Automatic analysis of malaria infected red blood cell digitized microscope images

Houari Abdallahi

A dissertation submitted in partial fulfilment

of the requirements for the degree of

Doctor of Philosophy

of

University College London

Department of Computer Science

University College London

Supervisor: Prof. Bernard F. Buxton

February 2016

Declaration of Authorship and Originality

I confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Department of Computer Science
University College London
Gower Street
London
WC1E 6BT
h.abdallahi@cs.ucl.ac.uk

Date: February 7, 2016

Houari Abdallahi

Abstract

Malaria is one of the three most serious diseases worldwide, affecting millions each year, mainly in the tropics where the most serious illnesses are caused by *Plasmodium falciparum*. This thesis is concerned with the automatic analysis of images of microscope slides of Giemsa stained thin-films of such malaria infected blood so as to segment red-blood cells (RBCs) from the background plasma, to accurately and reliably count the cells, identify those that were infected with a parasite, and thus to determine the degree of infection or parasitemia. Unsupervised techniques were used throughout owing to the difficulty of obtaining large quantities of training data annotated by experts, in particular for total RBC counts. The first two aims were met by optimisation of Fisher discriminants. For RBC segmentation, a well-known iterative thresholding method due originally to Otsu (1979) was used for scalar features such as the image intensity and a novel extension of the algorithm developed for multi-dimensional, colour data. Performance of the algorithms was evaluated and compared via *ROC* analysis and their convergence properties studied. Ways of characterising the variability of the image data and, if necessary of mitigating it, were discussed in theory. The size distribution of the objects segmented in this way indicated that optimisation of a Fisher discriminant could be further used for classifying objects as small artefacts, singlet RBCs, doublets, or triplets *etc.* of adjoining cells provided optimisation was via a global search. Application of constraints on the relationships between the sizes of singlet and multiplet RBCs led to a number of tests that enabled clusters of cells to be reliably identified and accurate total RBC counts to be made. Development of an application to make such counts could be very useful both in research laboratories and in improving treatment of malaria. Unfortunately, the very small number of pixels belonging to parasite infections mean that it is difficult to segment parasite objects and thus to identify infected RBCs and to determine the parasitemia. Preliminary attempts to do so by similar, unsupervised means using Fischer discriminants, even when applied in a hierarchical manner, though suggestive that it may ultimately be possible to develop such a system remain on the evidence currently available, inconclusive. Appendices give details of material from old texts no longer easily accessible.

Acknowledgements

First and foremost I want to thank my supervisor Professor Bernard Buxton. It has been an honour to be his last serving PhD student. I learnt a tremendous amount both professionally and personally for which I will always be grateful.

Thanks also to Dr Delmiro Fernandez-Reyes and the National Institute for Medical Research [NIMR]. The group has been a source of friendships as well as good advice and collaboration.

Last but by no means least, thanks go to my mum Yamina, dad Abdelkader and wife Ibtissam for their unconditional love and support.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The work described in this thesis arose from a collaboration with staff at the National Institute for Medical Research (NIMR) at Mill Hill in London where there have been several research groups studying various aspects of malarial infection. In humans, malaria may be caused by any one of five different species of parasitic protozoans that invade red-blood cells. Depending on the parasite species and on the age, health, well-being, clinical history of the patient, and the complications that may develop, malaria may have a range of effects from debilitating fevers to death. It is one of the world's most serious human diseases with an impact mainly in the poorer regions of the world in the tropics where the anophelene mosquito responsible for its transmission between victims proliferates. Since control of the mosquito by mass application of pesticides and other environmentally invasive procedures has become unacceptable malaria is a disease in urgent need of the development of improved clinical diagnosis and care and, more importantly, of rapid diagnostic tests and of a suitable vaccine. There is therefore much study in laboratories worldwide, including at the NIMR, of the fundamental host-parasite interactions. Though the long term aim remains the development of procedures that may be applied in the field in the detection, treatment and prevention of malarial infection, there is much to be gained from the development of processes that assist laboratory research.

It was the latter that provided the motivation for the work in this thesis; namely to develop image processing and computer vision algorithms that would relieve laboratory researchers of the time-consuming and tedious manual work required to determine the parasitemia of infected blood samples cultured in the laboratory. This manual inspection of thin-films of stained blood samples under a microscope and counting the relative numbers of infected and healthy red-blood cells (RBCs) is according to NIMR staff the gold-standard for determining the parasitemia or degree of infection [66, 177].

This is a much less ambitious aim than, for example, development of computer vision systems for the diagnosis of malaria from microscope slide images [176, 151], but one that

because it utilizes images taken under more controlled conditions and avoids some of the most difficult classification problems [175] is more achievable and, if successful, would have a more immediate impact as no lengthy medical approval is required. The determination of the degree of parsitemia of such laboratory blood samples is thus not only a useful objective it its own right but could also be instructive and provide much valuable experience and feed-back for the development of systems to be used in the field.

1.1 Problem statement

The basic requirement for development of a computer vision system for determining the parsitemia of laboratory samples is, as in the manual process, to count the relative numbers of infected and healthy RBCs. There are various ways in which this might be achieved [165] in an automatic system, but typically this requires discrimination of RBCs from the background plasma, white-blood cells, platelets and distracting artefacts, a decision as to whether each individual RBC is infected by a malaria parasite or not, and a processes for counting the number of each type of cell. This is the approach adopted in this thesis. It begins with – and is dependent on – detection and labelling of the pixels belonging to RBCs, irrespective of whether they are infected or healthy, in a segmentation process.

Segmentation is a classic problem in image processing and computer vision. It is discussed in most text books, both in image processing [87, 139, 68] and in computer vision [89, 183, 60] though there are often differences in approach. A wide variety of techniques are used in both kinds of applications ranging from hand crafted algorithms based on user evaluation or interaction, model-based methods, and machine learning.

Model-based approaches, including the construction of models of the object of interest, the imaging geometry and photometry [51, 100], models of the object's shape and appearance [30] and models of the effect of changing viewpoint [35, 206] are popular in computer vision research and applications. The application of machine learning methods has also become more popular recently both in image processing and in computer vision. This is a result of increased computer power, improved learning algorithms and improved statistical methods [14]. Of particular note is the application of capacity-controlled supervised machine learning algorithms such as those based on the support vector machine technology [137, 4]. Capacity control combats the need for the huge numbers of training examples that would otherwise be required owing to the very-high dimensionality of the pixel space of an image. Multi-resolution, pyramid and wavelet transform filtering methods are often employed to reduce the dimensionality and have the additional benefits of introducing the spatial coherence between neighbouring pixels in an

image and thereby making the statistics of the filtered images predictable [172, 70]. Reducing the resolution also helps combat the computational complexity of such supervised machine learning algorithms.

More recently, deep learning techniques which can automatically generate appropriate intermediate representations of image data, in particular for object recognition have been very effective [107]. Though such systems require less training data than older approaches would have even when capacity controlled, large training data-sets nevertheless seem to be required and it is often necessary to use specialised hardware (graphics accelerators) for its implementation.

Principal component analysis [92] is a very effective way of reducing the dimensionality of the image representation in particular when applied to specific types of object as by Cootes and Taylor [31] or, when incorporating hidden landmark point variables as in the flexible appearance methods, also developed by Cootes and Taylor [30]. The characteristics of an object or class of objects may then often be learnt from a few hundred or thousand examples, even when the object has internal degrees of freedom, as in the expression of a face [38].

In spite of these advances, an unavoidable disadvantage of such supervised learning approaches is the requirement for manual interaction to obtain training data. Even though an initial system may be constructed from a small training set to be used as a tool for assisting manual labelling of a large training set [31], acquiring training sets can be time consuming and laborious. Furthermore, if the training data is to be accurate, segmentation of the training images should be carried out by an expert. Their time is often a scarce resource and different experts may not always agree [193].

In machine learning terms in unsupervised methods labels are assigned according to a criterion that reflects the underlying characteristics of the problem. In image processing this often amounts to the choice of a predicate according to which segmentation is to be carried out and is most common in the clustering of pixels in a colour space. The choice of colour space involving a linear or non-linear transformation of the image data, clustering criterion, initialisation and the algorithm employed are all important in the design of such systems. Often unsupervised algorithms are based on statistical criteria which may be parametric, as in the application of Gaussian mixture models in colour image segmentation, or non-parametric when a statistical criterion may be used that reflects overall properties of the pixel data. Amongst the latter is the classical Fisher discriminant or linear discriminant analysis (LDA) according to which one seeks a labelling or classification of image pixels that maximises the ratio of a between-class variance to a within-class variance [54, 46]. When applied to a one-dimensional feature space,

such as pixel intensity, this leads to the Otsu algorithm whose implementation [89] seems rather better known than its origin [131].

In distinction to the approaches adopted by others, in particular at the University of Westminster [176, 175, 177, 154, 153, 155, 156, 151, 165], we have been exploring the application of such methods to address the segmentation of RBCs in images taken from thin-film slides of stained blood samples and detection of which cells are infected. In so doing we have developed and studied in detail a novel segmentation algorithm based on a combination of application of a multi-dimensional Fisher discriminant and the Otsu algorithm. This Fisher-Otsu algorithm seems especially well suited to the segmentation of pixels belonging to RBCs and, given such a segmentation, can be very effective when employed to count RBCs, especially when employed with a range of internal checks on the algorithm's performance. Unfortunately, in the preliminary work described here, it was less successful for the detection of stained pixels indicative of the infection of individual RBCs. Segmentation and counting of RBCs and detection of those containing stained pixels indicative of parasite infection are the fundamental steps whose solution is required in order to determine the degree of parasitemia of a sample. Before describing our work, we first give some further background on malaria and in particular on the thin-film blood slides images that are used throughout this thesis.

1.2 Background

Malaria parasites are micro-organisms that belong to the genus plasmodium. There are more than 100 species of plasmodium, which can infect many animals such as reptiles, birds, and various mammals. Only five species of plasmodium infect humans in nature, although there are some other species which can, exceptionally or under experimental conditions, infect humans. The five species which infect humans are: Plasmodium malariae, P. ovale, P. vivax, P. falciparum and recently, P. knowlesi. The first two species are less frequently encountered, while the fourth one is the only species that can cause severe, potentially fatal malaria [58].

A definite diagnosis of malaria infection is established by finding parasites in the blood [66]. The main component of blood is a yellowish fluid, called plasma. A great portion of the plasma is composed of water, a medium that facilitates the circulation of the many indispensable factors of which blood is composed. A cubic millimetre of human blood contains about 5 million red corpuscles called erythrocytes (but which we shall, as above, simply call red-blood cells, or RBCs for short); 5,000 to 10,000 white corpuscles (or white-blood cells – WBCs for short) called leucocytes; and 200,000 to 300,000 platelets properly called thrombocytes. The blood also carries many salts and organic substances in solution [37].

Even with all the serological and immunological techniques available and under development probably the only certain way of currently diagnosing malaria is by microscopical examination of the blood. This has been the case in the last century [66] and despite all the advancements in molecular biology still appears to be the case [177]. Microscopic examination is the most reliable test for malaria.

Malaria biomedical research work focuses on two species: humans and mice. Mice are often employed in the laboratory as genetic models because of the similarities between sections of human and mouse DNA [134]. Trager and Jensen [181] were the first to successfully culture red-blood cell stages of malaria. Two kinds of slides may be prepared for microscopic examination of blood samples, "thin" and "thick". When working in the laboratory, scientists prefer to make thin slides. These are produced by putting a single, small drop of blood onto a microscope slide and smearing it along the length of the slide with the edge of (say) another slide. In a thin slide, one thus sees only a single layer of mostly just red blood cells which are usually stained so that malarial DNA is visible as a distinctive colouration (figure 1.1). Red blood cells have no DNA and are thus not stained. The process most often used is Giesma staining which gives a purple colour to the malarial DNA. In the field both kinds of slides are prepared: thick as well as thin: thick to detect parasites, and thin for species identification.

A thick slide or smear is just a drop of blood on the microscope slide which, as noted above, are often not adequate for species identification of malaria parasites because cells overlap in the field of view and are difficult to distinguish, or are *lysed – i.e.* the cell membranes are disrupted. If a thick slide is positive for malaria parasites, a thin slide should be used for identifying the species of malaria infection and for quantitative work such as counting the fraction of infected cells and thus determining the severity of infection.

In the laboratory, thin-film slides may also be used for determining the stage of the malaria parasite's lifecycle but this quite difficult classification task was not one of our objectives. Determining the stage of the lifecycle is important in laboratory work because cultured samples of infected blood can be synchronous or asynchronous. Synchronous slides contain malaria parasites all at the same stage of the lifecycle. The malaria parasite undergoes three main stages during its infection of red blood cells. In this erythocytic cycle each parasite starts as a small ring stage with delicate cytoplasm plus one or two small chromatin dots and other occasional ornamental forms, followed by trophozoites of compact cytoplasm and dark pigment, which mature into schizonts with dark pigment which clump in one mass [58]. Figure 1.2 shows a more detailed and complete schema of the life cycle of malaria. Life cycle stages within the human liver and within the mosquito are included together with the 'off-shoot' of the erythrocytic

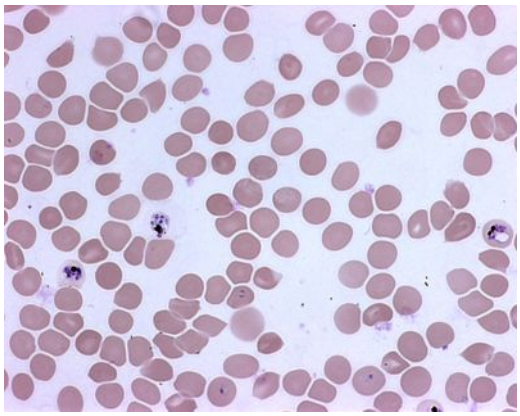


Figure 1.1: An example image of part of a thin-film microscope slide showing healthy RBCs and RBCs infected with malaria parasites. The parasites have been Giesma stained and appear dark blue/purple. This 1300×1030 pixel, 24-bit colour image courtesy of the NIMR taken at 1000 times magnification is typical of those used in this thesis with $\sim 100 - 150$ RBCs of diameter $\sim 70 - 80$ pixels visible of which a few $\%$ may be infected.

cycle in which gametocytes are formed. The part of interest to our work is the erythrocytic cycle in which the parasite resides within the RBCs. Different stages of the malaria parasite lifecycle can be observed in a single asynchronous slide.

In this thesis, we focus on working towards development of a system to measure the degree of infection by malaria parasites using a colour image of a thin slide of laboratory cultured malarial blood taken using a microscope in order to evaluate the parasitemia of the blood, (*i.e.* to count the number of parasites per number of red blood cells). A manual analysis of slides is tiring, time-consuming and requires expert technical staff. Our task is thus to automate the counting processes and any detection, segmentation, classification, labelling and other processes that may be required. It must be noted that while for a human detection, classification and counting are concurrent process they are usually separate processes in a machine vision system.

1.3 Aims and Objectives

The ultimate goal of our work is thus to contribute to the development, implemention and evaluation of an automatic system for the detection, location, identification, characterisation, and counting of red-blood cells in images of thin-film slides produced in the laboratory. By identification and characterisation in the above we mean determination of whether a cell is healthy or infected. Ideally, such a system would operate without error and with total reliability. It is, of course, extremely unlikely – in fact almost certainly impossible – that such a perfect,

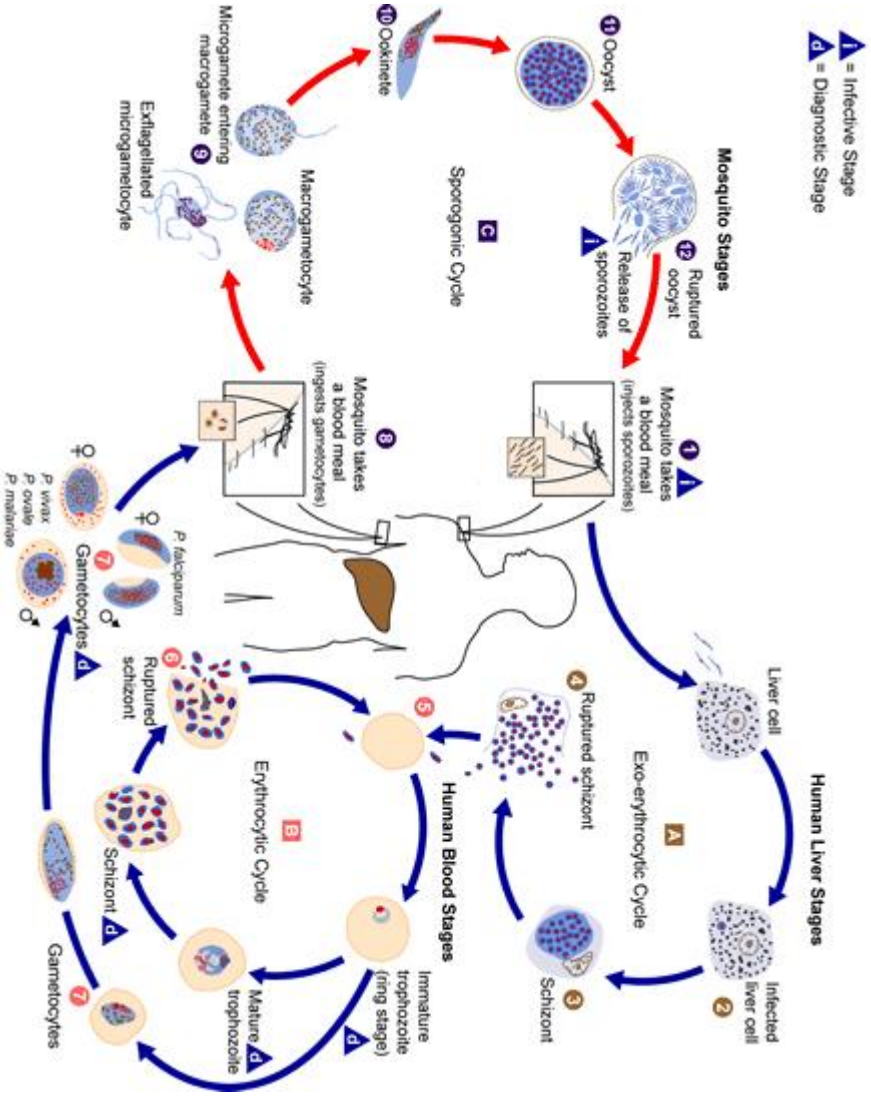


Figure 1.2: Schema of the Life Cycle of Malaria, from [58]. The three stages in the parasite erythrocytic life-cycle within the RBCs are illustrated towards the bottom right of the diagram. We note further for future reference that the degree of infection or parasitemia is defined using the number of parasites in asexual stage malaria, *i.e.* in the erythrocytic cycle.

ideal system could be developed. In practice, therefore it is necessary to establish a set of performance criteria according to the requirements of the users and of the medical applications in which the system might be utilised.

1.3.1 Central hypothesis

The main hypothesis statement of our research may be given as follows:

> *To contribute to development of an image processing/computer vision system for determining the parasitaemia of malarial infected samples of human blood used in laboratory biomedical research from images of microscope slides of thin-films.*

To build such a system and be confident it will meet laboratory requirements, we must be able to evaluate its performance on suitable test data and, in principle, be able reliably to predict its performance on images that may be taken in the future when the system is used in practice. For these purposes, images of slides in which the blood cells have been detected, located, classified and counted, all by a human expert or experts are required. Such annotated images may then be used in development of the algorithms and techniques to be used, in training (if required in a supervised system) and testing to determine the best algorithms and their optimal parameter settings, the overall system's expected performance and characterisation, and validation to verify that the expected performance is obtained. If required, such data could also be used in acceptance testing, normally used in commercial work to demonstrate that a pattern recognition system meets the contractual requirements of a customer or end user. In a system to be used in the laboratory a similar route would be required prior to its use as a tool. Such validation and acceptance testing may be regarded as analogous to clinical or more likely, to pre-clinical trials[1]. Such procedures require lots of data annotated by an expert or experts though cross-validation [99] and bootstrap techniques and, possibly, Bayesian statistical methods to integrate over the distribution of system parameters [14] may be used to improve the efficiency of data utilisation. Furthermore, in order to carry out such work efficiently, to make the best use of the available time of expert malaria researchers and, more importantly, to enable one to carry out the required research and development effectively, software tools are required for image capture and recording of experimental protocols and the image capture conditions, for annotation and storage of the images in a database, their retrieval, manipulation, processing *etc.*.

[1] Full-blown clinical trials are very extensive, have to be carried out according to strict medical standards and are usually expensive and time-consuming.

1.3.2 Sub-hypotheses

The central hypothesis above is very broad and has been addressed via a number of specific sub-hypotheses, *all involving completely automatic algorithms*:

(i) The detection of pixels belonging to red-blood cells, irrespective of whether they are healthy or infected by a malaria parasite, can be automated by use of an unsupervised classification algorithm applied globally to the whole of a thin-film microscope slide image.

(ii) The pixels detected in (i) may be grouped within individual RBCs and the locations of the cells within the image determined.

(iii) The grouped objects in (ii) may be counted in order to determine the number of RBCs in the image.

(iv) The pixels detected in (i) may be further processed by a second application of a similar unsupervised classification algorithm to detect those belonging to parts of RBCs infected by malaria parasites and used to indicate whether a RBC is infected or healthy.

(v) The outputs from (iii) and (iv) may be combined to determine the parasitemia of a blood sample to an accuracy and with a reliability that should meet the requirements of biomedical researchers and is competitive with results published by others in the literature.

There are a number of points relating to each of these sub-hypotheses that require clarification.

1.3.3 Detection of pixels belonging to red-blood cells (i)

The first point to emphasize is that this refers to classification of individual pixels as belonging to the image of RBCs (or parts of RBCs) or not. The hypothesis makes it clear that it is immaterial whether the RBC is infected or not and does not make any reference to classification of pixels deemed not to belong to the image of a RBC or part of a RBC. It is thus a two-class problem with the RBCs regarded as image 'foreground' or objects of interest and the rest of the image as 'background'. The 'excluded' pixels in the background are most likely to belong to the image of the blood plasma but could also belong to white-blood cells (WBCs), to platelets or to image artefacts, for example created by the staining process. In the wild, other bodies might be encountered in the blood including other parasites that may infect people living in tropical areas but this is not likely in the laboratory research samples of interest to us. Similarly, in the laboratory samples used in the current stage of malaria research we do not encounter

RBCs affected by other diseases. Our problem is thus much easier than it might be in a field application and we do not have to concern ourselves with, for example, the possibility of some RBCs being severely distorted in shape owing to conditions such as sickle-cell anaemia. We also thus refer to RBCs that have not been infected by malaria parasites as 'healthy'.

In addition, we should note that there is no reference to sub-parts of the RBCs, in particular their boundaries. Close inspection, for example of the image in figure 1.3, reveals that in many of our images one can often discern the outer membrane of the RBCs as a ring-shaped region that is lighter than the cell body and distinct from the background. It is not specified whether this boundary membrane is to be regarded as part of the cell or not. In principle it might be best if pixels on the boundaries of the RBCs were regarded as excluded from the cells or as a third class distinct from the foreground body of the cells and the background. However, as can be seen in figure 1.1 the parasite infections of interest for determination of the parasitemia are confined within RBCs. Thus, in practice within the approach we have followed, the point is largely irrelevant though we do need to keep an eye on the extent to which it may affect performance. More pertinently, in particular for an evaluation of performance at the level of classification of individual pixels [169], we do need to be aware of the fact that pixels on region boundaries will be *mixed*. If we include the RBC boundary membranes there are then in principle five or six classes to which pixels may be assigned: RBC body, RBC boundary, background, and mixed RBC body and boundary plus mixed RBC boundary and background, and depending on the width of the boundary membrane, possibly also mixed pixels spanning the RBC body, boundary and background (figure 1.3).

A second point to note is that the hypothesis refers to an unsupervised classification algorithm. Unlike a supervised system, such an algorithm does not require training data which has been labelled by a human expert or in which the classification may be determined by other means. Unsupervised systems are thus more convenient than supervised systems but care has to be taken to characterise the range of data over which an unsupervised algorithm will produce good results and, if appropriate, the settings or adaptation of any parameters required and initialisation conditions.

Thirdly, it is to be noted that the algorithm is to be applied globally over the whole of an image and that no pre-processing of the images is mentioned. Global application of an algorithm, in particular of decision thresholds is a point to bear in mind, especially as previous research using data similar to ours (and some of it from the same source at the NIMR) has questioned whether such global approaches can work because of variation in the intensity across an image caused by non-uniformity of the microscope illumination [176, 175]. We will propose

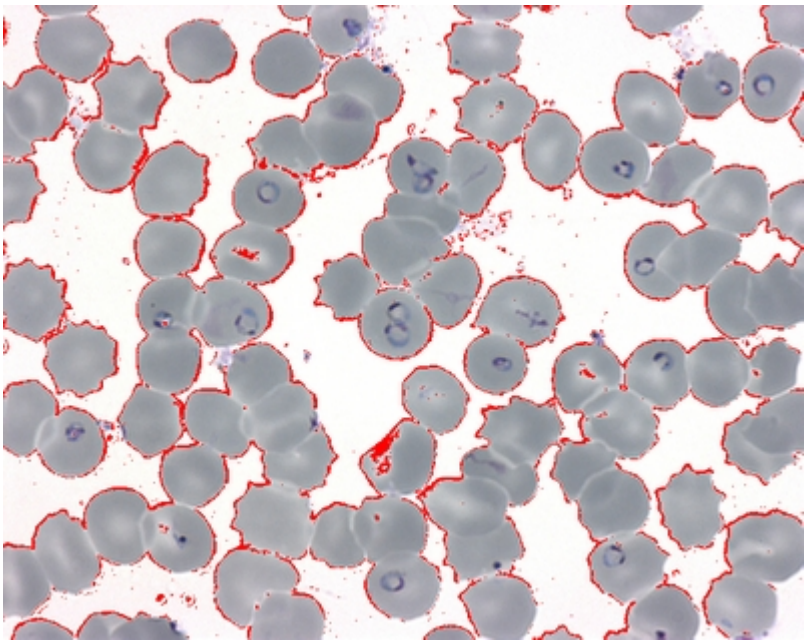


Figure 1.3: A memorable segmentation in which RBC boundaries were partially detected.

how one might examine and characterise the extent of such variation within an image and the extent and likely effect of variations, in particular of colour and intensity, between images. The latter may be affected by many things including the camera used as well as the microscope illumination and the details of the staining process and thin-film preparation. We note in passing that Tek [176, 175] points out that a Kohler illumination set up [170] of the microscope can and should be used to avoid significantly non-uniform illumination but that this is often neglected by researchers who are not expert in microscope optics because adaptation of the human eye to variations in the local illumination mean the problem is not perceived. Non-uniformity of the staining and in the procedure for making a thin-film slide may cause similar effects, though they are mitigated by the fact that, though our images are quite large (1300×1030 RGB pixels, figure 1.1), the magnification of 1000 times is such that the field of view is only a fraction of the area of a microscope slide.

