more elongated to focus on far objects [19]. The iris is the colored part of the eye which regulates the amount of light entering through the eye by closing when there is bright light and opening when there is low light. The darker center in the middle of the iris is called pupil and is through it that light rays are transmitted. The iris controls the pupil size [21].

Figure 3: Representation of the three important layers of the eye: the sclera which belongs to the external part of the eye, the choroid which is part of the eye's middle layer and the retina, the most inner coating of the eye [20].

The internal layer, the sensorial part of the eye, is called retina. The retina function is to receive light and proceed to its conversion into chemical energy. This chemical energy activates nerves that produce electrical impulses. These impulses are then transmitted to the brain through the optic nerve [15]. Light conversion into chemical energy is only possible due to two different cell types of photoreceptors: cones and rods. Rod cells are extremely sensitive to light but have very low spatial resolution. On the other hand, cone cells have high resolution but are relatively insensitive to light. The properties of cone cells allow us to distinguish colors. Differences in the

mechanisms of these cells contribute to the different reaction of rods and cones to different ranges of light intensity [22]. Throughout this work the images used are retinal images similar to Figure 4. In the center of the retina is the optic nerve, a circular white area. The major blood vessels of the retina radiate from this structure. On the left of the optic nerve a blood vessel-free reddish spot may be seen, the fovea, which is at the center of the area called macula [23]. The macula is responsible for a clear central vision whereas the fovea provides the detail vision [19].
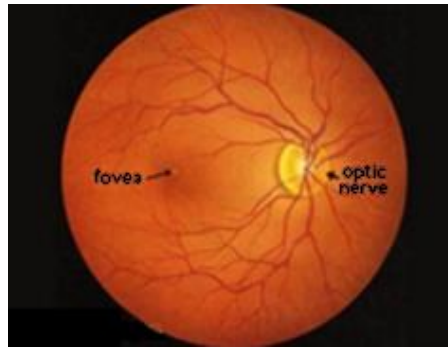
Figure 4: Human retinal image [23].

2.1.2. Eye Pathologies

Like any other organ of the human body, eyes are susceptible to pathologies as well. Many eye diseases have no early symptoms and no change in vision may be detected until the disease stage worsens [24]. Although eye disorders may appear at any age, most of them are more common among adults and especially among elder people [25].

Characteristic lesions of both Diabetic Retinopathy (DR) and Age-related Macular Degeneration (AMD) were found in the dataset of retinal images used in this work. Therefore a detailed description of these pathologies is presented below.

2.1.2.1. Diabetic Retinopathy (DR)

Diabetes is characterized by high blood glucose levels due to the fact that the body cannot properly use it. The reason is that the pancreas does not produce enough insulin (or that which is produced is not performing its function) which is responsible to help glucose, the main body fuel, enter the body's cells [26]. Diabetic Retinopathy (DR) is one of the common complications of Diabetes. It is characterized by the damage of the cells at the back of the eye, i.e. at the retina, due to the high blood glucose levels.

The retina is responsible for the conversion of photonic energy into electrical impulses which are then transmitted to the brain through the optic nerve. To properly function, the retina needs a constant supply of blood which is provided by blood vessels. A continuous high blood glucose level causes obstruction or leakage of the blood vessels, affecting the retina functions [27]. In some people abnormal new blood vessels may grow on the retina surface [5]. DR can cause vision loss when it progresses to an advanced stage.

This eye disorder has four stages:

Mild Nonproliferative Retinopathy: it is the earliest stage of the disease and it is characterized by the occurrence of microaneurysms. Microaneurysms are small areas of blood protruding from a vessel which may open and leak blood into the retinal tissue [28]. This results in the appearing of retinal hemorrhages, edemas and exudates. In Figure 5 hemorrhages are visible as reddish spots in the retina. Macular edema is the swelling or thickening of the part of the retina responsible for the central vision and appear in the retinal image as darker areas. Exudates are lipid residues of blood vessel leakage and appear as rounded yellow areas in the retinal image.
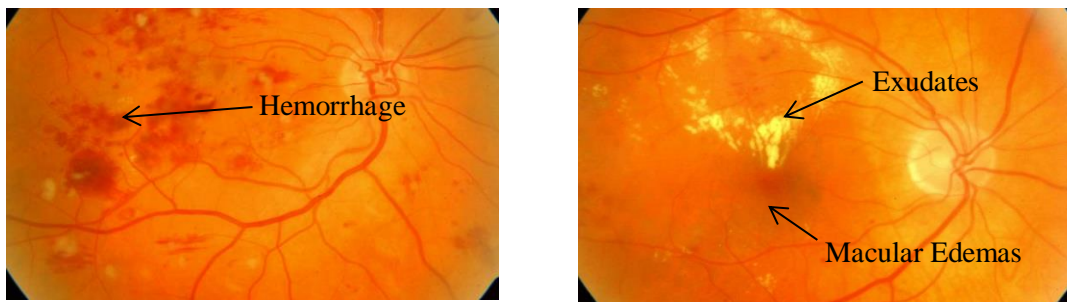
Figure 5: Representative retinal images of Nonproliferative Diabetic Retinopathy [29].

Moderate Nonproliferative Retinopathy: characterized by the blockage of blood vessels that nourish the retina [5].

Severe Nonproliferative Retinopathy: as the disease progresses many more blood vessels are obstructed depriving areas of the retina with blood supply. Those deprived areas transmit signals to the body to stimulate the growth of new blood vessels for the retina nourishment [5].

Proliferative Retinopathy: this is the most advanced stage of DR and is characterized by the growth of several new blood vessels which are abnormal and fragile. In case they suffer leakage it can result in severe vision loss or even blindness [5].

Symptoms of an advanced stage of DR include shapes floating in the vision field, blurred vision, reduced night vision and sudden blindness [27]. The difference between normal vision and that of a person with advanced DR is illustrated in Figure 6.

(a)   (b)

Figure 6: Representation of a scenario perceived by a person with normal vision (a) and by a person with Diabetic Retinopathy (b) [5].

2.1.2.2. Age-related Macular Degeneration (AMD)

Age-related Macular Degeneration (AMD) is the leading eye pathology causing blindness in the elderly. AMD essentially affects the retina area at the back of the eye called macula which is the center of vision. Although an exact cause for AMD is not known some factors may increase the chance of this disease development such as age, gender (more women suffer from AMD), genes, smoking, sunlight and diet [30].

The macula is a retina area that contains specialized photoreceptor cells designed cone cells. Cone cells allow color recognition and detail vision. When a person develops AMD these cells become damaged and do not properly perform their functions [30].

There are two forms of AMD, dry and wet AMD, and both can progress to advanced stages and cause severe vision loss.

<u>Dry AMD</u>

This form of AMD does not have many symptoms in the early stages. However, as the disease progresses the person may suffer from blurred vision and see objects more bright than they actually are. Usually people with dry AMD experience difficulty in face recognition and need more light to perform some tasks such as reading.

The most common early sign of dry AMD are drusens. Drusens are small yellow and white deposits in the retina made up of lipids and proteins. These substances are wasted products of photoreceptors cells which were not correctly disposed [31]. Drusens form on the retina, beneath the macula, causing it to deteriorate over time [32]. Figure 7 illustrates a retinal image affected with these lesions.

Dry AMD has three development stages. At an early stage there are no specific symptoms and only small or medium-sized drusens are formed. At an intermediate stage one or more larger drusens appear. At this stage some people see a blurred spot in the center of their vision and often need more light to perform some ordinary tasks. At an advanced dry AMD stage a breakdown of the light-sensitivite cells that support tissue in the macula occurs. This causes a blurred spot in the center of the vision which tends to get bigger and darker over time [33].
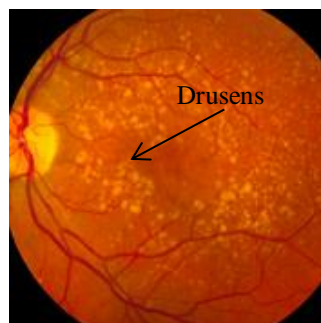
Figure 7: Representative retinal image of AMD with presence of drusens [34].

<u>Wet AMD</u>

The dry form of AMD can suddenly turn into the wet form which is what happens with 10% of all people suffering from AMD. This form of AMD is characterized by the growth of new blood vessels under the macula. The new blood vessels can be fragile and easily leak blood and fluids causing the macula swelling and damage. During the early stages of wet AMD straight lines may appear wavy [33].

(a)   (b)

<u>Figure 8</u>: Representation of a scenario perceived by a person with normal vision (a) and by a person with AMD (b) [33].

2.2. Automated Retinal Image Analysis Systems

2.2.1 The purpose of the Automated Retinal Image Analysis Systems development

The increase of the aged population worldwide means an increase of the number of people with eye pathologies as well. For this reason major importance has been given to ophthalmic services. Eye health includes several approaches such as raising awareness about eye health, early detection of the pathology, diagnosis and targeted prevention. The need in the clinical practice to find better ways of identifying and treating eye pathologies combined with recent advances in computer hardware and sophisticated image analysis techniques were the main driving forces to the development of methodologies and computer software for automated detection of retinal pathologies [8]. The aim of these automated systems is to aid ophthalmologists and physicians at their work by analyzing and measuring all features of interest in a retinal image with more precision than is routinely done. When the smallest abnormality is detected the system alerts the ophthalmologist to the need of a close scrutiny of the retinal image. Besides this, pairs of images are easily automatically compared for the detection of modifications over time [35].

Retinal images are a fascinating set of images to use in image analysis. They present a number of key features that may be automatically detected with ingenious image processing and pattern recognition techniques. Retinal pathologies present an even richer variety of patterns and

features to identify and classify. The automated detection of retinal pathologies is a challenging theme of study since some pathologic features are quite distinctive while many show similarities to others [8].

2.2.2 Understanding Automated Retinal Image Analysis Systems

An automated retinal image analysis system involves different stages of image analysis:

Image Enhancement: some image transformations allow certain features to be more emphasized, corrected or even removed, providing an output image which may be evaluated more easily for certain criteria [35].

Image Segmentation and Object Classification: pixels sharing a certain common properties are segmented from the image and can be associated with symbolic information. Posterior to image segmentation the segmented objects are identified. This stage presents a challenging task and uses pattern recognition techniques, specifically statistical classification, based on knowledge about each feature characteristics. The usual features detected are the blood vessels, the optic nerve and the fovea. The lesions are generally divided into objects brighter (drusens, subretinal fibrosis, amelanotic tumors) or darker (hemorrhage, pigmented scar, melanoma) than the background [35].

Diagnosis: in an automated diagnosis systems all possible diseases manifestations of interest must be coded into the computer knowledge base. Once classification is completed, a list of identified features in the image is reported. Based on the information provided to the computer knowledge base, the system analyzes the combination of features detected and provides a report with a potential diagnose [35].

Chapter 3

State-of-the-Art Methods

Person authentication is the process of verifying that a person is who he is supposed to be. In biometric authentication a physiological characteristic is used to perform that verification. A biometric feature may be useful for authentication if it is unique to the individual and do not suffers modifications over his lifetime. Biometric features that meet these requirements include fingerprints, retinal scans and iris scans [36]. It is important to note that person authentication differs from person identification. While in person authentication methods it is required to know in advance information about who the individual supposedly is in order to verify the veracity of that fact, in a person identification problem no information is associated to the biometric feature considered and the aim is the individual identification.

The use of the retina as a biometric feature presents certain advantages when compared to other biometric features. Retinal scan captures the blood vessels pattern of the eye. Retinal patterns are different for right and left eyes and are unique even for identical twins. The retina, being an inner layer of the eye, is extremely unlikely to be distorted by any environmental or temporal condition [37].

The main goal of this work is the improvement of an automated diagnosis system of eye pathologies using retinal images, more specifically the identification of database images belonging to a same patient and eye by solely considering image features. In this way, these images were used as biometric patient and eye identifiers. Although several retina recognition methods exist, which enable to verify if two retinas belong to the same individual and eye, no previous work was found on methods to search in large retinal image database, looking for retinal images belonging to the same patient and eye as those of an input image. However, related image retrieval methods were carefully studied in order to obtain a better insight about the techniques used in similar contexts.

3.1. Retina-based Person Identification Methods

Retinal images were used for person identification purposes since two famous studies were published. In 1935 Dr. Carleton Simon and Dr. Isodore Goldstein [38] confirmed the

uniqueness of blood vessels pattern of the retina and later in the 1950s Dr. Paul Tower [39] discovered that even among identical twins the uniqueness of the blood vessel patterns is maintained. Since then several methods were studied in order to enable a correct matching of retinal images belonging to the same person and reject otherwise.

The first solution of a person identification procedure based on retinal images was carried out by R. Hill in 1999 [40] and used features based on the Fourier spectrum of the retinal image. The similarity between two retinal images was measured by computing the correlation between their (Fourier) features. Other more modern approaches still use correlation although with a different preprocessing procedure. For instance Kabir et all published an approach [41] in 2004 in which a person identification is based on the color centroid of each retinal image radial segment obtained by dividing the image according to a polar grid. The match between the input image and a database image is evaluated by computing a feature correlation value. The results claimed by the authors were an indicator of this method robustness.

A novel retina feature, named Wavelet Energy Feature (WEF), was introduced in 2007 by Shahnazi et all [42]. Based on wavelets, a powerful tool of multi-resolution analysis, WEF describes the wavelet energy distribution of the blood vessels with different thickness and width in several directions at different wavelet decomposition levels. For each level, the wavelet energy in the horizontal, vertical and diagonal direction is computed. By dividing the retinal image into some non-overlapping blocks and then computing the previous energies for each one a feature vector is obtained for each retinal image. Similar retinal images, i.e. retinal images belonging to the same individual and the same eye, will have similar WEF vectors, which is illustrated in Figure 9. In a person identification procedure, in a training phase all training samples available of the same retina are captured and a template is obtained by averaging the corresponding WEFs. At the recognition phase, WEF of the query retina is compared with all the previously computed templates and the most similar one is selected. The authors reported a 100% recognition rate in a database with 400 images of 40 different people.
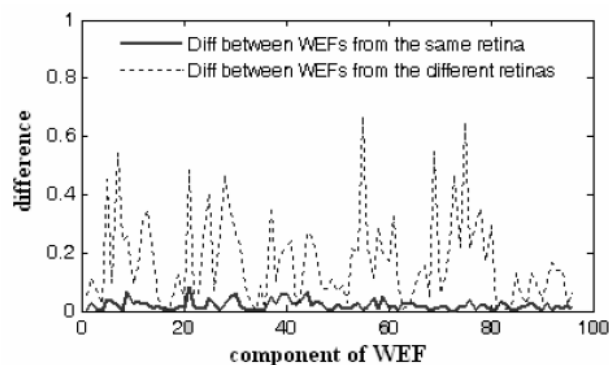
Figure 9: The connected line represents the differences between WEFs of retinal images belonging to the same individual while the dash line corresponds to the differences between WEFs of retinal images from different individuals [42].

Based on the work done by Shahnazi et all in 2007, Farzin et all [43] developed an approach in which the optical disk is initially located in the retinal image and a circular region of interest (ROI) around it is selected. Using a polar transformation, a rotation invariant template was created for the ROI. The template is then analyzed at three different scales using wavelet transformations to separate vessels according to their diameter size. In the last step, vessels position and orientation at each scale are used to define a feature vector for each

database image. For feature matching the authors introduced a modified correlation measure to obtain a similarity score for each scale of the feature vector. The final similarity score was computed by summing scale-weighted similarity scores of the different scales. The experimental results using this method on a database of 300 retinal images obtained from 60 people demonstrated an average error rate equal to one percent.

A different approach was published by S. Sukumaran and Dr. M. Punithavalli [44]. The authors used Fractal Dimension computation of the retinal image blood vessels based on a box counting algorithm. In order to obtain a more robust representation of the retina, the image was divided into subregions and for each one a feature vector was obtained. The correlation computed between the feature vectors of each pair of images was used for matching evaluation and a nearly 96% recognition rate was obtained. Although the recognition rate was lower than in other person identification methods, it was proved in 2009 by Zhao et all [45] that Fractal Dimension is an effective image feature computation procedure for image retrieval in large databases.

A different person identification approach was produced in 2008 by K. Fukuta et all [46]. The authors proposed a person identification method based on the retinal vessel tree images. In the first step, registration between the input image and each database reference image of each person was performed, including translational and rotational displacements between the pair of images. This method was based on the measure of similarity between retinal tree vessel images generated from the input and the reference image. The similarity measure is defined as the cross-correlation coefficient calculated from the pixel values. When the similarity is greater than a predetermined threshold, the image is identified as belonging to the same person of the reference image. With 162 retinal fundus images belonging to 41 different people, the authors presented a false rejection rate and a false acceptance rate of 9.9×10^{-5}% and 4.3×10^{-5}%, respectively.

As proved in the aforementioned method, image registration may not only be used to align two different images but may also be used in the similarity evaluation of two retinal images by measuring image similarities and differences of the aligned patterns. Marcos Ortega and Manuel G. Penedo [47] introduced in 2011 a robust method for person identification by retinal image matching using image registration. The retinal image registration is performed by finding a set of landmarks (bifurcations and crossovers of retinal vessel tree) in each image. Person identification is performed based on the number of landmarks matched between the pair of images. The authors take into consideration that due to disease progression or simply due to the conditions of the image acquisition, small differences between images may be accepted. For this reason they use the Global Affine Transformation model to perform image registration which works well in these conditions. The transformation is applied to the candidate image in order to register its landmarks with respect to the corresponding ones in the reference image. The transformation with the highest matching score will be accepted as the best transformation possible. This score is computed using a similarity measure between each pair of corresponding landmarks which takes into consideration the distance between the two landmarks in the two different images. Finally, the registration score will determine whether the retinal images belong to the same person. Figure 10 represents the main steps of this algorithm
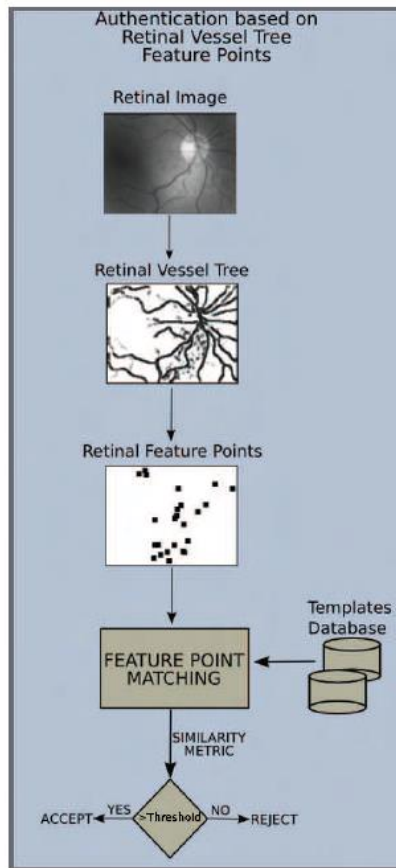
Figure 10: Representation of the main steps of Marcos Ortega and Manuel G. Penedo's [47] person identification method.

When using a person identification method based on image registration it must be ensured that the registration algorithm is efficient, i.e. it successfully registers retinal images belonging to the same patient. SIFT (Scale Invariant Feature Transform) [48] and SURF (Speeded Up Robust Features) [49] image descriptors are widely used in retinal image registration. In the approach developed by Wei et all [50], SIFT keypoints are obtained from the retinal image and a best-bin-first algorithm is applied to identify the corresponding keypoints between two images. The authors claimed a 98.75% success rate of retinal image registration. Similarly, SURF descriptors also proved to be adequate to be applied in retinal images. Cattin et all [51] used SURF keypoints to automatically fuse retinal images, proving that these descriptors perform well in retinal image registration tasks even when using retinal images with no discernible vascularity. In 100 pairs of retinal images this method failed in the registration of 6 pairs, thus showing its good performance.

Salient Feature Regions (SFR) are computed with a saliency measure that consists of both local adaptive variance and gradient field entropy. These regions were introduced by J. Zheng et all [52]. For each SFR region it is computed a local feature descriptor that combines gradient field distribution with the corresponding geometric information. These regions are then used in a cross-correlation-based local rigid registration method with well-aligned region centers as control points. The authors claimed a 95.3% success rate in image registration with this method.
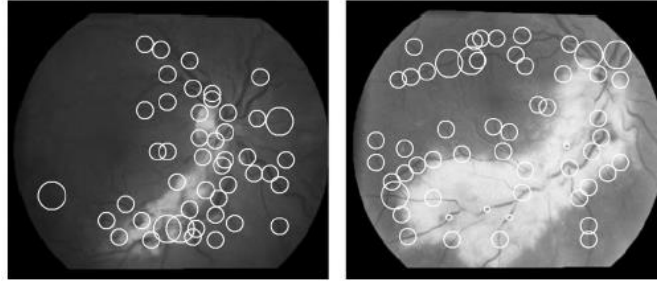
Figure 11: Localization of SFR regions from a pair of retinal images [52].

An approach published in 2012 by Condurache et all [53] uses SIFT feature descriptors to describe anatomical characteristics of the retinal vessel tree. The person identification is conducted with help of a sparse classifier. Sparse-representation based classification looks for the sparsest representation of a test vector in terms of a matrix of training vectors. Depending on the maximal number of examples per class, i.e. the maximal number of available retinal scans belonging to a person, a minimal size of the final feature is established in order to ensure the appropriateness of the sparse-classifier framework to the problem. The authors concluded that apart from high accuracy, around 99% efficiency rate, the proposed algorithm enjoys a set of invariance properties which make it robust to a set of issues affecting retina-based person identification systems.

Frequency analysis of the retinal images was also applied in person identification systems. In the approach proposed by Sabaghi et all [54] the Fourier spectrum of the retinal image is obtained and an angular partitioning of the spectrum is then computed. Initially the optical disc is localized using a template matching technique and it is used for rotating the retinal image into a reference position. This pre-processing step compensates the rotation effects which might occur during the scanning process. The energy of the Fourier spectrum and the sum of the phase angle per partition is used for the feature descriptor composition. The image matching is performed by euclidean distance computation between feature descriptors. The flowchart of this approach is figured in Figure 12.
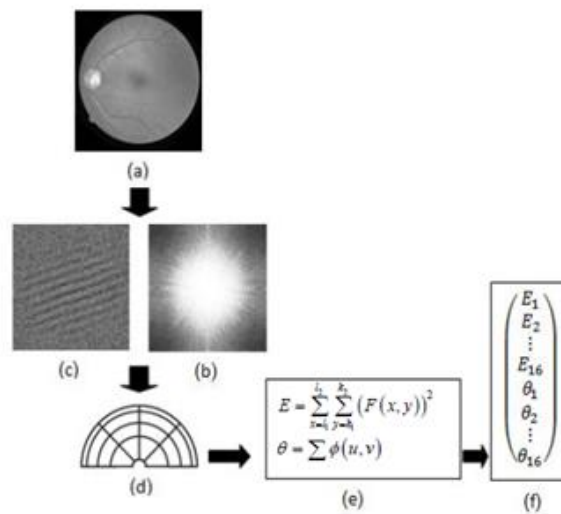
Figure 12: Flowchart for the feature computation process proposed in [54]. (a) Retinal image after rotation compensation (b) Fourier spectrum (c) phase angle (d) partitioning of the Fourier spectrum and the phase angle (e) computation of the Fourier spectrum energy and the sum of the phase angle per partition (f) descriptor vector computation.

Although the described person identification methods present high accuracy in verifying whether two retinal images belong to the same individual, all of them require the comparison of the query retinal image with all the database images, not being a good solution to the subject problem of this thesis.

