Detection of diabetic retinopathy lesions in color retinal images

Thesis submitted in partial fulfillment
of the requirements for the degree of

Master of Science (by Research)
in
Computer Science

by

Keerthi Ram
200607013
keerthiram @ research.iiit.ac.in

Centre for Visual Information Technology
International Institute of Information Technology
Hyderabad - 500 032, INDIA
January 2011

International Institute of Information Technology
Hyderabad, India

CERTIFICATE

It is certified that the work contained in this thesis, titled “Detection of diabetic retinopathy lesions in color retinal images” by Keerthi Ram, has been carried out under my supervision and is not submitted elsewhere for a degree.

Date

Adviser: Prof. Jayanthi Sivaswamy

to Lord Muruga and all my Gurus

Acknowledgments

This thesis is but a humble river, whose tributaries and formative springs span an intelligentsia of admirable brilliance. Foremost is my supervisor Dr.Jayanthi, whose sincerety, openness to inspiration, and whose capability to rise to the test, are traits worthy of emulative attempt. I have gained many a life lesson from her, on topics ranging from credibility to enthusiasm, punctuality to objective criticism – lessons which I am still trying to satisfactorily imbibe, and am indebted to humanity in my execution.

Much inspiration came in subtle forms, through many a cordial interaction, with my peers, my lab-mates, my teachers, my close friends, and totally unknown friendly people. A lot of appreciation is due to my wonderful teachers at IIIT-H, stellar enrapturing performers of a mystic and enviable art. It is but natural for a glossy-eyed witness such as me to wish to rise, to improve and to excel, when amidst them.

If this thesis ever amounts to anything more than a scientist's dissertation, it is symptomatic evidence of the charm of my teachers (school days onward), the faith of my well-wishers, the sheer attraction of "the Work", and the existential tautology of questions worthy of research in my chosen Field. I here refrain from naming all the exceptional individuals who have played a role, to any noticeable extent, in my thought process and my work. I shall, by way of gratitude, avail myself of their continued contact and inspiration, and be for them an ever-yielding well of friendship and trust.

Subtle memories of thanks unsaid, appreciations withheld, gratitude unshown,
Krishna! bless me that my smile convey all these - ssk

Abstract

Advances in medical device technology have resulted in a plethora of devices that sense, record, transform and process digital data. Images are a key form of diagnostic data, and many devices have been designed and deployed that capture high-resolution in-vivo images in different parts of the spectrum. Computers have enabled complex forms of reconstruction of cross-sectional/ 3D structure (and temporal data) non-invasively, by combining views from multiple projections. Images thus present valuable diagnostic information that may be used to make well-informed decisions.

Computer aided diagnosis is a contemporary exploration to apply computers to process digital data with the aim of assisting medical practitioners in interpreting diagnostic information. This thesis takes up a specific disease: diabetic retinopathy, which has visual characteristics manifesting in different stages. Image analysis and pattern recognition have been used to design systems with the objective of detecting and quantifying the extent. The quantitative information can be used by the practitioner to stage the disease and plan treatment, track drug efficacy, or make decisions on course of treatment or prognosis.

The generic task of image understanding is known to be computationally ill-posed. However adding domain constraints and restricting the size of the problem make it possible to attempt solutions that are useful. Two basic tasks in image understanding : detection and segmentation, are used. A system is designed to detect a type of vascular lesion called microaneurysm, which appear in the retinal vasculature at the advent of diabetic retinopathy. As the disease progresses it manifests visually as exudative lesions, which are to be segmented, and a system has been developed for the same.

The developed systems are tested with image datasets that are in the public domain, as well as a real dataset (from a local speciality hospital) collected during the course of the research, to compare performance indicators against the prior art and elicit better understanding of the factors and challenges involved in creating a system that is ready for clinical use.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In medical diagnosis, images are a source of in-vivo, painless observations that are visually analyzed. Complex imaging modalities are deployed in medical diagnosis and planning owing to the clinical value of visual information. Practioners and clinicians are trained to make decisions based on perceptible cues visually obtainable from medical images.

Images form the key information source in many medical decisions, for purposes such as diagnostics, surgery planning, therapy and follow up. In some applications, visual information is the only source of observations - brain FMRI for instance. It may even be the case that the information sought is perceivable only in visual information - lung nodules in MR images, for example, present strong likelihood of developing tuberculosis.

The information captured in images is best interpreted by human experts. The primary task towards attempting automated interpretation of visual information is object detection. This chapter introduces the problem of object detection in medical images, formulates a general image detector, and discusses about retinal images and diabetic retinopathy. The thesis develops analysis algorithms for color retinal images to perform automated detection of indicative lesions of diabetic retinopathy.

1.1 Object detection in images

Detection of objects in images is one example of an image understanding task. The object has a known visual manifestation, which is searched in the image. Manual search becomes a tedious task as the field of view increases. When accuracy, speed and unbiased detection are the requirements, the task calls for automation.

The object detection task can be considered as the first high-level abstraction of visual information. Higher abstraction tasks built upon object detection are object categorization and object identification [Ponce 06]. The detection task consists of localizing instances of the target object as projected on images. The challenge lies in constructing a detector that performs to stringent requirements of accuracy necessitated by the application.

Object detection is helpful in clinical decision-making. The object to detect could be disease-indicative lesions, hemorrhages, tumors, anatomic structures, or interesting patterns. In the case of diabetic retinopathy, initial stages of the disease are characterized on the retinal photograph by 'dot' lesions . The extent of affliction is indicated by the count of the lesions, and their locality. Object detection can provide quantitative information in each of these situations. The clinical decision-making process can be augmented by automatically analyzing visual information and transforming it into a presentable and measurable form. The information yielded could be useful in deciding the treatment, planning surgery, and for tracking progress.

1.1.1 The detection task

Detection in images is a task of finding the locality of a target object in a given image. In order to find instances of the target, a detector may invoke knowledge of the prototypical appearance of the target. The prototypes of the target are expressed as **characteristic patterns**, which are points in a measurement space or "feature-space". A metric is defined in the feature space to quantify the proximity of candidate samples to the known prototypes. The value of the metric helps to decide whether an observation is that of a target or not.

The **locality** and **number of occurrence** of the characteristic patterns may be directly utilized in deriving descriptive information about *the state of Nature*. For instance, in retinal image analysis, the spatial proximity of exudative lipids to the macula is an indicator of the criticality of non-proliferative diabetic retinopathy[Das 06]. Accurate localization of characteristic patterns may also be beneficial in improving precision of treatment and attentive care. Exhaustive localization of every instance of the target is laborious when performed manually, hence the task of object detection in medical images is of significance, and amenable to computer automation.

This chapter gives a general formulation of object detection in images, performance criteria necessitated by medical image analysis, and introduces diabetic retinopathy, the illustrative case taken up for this thesis. Also presented here is a dichotomy of the popular approaches for detection in the art.

1.1.2 Formulation

Define an *image detector* as a detector that localizes a specific target object in the given image I. Let I be decomposable into sub-images I_i in such a manner that in each I_i, the two possible states of Nature are 'target present' (ω_1) and 'target absent' (ω_0).

The general definition of detection is estimation of the current state of Nature, from among a finite set of possible states. Each prevailing state of Nature establishes a behaviour which may be observable. Considering the example of weather, defining the states of Nature as one of $\{rainy, sunny, cloudy\}$, a meteorological *observation* is a sample of *measurements* governed by the prevailing state of Nature, and detection involves estimating the state of Nature given a meteorological observation.

In terms of the observations, each state of Nature corresponds to a causal factor or distribution of observation probabilities, and detection involves estimating the source distribution for an observation.

If Ω is the state of Nature to be detected, define a hypothesis $H_0 : \Omega = \omega_0$, the null hypothesis, or the hypothesis which declares target to be absent, and $H_1 : \Omega = \omega_1$ as the alternative hypothesis, which declares target to be present in I_i.

The detector is then regarded as the tuple $D = (S, P, \Gamma)$ where S is a set of characteristic patterns known to be exhibited by the target object, P is a function which partitions input I into sub-images I_i, and $\Gamma = \{\gamma_i\}$ is a set of decision tests for each I_i to decide between the two states of Nature. Each test γ_i is of the form

$$\gamma_i(I_i) \gtrless_{H_0}^{H_1} 0 \tag{1.1}$$

accepting one among the two hypotheses, at I_i.

The members of set S are governed by the representation chosen to depict the target object. One method of providing S is inductive learning, by giving a set of training samples Y, which are sub-images with the state of Nature labeled as "target present" by a domain expert. In the absence of such training, data analysis techniques may be used to obtain the characteristic patterns. This is further elaborated in Section 1.3.

The partition function P may act such that the partitions I_i are of varying size and overlapping. This makes it possible to have combinatorial ways of partitioning I, among which those partition functions which do not fragment the target are of interest.

Thus the detector D consists of (S, P, Γ), and given an image I, outputs

$$\Psi = \{\psi | \gamma_\psi(I_\psi) > 0\} \subset \{i\}, \tag{1.2}$$

the indices of the sub-images in which the target object is posited to be present (called the *positives*). Design of a detector involves modeling the training subimages Y, obtaining the optimal partition function P and establishing the decision tests Γ. These elements are designed such that the detector meets some optimality criteria, elaborated next.

1.2 Performance characteristics

The performance of detectors is measured by two kinds of errors possible in the detection task: *false alarms* and *misses*. The nature of the deployment governs the performance requirement for detectors. For instance, if deployed in a screening scenario where a decision about the normalcy of the subject is to be made automatically, the system is expected to filter out normal cases (which are expected to constitute a majority) and earmark those cases with high probability of being abnormal, for manual analysis by experts. In this application, the detector is calibrated such that false alarm rate does not exceed a certain value (typically 1-4 false alarms per image [Abramoff 08]). If the detector is used in

treatment planning or tracking, and provides visual output to the clinician, the hit-rate (complement of miss-rate) is expected to be very high (typically above 80%).

The performance of a given detector D is ascertained by observing its output Ψ for a set of known images - images where the "truth" is known about the locality of the target objects (denote Ψ^*).

An accurate detector is one whose output closely matches Ψ^*. Comparison of Ψ and Ψ^* involves two sets defined here. A match TP is found as the set of one-to-one correspondences between Ψ and Ψ^*. TP is the set of *true positives*. The set $FP = \Psi - TP$ are the *false positives*.

The following relation can be stated about the sets Ψ^*, TP and FP:

$$0 \leq |TP| \leq |\Psi^*|, \text{ or } 0 \leq \frac{|TP|}{|\Psi^*|} \leq 1 \tag{1.3}$$

For a single image [1], the *sensitivity* of the detector is defined as $s = \frac{|TP|}{|\Psi^*|}$. Sensitivity is the power of the detector to accept H_1 when target is actually present ($\Omega = \omega_1$), expressed as a percent value. Other names for sensitivity are *true positive fraction, detection rate*, hit rate [2], and *recall*.

High sensitivity is achieved even when H_1 is accepted indiscriminately, irrespective of Ω (i.e., $\Psi \approx \{i\}$). It is desirable that $\Psi \to TP$, or $FP \to \emptyset$.

Over a dataset, s and average $|FP|$ are the two metrics to quantify a detector. Detector design is hence an unconstrained minimization of $(1 - s)$ and $|FP|$. Since detector design involves identification of decision tests Γ and the suitable partition function P, the process in essence is a minimization of functionals.

In practical applications however, requirements are set on the tolerable number of false alarms (τ), in which case detector design is a constrained optimization of s subject to $|FP| \leq \tau$ (the classical Neyman-Pearson task).

1.2.1 Detector design outline

The design of the image detector requires specification of the components below:

- S **the set of characteristic patterns**. S consists of appearance rules and sample feature vectors characterizing the target. The rules are captured in implicit form (embedded in the detector logic) or explicit form (rule-based knowledge system). In the typical situation, samples are created of subimages with target (independent of Ψ^* the evaluation set), called the training set.
- P **the partition function**. P is specified considering the variations in scale of the input image, and the size of the target.
- Γ **the set of tests**. Γ is designed for each variant of the characteristic patterns.
- Ψ^* **the evaluation set**.

[1] For a dataset of n images, sensitivity over the dataset $s = \{\sum_{i=1}^{n} |TP|_i\}/\{\sum_{i=1}^{n} |\Psi^*|_i\}$

[2] The quantity $1 - s$ is called the miss rate

1.2.2 Assumptions

The formulation above transforms the problem of detection into one of decision, and poses detector design as an optimization of some objectives. But the solution depends on the correctness of S. The detector has a fundamental dependence on S, and so good representative patterns are assumed to be made available to the detector.

The performance of the detector also relies on adequate coverage of the search space by the partition function P. Scale normalization of the input image should also be accounted for in the logic of P.

The evaluation set Ψ^* should be dependable and the set size statistically significant, since the detector is evaluated based on it.

1.3 Solution strategies

The formulation in Section 1.1.2 highlighted the key components of the image detector. This section relates some prominent solutions in the literature to computational methods of realizing the described components.

The state of art in general image detection can be categorized into two approaches: *learning-based*, and *unsupervised data analysis-based*. While the former approach transforms the optimization above into optimization of equivalent criteria, the latter exercises greater emphasis on domain and application knowledge in order to perform the task.

1.3.1 Learning-based approach

Target detection can be posed as a classification between 'Target' and 'Non-target', using training samples describing the target only. The problem is pertinent to outlier detection, and the taxonomy of [Hodge 04a] names it as single-class classification or outlier detection of *Type-3*. Detection is viewed as single-class classification since, for a given representation of S, 'non-target' is not rigorously defined and may encompass any number of classes based on the representation of S.

Several techniques [C.Papageorgiou 98] [Viola 01] [Dalal 05] solve the single-class problem through binary classification, considering a carefully selected, normalized sample set of the target instances and a "clutter" set populated by random selection of numerous partitions from several images known to not contain any instance of the target. The boot-strapping training method was used by [kay Sung 98], to accumulate negative training samples (the false alarms in training images devoid of the target), iteratively modifying the decision surface until satisfactory discrimination is achieved over a disjoint test set of images.

The optimization criterion in the formulation above is transformed by the type of classifier chosen. For instance, a Fisher discriminant classifier [Duda 00] *maximizes* inter-class distance while minimizing intra-class scatter. The SVM classifier [Burges 98] *maximizes* the margin of separation between the

optimal hyperplane and the labeled samples. A feed-forward neural network classifier performs a least-squares minimization of training residuals. A classifier when applied to the detection problem, finds the set of decision tests Γ which best separate the training samples as target and non-target.

It can be stated that *the correctness of the binary classifier decision surface ensures low false positive rate* ($|FP|/|\Psi|$) *and low miss rate* ($1-s$)

Let $y = \hat{g}(\mathbf{x})$ be the obtained decision hyperplane equation corresponding to the separating boundary of the two classes ω_1 and ω_0, with $\mathbf{x}$ the multivariate random variable corresponding to the feature measurements. $\hat{g}$ is obtained by the process of classifier design based on labeled training samples. Let $g(\mathbf{x})$ be the true decision hyperplane (assuming that it exists).

The decision test provided by this hyperplane $\hat{g}$ is: if $y > 0$ declare $\mathbf{x}$ to be in ω_1, else declare to be in ω_0.

Given $\mathbf{x}^*$ the observations at subimages Ψ^* of the ground truth, and l^* their labels, with elements of $l^* \in \{1, -1\}$, Consider $\mathbf{x_p^*}$, $\mathbf{x_n^*}$, such that $\mathbf{x_p^*} \cup \mathbf{x_n^*} = \mathbf{x}^*$ and

$$g(\mathbf{x^*}) = \begin{cases} 1 & \text{if } \mathbf{x}^* \in \mathbf{x_p^*} \\ -1 & \text{if } \mathbf{x}^* \in \mathbf{x_n^*} \end{cases}$$

Then, a particular evaluation sample $\mathbf{x_i} \in \mathbf{x}^*$ with label l_i is classified wrongly by $\hat{g}$ if

$$g(\mathbf{x_i})\hat{g}(\mathbf{x_i}) < 0 \text{ , or } l_i.\hat{g}(\mathbf{x_i}) < 0 \tag{1.4}$$

If ω_1 corresponds to the state of Nature accepting the alternative hypothesis (target present), then $|FP| = |\{\mathbf{x_i} \in \mathbf{x_n^*} \; ; \; \hat{g}(\mathbf{x_i}) > 0\}|$, and $|TP| = |\{\mathbf{x_i} \in \mathbf{x_p^*} \; ; \; \hat{g}(\mathbf{x_i}) > 0\}|$

A correct classifier (denote $\hat{g}$) tends to the hyperplane g as good training samples are provided. This means that $\hat{g}$ separates x_p^* and x_n^*. A correct classifier thus yields maximum $|TP|$, the number of true-positives, and $|TN|$, the number of true negatives.

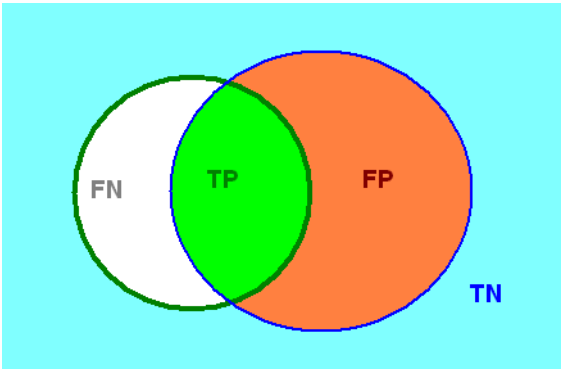


Figure 1.1 Venn Diagram showing the sets TP,TN, FP and FN

Fig. 1.1 shows a Venn diagram illustrating the sets $\Psi*$ (bounded by green circle), Ψ (bounded by blue circle), $TP = \Psi \cap \Psi*$ (green region), and TN (cyan region).

The classifier is capable of manipulating Ψ (blue circle) in order to achieve maximum $|TP|$ and maximum $|TN|$. Since TN is disjoint of $\Psi*$, from the Venn diagram it can be seen that maximizing $|TN|$ results in minimizing $|FP|$.

Hence for a fixed evaluation set of size $|\Psi|$, a correct classifier maximizes $s = |TP|/|\Psi|$ and minimizes $|FP|$.

The essential theme of the learning-based approach may be summarized as statistical inference : select a system that *best* models the target, based upon statistical evidence provided in the form of labeled samples. Unlike this approach, the following paradigm does not directly perform an explicit optimization, and relies more on domain knowledge, heuristics, assumptions and constraints.

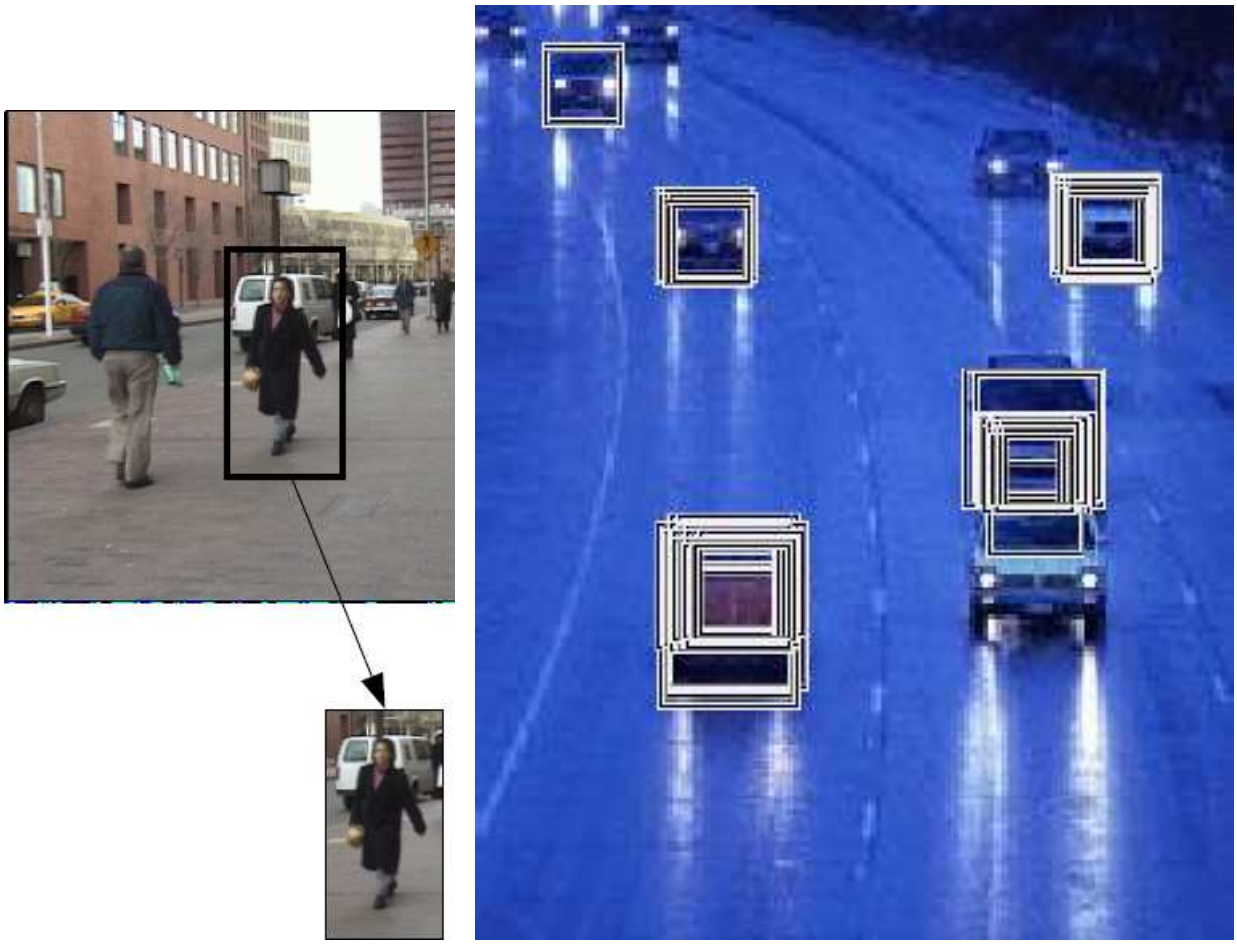
1.3.2 Unsupervised data analysis-based approach

The data analysis-driven approach consists of techniques such as normalized template matching, density estimation methods (including maximum aposteriori techniques such as random-field modeling), thresholding and clustering. Feature detectors such as edge detectors [D.Marr 80] [Canny 86], blob detectors, corner detectors [Harris 88], boundary and primal sketches [Asada 86] [Brady 85] [Haralick 83] were the earliest to use this approach.

Initial work on object detection used template matching [M.Betke 95] [A.Yuille 92], applying normalized correlation techniques and deformable templates to perform tasks such as detection of faces, pedestrians, cars and road signs. In the case of retinal images, unsupervised techniques have been used to detect various anatomical structures such as the vasculature [Garg 07] [Frangi 98], macula, optic disk [Singh 08] , vessel junctions [Ram 09], bright and dark lesions [Sinthanayothin 02] [Huang 05] [Bhalerao 08].

This approach relies on the factors, assumptions and the model considered by the algorithm developer. The problem is generally harder than the learning-based approach. The template or the model provides a representation of the object, and is hence expected to be versatile as well as discriminative. Techniques under this approach include significant amount of prior information and domain knowledge, constraints, heuristics and assumptions. The learning based approaches yielded better results, partly because they were developed later, but mainly because they rely less on assumptions about the input and enjoy greater flexibility in representation.