In fact, the magnification of the microscope images is such that many (thousands of) images could be taken per slide. Design and implementation of the systems required to automate and control the image acquisition would thus also need to be addressed for the ultimate, automatic laboratory (or field) applications. This would require control of microscope focus and x-y table movement and, given the huge amount of data that might otherwise be created, a strategy for characterising suitable fields of view and for selecting sample fields [176, 175]. A half-way house for our laboratory research application might be to retain user control of the image acqui-

sition – which thereby would have a degree of expert quality control – but to target the tedious (and therefore for a human, often error-prone) manual image analysis currently required for the determination of parasitemia. This is the situation most relevant to our work.

Two other issues implicit in the hypothesis are the possible effect of the density and distribution of RBCs within the images.

Finally, we note that the hypothesis makes no mention of the level of performance required. Its performance is important as the output of this first algorithm would be used in other parts of the system but, provided the performance is adequate to enable the overall determination of the parasitemia (say) to be sufficiently accurate, it is not of fundamental interest. However, following the precedent set by Southall *et al* [169] in systematically evaluating the intermediate algorithms and system components as well as its final output, and as our discussion of the issues regarding mixed pixels indicates, we shall evaluate this first step separately. We shall also try to ensure that the reasons underlying the performance of each step in the algorithm are understood and, where possible, characterise the robustness of the performance (chapter 5).

1.3.4 Pixel grouping within RBCs (ii) and counting the grouped RBC objects (iii)

The points that require clarification here include the extent to which it is necessary to group pixels within individual cells as opposed to clusters of two or more cells that are adjoining in the segmentation produced by the first step. This may occur because the cells are touching or because of poor performance of the first stage algorithm. In practice we need to discuss how a poor segmentation is recognised and, more importantly, to specify what is done to deal with touching RBCs which, as previous works have noted [154], will be a problem to be overcome if cells counts are to be accurate. The details are given in chapter 6. In addition we also have to decide what to do about cells that do not lie entirely within the image so as not unduly to perturb determination of the parasitemia and, in accord with our strategy to evaluate each of the intermediate steps, to permit a fair assessment of the performance. Although there is nothing on the way to evaluating the parasitemia for which the location of individual cells is explicitly required, it is convenient to have a record of the RBC locations for interrogation of the images, for use in understanding subsequent processing steps, display of results and for the testing of alternative approaches.

Finally, we note in particular that the locations of the RBCs are required in order to evaluate in detail whether each individual cell has been detected. In principle we would like to assess the true-positive rate or sensitivity of RBC detection, and its false-positive rate or (one minus the) specificity. Deciding whether cells have been correctly or incorrectly detected requires knowl-

edge of their locations and has to confront the question of what accuracy of location is required for a detection to be regarded as correct. This is a long-standing problem in the evaluation of computer vision algorithms familiar for example in early work on the detection of localised features such as edges or corners. An overall evaluation of the cell count however doesn't require such information but is less informative about algorithm performance as false positives and false negatives may cancel. A detailed evaluation also conveys greater understanding of the behaviour of the system and (hopefully) conveys greater confidence in the performance to be expected in practice. In addition, we also note that whilst it is easy to count the number of cells incorrectly detected as false-positives, it is more difficult to determine the false positive rate as the number of confusing negative objects in an image is not well-defined or known. This situation is encountered in many classification tasks not only in computer vision, but also in other areas as diverse as text mining [36].

1.3.5 Detection of infected RBCs (iv) and determination of parasitaemia (v)

The main point of clarification or emphasis here is that since we do not attempt to identify either the parasite species or stage of the erythrocyte life cycle (recall figure 1.2) an accurate segmentation of the stained pixels belonging to areas of the image corresponding to the presence of parasites within RBCs is not required. All we need is to gather sufficient evidence to classify each cell as infected or healthy. Similar issues have been considered by Ross *et al* [151] in the development of their system designed to detect whether results from the analysis of an image (or images) of a blood slide indicate whether a patient is healthy or infected. The species and life cycle stage are usually known under the controlled conditions in which samples are cultured in laboratory research (page 6).

Secondly, that a similar unsupervised classification algorithm to that developed in (i) is applied only to the pixels detected in the first stage as belonging to RBCs in order to detect pixels corresponding to parasites. There are two reasons why we can do this and two as to why we choose to do it. The first of the two that enable us to proceed in this manner is that the definition of parasitemia depends only on the fraction of RBCs infected by the parasite within the erythrocyte life cycle where the parasites are always within RBCs. The second of these two reasons is that since RBCs have no DNA of their own only the parasite can be stained within a RBC. The first of the two reasons as to why we choose to process only the pixels output from the first stage algorithm as belonging to RBCs is that pixels belonging to image regions corresponding to stained parasite comprise only a very small fraction of an image. They are thus a minority class far outnumbered by pixels belonging to the foreground RBCs and very far

outnumbered by pixels belonging to the image background 'plasma'. Secondly, a multi-class version of our Fisher-Otsu algorithm has not been developed so there is no alternative to using a hierarchical classification architecture if we wish ultimately to use this algorithm to distinguish more than two classes.

Finally we note that the count of the overall number of RBCs together with the decision as to whether each cell is healthy or infected with a parasite obviously provide sufficient information to calculate the parasitemia or degree of infection. The main questions remaining are: the way we should evaluate this final outcome and its reliability; whether the results would meet the requirements of potential laboratory research users and are competitive with results published by others in the literature [165]; the accuracy and reliability that therefore could be expected were the approach we have followed to be developed into an application; and any further work required in order to produce such an application.

1.3.6 General comments

Further context is of course required from a review of the relevant literature to be given in chapters 2, 3 and 4 but the above hypotheses and comments set much of the scene for the work described in this thesis. A few more general, additional comments may be made here however.

The first is, as mentioned in the last paragraphs above, that our approach is based on a hierarchical classification system. Such an approach of course means that the performances of later stages in the hierarchy are limited by the performances of stages that preceded them. It is therefore particularly important that segmentation of pixels belonging to RBCs in the first stage (i) (subsection 1.3.3) is accurate and reliable and that failures can be detected.

The second is that we haven't mentioned computational or hardware requirements. As will be seen, our main Fisher-Otsu algorithm is iterative, so the former could as usual be dealt with by consideration of the computational complexity of the main operations required and the convergence characteristics of the algorithm. The ways in which it can fail and the detection of such failures should also be considered. As for platform requirements, there are almost none. At the time of writing, desktop workstations, personal computers and laptops have all for some ten years or more had the memory, storage and processor capacities to deal routinely with image data of megapixel size. For the research carried out in this thesis there has thus been no emphasis or consideration of matters such as data transfer, storage or efficient coding other than to avoid obvious poor scaling of the main operational steps. For convenience and to enable the greatest attention to be given to the methods developed, Matlab has thus been used throughout. Nor would the target biomedical research users care greatly about detailed computational and

memory requirements issues. For them the main concern is to reduce the human effort required to complete a tedious task and to know that any such computer application will produce outputs of sufficient accuracy and reliability within a reasonable time which could be anything from minutes to hours on reasonably standard PC platforms. If necessary, they would be prepared to use multi-processor systems to speed-up the calculations [50].

1.4 Achievements and Contributions

In this section we summarise what the work carried out in the course of our studies has achieved and what contributions have, and may in the future, be made.

1.4.1 Achievements

Our main achievement is the unsupervised Fisher-Otsu algorithm developed initially for the first step of segmenting pixels belonging to RBCs. This algorithm is based on a combination of two well-established techniques both of which utilise a linear Fisher discriminant. The first is the classical use of the Fisher discriminant [54] to reduce the dimensionality of the RGB colour space by determination of the most discriminating feature vector as described in pattern recognition textbooks and review articles (see for example:Duda and Hart [45], Duda, Hart and Stork [46], Devijver and Kittler [43], Kittler [97] in the collection edited by Fu, and Sergios Theodoriois and Konstantinos Koutroumbas [178]) and resources such as CVOnline: Vision Geometry and Mathematics (http://homepages.inf.ed.ac.uk/rbf/CVonline/geom.htm), and of course Wikipedia (http://en.wikipedia.org/wiki/Linear_discriminant_analysis). We use the Fisher discriminant to project the data onto the most discriminating direction in the RGB colour space to generate a one-dimensional feature space to which the second ingredient, the Otsu algorithm [131], which is itself a consequence of application of the Fisher discriminant as we shall see in chapter 4 and in detail in appendix A, is applied. In our Fisher-Otsu algorithm these two steps are applied iteratively in alternation. The resulting algorithm which to the best of our knowledge is novel seems very well suited to segmentation of the pixels in our problem. Its development is described in chapter 4 together with its use for the segmentation of RBC image pixels.

From this, flow two achievements:

(i) An understanding of the algorithm with an exploration and characterisation of its convergence and robustness, the latter involving in particular an analysis of the effect of affine transformations of the RGB colour space on the algorithm (chapter 5).

(ii) Development of procedures for the grouping of RBC pixels into objects corresponding to

individual RBCs and clusters of contiguous RBCs and counting the number of cells these objects represent. The grouping procedures involve for the most part standard techniques but the counting, using a further application of the Otsu algorithm in conjunction with a number of *internal* cross checks is, to the best of our knowledge, original. Evaluation of the counting indicates it is both robust and reliable (chapter 6).

Furthermore, we propose, but have not implemented or evaluated:

(iii) Ways to characterise image data variability and the extent to which such variability, whether intra-image or inter-image, may be modelled by means of affine transformations of the RGB colour space to which the Fisher-Otsu algorithm is expected to be largely insensitive (chapter 6).

Finally, we have carried out a preliminary exploration of:

(iv) Further applications of the Fisher-Otsu algorithm to pixels segmented as belonging to healthy RBCs for the detection of pixels that belong to stained parasites.

(v) Using the detection of stained parasite pixels to identify RBCs that are infected and thus counting the numbers of infected and healthy RBCs as would be required for determination of the degree of parasitemia (chapter 6).

We may thus say that sub-hypotheses (i), (ii) and (iii) have been achieved and, given our limited preliminary investigations, that (iv) and (v) have been explored and remain as possibilities that further work might show were achievable by the approaches described in this thesis to accuracies useful to medical researchers.

1.4.2 Contributions

The main contribution arising from this thesis work is publication of the Fisher-Otsu algorithm [23]. This happened at an early stage of the research – hence the different name (Otsu3D) used there – in particular before we had developed a good understanding of the algorithm's behaviour and before much had been done to characterise its convergence and robustness or to explore its failures. In fact there were no failures of the algorithm as first constructed for the RBC pixel segmentation and initialised via use of PCA. From the outset however we always evaluated its performance and thus, for example, in [23] compared the algorithm with well-established alternative ways of using the Otsu procedure.

Given the deeper and more complete understanding we now have and that a system has been developed for the segmentation of RBCs and clusters of contiguous RBCs and for counting

the number of cells such objects in an image, it seems that there is potential perhaps for two further publications:

- one in the computer vision / pattern recognition literature on the Fisher-Otsu algorithm and its behaviour as illustrated by the RBC pixel segmentation and its hierarchical application for the detection of stained pixels belonging to parasites; plus
- a paper in the biomedical literature on the automatic counting of RBCs in thin-film microscope slide images.

1.5 Structure of the Thesis

The structure of the bulk of the remainder of this thesis is implicit in the cross-references in the achievements section 1.4.1 above. These indicate that chapters 4, 5, and 6 will in turn describe: development of the Fisher-Otsu algorithm; its use for segmentation of pixels belonging to RBCs; its convergence and robustness together with a discussion as to how variability of the data might be characterised; counting RBCs; and preliminary work on parasite detection and classification of individual RBCs as healthy or infected. In the final chapter 7 we discuss and summarise the conclusions to be drawn from our research and, as also indicated above, discuss what directions for further research our work prompts and what future work might be carried out.

As we shall see, the three most important elements needed for such research to be undertaken effectively and efficiently are: the provision of suitable data in sufficient quantities; the evaluation methodology or strategy; and the programs and software tools required to facilitate development of the algorithms and procedures, their testing, and assessment. Tools are also required for characterisation of the image data and its variability so that, given an *understanding* of how and why the algorithms work and how they may fail and failures be detected, the approach developed might, if desired, be applied in practice with confidence by biomedical researchers.

In chapter 2, we review the literature most relevant to the processing of images of thin slides of blood samples. This review opens with a discussion of general aspects of the problem based on papers on closely related topics, such as the processing of such images of blood slides for the diagnosis of leukaemia. We then use this background information to focus on discussion of a series of papers on the processing of microscope images of thin-film blood slides for the detection and characterisation of malarial infection. Of particular interest is work from 2000 to 2010 by the group at the University of Westminster: by Di Ruberto *et al* [154, 153, 155, 156] on

the detection of red blood cells infected by malarial parasites in thin-slide images, similar work by Rao *et al* [146, 147, 148, 149], and most recently by Tek [175] and Tek *et al* [176]. We then introduce and review other work relevant to our research, in particular two papers published in 2006 on systems developed for counting malarial infected cells [165] and on the diagnosis and classification of malaria on thin blood smears [151]. More recent image analysis work on the 'reading' of malaria blood slides (as it is called) is briefly discussed in the epilogue to the final chapter (chapter 7, section 7.3). We also briefly review the mostly 'standard' but fiddly techniques used to group pixels segmented as belonging to RBCs into individual cells and for counting the number of cells present in an image. Finally, a number of potential alternative techniques and approaches are briefly discussed, in particular in chapter 3 some modelling methods that we tested but found difficult to make work well.

Background and review of the pattern recognition theory on Fisher's linear discriminant and its application to histogram-based segmentation of images which leads to the well-known algorithm [89] first developed by Otsu [131] is mostly deferred to chapter 4 where these two techniques are brought together in development of the algorithm that is our main, core contribution.

With the comments in the section on achievements 1.4.1, the above set the scene or context for our research.

Chapter 2

Literature Review

In this chapter we review literature most relevant to processing of microscope images of thin-film slides of blood samples in malaria applications. We have already noted in the introductory chapter that our objectives are quite limited and that the images we shall be using are not very variable. The fundamental image processing/computer vision task to be carried out is segmentation, in particular of pixels belonging to RBCs and subsequently of stained pixels belonging to malarial parasites within RBCs. Grouping of pixels into objects representing RBCs, counting the number of cells such objects represent and deciding whether the cells are infected or healthy are also important. As we shall see, we thus do not require access to a large area of computer vision and image processing research and some of the techniques we have used are quite standard. It is nevertheless useful to set the context a little wider and to include in this review chapter brief discussion of:

(i) the processing of other biomedical imagery where it impinges on our research and/or on malarial applications,

in addition to more specific items on:

(ii) pre-processing techniques that may be used to reduce the range of image variation (limited though that may be) together with thresholding, the most commonly used image processing technique for segmentation in this and many other biomedical applications;

(iii) work by others on the processing of images of thin-film microscope slides for the detection and characterisation of malarial infection, including techniques required in order to build on the segmentation processes and finally obtain estimates of the degree of parasitemia;

(iv) classification techniques, feature selection, and, briefly, the use of neural network and other "black-box" classifiers and classifier combinations;

(v) performance measures, their limitations and practical application, in particular of the ROC, (receiver operating characteristic) curve to the problem; and

(vi) the fitting of deformable geometric models and templates, in particular the work of Liu and Sclaroff from around 1997 to 2001 which was applied to a similar problem in the processing of images of red blood cells in thin-film microscope slides and, again briefly, evolutionary search and optimisation techniques used in such approaches.

The items are numbered above to correspond with the main sections to follow in this chapter. Before embarking on those sections, we note that, in particular re (iii) above, a very useful review has recently been published by Tek *et al* [176]. Though that review is aimed at the automated diagnosis or screening of malarial infection it covers much of the background specific to processing of microscope images of blood thin-film slides, and in particular the decade of work by the group at the University of Westminster. The chapter concludes with a brief summary and mention of the flexible modelling approach to be discussed in chapter 3.

2.1 Biomedical image processing

To provide greater background and a wider context for the processing of 2D-images of thin-film microscope slides of malarial infected blood, we begin with a brief discussion of general aspects of biomedical image processing based on papers on closely related topics, such as the processing of such images of blood slides for the diagnosis of leukaemia. As the "2D" above indicates the processing of 3D medical images such as those produced by X-ray CT, MRI, SPECT, PET etc. and of time-varying, video sequences (of whatever dimension) are excluded from our discussion. In fact most microscope images of biological samples are considerably more challenging than our blood thin-film images as the example in figure 2.1 shows. Though for a different application and though their aim of a fully automated cell analysis and diagnosis of leukemia was not yet reached Hengen, Spoor and Pandit [79] covered in their work many important steps in the image acquisition and pre-processing of microscope slide images of blood (and in their case, bone marrow) smears (see figures 2.2 and 2.3). Indeed the focus of their next phase of work is described to be on classification by using features such as the mesh structure of the cells, the shape of cell and nucleus or the colour and granularity of the cytoplasm. Classifier choice is also not an easy step and, like many working on the processing of biomedical images, they have proposed a wide range of classification techniques including: a statistically optimal classifier with the presumption of a probability density of features (Bayes classification and special forms of maximum likelihood classification); classification by means of neural nets and

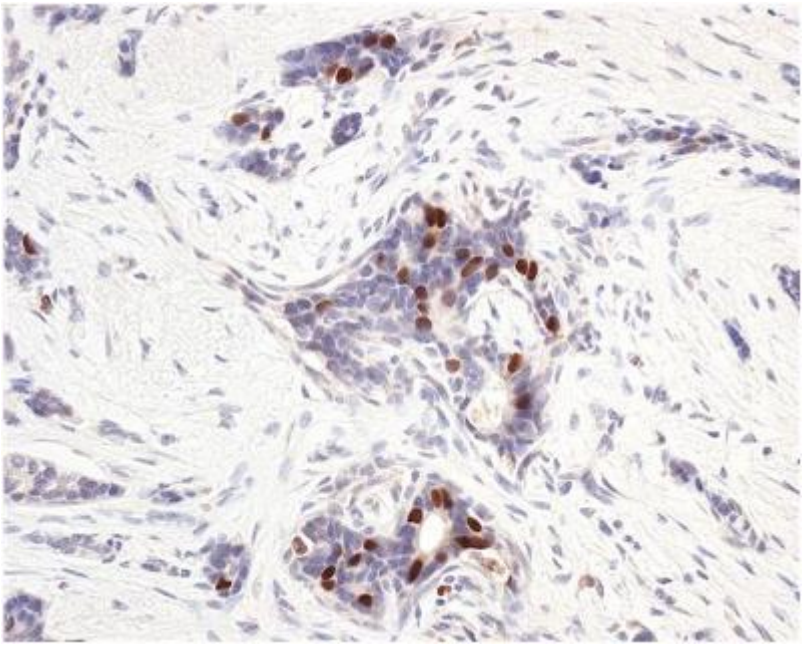


Figure 2.1: Example biomedical image courtesy of Dr. Marco Loddo at the Wolfson Institute for Biomedical Research, UCL. The image is of a stained, thin section of cancerous tissue. The aim is to count the numbers of healthy (purple) and diseased (brown-stained) cells. It can be seen from comparison with figure 1.1 that this is a much more difficult application than ours, amongst other things because of the background of connective tissue which is much more complicated than the background blood plasma in our case.

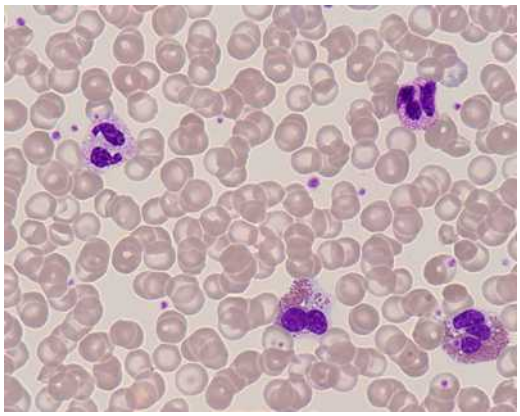


Figure 2.2: A blood smear image copied from the reference given in the text.

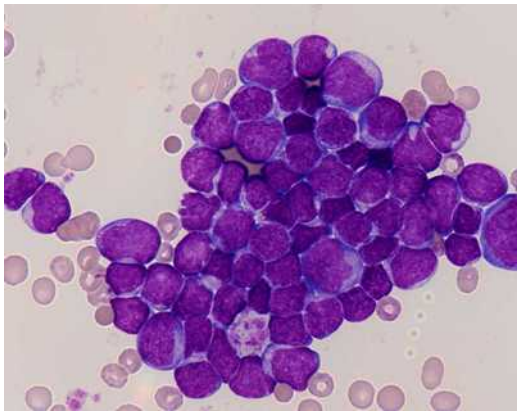


Figure 2.3: A bone marrow smear image copied from the reference given in the text.

by support vector machines (SVMs); and minimum distance classification. We shall review classification techniques in section 2.4 below. Results were described as still pending, but a further conclusion that may be drawn from this work is that consideration of such a range of classifiers does not seem untypical in biomedical image analysis. A quick scan of similar biomedical image processing literature confirms this – see for example [188, 129, 189]. Many are applications to dermatology, for example for skin cancer detection, and to white blood-cell analysis, for example in leukaemia studies.

In fact, control and automation of image acquisition for our application is discussed in the review by Tek *et al* [176] mentioned at the beginning of this chapter. Though automation would be the ultimate requirement for laboratory (or field) applications it raises many issues, such as how to control focus and in particular the $x - y$ movements of the microscope slide required to capture the thousands of fields of view needed to tessellate, or at least reliably to sample, the slide smear (see [165] and [175, 176] and *c.f.* [194]). Fortunately, such issues are not relevant for our work in which, as mentioned in the opening chapter, we assume images

are provided by biological researchers. The main issues for our work are variations in intensity and colour within and between images together with small variations in magnification which might change the apparent size of RBCs by (say) $\pm \sim 10\%$, numbers of RBCs and the fraction of RBCs which are infected. Such small variations in magnification, numbers of RBCs and the fraction of infected RBCs are intrinsic characteristics of the application with which we have to cope. However, variations in intensity and colour within and between images, together with large variations in magnification are extrinsic characteristics that would ideally either be eliminated by the capture process (*e.g.* by controlled, multispectral imaging techniques [71] or by standardisation of the imaging [200] and illumination [170]. Tek gives further details in his paper [176] and in his thesis [175]. We discuss in the next section some of the standard pre-processing techniques that may be used to reduce such extrinsic variations and we shall return to the issue of these extrinsic variations in chapter 5. We also note in passing that provision of public data sets, in particular if they were large-scale and sufficiently representative of thin-film microscope slide images of malarial infected blood [119] would also alleviate the problem – a point noted in our paper [23]. Before discussing image pre-processing, we end this section by noting that even images of blood samples in which red-blood-cells are the objects of interest often present more difficult application problems than ours. For example, Liu and Sclaroff seek to classify the shape of red blood cells in order to diagnose certain pathologies.

2.2 Image pre-processing and processing

Although image pre-processing does not usually increase image information content [168] (an obvious exception would be, for example, when model-based image processing is used to enhance or restore a poor quality, noisy or degraded image), image pre-processing can be useful in applications such as ours (and in biomedical applications in general) in order to reduce the variability owing to different factors such as changes in laboratory conditions, preparation of slides, staining and microscope view settings. For example, to deal with variability of acquired images two methods were investigated by Hengen *et al.* [79]: a prior and posterior standardization. The prior standardization was a simple check of the intensity and saturation of a piece of the image background to indicate whether the illumination is intense enough and altering the microscope settings accordingly. Both selection of the background region and alteration of the microscope setting were manual. The posterior standardization was applied to the intensity histogram. For an ideal, balanced camera the intensity of each pixel may simply be represented as

$$I(x, y) = R(x, y) + G(x, y) + B(x, y) \tag{2.1}$$

where R, G, B are the red, green and blue channels of the colour image and (x, y) denote the pixel locations. Images contain unevenly distributed grey values owing to their specialized content. Since images of stained thin-film microscope slides of blood and (in the paper by Hengen *et al*) bone marrow contain only a small number of different components, the resulting intensity values of the histograms of such images usually lie within a small range. Histogram equalization is a method for stretching the contrast of such images by uniformly redistributing the grey values [89]. This can result in a striking contrast improvement that makes otherwise hidden features visible to a human observer and can make manual threshold selection easier and more effective – *i.e.* to give a better segmentation of the image content. Depending on the algorithms used, histogram equalisation or other transformations of the grey-levels (or colour attribute values if thresholds are applied to each of the colour channels) may also lead to improved results in an automatic system.

A very simple form of image standardization is to "scale" the image grey-levels or colour values. This is a linear transformation. Slightly more general is an affine transformation which in addition to the scaling includes a constant additive bias value. The problem with such transformations is that the transformed values may under- or over-flow the $0 - 255$ allowed range and that the resulting histogram also has gaps between occupied grey-level sample bins. This problem, caused by quantisation of the grey-levels (or colour values) also occurs in histogram equalisation. Better methods stretch the histogram while filling all bins. A simple procedure would be a random redistribution of the pixels from occupied sample bins to neighbouring unoccupied sample bins [89]. Such methods may also be applied to pixels whose transformed values fall outside the allowed $0 - 255$ range . A range-limited adaptive histogram equalisation due to Karel [95] is used as a pre-processing step in the paper by Sio *et al.* aimed at the determination of parasitemia from cell-counts [165]. This paper is reviewed in detail below in section 2.3.4. Hengen *et al.* [79] chose to spread the histogram in such a way that the minimum value of the grey-levels is retained while the value at which the minimum preceding the background peak occurs is spread to the maximum grey-level value of 255, in effect cutting off the peak of the histogram and thus eliminating the background. This can be viewed as a method for separating the darker foreground from brighter background. It is noticeable that, although Hengen *et al* may have used a manually chosen value to select the minimum in the grey-level histogram immediately below the background peak, it would be straightforward to develop an algorithm to do this automatically. The paper however does not state clearly whether this was the case or not.

2.2.1 Colour transformations

In his work on the detection of malarial infection Tek [175, 177] also used separate transformations for foreground (RBC) and background (blood plasma) regions. In a manner somewhat similar to that of Hengen *et al*, the transformed background level for each colour channel was scaled to a maximum of 255 whilst in the foreground each channel was scaled to match the average of the foreground levels in twelve reference images selected by malaria researchers at the NIMR. Using the maximum value of 255 leads to some problems with overflow of the transformed pixel attributes and it would seem that the spatial inhomogeneity of the transformation may introduce artefacts at the boundary of the RBCs. The latter is particularly dependent on the quality of the initial foreground-background segmentation and might be alleviated by iteration of the segmentation and transformation processes. Examples are given in Tek's thesis though such iteration is not used in his application. Tek notes that the quality of the foreground-background segmentation may be strongly affected by non-uniformity of the background image intensity especially if global thresholding techniques are used. This is usually caused by failure to notice poor illumination of the microscope slides and thus to adjust it accordingly, for example by use of a standard Kohler illumination setting [170]. If a reference image of the illuminant alone (*i.e.* with a blank slide) is available the lighting may be corrected by image subtraction. As such an image is not usually available Tek uses a segmentation method based on morphological top hats to adapt to such image non-uniformities.