3.2. Image Database Search Methods

Identifying identical images in an image database may be considered a challenging problem due to the different viewpoint and lighting conditions of the image acquisition. Feature computation has always been the basis of content-based image retrieval. These features may include both text-based features (keywords or annotations for instance) and visual features (color, texture, shape, segmentation, etc.). These features shall be associated with the image content itself to allow a fast and efficient database searching. In order to make image retrieval systems scalable to large databases efficient multi-dimensional indexing techniques were explored. The most popular ones use multi-dimensional indexing techniques including k-d trees, clustering techniques and neural networks [55]. Typically an image feature is represented by a set of non-overlapping regions each represented by a vector computed from the region characteristics. Descriptors are computed for all images in the database and the search of a particular image feature proceeds by nearest neighbor matching of the descriptor vectors of that feature.

In 2003, J. Sivic and A. Zisserman [13] explored whether this type of image retrieval could be recast as text retrieval. In text retrieval [56] the documents are parsed into words. Posteriorly to the rejection of the most common words, each document is represented by a vector with components given by the frequency of occurrence of the words the document contains. The set of vectors representing all documents are organized in an inverted file to facilitate an efficient retrieval which is similar to a book index: each word is followed by a list of all the documents in which the word occurs. In text retrieval the documents with the closest vectors of word frequencies to that of a query text are selected. Adapting this approach to image retrieval J. Sivic and A. Zisserman computed region descriptors (Shape Adapted [57] and Maximally Stable [58] regions) for each image and constructed inverted files similar to the ones used in text retrieval. In order to quantize the descriptors into clusters, which will be the visual "words", a k-means clustering algorithm is applied to the set of region descriptors obtained. Instead of using a simple frequency vector, a weighting function known as "term frequency – inverse document frequency" is computed and used in image retrieval afterwards. In this approach images are ranked according to their similarity.

The use of text retrieval methodology in image retrieval gave rise to the development of efficient image retrieval algorithms. In 2006, D. Nistér and H. Stewénius [59] proposed an image retrieval scheme, based on J. Sivic and A. Zisserman approach [13], which efficiently retrieved images from a large image database by using Vocabulary Trees (VT). In this approach, SIFT descriptors were used to describe the image interest points. Those descriptors were used in a hierarchical k-means clustering in order to build the Vocabulary Tree. An initial k-means process is run on the set of descriptors defining k cluster centers. The same process is then recursively applied to each cluster of descriptors obtained, splitting each cluster into k new groups of descriptors. This process is illustrated in Figure 13.

In the querying phase, a traversal tree path is defined for each descriptor of the input image by comparisons, at each level, of the query descriptor vector with the k cluster centers choosing the closest one. This differs from the original approach of J. Sivic and A. Zisserman where a non-hierarchical visual vocabulary was used. This choice is more appealing in the search step and is more adequate to the subsequent scoring procedures proposed. The scoring procedure proposed is based in inverted files which are associated to each node of the Vocabulary Tree. The inverted files store the index numbers of the images with descriptors represented in a particular node, as well as the frequency of that image's descriptors in the node. The images retrieved are those represented in the terminal nodes of the paths defined by the query image descriptors. The implementation of the scoring system enables the ranking of the images retrieved according to their similarity to the input image. In this way the image with the highest score will be the most similar to the input image.
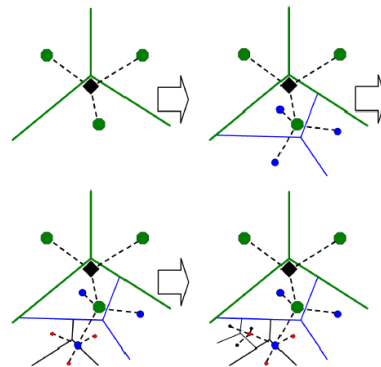
Figure 13: An illustration of the process of building the Vocabulary Tree by descriptor hierarchical quantization as proposed in [59].

Vocabulary Trees have been studied for several years and are still widely used in image retrieval. Such structures typically contain millions of terminal nodes making image retrieval very efficient. Inspired by this database search method several authors proposed different approaches of using Vocabulary Trees for database search [14]. In 2007, Chum et all [60] applied the query expansion to VTs for image retrieval purposes which reissues the initial retrieved results as new queries. In this way additional relevant terms can be added to the query. This method proved to substantially boost retrieval precision although requiring more computational time. The same authors introduced a post spatial verification by RANSAC which re-ranks the results returned. This method improved the search quality comparing to the state-of-the-art methods of that time though it did not show much improvement when used in large VTs [61]. A year later, the authors also proposed another method [62] in which each visual feature is associated to a weighted set of words, allowing the inclusion of features which were lost in the quantization stage of previous systems. It relies on a descriptor association to a combination of visual "words" rather than solely to a single "word" (the terminal node) as in previous approaches. This method of soft-assignment proved to exceed the image retrieval performance obtained by previous works. Figure 14 illustrates the improvement introduced by this approach in finding matching features.
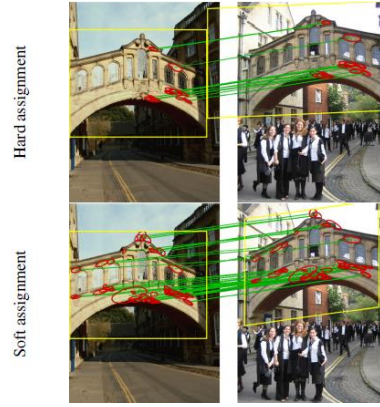
Figure 14: Comparison of matching results with and without soft-assignment technique proposed in [62].

A more recent approach proposed by Wang et all [14] improves the VT based search technique by introducing contextual weighting of local features in both descriptor and spatial domains. This procedure enhances the discriminative power of individual local features with a very small computational overhead. Based on other VT search approaches, this method implements a voting interpretation of the retrieved images from the initial query, i.e. images are ranked accordingly to their similarity to the input image. The scoring function includes a descriptor and a spatial contextual weighting. In the descriptor contextual weighting inverse weighted counts of a node path are considered, down-weighting the importance in image retrieval of nodes with many descriptors. Since local descriptors are not independent and their neighborhood contains important information, a simple statistics in the region surrounding a descriptor is applied to enhance its discriminative ability. The authors proved that this method performance in image retrieval reached 85.15% efficiency, although different results were obtained for the different datasets used.
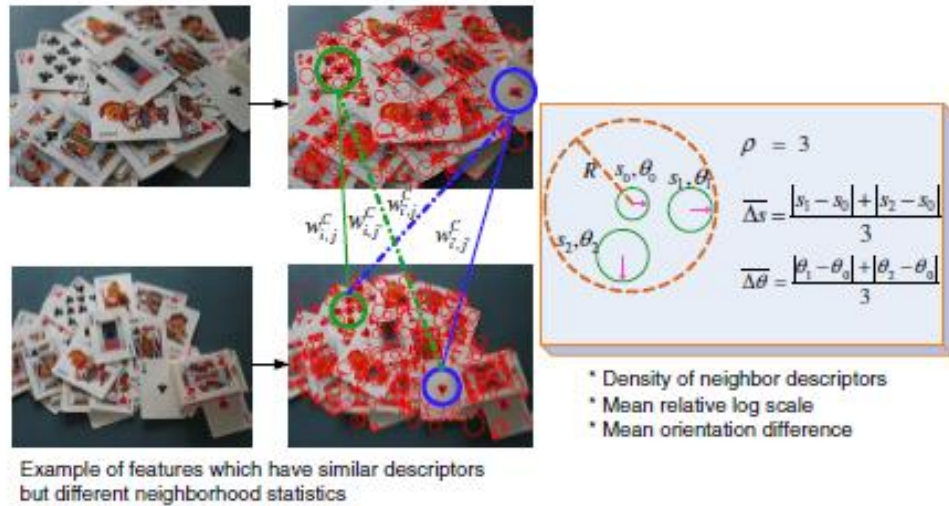
Figure 15: Illustration of the spatial contextual weighting approach proposed in [14]. The red circles represent local descriptors while the green and blue ones refer to the neighborhood of a local descriptor. ρ is the number of local descriptors inside the neighborhood region, Δs is the mean relative log scale of those descriptors and $\Delta \theta$ is their mean orientation difference.

Chapter 4

Methodology

The algorithms studied throughout this work may be divided into two different categories: image feature computation methods and database search techniques using image feature descriptors. Feature computation methods enable the image representation by feature descriptors. In a database search for the identification of the stored images belonging to the same individual and eye as a query retinal image, a subset of images is retrieved. Those images are the most similar ones to the query image. However, the subset of images retrieved contains not only the images sought but also other images belonging to different individuals or eyes, although being very similar to the query retinal image. Considering this, a person identification method based on the retina as biometric feature is used to distinguish the right images from the other images belonging to the subset of images retrieved. It is noteworthy that although the algorithm proposed still uses a retinal identification method, it is only applied to the reduced subset of images retrieved representing a significant computational time reduction. Moreover, as mentioned before this work solely focuses on the selection of the database retinal images most similar to the one being queried.

4.1. Image Descriptors

Four different image feature descriptors were studied: Fractal Dimension (FD), Wavelet Energy Feature (WEF), Scale Invariant Feature Transform (SIFT) and Speeded Up Robust Features (SURF). FD and WEF are image feature descriptors methods used in retina-based person identification systems which attained high efficiency in verifying whether two retinal images belong to the same person and eye. Hence we believe that these methods are appropriate to correctly represent retinal images and their vascular details. Besides this highly advantageous characteristic, both methods use a low complexity algorithm and the image features computed have low dimensionality. SIFT and SURF image descriptors were also studied in this work. From the research carried out, most references of VT in the literature use this type of descriptors in the organization of these search structures. Moreover, the distinctiveness and invariance of SIFT and SURF descriptors to image scale, image rotation, distortions, changes in viewpoint, noise and illumination [48] enable the development of robust algorithms for image retrieval purposes. Taking into consideration that the retinal image datasets used in this work include images with the presence of several eye pathologies characteristic lesions, the study of these methods was imperative as they could be robust to image changes due to the evolution of the pathologies.

4.1.1. Fractal Dimension (FD)

The Fractal Dimension concept is directly related to surface roughness being appropriate to use in content based image retrieval systems as shown in [45] where it is applied to person identification based on retinal images. As mentioned in Chapter 3 – State-of-the-Art Methods, the use of Fractal Dimension computation in retina-based person identification was previously studied by S. Sukumaran and Dr. M. Punithavalli [44]. The authors computed the Fractal Dimension value of image subregions obtaining a feature vector which was used to describe each retinal image.

Fractals are fragmented geometric shapes that can be subdivided into parts where each of those parts is a smaller copy of the original shape. These objects defy conventional concepts, such as length and area, since they exhibit similarity over all scales. For this reason, they are usually expressed by their Fractal Dimension value which expresses how completely a fractal appears when zoomed down to finer scales. Fractal Dimension is a quantity which determines how the fractal differs from a Euclidean object. In Euclidean objects the topological dimension associated to points is 0, to lines is 1, to surfaces is 2 and to volumes is 3. While Euclidean objects are expressed in integral dimensions, fractals are expressed in dimensions with non-integral values. The fractal concept may be better understood with Figure 16, where a Sierpinsky triangle is illustrated. This triangle is constructed excluding the middle triangle A'B'C', which is recursively done resulting in smaller triangles. The final triangle is not a one-dimensional object but it is not as well a bi-dimension object. In fact, it has a 1.58 dimension value, which corresponds to the object Fractal Dimension value [44].
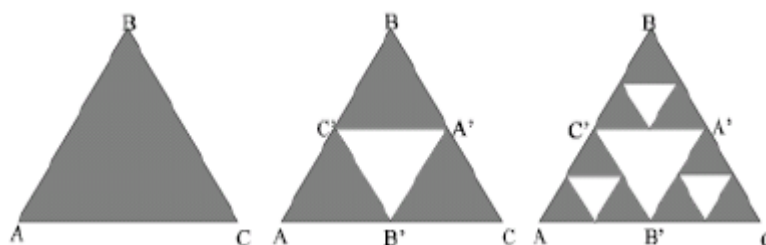
Figure 16: Sierpinsky triangle, a well-known point-set fractal [44].

Several ways of computing the Fractal Dimension of an object have been studied. From those, the box-counting method is one of the most widely used [63]. With this Fractal Dimension computation approach, the image pattern is evaluated in terms of presence of zero and nonzero pixel values. In this way it only makes sense to use it in retinal vessel tree images. The vessel tree computation is performed in the green channel of the retinal image, which was proved to be the channel containing more information about the retina vasculature [64]. Initially, the Curvelet Transform coefficients of this channel were obtained using the Fast Curvelet Trasform algorithm proposed by Candès et all [65], the smallest 10% were set to zero and the inverse Curvelet Transform coefficients were computed. The vascular tree segmentation is performed using the wavelet based approach proposed in 2012 by Bankhead et all [66]. However, other vascular tree computation methods may be used for this purpose. In the vascular tree image the vessels assume pixel value 1 and the remaining pixels have value 0 corresponding to the image background.
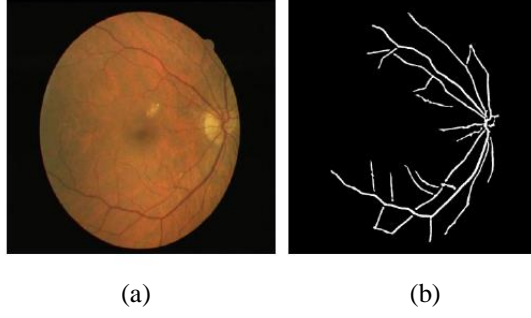
Figure 17: (a) Retinal image and (b) its corresponding retinal vessel tree image computed by using the aforementioned algorithm.

The algorithm of the box counting method starts by partitioning the M×M pixels image into square non-overlapping boxes of size s×s, where 1<s<M. The maximum image partitioning degree is given by $p = \frac{\log(M)}{\log(2)}$, where M corresponds to the number of pixels of the biggest side of the image. Therefore, the square image boxes assume sizes s= 2,4,8...2^p. Figure 18 illustrates different image partitioning degrees in a retinal vessel tree image.
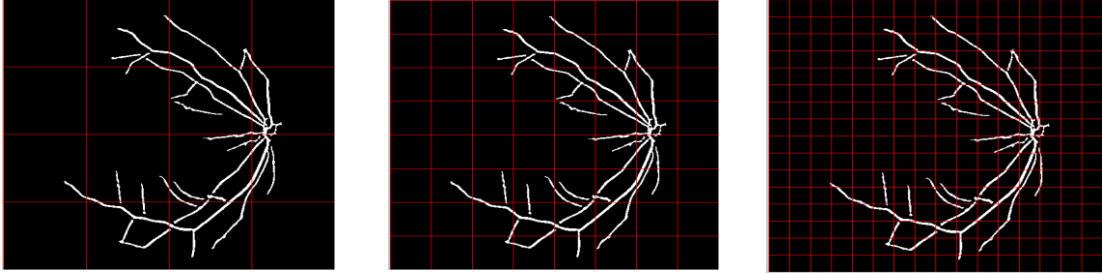
Figure 18: Image sampling with square boxes with size s=2^8 , s=2^7 and s=2^6, respectively.

The existence of nonzero pixels in each box is then verified. N_r represents the number of boxes with at least one nonzero pixel for each partitioning degree r. The partitioning degree is given by $r = \frac{s}{M}$. N_r is computed according to equation (1), where n_r assumes value 1 if the image box (i,j) contains at least one nonzero pixel and 0 otherwise.

$$N_r = \sum_{i,j} n_r(i,j), \tag{1}$$

Posteriorly to the computation of different N_r values for different partitioning degrees, the Fractal Dimension value of the image considered is given by the equation (2).

$$D = \frac{\log(N_r)}{\log(1/r)}, \tag{2}$$

Equation (2) may be interpreted as being the slope of the linear regression fitting of the plotting of $\log(N_r)$ along $\log(1/r)$, which is illustrated in Figure 19.

Plotting of log(Nr) along log(1/r)
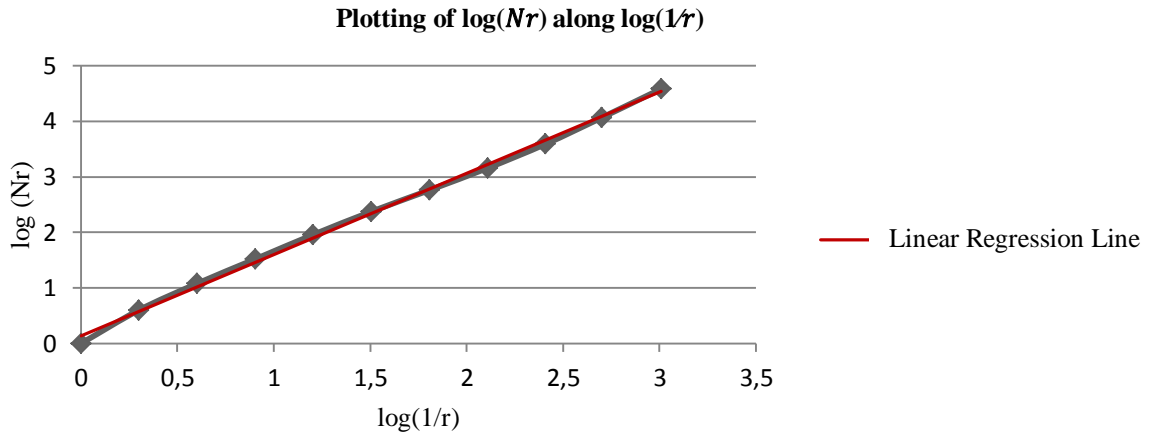
Figure 19: Plotting of log(Nr) along log(1/r).

The slope of the linear regression line represented in Figure 19 is 1.461, and it represents the Fractal Dimension of the vascular tree estimation image illustrated in Figure 17(a).

Representing an entire image by a single value does not enable an efficient representation of the image. In fact, Fractal Dimension values obtained for different retinal images were, as expected, very similar. For this reason and following [44] approach, the retinal image was divided into subregions and for each one the Fractal Dimension value was computed. To summarize, two different image partitioning types are performed: the retinal image is initially partitioned into subregions and for each of one the Fractal Dimension value is computed; when applying the box counting algorithm for Fractal Dimension computation of the image subregion, the partitioning of the image subregion into boxes is performed.
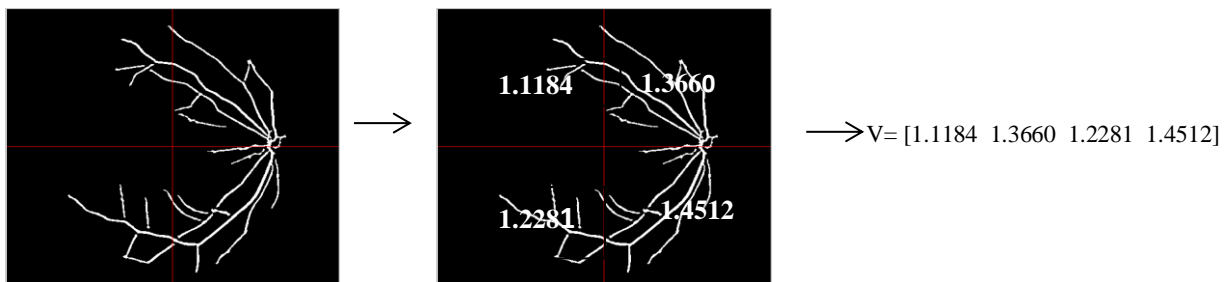
Figure 20: Illustration of this method of Fractal Dimension computation by image partitioning. The retinal vessel tree image was divided into four non-overlapping boxes and for each one a Fractal Dimension value was computed. The image may now be represented by a four-element descriptor vector V as illustrated.

From the research carried out, it was found that Fractal Dimension was also applied in image classification, more specifically, Hong et al. [67] used this feature computation method for X-ray medical images classification. Although the algorithm used by the authors for the Fractal Dimension computation was similar to the one described before, it differs in one aspect: besides using the original image for feature descriptor computation, the authors computed five transformed images in order to obtain more information about directional details. The original image is represented by I(i,j), where i and j are the coordinates of each image pixel. I1, I2, I3

and I4 , which are expressed by equations (3) to (6), represent the transformed images in the horizontal, vertical, diagonal and anti-diagonal directions, respectively.

$$I1(i,j)=|I(i,j)-I(i-1,j)| \qquad (3)$$

$$I2(i,j)=|I(I,j)-I(I,j-1)| \qquad (4)$$

$$I3(i,j)=|I(i,j)-I(i-1,j-1)| \qquad (5)$$

$$I4(i,j)=|I(i,j)-I(i-1,j+1)| \qquad (6)$$

Besides this, the image roughness, I5, can be expressed by equation (7).

$$I5(I,j) \begin{cases} I(i,j) - Lmin, if\ I(i,j) > Lmin \\ 0, others \end{cases} \qquad (7)$$

Lmin=min(I(i,j)+1/2*mean value of the original image
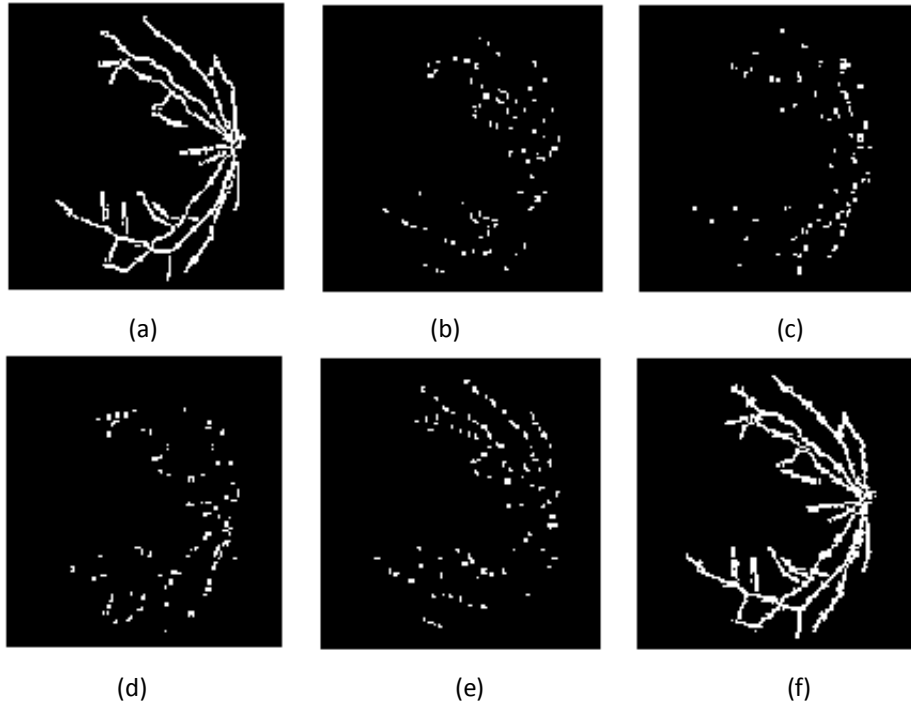
Figure 21: (a) Original vessel tree image, (b) transformed image I1, (c) transformed image I2, (d) transformed image I3, (e) transformed image I4, (f) transformed image I5.

This approach was followed when computing the Fractal Dimension descriptors of retinal vessel tree estimation images since it enables a better representation of vessel details rather than solely using the original image. By using the transformed images approach, each image subregion is represented by six Fractal Dimension values instead of a scalar vector, i.e. each subregion is described by a descriptor vector V=(x1, x2, x3, x4, x5, x6) where x1, x2…x6 are the Fractal Dimension values of the original image and of each transformed image obtained. The flowchart of the computation of this feature may be found in Appendix A.1.