The data analysis approach is suitable when samples are not straightforward to get or operate on (especially with reference to the partition function P). It is also useful where the object is simple, variations are less, and learning is counter-productive or superfluous.



(a) Pedestrian detection (b) Car detection

Figure 1.2 Sample outputs of pedestrian detection and car detection. Images courtesy of C.Papageorgiou, MIT 2000

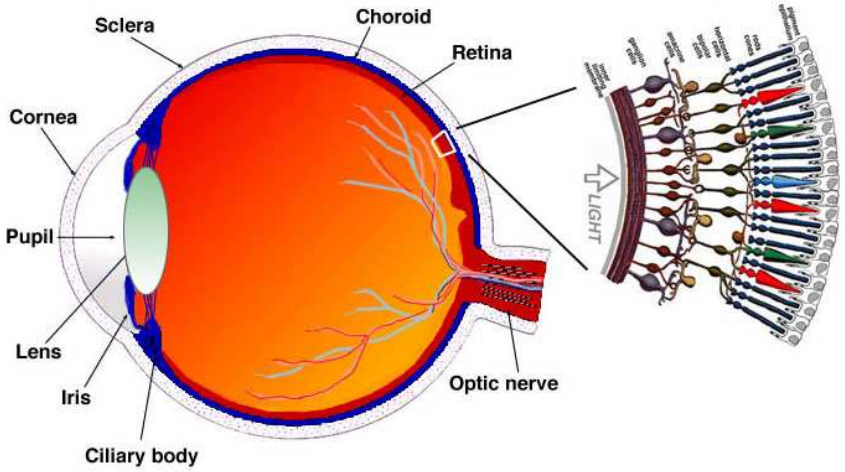


Figure 1.3 A schematic sagittal section of the human eye, with schematic enlargement of the retina. Image courtesy of *Webvision: The organization of the retina and visual system*, Helga Kolb, Eduardo Fernandez and Ralph Nelson, John Morgan Eye center, University of Utah. http://webvision.med.utah.edu

1.4 Focus of the thesis

This thesis aims to demonstrate each of the above two approaches by applying them for the detection of specific lesions indicative of diabetic retinopathy in color images of the retina.

A learning-driven approach is designed for the detection of microaneurysms. A multispace clustering based approach is discussed for the segmentation of retinal exudates. The designed systems and algorithms are a step towards achieving automated screening as a support tool for clinicians and medical practitioners. Further, insights from the state of art and the systems developed are presented in the hope of streamlining and accelerating further work in automated object detection in images.

1.4.1 Retinal Images

The human eye is structurally organized similar to a camera. Light that passes through the iris is focused onto the retina through a lens. Retina is the sensory membrane that lines most of the large posterior chamber of the vertebrate eye. The visual information is encoded in the retina, and transmitted to the brain through the optic nerve.

The human eye has a circular opening called the pupil through which light enters the eye and reaches the retina (see Fig. 1.3). Retinal imaging systems use this opening to capture the image of the retina. The diameter of the pupil adjusts itself so as to let an optimum amount of light enter the eye. However, the pupil can be dilated using drugs, in order to obtain a large diameter irrespective of the amount of light entering the eye. Often, in order to facilitate better illumination of the retina, the patients eyes are dilated before capturing the images.

As can be seen in Fig. 1.3, retina has the shape of an inner surface of a hemisphere. Because of this, it is not possible to capture the entire retina in a single image. Different parts are imaged by adjusting the camera into different positions. Typically, depending on the field of view of the camera, a number of images are obtained so that the part of the retina that is of interest is captured in at least one image.

1.4.2 Diabetic retinopathy

Diabetic retinopathy is an ocular manifestation of diabetes, and diabetics are at a risk of loss of eyesight due to diabetic retinopathy. Upto 80% of patients with diabetes tend to develop DR over a 15 year period. Worldwide, DR is a leading cause of blindness among working populations. DR is also the most frequent microvascular complication of diabetes. The eye is one of the first places where microvascular damage becomes apparent. Though diabetes is still incurable, treatments exist for DR, using laser surgery and glucose control routines. But early detection is key to ensure successful treatment.

For this disease, and consequently for this thesis, the retina is the most important part of the eye. Diabetes being a blood-related phenomenon, causes vascular changes, which can often be detected visually by examining the retina, since the retina is well-irrigated by blood vessels.

The vascular changes in diabetic retinopathy produces lesions, which hinder the working of the photoreceptive neurons lining the retina. Specific spatial regions exist in the retina, like the fovea, containing high concentration of photosensitive cells and is bereft of vasculature. Diabetic retinopathy leads to risk of vision loss if vascular changes occur near such regions.

DR presence can be detected by examining the retina for its characteristic features. One of the first unequivocal signs of the presence of DR is the appearance of microaneurysms.

MA appear due to local weakening of the vessel walls of the capillaries, causing them to swell. In some cases the MA will burst causing hemorrhages. As the disease and damage to the vasculature progresses, larger hemorrhages will appear. In addition to leaking blood, the vessels will also leak lipids and proteins causing small bright dots called exudates to appear. Next, a few small regions of the retina become ischemic (deprived of blood). These ischemic areas are visible on the retina as fluffy whitish blobs called cotton-wool spots.

As a response to the appearance of ischemic areas in the retina, the eye will start growing new vessels to supply the retina with more oxygen. These vessels (called neovascularizations) have a greater risk of rupturing and causing large hemorrhages than normal vessels.

Treatment of DR is still predominantly based on photo-coagulation, where a strong beam of light (laser) is applied to certain areas of the retina. The laser can be applied to leaking MAs to prevent further hemorrhaging. It can also be applied in a grid pattern over a larger part of the retina with the purpose of reducing the overall need for oxygen and diminishing the load on the damaged microvasculature. Photocoagulation can significantly reduce the risk of serious vision loss. However visual acuity already lost usually cannot be restored.

1.4.3 Analysis for detecting DR

Ophthalmologists can visually examine a patient's retina using a small portable instrument called an ophthalmoscope. It consists of a set of lenses and a light source, permitting the ophthalmologist to view regions of the patient's retina.

The pupil is narrow, thus it does not allow much light to enter the eye for illuminating the retina. The pupil may be dilated by administering eye drops (mydriasis).

An indirect way of examination is by using photographs of the retina captured using fundus cameras. This decouples the examination process into the disjoint tasks of image acquisition and interpretation. Further, modern fundus cameras are capable of capturing retinal images without mydriasis.

Digital fundus photography thus opens the possibility of large scale DR screening, where diabetic patients can be routinely checked for DR. The screening solution would automatically isolate abnormal cases by applying suitably calibrated detectors of disease indicators. Since the number of normal cases is expected to be greater than the abnormal, the screening process can reduce the work load of ophthalmologists, by having them examine only those cases which are hard to categorize as normal. This can also reduce the treatment costs and help to ensure treatment effectiveness amidst scale-up in the number of patients.

Further, the manual analysis may be augmented by using computer-based tools. For example, an image analysis system that automatically determines if lesions are present, can reduce the work load of ophthalmologists, by showing them only those cases which are abnormal, and directly archiving the normal cases.

1.5 Contributions

The thesis documents the following contributions:

- Two broad approaches for object detection in images are outlined, illustrated by applying them to the analysis of retinal images.
- The two developed systems include various novelties in terms of technique, dataset, validation methodology and interpretation of results.
- The systems are conceptualized as the core of an automated Diabetic Retinopathy screening solution.

1.5.1 Discussion

Medical images are an information source for making clinical decisions. The examples stated in this chapter pertain to visual information of medical significance. It is to be noted that the sensor capturing the information is not restricted to the visual spectrum, but the analysis by conventional methods is

visual. Humans can understand a scene not only by directly sensing it, but also by viewing a finite projection (image) of it. We can conjecture that visual representation through images is apt, convenient and informative for *manual* analysis.

The state of art in automated image understanding tasks indicates that such human-friendly visual information is challenging to analyze and derive information automatically. For an automatic analysis system, an image is a lattice of pixel (or voxel) values. The task of deriving higher abstractions from this representation is an inverse problem, and is generally ill-posed.

However analysis of medical images is not universally so. Medical imaging technologies such as tomography are capable of obtaining sectional views of objects – views that can not be sensed directly by the human visual apparatus. Human understanding of such images (as also microscopic images or images from non-visual spectra) is equally ill-posed. But nevertheless, an X-ray image of a fractured arm, for instance, conveys diagnostic information to a medical expert *trained* to analyze X-ray images. Human analysis of such images (projections of the scene) is built upon semantic understanding and information available about the causal factors at play in the scene. Such external information is necessary to better formulate the problem for automated analysis.

1.6 Organization

The issues involved in the design of detectors are introduced in Chapter 1, and a framework is described for detection in images. In Chapter 2, existing approaches are discussed and a detector is proposed for a class of lesions called *microaneurysms* in retinal images. Chapter 3 gives the detailed design of a successive rejection-based system for detection of microaneurysms. An extensive analysis of the system performance on two public datasets, and one dataset collected during this study, is presented in Chapter 4. To illustrate the alternative (unsupervised data analysis) paradigm, the problem of exudate segmentation is taken up in Chapter 5 and a clustering-based solution is proposed and evaluated. The thesis concludes with a discussion deriving insights from the state of the art and the systems developed during this study, and gives some open questions and directions to explore.

Chapter 2

Automatic screening for DR

The presence of microaneurysms (MAs) is an early sign of diabetic retinopathy (DR) and their automatic detection from color retinal images is of utility in screening and clinical scenarios. This chapter reviews the problem and previous efforts towards it. A new approach for automatic MA detection from digital colour retinal images is formulated, based on insights from the state of art.

2.1 Introduction

Diabetic Retinopathy (DR) is a major public health issue since it can lead to blindness in patients with diabetes. Microaneurysms (MAs) are the first clinical symptom of DR. They are swellings of capillaries caused by a weakening of the vessel wall [Fleming 06]. Their sizes range from $10\mu m$ to $125\mu m$ [Huang 05]. In the clinical scenario, experts rely either on direct manual examination or fluorescein fundus angiography (FA) where MAs appear with high contrast as bright white spots. Given the high cost and the cumbersome requirement of intravenous injection of a dye for this type of imaging, interest in the recent past has been on detecting MAs from colour fundus/retinal (CFI) images. In CFIs, MAs appear as tiny, reddish isolated dots. Automatic detection of MAs from digital CFIs can play an important role in DR screening at a large scale [Abramoff 08][Niemeijer 05]. It can significantly reduce the workload of the ophthalmologists and the health costs in the DR screening [Abramoff 08].

From computational point of view, MA detection from CFI requires extraction of tiny objects from a highly varying surround which is subject to many factors: large variability in colour, luminosity and contrast both within and across retinal images due to acquisition process; distinctive colour and background texture due to intrinsic characteristics of the patients, such as retinal pigmentation and iris colour; presence of other pathologies like cataract, etc; variable quality due to use of mydriatic or non-mydriatic fundus cameras of different make. The intensity profiles of two cases in Fig. 2.1 show contrast variations in the depth of the profile. Such variations make MA detection from CFIs very challenging. Notwithstanding these challenges, the performance of a MA detection method is assessed against expert markings on the CFI, in terms of its detection sensitivity and capability to handle the above mentioned variations.

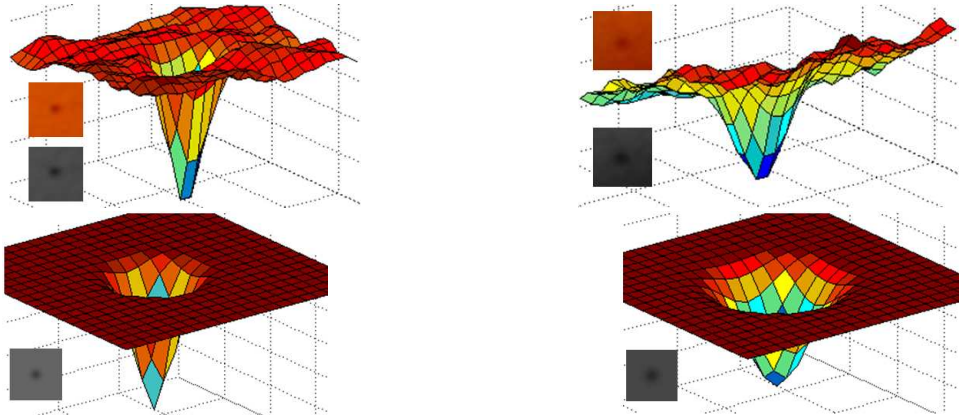


Figure 2.1 First row shows two sample MA profile obtained from CFI image. Second row shows the approximated MA profile using Gaussian model given in equation(1).

The chapter is organized as follows: the next section discusses the state of the art in MA detection, and lists some insights derived from the existing approaches. Section 3 gives the motivation for a new approach and Section 4 conceptualizes it. Section 5 illustrates a system developed upon the proposed approach. Section 6 details the experimental evaluation of the developed system. Section 7 analyzes the results and draw some conclusions.

2.2 State of the Art

Overview: Existing methods for MA detection generally consist of two-stages, where the first stage is aimed at obtaining potential MA candidates while the second stage is used to assign MA or non-MA category to the candidate using features computed around the candidate location. The main processing components include 1) pre-processing; selection of candidate MA and 2) feature extraction; classification.

The focus of the early methods has been on pre-processing and candidates selection steps. Later methods focus more on designing new sets of features and choosing of classifiers. Recently published work have re-examined the individual processing components and presented new improvements on certain aspects.

Due to the diversity in the presented techniques, in addition to their assessment carried out on different datasets, a quantitative comparison of various approaches is difficult.[1]

We now look at the existing approaches in detail. Early published work attempted to address the problem of MA detection in FA images of the retina [Lay 83][Spencer 96][Spencer 91][Frame 98][Cree 97]. Lay et al., [Lay 83] presented the first MA detection method for FA. In this method, MA candidates were obtained using top-hat transformation which eliminates the vasculature structure from the image yet left

[1]Recently, two public datasets have been made available to make quantitative performance assessment possible. A handful of methods have been evaluated on those datasets till date.

possible MA candidates untouched. Spencer et al., [Spencer 91] presented a shade correction technique and a candidate detection method using matched filtering.

However, potential mortality associated with the intravenous use of fluorescein [Yannuzzi 86][Niemeijer 05] prohibits the application of this technique for large-scale screening purposes. Instead, colour fundus image (CFI) has emerged as a preferred imaging modality due to its non-invasive nature [Yannuzzi 86]. A good amount of clinical studies show the effectiveness of CFI for large-scale DR screening [Abramoff 08].

Numerous algorithms have been proposed to detect early signs of DR (MAs) from CFI. The first such method was presented by Oien et al. [Oien 95]. The pre-processing used here is similar to the approach used by [Lay 83]. A rule-based classification step was added to the processing pipeline followed in [Spencer 96][Frame 98][Mendonca 99][Autio 05]. Usher et al.,[Usher 04] employed a neural network based classification after candidate selection based on recursive region growing and adaptive intensity thresholding.

Use of supervision: Niemeijer et al.,[Niemeijer 05] presented a supervised, pixel classification technique to extract red lesions to get MA candidates. A large set of features was added to the original feature set used in [Spencer 96]. A knn classifier was used for MA recognition. The recognition performance of individual MAs has been evaluated on 50 images collected from different screening programs and clinical hospitals.

Local information: Huang et al.,[Huang 05] presented a local adaptive approach to extract candidates, where multiple subregions of each image were automatically analyzed to adapt to local intensity variation and properties. This method was evaluated on 30 images taken from STARE retinal dataset [Hoover]. Fleming et al.,[Fleming 06] presented a local image contrast normalization technique to get more discriminative features for MA. A vessel-free region is obtained around each detected candidate using watershed segmentation. Vessel-free region is then used to enhance the contrast of candidate. A parametric model of a paraboloid is used for the MA and fitted on a set of pixels obtained by applying region growing on the candidate location. The model parameters are used to derive a new set of features for the candidate and finally classified using a knn classifier. The recognition performance of individual MAs is evaluated on a total of 71 images collected from a screening program.

Morphological processing: Walter et al.,[Walter 07] used a morphological (diameter) closing technique for detecting candidates. A supervised density-based classifier, trained on 21 images, is used for MA classification. The method has been evaluated on a database of 94 images. Huang et al.,[Huang 07] used edge-based information to delineate MA candidate regions and evaluate on 49 images collected in a clinical examination setup.

Template matching: Quellec et al. [Quellec 08] presented a method based on template matching with a generalized Gaussian template. The matching is performed in the wavelet domain to obtain MA candidates. The classification stage optimizes the selection of wavelet sub-bands in which maximum discriminative information exists for MAs versus non-MA regions. This scheme has been evaluated on 35 CFIs acquired for screening purposes.

Comparison of methods based on performance: The MA detection methods described above report sensitivity figures ranging from 30 to 89%. It is difficult to assess the merit of these methods based on these figures since each method uses a custom-built dataset of various sizes and reporting of results is not standard. A few datasets such as DRIVE [Staal], STARE [Hoover], MESSIDOR[Klein] are available in the public domain, yet these are not adequate for the evaluation of MA detection methods as they do not contain locational information about the MAs present in the images. Recently, towards bringing standardization in evaluating MA detection methods, two evaluation datasets have been made public: a) DIARETDB1 [Kauppi 07a] with 89 CFIs and b) Retinopathy online challenge (ROC) [Abramoff 07][Niemeijer 09] with 100 CFIs, respectively. These sets provide multi-observer (expert) information on locations of MA.

Prior to these two datasets, evaluation on a common dataset was not possible in the early MA detection methods due to a lack of standard evaluation dataset. Hence, it is difficult to conduct a quantitative performance comparison of individual processing steps presented by various methods [Winder 09]. Now, with the availability of 2 public datasets, it is desirable to assess existing methods or any newly developed method on these datasets. This will help in identifying the optimal series of processing steps and their best specifications for MA detection.

These public datasets have been available very recently and therefore only a limited number of methods have been tested on those dataset. Bhalerao et al., [Bhalerao 08] proposed an unsupervised technique evaluated on DIARETDB1 [Kauppi 07a]. It involves contrast normalization, blob detection by filtering with Laplacian-of-Gaussian filter, and complex filtering on an orientation map derived using gradient components. A sensitivity of 82.6% at 80.2% specificity is reported. Good automated screening solutions require high sensitivity at lower fppi (number of false positives per image). The attainable average fppi of this method is not inferable from the reported information.

Kande et al., [Kande 09] presented a relative entropy based thresholding to extract candidates and used SVM to perform classification. The evaluation was on a dataset of 80 images drawn selectively from STARE [Hoover], DIARETDB0 [Kauppi 07a] and DIARETDB1 [Kauppi 07a] datasets. Of these 80 images, 30 are used for training and remaining 50 are used for testing (no guidelines are given in [Kande 09] for image selection).

Retinopathy Online Challenge (ROC) presents a reference database for automated MA detection in CFIs for diabetic retinopathy screening [Abramoff 07][Niemeijer 09]. Five distinct MA detection methods (see Table 4.5) have been evaluated on this dataset and a comprehensive comparative analysis is available in [Niemeijer 09]. We examine this in greater detail in our experimental section.

In general, a good performance on a common dataset does not translate directly to a comparable performance on much larger unselected datasets [Niemeijer 09][Abramoff 08]. This is due to the following factors associated with a dataset:

a) population under consideration like {Asian, western},

b) source of the images - drawn from screening or clinical scenario,

c) ratio of normal images to images having DR pathologies,

d) camera used to acquire images,

e) retinal imaging protocol- including field of view, resolution and size of images and

f) total number of images in a dataset.

The two public datasets differ from each other on the above mentioned factors (a dataset-wise summary on these aspects is presented in Section 6A). Consequently, the reported performance on either of these datasets may not translate to a similar performance on an unseen dataset. In addition, these datasets contain not more than 100 images (DIARETDB1: 89 and ROC: 100) which implies that a method's performance on these datasets may be insufficient to estimate its performance on larger datasets.

It is understandable therefore that recent studies have concluded that

- the performance achieved by automated detection methods developed for early DR detection are not yet acceptable for inclusion in clinical practice [Abramoff 08] and
- there is a move towards evaluation of various methods rather than development of new methodologies to address the MA detection problem [Winder 09].

The strategy behind the existing methodologies is mainly aimed at getting a good characterization of the MA structure. Complex modeling of MA structure for candidate detection [Bhalerao 08][Quellec 08], local enhancement for illumination invariant MA features [Fleming 06], use of local context/statistics and color information [Fleming 06][Niemeijer 05][Walter 07] are all attempts to get a rich set of MA features. Different characterizations for MA can be evaluated on the following two aspects:

1. *robust modeling of MA*: ability to handle variations in MA profiles
2. *uniqueness of the characterization*: ability to discriminate from non-MA structures.

Overall, the existing methods are more successful in the first aspect with progressively different improvements in modeling. However, they are not successful at discrimination between MA and dark non-MA structures. This is addressed by most of the approaches using an explicit segmentation of dark structures to bring uniqueness in MA characterization. For example, suppression of candidates on vessels and optic disk is achieved using vessel and optic disk segmentation, respectively [Abramoff 08][Fleming 06][Walter 07]. These help eliminate false positives to a good extent but at the cost of rejecting true MAs in the proximity of dark non-MA structures. Various post processing steps are in turn devised (for example, [Fleming 06]) to address this problem.

In summary, discrimination between MA and non-MA structures remains an area that needs improvement and hence warrants fresh examination. In our work, we propose a new detection strategy which is motivated by the above conclusions.

2.3 Approach Formulation

MAs appear as tiny, reddish isolated dots, subject to small intensity- or structure-based transformations. As mentioned above, detection of MAs is compounded by the presence of similar looking structures or image noise, leading to high number of false positives. If we consider true MAs and non-MAs (similar structures) as two classes, in a given image, the probability that a candidate belongs to the true MA (P_T) class is substantially smaller, compared to that of belonging to non-MA class (P_C). Here, we can formulate the MA detection problem as a problem of detecting a target embedded in a background clutter, where the target occurs with a much lower probability compared to the clutter ($P_T \ll P_C$). From this formulation point of view, the earlier methods can be viewed as attempts towards getting better characterization of target class using various features and candidate detection techniques.

We are interested in exploring whether knowledge of the clutter class can play a positive role in MA detection. Thus, instead of the earlier formulations where MA is the *only* object of interest, we consider attempting to gain better understanding of objects in the clutter class, in addition to the target class. We believe that such understanding and characterization of commonly occurring clutter can lead to an alternative way to approach MA detection.

In order to illustrate the limitation in modeling the target exclusively, let us consider a Gaussian template matching solution to extract MA candidates. The Gaussian model G which is a good approximation of a true MA (target) profile, is defined as

$$G(x,y) = A\ \exp\Big(\frac{(x-x_0)^2}{2\sigma_x^2} + \frac{(y-y_0)^2}{2\sigma_y^2}\Big) \tag{2.1}$$

where amplitude 'A' models depth, (x_0, y_0) center location and standard deviation (σ_x, σ_y) captures variability of MA in x and y directions. This model is capable of characterizing fuzzy/good definition, low/high intensity and small/large size MAs typically found in a CFI. Figure 2.1 shows samples of MAs taken from a CFI image and corresponding profiles generated using Eqn. 2.1. A sample image shown in Fig. 2.2(a) contains three MAs highlighted using green boxes. Applying the template G on the sample image, and thresholding (done empirically) yields a binary output image, indicating the locations of the candidate MAs.