Though our Fisher-Otsu algorithm is a global thresholding technique we have not encountered the kinds of difficulties Tek describes, perhaps because we use a smaller data set comprised only of images supplied by malarial researchers at the NIMR. However, image non-uniformity is an important consideration especially as it may also result from the staining process and be affected by the thin-film slide preparation procedure. The caveat mentioned above on the spatially inhomogeneous transformations used by Tek may be overcome by alignment of the average image colours (or of the colours of each pixel in a LMS sense), or better by alignment of image histograms since the latter is not sensitive to the details of spatial distribution of the image content, the position of individual RBCs *etc.*.

Histogram alignment has been studied by Senanayake and Alexander [162]. They consider the general aspects of such alignment but with particular emphasis on the alignment of feature points in the histograms using a variety of image transformations from simple shifts and scaling of each colour channel to non-linear, third order transformations of each colour attribute. An important conclusion from their work is that affine transformation of each colour channel usually gives the best results in terms of alignment quality, robustness, numerical stability and

algorithm complexity. It is also notable that provided the 'range' and 'gap' problems noted above are appropriately dealt with[1] all such transformations form a group so they may be inverted, compounded, iterated etc. as required and thus form a natural, simple way of removing such extrinsic variations in the image data in a manner similar, for example, to the way Euclidean and similarity transformations are used in flexible shape and appearance modelling (see for example [33, 34]).

In fact, in flexible appearance and more especially texture or patch modelling [32], histogram equalisation or the mapping of image histograms to a reference is often used. Since cumulative density functions are monotonic such transformations will be invertible *etc.* and thus represent the most general group of non-parametric global image transformations that might be used. They are thus potential candidates – along with the affine and other parametric transformations – for implementation of a kind of *transfer learning* technique in which images are mapped as closely as possible into examples on which the algorithms to be used are known to work well. However, these transformations will suffer from the problems due to the choice of histogram sample bins noted above and, in view of this and the conclusions of Senanayake and Alexander, have not been pursued in our work.

The remaining types of colour transformations in the literature include *grey-world* and colour constancy [5, 6] both of which are mentioned in the context of the malaria application by Tek [175] although reference to the most sophisticated type of 'comprehensive' colour constancy algorithms [53] is overlooked. However, many of these algorithms do not seem well-suited to the application and may, given the restricted nature of the image content, not work well in practice. Moreover, at most all that the malarial application requires is a way of reducing the variation between images to the point where the algorithms and systems developed may be expected to work well rather than a strict colour constancy that is illumination and camera independent. For this reason, we consider in detail in chapter 5 the extent to which our Fisher-Otsu segmentation algorithm is robust to variations in the images and in particular invariant under affine colour transformations.

We noted above that there may be variations within images due to the staining process and the thin-film slide preparation procedure. Similarly there may be variations between images. Whilst the colour correction of pathological images for different staining-condition slides has been studied [1], given that all our data is sourced from biological researchers at the NIMR well-practised in such techniques it does not seem necessary to give this special consideration

[1] For example, by using 'continuous' real values instead of integers in the range $0-255$ and allowing transformed values to be negative or to exceed 255.

and we assume it would be encompassed sufficiently well by monitoring of non-uniformities within the images and within the affine image colour transformations and the robustness of our Fisher-Otsu algorithm. Similarly, Tek does not give it further special consideration in his work.

As a final point we note that Partida [132] in his MSc thesis demonstrated the effectiveness of grey-level and colour image pre-processing in the malaria application where the overall performance of his system was considerably improved by using a set of normalized and colour tuned images. Normalisation was carried out with respect to the number of pixels having each colour value, resulting in what is often known as the grey-world transformation [53] whilst the colour tuning was carried out via a transformation which made the colour of uninfected cells and parasites more uniform among all images. We stress, however, that the effectiveness of this particular pre-processing is dependent on the classification procedure used (*i.e.* on the features selected, choice of classifier – in this case an SVM (support vector machine), on the training process *etc.*) and that its usefulness cannot be generalised to other approaches. The most we can say is that if pre-processing is needed to reduce the range of image variability so as better to match those images on which the algorithms were developed and the algorithms do not have the flexibility to adapt to other circumstances, it may be expected to be useful. Tek's work [175, 165] provides an example.

2.2.2 Image processing - segmentation by thresholding

There is a variety of segmentation methods in the image processing literature and text books but one simple way a segmentation can be produced is by thresholding. In its simplest form applied to grey-scale images, thresholding produces a binary image by labelling each pixel in or out of the region of interest by comparison of the grey-level with a threshold. If colour imagery is used, a threshold may be applied to the overall intensity $I(x, y) = R(x, y) + G(x, y) + B(x, y)$, to the pixel's red attribute values $R(x, y)$, or to the G or B values. In particular, although the grey-level image was used for segmentation in the malaria thin-film image work at the University of Westminster by Rao, the green channel was used for extraction of stained objects and the difference $B - R$ for detecting mature parasites [146]. Other convenient attributes that can be computed from the RGB colour signal may be used such as the chromaticities or the hue and saturation (*c.f.* the work of Hengen *et al* [79]). As might be expected, thresholding using colour attributes often works better than thresholding on the grey-level intensity, especially in applications where it is desired to produce a segmentation into regions similar to those a human would designate. Automatic thresholding analyses the grey value distribution in an image, usually by using a histogram of the grey values (or of the colour attributes), and uses

knowledge about the characteristics of the application to select the most appropriate threshold [89]. Automatic thresholding also helps overcome the difficulty of simultaneously choosing several thresholds when the values in each colour channel are to be separately analysed and the results combined.

In principle, a good way to choose a threshold is to separate the distributions of some property values characteristic of the populations of each type of object or image component to be distinguished [168]. An example which as we shall see later maximizes the between-class variance of the grey-levels – a measure of the separation of the class conditioned mean grey-levels – is provided by the following iterative algorithm [89] originally due to Otsu [131]:

1. Select an initial threshold T, usually chosen as the average intensity μ of the image if no other information is available.

2. Segment the image into two regions using T, *i.e.* under an ideal scenario separate the histogram into two peaks.

3. Calculate the mean grey values $\mu(1)$ and $\mu(2)$ of the regions, *i.e.* separately for each of the two parts of the histogram.

4. Calculate a new threshold $T = (\mu(1) + \mu(2))/2$.

5. Repeat from 2 until the means $\mu(1)$ and $\mu(2)$ do not change.

Since the algorithm lies at the heart of our work a full discussion of it will be given in chapter 4. A derivation of it, including generalisation to the multi-category case, and its relationship to the Fisher discrimminant is given in appendix A.

If more than two classes are required, multiple thresholds may be used, for example on the grey-level intensity as in the multi-class version of the Otsu algorithm described in appendix A. However, such an approach may be insufficiently sensitive to minority classes so a sequential application in a tree-structured, hierarchical classification system is more usual. The latter is the thresholding approach adopted in the malaria work of Ross *al.* [151], except that the thresholding is applied to the green channel for detection of pixels belonging to RBCs (erythocytes) followed by detection of pixels belonging to parasites.

In addition, whatever thresholding algorithm is used, thresholding usually does not take into account the spatial characteristics of groups of neighbouring pixels of an image. In some circumstances, in particular when trying to segment objects of a very variable and distributed (filamentary) shape as for example the stained nuclear material in an infected blood cell, this can

be an advantage but when segmenting compact objects such as the red blood cells themselves, ignoring the spatial structure may be a disadvantage. Nevertheless, we stress that a segmentation based on inappropriate spatial information, such as a poorly chosen or poorly parameterised Markov random field [15] is likely to make systematic errors and produce worse results than one which uses no spatial information.

Work by [44], though for a very different application to TV video imagery, tried to extend the thresholding method from grey-level to colour images. Their proposed RGB thresholding method can be summarised as:

1. Apply a grey-level thresholding method to each of the R, G, B colour channels. The chosen grey-level thresholding method was Otsu's method. This results in eight clusters.

2. Calculate the mean for each class (*i.e.* cluster).

3. Calculate the within-class variance and the between-class variance for each class.

4. Two classes can be merged if either of their within-class variances is greater than or equal to the between-class variance.

The above is repeated until only two classes remain. Though attractive this procedure may not always result in maximizing the Fisher discriminant, say, of the final two classes.

This paper also illustrates two other points. First that manual manipulation of several thresholds may be tricky so automatic procedures become more attractive in such situations. If manual processing is essential, colour transformations for example to HSV, Hue-Saturation-Value, may be used to reduce the dimensionality of the feature-space and the number of thresholds required. Tek [175] does just this in his thesis to determine ground-truth segmentations of the RBCs and of the stained pixels. Second it illustrates the problem of over-segmentation in which the image is broken-up into too many, often small and fragmentary, regions. As in the algorithm summarised above region growing processes are frequently subsequently used to try to fix such problems. Similarly, under-segmentation in particular when multiple objects are included within a single segmented region can be a problem. A variety of methods are used to split such regions, many of them, as we shall see in the malaria work reviewed in the next section, tailored to the specifics of the application. Combined split and merge techniques may be used to tackle both problems. However, it is difficult to design such processes that will effectively fix segmentation errors without introducing further errors of their own. In general it is better if possible to design an accurate initial segmentation.

Finally, we recall Tek's critique of global thresholding. As we mentioned in section 1.3.3 Tek ascribes the problem to non-uniform illumination and, in general, uses morphological top-hats to overcome it which leads to a kind of locally adaptive thresholding. Other kinds of locally adaptive thresholding, for example utilising a parametrically defined spatially varying threshold function, have been used in other applications also to overcome problems caused by non-uniform illumination [169]. The Fisher-Otsu algorithm we have developed is applied globally to each of our images but such problems do not seem to trouble us. As already indicated we advocate monitoring this problem and checking for such non-uniform effects within our images. We also note that our approach could if necessary be used in a locally adaptive,'windowed' manner.

2.3 Work on malarial thin-film slide images

Images of malaria infected thin-film microscope blood slides may exhibit a range of appearances which makes them not easily interpretable, whether by a human being or by machine. This may be due to poor image quality, eye fatigue of the expert if a human interpreter, or the subtle nature of the findings required, in particular if the aim is to determine the infecting malarial species and/or the stage of the parasites' development [176, 175].

As described in the introduction and previously, our aim is to provide the research which could underpin building an automatic system to replace the expert counting of RBCs for the determination of parasitemia in the laboratory. Image processing holds the possibility of developing such a system and thereby improving the quality, speed and especially, the consistency of the analyses and simultaneously of reducing cost. Ultimately, there is the possibility of replacing human observers in hospital work or clinicians in the field. Although there is no effective system for the automatic detection of malaria that is sufficiently accurate, is robust to noise and behaves reproducibly, scales linearly, and runs for example on a personal computer, attempts began to be made from around 2000 to develop such systems for the analysis of malarial infection. We focus in this section in particular on the decade of work at the University of Westminster.

2.3.1 Work at the University of Westminster

Research at the University of Westminster has been carried out in the main in collaboration with the NIMR in the course of three PhD studies by Di Ruberto, Rao and, most recently, by Tek. In all of this work the ultimate aim is the computer diagnosis of malaria (CDM) [176]. The work by Tek is the most sophisticated, extensive and complete, addressing the ambitious and difficult aim of identifying the type of malarial infection and the stage of the parasite development within the RBCs. In all three of the above studies, in distinction to our work, there is an emphasis on

the use of morphological image processing techniques. Since such methods are not central to our work, we do not review here the details of such processing techniques used in each study but focus on what was achieved and the limitations of the work.

The experimental results of Di Ruberto *et al.* in their first publication [154] were limited to a single image. They displayed the result of applying the different steps of the system. Their figure, shown here as figure 2.4, however, is not of good quality and does not immediately make it clear precisely what they managed to do. The results of the second publication [153] were again limited to a single image and are similarly difficult to interpret. They did state, however, that: "Equally good results have been achieved on images with different magnification and exposure.". Unfortunately, there was no quantification as to what was meant by "equally good" and the figure they displayed (shown here as figure 2.5) was again hard to view. In their later publications [155, 156], they presented numerical results of processing 12 images of malarial blood. Each image was analysed by two biologists and the system developed. Counts of red blood cells and of immature trophozoites, mature trophozoites, and schizonts (the three stages in the parasite life-cycle within RBCs shown in figure 1.2) were noted. It was observed that the counting results of the two biologists were more different from each other than the counts of the system from either of them. They concluded that their system provided a "good compromise between the two experienced users".

The image used to describe their results was scanned from a colour photograph of stained malarial rodent blood. They also treated the different stages of infections of a particular parasite species the identity of which was not indicated [2]. The system described in [155] included the two systems from the earlier publications [154, 153], *i.e.* for cell segmentation and detection of parasites and also had a third step which involved classification of parasites by analysing the shape of the putative parasite regions after skeletonization or sequential thinning. In their final paper [156] a sample image object was used as a prototype. The system was then required to search the image and retrieve those objects whose colour distribution was globally similar to that of the query object. Though quite different, the two methods of parasite detection by shape analysis and by image search and retrieval were compared and shown to produce, as far as one can tell, very similar and it would seem identical results. Given such very similar results, no attempt was made to combine the decisions of the two seemingly very different detection systems and thereby obtain improved performance. This is useful early work, in particular in our context re the counting of cells to which we shall return later in section 2.3.4 and chapter 6.

[2] This is important because knowledge of the malarial species is required in order to identify the stage of development of the parasite. Nor was it specified if different species were present.

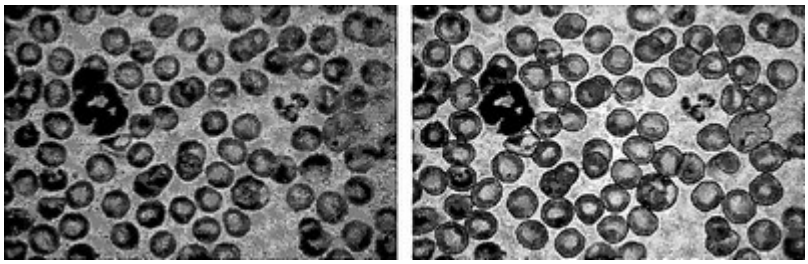


Figure 2.4: The initial sample image to the left and the final output of the system described by Di Ruberto *et al.* [154] shown to the right.

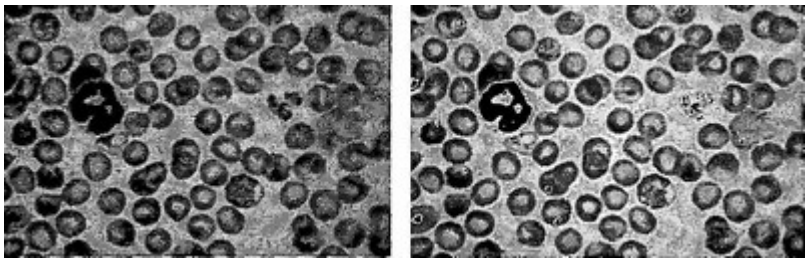


Figure 2.5: The initial sample image to the left and the final output of the system described by Di Ruberto *et al.* [153] shown to the right.

However, in this early work it is clear that evaluation and validation has to be more rigorous. It is also clear from the discrepancies in the analyses by the two biologists that identifying the stages of development is a difficult task .

Within the context of the computer diagnosis of malaria early work at the University of Westminster, appears to have set the strategy of 'detecting stained objects' which was continued in the work of the second research student, K. N. R. Mohano Rao. Rao used the detection of stained-objects as markers for a 'deductive' segmentation approach in which foreground and background regions were segmented, foreground regions were used to obtain individual RBCs and the RBCs finally analysed for the presence of parasites. Since white-blood cells (WBCs) contain DNA they will also (along with other possibly invasive cells and some artefacts) appear as stained-objects. Unfortunately, the presence of WBCs which are very much larger than RBCs was neglected when area granulometry was used to separate the detected objects by size resulting in the possibility of over-segmented WBC regions appearing as imposters for RBCs. The use of area granulometry [147] in preference to a fixed structuring element so that deformed RBCs would not be overlooked may have exacerbated this over-segmentation. A later publication [148] used an iterative segmentation algorithm based on the distance transformation to detect under-segmented clusters of touching or over-lapping RBCs and over-segmented regions by comparison of the area of each region to the average area of regions representing

individual RBCs. The latter was obtained from the main peak of the area granulometry. Unfortunately, according to Tek [175] this may not in Rao's work have prevented regions resulting from over-segmented WBCs being mistaken for RBCs.

Tek himself also makes extensive use of area granulometry and gives a good review of the technique and indicates an efficient way of calculating it [125]. This is a technique we could ourselves have used, for example for calculating the size of regions representing individual RBCs from the initial segmentation resulting from our Fisher-Otsu algorithm. However, since such size segmentation relies essentially on a 1-D histogram, a simpler (and intriguing) possibility was to apply Otsu's algorithm to it. As we shall see in chapter 6 a procedure using such a histogram can be made to work very well.

We noted above that identifying the stages of development is a difficult task even for experts. Indeed, in his later work, Tek [175] argues from the heuristic rules used by human observers that knowledge of the malarial species is required in order to identify the stage of development of the parasite – and vice-versa, *i.e.* for the diagnosis of the type of infection. In machine vision terms the two tasks may thus be expected to be linked in a multi-way classification. Tek therefore considered various classification strategies, the most extensive of which involved twenty classes[3]. Such a system is difficult not only to build but also to evaluate with lots of data required in order reliably to construct the full, 20×20 confusion matrix (see also section 2.5.2). Tek therefore also evaluated the performance of his system when used to perform less detailed classifications. The parasite/non-parasite classifier described earlier by Tek in [177] achieved 74% sensitivity when evaluated based on the acceptance of a single observation of a stained parasite component as sufficient evidence for infection. The paper itself states: "in a real diagnosis scenario a blood slide could provide thousands of stained objects".

Using neural network classifiers, Nicholas Ross *et al.* [151] achieved 85.13% sensitivity for parasite detection. For species identification, they managed 57% sensitivity for *Plasmodium falciparum*, 64% for *P. vivax*, 85% for *P. oval* and 29% for *P. malariae*. Since Ross *et al.* were interested in the diagnosis of malaria from thin-film slide blood smears, they presented a probabilistic model for estimation of the reliability of the diagnosis of malaria infection of a patient from the detection of parasites in the microscope image. This is an important consideration – and one relevant to similar issues in our aim of determining the parasitemia of laboratory sam-

[3]The reader may note that 20 is more than the product of the 4 different species known at the time to infect humans and, since Tek includes the gametocytes (*cf.* figure 1.2), the 4 stages of parasite development. Tek, who segments a 'hierarchy of stained objects' rather than RBCs, includes also classes for: white-blood cells, platelets, artefacts and incomplete samples.

ples – but we note Tek's comment that thick-film slides provide a much more sensitive means of determining the presence or absence of malarial infection [176, 175]. A related issue, discussed also by Tek in his paper and thesis, arises from the fact that only a small fraction of the RBCs may be infected. The positive predicted value (PPV) [176] thus becomes a very important consideration in building a system that is sufficiently sensitive and reliable. Essentially, PPV indicates the probability of a patient being infected given that parasites have been found in the classification of the blood sample. This is especially important in the field where it is highly desirable to detect infection in its early stages when the physical symptoms might not be severe and could be mistaken for flu or other mild illnesses. Again, we need to be aware of similar considerations for laboratory samples which may have a low-level of infection.

2.3.2 Morphology

Morphological operations are used extensively in all of the Westminster work reviewed above. These operations are logical transformations based on comparison of pixel neighbourhoods with a template [89] for processing image region shapes or, as we have mentioned above, in some more advanced algorithms, areas. They are most often used on binary images (binary morphology) but, with suitable extensions of the definition of the structuring element and, in some cases [4] use of grey-level supremum and infimum functions can also be applied to grey-level or even to colour imagery.

The first use of such morphological processing by the Westminster group was in a series of papers [154, 153, 155, 156] by Di Ruberto, Dempster, Khan and Jarra[5]. In the first publication [154] they proposed a system to segment the different cells in images of malarial Giemsa stained blood slides using size information. Size was evaluated using granulometries. A granulometry function can be obtained by a series of openings with increasing structuring element size. The granulometry function is simply a map between each structuring element size and the number of pixels removed during that opening operation. It is an efficient and well-tried way of evaluating the size of image objects. Their second publication [153], in the same conference proceedings as the first, was directed at the detection of parasites using size and colour information extracted by analysing the hue-saturation-value transformed image. They used granulometry for size then regional maxima to detect nuclei of the parasites. A regional maximum is a connected component set of constant level from which it is impossible to reach a point with higher level without descending. As noted above, the third paper [155] described a system that classified parasites by analysing the shape of the putative parasite regions after

[4] When the structuring elements are what is sometimes known as "cylindrical".

[5] Jarra was a member of staff at the NIMR who also collaborated with us, for example in [23].

skeletonization or sequential thinning.

As reflected in the title of his thesis [146] the use of morphological image processing operations was the focus of Rao's work which, as noted above, introduced the use of area granulometry and of a modified distance transform to tackle under- and over-segmentation and culminated in the system described in [149]. Tek built on the work of the preceding students and gives a thorough review of their work in his thesis [175]. Like Di Ruberto and Rao, Tek made extensive use of morphological processing techniques which are reviewed in some detail in his thesis and more briefly, somewhat as an explanatory glossary for non-experts, in the supplementary material to the culminating publication [176]. In addition to pointing out the efficient, fast implementation of area granulometry [147], Tek contributed a new minimum area watershed used in conjunction with area granulometry to provide an initial segmentation of foreground regions from the morphological gradient image and a new marker extraction method based on use of the Radon transform to calculate the centres of labelled regions obtained from the initial segmentation. The centres were used as control markers in a watershed transform used for the final segmentation. Since, unless they are lysed (broken-down) the size of individual RBCs does not vary much [199] and since, unless affected by other diseases such as sickle-cell anaemia, RBCs are roughly circular, these processes help combat under- and over-segmentation. Tek used manual marking of RBC centres obtained from the ground-truth segmentation to evaluate the performance of his segmentation approach and to show that it was superior to Rao's approach based on area top-hats and a distance transform. This is illustrated in Table 2.1 reproduced from Table 4.1 of Tek's thesis but with the methods re-named. A modification in which all stained objects were detected first and excluded from the marker extraction by the Radon transform further improved the performance of Tek's algorithms as shown in the last row of the table.

Method	% correct	% missed	% under	% over & redundant
Rao's approach	86.9	6.4	6.6	5.5
Tek's approach	90.8	6.2	2.9	4.2
Tek's modified approach	91.1	5.6	3.3	3.8

Table 2.1: Results reproduced from Table 4.1 of Tek's thesis showing the percentages of RBCs correctly segmented, those which were 'missed' or not detected, those which were under-segmented, and spurious results from over-segmentation of a RBC or from a redundant segmentation of an object as a RBC where none exists. The evaluation was carried out with respect to a ground truth obtained by manual thresholding. Details may be found in Tek's thesis [175].

We note that Tek has normalised the results so that the first three numerical columns of Table 2.1 approximately sum to 100% with the middle column representing the false-negative rate, *i.e.* the fraction of RBCs not detected, and the sum of the first and third columns the true-positive rate, *i.e.* the fraction of RBCs cells detected, irrespective of whether they were individual RBCs or clusters of two or more touching or overlapping cells. The final column indicates all labelled regions that did not correspond to manually marked RBCs. Such regions may correspond to pieces of over-segmented RBCs or to redundant artefacts which usually appear as small blobs.

The approaches used by the Westminster group, in particular as exemplified by the most recent work by Tek, seem to be the most sophisticated applications of morphological operations in the processing of malarial images. For example, though others such as Ross *et al* [151] use granulometry and the pattern spectrum, they do so only for disc-shaped structuring elements and do not use efficient algorithms or area granulometry. In general, morphological operations are of course widely used in malaria (and other biomedical image analysis) applications *e.g.* [156, 146] for example, for distance transforms, filling and watershed algorithms, and for median filtering [151]. Simple binary morphological closings and openings are used respectively to fill in holes in foreground regions (such as isolated pixels inside a foreground region classified as not belonging to the region) and to eliminate noise (isolated pixels in the background which have been selected as belonging to the foreground but are not in fact part of an object of interest). This is possible in the malaria application because objects of interest such as RBCs are known to be compact and of a certain (minimum) size. The use of such techniques, however as noted by Ross *et al*, is more problematic when applied to the segmentation of stained pixels. For example, though such techniques may usefully eliminate isolated pixel artefacts in the background plasma, they may eliminate very small objects representing very immature parasites and could affect the detailed shape of the parasite region which can be very important in identification of the type of infection.

2.3.3 Region growing

Simple segmentation processes such as thresholding, like many others [168], label individual pixels as belonging to the foreground objects of interest or to the background. In order to count or in general otherwise manipulate the objects of interest it is necessary to group neighbouring pixels in the foreground belonging to one object and label them as a single entity. The classic means of doing so in image processing is by connected components labelling. Region growing is a technique for extracting an image region that is connected based on some predefined criteria

[136]. These criteria can be based on intensity information and/or edges in the image [77]. Connected components labelling is a special case of a region growing algorithm that we shall use. In general, region growing is seldom used alone but usually within a set of image processing operations as demonstrated in the work of Hengen, Spoor and Pandit [79] for the diagnosis of leukaemia. This was because they had to deal with some of the thresholding and region growing techniques' drawbacks. In their work, the thresholding could lead to under-segmentation, in particular with blood cells that happened to touch being segmented as a single object. In general, region growing can be sensitive to noise, causing extracted regions to have holes or to be disconnected when they should be connected, whilst regions that should remain separate might become connected. Hengen *et al.* address these problems, in particular the clustering of neighbouring red blood cells into a single object by using a distance transform implemented efficiently by means of morphological processing operations.

Though most works on malaria applications try to identify objects which represent individual RBCs, often by the use of granulometry, it is particularly important to do so and to separate, or at least detect, clusters of two or more adjoining cells if one wishes to determine the parasitemia. For example, the MalariaCount system developed by Sio *et al.* [165] incorporated a clump splitting method proposed by Kumar *et al.* [101] to detect concavities in foreground regions which would be caused by the touching of adjacent RBCs. Similarly, Halim *et al.* in their work on estimating malaria parasitemia [73] had also to address this issue. Since they used a template-matching approach this involved consideration of the simultaneous matching of a group of RBC templates to an image region. As mentioned above, we shall in chapter 6 use optimisation of a Fisher discriminant to detect clusters of adjoining cells and count them appropriately.

2.3.4 Counting

After segmentation the labelling and thus counting of separate foreground objects is straightforward. Counting the number of RBCs is more difficult and as indicated above requires dealing with under and over-segmentation of the RBC foreground objects. However, segmentation errors are not the only difficulty because, as noted in section 1.3.3, the field of view of each image covers only a small fraction of the microscope slide smear and thus it is almost inevitable that there will be some RBCs near the periphery of the image which are only partly visible. Many such examples can be seen in figures 2.4 and 2.5 from the work of Di Ruberto *et al* and also in figure 1.1 from our own work. As the density of RBCs is higher in the work of Di Ruberto *et al*, there are more cells intersecting the image boundary in their images than in ours. The

potential for error in cell counts due to such effects is thus less in our work but, with ~ 100 RBCs in each 1.3 Mpixel image and, more importantly, often many fewer infected cells this could significantly affect computation of the parasitemia. In particular in our 1300×1000 pixel images with cells typically of diameter 81 pixels, $\sim 27\%$ of the image area lies in such border regions where a cell may be only partially visible. If half the cells in such border areas were independently incorrectly counted, the parasitemia could itself be in error by up to $\sim 19\%$.