4.1.2. Wavelet Energy Feature (WEF)

Wavelets provide a detailed multiscale analysis of a sequence or function, providing another perspective of analysis. Wavelet analysis enables the simultaneous localization of an event in time, defined by the wavelet localization, and in scale, defined by its shape. The continuous Wavelet Transform F(a,b) of a function f(t) is defined by

$$F(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t)\psi\left(\frac{t-b}{a}\right) dt \qquad (8)$$

The function $\psi_{a,b}$ is called the mother wavelet and is adjusted by parameters a and b which correspond to the wavelet scale and time shift respectively. In the discrete domain these parameters are discretized, as well as the mother wavelet [68]. Since this dissertation deals with image processing techniques only the discrete domain is considered.

The discrete wavelet analysis of an image yields the computation of subimages which are sets of coefficients that represent horizontal, vertical and diagonal oriented image details. These subimages contain information which enables a perfect reconstruction of the original image. For each decomposition level J, the image may be represented by 3J+1 subimages: the approximation coefficients of the last decomposition level (A_J) and the horizontal (H_i), vertical (V_i) and diagonal (D_i) coefficients of each decomposition level performed ({H_i,V_i,D_i}, i=1,...,J).

(a) (b)

(c) (d)
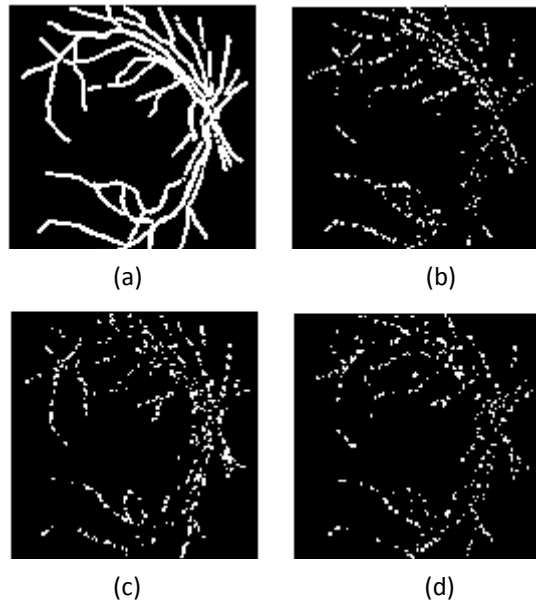
Figure 22: (a) Approximation, (b) Horizontal, (c) Vertical and (d) Diagonal coefficients subimages.

Subimages are obtained by filtering the original image using both a low-pass and a high-pass filter, which are applied in two different phases. First, the low-pass and the high-pass filter are applied in the original image resulting in the image approximation and the detail coefficients computation respectively. Posterior to this, the same set of filters are applied to both the approximation and detail coefficients. Filtering the subimage of approximation coefficients with a low-pass filter results in the computation of another subimage of approximation coefficients. When filtering the subimage of approximation coefficients with a high-pass filter, the subimage of horizontal coefficients is computed. By applying the same set of low-pass and high-pass

filters to the detail coefficients subimage, the vertical and diagonal coefficients are computed. A better understanding of this procedure is possible when analyzing the flowchart in Figure 23.
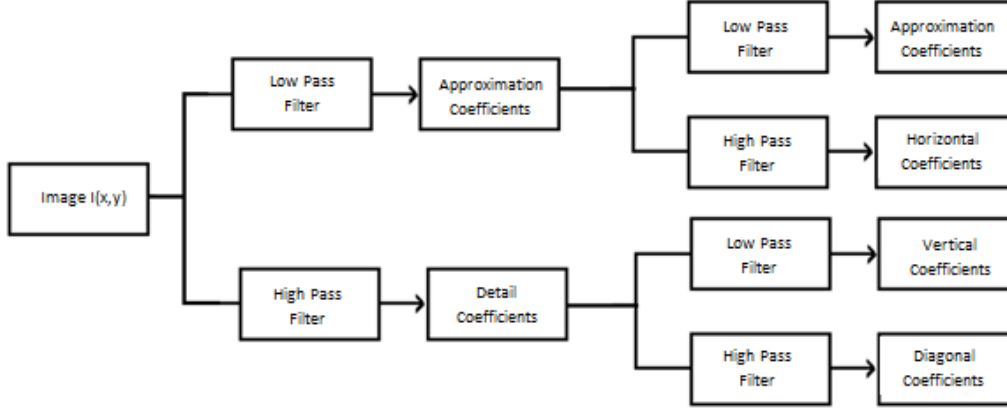
Figure 23: One wavelet decomposition level of image I(x,y).

Wavelet families are sets of different filters that may be used in image decomposition by wavelets. All of them enable a perfect reconstruction of the analyzed image by considering the approximation and detail coefficients generated [42]. The most used wavelet families are Haar, Morlet, Daubechies, Symlets and Coiflets wavelets.

The retina feature computation method based on image wavelet analysis considered in this work was based on the person identification procedure proposed by Shahnazi et all [42]. The authors computed horizontal (E^h), vertical (E^v) and diagonal (E^d) energies according to equations (9) to (11), where M and N correspond to the subimage width and height, respectively.

$$E_i^h = \sum_{x=1}^{M} \sum_{y=1}^{N} (H_i(x,y))^2 \qquad (9)$$

$$E_i^v = \sum_{x=1}^{M} \sum_{y=1}^{N} (V_i(x,y))^2 \qquad (10)$$

$$E_i^d = \sum_{x=1}^{M} \sum_{y=1}^{N} (D_i(x,y))^2 \qquad (11)$$

By using this image analysis technique with retinal vessel tree images, the energies reflect the blood vessel details, the main biometric feature of retinal images which is the crucial aspect in a person identification system.

Similarly to the procedure used in the Fractal Dimension descriptors computation, the retinal vessel tree image was partitioned into image subregions. For each image subregion the descriptor $\{E_i^h, E_i^v, E_i^d\}$ is computed. The number of descriptors per image depends on the image partitioning degree. For this reason different image subregion sizes as well as different wavelet families and decomposition levels were studied in order to find the most suitable combination which most efficiently describes the retinal images used in this work. This aspect is further explored in Chapter 5- Results and Discussion and a flowchart of the computation algorithm of this method may be seen in Appendix A.2.

4.1.3. Scale Invariant Feature Transform (SIFT)

In the state-of-the-art literature, several image retrieval methods use SIFT feature descriptors. The distinctiveness and invariance of these descriptors to image scale and rotation have proved to provide a robust image representation across a substantial range of distortions, changes in viewpoint, noise and illumination [48]. In retinal image registration these descriptors enable the identification of identical features in different images which is a crucial step to correctly align them. It is possible to infer whether two retinal images belong to the same individual by evaluating the number of correct matches found between them. This fact was the main motivation for exploring SIFT feature computation method in this dissertation.

The SIFT descriptor computation was introduced by Lowe [48] and the procedure used in this thesis was based on his work. This descriptor computation method was applied to a luminance image of the retinal images, which was normalized to a range pixel value between 0 and 1.

Detection of Scale-Space extrema

Scale-Space extrema are locations that can be identified under different views of the same image. Their detection is performed by searching from stable features across all possible scales. The image scale space $L(x,y,\sigma)$ of an image may be defined as the result of the convolution of a variable-scale Gaussian function $G(x,y,\sigma)$, which was proved by Lindeberg in [69] to be the most suitable one for this purpose, with the image $I(x,y)$.

$$L(x,y,\sigma)=G(x,y,\sigma)*I(x,y) \quad \text{with} \quad G(x,y,\sigma)=\frac{1}{2\pi\sigma^2}e^{-(x^2+y^2/2\sigma^2)} \qquad (12)$$

The scale space extrema in $D(x,y,\sigma)$, the difference-of-Gaussian function convolved with the image, which is computed from the difference of two nearby scales separated by a constant multiplicative factor k, are identified as stable keypoints.

$$D(x,y,\sigma)=(G(x,y,k\sigma)-G(x,y,\sigma))*I(x,y)=L(x,y,k\sigma)-L(x,y\sigma) \qquad (13)$$

Each octave of scale space is divided into an integer number s of intervals, so that $k=2^{1/s}$. For each octave s+3 images are computed so that the final extrema detection covers a complete octave. Adjacent image scales are then subtracted, producing the difference-of-Gaussian images. As for the local extrema detection, each sample point is compared to its eight neighbors in the current image and nine in the scales immediately above and below. The sample point is only selected if it is larger or smaller than all of those neighbors, constituting a local maximum or local minimum respectively. Besides this, maxima are only considered in the next steps if larger than a specified threshold value. Lowe's choice for this value was 0.01, however throughout this dissertation this value was studied in order to obtain different number of descriptors for the same image, exploring its contribution in person identification procedures. As it may be easily proven, smaller threshold values enable the computation of more feature descriptors.
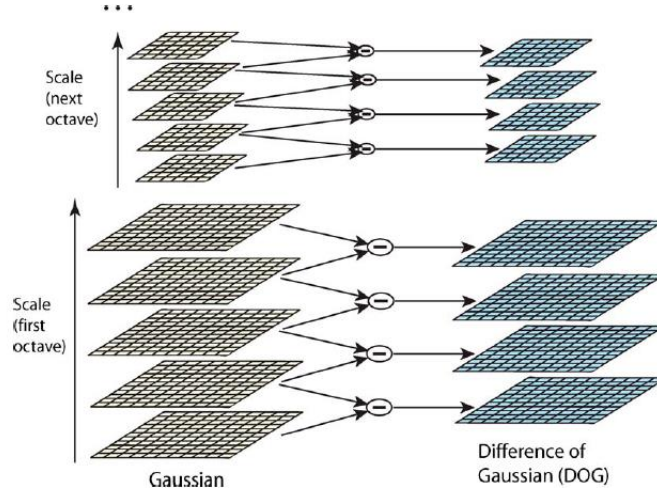
<u>Figure 24:</u> Representation of the set of scale space images (on the left), which are computed by repeatedly convolving the original image with a Gaussian function, and Difference-of-Gaussian images, which are the result of the subtraction of adjacent Gaussian images [48].

<u>Accurate keypoint location</u>

Once a keypoint is detected, a fit to the closest data is performed for its location, scale and orientation computation. This procedure allows the rejection of points that have low contrast or are poorly localized along the edge. A 3D quadratic function is fitted to the local sample points to determine the interpolated location of the maximum or minimum. This approach uses the Taylor expansion of the scale space function $D(x,y, \sigma)$, shifted so that the origin is at the sample point. This is expressed by equation (14), where D and its derivatives are evaluated at the sample point and $x=(x,y, \sigma)^{\mathrm{T}}$ is the offset from this point.

$$D(x) = D + \frac{\partial D^T}{\partial x} x + \frac{1}{2} x^T . \frac{\partial^2 D}{\partial x^2} x \tag{14}$$

The location of the extremum $\hat{x}$, is determined by taking the derivative of this function with respect to x and setting it to zero giving $\hat{x} = -\frac{\partial^2 D^{-1}}{\partial x^2} . \frac{\partial D}{\partial x}$. Derivatives of D are approximated by using differences of neighboring sample points. The resulting 3x3 linear system can be solved with minimal computational cost. If $\hat{x}$ is larger than 0.5 in any dimension it means it lies closer to other sample point and interpolation is performed around it. The final offset $\hat{x}$ is added to the location of its sample point to get the interpolated estimation for the location of the extremum. To discard unstable extrema it is considered the $D(\hat{x})$ function value at the extremum, which is calculated by equation (15).

$$D(\hat{x}) = D + \frac{1}{2} . \frac{\partial D^T}{\partial x} \hat{x} \tag{15}$$

If $D(\hat{x})<0.03$ the point is discarded (assuming that image pixels are in the range value [0,1].

The difference-of-Gaussian function will have a strong response along edges, which is an important aspect to be properly analyzed. A poorly defined peak in the difference-of-Gaussian function will have a large principal curvature across the edge but a small one in the perpendicular direction. Principal curvatures are computed from a 2x2 Hessian matrix, given by equation (16), computed at the keypoint location and scale.

$$H=\begin{bmatrix} Dxx & Dxy \\ Dxy & Dyy \end{bmatrix} \qquad (16)$$

The derivatives are estimated by taking differences of neighboring sample points. Eigenvalues are proportional to the principal curvatures of D. Since what is important is only the eigenvalues ratio, the eignvalue calculation can be avoid. Let α be the eignvalue with the largest magnitude and β the smallest. The trace and product from the determinant of H may be computed in the following way:

$$\text{Tr}(H)=Dxx+Dyy= \alpha + \beta \qquad (17)$$

$$\text{Det}(H)=Dxx.Dyy-(Dxy)^2== \alpha\beta \qquad (18)$$

If the Det(H) is negative, the curvatures have different signs and so the point is discarded. Let r be the ratio between α and β, so that $\alpha=r\beta$, then:

$$\text{Tr}(H)^2=\frac{(\alpha+\beta)^2}{\alpha\beta}=\frac{(r\beta+\beta)^2}{r\beta^2}=\frac{(r+1)^2}{r} \qquad (19)$$

This quantity is at minimum when the two eignvalues are equal and it increases with r. In this way, to check whether the ratio of the principal curvatures is below a threshold r, it is only necessary to verify if $\frac{Tr(H)^2}{Det(H)} < \frac{(r+1)^2}{r}$. The experiments performed use a value of r=10 which eliminates keypoints that have a ratio between principal curvatures greater than 10.

Orientation Assignment

Another important feature for the keypoint correct description is the orientation assigned to it. By being assigned an orientation to the keypoint, it achieves invariance to image rotation. The scale of the keypoint is used to select the Gaussian smoothed image L(x,y, σ), with the closest scale, so that all computations are performed in a scale invariant way.

For each image sample L(x,y) at this scale, the magnitude m(x,y) and orientation ɵ(x,y) gradients may be computed using pixel differences, as may be seen in equations (20) and (21).

$$m(x,y) = \sqrt{(L(x + 1, y) - L(x - 1, y))^2 + (L(x, y + 1) - L(x, y - 1))^2} \qquad (20)$$

$$\theta(x,y) = \tan^{-1}((L(x, y + 1) - L(x, y - 1))/(L(x + 1, y) - L(x - 1, y))) \qquad (21)$$

An orientation histogram is built from the orientation gradient of sample points within a region around the keypoint. The orientation histogram has 36 bins covering the 360 degree range of orientations. Each sample added to the histogram is weighted by its magnitude gradient and by a Gaussian-weighted circular window with a σ that is 1.5 times greater than the one of the keypoint scale. Histogram peaks correspond to dominant directions of local gradients. The highest peak is detected and any other local peak that is within 80% of the highest peak is used as well in the keypoint description. In this way, for the locations with multiple peaks with similar magnitude there will be multiple keypoints differing only in orientation. A parabola is then fitted to the three histogram values closest to each peak to interpolate the peak position for better accuracy. With these characteristics, SIFT features are resistant to even large amounts of pixel noise. The major cause of error in those conditions is the initial location and scale detection.

<u>Local Image Descriptor</u>

The previous steps assign to each keypoint a location, a scale and an orientation, providing a coordinate system in which to describe the local image region. The approach applied to compute the final SIFT descriptor was based on the one demonstrated by Edelman et all [70], differing only in the fact that this one allows for positional shift using a different computational mechanism.

Initially, image gradient magnitude and orientation are sampled around the keypoint location, using the scale of the keypoint to select the level of Gaussian blur for the image. The coordinates of the descriptor and the orientation gradient are rotated relative to the keypoint orientation in order to achieve invariance. Posteriorly, a Gaussian weighting function with $\sigma = \frac{1}{2} \times$(width of the descriptor window) is used to assign a weight to the magnitude of each sample point. This avoids sudden changes in the descriptor with small changes in the position of the window and gives less emphasis to gradients that are far from the center of the descriptor, since these are most affected by misregistration errors. The descriptor allows for significant shift in gradient positions by creating orientation histograms over 4x4 sample regions.

It is important to avoid all boundary effects in which the descriptor abruptly changes as a sample shifts smoothly from being within one histogram to another or from one orientation to another. Therefore, trilinear interpolation is used to distribute the value of each gradient sample into adjacent histogram bins. In other words, each entry into a bin is multiplied by a weight of 1-d for each dimension, where d is the distance of the sample from the central value of the bin measured in units of the histogram bin spacing. The descriptor is formed by a vector containing the values of all the orientation histogram bins. In this work, the descriptor is computed from a 4x4 array of histogram with 8 orientation bins in each, as illustrated in Figure 25, which result in 4x4x8=128 element feature vector for each keypoint.

Image gradientes   Keypoint descriptor
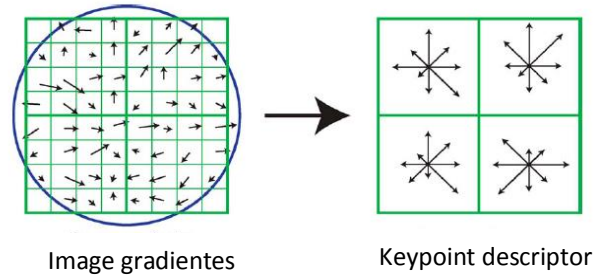
<u>Figure 25:</u> First it is computed the gradient magnitude and orientation at each image sample point in a region around the keypoint location (on the left). These are weighted by a Gaussian window, indicated by the overlaid circle. These samples are then accumulated into orientation histograms summarizing the contents over 4x4 subregions, as shown on the right, with the length of each arrow corresponding to the sum of the gradient magnitudes near that direction within the region. This figure shows a 2×2 descriptor array computed from an 8×8 set of samples, whereas the experiments in this work use 4×4 descriptors computed from a 16×16 sample arrays [48].

Finally, the feature vector is modified to reduce the effects of illumination change. First it is normalized to unit length, making the descriptor invariant to image contrast changes. Besides this it is also invariant to brightness change since the gradient values are computed from pixel differences. Non-linear illumination changes may cause a large change in relative magnitudes

for some gradients. Therefore, the influence of large gradient magnitudes is reduced by thresholding the values in the unit feature vector to be no larger than 0.2 and then renormalizing to unit length. In this way, orientation distribution has a greater emphasis.

The descriptor complexity may be adjusted by the number of orientations represented in the histogram and the width, n, of the $n \times n$ orientation histograms. As obvious, the more complex is the descriptor, the more discriminative it will be. On the other side, as complexity increases the descriptor becomes more sensitive to shape distortions and occlusions. The complexity degree was studied experimentally by analyzing the descriptor performance in enabling correct matches between images. It was proved that the best results were obtained with a 4x4 array of histograms with eight orientations each.

Figure 26 illustrates the location of SIFT descriptors computed from a retinal image. As previously mentioned, the threshold value which discards the maxima values of the scale space in the initial steps to select keypoint candidates was studied in order to choose one that properly enables the identification of retina vessels details in the image. The effect of this parameter in SIFT descriptors computation is further described in Chapter 5- Results and Discussion. The SIFT descriptors represented in Figure 26 were computed using a threshold value equal to 1, which was concluded to be the most suitable one for the images used in this work. As it may be analyzed, SIFT descriptors are mainly localized in the retina vessel tree, enabling an efficient representation of the retinal image details for a person identification procedure.
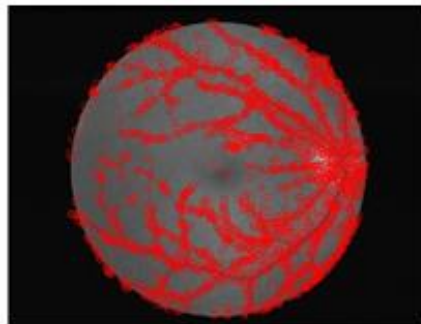
Figure 26: SIFT feature descriptors computed from a retinal image by using a threshold value 1 to discard the maxima values of the scale space in keypoint identification.

A flowchart of the main steps required for SIFT descriptor computation is present in Appendix A.3.

4.1.4. Speeded Up Robust Features (SURF)

The SURF descriptor computation method was based on the work developed by Lowe (SIFT descriptor computation method), using similar properties. As in SIFT descriptors computation, this approach consists of computing a keypoint descriptor in such a way that it makes that feature reproducible even in images with very different characteristics. The major steps of SURF descriptors computation are further described below.

<u>Keypoint Detection</u>

The keypoint detection is performed using a basic Hessian matrix approximation, which lends itself to the use of integral images. This enables a computational time reduction. Integral images allow for fast computation of box type convolution filters. In integral images $I_\Sigma(x)$ at the location x=(x,y)^T it is represented the sum of all pixels in the input image I within a rectangular region formed between the origin and x. In this way, by using integral images it only takes three additions to calculate the sum of intensities over any rectangular upright area, as illustrated in the figure below.
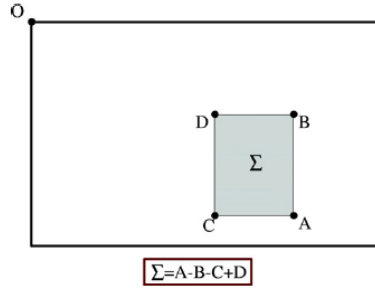
Figure 27: Illustration of the pixel intensity sum of the image inside a rectangular region, taking only three additions by using integral images [49].

The keypoint detector is based on the Hessian matrix since it has been proved to have a good performance in accuracy. Blob-like structures are detected at locations where the determinant is maximum. Given a point x=(x,y) in an image I, the Hessian matrix H(x,σ) in x at scale σ is defined as follows:

$$H(x,\sigma)=\begin{bmatrix} L_{xx}(x,\sigma) & L_{xy}(x,\sigma) \\ L_{xy}(x,\sigma) & L_{yy}(x,\sigma) \end{bmatrix} \tag{22}$$

$L_{xx}(x,\sigma)$ is the convolution of the Gaussian second order derivative $\frac{\partial^2}{\partial x^2}g(\sigma)$ with the image I in point x, and $L_{xy}(x,\sigma)$ and $L_{yy}(x,\sigma)$ are similarly computed. For scale space analysis, box filters are used, approximating second order Gaussian derivatives, as shown in the figure below.

(a) (b) (c) (d)
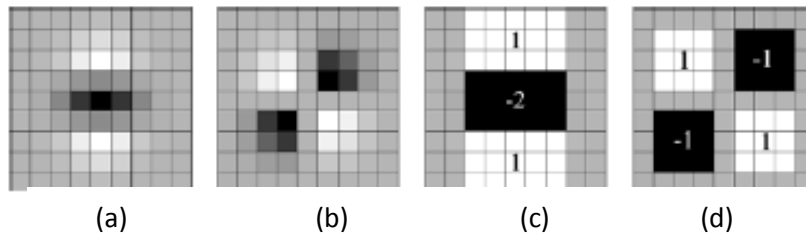
Figure 28: (a) Discretized Gaussian second order partial derivative in y (Lyy) and (b) xy (Lxy) direction; (c and d) box filtering approximations of (a) and (b) [49].

The box filtering approximations will be denoted by Dxx, Dyy and Dxy and so the determinant of the Hessian Matrix is computed by equation (23), where w is the weight of the filter responses, used to balance the expression for the Hessian's determinant.

$$Det(H_{approx})=D_{xx}D_{yy}-(wD_{xy})^2 \tag{23}$$

The approximated determinant of the Hessian represents the blob response in the image at location x. These responses are stored in a blob response map over different scales, from which local maxima are detected.