It is observed that for accomplishing good sensitivity, the model for the target also selects considerable amount of clutter into the set of candidate MAs. It can be seen from the result of thresholding in Fig. 2.2(b) that MAs are extracted, but at the cost of high number of false alarms. Among the clutter responses it could be possible to identify using knowledge of anatomy certain candidate locations at which MA can not occur. In the sample considered, many false alarms occur at vessel structures and general image background. Some unknown structures could also contribute to clutter. In Fig. 2.2(b), the sample false alarms highlighted in cyan arise due to noise.

At these two situations, we propose to model the clutter, attempting to address the discrimination aspect early, and postpone the target modeling. Such a strategy that aims at very early clutter labeling, can be beneficial to the overall detection as this can facilitate progressive rejection of clutter responses (us-

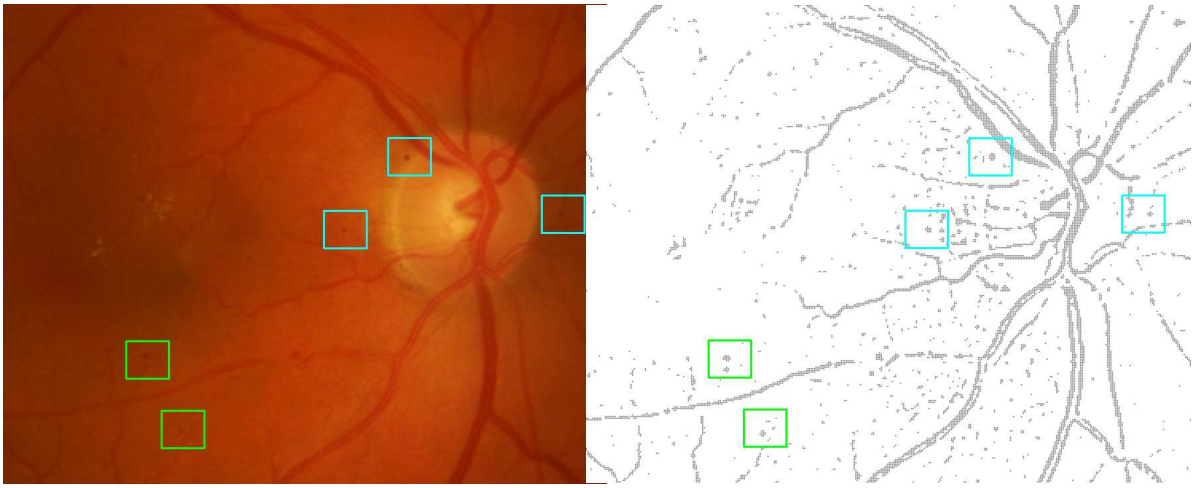


Figure 2.2 (a) A sample region of a CFI. Green box highlights the true MA locations and magenta box shows the similar looking image noise. (b) Template matching results using Gaussian model, given in Eqn. 2.1.

ing many rejectors sequentially), and target recognition may be performed when fewer clutter responses remain.

Each rejection stage can be implemented in supervised or unsupervised fashion, and responses classified as clutter can be removed from further consideration, retaining the remaining responses as putative targets. These are to be passed on to the subsequent rejector for further examination. The objective of such a cascade of rejectors is to reduce P_C while maintaining P_T. This approach is akin to the pattern rejection-based object recognition approach proposed by Baker et al.,[Baker 96]. The following chapter describes an approach to MA detection based on this idea, and provides specifications of an illustrative implementation of the approach.

Chapter 3

A successive rejection based approach for early detection of Microaneurysms in CFI

We propose a method for MA detection where the strategy is to *select* a set of candidate MAs using a simple threshold in a pre-processed image, and then culling the non-MA clutter among the candidates using a set of rejectors in cascade. Since the clutter class has multiple objects of different characteristics, the known and frequently occurring clutter objects are rejected first, and a second stage is designed to discriminate the remaining class of (largely unknown) clutter objects. In the final stage, the MA positives are assigned confidence values based on their similarity to true MAs.

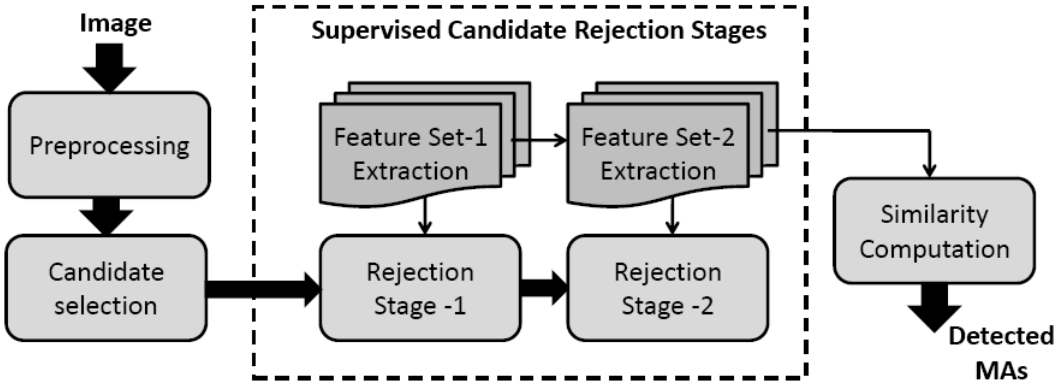


Figure 3.1 Outline of the proposed approach

Fig. 3.1 illustrates the processing pipeline of a MA detection method developed from the proposed idea. The candidate selection method may be a traditional algorithm such as template matching, matched filtering or morphological processing. Our focus is on handling rather than acquiring candidates. Subsequent stages aim at rejecting non-MA clutter from the candidate set.

The first stage rejection aims at eliminating candidates originating from dark structures like hemorrhages and vessels. Once such candidates are suppressed, the sources of remaining non-MA candidates could be due to local minima formed by image noise, region between two bright regions, optic disk, etc. Handling such candidates is the purview of the second rejector stage.

Culling of non-MA clutter by two stages in cascade is expected to result in a significant reduction in the number of reported candidates. In the final stage, the degree of similarity (confidence value) of each remaining candidate, to a true MA profile (which ranges from $[0-1]$) is to be computed. A final set of MA points can be obtained by applying a threshold on the confidence value. The confidence threshold is meant to be adjusted based on the desired performance in deployment. For instance, for a high-selectivity solution, threshold value should be set very high, so that only obvious (high confidence) MAs are reported by the system. In the forthcoming sub-sections, each of the processing stages is elaborated in detail.

Implementation: This section provides an illustrative implementation of the approach constructed above. Elaborated below are details of the components of a system as envisaged in Fig. 3.1.

3.1 Pre-processing (PP)

CFIs present variability in colour, luminosity and contrast both within and between retinal images due to the acquisition process. Pre-processing is an essential first step to normalize variations in order to aid in further processing.

Popularly deployed color fundus cameras produce a color image of the retina in 24-bit RGB. A heterogeneous set of CFIs from various commercially available fundus cameras is shown in Fig. 3.3.

The CFI can be considered as a tri-band image consisting of three channels, each capturing intensities in the red, green and blue bands of the visible spectrum. As seen in Fig. 3.3, the image is predominantly yellowish (additive composition of red and green). There is no blue content in the image due to the scene.

Compared to the red channel, local structural information with respect to background is better contrasted in the green channel. This is illustrated in Fig. 3.2.

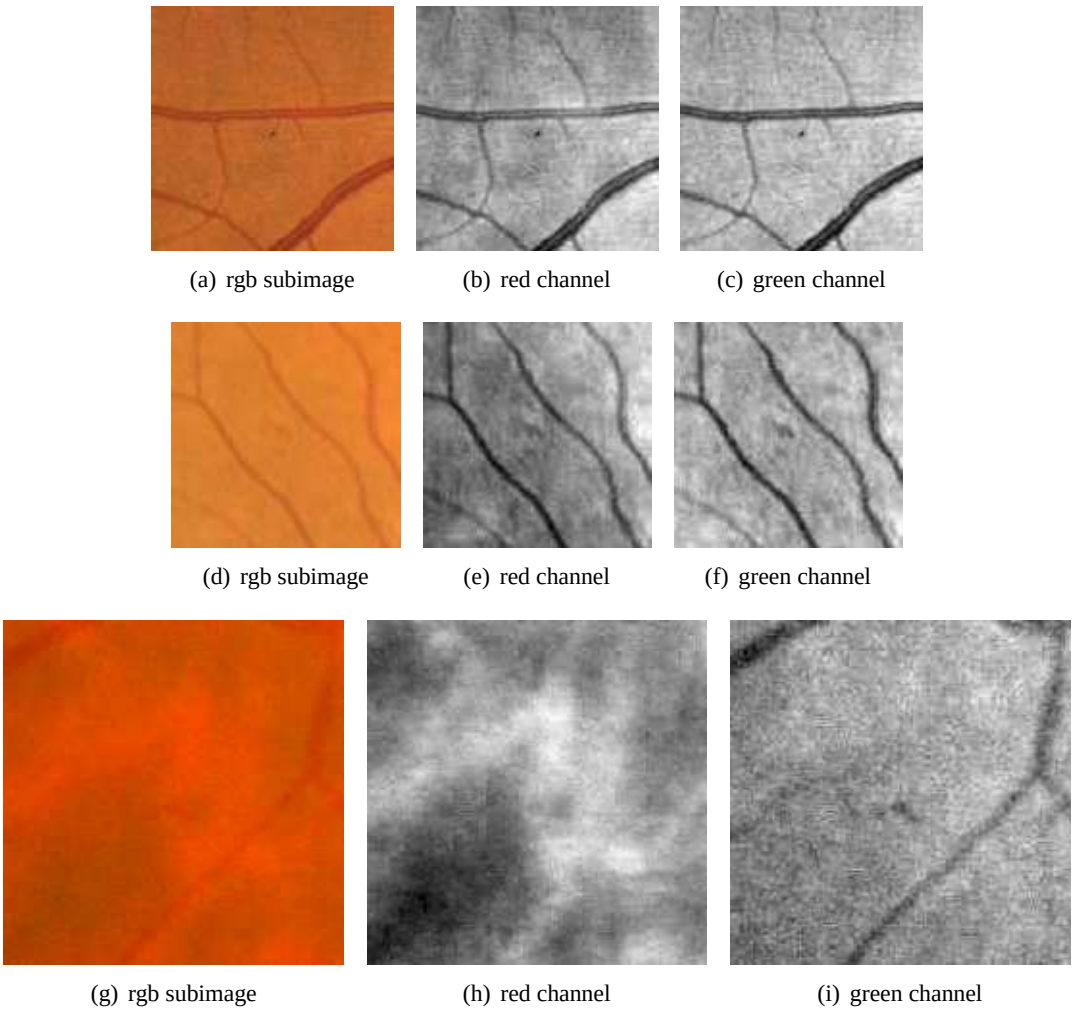
We therefore consider the green colour plane of CFI to carry out all our processing, as do most existing work (for eg.,[Niemeijer 05][Fleming 06][Quellec 08]).

The green channel I_g of retinal image I_{in} is modeled as a subtractively degraded image of a uniformly varying background illumination, as

$$M_1 : I_g = I_{bg} - I_{fg}. \tag{3.1}$$

By this model we intend to designate dark structures such as blood vessels and microaneurysms to the foreground (I_{fg}). The background (I_{bg}) is assumed to be a slowly-varying surface in a large domain. It is thus approximated by a using median filter of size about 25 to 30 pixels on I_g.

The fundus camera illuminates the retina and captures the image with the same aperture. Ambient illumination can not be imaged in this setup. Thus we treat illumination as a property of the *image*, not the scene. This permits us to perform background approximation using I_g itself, without considering the details of the structures in the scene, pose and magnification.



(a) rgb subimage (b) red channel (c) green channel

(d) rgb subimage (e) red channel (f) green channel

(g) rgb subimage (h) red channel (i) green channel

Figure 3.2 Selecting the channel to operate

The foreground estimate, $\hat{I}_{fg}$ is obtained by subtracting I_g from $\hat{I}_{bg}$:

$$\hat{I}_{fg} = \hat{I}_{bg} - I_g. \tag{3.2}$$

At bright regions, $\hat{I}_{bg} \leq I_g$, whereas at foreground regions, $\hat{I}_{bg} > I_{in}$. This subtraction thereby gives negative value to the bright pixels, and negligible positive value to the retinal background. The overall mean value of the difference is a small value. The pixels having value below the mean are quantized to 0.

$\hat{I}_{fg}$ contains high value at dark structures – vessels, microaneurysms, hemorrhages, which are anatomically identifiable, and some striated regions in the general retinal background, imaged due to retinal pigmentation, laser marks or streaks.

To exclusively enhance the MA, $\hat{I}_{fg}$ is match-filtered, the filter being an instance of an isotropic Gaussian density function defined on the radial distance from the filter origin. The standard deviation

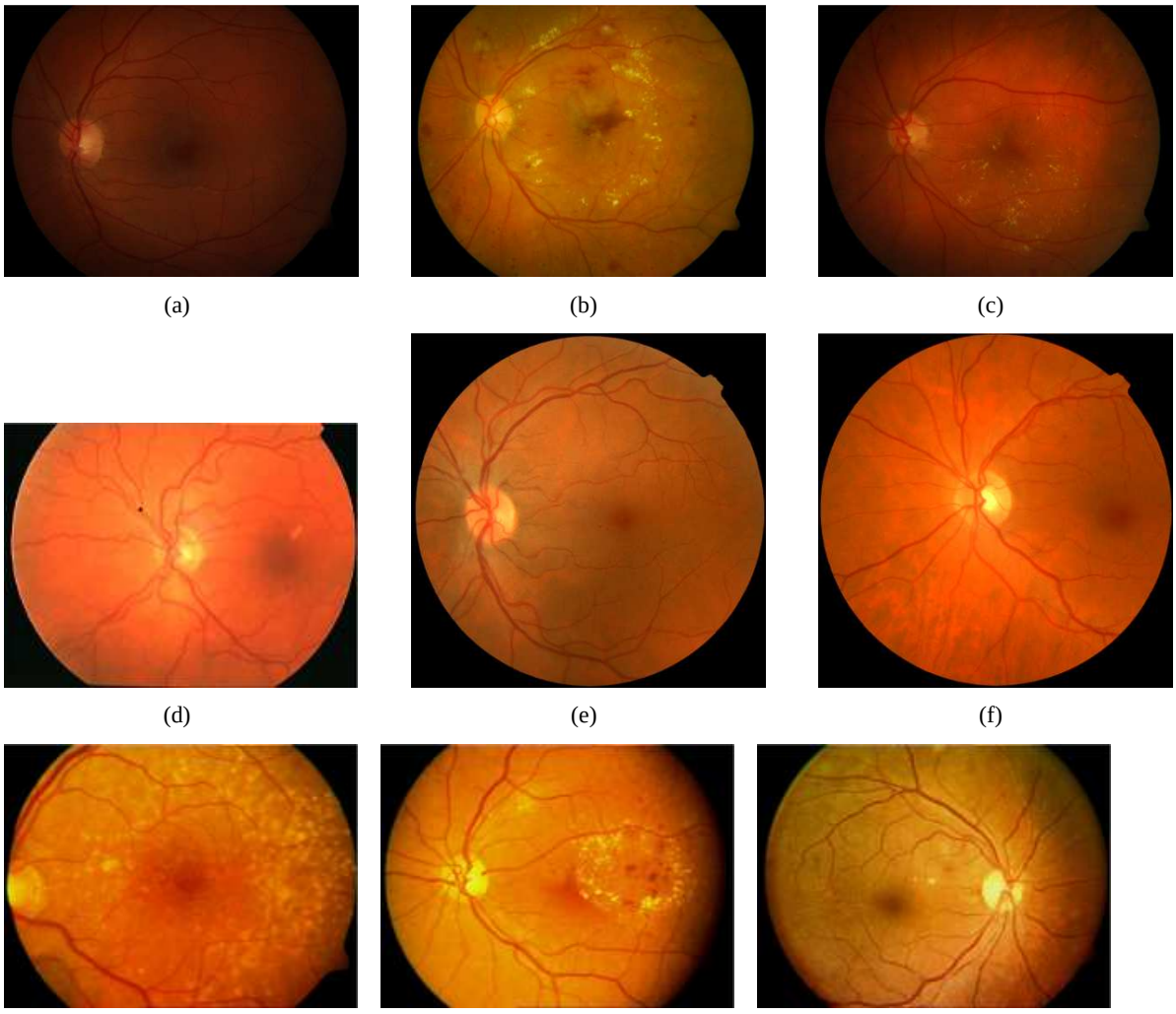


Figure 3.3 Representative CFIs from three different datasets. First row shows images taken from DIARETDB1 (PDS-1) [Kauppi 07a]; Second row shows images taken from ROC dataset (PDS-2) [Abramoff 07]; Third row shows sample images in CRIAS

(σ) of the filter is matched to the size of the lesion. For images of magnification from 50 to 30 degree, the range of MA profiles can be captured using $0.8 \leq \sigma \leq 2.0$.

$$I_{mf} = \hat{I}_{fg} * g(\sigma_{opt}) \tag{3.3}$$

Filtering results in high value at MA and similar-sized objects, whereas the striated regions are blurred due to the smoothing nature of the filter. To augment the relative contrast of microaneurysms further, we apply morphological top-hat filtering to I_{mf}, with a disk structuring element of radius 5 (i.e. object diameter is matched to 10 pixels). The resulting image I_{th} shows high value at the target lesions, in addition to some similar-structured noise, such as the border line of the prominent vessels (whose diameter was greater than the structuring element's), and locations on vessels having small local variation similar in morphological structure to the target.

To eliminate the linear structures in I_{th}, we use morphological opening with linear structuring element in 12 orientations [Spencer 96]. The suprema of the openings I_{su} is used as the marker, and with I_{th} as the mask, we perform morphological reconstruction, to get I_r. The final preprocessed image I_{pp} is obtained by subtracting I_{recon} from I_{bothat}, thereby suppressing linear structures. The potential candidate locations in I_{pp} have a high intensity. Fig. 3.4 shows the intermediate results of the processing occurring in this stage, and I_{pp} for a typical retinal image.

3.2 Candidate Selection (CS)

This stage is simple and very similar to the earlier presented candidate selection schemes [Spencer 96]. The goal of this stage is to apply a threshold on I_{pp} to get candidates.

The task of this module is to select candidate regions (C_0) from I_{pp}. The locations in I_{pp} having high value are potential candidates. I_{pp} is scaled to the range [0, 1], and quantized to 256 values by rounding. An integer threshold t can now be used on I_{pp}, to get a candidate set $C(t)$, as

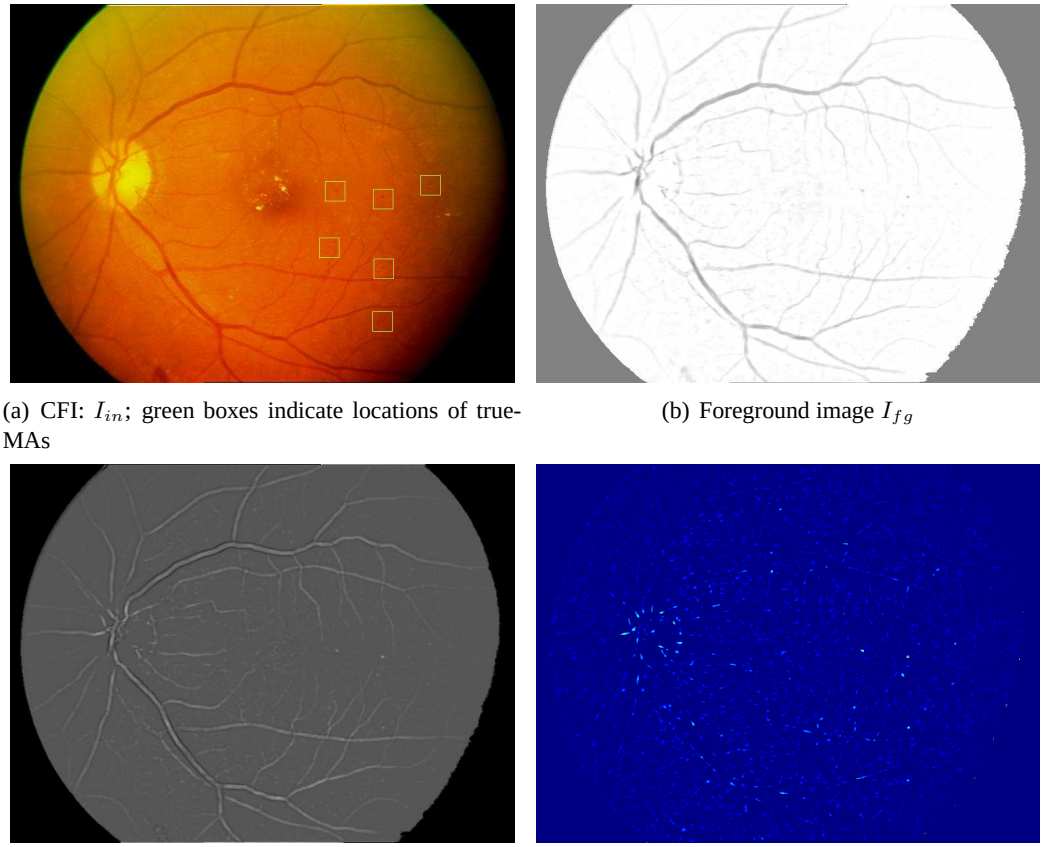
$$C(t) = \{p \mid I_{pp}(p) \geq t\}. \tag{3.4}$$

Candidates obtained in this fashion are actually small, finite, connected regions. We choose to assign the co-ordinates of the minima of each finite region to C.

Selecting a low threshold gives more number of candidates (denoted as $|C|$; see Fig. 3.5). We choose an optimal threshold t_{opt} as the least value of t such that the number of candidates does not exceed an upper bound tol (typical value 200). Then $C_0 = C(t_{opt})$

The selection of tol for a given dataset involves observing the threshold characteristics of t (variation of $|C_1|$ with t) in relation to the distribution of I_{pp} values at true lesions; graphs illustrating these are shown in Figs. 3.5 and 3.6.

The PP stage ensured that retinal background obtains a low value in I_{pp} and higher values at MA and optic-disk junctions. In Fig. 3.6, the peaks indicate that a value of $t > 50$ would cause the rejection of



(a) CFI: I_{in}; green boxes indicate locations of true-MAs

(b) Foreground image I_{fg}

(c) Bottom-hat enhancement for small dark objects

(d) Candidate soft-map I_{pp}

Figure 3.4 Processing occurring in PP stage

many true MAs . Reducing t to values below 50 adds exponentially more and more vessel pixels and background noise into C_1, which is undesirable. The peaks in Fig. 3.6 show that a value of t close to 40 is optimal. This corresponds to a value around 200 for tol.

Given a pre-processed image I_{pp}, the optimal threshold t_{opt} over I_{pp} is found as the minimum value of t at which if threshold is applied (as per Eqn. 3.4), the number of candidates $|C(t_{opt})|$ is less than tol.

Idealized detectors show exponential decrease in false candidates with increase in threshold, and constant decrease in sensitivity ($= \frac{|C_0|}{number\_of\_true\_lesions}$, at this module). The observed threshold characteristics of CS are shown in Figs. 3.7(a) and 3.7(b). The sensitivity monotonically decreases (nearly piecewise-linearly) with increasing threshold. This logarithmic trend variation between the sensitivity and fppi ensures reduced loss of true-MAs while searching for t_{opt}.