There is little mention of such effects in the literature as most work is directed primarily toward the development of computer vision techniques for the detection of malarial infection and determination of the parasite species. Cell counts, when they are made, are often incidental and for other purposes: for example in the work of Tek [175] to characterise the variation in cell density from image to image. Partida's work [132] provides another example where cell counts were made but were not the aim. However, when the aim is to count cells and to determine the parasitemia it is usually essential that effects at the image borders are considered and steps taken to overcome them. Thus, Sio *et al* [165] programmed the software of the MalariaCount system to exclude all cells touching the periphery of the image and in their analysis of a very large number of over 15000 RBCs were able to obtain results for the parasitemia well-correlated with those determined from manual counts over a wide range of parasitemia from below 1% to above 10% though the parasitemia was systematically under-estimated. For example, in one culture treated with chloroquine the under-estimation was by $\sim 6\%$ but in another treated with febrifugine it was only $\sim 0.2\%$. There is no discussion of this difference and analysis of it would require access to the raw machine and manual counts which are not given in the paper or in supporting supplementary information, but it was noted that the greatest difference in the parasitemia ($2.04 \pm 2.86\%$) occurred when cells were poorly separated whilst $0.25 \pm 0.18\%$ was achieved when cells were well-separated.

In the paper by Halim *et al* [73], in which the aim is to estimate the parasitemia, there is no mention of such effects. However, we note firstly that use of a template-matching technique with a threshold on the cross-correlation (or it would seem, on the correlation coefficient) of $\sim 0.65 - 0.70$ suggests that the system should be able to reject RBCs which are significantly outside the field of view though this cannot be verified from the pictures given in the paper. Secondly, the fact that parasite detection is only carried out within regions of successfully detected RBCs means that counts of the number of infected cells will not be independent of the total cell counts and correlation of the two counts will reduce errors in the estimate of the parasitemia. It thus appears that this system may not be badly affected by the presence of cells at the border of the image which are only partially visible. In general, however, if accurate cell counts are

required it seems best explicitly to mitigate such effects at the image borders wherever possible, for example, by excluding all cells touching the image boundary as in the the MalariaCount system.

2.4 Classifiers

Counting the number of healthy and infected RBCs in order to determine the parasitemia as above pre-supposes classification and labelling of cells as healthy or infected. An appropriate review of classification techniques and classifiers is thus required. In fact, in pattern recognition terms, algorithms such as thresholding classify each pixel on the basis of a set of characteristics or properties. These may be as simple as a pixel's grey-level, or more complicated properties such as histograms, colour co-occurrence matrices, moments and other shape measurements which may be calculated over the whole region to which a pixel belongs.

In general, classifier methods are pattern recognition techniques that seek to partition a feature space derived, in our application, from the image. The techniques may be unsupervised, and thus data driven, such as in the Otsu algorithm or maximisation of a Fisher discriminant function, or supervised when data with known labels may be used to train the system and set the classifier parameters, in particular, any decision thresholds [159]. A feature space is the range space of any function of the image, with the most common feature space being the image intensities themselves [136]. Training data for classifiers that are supervised is usually obtained from images that have been manually segmented, if possible by application experts, and then used as references for developing the system, *i.e.* the classifier(s), to be used for automatically segmenting new data. Since the automatic system is to be applied to previously unseen data, it is important in developing such classifier systems that they do not over-fit or adapt too well to the training data. Such over-fitting is likely if the classifier's performance is optimised over the training data alone – known as "empirical risk minimisation" [190]. In order to obtain good performance on previously unseen data, the classifier must generalise well. One way of ensuring this is by using "structural risk minimisation" [190] for the optimisation by imposing a classification margin, or gap separating the classes, and thereby limiting the classifier's capacity and preventing it from reacting to unimportant or chance peculiarities in the training data.

There are a number of ways in which a classifier may be developed in a supervised approach. A simple way is to construct a nearest-neighbour classifier, in which each pixel (say) is assigned to the same class as that of the training data with the closest intensity (say). The k-nearest-neighbour classifier is a generalization of this approach in which the pixel is classified into the same class as the majority of those of the k-closest training data. The k-nearest-

neighbour classifier is considered a non-parametric classifier because it makes no underlying assumption about the statistical structure of the data, in particular as to the functional form of the class conditioned distributions. A fortiori there is no assumption as to the form of the decision surfaces by which the classes are separated.

A commonly used parametric classifier is the maximum-likelihood or Bayes classifier. It assumes that the pixel intensities are independent samples from a mixture of probability distributions, usually assumed to be Gaussian. Classification of new data is obtained by assigning each pixel to the class with the highest posterior probability. If the class conditioned distributions are known the Bayes classifier makes the fewest classification errors. If the data from each class is assumed to be normally distributed, the decision surface for a binary classification will be a quadric. Since fitting quadrics is well-known to be unstable unless the data gives very good coverage it is not surprising that developing such a classifier which will generalise well and perform satisfactorily on new, unseen data is difficult. Furthermore, the higher the dimensionality d of the feature space the worse this problem becomes as the number of degrees of freedom inherent in the mean vectors and covariance matrices ($d(d+3)/2$ per class) indicates. The best results are thus often obtained by imposing special assumptions such as equality of the class-conditioned covariance matrices which makes the decision surfaces linear and therefore much less prone to over-fitting to training data. Such constraints also limit the capacity of the classifier by reducing the number of degrees of freedom.

Recent work by Boray Tek [177] utilised a Bayesian pixel classifier to mark stained pixels. The stained pixels were further processed to extract features (histogram, Hu moments [83], relative shape measurements [175], colour auto-correlogram [85]). A distance weighted k-nearest neighbour classifier [94] was then trained with these extracted features in order to construct a parasite/non-parasite classifier.

A disadvantage of the supervised approach to classifier development is the requirement of manual interaction to obtain training data. Training sets can be acquired for each image that requires segmenting, but this can be time consuming and laborious and, if the training data is to be accurate, should be carried out by an expert with the requisite clinical knowledge. Good training data is therefore often difficult to obtain, especially in large quantities annotated by an expert, as is well-illustrated by the publications to date on the malaria application. Our approach based on unsupervised techniques will enable us to avoid such problems whilst collaboration with the NIMR will enable us to address this deficiency when labelled data has to be used, for example for performance evaluation.

2.4.1 Feature selection

It can be seen from the above that having lots of features input to a classifier is not necessarily the best way to proceed. Certainly it is vital to include features which distinguish the classes of interest but the inclusion of irrelevant features that carry little or no discriminating power is likely to lead to a classifier that generalises less well. Judicious feature selection is thus often of great importance. Features may be selected by systematically adding them one by one and monitoring the resulting classifier performance (called forward feature selection, see for example [178]) or by similarly deleting them one by one (backward feature selection). A combination of such strategies may be used in a mixed approach. Collectively, these are known as "wrapper" techniques since the feature selection is carried out in a process that is aiming to optimise the performance of the classifier. Since such feature selection in general is a hard, *NP* problem, in principle any preferred applicable optimisation algorithm could be used to obtain an approximate solution – see section 2.6.1.

Alternatively, the feature selection might be attempted by use of a proxy measure whose optimisation it is believed will improve the classification performance [97]. One of the most popular of such techniques is PCA, but unfortunately PCA is designed to select the most compact description of the data as a whole rather than a description that is most discriminating for the classes of interest. A better proxy is thus the Fisher discriminant which is based on a measure of class separations. It may be optimised via a generalised eigenproblem that defines a sequence of the most discriminating features. This is also known as "linear discriminant analysis" or LDA. When combined with similar optimisation of a classifier based on such a selection of features this approach is central to our extension of the Otsu algorithm as will be discussed in chapter 4. We only note here that the Fisher discriminant and any resulting classifier based on the feature selection can, if desired, be optimised on a set of training data in a supervised manner.

In modern machine learning which focuses on the use of linear, capacity controlled classifiers such as the support vector machine or SVM via "structural risk minimisation" feature selection and optimisation takes the form of kernel design and optimisation, a technique that enables a set of non-linear feature mappings to be used in a mathematically well-defined manner and is computationally efficient [39]. In general, kernel design is a hard NP problem but thanks to the protection of the capacity limited classifiers against over-learning, standard recipes may usually be followed with good outcomes. Such kernel methods may be applied to non-linear generalisations of PCA (kPCA [160]) and to LDA (KDA [126]) though the latter seems not to lead to a procedure that is computationally efficient.

Finally, we note that from a pattern recognition perspective, image pre-processing and processing algorithms used to transform the input to a classifier should be regarded as feature selection and optimisation processes. These are usually designed either on an intuitive basis or, again according to optimisation of a proxy measure. Thus, we may seek to improve the image appearance to the eye of a human observer, either an expert or non-expert (though the former is obviously preferable), or to extract features such as edges and regions that are similarly intuitively appealing or which, like the filter design in the Canny edge detector, optimise a proxy measure [24, 135]. The ultimate test, of course, is whether such processes lead to a system that works well when the features are input to a particular classifier. In this sense, the pre-processing, image processing and choice of classifier design should be regarded as a single optimal design problem. In this context, image standardisation, either during the acquisition or by post-processing after acquisition, but particularly the latter, should also be regarded as feature optimisation processes. They could also be regarded as a simple type of *transfer learning* in which what works well on a set of test or development data for one purpose – optimisation of the acquisition, appearance, proxy quality measures *etc.* – is applied to all subsequent data for the classification.

2.4.2 Neural Networks

Neural networks [7] were originally developed to mimic biological learning [78], although there is now a well-established extensive literature containing mathematical analyses of the behaviour, performance and construction of a variety of neural network systems and architectures (see for example the book by Bishop and references therein [14]). These texts often describe the relationship of neural networks to pattern recognition problems without reference to biological systems (see for example the book by Duda, Hart and Stork [46]). In addition, much progress has been made in the last 15-20 years on understanding the capacity of neural network based decision classifiers and on developing learning algorithms that, in the supervised case, carry out structural risk minimisation rather than empirical risk minimisation [190]. Together with the development of kernel-based methods, which by use of the representer theorem and the theory of reproducing Hilbert space systems enable the understanding and analysis of linear classifiers to be applied to non-linear classifiers [39], such systems have recently become widespread with many successful applications. Traditionally, learning in neural networks is achieved through the adaptation of weights assigned to the connections between nodes in a multi-layer perceptron by use of the back-propagation algorithm [157]. Thanks to the simplicity and success of this algorithm, for some time now (at least until the development of the SVM and kernel methods,

and possibly still) neural networks have been the default, "black-box", method for constructing classifiers in many applications. They are thus widely used in medical imaging with the weights determined by using training data labelled by an expert or experts.

During the course of this thesis work there has been an upsurge of interest in the use of neural networks in image analysis, in particular for object recognition (see for example the recent review [107] and references therein). This has come about from the combination of two techniques: the use of convolution neural nets (CNNs), and construction of networks with many layers. The former are used to form local feature detectors and, in combination with the latter, to group these in ways insensitive to location and orientation. According to [107] this enables the construction of representations appropriate for object discrimination with the learning of the many weights required facilitated in a 'structural' manner, layer by layer. In spite of this, as noted in the introduction, large training datasets seem nevertheless to be required and graphics accelerator hardware is often used in implementations. Apart from the apparent requirement for large training datasets, the technique would seem well-suited to medical applications.

The use of neural networks was demonstrated by Ross *et al.* [151] for the classification of malaria. In their system, RBCs were first segmented from the plasma background by thesholding the image intensity and those with parasitic infections subsequently detected by setting a second, locally determined threshold within each of the RBC regions following application of morphological operations to remove spurious small objects and to fill holes in RBC objects. The local thresholding was combined with a global threshold by taking the union of the thresholded images and the result used as a marker image for morphological processing like that used by Di Ruberto *et al* though with a different series of morphological operations used to separate individual infected RBCs from clusters of cells. A three-layer multi-layer perceptron network was then used to establish whether a cell is infected or not from features describing the colour and texture of the possible parasite regions. The dimensionality of the feature space was reduced by using PCA to eliminate features that described less than 0.1% of the variation in the dataset. Finally if a cell was determined to be infected, a second similar three-layer network but with four output units, one for each type of parasite, was used to assign the species of malarial infection. Again, PCA was used to reduce the dimensionality of the feature set. Resilient backpropagation with training stopped when the error of a test/validation set began to increase was used in order to reduce over-fitting problems. A good aspect of this paper was the use, where possible – *i.e.* for samples of P. falciparum infection, of holdout data for the assessment of the system's performance (*i.e.* data used neither as the training nor as the test/validation set in adaptation of the neural network weights). Unfortunately, insufficient samples of other species of

infected blood cells were available to enable data on them to be held out of the weight adaptation process and the available samples were divided 50/50 between training and test/validation sets. A total of 350 images were used for the training and test phases with approximately 950 possibly infected erythrocytes (red blood cells). This 50/50 split into training and testing/validation data is a classical approach to the development of a neural network classifier. Nowadays it is likely that repeated 50/50 splits of the data taken, as far as any structure in the data allows, at random would be utilised in order to try to obtain greater accuracy and make better use of the available data. Alternatively, it is likely that n-fold cross-validation would be used with, following Kohavi [99] and depending on any structure in the data, n typically chosen to be $\sim 10-20$. Again, repeated selection of the data-folds may be made to improve accuracy and make better use of the available data.

2.4.3 Combining classifiers

The paper by Ross *et al* is also interesting because it utilises a number of classifier combinations. The basic architecture comprises two, two-stage tree classifiers which we will discuss further below. First however we recall that, as described above, subsequent to the detection of RBC pixels by the first stage intensity thresholding, the union of further local and global thresholding operations was used by Ross *et al* to indicate pixels potentially indicative of parasitic infections. This is itself a classifier combination though not represented explicitly as such in figure 2.6. In fact, this kind of pixel-wise combination is quite popular in image processing and, as noted in section 2.2.2, is frequently used to combine the outputs of separate thresholding operations on each of the colour channels or other attributes of an image. The main weakness of such methods is that the logical combination of outputs seems usually to be *ad hoc* or at best chosen after relatively little exploration and experimentation without any systematic attempt at optimising the system. For example, in Rao's research [146] the grey-level image was used for segmentation, the green channel to extract stained objects, whilst stained mature parasites were obtained from images formed of blue and red (or green) channel differences. The choice of attributes and number of possible combinations may seem daunting although it is sometimes possible [23] to choose an optimal combination of the outputs of separate thresholds applied to each of the R, G, B colour channels for RBC segmentation. In their work on video processing [44] described in section 2.2.2 Du, Chang, and Thouin also sought an optimal combination of outputs from the colour channels though they only considered pair-wise merging according to a variance criterion.

As noted above and sketched in figure 2.6 the first two-stage tree classifier in the system

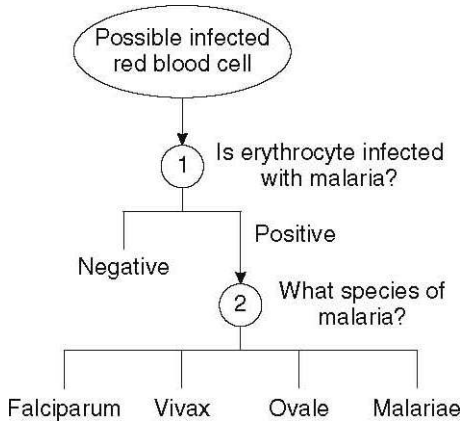


Figure 2.6: An extension of figure 10 from Ross *et al* showing the two-stage tree-classifier architecture used for the cell and infected region segmentation by pixel thresholding which provide inputs to the second two-stage neural net architecture used to determine whether a cell is infected and, if so, the malarial species.

developed by Ross *et al* uses intensity thresholding to segment RBCs and, within them, to detect pixels that may indicate parasite infestation, whilst the second utilise features calculated from those pixels in two three-layer neural networks to determine whether a cell is infected or not and, if it is, the parasite species. Since Ross *et al* are not interested in determining the parasitemia (for which they would have also to segment individual healthy RBCs), they can use the tree-structured architecture sketched above in which only information from the right-hand, positive decision channel of each classifier is passed forward for subsequent processing. This has the benefit of simplicity and importantly allows classifiers of greater sensitivity but reduced specificity to be used on data that has been filtered by preceding processes, *e.g.* for the detection of infected pixels and for the determination of malarial species. However, the performance of each stage in such an architecture is limited by the performance of the preceding stages – in this case since those preceding stages are always binary classifiers, by their true-positive and false-negative rates. At the pixel level, we can see that morphological operations are inserted after each classifier to improve these rates and thus the overall system performance.

The design of an optimal system would not only trade-off the limitations imposed on each subsequent stage by the filtering of preceding stages against the increased sensitivity of the classifiers that can be used but also relax the tree-structure in order to allow more complex decisions to be taken in a richer context. For example, according to Tek [175] classification of the parasite species should be carried out simultaneously with determination of the stage of parasite development in order to help avoid confusing parasite species that at different stages of their

development can exhibit some similar features. Such systems are obviously more complicated than those with a tree-structured or other fixed architecture and their design and optimisation requires the application of powerful algorithms, even when the individual classifiers are treated as fixed in a so-called, multi-classifier system [98]. Thus, for example, genetic algorithms have been used to optimise the combination of classifiers in a voting system [166] whilst genetic programming has been used to search for and to optimise the combination rules for classifiers operating on different feature sets [105] and of different types of classifiers [105] with application to the early stages of drug discovery [104].

More recently, genetic programming has been used successfully to develop multi-stage, 2D computer vision systems for detection and classification for a number of applications using a toolbox within which both the individual classifiers and the combination architecture can be evolved [128].

2.5 Performance measures

We noted in the previous section that the performance of a binary classifier could be characterised by its true-positive and false-negative rates. The terms "positive" and "negative" are used in an abstract manner to distinguish the two types of decision possible whenever a binary, two-way classification is to be made with "positive" usually denoting the class of greater interest. In our context of the analysis of thin-film microscope slides of malarial infected blood "positive" means, for example, that a cell is infected and a true positive decision thus means the cell is correctly identified as infected by the classifier. In such cases TP is the number of true-positive decisions made by the classifier. Alternatively, if the task is to determine the type of infection, "positive" may mean infection with a particular malaria parasite species such as *P. falciparum* and "negative" that the infection is caused by one of the other species. FN is the number of false negative outcomes produced by the classifier, *i.e.* negative, but incorrect decisions. Similarly, FP is the number of false positive outcomes, *i.e.* cases which were negative but incorrectly identified as positive by the classifier, and TN is the number of true negative outcomes. Since the number of cases or examples that were actually positive, P say, is fixed $TP + FN = P$. Similarly, if the number of negative cases is actually N, $TN + FP = N$.

The performance of a binary classifier may be characterised by one of the TP, FN pair and one of TN and FP. Usually, TP and FP are chosen, and their rates, *i.e.* $tpr = TP/P$ and $fpr = FP/N$ plotted as an ROC (receiver operating characteristic) curve within the unit square to describe the classifier's performance. An idealised ROC curve constructed from two classes characterised by a single feature (such as the image intensity) is sketched in figure 2.7.

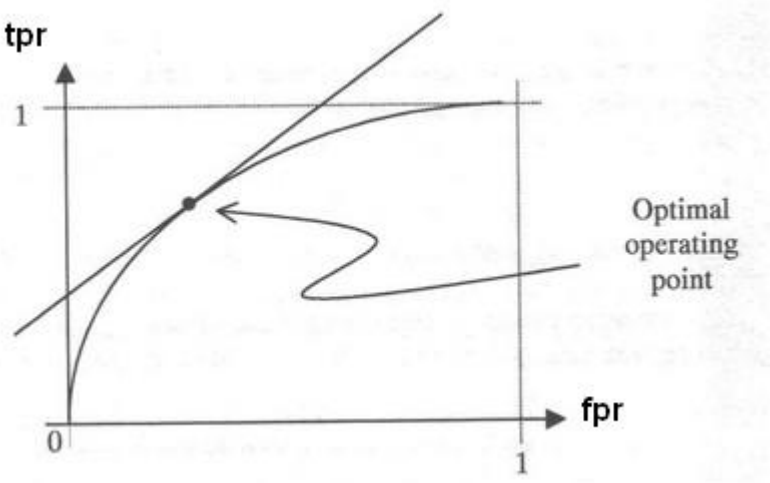


Figure 2.7: An idealised ROC curve from [19] produced when the decision threshold is varied to try to separate two classes characterised by a single feature whose one-dimensional normal distributions for each class overlap.

Sometimes sensitivity, $se = tpr$ and specificity $sp = 1 - fpr = tnr = TN/N$ are used as alternatives and, less frequently, the false positive and false negative rates. In particular, selection of an *operating point* on the ROC represents a particular trade-off between the two types of correct and two types of erroneous decisions that can be made [43]. Ideally the classifier would make no errors with $TP = P$ and $FP = 0$, but this is not achievable. An especially useful property of the ROC curve is that the area under the curve is the Wilcoxon (or Mann-Whitney) statistic and that it may be used as a measure of the performance of the classifier over all possible prior distributions, choices of decision parameters, and the range of possible trade-offs between the values ascribed to the correct decisions and losses associated with the incorrect ones [43, 46, 17, 57]. When combining the decisions of binary classifiers, a convex hull of the individual classifier ROCs elegantly defines the best possible performance that may be obtained in general and is known as the *Maximum Realisable ROC* or MRROC [161]. As the work of Southall *et al* [169] exemplifies, the MRROC may be used even when the classifier combination involves several adjustable parameters. Importantly, only when the decisions of the individual classifiers are logically combined or knowledge of the rationale underlying the decisions or other information, such as the confidence of the decisions (or a proxy for it), is used can the MMROC be bettered [49, 105]. Similarly, concavities in an ROC may be repaired to give performance beyond the MMROC only when certain dependency conditions are met [56, 49].

As Swets [174] indicates, the ROC is an extremely useful tool for characterising and

analysing the performance of a binary classifier though it must be remembered that such a characterisation is only complete and, in particular comparison of classifiers possible, when a measure of the variances is available. Methods have thus been developed in medical applications for averaging ROCs and for obtaining variance estimates from these averages (see for example [49] for a discussion of vertical averaging – which may be used when tp is a function of fp and the latter is under experimental control – and of threshold averaging), or via bootstrap techniques (for example, outlined in the appendix of [127]). These methods of course require access to labelled (ground-truth) data sets which, by their nature, must always be finite. Confidence bounds on the ROC curve and the area under it may be obtained from such estimates [75, 76] but, unlike the area under the area under the ROC, these confidence bounds depend on the underlying class-conditioned distributions. In medical applications, these are usually approximated as normal. Unfortunately, this is a level of sophistication rarely, if ever, to date achieved in image processing and computer vision research where presentation of individual ROC curves is the norm. In addition, there are several factors which limit the use of the ROC in our application.

2.5.1 Limitations

The first and most obvious limitation is that our malaria application is not a two-class problem. At the very least, if we ignore the stage of the parasite life cycle and the type of parasite or as in laboratory work these are known a priori, image pixels belong to one of three classes: background, RBC or stained parasite. In addition, in many cases the RBC membranes are visible (recall figure 1.3) so there should be at least four classes. Furthermore, some pixels will necessarily have a mixed origin giving rise to the possibility of a further three (background+membrane, membrane+RBC, RBC+parasite) or, depending on the resolution, more classes. Even if the presence of such mixed pixels is not acknowledged in the classification it can be important to allow for them in assessing the system's accuracy. If ignored they will give rise to systematic errors in the accuracy estimates [169]. Finally, as already noted the image background may not be solely plasma but also contain WBCs and various artefacts which may reduce performance even if we do not wish to distinguish them[6].

Extension of the ROC concept to a multi-class problem with say, C classes is, in principle, straightforward. Recall that when there are C classes an ideal Bayes classifier would assign an example with feature vector $\mathbf{x}$ to class I say according to whichever of the posterior probabilities $P(J|\mathbf{x})$ was the greatest [43]. Since the operating point of such a classifier can be biased by

[6] See for example [46]; if there are several possible outcomes a multi-category classifier should be used or inconsistent (and therefore, sub-optimal) decisions may be made.

weighting each of the probabilities $P(J|\mathbf{x})$ by some positive scalar $\phi(J)$ which acts roughly in a manner analogous to the inverse of a (positive) threshold $T(J)$ for each class, it can be seen that the classifier has $C-1$ degrees of freedom [103][7]. The key to generalising the ROC is that these $C-1$ degrees of freedom may conveniently be represented on a surface in the space of the C correct classification outcomes characterised by the rates $tr(J)$. In the familiar two-class case this would mean plotting tpr against tnr rather than fpr, but this is a trivial change since $tnr = 1 - fpr$ and the resulting curve is just a plot of sensitivity se against specificity sp which some prefer to use in any case. The area under this surface may then be used as an indicator of classifier performance, bounded above by one and below by $1/C!$ for a random classifier. As in the two-class case, a convex hull construction may similarly be used to construct a maximum realisable performance surface. There are, however, two important differences to the familiar two-class case. The first is that for $C > 2$ there is no counterpart to the equivalence of the area under an ROC to the Wilcoxon statistic.

Second, when $C > 2$ the performance of the classifier is no longer completely characterised by specification of an operating point in terms of the values of the C correct decision rates $tr(J)$. In particular, for each given $tr(J)$, there remain $C-1$ incorrect decision rates, $nr(I|J)$ with $I \neq J$), say, which may take any values satisfying the constraint [103]:

$$tr(J) = 1 - \sum_{I} nr(I|J). \tag{2.2}$$

When there are two classes, specifying an operating point on the ROC via given tpr and tnr determines the decision rates for all four possible outcomes but the incorrect decision rates are no longer completely specified by an equivalent procedure if $C > 2$. At each operating point a complete characterisation of the classifier's performance requires specification of the $C(C-1)$ off-diagonal elements of the confusion rate matrix $\xi_{IJ} = nr(I|J)$. The diagonal elements which are the correct decision rates $\xi_{II} = tr(J)$ are then determined via relationship 2.2 [103]. Thus, when tuning classifier performance, unless the costs of the incorrect decisions $\lambda(I|J), I \neq J$ (and values of the correct decisions, $\lambda(J|J)$) are known[8] [43], the best that can be done when $C > 2$ is to *Pareto* optimize the off-diagonal elements of the confusion matrix $nr(I|J)$ to the point where a decrease in any one of the off-diagonal elements $nr(I|J)$ would necessarily lead to an increase in one or more of the other $nr(I|J)$ [52]. Such a *Pareto*

[7]The argument is equally valid in practice for a classifier which bases its decisions on some imperfect estimate $P_D(J|\mathbf{x})$ of $P(J|\mathbf{x})$.