As in SIFT approach, the keypoints have to be found at different scales to ensure that a correct matching between two images may be found even when those images differ in several conditions such as lighting, viewpoint and noise. The images are repeatedly smoothed with a Gaussian, followed by a sub-sampling procedure. These images are usually represented in pyramids, in which the highest levels correspond to the most smoothed images. In SIFT approach the pyramid layers are subtracted in order to obtain the DoG (Difference of Gaussians) images where edges and blobs can be found. In this approach, instead of iteratively applying the same filter to the output of a previous filtered layer as in SIFT, box filters of any size can be directly applied on the original image. In this way, the scale space is analyzed by up-scaling the filter size rather than iteratively reducing the image size. This is represented in the image below. The initial box filter size used is a 9x9 filter, corresponding to the initial scale layer, which will be referred as scale s=1.2 (approximating Gaussian derivatives with σ=1,2). The following layers are obtained using successive bigger box filters. The main reason of this alternative procedure is computation efficiency.
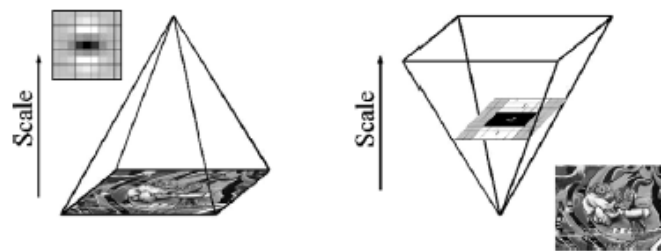
Figure 29: (a) Representation of the SIFT approach of iteratively reducing the image size; (b) Representation of SURF approach using integral images which enable the up-scaling of the filter [49].

The scale space is divided into octaves. An octave represents a series of filter response maps obtained by convolving the same input image with a filter of increasing size. In total, an octave encompasses a scaling factor of 2, which implies that one needs to more than double the filter size. Each octave is subdivided into a constant number of scale levels.

For accurate keypoint localization, filter responses are thresholded such that all values below a specific threshold are not considered. This threshold value was experimentally chosen in this dissertation as it may be seen in Chapter 5 – Results and Discussion. By increasing the threshold value the number of descriptors computed reduces, leaving the most stable ones for the further steps. Each pixel in the scale-space is compared to its 26 neighbors (8 pixels in the native scale space and 9 in each of the scales above and below). Those larger or smaller than all their corresponding neighbors are selected as keypoint candidates. A 3D quadratic fitting is used to accurately determine each keypoint space localization.

Orientation Assignment

It is crucial to assign an orientation to each keypoint to provide rotation invariance to the descriptors computed. For descriptor orientation computation, the Haar-wavelet responses in x and y direction are computed in a circular neighborhood of radius 6s around the interest point, where s is the scale at which the keypoint was detected. Once the wavelet responses are

computed and weighted with a Gaussian (σ=2.5s) centered at the keypoint, the responses are represented as vectors in a space with the horizontal response strength along the abscissa and the vertical response strength along the ordinate. To compute the dominant orientation it is used a sliding window through the responses space covering an angle of π/3. The horizontal and vertical responses within the sliding window are summed, yielding a new vector. The longest vector represents the keypoint orientation.

Descriptor Components

For the descriptor computation, a square region of size 20s is centered around the keypoint and oriented along the dominant orientation computed previously. The region is split up into smaller 4x4 square subregions, keeping important spatial information in. For each subregion a few simple features at 5x5 regularly spaced sample points are computed. For simplicity, dx refers to the Haar wavelet response in the horizontal direction in relation to the orientation assigned to the keypoint and dy refers to the response in the vertical direction. To provide robustness to geometric deformations and localization errors, the responses dx and dy are first weighted with a Gaussian (σ=3.3s) centered at the keypoint.

Then, the wavelet responses dx and dy are summed up over each subregion and form a first set of entries to the feature vector. In order to bring in information about the polarity of the intensity changes, the sum of the absolute values of the responses, |dx| and |dy|, is also computed. Hence, each subregion has a four-dimensional descriptor vector v for its underlying structure $v = (\sum dx, \sum dy, \sum |dx|, \sum |dy|)$. This results in a descriptor vector for all 4×4 sub-regions of length 64, since it was the approach proved to present the best results. The wavelet responses are invariant to different illumination characteristics and invariance to contrast (a scale factor) is achieved by converting the descriptor into a unit vector. Considering finer subdivisions appeared to be less robust and increased descriptor matching timed.
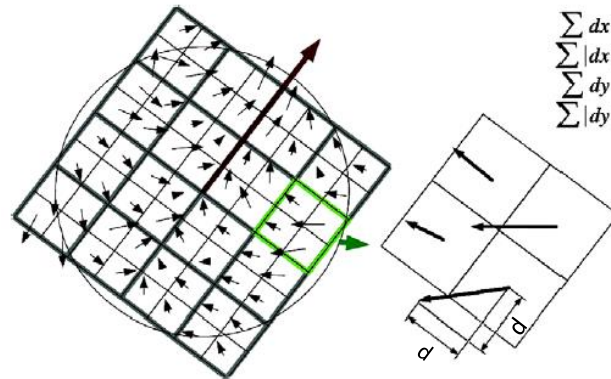
Figure 30: An oriented quadratic grid with 4 x4square sub-regions is laid over the interest point (left). For each square, the wavelet responses are computed from 5 x5 samples , which is represented in the figure by 2x2 sub-divisions for illustrative purposes. For each field, the sums dx, |dx|, dy, and |dy| are computed relatively to the grid orientation (right) [49].

Figure 31 illustrates the location of SURF descriptors computed in a retinal image by using a 1×10^{-5} threshold value to discard the filter response values below this value in the initial keypoint identification step. In Chapter 5 - Results and Discussion, the influence of different threshold values in descriptor computation may be analyzed.
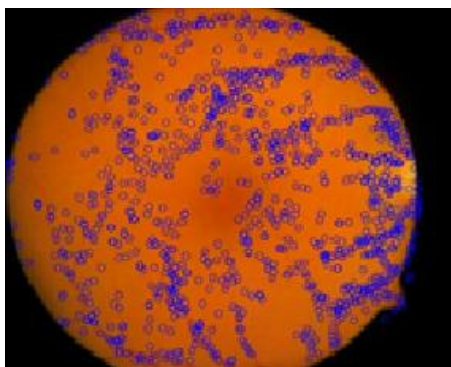
Figure 31: SURF feature descriptors computed from a retinal image and using a 1×10^{-5} threshold value which discards the filter response values below this value in keypoint identification step.

For a better understanding of the procedures used for SURF descriptors computation a flowchart of the main steps may be seen in Appendix A.4.

4.2. Database Image Search

4.2.1. Vocabulary Trees

Image database search is a fundamental issue in this dissertation being important to use an approach which scales efficiently to large databases. As aforementioned, the main motivation of this work is the identification in a large retinal image database of a subset of images that belong to the same individual and eye as those of a query image. One efficient solution would be the use of an identification method based on retinal images, which compares two retinal images and verifies whether they belong to the same person and eye. However, it was already concluded that this procedure is infeasible to use in the problem context since it would be necessary to compare the query image with each database image. Considering this aspect, Tree Data Structures were explored and used to the retrieval of a reduced list of the candidate images most likely to match the query image [71]. Posterior to this, an efficient person identification system based on retinal images can then be applied to the selected images in order to ensure the correct identification of images of the same individual and eye. For this type of image retrieval, database images need to be correctly represented in the Tree Data Structure. In the initial part of this chapter different feature descriptor computation methods were described. These descriptors were hierarchically quantized in a Tree Data Structure, more specifically a Vocabulary Tree (VT) [59], for an efficient image retrieval.

Vocabulary Trees define a hierarchical organization of the feature descriptors computed for each database image by performing a hierarchical k-means clustering on them. Initially, k large cluster nodes are defined, k representing the number of children nodes of each cluster (branch factor) [59]. Posteriorly, k-means clustering is recursively performed in each children cluster. This recursive division of the descriptor space is repeated until there are enough bins to ensure a good classification performance [71]. A VT with k branch factor and L levels of iteration has k^L

36

terminal nodes, which means that the descriptor vectors will be divided into k^L clusters according to their similarity degree.

Figure 32 illustrates a VT with three children nodes for each node (k=3) and two levels of iteration (L=2), meaning that it has 3^2=9 terminal nodes, in which the training descriptors are divided.
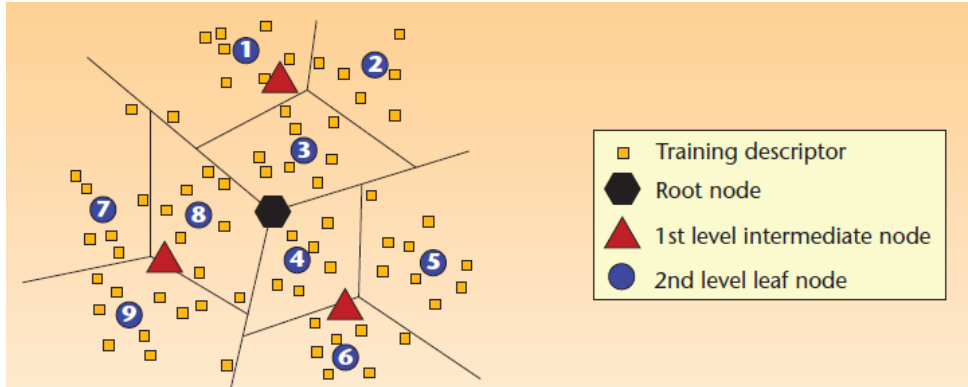
Figure 32: Vocabulary Tree representation [71].

At each terminal node it is associated a list with the number of feature descriptors of each image represented in the node. These lists, named inverted index lists, are represented in Figure 33, where i_{ab} refers to the image index b with feature descriptors in the node a and c_{ab} is the number of descriptors of the image with index b at the node a.
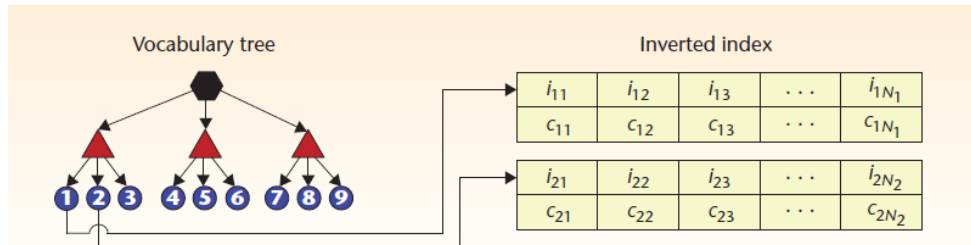
Figure 33: Inverted index lists associated to each node of a Vocabulary Tree [71].

Inverted index lists are only computed for terminal nodes since for inner nodes these lists correspond to the junction of their children inverted index lists. For this purpose inner nodes are considered to be associated to virtual inverted index lists as it may be seen in Figure 34.
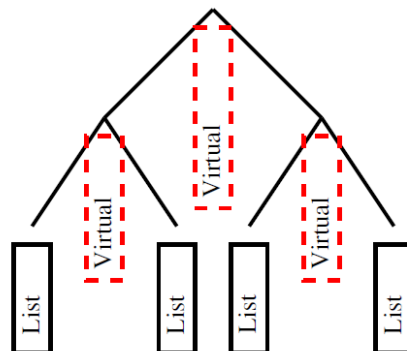
Figure 34: Real and virtual index lists in Tree Data Structures [59].

To search in this structure for the most similar images to a query one, the first step is the computation of the same type of descriptors (that were used in the VT training) for the image being queried. Each query descriptor traverses the VT until a terminal node is reached. The path of a descriptor is defined by its similarity to the mean descriptor of each node at each level. The similarity measure is performed by a Euclidean distance between the query descriptor and the node centroid descriptor, which is the mean descriptor vector of all feature descriptors represented in that cluster. The most similar node centroid to the query descriptor is identified and the same analogy is performed with its children nodes, in the level below. After trasversing all query descriptors in the VT, the images whose descriptors are represented in the terminal nodes selected by the query descriptors are retrieved. Since this subset of images may contain several images, it is important to exclude some of them and rank them in a similarity order to perform an efficient image database search.

It was noticed that there were terminal nodes with descriptors from almost all or even all images. Those terminal nodes do not allow an efficient image retrieval as they do not contribute to the discriminative power of the method. Similarly, terminal nodes with few descriptors do not significantly affect the search performance although they increase the search time. For this reason, a percentage of the terminal nodes with more and less descriptors were not considered in the search step, that is, images whose descriptors were in these nodes were not retrieved. The percentage of nodes not considered for this reason depends on the feature computation method applied as well as the VT used. In this way, these parameters were studied for several conditions which is further described in Chapter 5 – Results and Discussion.

Moreover, some descriptors may appear in almost all images while others may be very rare, not contributing to an efficient image retrieval. Due to this, a list of all possible descriptor vectors was organized and their occurrences in the VT were counted. In the same way and for the same reason as with the terminal nodes, the most and the least frequent descriptors were not considered in the search step, making part of stop lists as introduced by Sivic J. and Zisserman A. [13]. Once again the percentage of the least and the most frequent descriptors that can be ignored without affecting the search performance was studied for different feature computation methods and may as well be analyzed in Chapter 5 – Results and Discussion.

In addition to node and descriptor frequency analysis, a similarity scoring approach identical to the one proposed by Sivic J. and Zisserman A. [13] and Girod et all [71] was implemented. The score is based on a node weight computed by equation (24), where N is the total number of images represented in the VT and N_v is the number of images with descriptors in node v.

$$w(v) = \log(\frac{N}{N_v})$$
(24)

This weight is higher for terminal nodes with fewer descriptors and lower for terminal nodes with a larger number of descriptors, effectively down-weighting the less useful terminal nodes for image retrieval of visually similar images.

In the search step, a similarity scoring [71] is computed for each image retrieved so that the final list of images may be ranked and only the most similar ones are selected. For each node selected in each query descriptor path across the VT, the scores of the images represented in that node are incremented according to the similarity score expressed by equation (25). In this equation w(v) is the terminal node weight computed with equation (24), c_q and c_d correspond to the number of descriptors from the query image (q) and the database image (d) , respectively,

that fall into node v and $\sum d$ and $\sum q$ refer to normalization factors which are simply the sum of the node weights of all the k nodes in which there are descriptors from images d and q.

$$s(i_{kj}):= s(i_{kj}) + \frac{w_k^2 c_{kj} q_k}{\sum d \sum q}, j = 1, ..., N_k \tag{25}$$

Only the database images whose similarity score is above a similarity threshold value are selected for the next processing step, the person identification procedure based on retinal images, to efficiently only identify the target images. The threshold value was chosen experimentally according to the feature computation method used.

When considering SIFT and SURF descriptors, a different similarity scoring based on the one developed by Wang et all [14] was used for purposes of image ranking. This similarity score, which defines a similarity measure between two images q and d, is specified considering all the possible pairs between the query and the database image descriptors, i and j, which are represented in a same node v.

$$sim(q, d) = \frac{1}{|I_q||I_d|}\sum_{i\in I_q, j\in I_d} w_{i,j}^{q,d}(v) \tag{26}$$

The weighting function $w_{i,j}^{q,d}(v)$ is given by (27).

$$w_{i,j}^{q,d}(v) = w_{i,j}^C \times w_i^q \times w_j^d \times w(v) \tag{27}$$

As it may be easily analyzed, different aspects are taken into consideration in this weighting function. The first term expresses contextual statistics between descriptors.

$$w_{i,j}^C = w_{i,j}^\rho \times w_{i,j}^s \times w_{i,j}^\theta \tag{28}$$

$$w_{i,j}^\rho = \frac{\min(\rho_i, \rho_j)}{\max(\rho_i, \rho_j)}; \ w_{i,j}^s = \frac{\min(\overline{\Delta s_i}, \overline{\Delta s_j})}{\max(\overline{\Delta s_i}, \overline{\Delta s_j})}; \ w_{i,j}^\theta = \frac{\min(\overline{\Delta \theta_i}, \overline{\Delta \theta_j})}{\max(\overline{\Delta \theta_i}, \overline{\Delta \theta_j})} \tag{29}$$

For this term computation, three descriptor statistics ($\rho, \overline{\Delta s}, \overline{\Delta \theta}$) have to be obtained. Each SIFT or SURF descriptor (f_0) is represented by the descriptor vector itself x_0, its location u_0, its scale s_0 and its orientation θ_0. First, the image descriptors inside a circular neighborhood of the f_0 descriptor with radius $R = 12 \times 2^{s_0}$ are selected. The statistical terms may then be computed using equations (30) to (32), where $|C(f_0)|$ is the number of descriptors within the neighborhood of f_0.

$$\rho = |C(f_0)| \tag{30}$$

$$\overline{\Delta s} = \frac{1}{|C(f_0)|}\sum_{f\in C(f_0)} |s - s_0| \tag{31}$$

$$\overline{\Delta \theta} = \frac{1}{|C(f_0)|}\sum_{f\in C(f_0)} |\theta - \theta_0| \tag{32}$$

The second and third terms are related to the descriptor path in the VT and refer to the query image descriptor and database image descriptor respectively. Taking w_i^q as example, it may be computed using equation (33), where w(v) is obtained with equation (24) and consists in

the node weight of each node in the path $T(x_i)$ of descriptor x_i, and $n^q(v)$ represents the number of descriptors from image q represented in node v. For the database image d, w_i^d may be computed using the same analogy.

$$w_i^q = \sqrt{\frac{\sum_{v \in T(x_i)} w(v)}{\sum_{v \in T(x_i)} w(v) \times n^q(v)}} \tag{33}$$

Chapter 5

Results and Discussion

This chapter addresses two different algorithm performance analysis. The first one characterizes the effect of eye pathology lesions effect on the performance of the different image descriptors used to represent retinal images and their efficiency in performing person and eye identification. It was crucial to perform this initial evaluation since the second part of the study uses healthy and unhealthy retinal images. The second set of performance studies analyses the retrieval success rates as well as the processing time for the identification of retinal images belonging to the same patient and eye as those of the retinal image being queried in a large image database.

5.1. Robustness of the Feature Computation Methods to Eye Pathologies

As previously mentioned, the main motivation of this dissertation was the improvement of an automated clinical decision support system for eye pathology detection and monitoring. For this reason, it is important that the algorithm proposed is robust to the presence of the most common eye pathology characteristic lesions. The database search of retinal images from the same patient of the one being queried should not be affected by the presence of eye lesions in retinal images such as hemorrhages, drusens and edemas. In this analysis two diseases were considered: Diabetic Retinopathy (DR) and Age-related Macular Degeneration (AMD). The choice of these pathologies is justified by their prevalence among those that cause vision loss or total blindness.

Four different image datasets, which were all part of studies performed at Critical Health®, were used in this study. The datasets are described in Table 1.

Table 1: Characteristics of the retinal image datasets used

	Dataset 1	Dataset 2	Dataset 3	Dataset 4
Type of retinal images predominant in the dataset	AMD	DR	Healthy	AMD+DR+Healthy
Number of images	230	284	435	949
Number of different patients represented	58	81	200	339

In order to evaluate the robustness of each feature computation method to eye pathology lesions, the descriptors of each image, the query image, were compared to those of each remaining images in the dataset and their similarity was measured. For each dataset, a threshold value of similarity was established so that only the images whose similarity to the query image was higher than the value defined were considered to have high probability of belonging to the same patient and eye of the query image. The similarity threshold value was chosen so that for nearly 99% of the images queried all the images belonging to the same patient and eye were selected. This analysis was performed within each different dataset and for each of the feature computation methods analyzed in this thesis.

The four different types of descriptors used in this analysis were computed as described in Chapter 4- Methodology. The WEF descriptors were computed using Daubechies wavelet family filters and two levels of wavelet decomposition. Besides this, both WEF and FD descriptors were obtained by partitioning the image in 100x100 non-overlapped image subregions. For these two types of descriptors a row descriptor vector was built by joining all descriptor vectors computed for each image subregion. Since the images used in this analysis, within each dataset, have the same characteristics, regarding for example viewpoint, scale and illumination, this procedure of joining image subregion descriptors does not significantly affect the correct selection of the most similar images to a query image and enables a less complex evaluation of the method's performance. In order to measure the similarity between two images, a simple Euclidean difference between their respective descriptors is computed.

Although SIFT and SURF also enable the computation of several descriptor vectors per image, the procedure of joining the descriptors in a row vector is not adequate since these descriptors are computed for each interest point in the image and not for image subregions. Therefore, a different similarity measure was used with these two feature descriptor computation methods. To evaluate the similarity between a pair of images it is performed a keypoint descriptor matching between both images. The initial step of keypoint matching of the descriptors from image A to those of image B is the identification for each image A descriptor of the most similar descriptor in image B. The similarity between descriptors is measured by Euclidean distance between the two descriptor vectors. To ensure a correct keypoint matching, i.e. the keypoints matched represent the same feature in both images, the distance between the two matched descriptors is analyzed. A pair of matched descriptors D1 and D2 is only considered a correct match if the distance between then multiplied by 1.5 (value defined by Lowe for the same matching procedure [48]) is not greater than the distance of the descriptor D1 to all other descriptors from the other image. Considering this, the similarity measure between two images when using SIFT and SURF feature computation methods is given by the number of correct matches found between them.

The similarity between each dataset image and each remaining images from the same dataset was computed and the minimum percentage of the dataset that could be selected to ensure that all images of the same patient and eye of the query image were selected was determined. The selected images were those whose similarity measure with the query image was greater than the similarity threshold value, which was chosen experimentally for each feature computation method so that for at least 99% of the query images considered all the images belonging to the same patient and eye of the image being queried were selected.
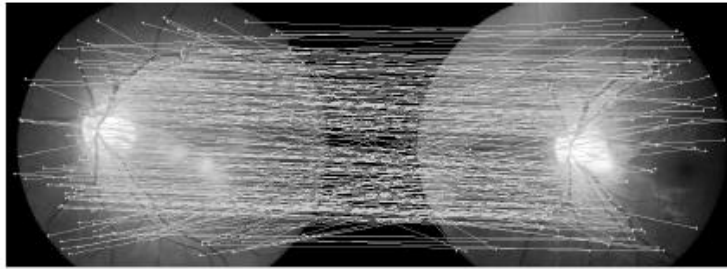
Figure 35: SIFT Keypoint matching between two retinal images [50].

The following tables summarize the results obtained in this study. FD and WEF feature computation methods were applied to retina vessel tree images for the reasons previously explained in Chapter 4 - Methodology. On the other hand, SIFT and SURF feature computation methods were applied to retinal images since these methods take into consideration the image gradients for descriptor computation.

Table 2: FD descriptors robustness to eye pathologies – % of dataset selection to ensure, in nearly 99% of the images queried, the selection of all images belonging to the same patient and eye of the image being queried.

Dataset	% of images selected from the entire Dataset
Dataset 1	97.51%
Dataset 2	79.61%
Dataset 3	51.06%
Dataset 4	47.07%

Table 3: WEF descriptors robustness to eye pathologies - % of dataset selection to ensure, in nearly 99% of the images queried, the selection of all images belonging to the same patient and eye of the image being queried.

Dataset	% of images selected from the entire Dataset
Dataset 1	87.27%
Dataset 2	77.92%
Dataset 3	73.49%
Dataset 4	38.91%

Table 4: SIFT descriptors robustness to eye pathologies – % of dataset selection to ensure, in nearly 99% of the images queried, the selection of all images belonging to the same patient and eye of the image being queried.

Dataset	% of images selected from the entire Dataset
Dataset 1	99.47%
Dataset 2	87.58%
Dataset 3	76.10%
Dataset 4	78.96%

Table 5: SURF descriptors robustness to eye pathologies – % of dataset selection to ensure, in nearly 99% of the images queried, the selection of all images belonging to the same patient and eye of the image being queried.