We apply a linear mapping to stretch gray-levels of I_{pp} in the range of [0-255]. This mapping normalises inter-image value variations usually found at MA locations. A sample soft-map obtained by this mapping is shown in Fig. 3.4(d). Since, MAs appear as bright structures in I_{pp}, an appropriate threshold is applied to retain bright pixels. This is followed by a connected component analysis to

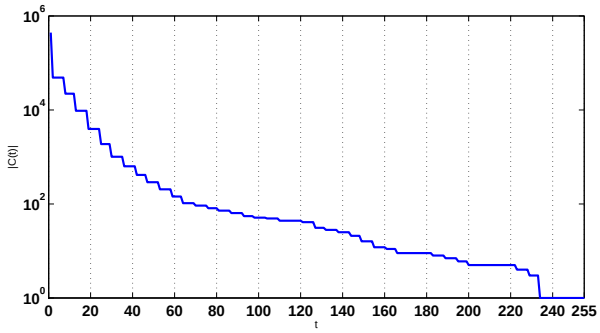


Figure 3.5 Relationship between t and $|C(t)|$, on a typical retinal image. Vertical axis is logarithmic scaled

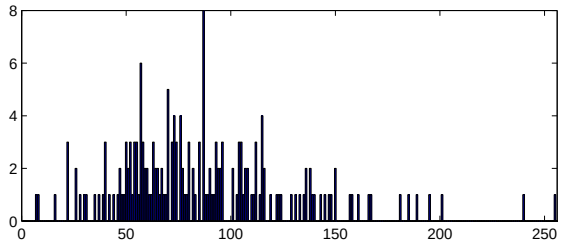


Figure 3.6 Histogram showing the distribution of I_{pp} values at true-MA locations in a dataset of 89 images

delineate candidates as small regions formed by connected pixels. The local minima of each component are used as candidates in further stages.

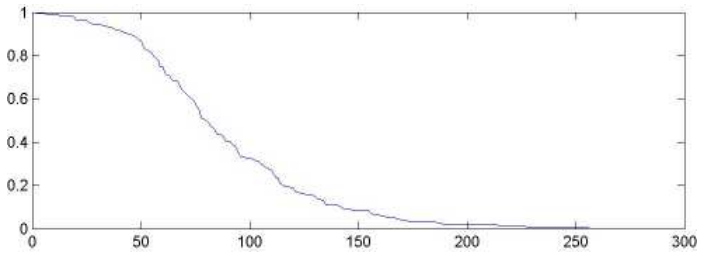
If n is the number of candidates obtained by applying a threshold t on I_{pp}, a low t gives many candidates, but selectivity is low. For each image, t is chosen such that n is below a desired bound tol. This is done to ensure selection of as many true-MAs as possible, while keeping the total number of possible false candidates below a tolerable value (tol). In our experiment section, we present an analysis about the role of this stage on the detection performance.

3.3 Successive Rejection

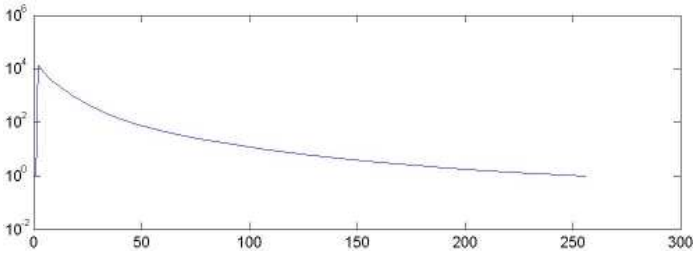
The set of selected candidates (designated as C_0) obtained from the CS stage will include many true MAs and several false candidates from the clutter class. As per the rejection approach, a cascade of rejectors which characterize non-MA is now designed. The goal is to reject the false candidates, while retaining as many of the true MA as possible.

Many of the false candidates occurring in C_0 are due to clutter such as

- points where vessels turn sharply
- depressions appearing on vessels



(a) Threshold characteristics : Sensitivity against t

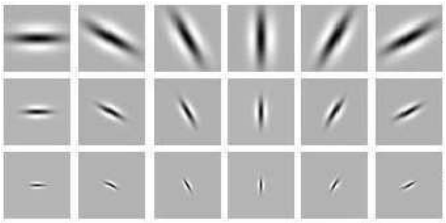


(b) Threshold characteristics : fppi against t. Vertical axis is logarithmic scaled

- junction of small vessels
- points in the optic disk
- small islands enclosed by bright lesions.
- depressions amidst hemorrhages
- small flame-shaped hemorrhages.
- noise pixels, laser artifacts, and other structures

Broadly the above clutter class can be grouped into two subclasses: vascular versus non-vascular clutter. Thus, we aim to design two rejectors to suppress them. The first rejector is intended to reject false candidates occurring on vasculature. Two reasons motivate this: (1) vascular structures are comparatively easier to model than non-vascular clutter; and (2) false candidates on vasculature occur very frequently in C_0. The list above is not exhaustive, and depending on the dataset, clutter may arise due to unforeseen factors or photometric variations. Hence, supervised learning techniques are chosen for the rejectors.

While true MA samples can be obtained out of expert-annotated data (ground truth), false positive samples have to be chosen in order to guide the learning. This is important since the rejectors are meant to model the clutter and discriminate them from true MA. The technique and the features used in the rejectors are elaborated below.



(c) Anisotropic filters

(d) Inverted Gaussian filters

Figure 3.7 Filters used in RJ_1

3.3.1 Rejection Stage 1 (RJ_1)

The task of RJ_1 is to identify from C_0, the known class of clutter namely, candidates in vasculature, hemorrhages, vessel junctions in the optic disk, etc.

The candidates in C_0 being local minima of I_{green}, are isolated points. Their local context in I_{green} provides a clue about their location of occurrence. Hence, information about the local context of each candidate is extracted and used to decide if a candidate is to be rejected. The information extracted from each candidate consists of responses to some specially designed filters, and scale-specific statistics, explained below. The local context of each candidate is a square neighborhood centered at the candidate, taken in I_{green}.

Feature Set-1

Anisotropic Filters: Vessel fragments can be modeled as elongated structures. A set of oriented (second-derivative of Gaussian) filters are designed to detect these elongated structures

The analytical expression for the second derivative in x-direction is found using 1-dimensional kernels, using the following relationships:

$$g_\sigma(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{x^2}{2\sigma^2}) \tag{3.5}$$

$$g'_\sigma(x) = -g_\sigma(x) \times \frac{x}{\sigma^2}$$

$$g''_\sigma(x) = g_\sigma(x) \times \frac{x^2 - \sigma^2}{\sigma^4} \tag{3.6}$$

A smoothed anisotropic Gaussian second derivative filter g_{xx} is constructed using separability as

$$g_{xx}(x, y, \sigma) = g''_\sigma(x) g_{\sigma_c}(y), \tag{3.7}$$

where σ_c is the standard deviation of a static 1-dimensional smoothing Gaussian function.

Such oriented filters should help in discriminating between false candidates on vessels and true MAs by way of high response to the former and low response to the latter.

A bank of filters at 6 equi-spaced orientations and 3 different scales are used at the output of which the maximum (r_m), variance (r_v) and sum (r_s) of the responses are computed. The following features are then derived for each candidate at each scale:

- $(r_s - r_m)$: this difference is high for true MA locations which are characterised by high r_s (about 6 times that of r_m) compared to clutter located on vessels.
- r_v: this is expected to be low at true MA locations, and high at vessel and junction locations.

The psf of the filters are depicted in Fig. 3.7. A total of 6 features is thus derived from the filters.

Scaled difference-of-Gaussians: A difference of Gaussian (DoG) filter acts as a blob detector, giving a high response to dark, isotropic structures. We introduce a variant of DoG, given by

$$f_d = \alpha\, g(\sigma_2) - g(\sigma_1) \tag{3.8}$$

where $\sigma_1 < \sigma_2$ and $\alpha > 0$ is a parameter controlling the height of the rim (see Fig. 3.8), σ_2 controls the width of the rim. At a candidate resembling a well-defined MA, this filter's response r_d is high. If a candidate lies on a vessel, r_d is low value (going negative if the vessel is thick). This is hence an informative feature for discrimination.

The anisotropic and DoG filters described above are similar to the centre-surround mechanisms, tuned to oriented structures, found in early stages of biological vision systems. Specifically, they are equivalent to centre-off types of ganglion cells.

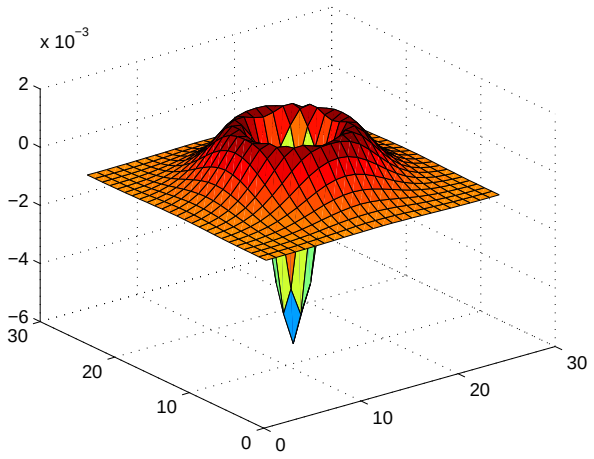


Figure 3.8 Scaled difference-of-Gaussians

Inverted Gaussians: While the first two type of features help detect clutter in the vasculature class, a second type of clutter structure that is similar to MAs are hemorrhages. In order to capture these, inverted Gaussian filters at high scale ($\sigma = 2, 4, 6$) are used. These filters will maximally respond to larger objects such as hemorrhages and thick vessels in contrast to well-defined MA. These responses r_{g_i} are hence included in our feature set.

The above features are intended to capture information to aid in discriminating candidates on vascular structures from true-MA. The model used for characterization is explained next.

Table 3.1 FS_1: Features extracted at each candidate, for RJ_1

Feature	Description
$(r_s - r_m)$	Difference between sum and max of responses from rotated $g_{xx}(\sigma_2)$ at 3 scales. $\sigma_2 = 2^{(i/2)}, i = 3, 4, 5$
r_v	Variance of responses from rotated g_{xx} at 3 scales (σ_2)
r_d	Response to scaled DoG filter
$r_g(\sigma)$	Response to inverted Gaussian ($\sigma = 2, 4, 6$)

Classifier-I

The design of the feature vector FS_1 is such that the feature-vectors corresponding to true samples occupy the positive (first) hyper-quadrant of the feature space and are agglomerated near the coordinate origin (have low positive values). In contrast, the feature vectors corresponding to false samples are scattered in the feature space away from the origin.

We use the nearest-mean classifier, which computes the mean of the true and false training samples, and stores them as prototypes. A new sample x_q is labeled by considering the distance to the prototypes and assigning the label of the nearest prototype to the new sample:

$$l_q = \arg\min(||x_q - \mu_i||), i = true, false \tag{3.9}$$

where μ_i is the prototype of class i in the training set.

The RJ_1 is trained offline using training data. Since both true and false MA samples are required for training these are obtained as follows. Given a set of training images, the candidates C_0 are selected first. Then the subset of true MA ($C_{0_{true}} \subset C_0$) is found. A random sampling is done over $C_0 - C_{0_{true}}$ to obtain false samples $C_{0_{false}}$.

3.3.2 Rejection Stage 2

The second stage is designed to handle the remaining class of (largely unknown) clutter objects. The function of this rejector is to identify from the candidates C_1 passed by RJ_1, the remaining class of clutter objects. These clutter arise due to a variety of reasons including image noise and are difficult to model. Hence, a very different strategy is required for their suppression.

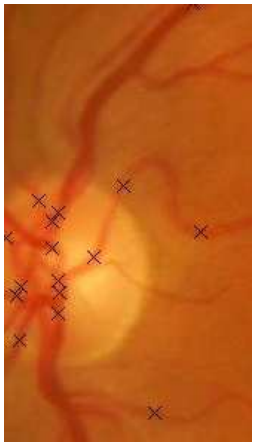


Figure 3.9 Subimage indicating candidates rejected by RJ_1 (indicated with cross)

A more general perspective of the problem of target-clutter separation, is outlier detection. Here, an outlier is defined as a sample which appears to be inconsistent with the remainder of the data [Barnett 94] (or abnormal). By modeling the target, outliers to the model can be isolated as clutter, and rejected.

Hodge and Austin [Hodge 04b] describe three fundamental approaches to outlier detection:

Approach-1: unsupervised clustering , with no prior knowledge of the data (no modeling of the source or underlying semantics)

Approach-2: model both the normal and abnormal (akin to supervised 2-class classification)

Approach-3: model only the normal, or in few cases model only the abnormal

The approach-3 is analogous to semi-supervised recognition. The method only learns the data marked as normal, and requires no abnormal data. A system based on approach-3 verifies if a query sample is within the boundary of normality. This method is capable of correctly labeling new abnormalities, as long as the sample is outside the boundary.

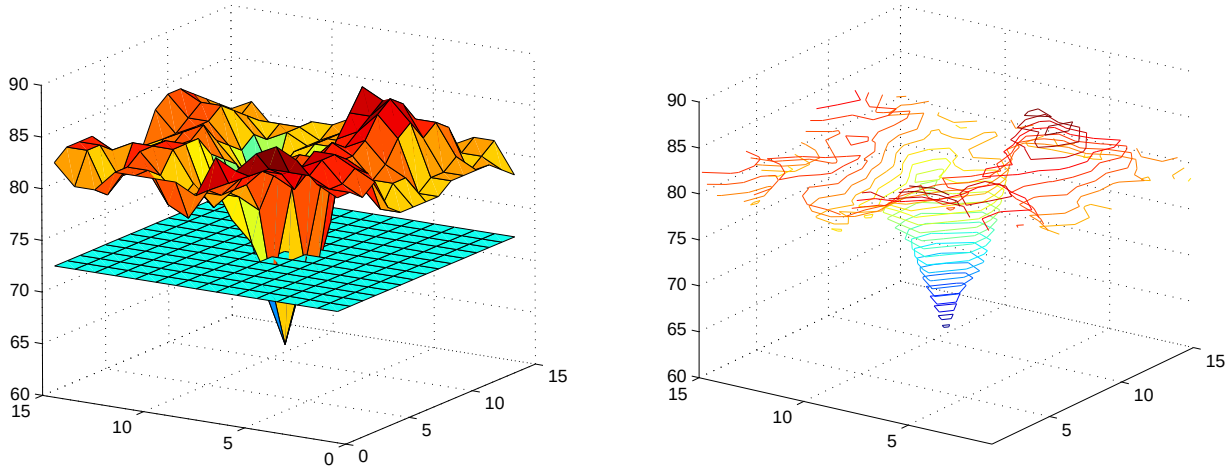
We follow the approach-3 and model the true MA to recognize clutter as outliers. True MA samples are thus used to build a model. The features that are extracted are designed to support the model, by verifying the isotropic nature and absolute topography of the candidates. A new set of features is proposed here for achieving this.

Feature Set-2

Distance feature: In FS_1, the distance between a sample x_p and the true-sample prototype of RJ_1 (denoted as $d_{true} = ||x_p - \mu_{true}||$) encodes a condensed information about the sample. The value of d_{true} is low for candidates that are similar in appearance to well-defined MAs. It is thus carried forward to RJ_2 as a feature.

Correlation features: Isotropy of a structure may be characterized by invariance of the topography with respect to in-plane rotation about its centre. A set of features to capture this information would be the correlation between a local neighborhood containing the structure, with itself after rotation. A high correlation at several orientations indicates a highly isotropic structure. The set of such correlation values is used to quantify the isotropy of the candidate.

Thus the second set of features for RJ_2 comprises of values obtained by correlating a square window (from I_{green}) around the candidate (just larger than the expected size of the lesion), with rotated versions of the window. The rotation is performed about the minima of the candidate. We use 5 equally-spaced orientations (each $\pi/5$ radians apart), to get 5 correlation values. These features are denoted as $\mathbb{R}_{\theta\theta}$.



(a) A plane at level $l = 72$ sectioning the surface defined by the grayvalues of a candidate neighborhood

(b) Contours of the above surface at 32 equally-spaced levels (M=32)

Figure 3.10 Illustrations of level cuts at a candidate

Features based on level cuts: The local grayscale topography around a candidate can be represented using iso-contours or level-curves of the local neighborhood considering it as a height map. We derive a set of features based on level "cuts", which we define as filled level-curves.

Structures resembling MA are local minima in I_{green}. Therefore the level curves at a MA-like candidate can be expected to be closed curves, making it is possible to perform filling within each level-curve, to obtain a finite area. We call this area a *level-cut*.

Fig. 3.10(a) depicts the topographic surface obtained by visualizing the local grayscale neighborhood of a candidate as a height map. A plane parallel to the ground plane (at level l) when intersecting with the surface, sections it and the intersection points define the level-curve (shown in Fig. 3.10(a)). A level cut is the closed area bound by a level-curve, containing within it the coordinate of the minima.

The area of a level-cut at level l_i is taken to be the number of pixels in the level cut and is denoted as $A(l_i)$. The features we propose are based on observing how A changes in the levels relevant to the candidate neighborhood.

At each candidate, the lowest and highest relevant levels, denoted as l_{min} and l_{max}, are found from the minimum and maximum gray values within a window (of radius 5) centered at the candidate minimum. M equi-spaced level cuts are chosen between these extrema and the area $A(l_i);\ i = 1, 2, ..., M$ of each level cut is determined and used to derive the following features:

$d_1 = l_{max} - l_{min}$: the estimated depth of the candidate grayscale topography

$l_c = \arg\max\{A(l_{i+1})/A(l_i); i = 1, 2, \ldots, M\}$: this denotes the level at which the level-cut area changes suddenly at the next level. (the approximate rim-level of the candidate).

ν: the ratio of volume of the candidate, to the volume of an inverted cone with base area $A(l_c)$, and height h:

$$\nu = \frac{V_c}{A(l_c)h/3} \tag{3.10}$$

$$\text{where } V_c = \sum_{i=0}^{l_c} A(l_i), \qquad\qquad h = d_1\ l_c/M$$

Table 3.2 FS_2: Features extracted at each candidate, for RJ_2

Feature	Description
d_{true}	Distance of sample from μ_{true} of FS_1
$\mathbb{R}_{\theta\theta}$	Correlation of candidate with small window at 5 angles of rotation (36^o)
m/d_1	Depth of the candidate
$A(l_1)$	Area at the first level above l_0
m/l_c	where l_c is the "rim-level" of the candidate
$A(l_c)$	Area of the candidate
Γ	measure of "jump"
Ω	measure of "overflow"
ν	Volume of the lesion relative to volume of cone of similar dimensions

Ω: This is a measure of "overflow". $\Omega = \frac{\partial A}{\partial l}\big|_{l_c}$
$= A(l_c + 1) - A(l_c)$.
Γ: This is a measure of "jump". $\Gamma = \frac{A(l_c+1)}{A(l_c)}$.

Figure 3.11 Subimage indicating candidates rejected by RJ_2 (indicated with blue squares)

Classifier-II

In this feature space (FS_2), the true samples are designed to agglomerate near the origin, and false samples are ideally scattered. The false samples are thus amenable to discrimination as outliers to a model dictated by the distribution of true samples in the feature space.

We model a hyper-cuboid H around the true samples, defined by the range occupied in each feature dimension for the true samples. The true samples ideally have a limited range and enclose the samples within H near the origin. False samples lie outside the hyper-cuboid obtained. The dimensions of the model H are stored. Given a new sample it is labeled as a true MA if it lies within H and rejected otherwise.

For training RJ_2, true samples are taken from the output of RJ_1 (ie. C_1) for known images.

3.3.3 Similarity Measure Computation (L)

The rejector cascade outputs a set C_2 of candidates which are likely to be true MAs. This module assigns a numerical confidence value to each sample in C_2, indicating the chance of it being a true lesion.

The confidence metric is based on similarity between the sample and a model of a true MA. This model can be obtained using supervised learning. We choose to perform the confidence assignment by considering the signed distance of a sample from the optimal hyperplane of a two-class SVM, in feature space. The technique we apply, and our chosen features are explained below.

Feature Set-3

To help in modeling true-MA, a few features are included from the previous stages. These are:

From RJ_2 : $d_{true}, A(l_1), A(l_c), \Gamma, \Omega, \nu$.

Additionally, some features based on context and symmetry are included, as described below.

Context features: A set of context features are also computed, which consider the pixels within the candidate, and a context surrounding it.

- Difference in mean value of the candidate region and its surround computed in 4 spectral bands: red, green, blue and hue. $msd_j = mean_j(cand) - mean_j(surround)$, where $j = \{red, green, blue, hue\}$
- The response of the candidate to a center-surround binary filter [Lienhart 02] with off-center. This is used as a rough descriptor of local minima along with its context.
- The perimeter p of the candidate, found as the number of pixels in the level curve at l_c (defined in FS_2)
- Mean response of derivative of Gaussian filter bank: $g_x, g_y, g_{xx}, g_{yy}, g_{xy}$ at pixels within the candidate (5 filters at 4 scales each, resulting in 20 features; scales used are $\sigma = \{1, 2, 4, 8\}$)
- standard deviation of response from the above filter bank

Symmetry features: A set of 8 features is obtained at each candidate by filtering with rotated Haar-like wavelets [C.Papageorgiou 98], shown in Fig. 3.12. The vertical 2-dimensional non-standard Haar wavelet is rotated in 16 orientations (each separated by $\pi/8$) to get 16 filters, as shown in Fig. 3.12. The axially anti-symmetric feature pairs (columns in Fig. 3.12) capture symmetry of the candidate along different axes, and the ratio of the pair responses is used as features (8 in number).

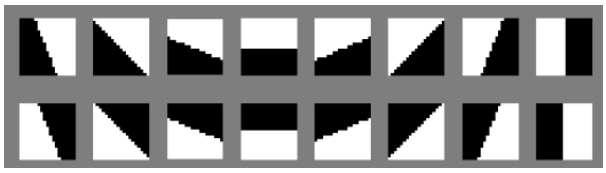


Figure 3.12 Ratio order of rotated Haar wavelets

The training data consists of the true MA, and FPs from the output of RJ_2 for known images. The feature values of the training data are normalized such that the mean of each feature is 0, and variance is 1.

Choice of classifier: Confidence values can be assigned using a simple supervised scheme with a k-nearest-neighbors classifier (knn). Let x_q be a novel sample to be ranked. For a given training set, let n_t be the number of true samples and n_f the number of false samples occurring as the k nearest neighbors of x_q , such that $n_t + n_f = k$. Then, the probability of x_q being a true sample can be taken to be n_t/k. This fraction relates to the confidence value for x_q, supported by the training data. Though practically simple, this classifier presupposes some properties over the feature space and the distribution of true and false training samples. Knn classification has been shown to perform exceptionally well when the training data is carefully selected, for multi-category problems [Boiman 08].