[8]Given there are C^2 such costs and values it would be rare for them to be known, even approximately when $C > 2$.

optimal solution is said to *dominate* other solutions. The situation is thus more complicated than the simple trade-off between the false-positive and false-negative rates in the two-class case. Fieldsend and Everson describe an evolutionary algorithm for finding Pareto-optimal misclassification rates as the decision vector comprised of the classifier parameters such as the weights $\phi(J)$ and the costs $\lambda(I|J)$ are varied. The resulting set of solutions are said to define the *Pareto front* in the $C(C-1)$-dimensional misclassification rate space. Furthermore it is suggested that the volume of the $C(C-1)$-dimensional hypercube dominated by solutions on the *Pareto front* could be used as a scalar measure of the classifier's performance. Fieldsend and Everson point out that projections of points from the *Pareto front* do not (necessarily) form a surface in the C-dimensional $tr(J)$ space. Thus, they do not necessarily lie on the ROC surface as described here for the multiclass case though they could be used to construct an MMROC surface. Finally, we note that multi-objective evolutionary algorithms were first introduced for optimising binary decision problems in medical imaging by [3] and [102].

There are many classification problems such as the extraction of facts of a particular scientific interest from text [36] where the number of actual negative examples (facts of no relevance to the given scientific context), N, may not be known and very difficult or even impossible to obtain. In image processing problems where for example a binary classification is carried out at the pixel level it is it easy to determine N essentially from the total number of pixels minus P and thus straightforward to construct an ROC although obtaining ground-truth data is time-consuming and tedious. Classification at the object level may thus be more attractive and, in many cases, is more closely related to application requirements [169]. However, in general, classification at the object level in image processing applications raises two issues: the difficulty of enumerating the negative instances and of determining the correspondence between ground-truth data and the output of a classifier – a problem that, amongst other things, is bedevilled by questions of location accuracy and resolution, well-known for example in attempts to evaluate the performance of edge detectors from ROC curves [59].

For the identification of RBCs in the malaria application Tek [175] describes how the latter may be overcome to some extent by marking the centres of RBCs as the ground truth and adopting the conventions used to generate the results reproduced in table 2.1. According to Tek's thesis: a 'hit' or correct outcome is recorded whenever an image region labelled as an RBC by the image processing system contains just one such centre, a 'miss' is recorded if a marker lies in the background region, a labelled region containing more than one ground-truth marker is recorded as *under-segmented*, whilst remaining labelled regions in the image processing output are regarded as *over & redundant* segmentations. Whilst this is a well-defined

procedure, as discussed in section 2.3.2 it does not unequivocally define which outputs should be regarded as 'true-positives' and which should be regarded as 'false positives'. Even if this ambiguity were sorted-out it would still not be possible to define a false-positive rate fpr as determination of N, the actual number of negative instances presented to the classifier, is not possible. The number of connected collections of pixels in the image that are not RBCs is undefined – and probably un-definable. Tek thus does not give any ROC curves describing the performance of this aspect of his system. The best that can then be done in order to characterise the system's performance is to use the sensitivity or recall rate, $se = TP/(TP + FN) = tpr$ (usually referred to in this context as the recall rate, rr) and the precision rate, $pr = TP/(TP + FP)$, neither of which requires knowledge of the number of actual negative examples, N. This is precisely what Ross *et al* did. Unfortunately, precision-recall curves have none of the nice properties of ROC curves. For example, though they lie in a unit square they do not necessarily pass through any of the vertices of the square, there is nothing with properties equivalent or similar to those of the area under the ROC, and precision-recall curves are sensitive to the relative abundances of the classes and thus to the weighting assigned to each class by the parameterisation of the classifier.

As noted earlier (section 2.3.1) when dealing with multi-class problems Tek [175] uses confusion matrices whose elements are just the counts of the numbers of correct and of the various types of incorrect outputs of the system. Since Tek attempts to identify both the malarial species and stage in the life cycle together with various objects (WBCs, platelets, artefacts and incomplete objects) that may have been included in the segmentation of stained-objects the confusion matrices are rather large (20×20) and thus the statistical significance of many of the off-diagonal scores is low – especially as the classes are very unbalanced. This is typical of performance analysis of multi-class problems but at least the problem of unknown incorrect decision rates is avoided. More significance can be attached to the scores when classification of the life-cycle stage is ignored and just the species identification considered in smaller 4×4 confusion matrices.

2.5.2 Practical uses of the ROC curve and other performance measures

The above may seem rather pessimistic but there are a number of very important ways in which ROC curves may nevertheless be used in practice in malaria image processing applications. We have already seen one – characterising the performance of a segmentation process at the pixel level. A second follows from the fact that in the stages of the malaria parasite life-cycle relevant to determination of the degree of infection or parasitemia, the parasite is wholly within

a red-blood cell. In this case, modulo the issues associated with cells in border regions of the image (section 2.5.1), the numbers of healthy and infected cells can be determined by expert observation and thus P and N are known, the rates tpr and fpr may be determined, and an ROC calculated. Tek [175] thus uses ROCs to characterise the performance of his system at both these levels; in particular detecting stained pixels and in "stained-object extraction" used in identifying healthy and infected RBCs. In both cases, Tek uses ROC analysis to help with feature selection and in the latter case also with choosing the most appropriate type of classifier to use. In both these cases, the decisions are dichotomous.

Since they are based on rates, ROC curves are insensitive to class abundances and to the decision costs and values. Flach [57] has thus proposed to extend the conventional two-class ROC to a third dimension so that the effect of class imbalances and of the relative decision costs and values are apparent. This is an interesting proposal but would take us too far afield here. The important point for us is that the conventional ROC enables the developer to focus on regions of the decision space that may be important because of such effects. For example, the part of the ROC where fp is very low is important when the number of negative cases far exceeds the number of positives and the slope of the ROC should be large so that the classifier may approach the ideal of high specificity, $sp = 1 - fp$ and high sensitivity $se = tp$. Since these requirements are in conflict and, as Fawcett points out [49] the performance of the conjunction and disjunction of two classifiers may, if they are independent, be expected to lie outside their convex hull, this is a good reason for combining classifiers.

This kind of reasoning also suggests that it may be advantageous to chain "conservative" classifiers in a sequential multi-classifier architecture but such chaining violates the independence assumption. Fawcett thus points out that the resulting combined classifier may have very different performance characteristics to any of the component classifiers – a fact which itself suggests such combinations may be worth exploring in practice. The results obtained by Southall and Buxton in their agricultural application bear this out [169] as does the increasingly widespread use of a 'forest' of decision trees in which each binary decision is based on a single feature [98, 133]. Since the individual classifiers are binary, their performances may be individually tuned using ROC analysis [169].

We have mentioned in passing in previous sections (*e.g.* 2.3.1) and above other terms for measuring performance such as the precision rate and its equivalent, the positive predicted value $ppv = TP/(TP+FP) = pr$. Tek utilises a variety of such measures [175] including: negative predicted value $npv = TN/(TN + FN)$, the accuracy rate $accr = (TP + TN)/(TP + FP + TN + FN) = (TP + TN)/(P + N)$, and the skew insensitive accuracy rate $accr* =$

$(se + sp)/2$ which is less sensitive to class imbalances. Another measure given by Fawcett [49] is the $F - measure = 2/(1/pr + 1/rr)$. Finally we note that some authors use the *Gini-coefficient G* which is a popular measure in many applications where the distribution of samples is asymmetric or far from normal (see for example the Wikipedia article [197]). For binary classification tasks, G is related to the area under the ROC, $G = 2 * AUC - 1$ (see for example [74] and also [52])[9]. It indicates how well the classifier does in comparison to random guessing and is of course independent of the class abundances. Finally, we note that the ROC provides more information than a net accuracy rate since it characterises both tpr and fpr whilst the effect of erroneous positive and negative decisions may cancel in the calculation of a net accuracy rate. This could be of importance in the estimation of parasitemia.

Although not relevant to our aims, we note that a good aspect of the paper [151] was the attempt to use probability theory to estimate the performance of the system on a blood sample when a large number, in our notation $P + N$, of red-blood cells is examined and the objective is to decide whether the blood sample is from a patient infected with malaria or not. Unfortunately, the attempt by Ross *et al* relied on the assumption that the identification of just one infected cell would suffice to classify a whole sample as infected. If the true-positive rate tpr were one such a conclusion would be valid, but when false negative decisions may be made, this is not so. A more sophisticated probabilistic analysis is required in which, for example, a confidence estimate is placed on the decision for the whole sample and the classification accuracy for a single cell used to estimate how many cells must be examined, or positive outcomes reported, *etc*, in order to reach the requirement [123].

2.6 Geometric and template fitting

Thresholding is the common segmentation scheme. However, there exist many other alternative methods. Garrido and Pérez [64] presented an automatic method using deformable templates. Their templates were built from geometric primitives, in particular ellipses, to represent the shape of the boundary of the cell, fitted to the location of cell boundary pixels obtained by edge detection. Traditionally, edge based segmentation has been divided into two independent stages: edge detection and edge linking. However, a one-way flow of information may yield wrong results because of error propagation due to the fact that information lost at one step cannot be recovered later and accumulation of false information obtained at an earlier step. Hence,

[9] Though the former's "simple generalisation" by averaging the area under the ROCs between all pairs of classes seems to us, unnecessarily complicated and the latter define an ROC in the $C(C-1)$ dimensional space of incorrect outcomes.

Garrido *et al.* proposed a new methodology, dividing the process into three parts: obtaining evidence about cell locations, calculating an elliptical approximation of these locations, and refining cell boundaries using locally deforming models. Visual inspection of their experimental results appears to show that their approach is quite robust but time consuming.

The segmentation performance can be greatly improved by incorporating prior knowledge about the specific type of images being processed. Jiang and Yang [90] also introduced an elliptical cell contour model to describe the boundary of the cells transforming the cell image segmentation problem into an optimization problem. They used images with fairly well separated cells. The results achieved were robust but with no indication of run time.

A better way of building and using flexible templates was developed by Liu and Sclaroff in a series of papers and laboratory reports from around 1997 to 2001 (see for example: [111, 112, 113, 110, 114]). They used region-based template models which may be bent, stretched, sheared, and tapered by means of linear (affine) and quadratic geometric transformations. The 'first' four degrees of freedom of the affine transformation determining the position, orientation and size of the template were usually left free whilst the remaining parameters $\mathbf{a}$ were constrained to belong to a prior distribution $p(\mathbf{a}|J)$ for objects belonging to class J as we have previously labelled them. These distributions were assumed to be normal so that $-log(p(\mathbf{a}|J))$ took the form of an energy term, E_{deform}, quadratic in the deformation parameters $\mathbf{a}$. The distribution $p(\mathbf{a}|J)$ was learnt during a supervised training stage, typically using $\sim 50 - 100$ image objects. If necessary, PCA was used to reduce the dimensionality of the deformation parameters $\mathbf{a}$ in order to avoid over-fitting.

Since their approach was region-based, Liu and Sclaroff used standard region-merging techniques which they referred to [10, 28] in order to produce an initial segmentation. This segmentation was used both for model training – corrected by hand where necessary – and in applications where the objective is to detect instances of the objects of interest in previously unseen images. The template model was deformed to match each region-grouping hypothesis i by minimising an energy cost function $E(i)$ comprised of a weighted sum of: a colour energy, E_{colour} equal to the norm of the colour covariance for pixels within the region grouping; an energy, E_{area}, computed from the region/template overlap; and E_{deform}. This energy $E(i)$ was minimised by a downhill simplex method, accelerated if necessary by a multi-scale approach. Selection of candidate region-grouping hypotheses was subject to local adjacency and edge-strength constraints. The final segmentation was then optimised for global consistency by minimisation of a global cost function constructed from the weighted sum of the $E(i)$ plus a penalty term proportional to the number of region-groupings in the segmentation [111, 112].

A best-first greedy algorithm, simulated annealing and a highest confidence first (HCF) algorithm [27, 93] were used for the optimisation with the HCF algorithm chosen as giving the best trade-off of performance and computational time.

Liu and Sclaroff initially developed and tested their approach using simple geometric shapes [111], images of fruit (in which the objects of interest are bananas), images of leaves and a set of synthetic fish images. They were able to obtain good results on the assumption that the initial region-based techniques over-segmented the images. Application to finding RBCs in microscope (thin-)slide images where RBC counting and morphology (*i.e.* their shape) are important indices for diagnosing certain pathologies, however, showed that their initial approach was inadequate in two respects. In their application, as is frequently the case in malaria applications:

(i) neighbouring RBCs may touch and sometimes overlap slightly and thus, as we have seen, may be under-segmented, and

(ii) computations were time consuming and thus application to the detection of a large number of RBCs in an image was unsatisfactory (*cf.* also the similar lengthy computations in the MalariaCount system [165]).

The final three articles referred to above addressed these two problems: first by using the quality of the template fitting to region-groupings to develop a split and merge technique [113], and second by using index-trees to speed-up the template-fitting [110] which was typically taking 90% of the computer time and subsequently [114] to speed-up and extend object retrieval. There are many important details in the three articles including, for example, in the template fitting:

1. the use of perceptual constraints to select hypotheses for the splitting of under-segmented regions – which may be compared to Kumar's clump-splitting technique used by Sio *et al* [165] as discussed in section 2.3.3;

2. comparison of the fitting energies $E(i)$ for competing grouping hypotheses i before and after splitting – which may be compared with the way template-matching was used also by Halim *et al* [73] as also discussed in section 2.3.3;

3. recursive splitting and a final use of the HCF algorithm to re-merge any regions which global consistency indicated should not have been split;

4. evaluation of the split and merge approach by computing the detection rates and the mean of template fitting costs $E(i)$.

In the development and use of the index trees, notable points included:

1. the eventual use of two index trees [114], one for the image segmentation introduced in [110], and one for similarity-based retrieval. Both could be pre-computed off-line and the retrieval tree obtained by re-organising the segmentation tree;

2. subtle changes to the weightings used in the fitting and global consistency energies which were finally defined in [114] as:

$$E(i) = \alpha E_{color} + (1-\alpha)[(1-\beta)E_{area} + \beta E_{deform}], \tag{2.3}$$

$$E = (1-\gamma)\sum_{i=1}^{n} r_i E(i) + \gamma n; \tag{2.4}$$

3. use of hierarchical clustering based on normalised central moment shape features to construct the segmentation index tree from a uniform sampling in the bounded deformation parameter space;

4. use of a three-layer neural net to provide a mapping from the feature space of seven central moments to the template-fitting cost $E(i)$ (or more simply, E_{area}), in particular at leaf nodes of the tree in order to preserve accuracy;

5. development of an image query system to retrieve shapes of particular interest by computing histograms of the shape deformation parameters;

6. evaluation by means of precision-recall curves computed for the retrieval of eight types of RBC clustered and classified according to their shape as normal, elliptical, *etc.* and comparison with previous methods of retrieval based on colour histograms and the integration of colour and shape information [88];

7. development of an interactive tool and GUI [114].

In the final article, examples were given of the retrieval of normal, elliptical (indicative of ovalocytes which may range in general from slightly egg-shaped to rod or pencil forms), helmet, sickle, and teardrop RBCs. The system could cope with multiple touching objects and a small amount of overlap of RBCs. It was also an impressive achievement in 2001 to build such a system with a response time that was sufficiently quick for it to be used to assist users in finding RBCs in previously unseen microscope images. It is possible an even faster system capable of handling more complicated queries in larger images could be developed using modern evolutional optimisation algorithms such as differential evolution (DE) or a self organizing

migrating algorithm (SOMA) which have been shown to be effective in object detection by flexible template matching and better than a simplex optimisation [207]. It is also notable that Zografos used pixel-wise matching of an object's colour, a prior on the object size or scale, and took care to ensure a proper Bayesian statistical interpretation of the whole image. This requires including not only foreground objects as Liu and Sclaroff do in the definition of the global consistency energy, E (equation 2.4), but also the background as pointed out by Sullivan and Blake *et al* [172].

2.6.1 Evolutionary search

Conventional local search minimization techniques are time consuming and tend to converge to whichever local minimum they first encounter. The key requirement of any global optimization method is that it should be able to avoid entrapment in local minima and continue the search to give a near-optimal final solution whatever the initial conditions. Hence, Jiang and Yang developed an efficient optimization algorithm by combining evolution into tabu search to develop a method which not only has the ability to find the global optimum, but also retains advantages of both tabu search and genetic or evolutionary algorithms [90].

Genetic algorithms have, in fact, been used some time ago for flexible template location and fitting in image processing and computer vision, in particular in one of the early active shape model (*ASM*) papers by Cootes and Taylor [82]. Although, modern evolutionary algorithms such as differential evolution (*DE*) and SOMA (*self organizing migratory algorithm*) mentioned in the previous section, [130, 171, 204, 203] have been used for some difficult, high-dimensional, template-based object detection, location and recognition problems [22, 207], multi-resolution techniques usually based on the use of image pyramids are much more common and, in particular are routinely used in the *ASM* model of Cootes and Taylor [32]. However, we note that such multi-resolution techniques are not guaranteed to converge to a global optimum and frequently may need good initialisation. Furthermore, our problem has the peculiarity, for example when we need to detect healthy red blood cells, of there being many objects of interest in the image when problems of the kind noted in [22] may arise (see section 3.3). The index tree approach of Liu and Sclaroff discussed in detail above offers an interesting potential solution.

Finally, we recall that Giemsa staining gives the parasites a characteristic blue colour. Use of colour information should therefore not be overlooked, perhaps even when trying to locate healthy cells, and methods based on colour histograms (and bihistograms) [173, 120, 191] might therefore be useful, for example in order to provide starting points for a multi-

resolution algorithm, or perhaps for a hybrid multi-resolution/evolutionary method. If colour-based techniques are to be used, selection of colour attributes, use of chromaticity, and colour space normalisation and transformations, such as the grey-world transformation, comprehensive normalisation of Finlayson *et al*, and Procrustes colour alignment are again all possibly relevant. However, to be useful such algorithms would need to detect objects with a high tpr and low fnr and, unfortunately, the former is rarely given in publications and, as noted in section 1.3.4, the latter inaccessible. Furthermore, the rather low values of the precision and retrieval rates (given for example in the experimental comparisons carried out by [114]) with the colour and shape method of Jain and Vailaya [88] do not generate confidence that such an approach would be likely to be successful.

2.7 Summary and Conclusions

As indicated in the opening sentence, in this chapter we have reviewed literature relevant to the processing of microscope images of thin-film slides of blood samples in malaria applications. This has included a brief discussion some wider aspects of the processing of other thin-film microscope images from other biological applications to set the context, a section on image pre-processing and low-level processing – the latter in particular for segmentation by thresholding which will be most relevant for the approach we have explored in greatest depth. A separate section was dedicated to work on malarial thin-film slides with subsections on work carried out over several years by the group at the University of Westminster (2.3.1) and on the morphological and region-based segmentation techniques used extensively by them, and by others (2.3.2). Since we are ultimately interested in determining the parasitemia or degree of infection determined as the fraction of RBCs which are infected, a subsection (2.3.4) was included on counting the number of RBCs and the number of those that are deemed to be infected.

This was followed by a review of classification techniques with subsections on feature-selection, the use of neural network classifiers in malaria detection, and on combining classifiers as several decisions are usually required in the detection of RBCs and as to whether parasites are present or not. Characterisation of a classifier or, more generally, of any system's performance is vital and a number of measures were reviewed in the next section. This began with and concentrated on the familiar ROC curve and included its limitations, some discussion of its extensions to multi-class problems (as the malaria problem always properly is), and its practical use in such problems. The final section described geometric and template-fitting approaches, in particular the extensive work by Liu and Sclaroff which included and eventually focussed on a closely related application for the detection of RBCs and characterisation of their shape as an

indication of certain pathologies. Since model-based approaches involve a search for a best-fitting model or template or optimisation of model parameters a brief review of evolutionary search methods was included.

The main conclusions from the above may be summarised as:

(i) Computer vision researchers need to be aware of the potential variability of microscope images of biological thin-film samples, both between images from different laboratories, from different slides, from different parts of the same slide and even within one image. Where possible images should be provided by expert biological users and standardised as far as possible to avoid extraneous, extrinsic variability.

(ii) The intrinsic variability and complexity of many microscope images of thin-film biological samples can nevertheless be very high (recall for example the image reproduced in figure 2.1) and even though in some applications (recall figures 2.2 and 2.3) the complexity and variability is not so high and in particular in our malaria application it is even less, it is nevertheless apparent that reliable, automatic processing is likely to be quite difficult.

(iii) In this context, the processing of Giemsa stained thin-film slides for malaria applications seems to be one of the simplest such problems. In particular, RBCs usually have both a characteristic colour and shape, especially if they are healthy. However, there are still a number of difficulties to be overcome in the accurate segmentation and detection of RBCs: there may be some extrinsic and intrinsic variability in their appearance, they may touch each other and even in thin-film images occasionally overlap a little, they may not be completely visible within the field of view, and they may be confused with white-blood cells and other artefacts, in particular if the latter are not properly segmented from the background blood-plasma.

(iv) The characteristic colour and shape, in particular of healthy RBCs, means that model-based techniques may be used to detect them. The work of Liu and Sclaroff seems to be the most successful and well-developed model-based approach. However, the way they took shape variation, which is essential in their application for diagnosis of certain pathologies (not including malaria) , into account seems unnecessarily complicated given the flexible shape and appearance modelling techniques already developed by that time ($\sim 1997 - 2001$) by Cootes and Taylor and their colleagues.

(v) The appearance, and especially the size and shape, of a malarial parasite in a Giemsa stained thin-film slides is highly variable. As noted by Tek in his thesis and review article

this depends on the malarial species and on the stage of development of the parasite in a manner that can only properly be addressed as a complicated multi-class classification problem. The only general simplifications are that the Giemsa stain produces a very characteristic dark blue colour from the parasite DNA (whilst healthy RBCs of course do not contain any DNA) and that the infection is *inside* the cell and almost always due to only one parasite. In our application concerned with determining the degree of parasitemia of laboratory samples for research purposes we have the additional simplifications that the infecting species and stage of its life cycle are usually known.

(vi) The variability in the appearance, size and shape of the parasite within infected RBCs makes a model-based approach difficult. In common with most work on the processing of microscope images of thin-film biological samples, neural nets have thus often been used to detect infected cells though the choice of features to be used is then a problem. One of the best pieces of work is by Tek with a great deal of detail given in his thesis. Another good piece of work was by Ross *et al*. Our review has thus included a section on classifiers with subsections on feature selection, neural networks and, since most of the systems developed for the malaria problem used some type of sequential classifier combination architecture, on combining classifiers.

(vii) The performance of any computer vision system must be assessed and characterised in some manner. Performance measures, in particular use of the ROC (receiver-operating characteristic) curve were therefore included as a separate section in this review chapter with subsections on their limitations and practical use in the malaria problem. An interesting aspect of this was the possible extension of the ROC curve to higher-dimensional multi-classifier problems. Though this was largely an aside, it is evident that awareness of these issues and of recent extensions seems very relevant to much medical (and other) image processing and computer vision.

(viii) In order to construct an ROC curve (or surface), labelled 'ground-truth' data is (usually[10]) required. Since provision of reliable 'ground-truth' data requires images to be annotated by an expert such data is rarely available in medical applications where an expert's time is scarce and expensive. Since such data is needed also for the training of neural network or other supervised classification systems and should be used independently in a holdout

[10]It seems that the EM algorithm may not only be used to combine estimates of 'ground-truth' from different experts, but also to enable a 'ground truth' to be established as an optimal consensus and thus to avoid the need for access to a definitive expert at all. See for example [72].

or cross-validation manner for performance evaluation, its scarcity becomes even more of a problem when these approaches are adopted.

(ix) Apart from some counts of healthy and infected RBCs in some of the early parts of Tek's work [175] the two main early works on determining the parasitemia of malarial infections are the Malariacount system developed by Sio *et al* ([165]) and the template-matching system of Halim *et al*) [73]. Under good conditions when RBCs were well separated the former achieved an accuracy of better than one percent in the parasitemia (section 2.3.4) whilst the latter ultimately achieved precision and recall rates of 92% and 95% respectively.

(x) Much of the image processing of thin-film slide images in the malaria application has been on detecting whether the sample is infected or not and, if infected, by which malarial species. This is not only a difficult problem as indicated in item (v) but also seems a somewhat odd thing to do for, as Tek points out, thick-film slides are a much more sensitive indicator of the presence or absence of malarial infection and, because they contain many layers of over-lapping RBCs, would require rather different processing techniques.

It might seem there are two major omissions from this review chapter. The first, in particular from (iv) above is discussion of the flexible modelling techniques developed by Cootes and Taylor and their co-workers at the University of Manchester. This is such an important and obvious approach for the detection of RBCs that a considerable amount of preliminary work was, as will be described in the next chapter, carried out to investigate its application to the malaria problem. Owing to the variability in the appearance of parasites this approach was not expected to be easily applicable to the modelling of infected cells but as we shall see, at least for the simplest kinds of patch models considered, this approach also did not work very well work even for the detection of healthy RBCs.

Given this failure and the difficulties noted above of characterising the appearance of parasite infections and of using supervised techniques to train classifiers, our main approach was thus to use an unsupervised method based on optimisation of Fisher's dscriminant and Otsu's algorithm [131] for the detection of healthy and infected RBCs. This approach, described in [22] relies on minimal modelling of cells' appearance – namely on their colour in Giemsa stained thin-film slides. The second major omission from this review chapter is therefore discussion of these well-established techniques and related clustering methods. This is deferred to chapter 4 where description of our use of this approach commences.

Chapter 3

Preliminary Work – Flexible Modelling

This chapter mainly describes how we explored the use of flexible modelling techniques developed by Cootes and Taylor for the detection of RBCs in our malaria problem. As noted under item (iv) in the concluding section 2.7 of the previous review chapter and as can be seen in figures 1.1 and 1.3, RBCs appear approximately circular in our microscope images, are approximately the same size and are often of a similar colour, though the latter can vary somewhat especially if a cell is infected with a parasite. The simplest model developed by Cootes and Taylor is one in which principal component analysis (PCA) is applied pixel by pixel to a window or small patch of the image just big enough to contain a RBC as the object of interest. This kind of model was first used by the Manchester group in [31] for a PCB inspection application and described as a statistical grey-level model though, following Cootes' notes [29] and in keeping with the conventions used by Cootes and Taylor, we shall refer to such a model as a "flexible patch model" or *FPM* and to its use for finding objects of interest in unseen (new) images as an "active patch model" or *APM*.

This kind of model seems to have first appeared in modern pattern recognition, computer vision and image processing literature when it was applied to face recognition by Sirovich and Kirby [167, 96] and to have become popular when Pentland and Turk [186, 185] showed that a face recognition system developed using such an approach could out-perform the feature-based techniques that had previously dominated the field. In face recognition, following Turk and Pentland, such models are often referred to as "eigenfaces". They quickly supplanted feature based methods [16] and have themselves been further developed using a variety of statistical pattern recognition techniques including: probabilistic principal component analysis (pPCA) [180]; linear discriminant analysis (LDA) leading to the technique known as "Fisherfaces" [8] and also pLDA [86, 143]; the use of non-linear kernel methods, and of graphical models [14]. For details, see the reviews by Chellappa *al* [205, 26] and especially the recent book by Prince [142]. Much useful information can also be obtained from a number of Wikipedia articles, in

particular that on "Eigenfaces" [196] and other on-line resources such as CVonline [55].

In the following sections we first review PCA prior to briefly describing the flexible patch model *(FPM)* and how it may used as an active patch model *(APM)* to find and locate objects of interest in unseen, new images. We review PCA as a linear modelling approach before discussing flexible patch models in the context of our malaria problem. Particular attention is given to the modelling of healthy RBCs to which the *FPM* seems best suited and its application as an active patch model or *APM* to new image data. A number of conclusions are drawn which suggest that building detailed models is not the best way to address the malaria problem.