Dataset	% of images selected from the entire Dataset
Dataset 1	97.32%
Dataset 2	88.15%
Dataset 3	77.81%
Dataset 4	77.03%

Percentage of each dataset selection to ensure that all images belonging to the same patient and eye of the image being queried are selected
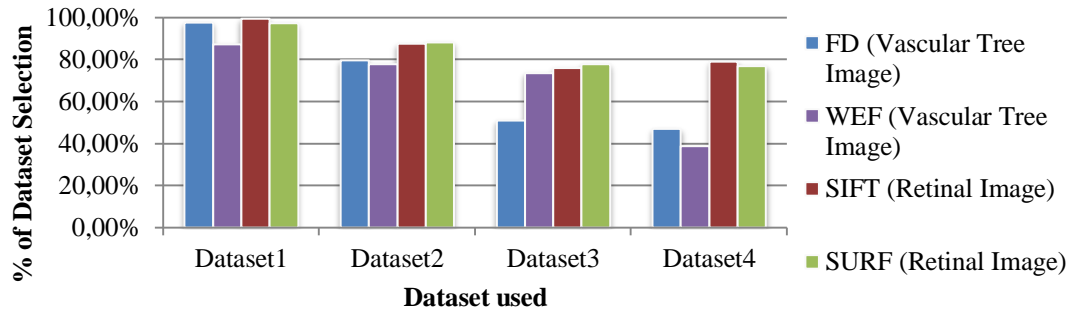
Figure 36: Representation of the percentage of each dataset selection needed to ensure that all images belonging to the same patient and eye of the image being queried are selected.

Taking into consideration the results from the previous tables, the effect of AMD and DR characteristic lesions in the retina obviously influence the feature descriptor computation. When considering images from Dataset 1, with prevalence of retinal images with AMD characteristic lesions, almost the entire dataset needs to be selected for the correct selection of images belonging to the same patient of the image being queried. The large yellow areas, drusens, which are consequences of this eye pathology, are overlapped with the retina vessel tree explaining the small dataset reduction obtained.

Images from Dataset 2 are associated with DR characteristic lesions. These lesions although smaller than the AMD drusens also affect the retina vessel tree, modifying this biometric feature. By a careful examination of the results, one may easily perceive the effect of these lesions in the identification of images belonging to the same patients and eye, though not as severe as with AMD lesions.

When considering the results obtained with Dataset 4, which contains healthy and unhealthy retinal images, one would not expect that a more reduced dataset selection would be obtained with this dataset than with Dataset 3, which only contains healthy retinal images. However, a reasonable explanation is the fact that these type of lesions, when present in all images of the same patient, are one more unique feature of the patient retina. Therefore, the fact that some images were associated with different types of lesions while others not enabled to discard a larger part of the dataset than when using solely healthy retinal images in Dataset 3. Even so, a robust method for the identification of images of the same patient and eye in a database may not be based in this fact since eye lesions may decrease or even disappear with treatment through time or, on the contrary, increase in size and number in the worst scenario.

With this preliminary analysis of the feature computation methods implemented it is easily concluded that FD and WEF descriptors are the most adequate ones to use in the retrieval of images of the same patient and eye of a query retinal image. One reason for the good performance of these feature computation methods in this analysis is the fact they were applied to retinal vessel tree images. In these images the influence of eye pathology lesions is not as evident as in retinal images since only the vessels are figured in the image and only the situations where lesions are overlapped to the vessel tree are portrayed.

5.2. Image Descriptors Efficiency in Image Retrieval

The novel aspect introduced by this dissertation is the use of VTs combined with image feature computation methods for the identification, in a large database, of the most similar images of a query retinal image. This subset of images may be as reduced as possible since it is posteriorly used in an efficient retina-based person identification method which enables to accurately verify whether two retinal images belong to the same patient.

In image retrieval performance analysis two different approaches were studied. If considering that the images stored in the database are not labeled or indexed in any way, not being possible to establish any type of correspondence between the images and their respective patient information, it is crucial that the search technique applied enables the retrieval of all images from the same patient. On the other hand, database images may be correctly labeled and the lack of information is only verified in the acquisition of a new retinal image, the query one. In this scenario it is only necessary to ensure the retrieval of one stored database image from the same patient of the image being queried. Doing so, a simple database index search is performed to retrieve the remaining images belonging to that patient. For these reasons, both scenarios were included in the analysis conducted. All the parameters chosen throughout the study were based on the fact that for nearly 99% of the query images used all or at least one, depending on the analysis approach considered, of the stored database images of the same patient and eye of the image being queried should be retrieved in the final outcome.

To ensure that the training and testing dataset choice do not significantly affect the results obtained, a k-fold procedure was used in all computations performed. The total image dataset (1989 images) available for these experiments with correct patient information associated to each image was divided into five groups (with 397,397,398 398and 399 images respectively). In each analysis, one of these groups was used as a testing dataset and the remaining four groups were used for the algorithm training. Since the VT retrieval efficiency strongly depends on the number of descriptors used in its training, more 3467 retinal images were used for that purpose. This means that a training dataset with more than five thousand retinal images is used for the VT training. Considering this, the results shown in this thesis correspond to the mean result obtained with the five testing image datasets.

5.2.1. Fractal Dimension (FD)

In Chapter 4- Methodology, the FD feature computation method is described in detail. Image partitioning is one of the major steps in this method. Since the image subregions may assume different sizes, an initial analysis was performed to evaluate the image subregion size effect in the image retrieval efficiency. As expected, by decreasing the image subregion size the number of subregions in each image increases, as well as the number of descriptors associated to each image. In Table 6 it is shown the number of descriptors per image obtained with different image subregion sizes for image partitioning, considering that the feature computation method was applied to 1024x1024 retinal vessel tree images.

Table 6: Number of FD descriptors per image with different image partitioning degrees.

Image Subregion Size in Pixels	Number of Descriptors per image
200x200	25
150x150	36
125x125	64
106x106	81
100x100	100
75x75	169
50x50	400
32x32	1089

For each image partitioning degree, the descriptors computed were organized in VTs with k=5 and different levels (L). The following tables show the database percentage retrieved when using the VT to search for the most similar images to a query image, as well as the percentage of the query images for which all or at least one image of the same patient and eye were correctly retrieved.

Table 7: FD - Image retrieval performance using 200x200 image subregions.

Structure (KxL)	% of the database selected	Approach 1: % selection of ALL images of the same patient and eye	Approach 2: % selection of at least ONE image of the same patient and eye
Tree 5x4	56.27%	99.80%	99.95%
Tree 5x5	43.06%	91.90%	98.90%
Tree 5x6	18.22%	47.66%	85.92%

Table 8: FD - Image retrieval performance using 150x150 image subregions.

Structure (KxL)	% of the database selected	Approach 1: % selection of ALL images of the same patient and eye	Approach 2: % selection of at least ONE image of the same patient and eye
Tree 5x4	60.81%	100.00%	100.00%
Tree 5x5	49.85%	98.84%	99.95%
Tree 5x6	28.43%	67.52%	94.97%

Table 9: FD - Image retrieval performance using 125x125 image subregions.

Structure (KxL)	% of the database selected	Approach 1: % selection ALL images of the same patient and eye	Approach 2: % selection of at least ONE image of the same patient and eye
Tree 5x5	48.15%	99.95%	100.00%
Tree 5x6	37.93%	91.86%	98.84%
Tree 5x7	17.06%	43.28%	83.71%

Table 10: FD - Image retrieval performance using 106x106 image subregions.

Structure (KxL)	% of the database selected	Approach 1: % selection ALL images of the same patient and eye	Approach 2: % selection of at least ONE image of the same patient and eye
Tree 5x5	49.48%	100.00%	100.00%
Tree 5x6	42.41%	97.03%	99.60%
Tree 5x7	22.90%	58.47%	90.85%

Table 11: FD - Image retrieval performance using 100x100 image subregions.

Structure (KxL)	% of the database selected	Approach 1: % selection ALL images of the same patient and eye	Approach 2: % selection of at least ONE image of the same patient and eye
Tree 5x5	56.47%	100.00%	100.00%
Tree 5x6	49.89%	99.04%	99.90%
Tree 5x7	29.70%	68.88%	93.37%

Table 12: FD - Image retrieval performance using 75x75 image subregions.

Structure (KxL)	% of the database selected	Approach 1: % selection ALL images of the same patient and eye	Approach 2: % selection of at least ONE image of the same patient and eye
Tree 5x6	53.74%	99.95%	100.00%
Tree 5x7	41.62%	90.85%	98.84%
Tree 5x8	20.05%	43.89%	84.82%

Table 13: FD - Image retrieval performance using 50x50 image subregions.

Structure (KxL)	% of the database selected	Approach 1: % selection ALL images of the same patient and eye	Approach 2: % selection of at least ONE image of the same patient and eye
Tree 5x6	59.95%	100.00%	100.00%
Tree 5x7	52.24%	99.55%	99.95%
Tree 5x8	36.70%	76.44%	94.93%

Table 14: FD - Image retrieval performance using 32x32 image suubregions.

Structure (KxL)	% of the database selected	Approach 1: % selection ALL images of the same patient and eye	Approach 2: % selection of at least ONE image of the same patient and eye
Tree 5x7	59.16%	100.00%	100.00%
Tree 5x8	52.91%	99.30%	99.90%
Tree 5x9	41.77%	84.77%	96.23%

In order to provide a better analysis of the results obtained, the VTs which enabled the best image retrieval performance for each image partitioning degree are shown in Table 15 and Table 16, for image retrieval of all and at least one image of the same patient, respectively.

Table 15: FD - Best image retrieval performance, of all images from the same patient and eye, for each image partitioning degree.		
Image Subregion Size	**Tree Structure (kxL)**	**Approach 1: % of the database selected to ensure the retrieval of ALL images of the same patient and eye**
200x200	Tree 5x4	56.27%
150x150	Tree 5x5	49.46%
125x125	Tree 5x5	48.07%
106x106	Tree 5x5	49.20%
100x100	Tree 5x6	50.66%
75x75	Tree 5x6	53.74%
50x50	Tree 5x7	52.24%
32x32	Tree 5x8	53.27%

Table 16: FD - Best image retrieval performance, of at least one image of the same patient and eye, for each image partitioning degree.		
Image Subregion Size	**Tree Structure (kxL)**	**Approach 2: % of the database selected to ensure the retrieval of at least ONE image of the same patient and eye**
200x200	Tree 5x5	43.06%
150x150	Tree 5x5	49.85%
125x125	Tree 5x6	37.93%
106x106	Tree 5x6	42.41%
100x100	Tree 5x6	49.89%
75x75	Tree 5x7	41.62%
50x50	Tree 5x7	52.24%
32x32	Tree 5x8	52.91%

% of the database selected to ensure the retrieval of ALL images of the same patient and eye
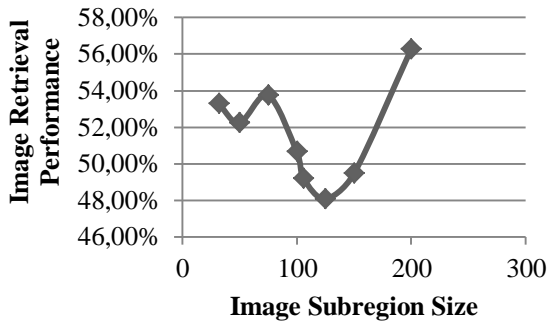
Figure 37: Best image retrieval performance, of all images of the same patient and eye, for each image partitioning degree.

% of the database selection to ensure the retrieval of at least ONE image of the same patient and eye
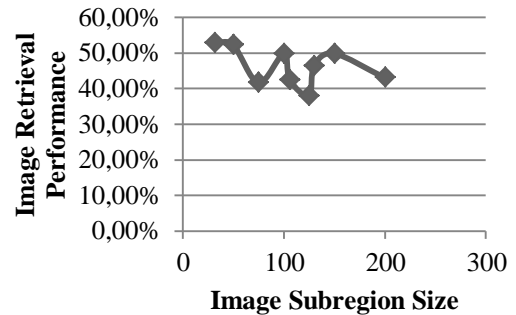
Figure 38: Best image retrieval performance, of at least one image of the same patient and eye, for each image partitioning degree.

Analyzing Figures 37 and 38, it may be concluded that image partitioning significantly affects the image retrieval performance. In Figure 37, it is noticeable that when decreasing the image subregion size a better image retrieval performance is achieved, i.e. a smaller percentage of the database is retrieved. However, when using too small image subregions the image information represented in them is not enough for a correct descriptor computation and the image retrieval becomes less efficient. By analyzing the results from both approaches, one may conclude that 125x125 image subregions are the most appropriate ones for this type of image retrieval.

For purposes of obtaining a better image retrieval performance using VT, different k values were used with the three best image partitioning degrees obtained from the previous analysis (150x150, 125x125 and 106x106 image subregions).

Table 17: FD - Image retrieval performance of VTs with different k factors in retrieving All images of the same patient and eye in nearly 99% of the images queried

k	150x150		125x125		106x106	
	Level	% of the database selected	Level	% of the database selected	Level	% of the database selected
2	10	62.73%	11	49.31%	12	47.39%
3	6	55.91%	7	47.79%	8	45.61%
4	5	55.72%	5	51.81%	6	47.71%
5	5	49.46%	5	48.07%	5	48.07%
6	4	56.05%	5	44.92%	5	46.48%
7	4	53.67%	4	49.80%	4	53.26%
8	4	50.87%	4	47.89%	4	50.12%
9	3	64.72%	4	47.15%	4	48.16%
10	3	62.23%	4	45.71%	4	47.19%
11	3	62.88%	3	59.66%	4	46.18%
12	3	58.16%	3	57.01%	4	46.02%
13	3	57.10%	3	55.58%	4	44.03%
14	3	55.98%	3	53.45%	3	57.25%
15	3	55.17%	3	51.53%	3	55.85%

Table 18: FD - Image retrieval performance of VTs with different k factors in retrieving at least ONE image of the same patient and eye in nearly 99% of the images queried

k	150x150		125x125		106x106	
	Level	% of the database selected	Level	% of the database selected	Level	% of the database selected
2	11	58.01%	13	38.66%	13	42.76%
3	7	49.48%	8	42.29%	9	36.24%
4	6	45.68%	5	51.81%	7	40.08%
5	5	49.85%	6	37.93%	6	42.41%
6	4	56.05%	5	44.92%	5	46.48%
7	4	53.67%	5	40.32%	5	43.25%
8	4	50.87%	4	47.89%	5	39.05%
9	4	47.87%	4	47.15%	4	48.16%
10	3	44.30%	4	45.71%	4	47.19%
11	4	41.01%	4	44.26%	4	46.18%
12	3	58.16%	4	42.35%	4	46.02%
13	3	57.10%	4	41.69%	4	44.03%
14	3	55.98%	4	39.11%	4	43.20%
15	3	55.17%	4	39.01%	4	41.25%

In Tables 17 and 18, which summarize the results of this analysis for approach 1 and approach 2 respectively, only the best image retrieval performance for each k value is shown. Taking into consideration this analysis, 106x106 image subregions and a 13x4 VT were found to enable the best image retrieval performance for retrieval of all images of the same patient and eye (it enables the selection of 44.03% of the entire database). On the other hand, for image retrieval of at least one image of the same patient and eye 125x125 image subregions and a 5x6

VT were the combination that enabled the best retrieval efficiency (selection of 37.93% of the entire database).

As mentioned in the previous chapter, the most and the least populated nodes, that is, those with more and less descriptors, can be removed from the VT search step in order to increase image retrieval performance. Tables 19 and 20 summarize this analysis for both approaches studied.

Table 19: FD - Image retrieval performance in retrieving all images of the same patient and eye by removing the most and the least populated nodes

	% of Nodes removed	% of the database selected	Approach 1: % of situations in which ALL images of the same patient and eye are selected
Most Populated Nodes	0.02%	43.60%	98.79%
	0.03%	43.60%	98.79%
	0.04%	43.60%	98.79%
	0.05%	43.29%	98.69%
Least populated nodes (After removal of the 0.04% most populated nodes)	0.1%	43.60%	98.79%5
	0.2%	43.60%	98.74%
	0.5%	43.60%	98.74%
	1.0%	43.60%	98.74%

Table 20: FD - Image retrieval performance in retrieving at least one image of the same patient and eye by removing the most and the least populated nodes

	% of Nodes removed	% of images selected from the database	Approach 2: % of situations in which at least ONE image of the same patient and eye is selected
Most Populated Nodes	0.01%	37.76%	98.84%
	0.05%	37.72%	98.79%
	0.07%	37.67%	98.79%
	0.1%	37.63%	98.79%
	0.2%	37.52%	98.74%
	0.5%	37.15%	98.49%
Least populated nodes (After removal of the 0.1% most populated nodes)	5.0%	37.63%	98.79%
	6.0%	37.63%	98.79%
	7.0%	37.63%	98.74%
	10.0%	37.62%	98.74%

Although this analysis has not enabled a significant increase in image retrieval performance, removing the most and the least populated nodes enables a faster search step. In order to ensure that the percentage of situations in which all images from the same patient were retrieved (approach 1) or at least one image was retrieved (approach 2) did not suffer a significant decrease, in Approach 1 it was concluded that the 0.04% most populated nodes and the 0.1% least populated nodes could be removed from the search step and in Approach 2 the 0.1% most populated and the 6.0% least populated nodes could as well be removed.

Besides the node analysis, descriptor frequency analysis was also performed. As previously explained in Chapter 4 - Methodology, the most and the least frequent descriptors may not be used in the search step since they do not contribute to an efficient image retrieval performance. However, in this method of FD feature descriptor computation it was concluded that the descriptors were very distinct and although a group of descriptors were found to be more common in retinal images, the remaining ones were found to have only one occurrence in the total database of descriptors. Taking this into consideration the 10% most frequent descriptors were selected. Before traversing each query descriptor in the VT, the descriptor is compared to each of the 10% most frequent descriptors and if its difference to one of them is smaller than a specific threshold value of similarity, then that descriptor is not used in the VT traversal. This threshold value was experimentally chosen as it may be seen in Tables 21 and 22.

Table 21: **FD - Image retrieval performance in retrieving ALL images of the same patient and eye by choosing different difference threshold values for the 10% most common descriptor discard**

Threshold Value	% of images selected from the database	% of situations in which ALL images of the same patient and eye are selected
0	43.60%	98.79%
6	43.37%	98.79%
6.3	43.04%	98.69%
6.5	42.51%	98.49%
6.7	41.62%	97.94%
7	37.43%	92.06%

Table 22: FD - Image retrieval performance in retrieving at least ONE image of the same patient and eye by choosing different difference threshold values for the 10% most common descriptor discard

Threshold Value	% of images selected from the database	% of situations in which at least ONE image of the same patient and eye is selected
0	37.63%	98.79%
5	37.63%	98.79%
5.5	37.63%	98.79%
5.7	37.62%	98.74%
6	37.53%	98.69%

Once again, this analysis has not enabled a significant increase of the image retrieval performance although faster queries may be performed since less descriptors need to be traversed in the VT. It may be concluded that using a 6 or 5.5 threshold value for image retrieval of all images or of at least one image of the same patient and eye, respectively, is appropriate to reduce the amount of descriptors to traverse in the VT.

The last analysis performed was the use of a similarity score which is given by equation (24) in Chapter 4- Methodology. In this analysis each retrieved image is associated with a similarity score. Images whose similarity score is above a specified threshold value are considered the most similar ones to the image being queried and have high probability of belonging to the same individual. Different threshold values were tested in order to reduce the percentage of images retrieved while still ensuring that nearly in 99% of the images queried all

images (approach 1) or at least one image (approach 2) from the same patient and eye were retrieved.

Table 23: FD - Similarity score analysis for image retrieval of ALL images of the same patient and eye performance

Similarity Threshold Value	% of images selected from the database	Approach 1: % of situations in which ALL images of the same patient and eye are selected
5×10^{-6}	38.85%	98.79%
1×10^{-6}	38.85%	98.79%
5×10^{-5}	38.79%	98.64%

Table 24: FD - Similarity score analysis for image retrieval of at least ONE image of the same patient and eye performance

Similarity Threshold Value	% of images selected from the database	Approach 2: % of situations in which at least ONE image of the same patient and eye is selected
1.350×10^{-3}	32.80%	98.79%
1.355×10^{-3}	32.78%	98.79%
1.360×10^{-3}	32.76%	98.74%

This analysis enabled an increase in the image retrieval efficiency while ensuring the image retrieval of the correct images. For image retrieval of all images of the same patient and eye the similarity threshold value which enabled the best image retrieval performance (selection of 38.85% of the entire database) was a 1×10^{-6} threshold value while in approach 2 for image retrieval of at least one of those images a threshold value of 1.355×10^{-3} enabled the selection of only 32.78% of the database.

5.2.2. Wavelet Energy Feature (WEF)

To improve WEF feature computation method performance in image retrieval for identification of images of the same patient and eye of a query image, a few initial analysis were performed. In these preliminary studies simple descriptor Euclidean differences were computed to distinguish whether or not images belong to the same patient and eye.

Similarly to the procedures considered in the FD method, the effect of image partitioning size was the first parameter analyzed. Different image subregion sizes were used to compute WEF descriptors. By computing Euclidean descriptor differences, the database images were sorted by their similarity to the query image. Considering that it is necessary to select all images of the same patient and eye or, in a different analysis, at least one image of the same patient and eye for 99% of the query images, the parameter that was evaluated was the percentage of the database that could be selected while meeting the these requirements. Table 25 expresses the results obtained.

Table 25: **WEF - Image subregion partitioning effect in the retrieval performance.**

Image Subregion size	% Selection of the database to ensure the retrieval of ALL images of the same patient and eye	% Selection of the database to ensure the retrieval of at least ONE image of the same patient and eye
400x400	96.61%	87.36%
256x256	92.80%	63.84%
150x150	72.53%	41.31%
126x126	55.65%	25.78%
116x116	66.10%	33.26%
100x100	57.52%	20.13%
90x90	55.37%	23.09%
80x80	52.61%	29.10%
50x50	61.86%	27.97%
26x26	77.12%	40.61%
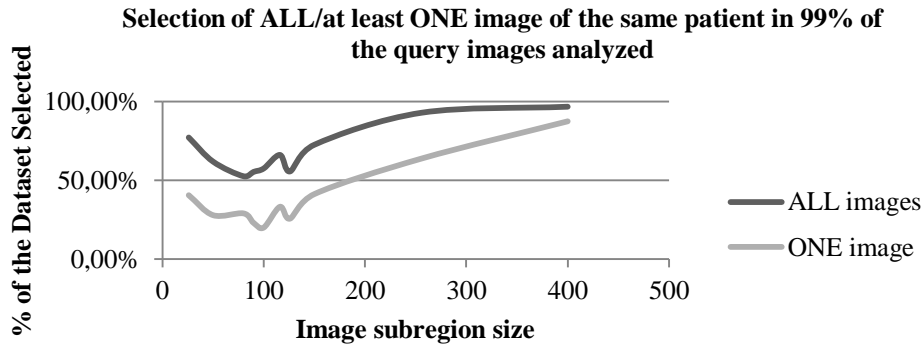
Figure 39: Image subregion partitioning effect in retrieval performance.