Our training data however, is bi-partite labeled, and is imbalanced (false samples are more numerous than the available number of true samples for training). Thus we propose an alternative, which is based on the distance of a sample from the optimal hyperplane of a support vector machine (SVM). A

Table 3.3 FS_3: Features extracted at each positive, for L

Feature	Description
d_{true}	Distance of sample from μ_1 of FS_1
$A(l_1)$	Area at the first level above l_0
$A(l_c)$	Area of the candidate
Γ	measure of "jump"
Ω	measure of "overflow"
ν	Volume of the lesion relative to volume of cone of similar dimensions
md_r	Difference in mean red band values within the candidate and its surrounding region
md_g, md_b	Same as previous, in green and blue band
md_h	same as previous, in hue plane
c.s	Response of center-surround binary filter
p	Perimeter of the candidate
$mr(\sigma), sd(\sigma)$	Mean and std. deviation of response to Gaussian derivative filters $g_x, g_y, g_{xx}, g_{xy}, g_{yy}$ within the candidate. $\sigma = 1, 2, 4, 8$
s.f	symmetry features from non-standard Haar wavelet

strength of SVM is its ability to handle imbalanced distributions of true and false samples [Cortes 95]. Additionally, it permits the use of non-linear kernel transformations, to overcome hyperplane linearity assumption. The SVM is trained on C_2 obtained at the output of RJ, with a set of known training images. The confidence value is set to be proportional to the distance from the hyperplane. The derivation of the approach is detailed below. To summarize the outcome, the confidence measure ψ (a function of x) obtained is such that it models a posterior probability of the two-class SVM assigning a label "true-MA" to x_q, given its feature values, i.e, $\psi(x_q) = p(y_q \leftarrow true|x_q)$.

3.3.3.1 Hyperplane-distance based confidence assignment

The operating condition for a supervised linear classifier can be expressed as

$$y_i(w^T x_i + b) \geq 1 \tag{3.11}$$

where x_i is a sample in feature space, $y_i \in \{1, -1\}$ its determined label, and w is the normal to the separating hyperplane of the true and false samples in feature space (assuming separability). Supervision in the form of labeled samples $D = \{(x_i, y_i)\}$ helps to find the optimal w that separates the samples in D.

For a novel (query) sample x_q classification determines the label y_q using the following condition:

$$y_q = \begin{cases} 1 & \text{if } w^T x_q + b > 1 \\ -1 & \text{if } w^T x_q + b < -1 \end{cases}$$

This may be considered as a linear projection followed by a thresholding step, where the projection defines a function $h(x_q) = w^T x_q$ on the real line. This is a signed distance between x_q and the hyperplane; h is positive for true training samples and negative for false training samples.

Our interest is in obtaining a function ψ which maps from x_q to a confidence value. If we choose the range $[0, 1]$ for the confidence value, the desired function can be formulated as a probability mass function. An interesting formulation that captures the notion of lesion confidence is the posterior probability of y_q being assigned the value "true-MA", given the query sample, that is

$$\psi(x_q) = p(y_q \leftarrow true | x_q). \tag{3.12}$$

For a sample lying close to w with positive h, the desired value of $\psi \geq 0.5$, and a sample with negative h the desired $\psi < 0.5$.

Let us consider a simple piecewise-linear model for mapping from the signed distance $h \in [-\infty, \infty]$ to $\psi \in [0, 1]$. The line $h = m(\psi - 0.5)$ linearly stretches ψ using a slope of m, with range $[-m/2, m/2]$ and zero intercept at $\psi = 0.5$. Fig. 3.13(a) shows this model for a value of $m = 100$. It is clear from the figure that the mapping is odd-symmetric.

The mapping from h to ψ using this model can be expressed as a saturating linear function given as the following piecewise function:

$$\psi = \begin{cases} h/m + 0.5 & \text{if } h \in [-m/2, m/2] \\ 1 & \text{if } h > m/2 \\ 0 & \text{if } h < m/2 \end{cases}$$

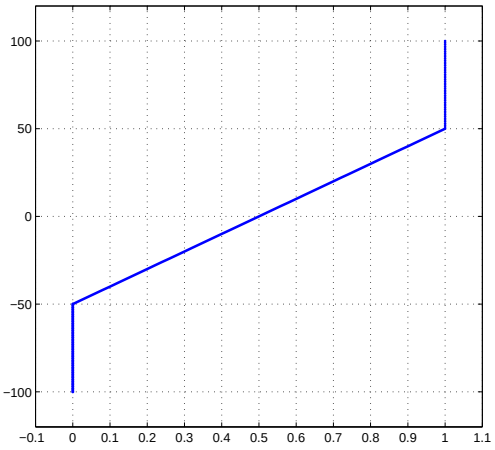
This model has a free parameter m which determines the saturation of ψ. Apart from being discontinuous and hence indifferentiable (relevant in the ensuing context), m also restricts the saturation at both the extremes owing to symmetry. A better model to link ψ to h without discontinuities is the logit function, given as

$$\hat{h}(x_q) = \log\left\{\frac{\psi(x_q)}{1 - \psi(x_q)}\right\} \tag{3.13}$$

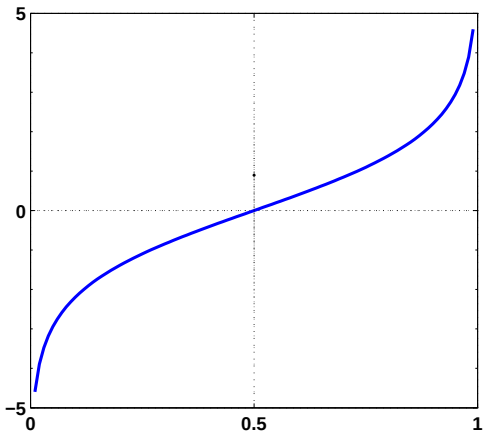
This is a monotonic model (see Fig. 3.13(b)), and $\hat{h}(x_q)$ emulates a signed distance with the desired properties over ψ, since

$$\hat{h}(x_q) \begin{cases} = 0 & \text{if } \psi(x_q) = 0.5 \\ > 0 & \text{if } \psi(x_q) > 0.5 \\ < 0 & \text{if } \psi(x_q) < 0.5 \end{cases}$$

Eqn. 3.13 can be re-written to express ψ in terms of $\hat{h}$ (dropping the parameter x_q for notational convenience) as



(a) linear model



(b) The logit model. The horizontal axis denotes ψ and the vertical axis $\hat{h}$. The model is asymptotic in the vertical axis

$$\begin{aligned} \hat{h} &= \log\left\{\frac{1}{1/\psi - 1}\right\} = -\log\left\{\frac{1}{\psi} - 1\right\} \\ \exp(-\hat{h}) &= \frac{1}{\psi} - 1 \\ \psi &= \frac{1}{1+\exp(-\hat{h})} \end{aligned} \tag{3.14}$$

The model thus evaluates the posterior using a sigmoid function (Eqn. 3.14). Including a bias term ($B \geq 0$) and a scale factor ($S > 0$) in Eqn. 3.14, a more general posterior density can be expressed as

$$\psi = \frac{1}{1+\exp(-S\hat{h} - B)}. \tag{3.15}$$

We now replace $\hat{h}$ with h, the actual linear projection trained from a linear classifier using D. Thus the problem of finding a confidence function ψ has been reduced to finding optimal values of scalars S and B, which are consistent with D.

We pose this as a maximization of ψ over the true samples and minimization of ψ (equivalently, maximization of $1-\psi$) over the false training samples. Let D_t be the true samples and D_f be the false samples in D (such that $D = D_t \cup D_f$). In the training set, if $h_i \triangleq h(x_i)$ and $\psi_i \triangleq \psi(x_i)$,

$$(S, B) = \arg\max\Big\{\prod_{i \in D_t} \psi_i \prod_{j \in D_f} (1 - \psi_j)\Big\} \tag{3.16}$$

Equivalently, we consider the logarithm of the above expression, to transform the product to summation:

$$(S, B) = \arg\max \Big\{ \sum_{i \in D_t} \log(\psi_i) + \sum_{j \in D_f} \log(1 - \psi_j) \Big\} \tag{3.17}$$

Using Eqn. 3.15, $\log(\psi) = -\log(1 + \exp(-Sh - B))$, and

$$\begin{aligned} 1 - \psi &= 1 - \frac{1}{1 + \exp(-Sh - B)} = \frac{\exp(-Sh - B)}{1 + \exp(-Sh - B)} \\ \text{Thus } \log(1 - \psi) &= (-Sh - B) - \log(1 + \exp(-Sh - B)) \\ &= \log(\psi) - Sh - B \end{aligned} \tag{3.18}$$

Thus Eqn. 3.17 may be simplified as

$$\begin{aligned} (S, B) &= \arg\max \Big\{ \sum_{i \in D_t} \log(\psi_i) + \sum_{j \in D_f} (\log(\psi_j) - Sh_j - B) \Big\} \\ &= \arg\max \Big\{ \sum_{i \in D} \log(\psi_i) - \sum_{j \in D_f} (Sh_j + B) \Big\} \\ &= \arg\min \Big\{ \sum_{i \in D} \log(1 + \exp(-Sh_i - B)) + \sum_{j \in D_f} (Sh_j + B) \Big\} \end{aligned} \tag{3.19}$$

We perform Newton descent to iteratively solve for (S, B). The update rule is given by

$$x_{k+1} = x_k + \eta H^{-1} F(x_k) \nabla F(x_k) \tag{3.20}$$

where $x = [S\ B]^T$, η is the step size, $H^{-1}F$ is the inverse of the Hessian of the objective function F (RHS of Eqn. 3.19), and ∇F is the gradient of the objective function.

3.3.3.2 Training the SVM-based confidence assignment stage

Let the number of training samples $|D| = n$. Training the SVM consists of minimizing the following Lagrangian expression with respect to w:

$$L_P \equiv \frac{1}{2}||w||^2 - \sum_{i=1}^{n} \alpha_i \{ y_i (w^T \phi(x_i) + b) - 1 \} \tag{3.21}$$

where ϕ is a non-linear transformation defining the kernel function as $k(x_i, x_j) = \phi^T(x_i)\phi(x_j)$, and $\alpha_i \geq 0$ are Lagrangian multipliers.

Enforcing $\partial L_P / \partial w = 0$ yields $w = \sum_i \alpha_i y_i \phi(x_i)$. Let $D_s = \{x_{s_i}\} \subset D$ be the subset of training samples with $\alpha_{s_i} > 0$ (for all other samples, $\alpha_i = 0$). D_s is the set of support vectors (the samples which

influence w). Training the SVM consists of determining the support vectors and their corresponding Lagrangians α_{s_i} (and the scalar bias term b).

The classification condition is thus

$$y_q(w^T\phi(x_q) + b) \geq 1. \tag{3.22}$$

Comparing Eqn. 3.22 with Eqn. 3.11, we see that

$$\begin{aligned} h(x_q) &= w^T\phi(x_q) && (3.23)\\ &= \{\sum_{x_{s_i}} \alpha_{s_i} y_{s_i} \phi(x_{s_i})\}^T \phi(x_q) \\ &= \sum_{x_{s_i}} \alpha_{s_i} y_{s_i} \phi^T(x_{s_i})\phi(x_q) \\ &= \sum_{x_{s_i}} \alpha_{s_i} y_{s_i} \kappa(x_{s_i}, x_q) && (3.24) \end{aligned}$$

On a dataset of 50 images (PDS-2 dataset), the training set extracted from these images had 336 true and 3541 false samples. The trained SVM had 523 (142 true + 381 false) support vectors.

Once the support vectors and their α_{s_i} are computed, we use Eqns. 3.15, 3.24 in Eqn. 3.19, to find S and B. For this we use only the non-support vectors (i.e., $D - D_s$). Geometrically, this ensures that within the SVM margin the distance function evaluates to 0 (and samples within the margin receive a confidence of 0.5). The non-support vectors are away from the margin, and hence contribute to faster convergence while minimizing Eqn. 3.19.

Figure 3.13 An image showing detected MAs with confidence values. Candidates rejected in RJ-1 and RJ-2 are shown in dark cross and square

Chapter 4

Experimental Evaluation

The proposed method was evaluated against different data sets to study its performance against possible variations and challenges that confront an automated MA detection system. For the purpose of evaluation three datasets were considered: two are the publicly available datasets namely, the DIARETDB1 [Kauppi 07a] and ROC [Abramoff 07] datasets. We will henceforth refer to these as PDS-1 and PDS-2 respectively to emphasise the fact they are public datasets and to distinguish them from a custom-built set called CRIAS. This dataset is composed of images collected for clinical purposes in a local hospital and hence represents a homogeneous population.

4.1 Datasets and Ground Truth

PDS-1 consists of 89 images in uncompressed PNG format, of which 5 images do not contain any DR-indicative lesions.The images were collected from a screening program and taken under a fixed imaging protocol. The images were selected by the medical experts, but their distribution does not correspond to any typical population. The annotation supplied with this dataset is a soft map consisting of regions indicating expert consensus level information averaged from multiple experts. A bright region thus indicates high consensus about the presence of MA. According to the guidelines given with the dataset, our evaluation of the presented method is done on a 75% consensus level (relative to maximum) as the ground truth. A total of 182 MAs are obtained at 75% consensus level.

PDS-2 consists of 50 training images with associated ground truth, and a test set of 50 images whose ground truth is retained by the organizers of the ROC competition [Abramoff 07][Niemeijer 09]. The images are taken from a DR screening program across multiple sites, and hence captured with different cameras, fields of view and resolution. The images in this set are relatively heterogeneous [Niemeijer 09] and in compressed JPEG format. The supplied annotation for the training set is obtained by merging the annotation of 4 retinal experts: if at least one expert has identified a lesion, it is recorded in the annotation. The images were acquired without dilating the pupil, which leads to variations in image quality. The number of lesions in the training set is 336. The test set contains a total of 343 MAs.

Table 4.1 Dataset specifications under different related factors. Abbreviations used: IVW: Illumination variation with-in images; IVA: Illumination variation across images; BLA: Blurring and lighting artifacts; CP: Images taken under a common protocol; ICT: Image compression type[UC:uncompressed/ C: compressed]

	No. of images	**Imaging Factors**						
		Cameras	**FOV**	**IVW**	**IVA**	**CP**	**Image resolution**	**Mydriatic**
PDS-1	89	fixed	50°	high	low	yes	fixed	no
PDS-2	100	varying	45°	low	medium	no	mixed	no
CRIAS	288	fixed	$30^\circ - 45^\circ$	low	high	yes	fixed	yes

	Image Quality				**Pathological type ratio**		**Ground Truth**	
	Clarity	**Contrast**	**ICT**	**BLA**	**High**	**Mild**	**Type**	**Number of experts**
PDS-1	low	low	PNG (UC)	low	none	high	soft	multiple
PDS-2	medium	medium	JPG (C)	low	low	medium	hard	4
CRIAS	high	high	TIF (UC)	high	high	low	hard	2

CRIAS is a dataset of 288 images taken from a local hospital. These images are mainly collected for clinical documentation and patient profiling. These images are of diabetic patients who have been diagnosed with DR. Mostly, these images have high pathology occurrence, several blurring and lighting artifacts, laser marks, pigmentation and illumination variations. A dilation is performed before imaging thus images having quite uniform illumination across images. Annotation was obtained from two training experts and total number of marked MAs is 1436 which is the highest among the three datasets.

The detailed specifications and other variability occurring in each of the selected datasets is summarized in Table 4.1.

4.2 Practical specifications

As mentioned earlier, by design, the proposed MA detection system is data driven and hence, there are no parameters to be tuned for evaluating the system. While training on each dataset, a value tol has to be provided for the CS stage. The value of tol applied for each dataset is shown in Table 4.2. The fourth column of this table gives the rate of occurrence of true lesions (MAs) per image in each dataset. This factor can be used to choose tol for a new dataset.

Table 4.2 Selection of tol

Dataset	n_{true}	N	n_{true}/N	tol
PDS-1	182	89	2.044	150
PDS-2	336	50	6.72	250
CRIAS	1436	288	4.98	250

The points captured in C_0 (refer Section 5B) may not be accurately localized within the candidate (the local minima might not be the center of the candidate). This can cause filter responses (used in RJ_1 and RJ_2) to deviate from the desired responses. To adjust for inaccuracies, we average the responses obtained by filtering with the center positioned at each of the 8-neighbors of the candidate location, with a weight of 1 for the 8 neighbors, and 1.2 for the coordinate stored in C_0.

Among the n_{true} lesions in each dataset, training is performed using 90% of the lesions, holding out the rest for evaluation. Each training stage in turn performs folded validation with 8 folds, and the best performing model is retained for each stage. False samples for each supervised stage are included at random (boot-strapped) from the output of its previous stage. The ratio of true- to false-samples used for training RJ_1 is 1:15. True-to-false ratio used for training the L stage is 1:5.

For the L-stage, a radial-basis kernel was chosen: $\kappa(x_1, x_2) = \exp(-\gamma||x_1 - x_2||^2)$, and a L2-soft-margin kernel-SVM (with slack coefficient=10) [Cortes 95] was trained.

Evaluating stage-wise performance

The performance of RJ_1 is evaluated using sensitivity ($s_1 = n_1/|C_{0_{true}}|$) and rejection rate ($rr_1 = n_2/|C_{0_{false}}|$) where n_1 is the number of samples labeled as true, n_2 is the number of samples labeled as false and $|C_x|$ denotes the number of candidates in C_x.

The performance of RJ_2 is evaluated using sensitivity ($s_2 = n_1/|C_{1true}|$) and rejection rate ($rr_2 = n_2/|C_{1false}|$) where n_1 is the number of samples labeled as true ($l_q = 1$), n_2 is the number of samples labeled as false ($l_q = 2$).

The net rejection achieved by RJ cascading RJ_1 and RJ_2 can be found by applying the relation

$$rr = rr_1 + (1 - rr_1/100)rr_2\%. \tag{4.1}$$

To evaluate the performance of the SVM, we used histograms depicting the likelihood (sample density) function of the true and false validation samples. A desired likelihood for the true class should have mean and mode above 0.5. A desired likelihood for false class should have low mean and mode close to 0. Fig. 4.1 shows the likelihood functions obtained on a validation set (from PDS-2).

4.3 Performance evaluation measure

The ground truth available with a dataset is used to determine the true-positives and the false-positives obtained overall by our MA detection method. The performance of the method is assessed by two measures: sensitivity, and fppi, based on the number of true- and false-positives encountered over images in the test set, in the following manner:

$$sensitivity = \frac{\sum_i TP_i}{\sum_i GT_i}, \qquad fppi = \frac{\sum_i FP_i}{N} \tag{4.2}$$

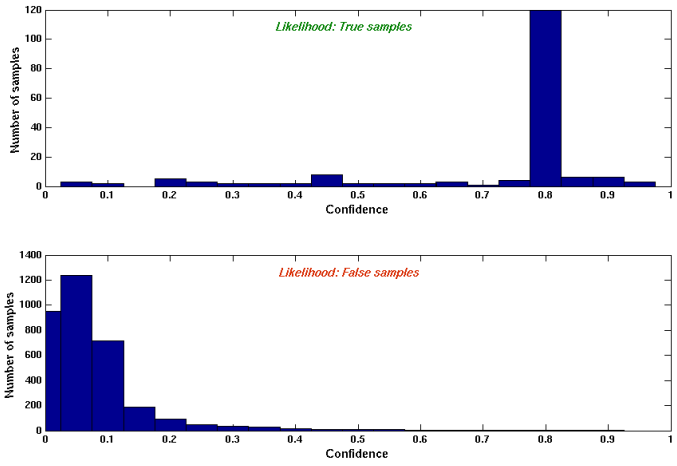


Figure 4.1 Likelihood functions for the true- and false-samples in validation set. At a confidence threshold of 0.5, the area (conditional density) under the true-sample likelihood is 84.66%, false-sample area is 98.72%. This shows that 85% of the true-MAs get a confidence value higher than 0.5, and 99% of false samples get assigned a confidence lower than 0.5, for the selected dataset).

where N is the total number of images, TP_i and FP_i are the number of true positives and false positives, respectively, obtained in the i^{th} image of the test set, and GT_i is the number of true-MA in the i^{th} image. As well known, sensitivity is independent of dataset size (N), but not fppi. Yet fppi is an informative measure with respect to lesion-level detection, and the two measures capture the detection-error trade-off. For an ideal detection, it is desired to achieve high sensitivity at low fppi.

Typically, detection methods are evaluated by computing sensitivity and fppi for each possible input parameter setting. Varying some control parameters results in different detector responses, thus yielding different values of sensitivity and fppi. These values are then plotted to obtain the free-response receiver operating characteristics (FROC) curve. A point in the FROC curve shows sensitivity obtained at the respective fppi for a single parameter setting. Traditional computation of FROC curve involves multiple runs of the detection method at different parameter settings.

Our presented approach for MA detection permits us to obtain a FROC curve in a simpler, straightforward manner. A single run of our method yields MAs, as well as a confidence value associated with each positive. From this set, we first take into consideration only those positives of each image receiving highest confidence value, and compare with expert annotation to compute sensitivity and fppi. By using a threshold k varying from 1 to 0 on the confidence values, we gradually increase the number of detection, and compute sensitivity, fppi value pairs at each k. Each pair gives a point on the FROC plot, and the points are then connected using straight lines to get the FROC curve. The trend of this curve matches the expected trend in a typical FROC.

This evaluation method is more informative compared to the traditional method. It is possible to attribute each point on our FROC to a confidence setting k_i, and a sequence exists in the curve; as k is reduced, fppi and sensitivity increase. The lowest point in our FROC is attributed to maximum

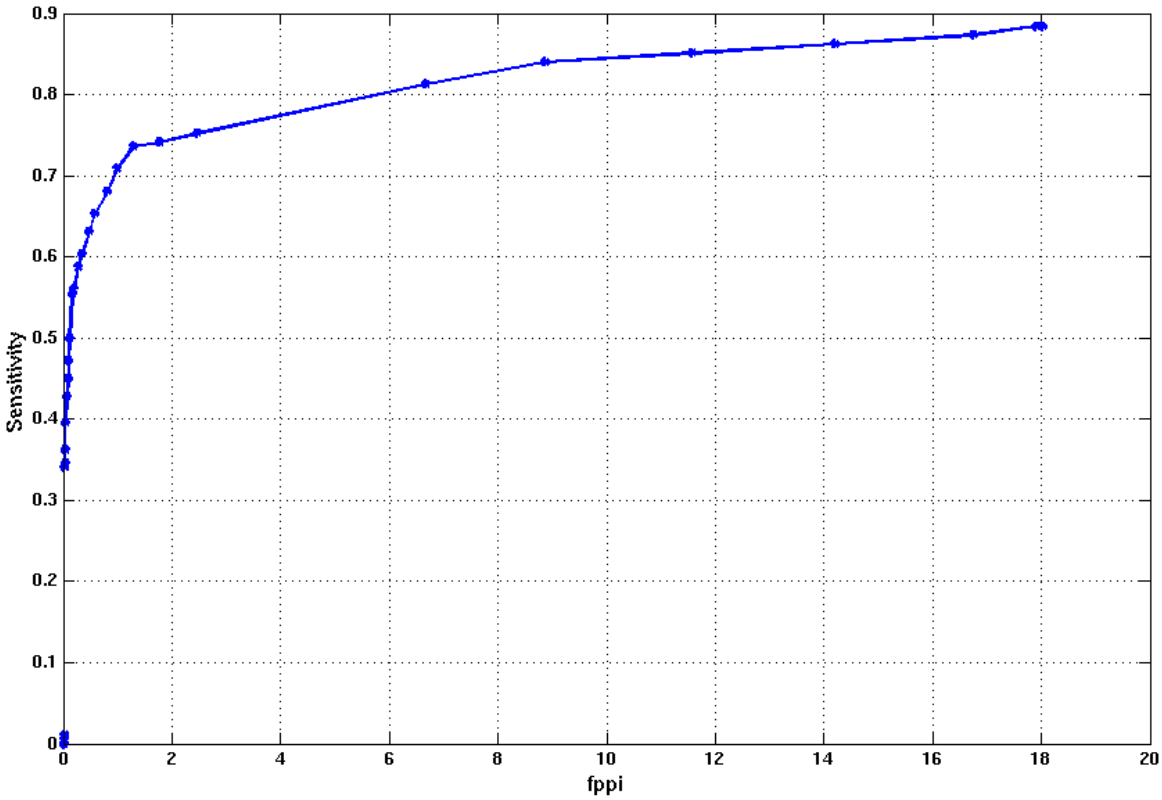


Figure 4.2 FROC curve on PDS-1

confidence ($k \rightarrow 1$), and the highest point corresponds to $k = 0$, i.e. the entire set C_2 of detected MA. Such an understanding cannot be elicited from the traditional FROC, where each point is obtained by varying one or more control parameters, and then connecting adjacent points. Moreover, in this new method, no specific knowledge of the working of the detector needs to be known to evaluate it or analyze the performance.