3.1 PCA – A Linear Modelling Approach

One way to introduce PCA is to consider a bilinear model in which a d-component signal or data variable, $\underline{x} = \{x_1 \cdots x_d\}$ is represented as:

$$\underline{x} = \langle \underline{x} \rangle + \sum_{k=1}^{d} \underline{p}(k)\underline{b}(k, \underline{x}) \tag{3.1}$$

where $\langle \underline{x} \rangle$ is the mean of $\underline{x}$.

This model is underdetermined. If we write 3.1 in matrix form as

$$\underline{x} = \langle \underline{x} \rangle + P\underline{b} \tag{3.2}$$

then the $d \times d$ matrix P whose columns are the modes $\underline{p}(k)$ and the vector $\underline{b}$ of components can both be transformed by inserting a non-singular matrix T and its inverse T^{-1} in 3.2:

$$\underline{x} = \langle \underline{x} \rangle + PTT^{-1}\underline{b} \tag{3.3}$$

There are thus d^2 degrees of freedom undefined in 3.1. To make the model well-determined we need to impose auxiliary conditions such as:

(i) normalisation $$\underline{p}^T(k)\underline{p}(k) = 1 \qquad k = 1 \cdots d \tag{3.4}$$

(ii) orthogonality $$\underline{p}^T(k)\underline{p}(k') = 0 \qquad k' < k \tag{3.5}$$

(iii) uncorrelatedness $$\langle b(k)b(k') \rangle = 0 \qquad k' < k. \tag{3.6}$$

The total number of constraints above is

$$d + \frac{1}{2}d(d-1) + \frac{1}{2}d(d-1) = d^2 \tag{3.7}$$

so the model is then well-determined.

To show that the model then leads to PCA with the modes $\underline{p}(k)$ defined as eigenvectors we consider the $d \times d$ *covariance matrix*

$$\begin{aligned} S &= \langle (\underline{x} - \langle \underline{x} \rangle)(\underline{x} - \langle \underline{x} \rangle)^T \rangle \\ &= \langle \sum_{kk'} \underline{p}(k) b(k) b(k') \underline{p}^T(k') \rangle \\ &= \sum_{kk'} \underline{p}(k) \langle b(k) b(k') \rangle \underline{p}^T(k') \\ &= \sum_{k'} \underline{p}(k') \langle b^2(k') \rangle \underline{p}^T(k'). \end{aligned} \tag{3.8}$$

On using 3.4 and 3.5 we see that

$$S\underline{p}(k) = \langle b^2(k) \rangle \underline{p}(k) \tag{3.9}$$

which is an eigenvalue-eigenvector problem:

$$S\underline{p}(k) = \lambda(k) \underline{p}(k) \tag{3.10}$$

with

$$\lambda(k) = \langle b^2(k) \rangle. \tag{3.11}$$

It is convenient to note:

(i) The modes $\underline{p}(k)$ depend only on averages of the data so we were able to take them outside the expectation $\langle \cdots \rangle$ in 3.8 above.

(ii) It follows from application of 3.4 and 3.5 to 3.1 that the components:

$$b(k, \underline{x}) = \underline{p}^T(k)(\underline{x} - \langle \underline{x} \rangle). \tag{3.12}$$

3.1.1 Alternative approaches

The outcome would be the same if the modes $\underline{p}(k)$ were *defined* a priori as the eigensolutions of 3.10 when, since S is a real symmetric matrix, we may always choose a set of orthonormal eigenvectors satisfying 3.4 and 3.5. Thus, in such an approach we immediately have 3.12 above and, since the coefficients $b(k, \underline{x})$ are scalars it follows that:

$$\begin{aligned} \langle b(k, \underline{x}) b(k', \underline{x}) \rangle &= \underline{p}^T(k) \langle (\underline{x} - \langle \underline{x} \rangle)(\underline{x} - \langle \underline{x} \rangle)^T \rangle \underline{p}(k') \\ &= \underline{p}^T(k) S \underline{p}(k') \\ &= \lambda(k) \delta_{kk'}. \end{aligned} \tag{3.13}$$

The un-correlatedness of the components 3.6 is then a consequence of the definition of the $\underline{p}(k)$.

3.1.2 The modes

Whichever approach is adopted, the $\underline{p}(k)$ are often referred to in image processing and computer vision as *modes of variation* (or 'modes' for short). In statistics the $b(k, \underline{x})$ are often referred to as the principal components (PCs for short) whilst the $\underline{p}(k)$ are usually referred to as the vector of coefficients or loadings (see *e.g.* chapter 1 of [92]). We will not use the latter terminology for the $\underline{p}(k)$ but, whilst we may use the term 'PCs' as above, following Cootes and Taylor's tutorial [32] we may occasionally also refer to the $b(k, \underline{x})$ as 'parameters'.

The orthogonality of the modes 3.5 is very important since it means each makes a contribution to the data

$$\underline{x}(k) = \langle \underline{x} \rangle + \underline{p}(k) b(k, \underline{x}) \tag{3.14}$$

which is unaffected by the contributions of the other modes. In particular:

$$\langle [\underline{x}(k) - \langle \underline{x} \rangle]^2 \rangle = \langle b(k, \underline{x})^2 \rangle = \lambda(k) \tag{3.15}$$

and we may say that

$$\langle b(k, \underline{x})^2 \rangle = \sigma^2(k) = \lambda(k) \tag{3.16}$$

is the variance of the $k'th$ mode.

3.1.3 Spectral decomposition of the covariance matrix

It follows from subsection 3.1.2 and the orthonormality of the modes 3.4 and 3.5 that the *total variance*

$$\begin{aligned} \sigma_T^2 = tr(S) = \langle [\underline{x} - \langle \underline{x} \rangle]^2 \rangle &= \sum_{k=1}^{d} \langle [\underline{x}(k) - \langle \underline{x} \rangle]^2 \rangle = \sum_{k=1}^{d} \langle b(k, \underline{x})^2 \rangle \\ &= \sum_{k=1}^{d} \sigma^2(k) = \sum_{k=1}^{d} \lambda(k). \end{aligned} \tag{3.17}$$

Furthermore, S being a covariance matrix is positive semi-definite. Its eigenvalues $\lambda(k)$ are therefore non-negative and may be ordered in a non-increasing sequence with the first[1] labelled $\lambda(1)$ and the last $\lambda(d)$. Thus, if it is unique, the first mode $\underline{p}(1)$ describes the direction in the d-dimensional signal space in which the variance is the greatest. Similarly, if it is unique, the second mode $\underline{p}(2)$ describes the direction in the subspace orthogonal to $\underline{p}(1)$ in which the remaining variance is greatest – and so on. Jolliffe (chapter 1) describes this as the 'standard derivation' of PCA and points out (chapter 2) that it "emphasizes that the PCs *explain* succinctly, as much as possible" of the total variance, σ_T^2. As Jolliffe also points out (chapter 2) a similar statement can be made about the whole of the covariance matrix S.

[1] And if, as is usually the case, it is unique, the largest.

To see this, recall from matrix algebra that if we construct the $d \times d$ matrix P whose columns are the eigenvectors $\underline{p}(k)$ as in 3.2 and let $\Lambda = diag\{\lambda(1), \cdots \lambda(d)\}$ be the diagonal matrix whose elements are the eigenvalues $\lambda(k)$ in the usual non-increasing order, then since P is an orthogonal matrix, the covariance

$$S = P\Lambda P^T = \sum_{k=1}^{d} \lambda(k)\underline{p}(k)\underline{p}^T(k). \tag{3.18}$$

Equation 3.18 is known as the *spectral decomposition* of the covariance matrix.

Finally, we note there are a number of criteria that are optimised by PCA. Jolliffe (chapter 2) describes several such properties, both algebraic and geometric, and gives references to a variety of other optimality properties, including of entropy and information measures.

3.1.4 Dimensionality reduction

The above 3.17 and 3.18 strongly motivate the application of PCA in *dimensionality reduction* when a subset of the PCs, $k = 1, \cdots t$ (say) with (hopefully $t << d$) may be used to describe most of the data variation encountered in a problem more succinctly and transparently (*i.e. intelligibly*) than the original signal or data variables $\underline{x}$. This immediately raises the question of how to choose the number of modes t that will suffice. How rapidly the eigenvalues $\lambda(k)$ decrease as k increases and their magnitudes are obviously important considerations but, as we shall see in section 3.2.4, are by no means the whole story.

We also note that the d-modes $\underline{p}(k), k = 1, \cdots d$ span the space of the d-dimensional data $\underline{x}$. They form an orthonormal basis and *any* signal or datum $\underline{x}$ may be represented by the modes as in 3.1 and 3.2 *without error*. If the number of modes is reduced this will no longer be the case.

3.1.5 Calculating the modes and PCs

As Jolliffe notes, the advent of electronic computers and especially the availability of reliable and efficient matrix eigensolvers (and related and other procedures – see later sections 3.2.2 and 3.1.6) made carrying out PCA practicable and the technique is now very widely used. In particular, since the mid-1990s when personal computers became sufficiently powerful routinely to process images, it has been very frequently used in image processing and computer vision – a trend accelerated by its early success in face recognition and by the flexible modelling techniques developed from around that time by Cootes and Taylor at the University of Manchester.

The problem nowadays is not computational but obtaining an estimate of the correlation matrix from which to calculate the PCs and coefficients by standard numerical matrix procedures. In particular, the fact that S may be so easily estimated by a version of 3.8 for a finite

sample dataset of, say, n observations, $\underline{x}(j), j = 1, \cdots n$:

$$S = \langle [\underline{x}(j) - \langle \underline{x} \rangle_j][\underline{x}(j) - \langle \underline{x} \rangle_j]^T \rangle_j \tag{3.19}$$

where

$$\langle \underline{x} \rangle_j = \frac{1}{n} \sum_{j=1}^{n} \underline{x} \tag{3.20}$$

denotes a sample average, often obscures a few relevant points.

First, 3.19 is a maximum likelihood estimate of S (Jolliffe, section 3.7.1) and, though as usual for an MLE asymptotically correct, is biased slightly too small by a factor $n/(n-1)$ (Jolliffe, section 3.1) if, as assumed in 3.20, the mean $\langle \underline{x} \rangle$ is not known a priori. This difference is usually negligible, trivial and of little other than theoretical interest especially as we shall often effectively work with an un-normalised version of S (Jolliffe section 3.1).

Second, it should be emphasized that the data set $\{\underline{x}(j)\}$ used to provide an estimate of the covariance must properly be regarded as a *training set*. It must therefore be representative of the population and free from contamination. It is important therefore to be aware of 'structure' or 'pattern' in the data (Jolliffe, section 3.8) and the data should be free of outliers which can have a disproportionate effect on the PCs (*cf.* Jolliffe, section 10.4 where robust estimation of principal components is discussed). The only safe way to ensure training data is free from outliers is to have an expert select it manually. This can be an expensive and time consuming process, especially in medical applications where experts' time is a scarce resource. The provision of sufficient quantities of good-quality training data is thus often one of the most difficult problems in such applications. As we have already noted in the opening chapters this is the case in the development of computer vision systems for processing malaria thin-film slide images and, as in Tek's work [175], one often finds training data is provided by the computer vision researchers themselves acting as a semi-expert. While likely to be okay for the identification of RBCs, this is a more doubtful procedure when it is necessary to identify infected RBCs and especially if identification of the infecting species and stage of the parasite life-cycle is required. Whether carried out by an expert or a computer vision researcher as semi-expert, it is important to streamline the work as much as possible by provision of software and a suitable GUI so that the manual role is mainly one of checking [175, 176, 110]. In favourable circumstances, the provision of annotated training data might be 'boot-strapped' with the system development as in the early work by the Manchester group in [31] for a PCB inspection application.

In general, as Golub and van Loan (1996, section 7.2, [67]) say, the eigenvalues of a real, symmetric matrix (which is necessarily normal[2]) are not "extremely sensitive" to perturbations.

[2] A matrix M is normal if $MM^T - M^TM = 0$.

However, this is not true of the eigenvectors (of a real, symmetric matrix or any other normal matrix) belonging to an eigenvalue which is degenerate or to two or more eigenvalues which are nearly degenerate. Fundamentally, this sensitivity arises because the eigenvectors belonging to a degenerate eigenvalue are not uniquely defined, even up to the usual sign (or more generally phase) ambiguity, and form an orthonormal set that may be in any orientation within an invariant subspace. For the large eigenvalues labelled in our convention by low values of k such degeneracies or near degeneracies are unlikely unless, for example, the data is structured (*e.g.* due to some underlying symmetry of the application). However, the situation is completely different for the eigensolutions labelled by high values of k when the eigenvalues are likely to be small, similar to each other and close to or even zero[3]. The last may only happen when S is rank deficient which usually occurs when insufficient training data is used in 3.19 with $n < d$. In this case the eigenvectors $\underline{p}(k)$ belonging to $\lambda(k) = 0$ may be chosen at random in any manner consistent with the orthonormality constraint.

In general therefore, we may expect eigenvectors $\underline{p}(k)$ labelled by high values of k to look like random noise and to be sensitive to the selection of the data used in the training set and to any perturbations that may inadvertently have crept into it. Both are good reasons for ignoring such 'high' PCs and for using PCA for dimensionality reduction using only $t < d$ modes and for treating as noise the unexplained variation:

$$\underline{x}_{un} = \sum_{k=t+1}^{d} p(k)b(k, \underline{x}). \tag{3.21}$$

3.1.6 Probabilistic *PCA* and other models

If the noise is assumed to have an isotropic normal distribution with variance σ^2 we arrive at the probabilistic PCA (pPCA) model of Tipping and Bishop [180] in which equation 3.1 is replaced by

$$\underline{x} = \langle \underline{x} \rangle + \sum_{k=1}^{t} p(k)b(k, \underline{x}) + \underline{\epsilon} \tag{3.22}$$

where

$$\langle \underline{\epsilon}^2 \rangle = \sigma^2 = \frac{1}{d-t} \sum_{k=t+1}^{d} \lambda(k) = \frac{1}{d-t} \sum_{k=t+1}^{d} \langle b(k, \underline{x})^2 \rangle. \tag{3.23}$$

According to Tipping and Bishop equation 3.23 is, like all other quantities in the model, a maximum likelihood estimate of the noise level σ^2. In matrix-vector form equation 3.22 may be written as:

$$\underline{x} = \langle \underline{x} \rangle + P\underline{b} + \underline{\epsilon} \tag{3.24}$$

[3] All magnitude measures may, for example, be taken relative to $tr(S)$ or some other matrix norm.

where P is the $d \times t$ rectangular matrix whose columns are formed from the first t modes $\underline{p}(k), k = 1 \cdots t$ and projects into the space of the explained variation. It is important to note that this is essentially a *best-fit* model and that, because the noise is assumed to be isotropic, no signal or datum $\underline{x}$ is now assumed to be represented exactly by the modes $\underline{p}(k), k = 1 \cdots t$ even in the space of the explained variation.

Roweis [152] introduced a slightly more general model in which the error covariance matrix $\langle \underline{\epsilon}\underline{\epsilon}^T \rangle$ may be allowed to take any form. In both works Roweis and Tipping and Bishop describe how an EM algorithm [42] may be used to fit these models to a dataset. Roweis emphasizes the computational advantages of using an EM algorithm of complexity $O(dnt)$ when the dimensionality d of the data or the number of training examples n is very high (and most especially when both are high, but see section 3.2.2). Jolliffe [92], section 3.9, gives other reasons for the introduction of the probabilistic PCA model; in particular that a Bayesian treatment of the model [180] may facilitate determination of the number of modes required, t (see also [11] and [12] and that use of the EM algorithm allows extension to the imputation of missing data values and to data drawn from mixtures of distributions).

Formally, the matrix-vector form of pPCA appears similar to factor analysis, though the constraints or assumptions used to make factor analysis unique differ from those, (i) – (iii), that we used in section 3.1. Joliffe (chapter 7) describes the differences and stresses (section 7.5) that whilst PCA concentrates on explaining the diagonal elements of the covariance matrix S (recall maximisation of the variances in section 3.1.3), factor analysis, unless specific factors are introduced, concentrates on finding common factors relating two or more components of the data variables $\underline{x}$, *i.e.* on explaining the off-diagonal elements of S.

3.2 Flexible Patch Models in the malaria application

In a flexible patch model the image intensities or, more usually nowadays, colours $\underline{x}$ are represented pixel-by-pixel by an equation of the form 3.1 over a region of fixed size and shape. A rectangular region of $P \times Q$ pixels is most often used in patch models, but in our malarial application, a square region with $P = Q = 81$ can be used to enclose individual RBCs which are usually approximately circular and do not vary much in size (figure 3.1). Use of a bounding box rather than an inscribed grid as in Cootes's illustration of the patch model for face recognition [29] means that in addition to pixels belonging to the RBC a fraction $\sim 20\%$ of the pixels in these square regions lie in the background which is usually blood plasma. The variation of such pixels whether due, for example: to small variations in RBC size or deviation from a circular shape, to RBCs touching each other or an artefact in the plasma, or to variation in the

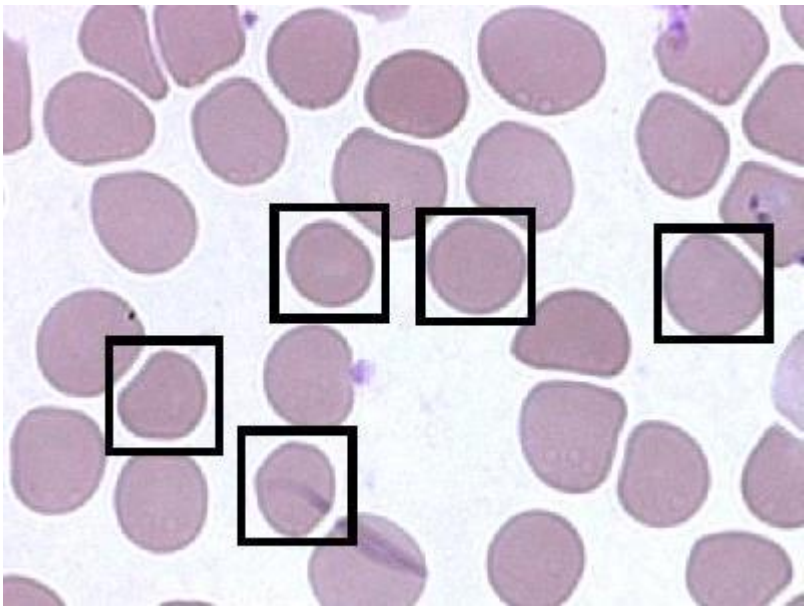


Figure 3.1: A blood smear image in which 81×81 pixel bounding boxes have been placed over a number of RBCs.

appearance of the background plasma, will be included in the patch model.

This obvious weakness of the model in comparison for example to a flexible appearance model is offset by its simplicity. In particular, no landmarks are required to define the boundary of a RBC. Furthermore, our patch model was especially simple as we did not find it necessary to place the bounding boxes with sub-pixel accuracy or to vary their size and thereby avoided the need to interpolate intensities or colour attributes between pixel centres [29]. In addition, for greatest simplicity, we concatenated the pixel RGB colour values directly into a vector $\underline{x}$ of dimension $d = 3PQ$ without making use of any linear (strictly, affine) colour transformation or histogram equalisation procedure to allow for global variations due to microscope illumination or staining effects as discussed in section 1.3.3. Any such variations thus also became an intrinsic part of our patch model.

3.2.1 Selection of a training set of healthy RBCs

A GUI was developed to facilitate selection of RBCs in an image, placing of bounding-boxes over them, and extraction of the vector $\underline{x}$ of pixel values via a raster scan. Even with such a tool, creation of a training set with over $3PQ = 19683$ examples would have been tedious and difficult. Fortunately, as is typically the case in image processing and computer vision applications [20] the patch has many fewer degrees of freedom and a much smaller number of training examples n may be used. Pixel values within the RBC (especially if it is healthy) are highly correlated, as to a lesser extent are pixel values from the small parts of the patch outside

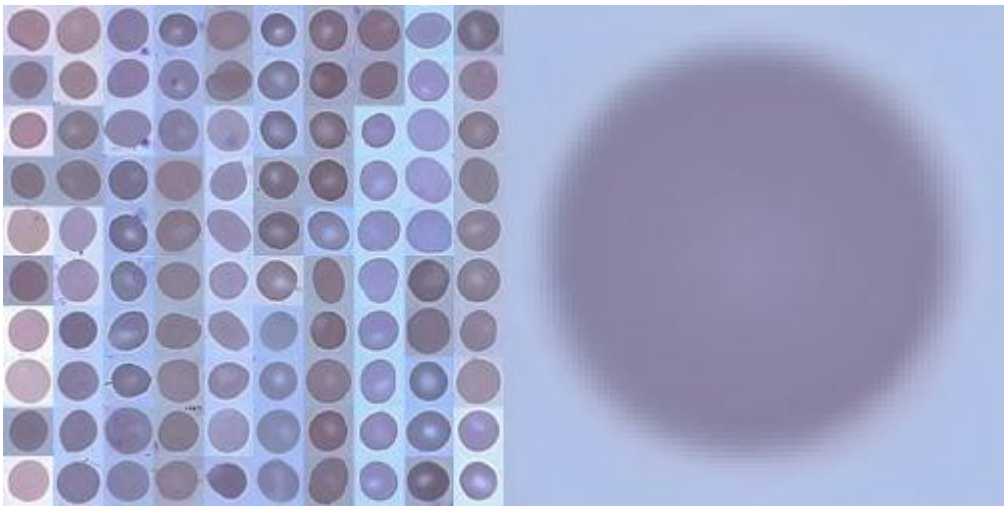


Figure 3.2: A montage of the training set $\underline{x}(j)$ of 100 patches each containing a healthy RBC and, on the right their mean $\langle \underline{x} \rangle_j$.

the cell especially if they are plasma background. A training set of 100 patches each containing a healthy RBC was therefore constructed using the GUI (figure 3.2). It can be seen from their different colours that the healthy RBC were selected from several of the images in our dataset.

3.2.2 Use of the *SVD*

In principle, the long vectors $\underline{x}(j)$ of dimension $d = 3PQ$ could be used to construct a huge $3PQ \times 3PQ = 19683 \times 19683$ covariance matrix S as in 3.19 and its eigensolutions calculated. However this is unnecessary. A covariance matrix S is a very special kind of real, symmetric matrix and the eigensolutions required can, in such cases where $d > n$, be generated much more efficiently. To see this, we first form a $d \times n$ data matrix X whose columns are comprised of the deviations of each of the training samples $\underline{x}(j)$ from the mean:

$$x_{ij} = x_i(j) - \langle x_i \rangle_j. \tag{3.25}$$

The covariance matrix may thus be written from 3.19 and 3.20 as $S = \frac{1}{n} X X^T$ and equation 3.10 for the modes $\underline{p}(k)$ and eigenvalues $\lambda(k)$ as:

$$S\underline{p}(k) = \frac{1}{n} X X^T \underline{p}(k) = \lambda(k)\underline{p}(k). \tag{3.26}$$

Pre-multiplication of 3.26 by X^T yields

$$X^T X X^T \underline{p}(k) = n\lambda(k) X^T \underline{p}(k). \tag{3.27}$$

From this it can be seen that effectively we have only to solve an eigen-equation of much lower dimensionality $(n \times n)$ defined by the number of training samples n:

$$X^T X \underline{q}(k) = n\lambda(k)\underline{q}(k) \tag{3.28}$$

where we have introduced the vector:

$$\underline{q}(k) \propto X^T \underline{p}(k). \tag{3.29}$$

Moreover, provided $\lambda(k) \neq 0$ the first n modes $\underline{p}(k)$ may in principle be obtained, up to normalisation, from the $\underline{q}(k)$:

$$\underline{p}(k) \propto \frac{1}{n\lambda(k)} X\underline{q}(k) \text{ provided } \lambda(k) \neq 0. \tag{3.30}$$

The remaining $d - n$ modes $\underline{p}(k)$ for $k > n$ (and any that it may turn out for smaller k also have zero eigenvalue) are degenerate with zero eigenvalue $\lambda(k) = 0$ and could be constructed from the orthogonality requirement 3.5, but as discussed in section 3.1.5 these are uninteresting as they represent noise.

Obviously, solving 3.28 requires much less computational effort (and storage) than solving 3.10 – even if in the latter case only n modes $\underline{p}(k)$ were calculated. However, there is a simpler, yet more direct route; namely to recognize that 3.29 and 3.30 form a *singular value decomposition* or *SVD* pair [67] which we write as:

$$\begin{aligned} X^T\underline{p}(k) &= \sqrt{n}\sigma(k)\underline{q}(k) \\ X\underline{q}(k) &= \sqrt{n}\sigma(k)\underline{p}(k) \end{aligned} \tag{3.31}$$

so that $\sigma^2(k) = \lambda(k)$ is the variance of the $k'th$ mode as defined in 3.10 and 3.11. In matrix form, the *SVD* is usually written as [67]:

$$XQ = P\Sigma X^T P = Q\Sigma^T \tag{3.32}$$

where P and Q are respectively $d \times d$ and $n \times n$ orthogonal matrices and, in our case following 3.31:

$$\Sigma = diag\{\sigma(1), \sigma(2), \cdots \sigma(m)\} \tag{3.33}$$

with $m = min\{n, d\}$, and

$$\sigma(1) \geq \sigma(2) \geq \cdots \sigma(r) > \sigma(r+1) = \sigma(r+2) = \cdots \sigma(m) = 0 \tag{3.34}$$

where $r \leq m$ is the rank of X. In our case with $n < d$, $m = n$. We shall often ignore the $\sqrt{n}$ factors in equations 3.31 above which – as noted in section 3.1.5 – is equivalent to working with an un-normalised version of S.

Finally, we note that in imaging applications such as ours when $n < d$ equations 3.1 and 3.2 can no longer be exact unless the high modes $\underline{p}(k)$ with $k > n$ are included. Since these modes belonging to eigenvalues $\lambda(k) = 0$ represent noise this would seem to be a good reason for preferring the probabilistic *pPCA* model mentioned in section 3.1.6.