From both Table 25 and Figure 39, one may easily conclude that the best image retrieval performance is achieved by using 100x100 image subregions if the aim is to select all images of the same patient and eye and 80x80 image subregions if the selection of at least one image of the same patient and eye is enough. In addition to the image partitioning degree, there are other parameters which may be adjusted in order to increase image retrieval performance such as the filters wavelet family and the number of decomposition levels used.
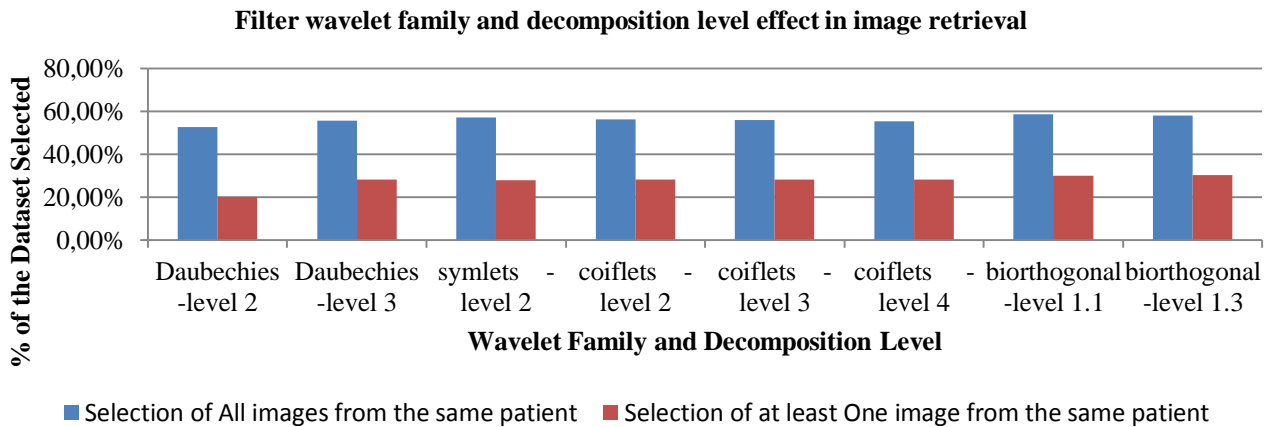
Figure 40: Filter wavelet family and decomposition level effect in image retrieval performance.

The best image retrieval performance for both approaches analyzed (retrieval of all images of the same patient and eye and retrieval of at least one of those images) was obtained using the Daubechies filter wavelet family and two levels of decomposition. The studies performed posteriorly to this analysis were computed using these parameters.

Considering the aspects previously concluded, VTs with WEF descriptor vectors were organized. Each image subregion is represented by a 6-length vector: the three energies for each level of decomposition as described in detail in Chapter 4 - Methodology. Similarly to the analysis performed with the FD feature computation method, different values of k, the number of branches of each VT node, were analyzed. Since in the previous study about the effect of the image partitioning degree on image retrieval the best image retrieval performances were obtained with two different image subregions sizes for the different analysis approaches, both image subregion sizes (100x100 and 80x80 image subregions) were used in this analysis.

Table 26: **WEF - Image Retrieval Performance by using VT with WEF descriptors computed from 100x100 image subregions**

K	Selection of ALL images of the same patient and eye		Selection of at least ONE image of the same patient and eye	
	Level	% of the Dataset selected	Level	%of the Dataset selected
3	7	96.84%	8	79.67%
4	6	86.36%	6	86.36%
5	5	89.45%	5	89.45%
6	4	98.12%	5	70.55%

Table 27: **WEF - Image Retrieval Performance by using VTs with WEF descriptors computed from 80x80 image subregions**

K	Selection of ALL images of the same patient and eye		Selection of at least ONE image of the same patient and eye	
	Level	% of the Dataset selected	Level	%of the Dataset selected
3	7	100.00%	9	71.91%
4	6	96.45%	7	73.24%
5	5	97.17%	6	71.60%
6	4	99.41%	5	81.73%

Due to the fact that WEF descriptors are very similar to each other, when using them to build a VT a low image retrieval performance is obtained. In fact, to ensure the selection of all images of the same patient and eye almost the entire image database needs to be considered (86.36% was the best image retrieval performance obtained with a 4x6 VT built with WEF descriptors computed from 100x100 image subregions). Although the selection of only one image of the same patient and eye enabled a better performance (70.55% with a 6x5 VT also built with WEF descriptors computed from 100x100 image subregions), it still represents a great part of the image database. Similarly to the analysis performed with the FD method, the most and the least populated nodes were ignored from the VT search step.

Table 28: **WEF - Image retrieval performance in retrieving ALL images of the same patient and eye by removing the most and the least populated nodes**

	% of Nodes removed	% of images selected from the database	Approach 1: % of situations in which ALL images of the same patient and eye are selected
Most Populated Nodes	2.0%	81.57%	97.98%
	1.0%	83.68%	98.79%
	0.5%	84.94%	98.49%
Least populated nodes (After removal of the 1.0% most populated nodes)	20.0%	83.45%	98.79%
	25.0%	83.30%	98.79%
	30.0%	83.11%	98.23%

Table 29: **WEF - Image retrieval performance in retrieving at least ONE image of the same patient and eye by removing the most and the least populated nodes.**

	% of Nodes removed	% of images selected from the database	Approach 2: % of situations in which at least ONE image of the same patient and eye is selected
Most Populated Nodes	2.0%	68.83%	98.79%
	4.0%	68.35%	98.79%
	6.0%	68.15%	98.79%
	7.0%	67.94%	98.44%
Least populated nodes (After removal of the 6% most populated nodes)	10.0%	68.02%	98.79%
	15.0%	67.95%	98.58%
	20.0%	67.60%	98.44%

Taking the results from Tables 28 and 29 into consideration one may conclude that although removing the most and the least populated nodes from the VT search step the image retrieval performance almost remains the same in both approaches. Although an increase in image retrieval performance is not expected by performing an analysis of the most common descriptors, since that analysis was performed for the other methods it was performed with this one as well.

Table 30: WEF - Image retrieval performance in retrieving all images of the same patient and eye by choosing different difference thresholds for the most common descriptor discard

Threshold Value	% of images selected from the database	% of situations in which ALL images of the same patient and eye are selected
0.7	82.81%	98.79%
0.5	82.55%	98.54%
0.4	81.16%	97.98%

Table 31: WEF - Image retrieval performance in retrieving at least one image of the same patient and eye by choosing different difference thresholds for the most common descriptor discard

Threshold Value	% of images selected from the database	% of situations in which at least ONE image of the same patient and eye is selected
1.3	65.91%	98.58%
1.4	67.09%	98.78%
1.5	67.34%	98.79%

As expected no major improvement of the image retrieval performance was obtained with this analysis, apart from the reduction in the computation time required for the VT search step. The final study performed was the similarity score attribution for image ranking. This similarity score is described by Equation (24) in Chapter 4- Methodology. Although not very significant, this procedure enabled a slightly better image retrieval performance as it may be seen in the following tables.

Table 32: WEF - Similarity score analysis for image retrieval of ALL images of the same patient and eye performance

Similarity Threshold Value	% of images selected from the database	% of situations in which ALL images of the same patient and eye are selected
1×10^{-6}	83.30%	98.79%
1×10^{-5}	80.02%	98.79%
2×10^{-5}	79.97%	97.73%

Table 33: WEF - Similarity score analysis for image retrieval of at least ONE image of the same patient and eye performance

Similarity Threshold Value	% of images selected from the database	% of situations in which at least ONE image of the same patient and eye is selected
1×10^{-6}	64.38%	98.78%
1×10^{-5}	60.24%	98.75%
2×10^{-5}	59.47%	98.57%

Considering the previous analysis, the best image retrieval performance obtained for the selection of all images of the same patient and eye was the selection of 80.02% of the image database. While, on the other hand, the selection of at least one of those images allows the use of only 60.24% of the total database.

5.2.3. Scale Invariant Feature Transform (SIFT)

In the SIFT descriptor computation method, several parameters could be chosen in order to adjust this descriptor computation method to the images used. In this dissertation only the threshold value that discards the maxima values of the scale space in the initial steps in order to select keypoint candidates is studied. Lowe used a 0.01 threshold value. However, by a careful analysis of the images from Figure 41 it is easily concluded that this threshold value is not appropriated to be used in this type of images, since too many descriptors are extracted.
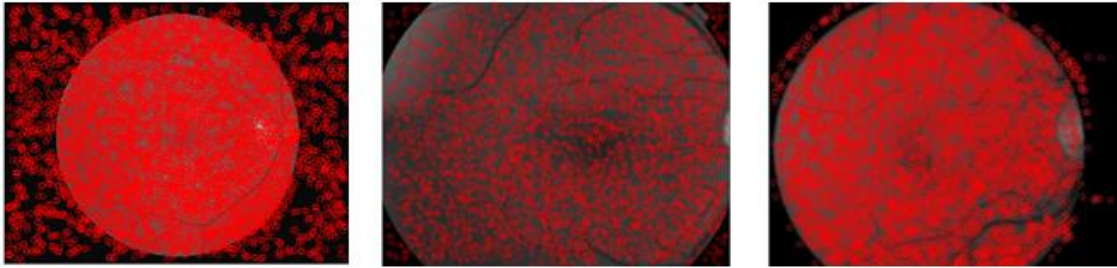
Figure 41: SIFT descriptors localizations by using a threshold value of 0.01 (Lowe's choice) for the scale space maxima discard.

Several threshold values were then analyzed to reduce the number of SIFT descriptors computed, although ensuring that important image features were still detected, such as the retinal vessel tree. By increasing the threshold value, the number of descriptors reduces. Figures 42 and 43 show the localization of SIFT descriptors computed with threshold 1 and 2, respectively.
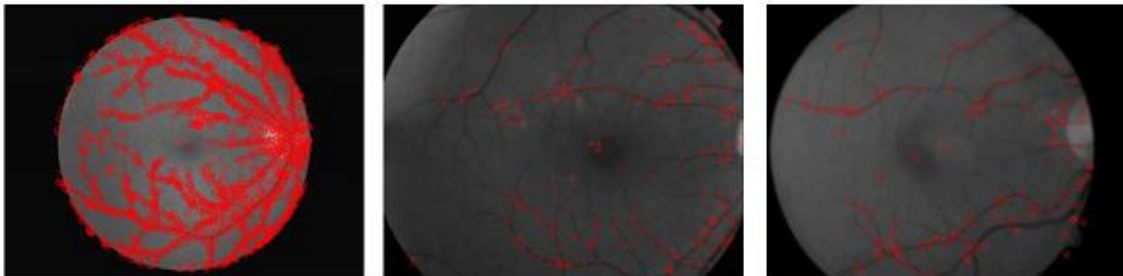
Figure 42: SIFT descriptors localizations by using a threshold value of 1 for the scale space maxima discard.
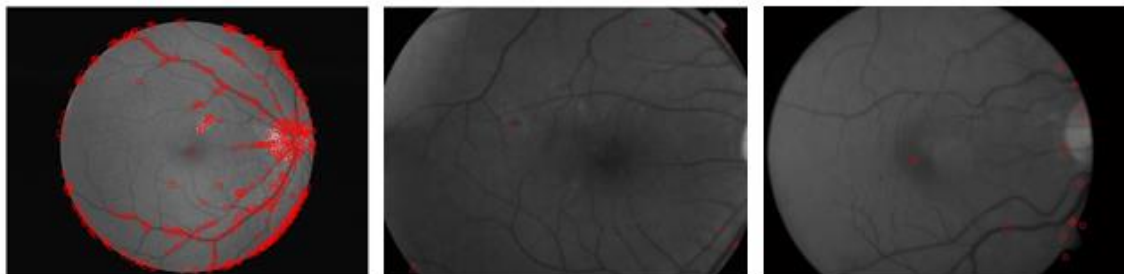
Figure 43: SIFT descriptors localizations by using a threshold value of 2 for the scale space maxima discard.

Using a threshold value equal to 1 seemed to be the best option from the threshold values analyzed since the keypoints detected are located in the retinal vessels, which is the most

important part of the image in a person identification method. Using a threshold value equal to 2 was not considered a good option due to the fact the number of keypoints detected is not enough to an efficient VT search. Moreover, important details of the tree vasculature of the retinal image were not detected with this threshold value. In this way, the further studies conducted are performed with SIFT descriptors computed with threshold value 1.

Posterior to the SIFT descriptors computation from all database images, VTs with k=5 were built. It was noticeable that the average number of SIFT descriptors computed for each image was significantly higher than the number of FD or WEF descriptors obtained for the same images. Therefore, fewer images were used in the VT training. Besides the use of the k-fold sets of images for which ones the information about each image patient is available, more 1100 images were used in this study instead of the 3467 images used for the VT training in the previous two methods analyzed.

Table 34: SIFT - Image retrieval performance with SIFT descriptors using retinal images

K	Level	%of the Dataset Selected	Approach 1: % of situations in which ALL images of the same patient and eye are selected	Approach 2: % of situations in which at least ONE image of the same patient and eye is selected
5	5	99.86%	100.00%	100.00%
	6	99.20%	98.77%	99.83%
	7	88.28%	74.78%	94.05%

Taken into consideration the excellent performance results obtained in published works using this type of descriptors in VTs for image retrieval, it was assumed that some characteristics of the retinal images used were not carefully examined which could explain the weak performance of this method with the images used. One of the image characteristics that was carefully examined was the image noise. The background pixel variance of the retinal images used was computed. For this computation a retina segmentation was performed using a threshold technique. A 0.5 pixel value threshold was applied to the image, segmenting the background from the retina itself. Doing so, only the background pixels are considered in the computation of the background variance intensity.

(a)   (b)
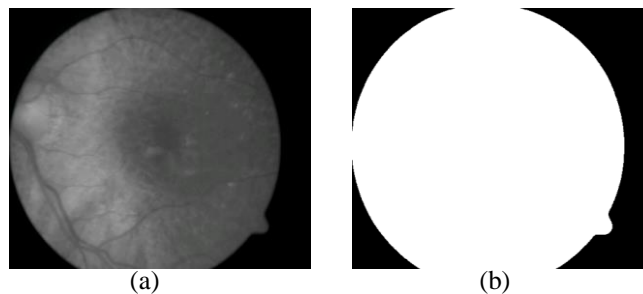
Figure 44: (a) Retinal image and (b) its segmentation from the background for pixel variance computation.

The images from the datasets used were analyzed with respect to their background variance and it was found that the pixel intensity variance obtained could assume different values as it may seen in Figure 45
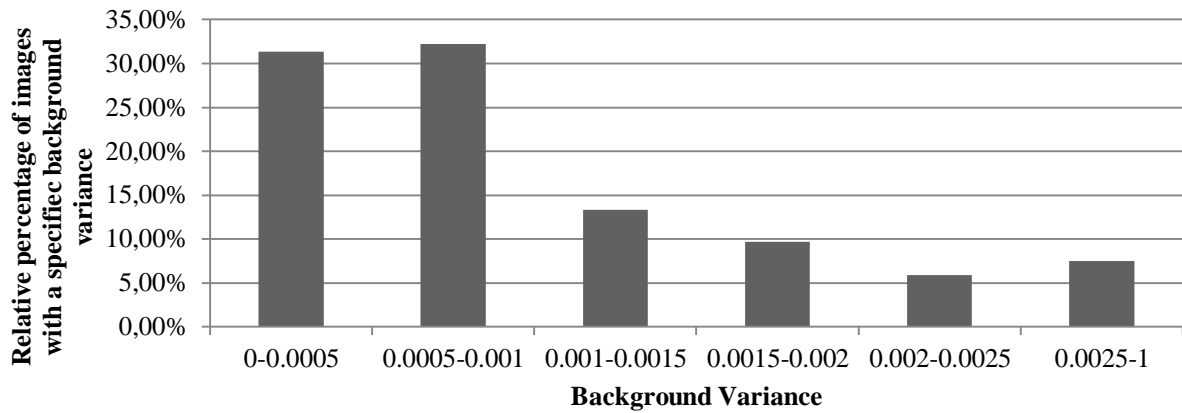
Figure 45: Plotting of the background variance of the retinal images used in this study.

To avoid this undesirable effect, retinal vessel tree images were used instead of the retinal images themselves. SIFT feature computation method is based on image gradients. In this way and since retina vessel tree images are logical images, a low-pass filter was applied in order to blur the binary representation of the tree vasculature so that a smooth gradient is obtained instead of a rough transition from pixel value 0 to pixel value 1. The filter used for this purpose was a Gaussian low-pass filter and its effect on the vascular tree image can be analyzed in Figure 46.
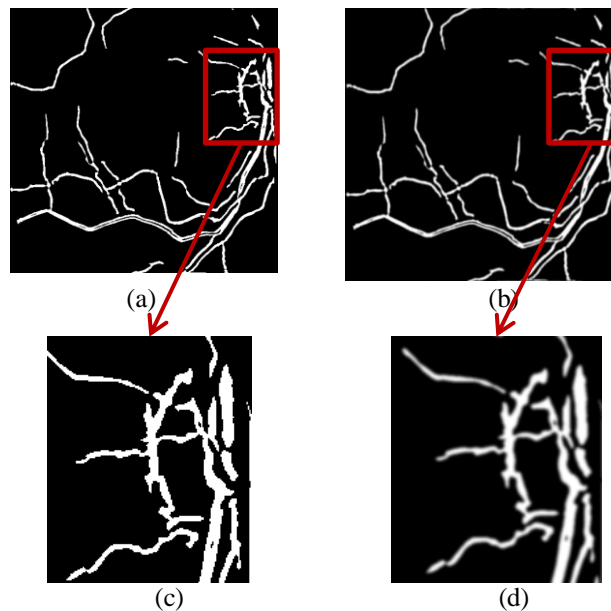
Figure 46: (a) Original retina vessel tree image, (b) Filtered retinal vessel tree image, (c) Detail of the original retina vessel tree image, (d) Detail of the smoothed retina vessel tree image.

VTs were built with the SIFT descriptors computed from the smoothed retinal vessel tree images. The retrieval performance results may be analyzed in Table 34. Since less SIFT descriptors are computed when using these images, 3467 images were used in this analysis apart from the k-fold sets of images.

Table 35: SIFT - Image retrieval performance of VT built with SIFT descriptors computed from smoothed retinal vessel tree images

K	Level	%of the Dataset Selected	Approach 1: % of situations in which ALL images of the same patient and eye are selected	Approach 2: % of situations in which at least ONE image of the same patient and eye is selected
	10	66.36%	100.00%	100.00%
5	11	65.94%	98.86%	99.38%
	12	63.17%	98.43%	98.95%

From the results represented in Table 35 it is noticeable the increase in image retrieval of the VTs when using SIFT descriptors computed from the smoothed retinal vessel tree images. The use of a 5x11 VT enables the selection of 65.94% of the entire database to ensure the retrieval of all images belonging to the same patient and eye of the image being queried. When only ensuring the retrieval of one of those images, 63.17% of the database is selected by using a 5x12 VT.

Similarly to what was done for other descriptors, the removal effect of the most and the least populated nodes was studied in detail, as well as the discard of the most and the least common descriptors. The effect of these parameters may be further understood by examining Tables 36 and 37.

Table 36: SIFT - Image retrieval performance in retrieving ALL images of the same patient and eye by removing the most and the least populated nodes

	% of Nodes removed	% of images selected from the database	Approach 1: % of situations in which ALL images of the same patient and eye is selected
Most Populated Nodes	0.5%	65.62%	98.86%
	1.0%	65.21%	98.81%
	1.3%	64.87%	98.79%
	1.5%	64.80%	98.74%
Least populated nodes (After removal of the 1.3% most populated nodes)	5.0%	64.85%	98.79%
	10.0%	64.77%	98.79%
	13.0%	64.72%	98.75%

Table 37: SIFT - Image retrieval performance in retrieving at least ONE image of the same patient by removing the most and the least populated nodes.

	% of Nodes removed	% of images selected from the database	Approach 2: % of situations in which at least ONE image of the same patient and eye is selected
Most Populated Nodes	0.5%	63.16%	98.95%
	1.0%	62.93%	98.89%
	1.5%	62.77%	98.81%
	1.7%	62.68%	98.75%
Least populated nodes (After removal of the 1.5% most populated nodes)	12.0%	62.47%	98.81%
	14.0%	62.41%	98.78%
	15.0%	62.38%	98.75%

The removal of the 1.3% and the 10.0% most and least populated nodes, respectively, enables the use of 64.77% of the entire database to ensure that all images belonging to the same patient and eye of the image being queried are retrieved. On the other hand, by removing the 1.5% and the 14.0% most and least populated nodes, 62.41% of the database may be selected to ensure that at least one of those images is retrieved.

By a careful examination of the SIFT descriptors computed, it was concluded that they were very distinct, most of them only appearing once in the database. Therefore, it was not possible to identify frequent and infrequent descriptors since most of them were unique in the set of descriptors.

As previously explained in Chapter 4 - Methodology, a different similarity score was used with SIFT and SURF descriptors, which is expressed by Equations 26 to 31. Using this score which is associated to each image retrieved, a similarity threshold may be used to reduce the percentage of the dataset selected. This threshold was experimentally chosen as it may be seen in Tables 38 and 39.

Table 38: SIFT - Similarity score analysis for image retrieval of ALL images of the same patient and eye performance

Similarity Threshold Value	% of images selected from the database	% of situations in which ALL images of the same patient and eye are selected
4×10^{-4}	61.35%	98.78%
8×10^{-4}	59.77%	98.77%
9×10^{-4}	58.98%	98.63%

Table 39: SIFT - Similarity score analysis for image retrieval of at least ONE image of the same patient and eye performance

Similarity Threshold Value	% of images selected from the database	% of situations in which at least ONE image of the same patient and eye is selected
9×10^{-4}	58.17%	98.78%
1×10^{-3}	57.84%	98.78%
1.3×10^{-3}	57.09%	98.66%

From Tables 38 and 39 one may conclude that the best retrieval performance was obtained by using a similarity threshold value of 8×10^{-4} for retrieval all images of the same patient and eye of the image being queried and 1×10^{-3} for retrieval of at least one of those images. Doing so, 59.77% and 57.84% of the entire database may be selected in approach 1 and approach 2, respectively.

5.2.4. Speeded Up Robust Features (SURF)

In SURF descriptor computation method, similarly to the SIFT method, there are some parameters that may be adjusted in order to adjust the characteristics of the SURF descriptors computed. In this work the parameter analyzed may be compared to the one studied in the SIFT feature descriptor method, since both regulate the number of keypoints computed. In Chapter 4 – Methodology the algorithm for SURF descriptor computation is carefully described and it is mentioned that one of the steps for keypoint localization is the use of a Hessian matrix approximation. The image blob response map over different scales, from which local maxima are detected, is computed from the approximated determinant of the Hessian matrix at each location x in the image. The maxima points are only considered in the following steps if greater than a threshold value, which assumes 0.0002 value in the SURF descriptor computation

algorithm developed by Bay et all [49]. As it is illustrated in Figure 47, when using this threshold value, very few SURF descriptors are computed. Therefore, other threshold values, smaller than this one, were studied and their effect in the descriptors computation was evaluated.