4.4 Experimental results

The data available in PDS-1 and CRIAS have associated ground truth. Hence a part of the training data is held out as the evaluation or test set. We use 10% holdout in these two datasets. In the case of PDS-2, the organizers of ROC have explicitly set aside a set of 50 test images (training on the test set is not possible since their associated ground truth is not revealed by ROC).

The overall performance for each dataset in terms of the FROC curve is discussed next.

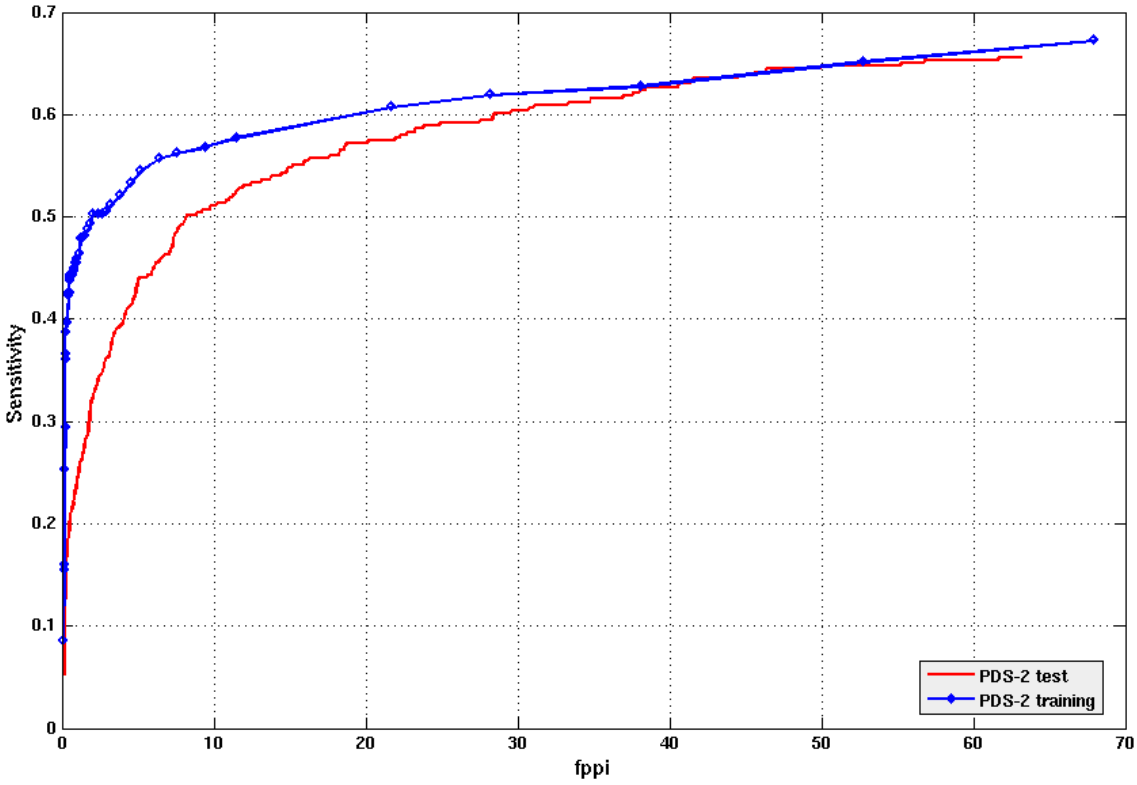


Figure 4.3 FROC curve on PDS-2 training and test set

4.4.1 Performance Analysis: PDS-1

Fig. 4.2 shows the FROC curve obtained for PDS-1. The FROC rises quickly at very low fppi. The highest sensitivity achieved is 88.46% at 18.02 fppi. The sensitivity figures at $fppi = 1, 2, 4, 8, 12, 16, 20$ are shown in Table 4.3 for convenience. At 1 fppi, the sensitivity achieved is 70.8%. The optimal point on the froc is at 1.2 fppi, with a sensitivity of 73.6%.

4.4.2 Performance analysis: PDS-2

Fig. 4.3 shows the FROC obtained for the PDS-2 training (in blue) and test (in red) datasets [Abramoff 07] [Niemeijer 09]. The maximum sensitivity achieved in the training set is 67.26% at 67.9 fppi. The best performance (among different algorithms [Niemeijer 09]) in the test set is also similar: 65.6% sensitivity at 63.1 fppi (refer Table 4.5).

The optimal performance point on the training curve is 54.46% sensitivity at 5.14 fppi. On the test curve, the optimal point is 50.15% sensitivity at 8.2 fppi. It is clear that the initial rising part of the FROC is receded in the test set, especially for fppi < 5. For fppi values beyond 30, the training and

test set performance are very similar. The initial lag is caused by misses at high k values in the test set compared to training set. Similar performance (over test and training sets) towards the end of the FROC curves indicates that CS and RJ stages perform equally well whereas the performance of the confidence assignment stage is sub-optimal.

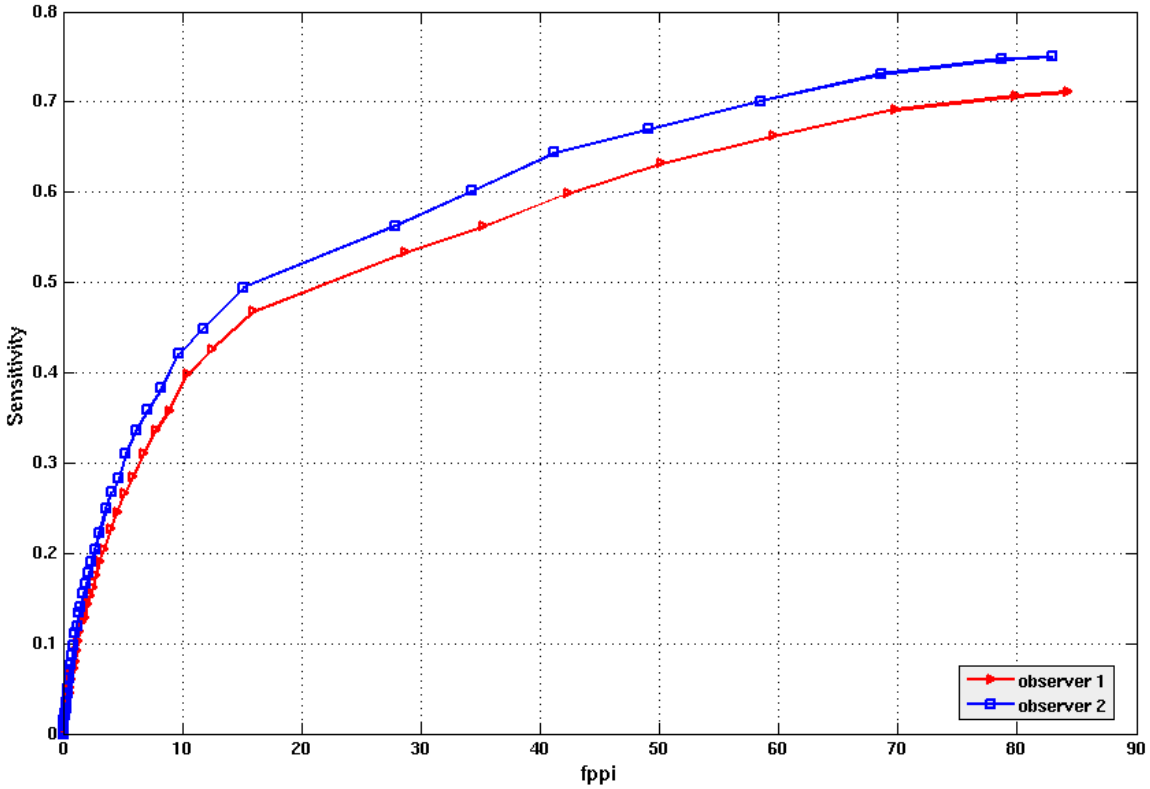


Figure 4.4 FROC curve on CRIAS dataset set (2 observers)

4.4.3 Performance analysis: CRIAS

The evaluation of our method against public datasets PDS-1 and PDS-2 was against multiple experts. In contrast, the performance on CRIAS dataset is assessed individually, against two observers. The training was performed using annotations of observer-1. Fig. 4.4 shows the FROC obtained against the two observers. The maximum (and optimal) sensitivity achieved against observer-1 is 71.17% (46.8%) at 84.1 (15.9) fppi. These figures against observer-2 are 75.03% (49.47%) at 82.8 (15) fppi.

Though the system was trained with the annotations of observer-1, the evaluation against observer-2 gave a consistently better performance. The last two rows of Table 4.3 show about 2 to 4% increase in sensitivity against observer-2 at each fppi value. Viewed alternatively, at a given sensitivity, the system is able to achieve lower fppi when evaluated with the annotation of observer-2.

The difference could be explained with reference to the sensitivity of the observers. The number of MAs marked by observer-1 on the CRIAS dataset is 1436. Observer-2 has marked 1510 MAs. The selectivity of observer-2 is thus lower (observer-2 marks one lesion more than observer-1 for every 4 images). Thus false positives will be lower when evaluated with observer-2. This results in the observed behavior of the FROC.

4.4.4 Comparative Performance Analysis

Fig. 4.5 indicates the FROC curves of all 3 datasets in a combined plot. It can be seen that the performance on PDS-1 is highest among the selected datasets. The performance over PDS-2 and CRIAS converge beyond 30 fppi, but at lower values of fppi, PDS-2 has better sensitivity. This section analyzes the proposed system to identify some reasons and factors governing performance.

Table 4.3 Performance on different datasets

Dataset	FPPI						
	1	2	4	8	12	16	20
PDS-1	0.708	0.742	0.78	0.83	0.85	0.87	0.88
PDS-2	0.45	0.503	0.520	0.562	0.57	0.59	0.6
CRIAS-1	0.09	0.14	0.22	0.34	0.42	0.47	0.49
CRIAS-2	0.11	0.17	0.26	0.38	0.45	0.5	0.52

The end-to-end performance of the system can be understood by examining the performance at individual stages. Table 4.4 shows the performance of the RJ (consisting of 2 rejectors) across the 3 datasets.

Table 4.4 Performance of RJ stage in the 3 datasets

	Sensitivity			**Rejection Rate**		
Dataset	RJ_1	RJ_2	overall RJ	RJ_1	RJ_2	overall RJ
PDS-1	98.9	97.35	96.23	33.14	40.75	60.38
PDS-2	98.9	98.32	97.23	33.14	22.18	47.96
CRIAS	95.55	99.72	95.28	35.12	21.09	48.8

It is seen that RJ_1 and RJ_2 maintain very high sensitivity, leading to 95-97% sensitivity for RJ. This comes by design (RJ is expected to pass the maximum number of true MAs, while rejecting specific types of non-MA). The high sensitivity levies a limit on rejection rate achieved in RJ. Table 4.4 shows that the rejection rate ranges from 60% in PDS-1, to 47% in PDS-2. The performance (sensitivity,fppi) achieved by CS and RJ combined (excluding the confidence assignment stage) are as follows. PDS-1: (88.4%, 18), PDS-2: (67.26%, 67.9), CRIAS: (75.03%, 82.8). These values show that although sensitivity is retained above 65%, the number of positives passed on to the L stage is very high in the case of PDS-2 and CRIAS (3-4 times the positives in PDS-1).

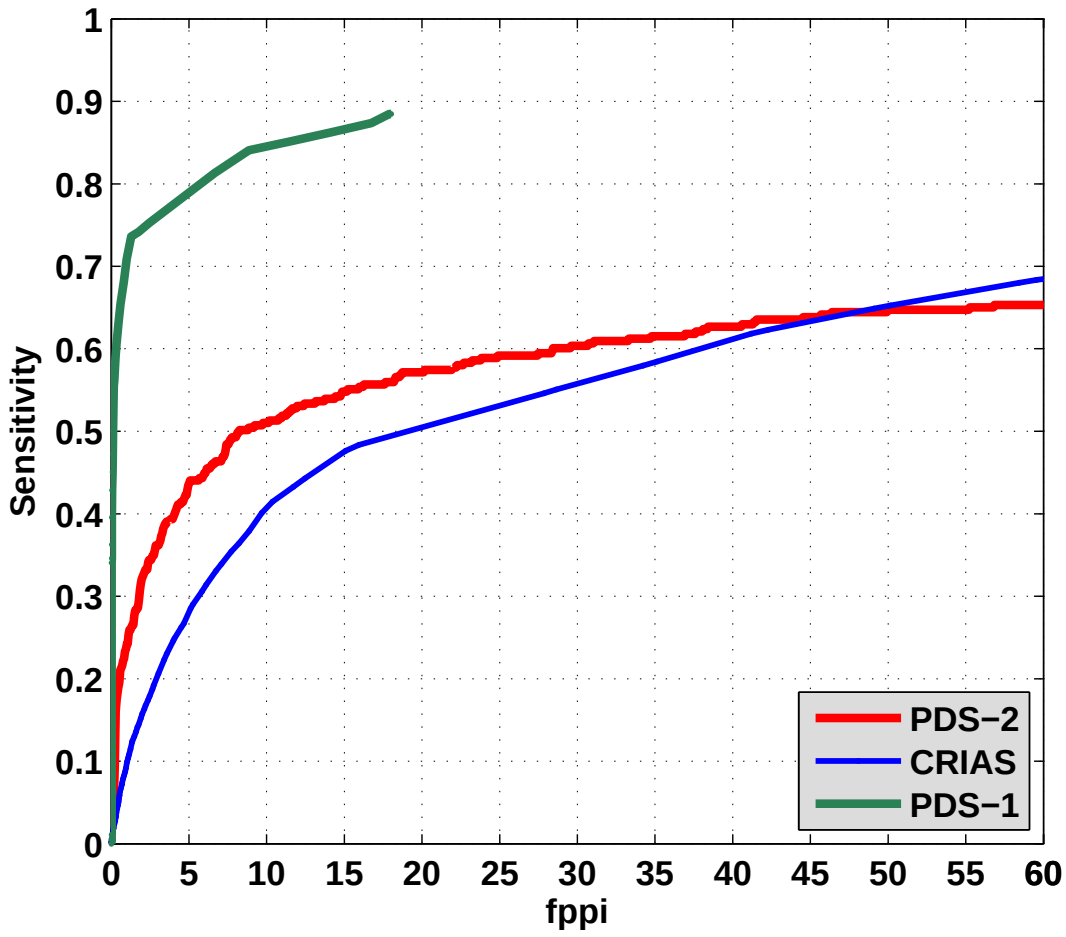


Figure 4.5 Performance curves over 3 datasets

The nature of the dataset has a role to play in this observation. PDS-1 having been obtained with restricted photometric variations, fewer lighting artifacts, blurs and fewer pathologies, is conducive for the candidate selection (CS) stage to achieve above 90% sensitivity at about 80 candidates per image. The task of RJ is simpler in PDS-1. Thus RJ achieves 60% rejection rate in PDS-1, sparing about 30 positives for L stage.

In PDS-2 and CRIAS, the CS stage is intimated of the challenge (see Fig. 3.3) in the dataset by permitting greater number of candidates (through the tol parameter). In CRIAS for example, CS stage achieves 78.9% sensitivity, by passing about 157 candidates per image. The RJ stage manages to maintain sensitivity at 75%, rejecting about 70 candidates. The fact that 82 positives arise from RJ indicates that the rejection achieved is insufficient. The onus is thus on the L stage to assign low confidence to the non-MAs. However, the non-MA passing to L stage are assured to be hard to classify (with the usual feature set), since easier false samples would have been rejected earlier.

Table 4.5 Performance by different methods on ROC (PDS-2) test image dataset [Abramoff 07] [Niemeijer 09]

	FPPI							
Method	$1/8$	$1/4$	$1/2$	1	2	4	8	**Score**
Niemeijer et al. [Abramoff 07]	0.243	0.297	0.336	0.397	0.454	0.498	0.542	0.395
Waikato Retinal Imaging Group [Niemeijer 09]	0.055	0.111	0.184	0.213	0.251	0.300	0.329	0.206
Fujita Lab [Mizutani 09]	0.181	0.224	0.259	0.289	0.347	0.402	0.466	0.310
LaTIM [Quellec 08]	0.166	0.230	0.318	0.385	0.434	0.534	0.598	0.381
OKmedical [Zhang 09]	0.198	0.265	0.315	0.356	0.394	0.466	0.501	0.357
GIB Valladolid [Sanchez 08]	0.190	0.216	0.254	0.300	0.364	0.411	0.519	0.322
ISMV [Abramoff 07]	0.134	0.146	0.202	0.249	0.286	0.345	0.430	0.256
Proposed Method on test	0.041	0.160	0.192	0.242	0.321	0.397	0.493	0.264
Proposed Method on training	0.1722	0.3903	0.4405	0.4592	0.5010	0.5238	0.564	0.436

While features play a key role in the performance of the L stage, the training data used is also a factor in our approach, since the system is largely data driven. The training data in the case of PDS-1 consisted of 75% expert-consensus lesions. Many of the true samples used for training L in PDS-1 dataset are thus bound to be well-defined MAs. This is important, and expected by L, since it models the probability of the given sample being labeled as "true" by the 2-class classifier h trained on the same data. Well defined MAs help to move the distribution-modes apart, whereas ambiguous MAs tend to bring them closer to 0.5 (equi-probable case).

The lower performance over PDS-2 and CRIAS is primarily due to the lack of consensus-based annotations as used in PDS-1. The annotations are very sensitive to inter-observer variability PDS-2 has multiple observers and a union of their markings has been taken to annotate an MA. In CRIAS only 2 experts have annotated the images and the variability between them has already been pointed out. The net result of this is that in both these datasets more fuzzy MAs are part of the true MAs. Additionally, these datasets exhibit variations in many other respects (refer Table 4.1) as well. Hence, these datasets do not satisfy the assumption of presence of many well-defined MAs in the true samples (see construction of Eqn. 3.16). Next, we do a comparative study of performance against PDS-2.

4.4.5 Performance comparison against other methods tested on PDS-2

The ROC technical report [Niemeijer 09] gives a detailed comparative analysis of 5 different detection methods on various performance aspects. A performance score is computed by taking an average of the sensitivities reported at fppi of [0.125, 0.25, 0.5, 1, 2, 4, 8] for each method. Table 4.5 shows the sensitivities and scores reported for the 5 different methods [Niemeijer 09] and by our method (last row). [1]

[1] Performance obtained on PDS-2 training dataset is also included in the table

The maximum and minimum reported scores in the ROC challenge are 0.381 and 0.206 respectively, and the mean is 0.315. In comparison, our proposed method obtains a score of 0.264. An analysis of sensitivity at different fppi reveals that at fppi values above 1, our method performs at par with other methods.

PDS-2 has images of mixed sizes and resolutions which poses a problem in detection. Most of the reported methods [Niemeijer 09] explicitly handle this problem. Our objective of this experiment is to assess detection performance on an unseen dataset or screening scenario where prior knowledge about the dataset is not available. Thus, we did not attempt to handle these variations or perform any special parameter tuning for each dataset. The performance of our method, viewed from this perspective, can be considered to be reasonably good.

4.5 Discussion

In this work, we formulate MA detection as a target detection in clutter problem and have identified a potential role for learning non-MA structures in the detection process. An approach has been developed utilizing the insights and complexities reported in the earlier work. A successive rejection based approach is proposed where rejection stages are arranged based on the occurrence frequency and discriminability of the underlying non-MA structures. A new set of morphological and appearance-based features are presented to characterize various non-MA and MA structures.

This approach has some inherent advantages over the strategy used in earlier work. First, it eliminates explicit segmentation of optic disk (OD) and vessels to suppress candidates arising on such structures. Over- or under-segmentation of OD or vessels affect the MA detection sensitivity and fppi [Abramoff 08][Fleming 06]. For instance, thin vessel junction and end points are numerous among the obtained candidates. Vessel segmentation employed to segment thin vessel may also include MAs (which are referred to as noise structures, from vessel segmentation point of view). In vessel segmentation, it is called a trade-off of thin vessel and small dark noise. Most of the segmentation methods use elongated property of vessel to suppress such noise. Quantitative vessel segmentation assessment can be performed using DRIVE [Staal] dataset but will not be adequate to quantify its effect on MA detection sensitivity. Another limitation is that MAs near vessels usually get suppressed due to their inclusion in the segmented vessel map. Overall, segmentation-based suppression of non-MA candidates is not an optimal strategy to employ.

In the proposed method, a learning-driven rejector (RJ_1) was designed to eliminate vascular non-MA candidates. Features based on specialized filters were used to characterize them. Table 4.4 summarizes individual stage performance achieved on 3 different datasets. It can be seen that the first stage gives a good performance. On an average it allows 97% of MAs to the next stage while rejecting 34% non-MA candidates, belonging to the vessel structure.

The second rejector also uses a learning-based strategy for rejecting the FP. In this stage, the choice of features for modeling the structures (in clutter class) is quite difficult as the structure classes associated

with the FP passed by the first rejector are mostly unknown. The type of structures is different in each dataset: in PDS-1, the FP are typically due to image noise, camera noise, small laser marks, exudate clusters. In PDS-2, the FP are mainly due to blurred image regions and noise. The CRIAS dataset on the other hand contains images with high number of pathologies which generate FP candidates from the following structures: small and big hemorrhages; region between two bright regions (usually inside a hard exudate cluster). This results in a high number of clutter classes. Dependency on the underlying dataset is clearly seen in the performance of RJ_2 (refer Table 4.4). The rejection rate obtained by second stage is 40%, 22%, 21% on PDS-1, PDS-2, CRIAS respectively, while the sensitivity remains high at approximately 98%. The rejection performance degrades by 50% for the two datasets which are more challenging. We can conclude that the appearance based features used in this stage are not robust to scale variations found in the structures.

Nevertheless, the high sensitivity signals a scope for further investigation towards better understanding and modeling of clutter classes. Experiments on large number of images collected from typical screening programs can give some useful insights to address this.

The technique of confidence measure assignment to the detected MA produces an outcome matched to the perceptual capabilities of medical experts. The associated confidence measure of MAs gives a means to classify them into perceptually defined categories such as subtle, general and well-defined/ obvious MA. This enables us to quantitatively analyze the behavior of the detector in different MA categories . The added benefit is this technique enables obtaining a FROC type of curve in a single run, unlike the parameter controlled multiple runs required in earlier approaches. This makes it possible to test the method's behavior on an unseen data in a single run.

The experimental evaluation offers some new insights about challenges for MA detection method development. Our experiments on different datasets were carried out without any tuning. This helps to predict performance on unseen data.