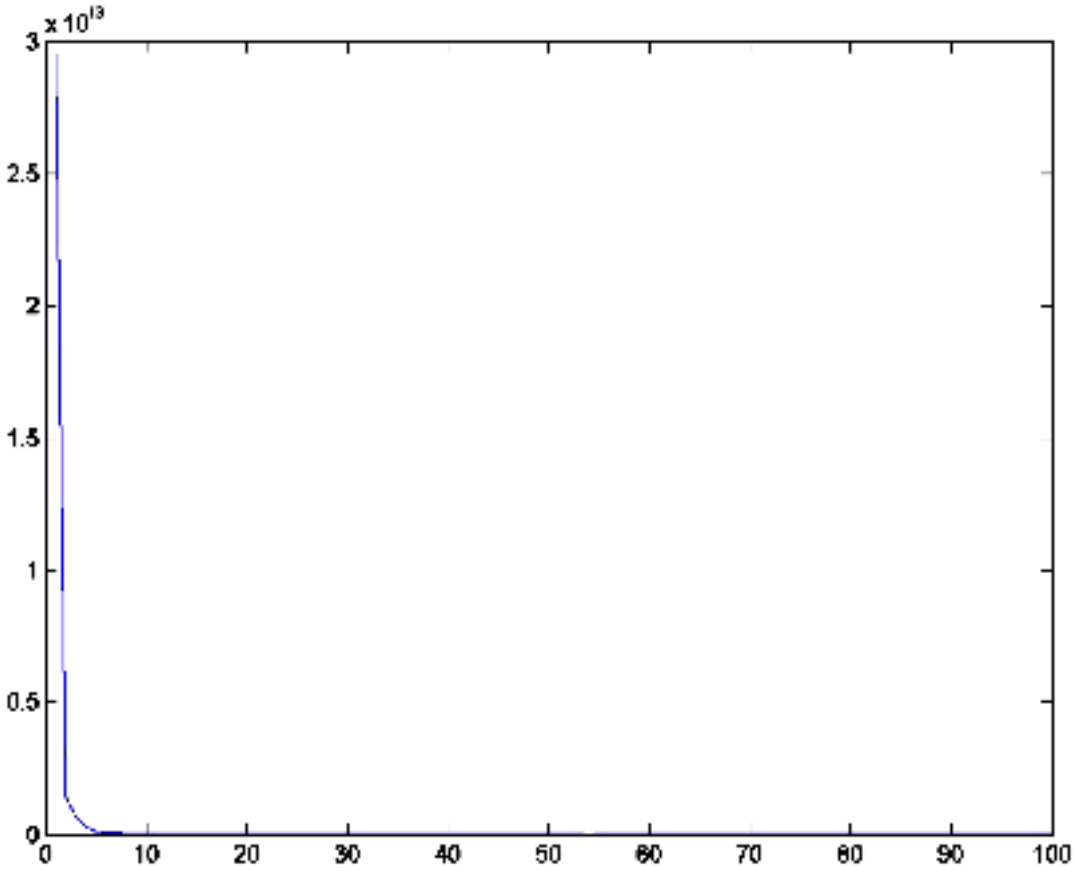


Figure 3.3: The variances (y-axis) obtained from the *SVD* of the un-normalised data matrix X for the training set of 100 healthy RBC image patches shown in figure 3.2 as a function of the mode number (x-axis).

3.2.3 Modes for representing healthy RBCs

Matlab was used to calculate modes $\underline{p}(k)$ by SVD of the $d \times n$ data matrix X constructed from the $n = 100$ healthy RBCs selected as a training set as described in section 3.2.1. The standard deviations $\lambda(k) = \sigma(k)^2$ decrease rapidly as k increases as shown in figure 3.3 with a small number of modes accounting for most of the total variation $\sigma_T^2 = tr(S)$.

It is useful to display the modes $\underline{p}(k)$ as image patches

$$\underline{x}(k) = \langle \underline{x} \rangle + \alpha(k)\underline{p}(k) \tag{3.35}$$

with the pre-factor $\alpha(k)$ chosen so that the variation of each mode is visible. Figure 3.4 contains a montage of the modes displayed in this manner. If desired, $\alpha(k)$ may be chosen proportional to $\sigma(k)$ to show the relative strengths of the modes.

It is immediately apparent from figure 3.4 that image patches corresponding to modes with low values of k are highly structured especially around the almost circular cell boundaries and show distinctive colouration. In modes with higher values of k the boundary variation

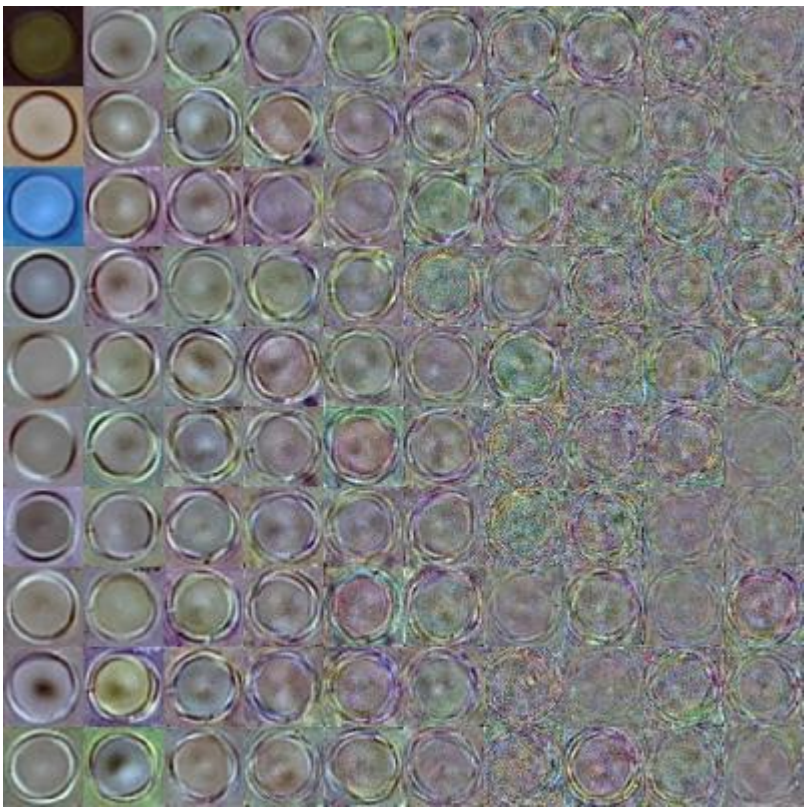


Figure 3.4: The modes $\underline{p}(k)$ displayed as image patches $\underline{x}(k)$ as described in the text. First column, top to bottom: $\underline{p}(1) \cdots \underline{p}(10)$; second column: $\underline{p}(11) \cdots \underline{p}(20)$; *etc.* to $\underline{p}(100)$.

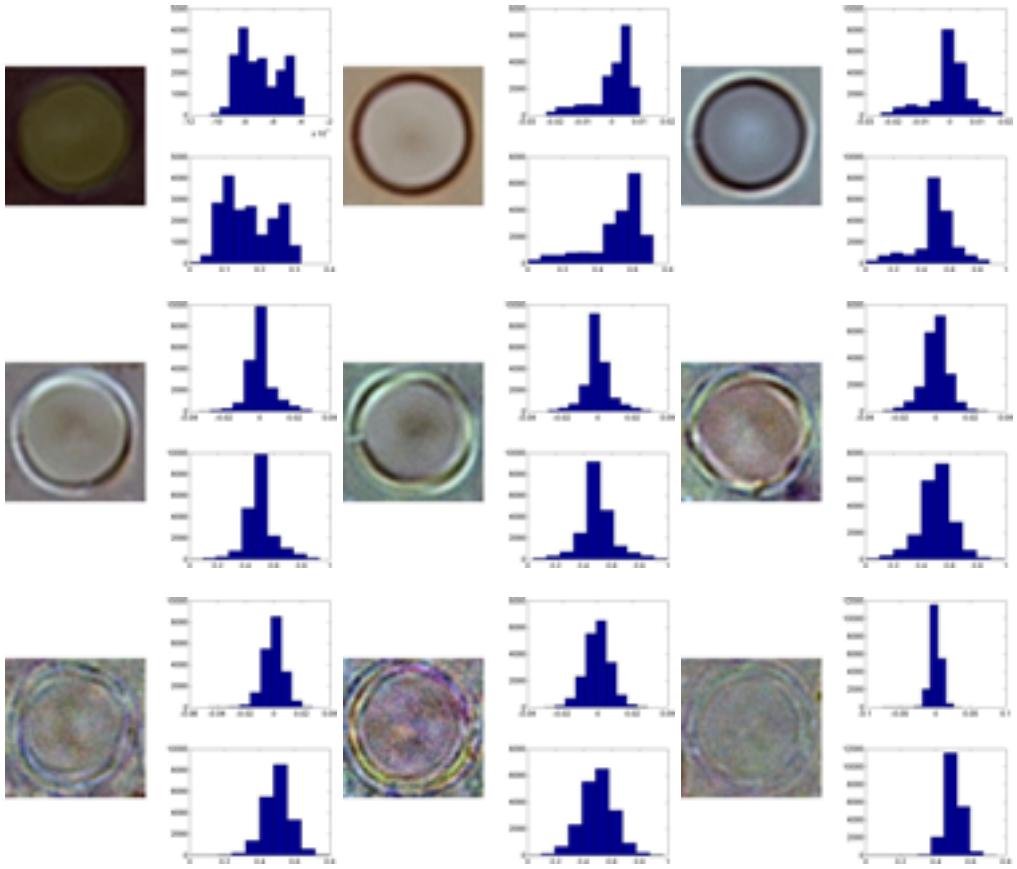


Figure 3.5: Selected modes $\underline{p}(k)$ for $k = 1, 2, 4, 8, 16, 32, 64, 80, 96$ displayed (top-left to bottom-right) as image patches $\underline{x}(k)$ as described in the text and the caption to figure 3.4 with alongside each histograms of the elements of the mode $p_i(k)$ (upper) and of the pixel values $x_i(k)$ (lower).

becomes increasingly distorted and noisy, until eventually at high k the whole mode becomes quite noise-like in appearance. This is reinforced by histograms of the elements $p_i(k)$ and of the pixel values $x_i(k)$ shown alongside selected modes in figure 3.5. For the first few modes the histograms are structured (and in particular all the $p_i(k)$ for the lowest mode $k = 1$ have the same sign), but they rapidly (*e.g.* for $k \geq 10$) become symmetrical and to look like zero-mean Gaussian distributions. It is notable that the histograms begin to look Gaussian for much smaller values of k than the patches themselves look like noise. This seems to be due to the structure of the cell boundary though there is also other structure, for example in the centre of a cell, in some modes.

3.2.4 Choosing the number of useful modes t

In image processing and computer vision it has become customary following Cootes and Taylor [32] to use t modes which explain a fraction f of the total variance, *i.e.* such that:

$$\sum_{k=1}^{t} \sigma(k)^2 \geq \text{f}\, \sigma_T^2 > \sum_{k=1}^{t-1} \sigma(k)^2 \tag{3.36}$$

with f chosen as appropriate for the application, but often in the range $0.7 - 0.95$. Choosing a value f dependent on the application may be regarded either as an advantage or as a disadvantage but it is clear that once a value for f has been chosen the procedure is straightforward to apply. It is undoubtedly a convenient 'rule of thumb' but opting immediately to use it ignores that fact that a whole chapter of Jolliffe's book (chapter 6) is devoted to this question and that much work has been done by statistical researchers to try to provide an answer. Those interested are referred to Jolliffe's chapter for details but brief consideration of several of the approaches described is worthwhile.

The underlying aim in choosing t is to use only modes that describe interesting, significant variations in the data and to ignore modes which describe uninteresting variations such as noise or artefacts peculiar to the particular training sample used. This is consistent with the remarks made in previous sections about the modes with high k. However in a pattern recognition application choosing t is less straightforward than it might otherwise seem. Discriminating characteristics may be quite subtle and easily lost if too many of the high modes with $k > t$ are rejected.

In chapter 6 of [92] Jolliffe presents several 'rules of thumb', describes attempts to develop statistical tests that may be used to calculate t, and discusses briefly partial correlation, 'bootstrap' and cross-validation methods. Many of the statistical methods are criticised as based on over-simplistic (often Gaussian) assumptions, for often being more appropriate to factor analysis than PCA, and for being unreliable in that they can fail in certain circumstances (*i.e.* may lead to the rejection of significant, important modes) and frequently produce estimates of t which are unrealistically large or unrealistically small. Bootstrap and cross-validation methods tend to be based on low-rank approximations to the data matrix X or to the covariance matrix S (recall the spectral decomposition of equation 3.18). With the possible exception of some jackknife methods, these are too complicated and computationally intensive to be useful, as well as still being potentially unreliable and in some cases more applicable to factor analysis than PCA. In practical applications it seems we are left with the 'rules' of thumb.

The first rule is to choose the number of modes t that explain a fraction f of the total variance as in 3.36 above. If an independent estimate of the noise level σ^2 is available, this could be

improved by adopting the probabilistic, pPCA, point of view and choosing t such that the average of the unexplained variation over the rejected modes with $k = t+1, \cdots n$ when $n \leq d$ is consistent with the expected noise level of a patch , σ_P^2 , *i.e.* such that:

$$\sum_{k=t+1}^{n} \sigma(k)^2 \leq (n-t)\sigma_p^2 < \sum_{k=t}^{n} \sigma(k)^2. \tag{3.37}$$

In imaging applications, it would be ideal to estimate the noise level of a patch σ_P^2 from images taken at different times but otherwise of exactly the same scene. Unfortunately such images are rarely available, even in applications where successive images in a video sequence may be taken, so other estimates have to be used as a fallback. For example, in the malaria application one could try to estimate σ_P^2 from the background, plasma region but even if done manually this is not ideal as the background itself must be modelled satisfactorily and imaging noise may have multiplicative characteristics rather than being purely additive.

The second rule discussed by Jolliffe focuses on the magnitudes of the variances of the PCs, *i.e.* on the magnitudes of the eigenvalues $\lambda(k) = \sigma(k)^2$. A simple approach following Kaiser's rule for PCA of correlation matrices would be to choose the cut-off λ_c for the PCs to be retained according to the average of the eigenvalues, $\bar{\lambda}$. For a full rank covariance matrix S this average would be:

$$\bar{\lambda} = \frac{1}{d}\sum_{k=1}^{d} \lambda(k) = \frac{1}{d} tr(S) \tag{3.38}$$

but if S is rank-deficient with rank $r < d$ (and, in particular in image processing and computer vision when rank $r \leq n << d$) it could be argued that it would be preferable to take:

$$\bar{\lambda} = \frac{1}{r}\sum_{k=1}^{r} \lambda(k) = \frac{1}{r}\sum_{k=1}^{r} \sigma^2(k) = \frac{1}{r} tr(S). \tag{3.39}$$

Jolliffe [92] and [91] suggests that a lower cut-off at $\lambda_c = 0.7\bar{\lambda}$ may be more appropriate in order to avoid 3.38 or 3.39 selecting too few PCs which, given the rapid decrease of the $\lambda(k)$ with k shown in figure 3.3, seems likely in our application.

Finally, we note that Jolliffe mentions the "broken-stick" model in which the arbitrary cut-off $\lambda_c = 0.7\bar{\lambda}$ above is replaced by (for a full rank matrix S):

$$\lambda_c = \frac{\bar{\lambda}}{d}\sum_{k=t}^{d} \frac{1}{k} \tag{3.40}$$

with $\bar{\lambda}$ as in 3.38. It represents a kind of 'parallel analysis' (see below) in which the $\lambda(k)$ are being compared with the expected magnitude of the t th largest segment obtained were $tr(S)$ broken at random into d segments. If S is of rank $r < d$ it would seem reasonable to replace d by r in equation 3.40 and to use equation 3.39 instead of equation 3.38.

The third rule is concerned with the rate of decrease of the $\lambda(k)$ as k increases and requires judgement of when $\lambda(k) - \lambda(k+1)$ stops being large. This depends on both the relative magnitudes of $\lambda(k) - \lambda(k+1)$ and its predecessor $\lambda(k-1) - \lambda(k)$ and on their absolute values. It is obviously motivated by the fact that as we have seen the $\lambda(k)$ are expected to become almost constant at high k if they represent noise, but this doesn't specify how to decide when $\lambda(k) - \lambda(k+1)$ is not large. Regarding a plot of $\lambda(k)$ against k as a *'scree-graph'* and looking for the 'knee' on the curve where it stops being steep can be highly subjective, although for graphs like that shown in figure 3.3 it is possible that most researchers might choose a similar cut-off t.

One way of making this less subjective and even of potentially automating it, is to look for the point on the curve where its slope is equal to the slope of the chord $(\lambda(1) - \lambda(r))/(r-1)$ or, more simply and more conservatively since more modes would be retained, to $\lambda(1)/d$. Should these yield values of t explaining similar fractions of the total variance one might be confident that a reasonable cut-off had been specified. If there are several points where the slope is equal to the chord a conservative choice to take the largest resulting t can be made. Another way would be to carry out a 'parallel analysis' of the eigenvalues of a (suitably defined) random matrix and to compare the two.

Other methods described by Jolliffe include testing the hypothesis that successive eigenvalues are equal commencing with the last two that are non-zero (Bartlett's test) or its reverse commencing with the two largest eigenvalues (Jackson's test), cross validation (and bootstrap) methods, and a partial correlation test. Cross-validation, bootstrap and jackknife approaches are essentially looking at how well the retained modes can approximate data that was not in the training set. This is a reasonable criterion, but unfortunately the goodness of fit has itself to be specified and we are returned to the kind of issues that confronted the first rule and, except for some jackknife methods (see Jolliffe [92], section 6.1.5), are confronted with vastly more computational work. Similarly, the level of significance has to be specified in methods that utilise hypothesis testing. The only other method Jolliffe discusses (section 6.1.6) which avoids this kind of difficulty is Velicer's partial correlation criterion in which one looks for a minimum of the average V of a particular squared partial correlation measure.

Unfortunately, Jolliffe describes Velicer's method as most suited to factor analysis of correlation matrices rather than to PCA. It would seem therefore that only the first three rules described above are worthy of much consideration, with the first having the obvious merit of being the simplest. It is also the most widely used in computer vision and image processing applications and, though it would seem interesting to explore the other two, is the only rule

used in this work.

3.3 Active patch model or *APM*

In this section we review how the PCA flexible patch model (*FPM*) may be used as an active patch model (*APM* for short) in our malaria application. Cootes covers similar ground in his presentation [29] though we do not need to utilise all the steps Cootes describes, in particular since we have made no attempt to align the training data in colour space nor to align RBCs spatially to sub-pixel accuracy. The procedure is based on the same principles used in many of the Cootes and Taylor flexible models and their corresponding 'active' models, from the original *FSM* and *ASM* for shape, the *FTM* and *ATM* for texture (usually used within a mean shape), to the *FAM* and *AAM* for appearance [32].

In essence, in the active modus operandi the model is used to find an instance or instances of the object of interest in an image. This is of course most useful when the images are of previously 'unseen' objects and 'hits' representing possible instances of the object of interest declared where the fit to the underlying image patch is good even though, as noted in section 2.6, this is not a statistically correct approach [172].

3.3.1 Fitting the *APM* to image data

In this section we explore how well the model, constructed from the most significant modes as discussed above, fits or describes image data. In principle the patch model should be scanned across an entire image and the goodness of fit recorded for each position of the centre (say) of the patch. However, as described by Buxton and Zografos [22] this would result in a very large, complicated error surface so, in order to focus in detail on what happens, the centre of the *APM* was scanned over a 31×31 pixel square resulting in a search window 111×111 pixels centred on a RBC or some other selected part of an image. The sum of squared error of fit of the *APM* to the image data within this window was calculated according to:

$$\underline{\epsilon}^2_{PCA} = [\underline{x} - \langle \underline{x} \rangle - \sum_{k=1}^{t} \underline{p}(k) b(k, \underline{x})]^2 = (\underline{x} - \langle \underline{x} \rangle)^2 - \sum_{k=1}^{t} b(k, \underline{x})^2 \tag{3.41}$$

and displayed as a surface over the 31×31 pixel search square using Matlab (figure 3.6). In 3.41 we note that $\underline{x}$, $\langle \underline{x} \rangle$ and the modes $\underline{p}(k)$ are long vectors of dimension d. The scalar products thus automatically imply a sum over all pixels in the patch and over the colour channels. The mean $\langle \underline{x} \rangle$ is estimated as in 3.20.

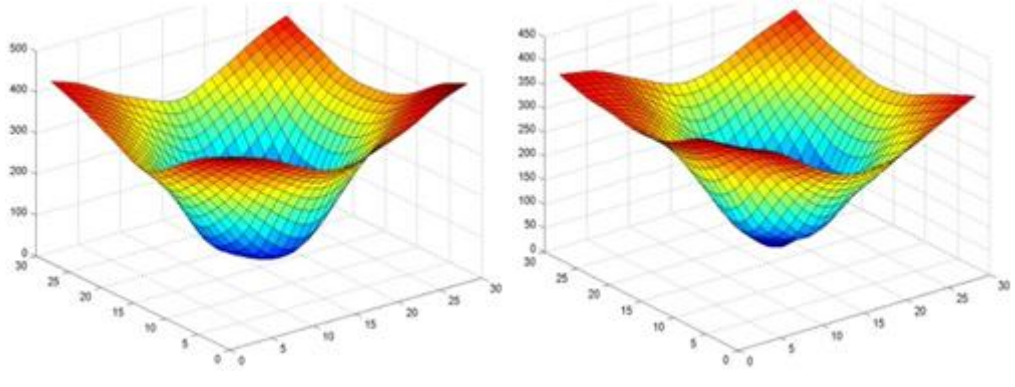
A RBC used in training was chosen so that there were no others or artefacts nearby in the image. The squared error (SE) surface shown in figure 3.6 thus has the expected bowl-shape with a minimum at the centre of the search square when the model template lies precisely over

the cell in the image. However, this minimum is rather broad so setting a threshold on the error would lead to many 'hits' for the possible location of the cell as shown in figure 3.6 (d). The cell selected and the model fit to it at the correct location is shown in figure 3.7 (a). Two other RBCs not used in training were similarly chosen, one from an image that contained other cells which were used in constructing the training set and one from an image not used at all in training. The error surfaces for these cells are shown in figure 3.6 (b) and (c) with the corresponding cells and the model fits to them in figure 3.7 (b) and (c). These images and the quantitative results in figure 3.6 illustrate that the model fits the training data and other healthy RBCs with little error. In particular, the results in (b) and (c) in figures 3.6 and 3.7 are similar to those obtained with an isolated cell used in training (see (a) in the figures) though the SE is roughly twice as large for an isolated cell in an image that was not used for training (figure 3.6 (c)).

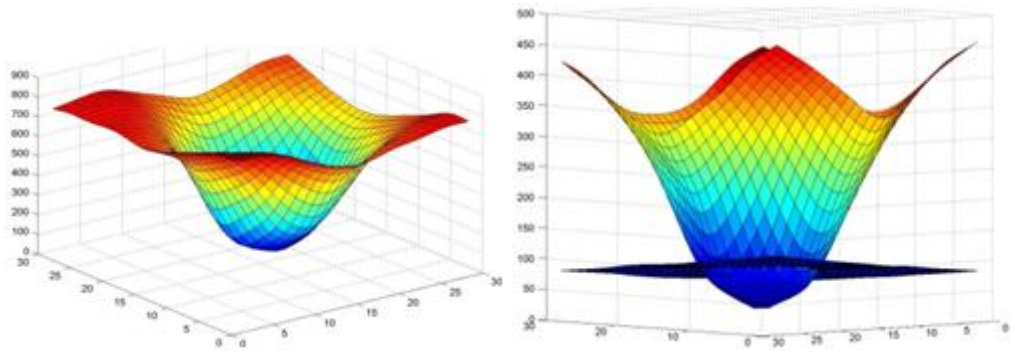
The results in figure 3.6(d) indicate that non-maximal suppression would be needed to locate the cell to pixel accuracy whilst interpolation could be used to obtain the location to sub-pixel accuracy. However, before an *APM* could be used in this way to detect and locate healthy RBCs in an image, we need to know how it performs when located over other regions of an image, in particular the background. Figure 3.8 (a) and (b) show respectively the error surfaces as the *APM* is scanned over windows placed in background (plasma) regions of an image used in training and one not used in training whilst (c) shows the error surface when the model was scanned over a window centred on a part of a cell.

The results are qualitatively different. None of the error surfaces in figure 3.8 have well-defined, bowl-shaped minima like those in figure 3.6. For the background regions they are also quantitatively different with a range of errors from $\sim 80 - 85$ for the background region from a training image and $\sim 130 - 135$, for a background region from an image not used in training. It was because of the former that the threshold in figure 3.6 (d) was chosen as ~ 80 but it is apparent that this threshold is *below* the minimum in figure 3.6 (c) obtained for the best fit of the *APM* to a healthy RBC in an image that was not used in construction of the training set. The *APM* used in this manner would thus not be able reliably to detect RBCs and would generate many false negatives (*i.e.* many cells would be missed). Raising the threshold to reduce the false negative rate would generate many false positives.

It is an open question whether introducing a colour alignment step in the model building and in application of the active model would improve matters. Aspects of the colouration of each image would then become extrinsic to the flexible model and fitting the active model would have additional degrees of freedom. The flexible model itself would thus become more specific – making it more likely that a suitably discriminating threshold might be found – but the

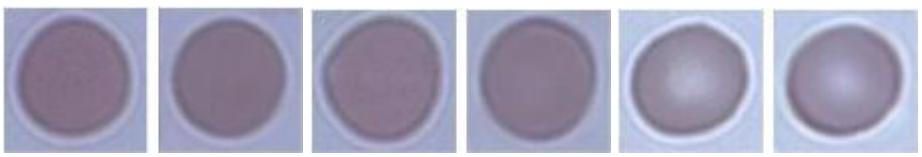


(a) (b)



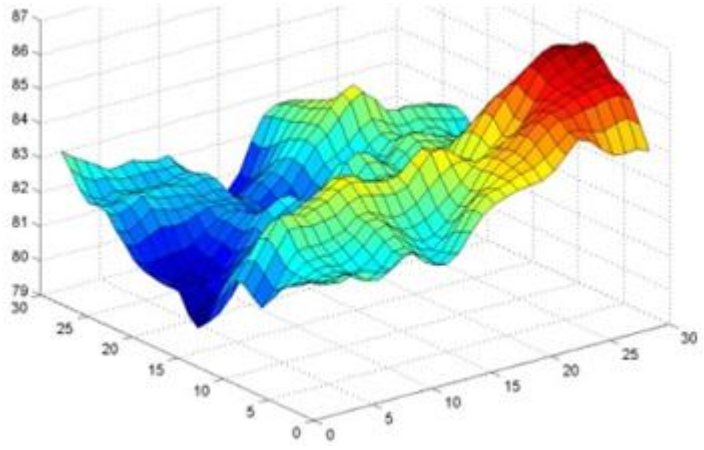
(c) (d)

Figure 3.6: (a) The SE surface obtained when the *FPM* is scanned as a flexible template or *APM* across a window centred on a RBC selected for training; (b) the surface obtained when the *APM* was scanned across an isolated cell in an image used in training but where the cell itself was not selected as a member of the training set; and (c) the surface obtained when the *APM* was scanned across an isolated cell from an image not used in training. (d), the same error surface as in (a) showing the many 'hits' for the possible location of the cell resulting when a threshold ~ 80 is chosen.