Figure 47: SURF keypoints obtained by discarding the interest points in the image response map computed with a Hessian matrix approximation lower than 0.0002.
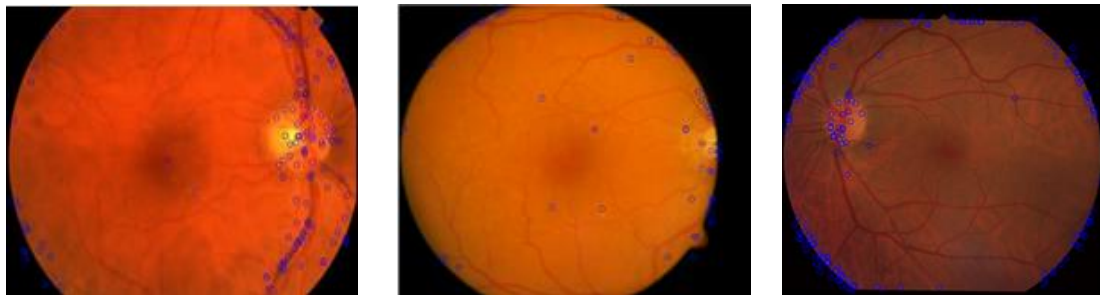
Figure 48: SURF keypoints obtained by discarding the interest points in the image response map computed with a Hessian matrix approximation lower than 0.0001.
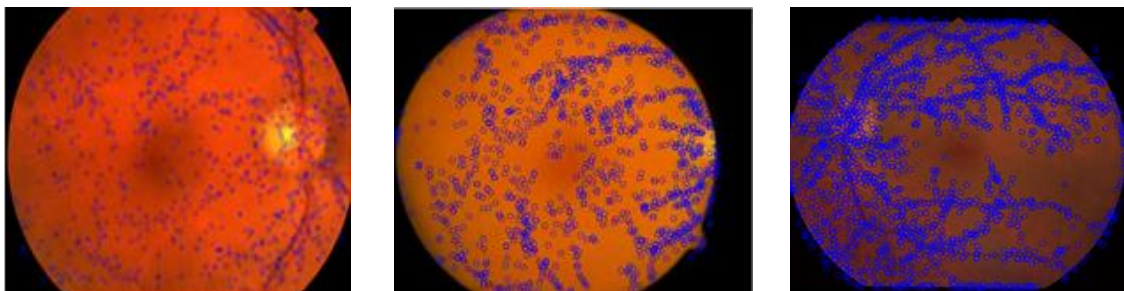
Figure 49: SURF keypoints obtained by discarding the interest points in the image response map computed with a Hessian matrix approximation lower than 0.00001.

Taking the above figures into consideration, it is easily concluded that smaller threshold values enable the computation of more SURF descriptors since less maxima points in the Hessian response map are discarded. . Although more threshold values were studied, only three are represented for illustrative purpose. Threshold value 0.00001 enables the computation of a reasonable number of SURF descriptors, considering their use in VTs. Moreover, most of the descriptors computed represent vessel tree details, which is one of the most important aspects that needs to be taken into consideration in this work. The posterior studies performed use SURF descriptors computed with this threshold value to discard those whose Hessian response is low.

Likewise to the procedure used in the others feature computation methods, SURF descriptors were organized in VTs for a fast search in the database of retinal images. VT with k=5 were built and their image retrieval performance was evaluated. Similarly to what was concluded with the SIFT method, this feature descriptor computation method enables the computation of a significant higher number of descriptors per image than regarding FD and WEF methods. Therefore a fewer number of images was used in this study. Besides the four k-fold groups of images for which there are available information about the corresponding patient of each image, more 1100 images were used to the VT training, instead of using more 3467 images as with FD and WEF method.

Table 40: SURF - Image retrieval performance of VT built with SURF descriptors computed from retinal images

K	Level	%of the Dataset Selected	Approach 1: % of situations in which ALL images of the same patient and eye are selected	Approach 2: % of situations in which at least ONE image of the same patient and eye is selected
	5	55.07%	99.89%	100.00%
5	6	54.85%	99.40%	99.79%
	7	52.31%	95.22%	97.83%

The use of VTs with SURF descriptors has proved to be more efficient in the retinal image retrieval than using SIFT descriptors. However, considering the increase in the image retrieval efficiency obtained when considering the retina vessel tree images with SIFT descriptors, the same approach was studied with SURF descriptors. Retinal vessel tree images were obtained in the same way as previously described, by using a Gaussian filter to smooth the binary vascular tree image. It is worth to mention that when applying SURF method to this type of images the threshold value considered to discard the less stable keypoints of the Hessian response map was 0.0002. Table 41 shows the image retrieval performance obtained with VTs using SURF descriptors computed from the smoothed retinal vessel tree images.

Table 41: SURF - Image retrieval performance of VT built with SURF descriptors computed from smoothed retinal vessel tree images

K	Level	%of the Dataset Selected	Approach 1: % of situations in which ALL images from the same patient are selected	Approach 2: % of situations in which at least ONE image from the same patient is selected
	11	49.19%	99.57%	99.83%
5	12	44.61%	99.24%	99.75%
	13	38.86%	98.63%	98.89%

By using SURF descriptors computed from the retinal vessel tree images a lower percentage of the database may be retrieved. For the retrieval of all images from the same patient of the image being queried 44.61% of the database is selected with a 5x12 VT. If considering the retrieval of only at least one of those images, 38.86% of the database may be selected with a 5x13 VT.

The removal of the most and the least populated nodes was performed and the results obtained are summarized in Tables 42 and 43.

Table 42: SURF - Image retrieval performance in retrieving ALL images of the same patient and eye by removing the most and the least populated nodes

	% of Nodes removed	% of images selected from the database	Approach 1: % of situations in which ALL images of the same patient and eye are selected
Most Populated Nodes	0.5%	44.47%	99.24%
	1.0%	44.21%	98.85%
	1.5%	43.89%	98.80%
	1.6%	43.89%	99.74%
Least populated nodes (After removal of the 1.5% most populated nodes)	12.0%	43.67%	98.80%
	14.0%	43.64%	98.78%
	16.0%	44.63%	98.73%

Table 43: SURF - Image retrieval performance in retrieving at least ONE image of the same patient and eye by removing the most and the least populated nodes.

	% of Nodes removed	% of images selected from the database	Approach 2: % of situations in which at least ONE image of the same patient and eye is selected
Most Populated Nodes	1.0%	38.32%	98.81%
	1.5%	38.07%	98.78%
	1.8%	37.99%	98.78%
	2.0%	37.91%	98.75%
Least populated nodes (After removal of the 1.8% most populated nodes)	10.0%	37.95%	98.78%
	12.0%	37.93%	98.78%
	13.0%	37.92%	98.74%

The removal of the 1.5% most populated and the 14.0% least populated nodes from the search step enabled a slight increase in the image retrieval performance in approach 1. The removal of the 1.8% most populated and the 12.0% least populated nodes from the search step enabled, as well, a slight increase in the image retrieval performance in approach 2.

As in SIFT method, the descriptors computed are very distinct not being possible to distinguish frequent and infrequent nodes. In fact more than 90% of the descriptors computed are unique in the set of descriptors. Therefore, similarly to the SIFT method, the analysis of the effect of the removal of frequent and infrequent descriptors from the search step was not performed. On the other hand, the attribution of a similarity score to the images retrieved was studied in this feature computation method. The similarity score computation is described in

Chapter 4 – Methodology. The similarity threshold defined to distinguish the most similar images to the image being queried from the other retrieved images is shown in Table 44 and 45 for both approaches of analysis.

Table 44: SURF - Similarity score analysis for image retrieval of ALL images of the same patient and eye performance		
Similarity Threshold Value	% of images selected from the database	% of situations in which ALL images of the same patient and eye are selected
1.0×10^{-4}	39.85%	98.78%
3.0×10^{-4}	39.29%	98.78%
5.0×10^{-4}	38.91%	98.73%

Table 45: SURF - Similarity score analysis for image retrieval of at least ONE image of the same patient and eye performance		
Similarity Threshold Value	% of images selected from the database	% of situations in which at least ONE image of the same patient and eye is selected
5.0×10^{-4}	33.90%	98.78%
5.5×10^{-4}	33.48%	98.77%
6.0×10^{-4}	33.07%	98.69%

From this last analysis, one may conclude that to ensure the retrieval of all images belonging to the same patient of the image being queried, a similarity threshold value of 3.0×10^{-4} may be used, resulting in the selection of 39.29% of the entire database. When considering approach 2, the retrieval of at least one image of the same patient and eye, the most appropriate similarity value is 5.5×10^{-4} resulting in the selection of 33.48% of the database.

5.3 Other methods to increase the image retrieval performance

Table 46 expresses the best image retrieval performances obtained for the retrieval of all or at least one image of the same patient and eye of an image being queried, using the different feature computation methods studied associated to VTs.

Taking these aspects into consideration, Fractal Dimension feature descriptor method is the one which enables the selection of a most reduced subset of images from the entire database to ensure a correct image retrieval in both of the approaches analyzed. The following studies were performed considering only approach 2, i.e. only the retrieval of one image belonging to the same patient and eye of the image being queried was required. This aspect not only enables a more efficiency image retrieval performance with less computational time but it also enables to explore other techniques for reducing even more the number of images that needs to be verified with a retina-based person identification method.

Table 46: **Image retrieval performance of the different methods employed considering the two different approaches.**

	Approach 1: % of the dataset selected to ensure the retrieval of all images of the same patient and eye	Approach 2: % of the dataset selected to ensure the retrieval of at least one image of the same patient and eye
FD	38.85%	32.78%
WEF	80.02%	60.24%
SIFT	59.77%	57.84%
SURF	39.29%	33.48%

5.3.1. Retrieved Images Ranking with WEF descriptors

Besides the fact that WEF descriptors do not perform well when used in VTs, the initial analysis performed has shown that it is very efficient in image ranking accordingly to their similarity to a query image. In this way, the algorithm herein proposed enables the retrieval of a subset of images, which is performed by using FD descriptors in VTs, with a posterior ranking, using WEF descriptors, of the images retrieved.

.
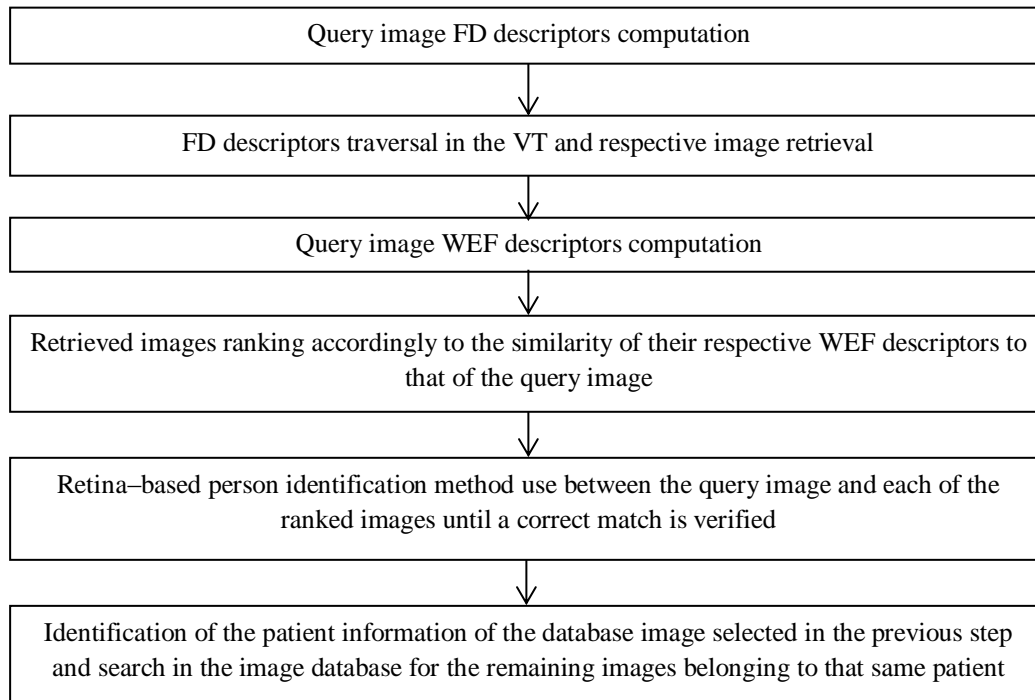
Figure 50: Flowchart of the algorithm used for image retrieval using FD descriptors in a VT and image ranking with WEF descriptors.

For this algorithm it is considered that database images are correctly indexed and solely one image of the same patient and eye of the query image needs to be selected from the database. Using FD descriptors with VTs was proved to enable the use of 32.78% of the entire database in a retina-based person identification method for the correct images identification. By ranking the retrieved images using WEF descriptors, that is, considering the WEF descriptor differences between each retrieved image descriptor and the query image descriptor, it is easily understood which are the most similar images to the one being queried. If when applying the retina-based person identification method one begins the search by the most similar retrieved images until a correct match is found, a major reduction of the total amount of images used in this step may be obtained. This fact was analyzed in the different subsets of images used in these studies and it was found that on average only 3.15% of the entire database needed to be used in the person identification method, until a correct match was verified. The flowchart illustrated in Figure 50 represents the major steps of this algorithm.

5.3.2. KNN Classification

For image analysis and classification some authors consider the use of various classifiers which are used in algorithms that involve object recognition. These classifiers include Bayes classifier, Support Vector Machines (SVM) and K-Nearest Neighbors (KNN) among others. In this work, the one considered was the KNN classifier. This method is used for object classification based on the k closest training examples. Although it is one of the simplest machine learning algorithms, it is suitable for the problem of patient identification due to the number of classes considered, that is, the number of patients and eyes considered. In other classifiers, SVM for instance, that map the samples in a space in which it is possible to define decision boundaries to separate the samples of each class, it is important to use a significant number of examples of each class in the training phase of the classifier. However, in this work the number of retinal images belonging to the same patient and eye is not sufficient taking into consideration the number of classes: there are a maximum of five images of the same patient and 573 different patients. Due to the significant number of classes it is not expected an excellent classification. Besides this fact, the disadvantages of SVM classification include the significant computational time required for both training and testing phase. Considering this, the use of a KNN classification was found to be suitable for patient and eye identification since the decision is based on the most similar training example, which has high probability of belonging to the same patient and eye of the one being queried [72].

The set of WEF and FD descriptor vectors, computed for each of the image subregions, were joined together to origin a single row vector. Although rotation invariance may be compromised with this technique, it was found that this technique still performed reasonably well taking into consideration the characteristics of the datasets used. The same datasets of images used for the VT training and testing were used in this study. In Table 47 it is represented the classifier performance in identifying the correct patient and eye by searching for the k-closest training descriptors. The best classification of the testing descriptors was obtained when using WEF descriptors to represent the image and considering only the class (patient and eye) of the closest training descriptor. It was already mentioned in this work that although WEF descriptors enable a very efficient image ranking accordingly to the images similarity to the one being queried, they differ very little from each others. Considering this, it is not surprising that

the use of only the closest training descriptor enables a better descriptor classification than using a higher number of neighbor descriptors in which to base the testing descriptor classification. By using WEF descriptors with a KNN classifier with k=1, 54.97% of the testing descriptors respective patients were correctly identified.

Table 47: Patient and Eye Identification Performance with a KNN classifier

		% of query images for which was performed a correct patient and eye identification
WEF	K=1	54.97%
	K=3	54.62%
	K=5	49.09%
Fractal Dimension	K=1	38.39%
	K=3	38.39%
	K=5	36.82%

Although the classification performance was not enough for the problem of patient and eye identification solution, it was considered in a preliminary classification in the final algorithm proposed in this thesis, as it may be seen below.

5.4 Proposed Algorithm

Considering the performance of the methods studied for an efficient image retrieval, it was concluded that the best algorithm for this purpose would be a combination of the most efficient methods studied. As aforementioned, using VTs for retinal image retrieval has proved to be a good approach in the identification of retinal images belonging to the same patient and eye of an image being queried. From the four descriptor computation methods used to build VTs, FD method was the one which enabled the selection of a smaller percentage of the database to use in a retina-based person identification method. Moreover, other studies performed have indicated that WEF descriptors may be useful in an efficient images ranking accordingly to their similarity to a query image and that KNN classification could be used in a preliminary classification for patient and eye identification.

Considering all the studies performed in this work, a final algorithm is proposed for the correct retrieval of database images belonging to the same patient of a retinal image being queried. Since the goal was to obtain the most efficient algorithm possible, it was considered that the database images are correctly labeled and labeling errors may only occur with the query images.

Initially, the binary retinal vessel tree image of the query retinal image is computed. WEF descriptors of this image are obtained according to the algorithm described in Chapter 4-Methodology. These descriptors are joined in a single row vector descriptor and it is used for a KNN classification with k=1 in order to identify the patient and eye of the query retinal image. From the analysis performed, there is a nearly 55% chance the correct patient is identified in this procedure. To verify whether the identified patient and eye is the same of that of the query image, a retina-based person identification method is used to compare the query retinal image with one of the database images belonging to that patient. In case a correct match is verified, the remaining database images belonging to that patient are retrieved. Otherwise, FD descriptors are computed from the retina vessel tree image in order to search in a VT for the most similar images to the image being queried. Posteriorly to the VT image retrieval, the retrieved images are ranked accordingly to the similarity of their WEF descriptors to that of the query image. A retinal-based person identification algorithm is used to verify if each of the retrieved and ranked retinal images belong to the same patient of the image being queried. The first image considered in this verification step is the most similar one in the ranked list of retrieved images and so on. When a correct match is verified the remaining database images belonging to the same patient of the image considered are retrieved. The retina-based person identification method used, image registration, has an efficiency performance of 98.79% in identifying correct matches in images belonging to the same individual and eye and incorrect matches otherwise. It is important to remember that the final algorithm proposed must present a nearly 99% efficiency in the retrieval of the correct images. Therefore, considering the performance efficiency of the retina-based person identification method used it was defined that, in case any correct match is verified between the query image and each of the retrieved images, the entire database is used in this step, which only occurs with nearly 1% of the query images studied. In this way, the algorithm performance in retrieving the correct images is 98.79%, which corresponds to the person identification method performance.

One more aspect needs to be considered as well. Whenever new retinal images are added to the database it is necessary to re-train the KKN classifier but not the VT structure. To actualize the VT with the new descriptors, they are traversed in the VT in the same way the query descriptors do and the centroid of each node is simple re-computed. Figure 51 illustrates the major steps of the proposed algorithm.
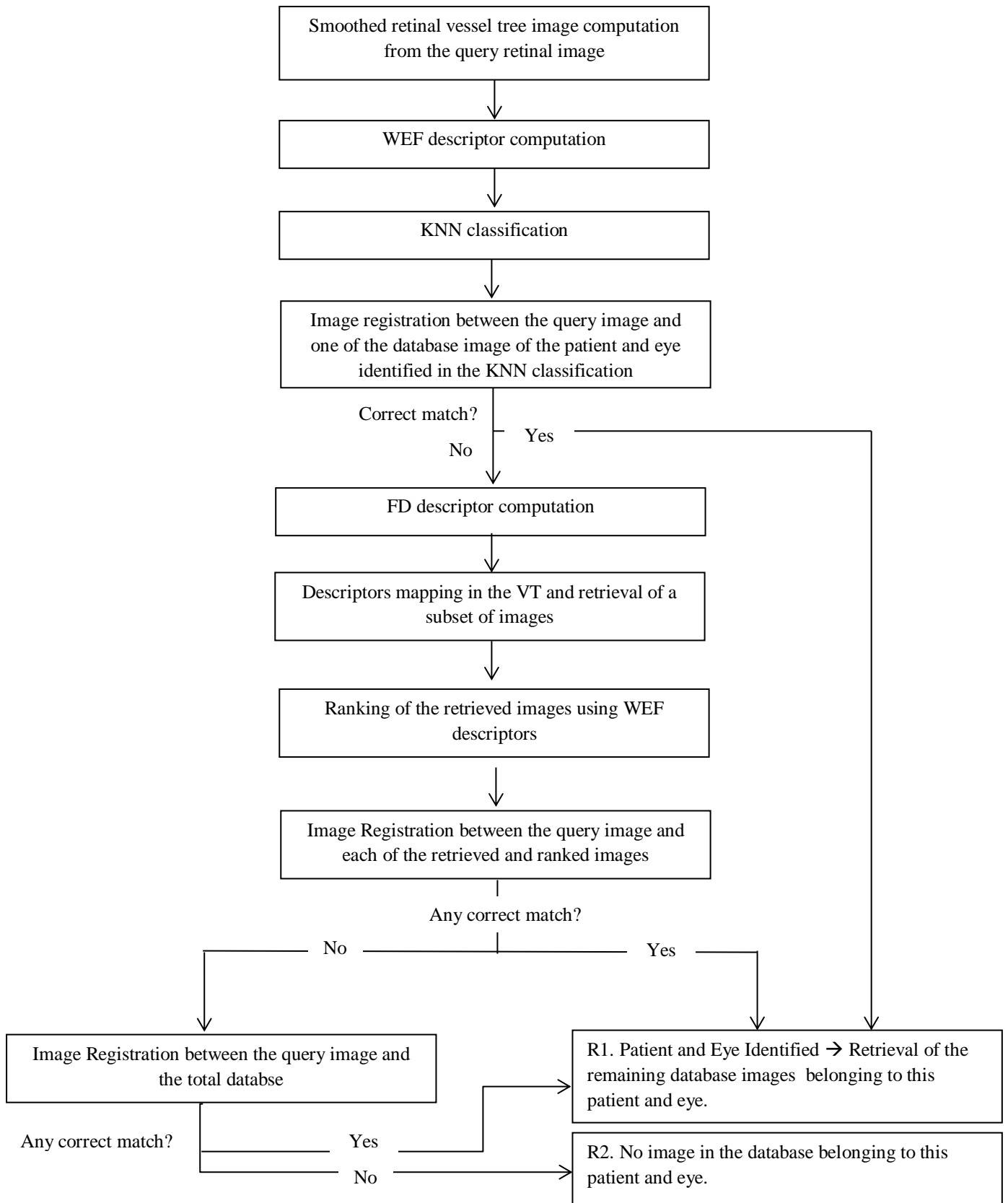
Figure 51: Flowchart of the proposed algorithm.

5.5 Algorithm Computation Complexity

As mentioned before, the main purpose of the development of this work, i.e. the development of an algorithm for the identification in a large image database of the retinal images belonging to the same patient and eye of a retinal image being queried, is the integration in an automated clinical support decision software for eye pathology detection and monitoring. It is important to note that the application environment of the target software puts some constrains on the computational power required by the proposed algorithm. To efficiently verify whether the retrieved images in fact belong to one patient and eye, a retina-based person identification method is used between the query image and each of the retrieved images. The person identification method used is out of the scope of this project, but as it may be confirmed, the computation time required for this method directly influences the overall final algorithm computation time. The retina-based person identification method used in this work was the quality assessment of the image registration between the query and the retrieved images. Although this method requires on average 28.92 seconds to compare a pair of images, since it was not included on the goals of this work, it was used to evaluate the proposed algorithm performance. However, for the integration of the algorithm in a real-time application environment, further developments in this step need to be considered. That said, the proposed algorithm performance was evaluated in comparison to the use of the retina-based person identification method with the entire image database instead of only considering the subset of retrieved images.

Table 48 shows the performance results of the different approaches considered throughout this work: the use of the person identification method alone, the use of the person identification method with the retrieved images obtained by using a VT with Fractal Dimension descriptors, the posterior ranking of the retrieved images with WEF descriptors and the use of a KNN classifier for a preliminary patient and eye identification. For each approach two parameters were analyzed: the percentage of the total image database used in the verification step, i.e. the retina-based person identification method, and the overall computation time required for the algorithm to retrieve the correct images. The computation time reduction is computed in comparison to the first approach: the use of the person identification method in the entire image database. As expected, significant computation time reductions were obtained in the approaches developed. The use of a VT with Fractal Dimension descriptors to represent each database image and the search in this structure for the most similar images to the one being queried enables the use of 32.78% of the entire database in nearly 99% of the query images analyzed and the total database in the remaining 1% . Therefore, in Table 48 the performance values expressed are mean performance values enclosing these two outputs. Although additional computations are required for the VT use, their computation time is insignificant when compared to the overall computation time reduction of the algorithm. Since the VT training and the image database descriptors computation are performed offline, not contributing to the computation time of the algorithm, it only takes 1.893s to compute the Fractal Dimension descriptors of the query image and 0.1711s to search in the VT for the most similar images to the one being queried. A 66.54% computation time reduction was obtained with this approach. The use of WEF descriptors to rank the retrieved images accordingly to their similarity to the query image enables the use, on average, of 3.15% of the image database, in 99% cases, since

the algorithm ends when a correct match between the query and a retrieved image is found. In this approach only additional 0.3973s, for WEF descriptors computation of the query image, and 0.016s, for the retrieved images ranking, are required in the total computation time of the algorithm. The total time of computation of this approach was found to represent a 95.88% time reduction when compared to the use of image registration with all database images. The final proposed algorithm only differs from the previous one in the initial KNN classification. As mentioned before, the KNN classification proposed enables the correct identification of the patient on 54.97% of the classifications performed. Therefore, it is used in a preliminary phase of the algorithm and 0.1224s are required to perform this classification. If a correct match between the query image and one of the images from the patient identified in the KNN classification is found, the algorithm ends and it is only performed an image registration between a pair of images, which takes 29.43s (considering the computation required for WEF descriptors computation and KNN classification). This represents a 97.57% time reduction of the overall computation time of the algorithm. In case it is not found a correct match, in 45.03% of the cases, with the KNN classification, the previous approach is considered.