(i)The best performance of our method on PDS-1 reveals a few insights: Due to a fixed imaging protocol used, image size and resolution are fixed and intra-image variation is low. Consequently, learning by the proposed method is better in PDS-1 compared to other datasets. More generally, this indicates that performance of automated screening solution is better when a fixed protocol is used to acquire images.

(ii) The performance against CRIAS dataset (in Fig. 4.4) shows that at lower fppi, almost equal sensitivities against both observers, indicating consensus on obvious MAs. Whereas, there is a good amount of disagreement between experts in detecting general and subtle MAs. Thus, for an ideal evaluation of the method, ground truth should be collected from multiple experts and the evaluation technique should account for consensus level among experts.

A recent survey [Winder 09] indicated that creation of a large training set is an important step to ensure that evaluation is close to a real screening scenario. This study has also concurred that the development of successful screening solution for DR screening would be greatly facilitated by the adoption of a standard format for evaluating detection methods.

Chapter 5

Exudate segmentation

Exudates are a class of lipid retinal lesions visible through optical fundus imaging, and indicative of diabetic retinopathy. Unlike microaneurysms which are small by nature, exudates are of variable size and appearance. The need is thus not detection but segmentation, or demarcating exudate clusters from background. To illustrate the data analysis-driven approach described in Chapter 1, we devise a clustering-based method to segment exudates, using multi-space clustering, and colorspace features. This chapter gives the design and validation of the method on a set of 89 images from a publicly available dataset.

5.1 Introduction

Exudates are a class of lipid lesions visible in optical retinal images, which are clinical signs of DR. Two manifestations of exudates are known: hard exudates, that appear as bright yellow regions, and soft exudates or cotton-wool spots, which have fuzzy appearance. Automatic detection of exudates is of interest as it can assist ophthalmologists in DR diagnosis and early treatment.

The common approaches to lesion-level exudate detection follow a bottom-up strategy [Zhang 05], beginning with pixel classification, followed by region-level classification. Color values are used in pixel classification, since exudate pixels exhibit a limited range of color. Region-level classification has been attempted with features like edge-strength [Zhang 05], mean intensity within the region[Osareh 01, Garca 07], and contrast features[Niemeijer 07]. The optic disk is a structure with similar color characteristics as exudates, imaged in the central views of the retina. Optic disk has been distinguished by using entropy features[Sopharak 09], or using dedicated methods like active contours[Kande 08].

Existing work use supervised classification like k-nearest neighbor[Niemeijer 07], neural networks or SVM [Zhang 05]. In these methods, color normalization is performed as a common step in order to reduce the variability within retinal images, occuring due to imaging conditions, pigmentation, and presence of other pathology. Color is a prominent characteristic of exudates, and the performance of existing approaches rely on the ability of the normalization technique to handle the variability effectively.

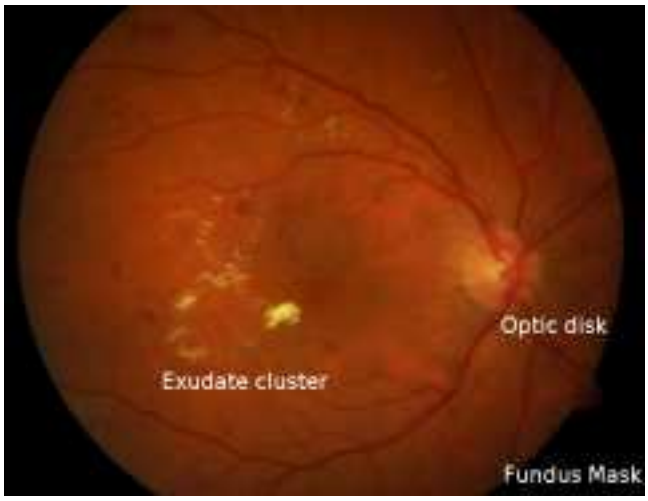
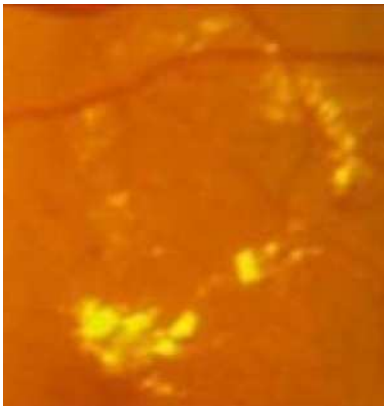
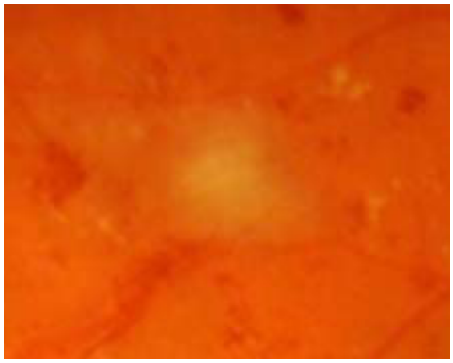


Figure 5.1 Retinal image indicating an exudate cluster, optic disk and the fundus mask



(a) Hard Exudate cluster in a retinal CFI sub-image



(b) A Soft exudate

Segmentation may also be performed in an unsupervised manner. Low-level segmentation has been performed by clustering using multispectral images and uniform-sized neighborhoods [Amadasun 88]. To achieve segmentation using clustering, features are computed at each pixel (or its neighborhood), thereby yielding data points in feature space. The clustering algorithm then assigns labels to each data point by optimizing over a cost function [Zhang 05, Osareh 01, Kande 08]. Segments are contiguous regions of pixels receiving the same label.

Two factors play a role in the clustering method: the feature space used, and the distance metric defined upon that space. The cost function uses the distance metric to find the proximity of cluster prototypes to each pixel, and assigns labels based on optimal values of the cost function.

Multi-space clustering is a technique which uses multiple feature spaces, with potentially each feature space using a different clustering algorithm and distance metric[Bickel 04]. This technique has shown improvement in performance compared to single-space clustering [Bickel 04, Pensa 08]. The

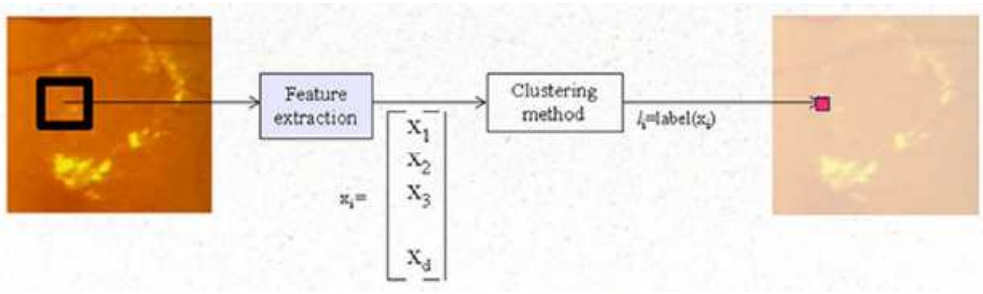


Figure 5.2 Traditional segmentation by clustering

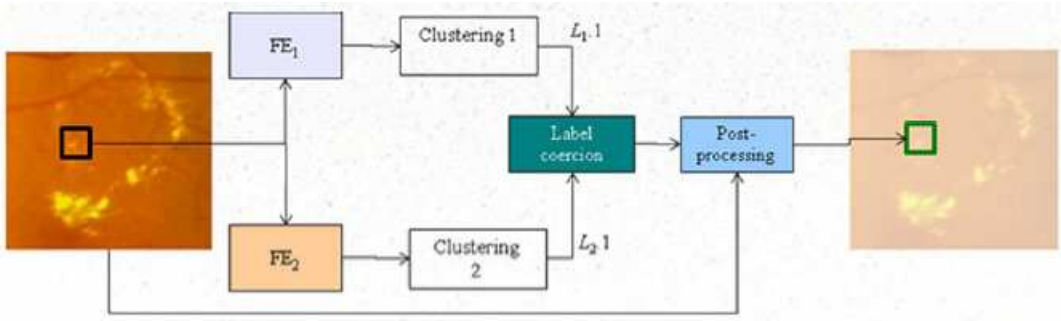


Figure 5.3 Flow diagram of Segmentation by multi-space clustering

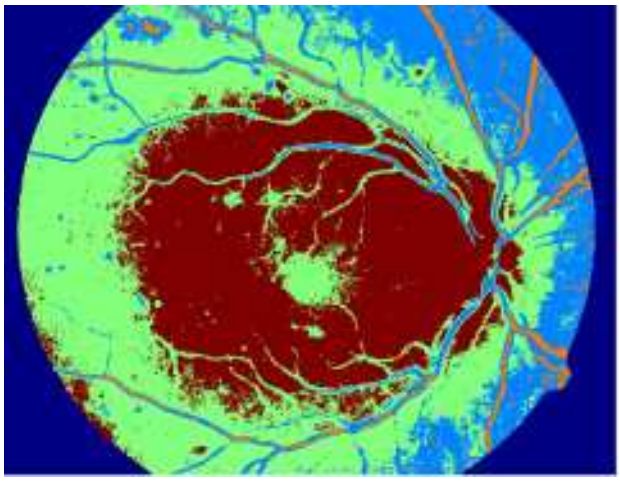
improvement is achieved by coercing the outcomes of the individual clusterings in a constrained fashion.

In this work, we propose a multi-space clustering approach to exudate segmentation, which does not use color normalization or preprocessing. The proposed method uses colorspace features constituting two feature spaces. Clustering is performed individually in each feature space, and the obtained labels are combined in a special manner to yield exudate segments.

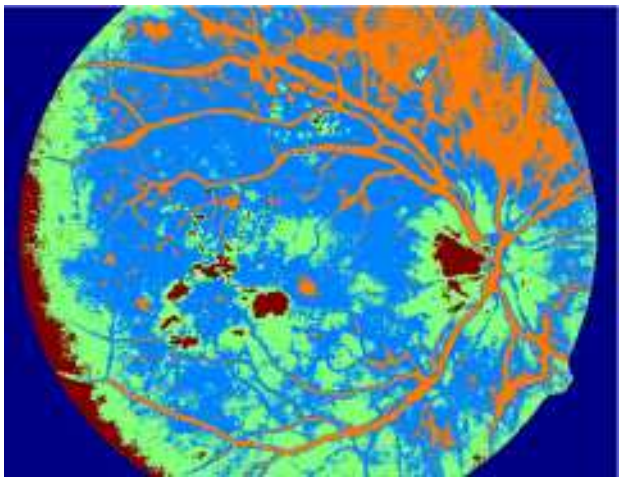
5.2 Proposed method

The proposed method is a bottom-up approach consisting of the following steps:

1. suppressing the fundus mask
2. obtaining pixel values in multiple color spaces
3. constructing the feature spaces to perform clustering
4. clustering to obtain labels
5. combining the clustering outcomes, to get candidate regions
6. suppressing false candidates



(a) Clustering in $\mathbf{f_1}$



(b) Clustering in $\mathbf{f_2}$

Figure 5.4 Clustering based segmentation in two feature spaces

The fundus mask is the dark peripheral part of the RGB retinal image, which does not contain informative pixels (fundus pixels). The fundus mask can be excluded by thresholding the brightness (V of HSV space). The next step computes color transformations into four color spaces, for each of the fundus pixels. Conditionally independent feature spaces are constructed from the colorspace values. We have considered the following colorspaces: RGB, CIE L*u*v*, HSV, HSI, and constructed two feature spaces: $\mathbf{f_1} : (H, S, V, I)$, and $\mathbf{f_2} : (R, G, L*, u*, v*)$. It can be seen that conditional independence is ensured among the two feature spaces. This is essential [Bickel 04] for the multi-space clustering framework.

We use k-means clustering, $1-$cross-correlation being the distance metric, treating data points as sequences. The values forming each such sequence are normalized to have zero mean and unit standard deviation. The clusters are initialized by performing a preliminary clustering with random 10% sub-

sampling of the data, and k centroids at random. We set $k = 4$, thereby partitioning the image into 4 segments.

Clustering in $\mathbf{f_1}$ results in segments corresponding to the following structures in the retinal image:

1. Bright lesions and bright background
2. Vessels, dark background, macula region
3. General retinal background
4. Peripheral region,

whereas clustering in $\mathbf{f_2}$ results in the following segmentation:

1. Optic disk, hard exudates, peripheral bright regions
2. Vessels, dark lesions, dark background
3. Regions surrounding the bright objects (enclosing regions of (1))
4. Other background pixels

Regions of $\mathbf{f_1}$ with label 1 (denote $L_1.1$) miss some minute, isolated exudates and some faint exudates, which are however picked up in $\mathbf{f_2}$ label 1 (denote $L_2.1$), at the cost of picking several periphery pixels. The choice of feature space has yielded this complementary nature to the clustering relevant to the region of interest. The labels, if combined appropriately, help to maximally identify the exudate regions and optic disk. We subsequently show a scheme devised to achieve this.

Clustering results in separation of the four clusters, from which we identify the cluster corresponding to $L_1.1$ and $L_2.1$ using the following observations:

1. exudates are bright lesions: $\max(I)$ value (of HSI) will be high in the exudate cluster.
2. exudates exhibit a yellowish color: $\max(R) - \max(G)$ in exudate cluster should have a low value.
3. cluster having minimum $\max(I)$ can be rejected as being $L_1.2$ or $L_2.2$. Similarly cluster having maximum $\max(R) - \max(G)$ can be rejected as non-exudate cluster.

This logic is summed up in Table 5.1. $L_1.1$ and $L_2.1$ are shown in brown in Figure 5.4

Having identified $L_1.1$ and $L_2.1$, their complementary nature is now used to extract the most likely exudate regions. For this we have devised the following scheme:

$L_1.1$ contains slightly over-segmented exudate regions, and several bright background pixels surrounding and including the optic disk. $L_2.1$ contains well-segmented exudates, minute exudates, and several peripheral pixels. The exudate regions can thus be extracted by finding all $L_2.1$ regions present in $L_1.1$, and the other $L_1.1$ regions not present $L_2.1$. Connected components analysis is done to enumerate the regions and find their presence in $L_1.1$ and $L_2.1$. The desired regions are then extracted, as shown in Figure 5.5.

Table 5.1 Identifying $L_1.1$ and $L_2.1$ clusters

$\max(I)$	$\max(R) - \max(G)$	Possible clusters
maximum	minimum	$\mathbf{L_1.1}$, $\mathbf{L_2.1}$
minimum	X	$L_1.2$, $L_2.2$, $L_1.4$
X	maximum	$L_1.2$, $L_1.3$, $L_2.2$
X	X	$L_1.3$, $L_2.4$, $L_2.3$

Parts of the optic disk and a few bright background regions near it now remain to be identified and suppressed. It can be seen that these superfluous regions are bounded by or cut across by blood vessels. Yet the contrast between the vessels and the candidate region is not prominent, leading to the regions enclosing some vessel segments. The optic disk is one such region, where the major vessels are incident.

We use band decorrelation[Gillespie 86] among the RGB bands in the candidate regions. This results in strong accentuation of the vessel contrast. The red component of the decorrelated result shows a very high value at blood vessels, whereas green component assumes high value at exudates and bright regions (see Figure 5.6(a)). As per this observation, optic disk regions and regions with vessel crossings assume higher mean decorrelated red value. Thus we find the difference between the mean of red value before and after decorrelation, and suppress candidates yielding a negative value of this difference.

5.3 Evaluation

5.3.1 Dataset

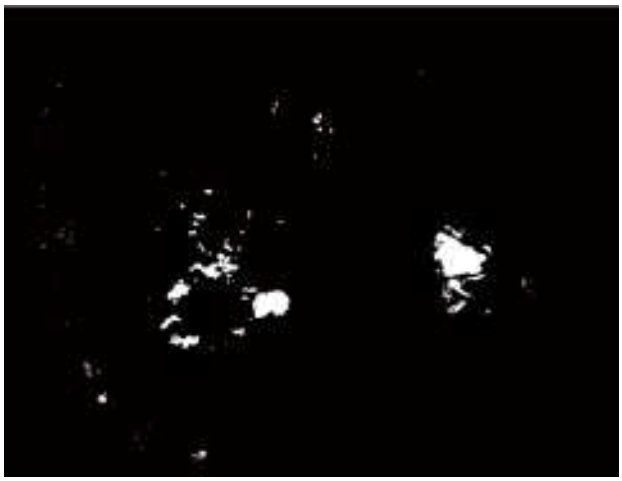
Our method was evaluated against the publicly available DIARETDB1 dataset, consisting of 89 images, of which 38 images contain hard exudates, and 20 contain soft exudates. All images are of same size (1500x1152) and captured using 50 degree field-of-view digital fundus camera[Kauppi 07b]. The accuracy of the method is reported in terms of sensitivity and positive predictive value [G.Altman 94] (PPV).

Since the ground truth is available in terms of polygonal regions, and the polygons are not an exact annotation of the lesions, to calculate sensitivity we thresholded the green band enclosed by each annotated polygon at highest consensus level (thresholding was done using Otsu method) and used the resulting pixels as the true exudate regions.

Positive pixels are identified as those which coincide with pixels in the true regions. PPV is found as the ratio of number of positive candidate pixels, to the total number of candidate pixels. Sensitivity and PPV are computed across the dataset, as against averaging a per-image evaluation (as reported in [Garca 07]). Our method shows a sensitivity of 71.96% and PPV of 87%.

Applying a region overlap accuracy metric, where each ground truth region is considered as segmented if positive candidates coincide with at least 50% of the ground truth region, our approach has

(a) $L_1.1$ regions not present in $L_2.1$



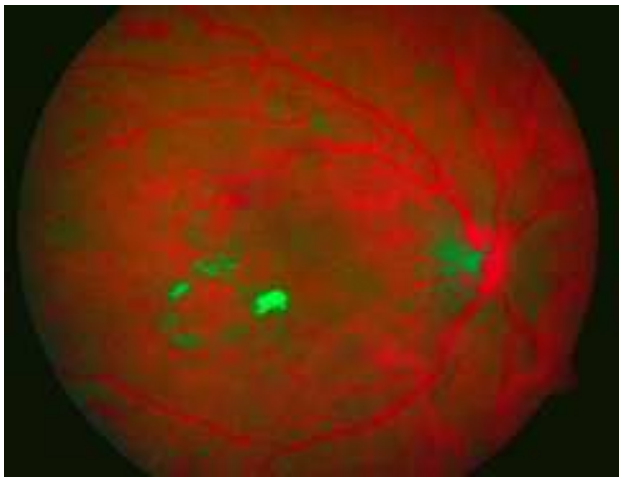
(b) $L_2.1$ regions present in $L_1.1$

Figure 5.5 Candidate regions identified by coercing the clusterings

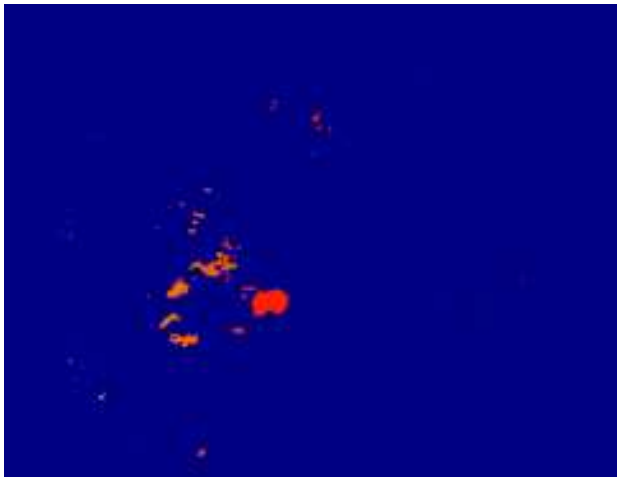
an accuracy(recall) of 89.7%. This can be compared with the supervised methods of [Sopharak 09], which reports a recall of 87.28% on a dataset of 40 images from a local hospital, and [Garca 07], reporting recall of 84.4% at PPV of 62.7% on a set of 50 images. However the performance is lower than [Osareh 01], reporting 92% recall on a set of 42 images. The criterion of at-least-50% spatial overlap may be justified considering that several exudates are small, irregular-shaped and appear in clusters.

5.4 Discussion

The use of correlation as distance metric, and use of well-selected feature spaces has compensated for the commonly performed image pre-processing step. To perform clustering, other work in literature have used fuzzy c-means, but have not capitalized on the fuzzy membership values. Moreover, time taken to



(a) RGB band-decorrelated image



(b) Positive regions after false candidate suppression

Figure 5.6

process a single image is reported to be considerably high (for example, 18 minutes [Sopharak 09] for 752x500 image in Matlab platform) in fuzzy c-means method.

We perform k-means clustering at pixel level using only color information at each pixel, owing to the bottom-up strategy. Processing using a pixel color list data structure enabled much faster clustering (less than 20 seconds on average. See Table 5.2). This comfortably permits clustering on two feature spaces.

For evaluation, since accurate lesion demarcation was not available in the dataset, we have used a threshold within the expert-annotated polygons. Though this step has helped to capture the lesion boundary in most cases, bright background pixels surrounding some small lesions have also been captured. This has affected the overall sensitivity value.

Table 5.2 Average running time for a single image: 1500x1152, in Matlab platform

Stage	Avg. time taken(sec)
Fundus mask suppression	0.04
Construction of color list	10.51
Color transformations	7.96
Clustering in $\mathbf{f_1}$	12.51
Clustering in $\mathbf{f_2}$	19.48
$L_1.1$ and $L_2.1$ identification	1.282
Coercion of labels	0.302
Suppression of false regions	11.802
Total	63.886

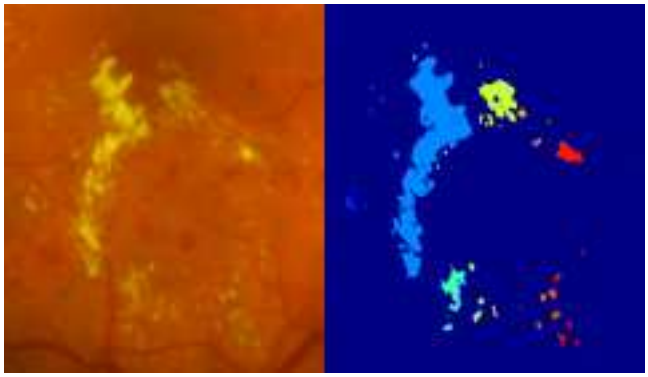


Figure 5.7 Sub-image indicating segmented exudates

The high value of PPV indicates that false-positives are few in our approach. Some bright imaging artifacts appearing as small blobs, and laser marks, which appear away from the macula in central views, are two observable false alarms. In few images, exudates appear close to the optic disk, leading to forming a single region enclosing both a true lesion and the optic disk. In this case only parts of the optic disk are suppressed by the method.

Overall, the results obtained indicate that there is good potential for multi-space clustering to be applied as a segmentation technique. Our method is significantly faster than the state of the art, achieves comparable accuracy, and segments are visually well-correlated with the lesion.

Chapter 6

Conclusions

This thesis described novel methods for automatic analysis of retinal images with the goal of automated diabetic retinopathy screening. In automated screening a computer system analyzes retinal images before an ophthalmologist does, and only the images that are suspect for the presence of diabetic retinopathy are presented to the ophthalmologist. To achieve this objective, the thesis presented methods for detection of diabetic retinopathy-indicative lesions in retinal images.

Chapter 1 formulated detection as a decision between two states "target-present" and "target-absent". Observations are made consisting of measurements, and the observations help to make decisions by comparison with prototypical patterns in the observation space or feature space. A detector hence becomes a set of decision functions, where the task of each decision function δ_i is a transformation from the observation space to *its* decision space S_i. The observation x is transformed into decision space S_i, and the power of decision comes from the existence of numerical ordering in S_i.