Table 48: **Computation time for comparison purposes of the different approaches considered**

	% of the Dataset used in the Registration Step	Total Time of computation for identifying one image of the same patient and eye in the database (s)	Computation time reduction (s)
Image Registration	100.000%	1.463×10^5	----
VT + Image Registration	33.452%	4.895×10^4	66.54%
VT + retrieved images ranking+ Image Registration	4.115%	6.029×10^3	95.88%
KNN classification + VT + retrieved images ranking+ Image Registration	0.042%	3.548×10^3	97.57%

Chapter 6

Conclusion

The main contribution of this thesis was the use of Vocabulary Trees in the search of retinal images belonging to the same patient and eye of a retinal image being queried in a large database. Although VTs are widely used in large image databases, their use with retinal images was found to be an unexplored issue. Therefore, this thesis contributes with a novel approach for retinal image database search. The significant efficiency of VTs in image analysis and classification were the main motivation for the use of these database search structures in this work. Two different types of methods for image representation in the VTs were studied. Fractal Dimension and WEF descriptor computation methods were studied due to the fact these descriptors are used in retina-based person identification methods. This means that they enable an efficient retinal image representation, which is a crucial aspect in this work. Apart from the FD and WEF methods, SIFT and SURF descriptor computation methods were also studied. The use of these descriptors with VTs is referred in the literature with excellent performance results. Besides this, they enable the computation of a different type of descriptors. These methods detect interest points in the image and represent them with descriptors that are invariant to several conditions related to image acquisition, such as image rotation, scale and illumination.

The retina vessel tree is the biometric feature considered in this work, which enables to distinguish retinal images from the same individual from retinal images from different people. This biometric feature is unique for each person and usually do not suffer modifications through time. However, it was found that some eye pathology characteristic lesions could introduce modifications in the vascular tree. Since the target application which was the main motivation for the development of this work is an automated clinical support decision software for eye pathologies detection and monitoring, it was studied the effect of AMD and DR characteristic lesions, two of the most common eye pathologies, on the image retrieval performance by using different descriptor computation methods. Fractal Dimension and WEF descriptors were found to be the most robust methods to represent retinal images with eye lesions, mainly due to the fact they are computed from the binary retina vessel tree image. In this way, the eye lesions effect in the retina vascular tree is not so evident as in the retinal image.

In order to use the proposed algorithm in a real-time application environment the retina-based identification method used needs to be optimized, which was out of the scope of this thesis. Even so, the developed algorithm that this document addresses represents a contribution to retinal image retrieval for the identification of images belonging to the same individual of the one being queried, whose performance was demonstrated by the already discussed results.

The VT search with VT built with Fractal Dimension descriptors proved to be highly efficient in the retrieval of images belonging to the same patient of an image being queried. As

previously mentioned in this work, the final algorithm includes not only the VT image retrieval but also the retrieved images ranking accordingly to their similarity to the one being queried. Moreover, an initial KNN classification is performed and in nearly 55% cases it is enough for the correct image retrieval. The final algorithm proposed in this thesis, which is carefully described in Chapter 5- Results and Discussion, enables the retrieval of database images belonging to the same patient of the query image with a 98.79% efficiency performance and representing a 97.57% computation time reduction when compared to the use of retina-based person identification method solely.

Chapter 7

Future Work

The selection in a large retinal image database of the most similar images to a query retinal image is a problem for which there is not any significant proposed solution in the literature. Considering the fact this is an unexplored issue, future research for the development of a different solution from the one proposed in this work would be interesting.

Even considering this proposed algorithm, several other image feature computation methods could be implemented for retinal image accurate representation. Other Vocabulary Tree search techniques could as well be studied to improve either the image retrieval efficiency and the overall computation time of the algorithm.

The main aspect that needs to be addressed in order to adjust the proposed algorithm to a real-time application is the retina-based method used to verify whether two retinal images belong in fact to the same individual and eye. In this work, image registration was implemented since it enables an efficient identification of correct matches between pairs of retinal images. Although efficient, a less computationally complex method must be implemented to make it feasible to use this algorithm in a real-time application.

Appendix A

A.1. FD Descriptor Computation Method Flowchart

1 - Vascular Tree Extraction

2-Transformed images computation

3- Image Partitioning in n blocks

for m=1:6

for k=1....n

4- Fractal Dimension value computation (x_{mk})

5- Computation of the descriptor Vector $D_k=[x_{1k}, x_{2k}, x_{3k}, x_{4k}, x_{5k}, x_{6k}]$ associated to block k.

X_{m1} X_{m2}

X_{m3} X_{m4}

A.2. WEf Descriptor Computation Method Flowchart

for L=1..l   for m=1:3   for k=1….n

A.3. SIFT Descriptor Computation Method Flowchart

1 - Vascular Tree Extraction and low-pass Gaussian filtering

2 – Image scale space computation, $L(x,y,\sigma)$.

3 – Difference-of-Gaussian computation, $D(x,y,\sigma)$.

4 – Extrema detection in $D(x,y,\sigma)$ of n keypoints

5 – Accurate keypoint location (x,y) and scale s computation

6 – Keypoint orientation ɵ assignment

7 – Computation of the Keypoint 128-length descriptor

For i=1:n   If $n_i > 0.01$

A.4. SURF Descriptor Computation Method Flowchart

1 - Vascular Tree Extraction and low-pass Gaussian filtering

2 – Image scale space computation, using a Hessian matrix approximation

$$\mathcal{H}(\mathbf{x}, \sigma) = \begin{bmatrix} L_{xx}(\mathbf{x}, \sigma) & L_{xy}(\mathbf{x}, \sigma) \\ L_{xy}(\mathbf{x}, \sigma) & L_{yy}(\mathbf{x}, \sigma) \end{bmatrix}$$

3 – Extrema detection of n keypoints

4 – Accurate keypoint location (x,y) and scale s computation

For i=1:n

If $n_i > 0.0002$

5 – Keypoint orientation assignment ө

6 – Computation of the Keypoint 64-length descriptor

$$\sum dx$$
$$\sum |dx|$$
$$\sum dy$$
$$\sum |dy|$$

References

[1] Harvey PT, "Common Eye Diseases of Elderly People: Identifying and Treating Causes of Vision Loss", *in PubMed - US National Library of Medicine, National Institutes of Health,* vol. 49, No. 1, pp. 1-11, Jan - Feb 2003.

[2] American Optometric Association, "Diet, Nutrition and Eye Health", URL:*<http://www.aoa.org/documents/nutrition/Diet_Nutrition_Eye_Health_booklet.pdf>,* Last accessed date: Jun. 2013.

[3] Bupa, "Age-related Macular Degeneration", URL:*<http://www.bupa.co.uk/individuals/health-information/directory/a/age-related-macular-degeneration>,* Last accessed date: Jun. 2013.

[4] International Diabetes Federation, "Diabetes Atlas 5th edition", URL: *<http://www.idf.org/media-events/press-releases/2011/diabetes-atlas-5th-edition>,* Last accessed date: Jun. 2013.

[5] National Institutes of Health - National Eye Institute, "Facts About Diabetic Retinopathy," URL:*<http://www.nei.nih.gov/health/diabetic/retinopathy.asp>,* Last accessed date: Jun. 2013.

[6] J. Goh, L. Tang, G. Saleh, L. Al turk, Y. Fu and A. Browne, "Filtering Normal Retinal Images for Diabetic Retinopathy Screening Using Multiple Classifiers", *in Proceedings of 9th International Conference on Information Technology and Applications in Biomedicine (ITAB 2009),* Larnaca, pp.1-4, Nov 2009.

[7] J. M. Dias, C. M. Oliveira and L. A. Cruz, "Retinal Image Quality Assessment Using Generic Image Quality Indicators", *in Information Fusion Journal, Special Issue on Medical Image Computing, vol.13, No.0, pp.1-18, Aug.* 2012.

[8] H. F. Jelinek and M. J. Cree, "Automated Image Detectino of Retinal Pathology", *CRC Press*, 2010.

[9] N. Patton, T. M. Aslam, T. MacGillivray, I. J. Deary, B. Dhillon, R. H. Eikelboom, K. Yogesan and I. J. Constabke, "Retinal image analysis: Concepts, applications and potencial", *in Elsevier - Progress in Retinal Eye Research,* vol. 25, pp. 99-127, 2006.

[10] Critical Health, "Retmarker", URL: *<http://www.retmarker.com/>*, Last accessed date: Jun. 2013.

[11] A. Bhuiyan, E. Lamoureux, B. Nath, K. Ramamohanarao and T. Y. Wong, "Retinal Image Matching Using Hierarchical Vascular Features", *in Computational Intelligence and Neuroscience, vol.2011, Article ID 749054, 7pages,* Oct. 2011.

[12] M. F. Zibran, "Eye Based Authentication: Iris and Retina Recognition", Technical Report #2011-04, Univesity of Saskatchewan, 2009.

[13] J. Sivic and A. Zisserman, "Video Google: A Text Retrieval Approach to Object Matching in Videos", *in Ninth IEEE International Conference on Computer Vision*, vol.2, pp.1470-1477, France, Oct. 2003.

[14] X. Wang, M. Yang, T. Cour, S. Zhu, K. Yu and T. X. Han, "Contextual Weighting for Vocabulary Tree based Image Retrieval", *in Proceedings of IEEE International Conference on Computer Vision*, Barcelona, Spain, Nov. 6-13, 2011.

[15] K. Rogers, "The Eye - The Physiology of Human Perception", *Britannica Educational Publishing*, 2011.

[16] Helga Kolb, "Gross Anatomy of the Eye", *Webvision - The Organization of the retina and Visual System,* Jan. 2012, URL: *<http://webvision.med.utah.edu/book/part-i-foundations/gross-anatomy-of-the-ey/ >*, Last accessed date: Jun. 2013.

[17] R. E. Hamor, "Medical and surgical Management of Corneal Disease", URL:*<http://www.dcavm.org/12junenotes.pdf>*, Last accessed date: Jun. 2013.

[18] Glaucoma Research Foundation, "What is Glaucoma - Eye Anatomy", Jan. 2012, URL:*<http://www.glaucoma.org/glaucoma/anatomy-of-the-eye.php>*, Last accessed date: Jun. 2013.

[19] VisionAware-Resources for Independent Living with Vision Loss, "Eye Health: Anatomy of the Eye", URL:*<http://www.visionaware.org/section.aspx?FolderID=6&SectionID=116&DocumentID=5787>,* Last accessed date: Jun. 2013.

[20] St. Luke's - Cataract Laser Institute, "St- Luke's Eye - Anatomy", URL:*<http://www.stlukeseye.com/anatomy/choroid.html>*, Last accessed date: Jun. 2013.

[21] University of Michigan Kellogg Eye Center, "Patient Care - The Eye", URL:*<http://www.kellogg.umich.edu/patientcare/conditions/anatomy.html>,* Last accessed date: Jun. 2013..

[22] Purves D., Augustine G.J., Fitzpatrick D., "Neuroscience-Functional Specialization of the Rod and Cone Systems", Sunderland (MA), 2011.

[23] Kolb H., "Simple Anatomy of the Retina", In *Webvision: The organization of the Retina and Visual System*, UT: University of Utah Health Sciences Center, URL:<http://webvision.med.utah.edu/book/part-i-foundations/simple-anatomy-of-the-retina/ >, Last accessed date: Jun. 2013.

[24] Bausch + Lomb, "Your Eye Concerns - Diseases and Disorders", URL:*<http://www.bausch.com/en/Eye-concerns/eye-diseases-and-disorders>*, Last accessed date: Jun. 2013.

[25] WebMD Medical Reference, "Eye Health Center - Common Eye Problems", URL:*<http://www.webmd.com/eye-health/common-eye-problems>*, reviewed by Alan Kozarsky, Last accessed date: Jun. 2013.

[26] Diabetes UK, "Guide to diabetes - What is diabetes?", URL: *<http://www.diabetes.org.uk/Guide-to-diabetes/Introduction-to-diabetes/What_is_diabetes/>,* Last accessed date: Jun. 2013.

[27] NHS choices, "Diabetic Retinopathy", URL: *<http://www.nhs.uk/conditions/diabetic-retinopathy/Pages/Introduction.aspx>*, Last accessed date: Jun. 2013.

[28] Florida Retina Center, "Vascular Retina Disorders - Retinal Microaneurysm", URL:*<http://www.retinaandmacula.com/retina/retinal_microaneurysm_bonita_springs.htm>, Last accessed date: Jun. 2013.*

[29] Centers for Disease Control, "Diabetes Disabling Disease to Double by 2050", URL:<http://www.cdc.gov/nccdphp/publications/aag/ddt.htm >, Last accessed date: Jun. 2013.

[30] RNIB, "Age-related macular degeneration (AMD)", URL:*<http://www.rnib.org.uk/eyehealth/eyeconditions/conditionsac/Pages/amd.aspx>,* Last accessed date: Jun. 2013.

[31] AMD.org - Macular Degeneration Partnership, "What are Drusens", URL:*<http://www.amd.org/what-is-amd/dry-amd/119-drusen.html>,* Last accessed date: Jun. 2013.

[32] Lylas G. and Mogk M.D., "The difference between wet and dry age-related macular degeneration," *American Foundation for the Blind - Reader's Digest Partners for Sight, URL:<http://www.visionaware.org/section.aspx?FolderID=6&SectionID=134&DocumentID=5972>,* Last accessed date: Jun. 2013.

[33] National Institutes of Health - National Eye Institute, "Facts about age-related macular degeneration", URL:*<http://www.nei.nih.gov/health/maculardegen/armd_facts.asp>, Last accessed date: Feb.2013*

[34] Master Eye Associates, "Drusen", URL: *<http://www.mastereyeassociates.com/eye-diseases-treatments/drusen/>*, Last accessed date: Jun. 2013.

[35] M. H. Goldbaum, N. P. Katz, S. Chaudhuri and M. Nelson, "Image Understanding for Automated Retinal Diagnosis", *in Proceedings of Annual Symposium on Computer Application in Medical Care*, pp.756-760, Nov. 1989.

[36] V. Malmsten, "Eye Scans - Authentication with Biometrics", *Global Information Assurance Certification Paper, SANS Institute 2001-2002.*

[37] K. Saraswathi, B. Jayaram and D. R. Balasubramanian, "Retinal Biometrics based Authentication and Key Exchanged System", *in International Journal of Computer Applications ,* Vol. 19, No. 1, Apr. 2011.

[38] C. Simon and I. Goldstein, "A New Scientific Method of Identification", *in New York State Journal of Medicine,* Vol. 35, No. 18, pp. 901-906, 1935.

[39] P. Tower, "The fundus Oculi in monozygotic twins: Report of six pairs of identical twins", *Archives of Opthalmology,* Vol. 54, pp. 225-239.

[40] R. R. Hill, "Retina Identification", *Biometrics: Personal Identification in Networked Society. Kluwer Academic Press, Boston, Springer US,pp.123-141,* 2002.

[41] S. M. Kabir, R. Rahman, M. Habib and M. R. Khan, "Person Identification by Retina Pattern Matching", *in Proceedings of 3rd International Conference on Electrical & Computer Engineering ICECE*, Dhaka, Bangladesh, pp.522-525, Dec. 2004.

[42] M. Shahnazi, M. Pahlevanzadeh and M. Vafadoost, "Wavelet Based Retinal Recognition", *in Proceedings of 9th International Symposium on Signal Processing and Its Aplications ISSPA*, Sharjan, pp.1-4, Feb. 2007.

[43] H. Farzin, H. Abrishami-Moghaddam and M.-S. Moin, "Research Article - A Novel Retinal Identification System," *in Hindawi Publishing Corporation EURASIP Journal on Advances in Signal Processing,* Vol. 2008, Feb. 2008.

[44] S. Sukumaran and D. M. Punithavalli, "Retina Recognition Based on Fractal Dimension", *in IJCSNS International Journal of Computer Science and Network Security,* Vol. 9, no. 10, Oct. 2009.

[45] Z. Hai-ying and X. Z.-g. Penghong, "A Texture Feature Extraction Based on Two Fractal Dimensions for Content-based Image Retrieval", *in Proceedings of World Congress on Computer Science and Information Engineering, Vol.3, pp.117-121, Mar.*2009.

[46] K. Fukuta, T. Nakagawa, Y. Hayashi, Y. Hatanaka, T. Hara and H. Fujita, "Personal Identification Based on Blood Vessels of Retinal Fundus Images", *in Proceedings of Medical Imaging 2008: Imaging Processing. Proc. of SPIE,* Vol. 6914, Mar. 2008.

[47] M. Ortega and M. G. Penedo, "Retinal vessel tree as biometric pattern Biometrics*", edited by Jucheng Yang*, Jun. 2011.

[48] D. G. Lowe, "Distinctive image features from scale-invariant keypoints", *in International Journal of COmputer Vision,* Vol. 60, no. 2, pp. 91-110, 2004.

[49] H. Bay, A. Ess, T. Tuytelaars and L. V. Gool, "Speeded-Up Robust Features (SURF)", *in Computer Vision and Image Understanding*, Vol.110, issue:3, pp.346-359 , Jun. 2008.

[50] L. Wei, L. Pan, L. Lin and L. Yu, "The Retinal Image Registration Based on Scale Invariant Feature", *in Proceedings of the 3rd International Conference on Biomdecial Engineering and Informatics (BMEI 2010) Vol.2, pp.639-643,* Oct. 2010.

[51] P. C. Cattin, H. Bay, L. V. Gool and G. Székely, "Retina Mosaicing Using Local Features", *in Proceddings of the Medical Image Computing and Computer-Assisted Intervention*, Vol.9, issue:2 , pp.185-192, 2006.

[52] J. Zheng, J. Tian, K. Deng, X. Dai, X. Zhang and M. Xu, "Salient Feature Region: A New Method for Retinal Image Registration", *in IEEE Transactions on Information Technology in Biomedicine,* Vol. 15, no. 2, pp. 221-232, Mar 2011.

[53] A. P. Condurache, A. Mertins and J. Kotzerke, "Robust Retina-Based Person Authentication Using the Sparse Classifier", *in Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*, pp.1514-1518, Bucharest, Aug. 2012.

[54] M. Sabaghi, S. R. Hadianamrei, A. Zahedi and M. N. Lahiji, "A New Partitioning Method in Frequency Analysis of the Retinal Images for Human Identification", *in Journal of Signal and Information Processing,* Vol. 2, pp. 274-278, Nov 2011.

[55] R. Yong, T. S. Huang and S.-F. Chang, "Image Retrieval: Current Techniques, Promising Directions and Open Issues", *in Journal of Visual Communication and Image Representation,* Vol. 10, pp. 39-62,

1999.

[56] R. Baeza-Yates and B. Ribeiro-Neto, "Modern Information Retrieval", *Addison-Wesley Longman Publishing Co.*, Inc. Boston, USA, 1999.

[57] K. Mikolajczyk and C. Schmid, "An affine invariant interest point detector", *in Proceedings of the 7th European Conference on Computer Vision (ECCV),* Vol.2350, pp.128-142, Springer-Verlag, 2002.

[58] J. Matas, O. Chum, M. Urban and T. Pajdla, "Robust wide baseline stereo from maximally stable extremal regions", *Image and Vision Computing,* Vol.22, issue:10, pp. 761-767, Sep. 2004.

[59] D. Nistér and H. Stewénius, "Scalable Recognition with a Vocabulary Tree", *in IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol.2, pp.2161-2168, 2006.

[60] O. Chum, J. Philbin, J. Sivic, M. Isard and A. Zisserman, "Total recall: Automatic query expansion with a generative feature model for object retrieval", *in Proceedings of the 11th International Conference on Computer Vision (ICCV),* pp.1-8, Rio de Janeiro, Oct. 2007.

[61] J. Philbin, O. Chum, M. Isard, J. Sivic and A. Zisserman, "Object retrieval with large vocabularies and fast spatial matching", *in Proceedings of Conference on Computer Vision and Pattern Recognition*, pp.1-8, Minneapolis, Jun. 2007.

[62] J. Philbin, O. Chum, M. Isard, J. Sivic and A. Zisserman, "Lost in Quantization: Improving Particular Object Retrieval in Large Scale Image Databases", *in Proceedings of Conference on Computer Vision and Pattern Recognition,* pp.1-8, Jun. 2008.

[63] S. Sankar and T. Thomas, "Fractal Features Based on Differential Box Counting Method for the Categorization of Digital Mammograms", *in International Journal of Computer Information Systems and Industrial Management Applications (IJCISIM),* Vol. 2, pp. 11-19, 2010.

[64] S. H. Rezatofighi, A. Roodaki and H. A. Noubari, "An Enhanced Segmentation of Blood Vessels in Retinal Images Using Contourlet", *in 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Vancouver, British Columbia, pp.3530-3533, Vancouver, Aug. 2008.

[65] E. Candès, L. Demanet, D. Donoho and L. Ying, "Fast Discrete Curvelet Transforms", *Society for Industrial and Applied Mathematics (SIAM), Multiscale Modeling & Simulation,* Vol. 5, No. 3, pp-861-899, Mar. 2006.

[66] P. Bankhead, C. N. Schofield, J. G. McGeown and T. M. Curtis, "Fast Retinal Vessel Detection and Measurement Using Wavelets and Edge Location Refinement", *PLoS ONE,* vol. 7, no. 3, Mar. 2012.

[67] S. Hong and D. Huidong, "Fractal Dimension Applied in Texture Feature Extraction in X-Ray Chest Image Retrieval", *in International Conference on Inforamtion and Automation*, pp. 841-845 Shenyang, Jun. 2012.

[68] D. T. Lee and A. Yamamoto, "Wavelet Analysis: Theory and Applications", *in Hewlett-Packard Journal,* Vol.45, pp. 44-52, Dec. 1994.

[69] T. Lindberg, "Scale-space theory: Abasic tool for analysing structures at different scales", *in Journal of Applied Statistics,* Vol. 21, No. 2, pp. 224-270, 1994.

[70] S. Edelman, N. Intrator and T. Poggio, "Complex Cells and Object Recognition", *in Proceedings of Neural Information Processing Systems (NIPS'97) - Visual Processing,* 1997.

[71] B. Girod, V. Chandrasekhar, R. Grzeszczuk and Y. A. Reznik, "Mobile Visual Search: Architectures, Technologies and the Emerging MPEG Standard", *IEEE MultiMedia ,* Vol. 18, No. 3, pp. 86-94, Mar. 2011.

[72] J. Kim, B. S. Kim and S. Savarese, "Comparing Image Classificatino Methods: K-Nearest-Neighbor and Support-Vector-Machines", *in Proceedings of the 6th WSEAS International Conference on Computer Engineering and Applications and in Proceedings of the 2012 American Conference on Applied Mathematics*, pp.133-138, Wisconsin, USA, 2012.