In theory achieving such ordering helps to formulate the detection problem as an optimization in the decision space. Such structural restriction as in S_i is however not stated for the observation space. As in the task at hand, for most real world decision problems it is necessary to identify what measurements qualify to be considered as an observation amenable to statistical analysis and inference. This is the transport between the semantics of the real world and the observations in the mathematical/ statistical world.

Science is built up of facts, as a house is built up of stones. But an accumulation of facts is no more a science than a heap of stones is a house **– Henri Poincare**.

It is only through empirical studies, experimentation and experience in the domain, that good observations can be made. Assuming a set of features to be available for the application, a classifier-approach would try to search for a separating hyperplane in the observation space, projecting upon which an ordering would arise (the projection maps from the observation space to the decision space).

Chapter 3 considered a different perspective. Classifier design builds a single projection function g for a variety of samples. In the case of a multiple-discriminant or perceptron learning there are multiple

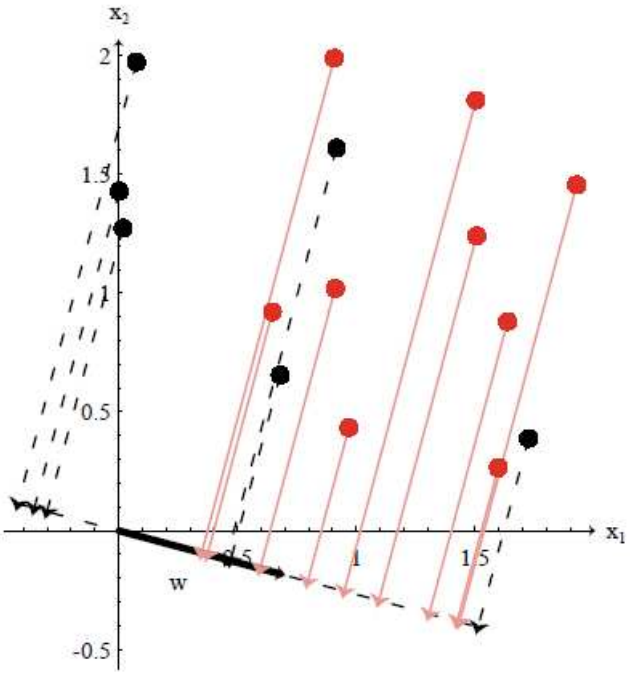


Figure 6.1 An illustrative projection in a Fisher linear discriminant. Image courtesy of Richard O. Duda, Peter E.Hart and David G.Stork, Pattern Classification, Second Edition, Wiley 2001

projections but in the same observation space. By its very formulation statistical modeling is reductive, necessarily missing part of the complexity of the real world to give a simplified representation permitting decision making. There is thus a possibility that a part of the training samples may not conform to their generative hypothesis. Essentially it can be claimed that Chapter 3 attempted to circumvent this by considering **feature sub-spaces**, instead of performing the projection in a unified feature space.

The subspaces may be constructed manually, or by applying an automatic feature selection technique combined with a clustering algorithm, where each cluster denotes a representative of a possible expression of the target to be detected. A problem certain to arise in this direction is that the number of expressions exhibited by the target must be enumerated and annotated, in order to verify the outcome. But for the detection task the annotation available is only indicative of the target, and the different expressions of the target, such as “faint, distinct, uneven, off-shape” and so on, are not straight-forward to obtain, and are subjective. Further this strategy might not help to make generalizations.

We thus shifted our focus from the target object to the non-target, in the fashion of the knowledge-discovery technique known as *iterative denoising* [Giles 08]. The subspace idea along with the focus on non-MA culminated in the successive-rejection strategy for MA detection presented in Chapters 2 and 3.

The presented approach exploits the high occurrence probability of clutter structures and derives a novel, successive rejection-based method. A system developed using the proposed approach has been evaluated on three datasets, and a comparison of performance with other approaches has been presented.

To the best of our knowledge, this is the first lesion-level MA detection method to be evaluated against an extensive set of images from multiple datasets.

The practical implementation of the system opened up new challenges, foremost was the need to identify and model various frequently occurring FP structures (among the candidates), other than the ones we have identified in this work, for better rejection. The modular design of the proposed approach offers flexibility to include additional rejection stages independently, without affecting the functionality of other stages. The performance obtained indicated that a more sensitive candidate selection component is of need, to get desirable performance, since our experiments show that the candidate selection determines an upper bound on the maximum sensitivity which can be achieved. For instance, the wavelet-domain template matching approach of [Quellec 08] could replace the morphological approach we applied, towards this goal.

Some questions that emerge from this work are:

1. Should the training set strive to include the entire gamut of expressions exhibited by the characteristic patterns (without enumeration)? This would be an effort towards generalization, but would it penalize the performance of the system by posing greater variability ? What are the ways by which these opposing objectives can be reconciled?

2. Given a detector, having only its parameters to control, how can it be found if there exists a parameter value at which the detector performance is satisfactory?

Our results indicate that for the inductive learning approach espoused in this thesis for MA detection, greater variations in the training set only result in reduced performance. This is evidenced in the gap between the PDS-1 and CRIAS performance. This is concordant with the assumption expressed in Chapter 1, that the set of characteristic patterns be representative and "good".

The assumption is also endorsed by the "Compactness hypothesis" of training data –In order to derive, from a training set, a classifier that is valid for new (unseen) objects, the representation should fulfill the condition that *representations of similar real world objects should have to be similar as well* [1].

By training with greater intra-class variations, the classifier arises as a model with low statistical bias, but high variance (a model with an inclination to flexibility/generalizability). Such a model permits greater false alarms leading to reduced performance, especially in medical applications which are required to have very low false alarms, more so in a screening scenario.

A classifier trained with fewer intra-class variability could lead to high performance, as seen in PDS-1, but may lead to a rigid model (low variance, but high bias) with the possibility of having overfitted the training data.

These issues are traded away in the unsupervised approach detailed in Chapter 4. Here clustering is performed as a pixel labeling step, with an intended partitioning of the image into specific types of

[1] from the Handbook of pattern recognition and computer vision, CH Chen and PSP Wang (eds)

partitions. Connected components form the potential regions of exudates, which are analyzed further based on RGB band decorrelation information, to reject some regions. This approach uses two feature spaces where pixel labeling is performed independently and the results are merged in a manner driven by heuristics and a knowledge of the outcomes. The approach is unsupervised and relies on domain knowledge and assumptions on the expressions exhibited by exudates in CFI.

Future work:

The MA detection algorithm was also evaluated against a dataset collected from L.V.Prasad Eye Hospital, part of which is called CRIAS (colour retinal image analysis set). The dataset contains CFI of 288 patients, containing at most 2 images per patient (left and right eye). The dataset also contains fluorescin angiogram images (FA) of the patients. It is thus possible to attempt automated multimodal registration of CFI and FA, and study the turnover of MA in the two modalities. It is well-known that MA appear as easily detectable bright spots in the FA. But FAs show about twice as many MAs as seen in CFI. This is because the retina is multi-layered, and CFIs capture reflected light from only the first few layers. MAs located in further layers show up in the FA but not in CFI. By obtaining image registration between FA and CFI, a number of insights might arise regarding better ways to obtain unequivocal ("gold-standard") ground truth, and strategies for developing better MA candidate selection algorithms for CFI.

Regarding the achievement of desired performance, detection theory only gives an objective for optimization. The decision capacity of the detector may be ascribed to a likelihood ratio test on the observations. The detector model is thus data driven, highly dependent on the training data and the measurements constituting observations for achieving performance. This being the case, the theory directs the algorithm designer to experimentation at various parameter settings as the only means to verify the power of the detector. By relooking at the formulation of detection it is possible to evolve an alternative way to setup the detector objective function.

The room is but a resource, I hold its lock and key
Now they call it my room, but it is what holds me.

Related Publications

1. Keerthi Ram, Gopal Joshi and Jayanthi Sivaswamy, **A successive clutter-rejection based approach for early detection of diabetic retinopathy**, *IEEE Transactions on Biomedical Engineering*, (submitted May 2010)

2. Keerthi Ram and Jayanthi Sivaswamy, **Multi-space clustering for segmentation of exudates in retinal color photographs**, *in Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, September 2009.

Bibliography

[Abramoff 07] Michael D. Abramoff, Bram van Ginneken & Meindert Niemeijer. *Retinopathy Online Challenge*, December 2007.

[Abramoff 08] M.D. Abramoff, M. Niemeijer, M.S. Suttorp-Schulten, M.A. Viergever, S.R. Russell & B. van Ginneken. *Evaluation of a system for automatic detection of diabetic retinopathy from color fundus photographs in a large population of patients with diabetes*. Diabetes Care, vol. 31, no. 2, pages 193–198, 2008.

[Amadasun 88] M. Amadasun & R. A. King. *Low-level segmentation of multispectral images via agglomerative clustering of uniform neighbourhoods*. Pattern Recognition, vol. 21, no. 3, pages 261 – 268, 1988.

[Asada 86] H. Asada & M. Brady. *The Curvature Primal Sketch*. PAMI, vol. 8, no. 1, pages 2–14, 1986.

[Autio 05] I. Autio, J.C. Borra, I. Immonen, P. Jalli & E. Ukkonen. *A voting margin approach for the detection of retinal microaneurysms*. Proc. Visualization, Imaging and Image Processing, 2005.

[A.Yuille 92] A.Yuille, P.Hallinan & D.Cohen. *Feature extraction from faces using deformable templates*. IJCV, vol. 8, no. 2, pages 99–111, 1992.

[Baker 96] Simon Baker & Shree K. Nayar. *Pattern Rejection*. In Proc.CVPR, pages 544–549, 1996.

[Barnett 94] V. Barnett & T.Louis. Outliers in statistical data. John Wiley & Sons, 1994.

[Bhalerao 08] A. Bhalerao, A. Patanaik, S. Anand & P. Saravanan. *Robust Detection of Microaneurysms for Sight Threatening Retinopathy Screening*. In ICVGIP08, pages 520–527, 2008.

[Bickel 04] Steffen Bickel & Tobial Scheffer. *Multi-view Clustering*. In Proc. IEEE International Conference on Data Mining (ICDM), 2004.

[Boiman 08] Oren Boiman, Eli Shechtman & Michal Irani. *In Defense of Nearest-Neighbor Based Image Classification*. In Proc.CVPR, 2008.

[Brady 85] M Brady, J Ponce, A Yuille & H Asada. Describing surfaces. 1985.

[Burges 98] Christopher J. C. Burges. *A tutorial on support vector machines for pattern recognition*. Data Mining and Knowledge Discovery, vol. 2, pages 121–167, 1998.

[Canny 86] J. F. Canny. *A computational approach to edge detection*. IEEE Trans. PAMI., vol. 8, pages 679–697, 1986.

[Cortes 95] Corinna Cortes & Vladimir Vapnik. *Support-Vector Networks*. Machine Learning, vol. 20, no. 3, pages 273–297, 1995.

[C.Papageorgiou 98] C.Papageorgiou, M.Oren & T.Poggio. *A general framework for object detection*. In Proc.ICCV, pages 555–562, 1998.

[Cree 97] M. Cree, J. Olson, K. McHardy, P. Sharp & J. Forrester. *A fully automated comparative microaneurysm digital detection system*. Eye, vol. 11, pages 622–628, 1997.

[Dalal 05] Navneet Dalal & Bill Triggs. *Histograms of Oriented Gradients for Human Detection*. In Proc. CVPR, volume 1, pages 886–893, 2005.

[Das 06] Taraprasad Das & Alka Rani. Diabetic eye diseases, chapter Foundations in Vitreo-Retinal Disease. 2006.

[D.Marr 80] D.Marr & E.Hildreth. *Theory of edge detection*. In Proceedings of the Royal Society of London, B, volume 207, pages 187–217, 1980.

[Duda 00] Richard O. Duda, Peter E. Hart & David G. Stork. Pattern classification (2nd edition). Wiley-Interscience, November 2000.

[Fleming 06] A. D. Fleming, S. Philip, K. A. Goatman, J. A. Olson & P. F. Sharp. *Automated microaneurysm detection using local contrast normalization and local vessel detection*. Proc. Norwegian Signal Process. Symp., vol. 25, no. 9, pages 1223–1232, 2006.

[Frame 98] A. Frame, P. Undrill, M. Cree, J. Olson, K. McHardy, P. Sharp & J. Forrester. *A comparison of computer based classification methods applied to the detection of microaneurysms in ophthalmic fluorescein angiograms*. Comput. Biol. Med., vol. 28, pages 225–238, 1998.

[Frangi 98] Alejandro F. Frangi, Wiro J. Niessen, Koen L. Vincken & Max A. Viergever. *Multiscale Vessel Enhancement Filtering*. Lecture Notes in Computer Science, vol. 1496, 1998.

[G.Altman 94] Douglas G.Altman & J Martin Bland. *Diagnostic tests 2: predictive values*. BMJ, vol. 309, no. 6947, page 102, July 1994.

[Garca 07] Mara Garca, Roberto Hornero, Clara I. Snchez, Mara I. Lpez & Ana Dez. *Feature Extraction and Selection for the Automatic Detection of Hard Exudates in Retinal Images*. In Proc. International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS), pages 4969–4972, 2007.

[Garg 07] S Garg, J Sivaswamy & S Chandra. *Unsupervised Curvature-Based Retinal Vessel Segmentation*. pages 344–347, 2007.

[Giles 08] Kendall E. Giles, Michael W. Trosset, David J. Marchette & Carey E. Priebe. *Iterative denoising*. computational statistics, vol. 23, no. 4, pages 497–517, 2008.

[Gillespie 86] Alan R Gillespie, Anne B Kahle & Richard E Walker. *Color enhancement of highly correlated images. I. Decorrelation and HSI contrast stretches*. Remote Sens. Environ., vol. 20, no. 3, pages 209–235, 1986.

[Haralick 83] R Haralick, L Watson & T Laffey. *The Topographic Primal Sketch*. volume 2, 1983.

[Harris 88] C. Harris & M. Stephens. *A combined corner and edge detector*. In Proc. 4th Alvey Vision Conference, pages 147–151, 1988.

[Hodge 04a] Victoria Hodge & Jim Austin. *A Survey of Outlier Detection Methodologies*. Artif. Intell. Rev., vol. 22, no. 2, pages 85–126, 2004.

[Hodge 04b] Victoria Hodge & Jim Austin. *A Survey of Outlier Detection Methodologies*. Artif. Intell. Rev., vol. 22, no. 2, pages 85–126, 2004.

[Hoover] A. Hoover & M. Goldbaum. *STARE dataset*.

[Huang 05] Ke Huang & Michelle Yan. *A local adaptive algorithm for microaneurysms detection in digital fundus images*. In Proc. CVBIA, pages 103–113, 2005.

[Huang 07] K. Huang, M. Yan & S. Aviyente. *Edge Directed Inference for Microaneurysms Detection in Digital Fundus Images*. Proc. SPIE, vol. 6512, 2007.

[Kande 08] Giri Babu Kande, P. Venkata Subbaiah & T. Satya Savithri. *Segmentation of Exudates and Optic Disk in Retinal Images*. In Proceedings of Indian Conference on Computer Vision, Graphics and Image Processing, 2008.

[Kande 09] G. B. Kande, T. S. Savithri, P.V. Subbaiah & M.R.N. Tagore. *Detection of red lesions in digital fundus images*. Proc. Int. Symp. on Biomedical Imaging (ISBI)., 2009.

[Kauppi 07a] Tomi Kauppi. *DIARETDB1 - standard diabetic retinopathy database - calibration level 1*, April 2007.

[Kauppi 07b] Tomi Kauppi, Valentina Kalesnykiene, Joni-Kristian Kamarainen, Lasse Lensu, Iiris Sorri, Heikki Kalviainen & Juhani Pietila. *DIARETDB1 diabetic retinopathy database and evaluation protocol*. In Proc. 11th Conf. on Medical Image Understanding and Analysis, 2007.

[kay Sung 98] Kah kay Sung & Tomaso Poggio. *Example-based learning for view-based human face detection*. IEEE Trans. PAMI, vol. 20, pages 39–51, 1998.

[Klein] Jean-Claude Klein. *Messidor Dataset*.

[Lay 83] B. Lay. *Analyse automatique des images angiofluorographiques au cours de la retinopathie diabetique*. Ph.D. Thesis, Centre of Mathematical Morphology, Paris School of Mines, June, 1983.

[Lienhart 02] R. Lienhart & J. Maydt. *An extended set of Haar-like features for rapid object detection*. In Proc.ICIP, volume 1, pages I–900–I–903 vol.1, 2002.

[M.Betke 95] M.Betke & N.Markis. *Fast object recognition in noisy images using simulated annealing*. In Proc. ICCV, pages 523–20, 1995.

[Mendonca 99] A. M. Mendonca, A. J. Campilho & J. M. Nunes. *Automatic segmentation of microaneurysms in retinal angiograms of diabetic patient*. Proc. Int. Confe. Image Anal. Process., pages 728–733, 1999.

[Mizutani 09] A. Mizutani, C. Muramatsu, Y. Hatanaka, S. Suemori, T. Hara & H. Fujita. *Automated microaneurysm detection method based on double ring filter in retinal fundus images*. Proc. SPIE Medical Imaging 2009: Computer-Aided Diagnosis, vol. 7260, page 72601L, 2009.

[Niemeijer 05] M. Niemeijer, B. van Ginneken, J.Staal, M.S.A.Suttorp-Schulten & M.D.Abramoff. *Automatic detection of red lesions in digital color fundus photographs*. IEEE Trans. Medical Imaging, vol. 24, no. 5, pages 584–592, 2005.

[Niemeijer 07] Meindert Niemeijer, Bram van Ginneken, Stephen R. Russell, Maria S. A. Suttorp-Schulten & Michael D. Abrmoff. *Automated Detection and Differentiation of Drusen, Exudates, and Cotton-Wool Spots in Digital Color Fundus Photographs for Diabetic Retinopathy Diagnosis*. Investigative Ophthalmology and Visual Science, vol. 48, no. 5, pages 2260–2267, May 2007.

[Niemeijer 09] M. Niemeijer, B. Ginneken, M. J. Cree, A. Mizutani, G. Quellec, C. I. Sanchez, B. Zhang, R. Hornero, M. Lamard, C. Muramatsu, X. Wu, G. Cazuguel, J. You, A. Mayo, Q. Li, Y. Hatanaka, B. Cochener, C. Roux, F. Karray, M. Garca, H. Fujita & M. D. Abramo. *Retinopathy Online Challenge: Automatic Detection of Microaneurysms in Digital Color Fundus Photographs*. Technical Report, 2009.

[Oien 95] G. E. Oien & P. Osnes. *Diabetic retinopathy: Automatic detection of early symptoms from retinal images*. Proc. Norwegian Signal Process. Symp., pages 135–140, 1995.

[Osareh 01] Alireza Osareh, Majid Mirmehdi, Barry Thomas & Richard Markham. Automatic recognition of exudative maculopathy using fuzzy c-means clustering and neural networks, pages 49–52. 2001.

[Pensa 08] R.G. Pensa & Mirco Nanni. *A Constraint-Based Approach for Multispace Clustering*. In Proceedings of LeGo-08 Workshop (From Local Patterns to Global Models), ECML/PKDD, 2008.

[Ponce 06] Jean Ponce, Martial Herbert, Cordelia Schmid & Andrew Zisserman, eds. Towards category-level object recognition. LNCS, 2006.

[Quellec 08] G. Quellec, M. Lamard, P.M. Josselin, G. Cazuguel, B. Cochener & C. Roux. *Optimal Wavelet Transform for the Detection of Microaneurysms in Retina Photographs*. IEEE Trans. Medical Imaging, vol. 27, no. 9, pages 1230–1241, September 2008.

[Ram 09] Keerthi Ram, Yogesh Babu & Jayanthi Sivaswamy. *Curvature orientation histograms for detection and matching of vascular landmarks in retinal images*. In Proc. SPIE. Medical Imaging, 2009.

[Sanchez 08] C. I. Sanchez, R. Hornero, A. Mayo & M. Garcia. *Mixture Model-based Clustering and Logistic Regression for Automatic Detection of Microaneurysms in Retinal Images*. Proc. SPIE, vol. 7260, 2008.

[Singh 08] Jeetinder Singh & Jayanthi Sivaswamy. *Appearance-based object detection in colour retinal images*. In Proc. ICIP, 2008.

[Sinthanayothin 02] C. Sinthanayothin, J. F. Boyce, T. H. Williamson, H. L. Cook, E. Mensah, S. Lal & D. Usher. *Automated detection of diabetic retinopathy on digital fundus images*. Diabetic Med., vol. 19, no. 2, pages 105–112, 2002.

[Sopharak 09] Akara Sopharak, Bunyarit Uyyanonvara & Sarah Barman. *Automatic Exudate Detection from Non-dilated Diabetic Retinopathy Retinal Images Using Fuzzy C-means Clustering*. Sensors, vol. 9, no. 3, pages 2148–2161, 2009.

[Spencer 91] T. Spencer, R. P. Phillips, P. Sharp & J. Forrester. *Automated detection and quantification of microaneurysms in fluorescein angiograms*. Graefe's Archive for Clinical and Experimental Ophthalmology, vol. 230, pages 36–41, 1991.

[Spencer 96] T. Spencer, J. Olson, K. McHardy, P. Sharp & J. Forrester. *An image processing strategy for the segmentation and quantification in fluorescein angiograms of the ocular fundus*. Comput. Biomed. Res., vol. 29, pages 284–302, 1996.

[Staal] J.J. Staal, M.D. Abramoff, M. Niemeijer, M.A. Viergever & B. van Ginneken. *DRIVE dataset*.

[Usher 04] D. Usher, M. Dumskyj, M. Himaga, T.H. Williamson, S. Nussey & J. Boyce. *Automated detection of diabetic retinopathy in digital retinal images: a tool for diabetic retinopathy screening.* Diabet. Med., vol. 21, pages 84–90, 2004.

[Viola 01] Paul Viola & Michael Jones. *Robust Real-time Object Detection*. IJCV, 2001.

[Walter 07] Thomas Walter, Pascale Massin, Ali Erginay, Richard Ordonez, Clotilde Jeulin & Jean-Claude Klein. *Automatic detection of microaneurysms in color fundus images*. Medical Image Analysis, vol. 11, no. 6, pages 555–566, 2007.

[Winder 09] R.J. Winder, P.J. Morrow, I.N. McRitchie, J.R. Bailie & P.M. Hart. *Algorithms for digital image processing in diabetic retinopathy*. Comput Med Imaging Graph., 2009.

[Yannuzzi 86] L. Yannuzzi, K. Rohrer & L. Tindel. *Fluorescein angiography complication survey*. Ophthalmology, vol. 93, pages 611–617, 1986.

[Zhang 05] Xiaohui Zhang & Opas Chutatape. *Top-Down and Bottom-Up Strategies in Lesion Detection of Background Diabetic Retinopathy*. In IEEE Computer Society Conference on Computer Vision and Pattern RecognitionCVPR, volume 2, pages 422–428, 2005.

[Zhang 09] B. Zhang, X. Wu, J. You, Q. Li & F. Karray. *Hierarchical detection of red lesions in retinal images by multiscale correlation filtering*. Proc. SPIE Medical Imaging 2009: Computer-Aided Diagnosis, vol. 7260, page 72601L, 2009.