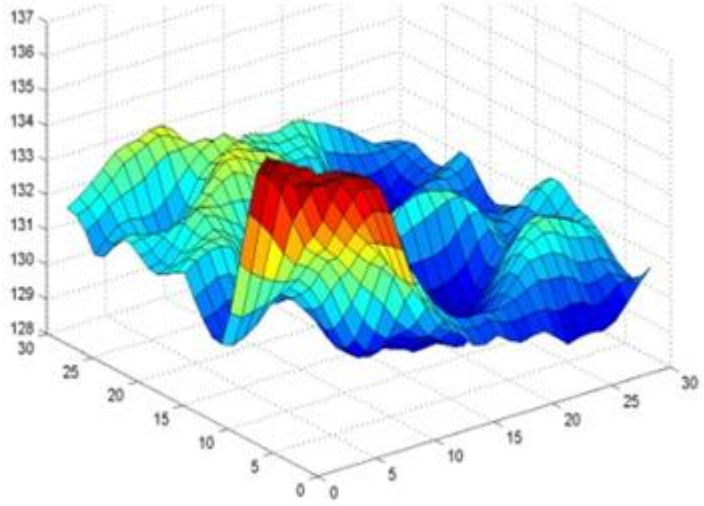


(a) (b) (c)

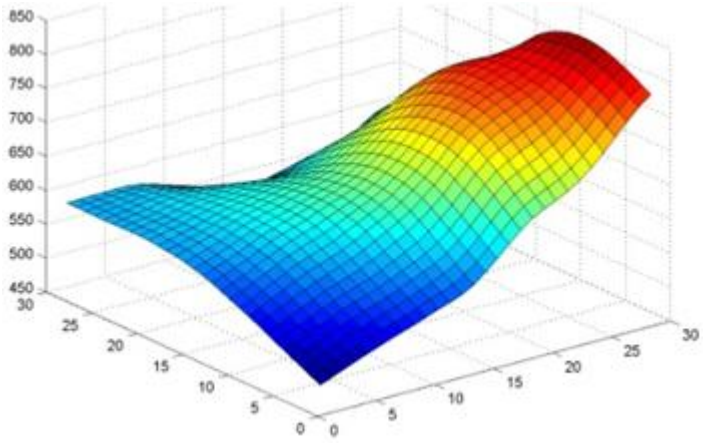
Figure 3.7: (a): Left: the cell from the training set used to construct the error surface shown in figure 3.6(a) and, right, the model fit to it. (b): As in (a) but with a cell not selected in the training (left) and the model fit to it (right) corresponding to figure 3.6(b). (c): Similarly for the cell used in figure 3.6(c) and the model fit to it. Note the lighter background.



(a)

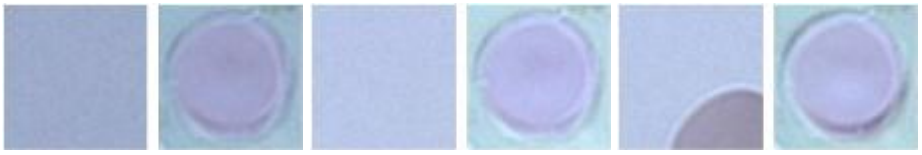


(b)



(c)

Figure 3.8: The SE surfaces obtained when the *APM* is scanned across a window centred on: (a) a background region of an image used in training; (b) on a background region of an image not used in training, and (c) on a region containing a part of a cell from an image not used in training.



(a) (b) (c)

Figure 3.9: (a): Left: A background, plasma region taken from an image used in training used to construct the error surface shown in figure 3.8(a) and, right, the model fit to it when the patch was accurately located over this region. (b): As in (a) but from an image not used in training corresponding to the error surface in 3.8(b). (c): Similarly for a region that contained part of a cell from an image not used in training corresponding to the error surface in 3.8(c).

danger of over-fitting the model and extrinsic parameters in the active modus operandi would be increased which might reduce the discriminating power and increase the false positive rate. It can be seen from figure 3.9 that the active model 'hallucinates' the appearance of a RBC when there is really no cell present in the input region. This may be expected to happen whenever the mean colour levels within a patch $\underline{\bar{x}}$ are different from the average colour over the patch of the model mean $\langle \underline{\bar{x}} \rangle$ because, as noted in section 3.2.3, all components of the first mode $\underline{p}(1)$ are of the same sign and thus

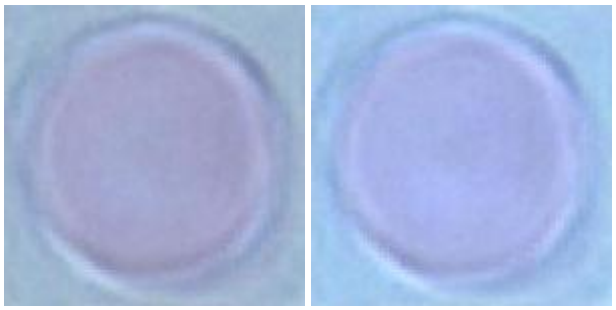
$$b(1, \underline{x}) = \underline{p}(1)^T (\underline{x} - \langle \underline{x} \rangle) \tag{3.42}$$

will be non-zero. It can be seen from figure 3.10 that extending the *FPM* to include a locally adaptive additive colour bias with the principal components given by:

$$b(k, \underline{x}) = \underline{p}(k)^T [\underline{x} - \langle \underline{x} \rangle - (\underline{\bar{x}} - \langle \underline{\bar{x}} \rangle)] \tag{3.43}$$

may slightly reduce the tendency of the active model to 'hallucinate' RBCs in a background region but does not eliminate the effect entirely. In the above $\underline{\bar{x}}$ denotes the colour averaged over the patch and $\langle \underline{\bar{x}} \rangle$ denotes it averaged also over the training set.

The 'bottom-line' is that the active model is adaptive and may be expected to share the tendency of such models to adapt and fit well to background regions of an image as noted by Buxton and Zografos [22]. In fact, according to Buxton and Zografos an adaptive template model will only grossly fail to fit well to an image patch if the patch contains some features *not* corresponding to the object of interest. They point out this is most likely when the patch in the input image includes part of an object of interest and parts of other things, background or other objects, and will generate a 'rim' in the error surface where the model fits poorly around the correct location of an object. This effect is consistent with the results above. The largest errors in the background regions (~ 85 or ~ 135 depending on whether they are from images



(a) (b)

Figure 3.10: (a): The model fit to the plasma region shown in figure 3.9(a) when a locally adaptive additive colour bias was included. (b): As in (a) but with the background plasma region from figure 3.9(b). It can be seen that the colours are slightly different in the two cases.

used in training or not) are much smaller than the largest errors ($\sim 400 - 450$ and ~ 700 respectively) in figure 3.6 (a)– (c) whereas these latter errors are comparable to those in figure 3.8(c). This is borne out by the error surfaces obtained when the *APM* is scanned over larger regions surrounding healthy RBCs as shown in figure 3.11.

3.3.2 The likelihood of an *APM*

Principal component analysis – in particular as it has been introduced and utilised in this thesis – need make no assumptions about the statistical properties of the data or of the modes $\underline{p}(k)$ or, apart from the requirement that they are uncorrelated, of the PCs $b(k)$. We may thus introduce whatever statistical assumptions are appropriate (or are convenient simplifications or idealisations). One of the most widely used is that the significant PCs used in the model with $1 \leq k \leq t$ are normally distributed according to:

$$p(\underline{b}) = \prod_{k=1}^{t} \frac{1}{\sqrt{2\pi\sigma(k)^2}} e^{-\frac{b(k)^2}{2\sigma(k)^2}} \tag{3.44}$$

where $\underline{b}$ represents these PCs as a t component vector. As usual the variances are given by the eigenvalues of the covariance matrix S calculated in training (recall equations 3.9 – 3.11). We note that for use in 3.44 the eigensolutions must be calculated from a properly normalised covariance matrix or its equivalent. Since the PCs are given by 3.10, 3.44 amounts to assuming that the modelled parts of the data $\underline{x}$ are normally distributed but, of course unlike the PCs, correlated. We also note that 3.44 says nothing about the distribution of the high PCs with $k > t$ not retained in the model. Thanks to the fact that each PC may, according to 3.12, be computed independently of all other PCs we are free to assume that *outside the model subspace* they are, for example, normally distributed as noise[4].

[4]In image processing and computer vision where the number of training examples is usually considerably less

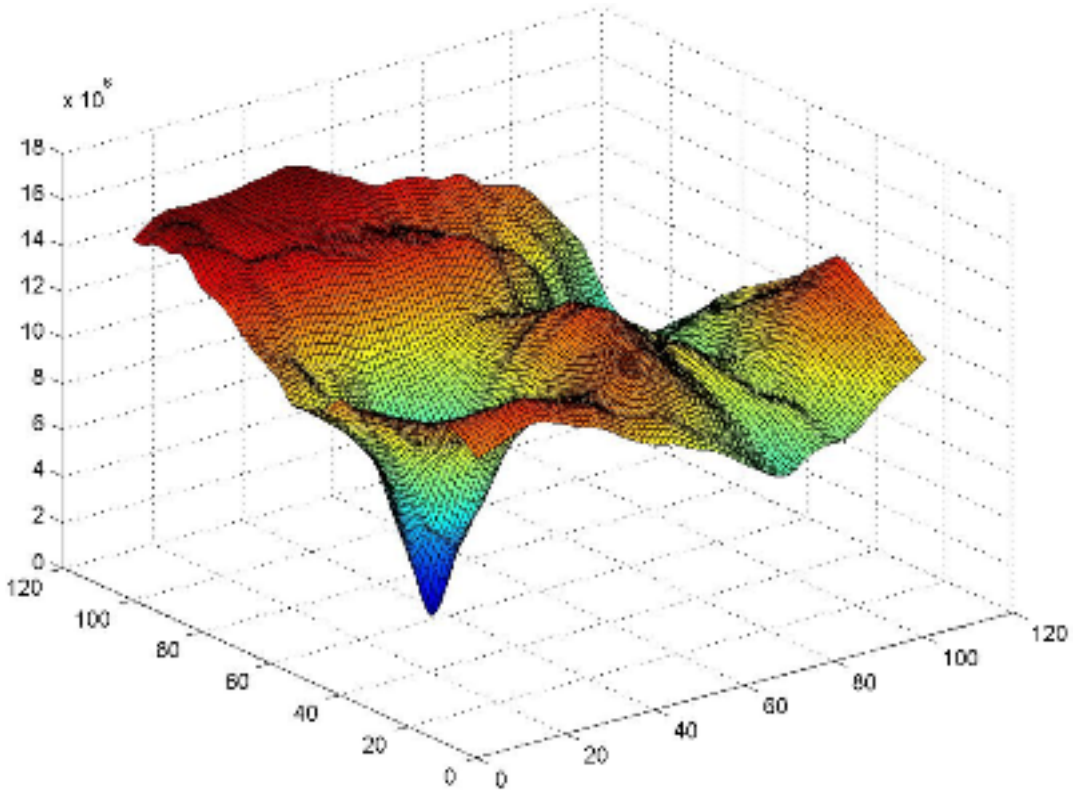


Figure 3.11: The SE surface over a larger search region around a healthy RBC. A nearby RBC causes the second dip on the right of the surface and the col between it and the imit on the left.

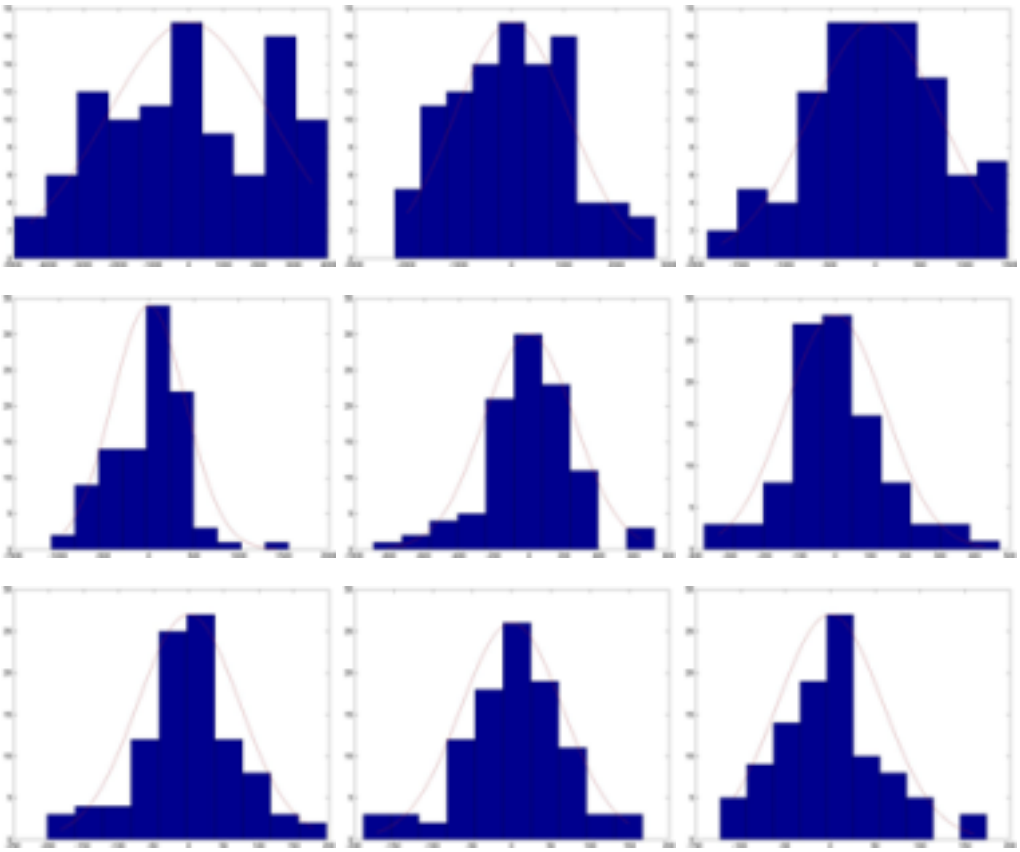


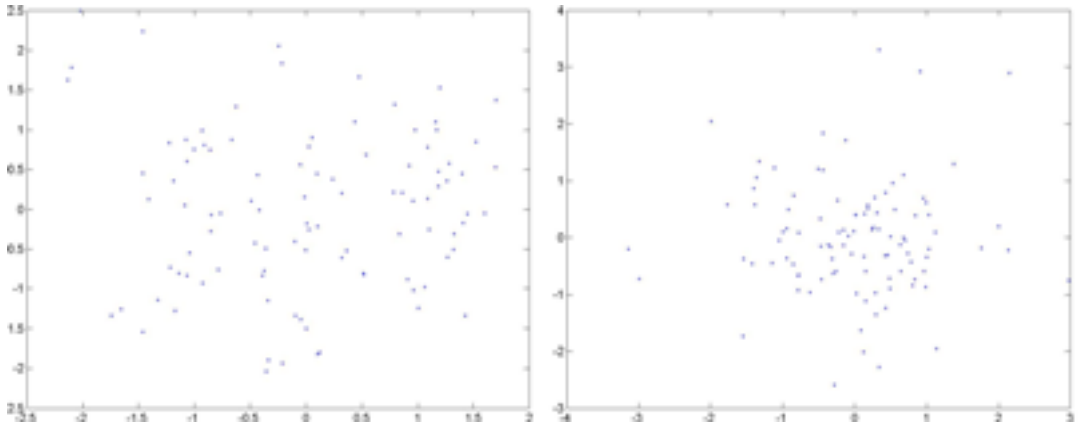
Figure 3.12: From top left to bottom right: histograms of selected PCs, $b(k, \underline{x})$ for $k = 1, 2, 4, 8, 16, 32, 64, 80, 96$ computed according to equation 3.12 for 100 images $\underline{x}$ from the training set. Also shown is a fit of a normal distribution to each of the histograms.

The results in figure 3.12 illustrate that, apart from the first few modes with low k, the assumption that the PCs of our *FPM* are multivariate normal seems to be quite good. Although the distributions are sometimes rather asymmetric, the average of each is, as required by 3.12 always zero.

In addition, we show in figure 3.13 scatter-plots of pairs of PCs calculated for the cell patches used in training which show that the components, both at low k and high k seem to be independent – as required by the multi-normal distribution 3.44 in addition to their being uncorrelated which is guaranteed by the PCA process.

When fitting an *APM* to data it is important not only that the fit is good with a small error, but also that the model thereby inferred is *not* one that, according (say) to 3.44 would be very unlikely. In terms of the data variables $\underline{x}$ an unlikely model is one that, within the t-dimensional

than d it may *not* be assumed that all the PCs are normally distributed according to a version of 3.44 in which t is replaced by d.



(a) (b)

Figure 3.13: Scatter-plots of pairs of PCs calculated for the set of cell image patches used in training for: (a) two low values of k, $\underline{b}(1)$ and $\underline{b}(2)$; and (b) for two high values of k, $\underline{b}(50)$ and $\underline{b}(90)$.

model subspace, isn't like the examples selected for training. 3.44 thus reflects the *model specificity* which may be tested by generating samples from the distribution and calculating a measure of their distance from the training examples[5].

The values generated by a normal distribution – especially if it is not low-dimensional – are often very small so it is convenient to consider the negative of the model log-likelihood:

$$L_m = \sum_{k=1}^{t} \left\{ \frac{b(k)^2}{2\sigma(k)^2} + \frac{1}{2} log(2\pi\sigma(k)^2) \right\}. \tag{3.45}$$

The last term does not depend on the data $\underline{x}$ to which an *APM* is being fitted so for the purpose of assessing the likelihood of the model inferred from the data we may use a weighted sum of squares of the PCs:

$$L(\underline{b}) = \sum_{k=1}^{t} \frac{b(k)^2}{2\sigma(k)^2}. \tag{3.46}$$

Since the variances $\sigma(k)^2 = \lambda(k)$ decrease rapidly as k increases it is apparent that a fitted model will be unlikely unless the PCs for moderately high k (but of course not greater than t) are correspondingly small. It is plausible therefore to set a threshold on $L(\underline{b})$ to reject models that would be unlikely and in particular when the model, like our *FPM*, can be adaptive to help reject spurious fits to bland data in the background of an image. Since only log-likelihoods relative to the threshold are required for this purpose, we note that variances calculated from training data matrices that have not been normalised may be used (recall section 3.2.2). Examples of

[5] Such a procedure is recommended by Cootes and Taylor who calculate a distance from the sample to the nearest example in the training set.

such likelihood surfaces are shown in figure 3.14. In the figure (a), (b) and (c) correspond to the SE surfaces shown in figure 3.6 whilst (d) shows a generous threshold on the likelihoods that is well above their minimum values of ~ 7 near to the correct cell locations. Further examples for other image regions in the plasma background and containing part of a RBC are shown in figure 3.15. These correspond to the SE surfaces in figure 3.8. It can be seen that the threshold of ~ 42 illustrated in figure 3.14(d) would reject each of these model fits as unlikely.

3.3.3 MAP estimates

In a Bayesian framework 3.44 represents the *prior probability* of a patch model characterised by the PCs $b(k), k = 1 \cdots t$. In such a framework a model should be sought which best explains the fit to a datum $\underline{x}$ *a posteriori* by:

$$\max_{\underline{b}} \{p(\underline{b}|\underline{x})\} . \tag{3.47}$$

Since

$$p(\underline{b}|\underline{x})p(\underline{x}) = p(\underline{x}|\underline{b})p(\underline{b}) \tag{3.48}$$

and the distribution of the data $p(\underline{x})$ is independent of the model, 3.47 is equivalent to:

$$\max_{\underline{b}} \{p(\underline{x}|\underline{b})p(\underline{b})\} . \tag{3.49}$$

$p(\underline{x}|\underline{b})$ is just the probability of observing the noise $\underline{\epsilon}_{PCA}$ defined by 3.41 which may be modelled as normal:

$$p(\underline{x}|\underline{b}) = p(\underline{\epsilon}_{PCA}) = \frac{1}{(2\pi\sigma^2)^{\frac{(d-t)}{2}}} e^{\frac{-\underline{\epsilon}_{PCA}^2}{2\sigma^2}} . \tag{3.50}$$

In 3.50 we have been careful to remember that, in PCA with the PCs determined by 3.12, the model *always* fits a datum perfectly in the model subspace and that $\underline{\epsilon}_{PCA}$ lies entirely in the remaining $d - t$ dimensions.

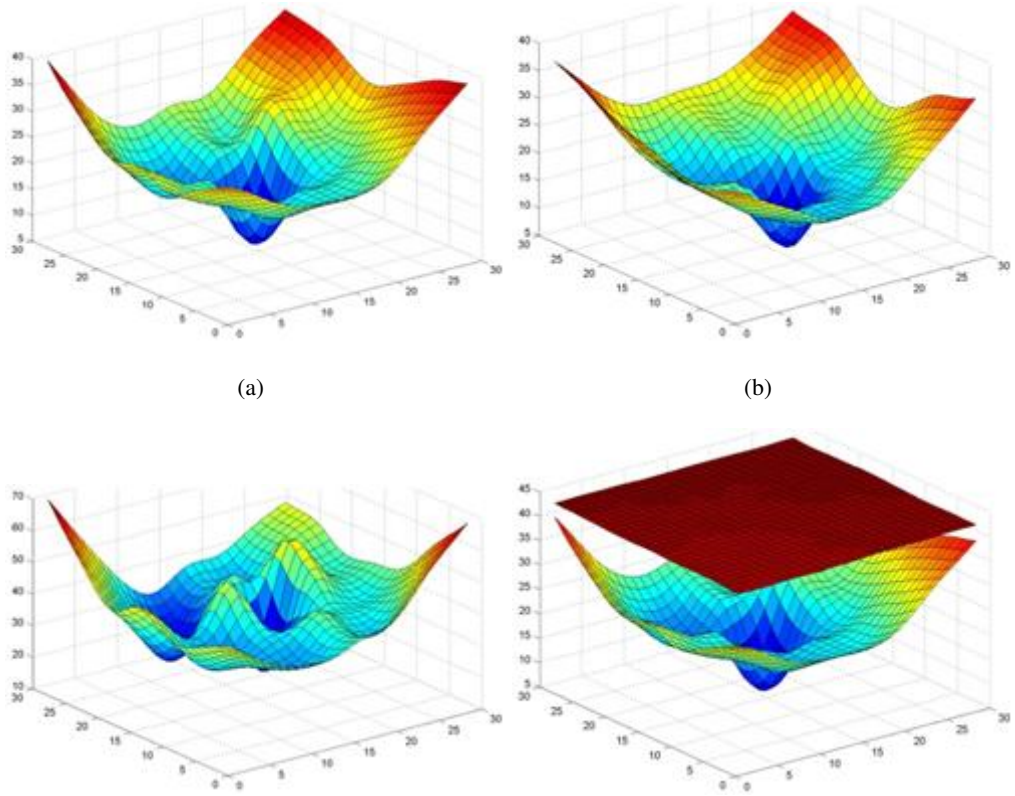
The negative log-likelihood corresponding to $p(\underline{x}|\underline{b})p(\underline{b})$ in 3.49 is thus:

$$L = \frac{\underline{\epsilon}_{PCA}^2}{2\sigma^2} + \frac{(d-t)}{2} log(2\pi\sigma^2) + \sum_{k=1}^{t} \left\{ \frac{b(k)^2}{2\sigma(k)^2} + \frac{1}{2} log(2\pi\sigma(k)^2) \right\} . \tag{3.51}$$

According to 3.51 the maximum likelihood estimate for the noise level is:

$$\sigma^2 = \frac{1}{d-t} \langle \underline{\epsilon}_{PCA}^2 \rangle_j . \tag{3.52}$$

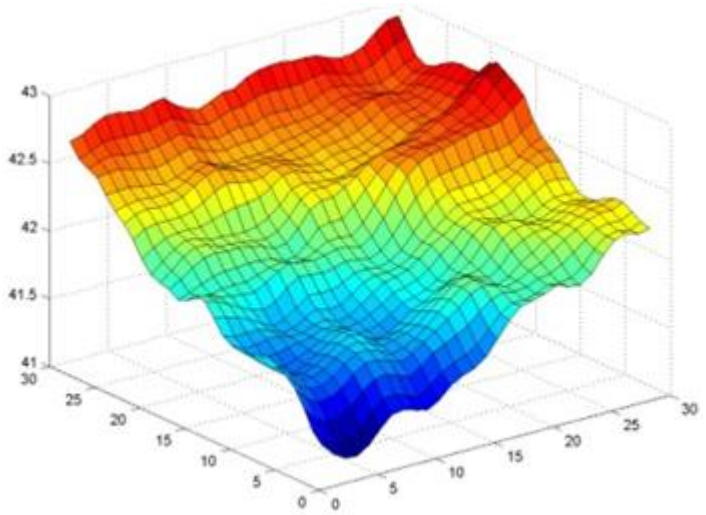
This is similar to the maximum likelihood estimate of the noise level that should be made in a probabilistic principal component analysis, pPCA (recall equation 3.23), but is *not* consistent when the rank r of the covariance matrix S used in training is less than d. In particular if



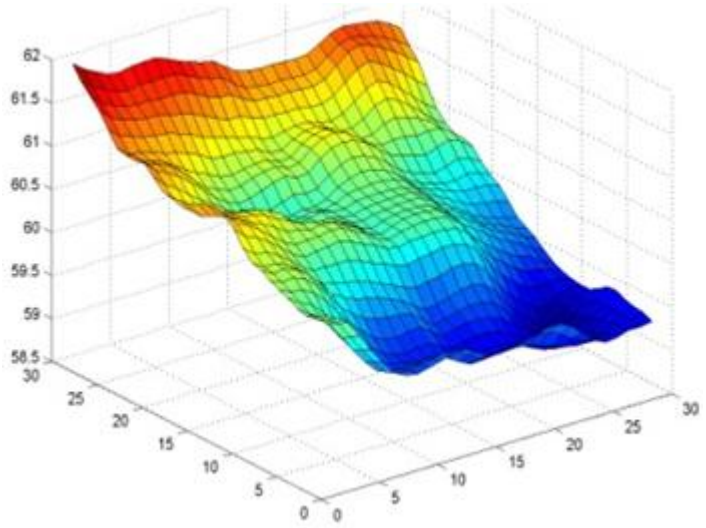
(a) (b)

(c) (d)

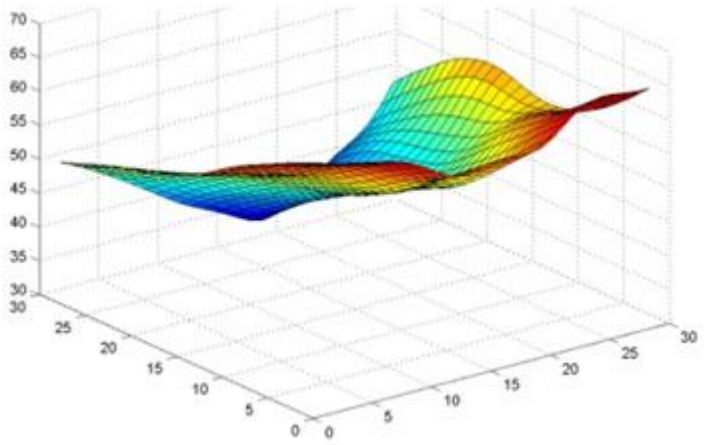
Figure 3.14: The likelihood surfaces $L(\underline{b})$ obtained when the *FPM* is scanned as a flexible template or *APM* across a window centred on a RBC as described in figure 3.6. (a) for a RBC selected for training; (b) the surface obtained when the *APM* was scanned across an isolated cell in an image used in training but where the cell itself was not selected as a member of the training set; and (c) the surface obtained when the *APM* was scanned across an isolated cell from an image not used in training. (d), the same error surface as in (a) showing that the likelihood at the correct cell location is well below a threshold of ~ 42 – as they are also in (b) and (c).



(a)



(b)



(c)

Figure 3.15: The likelihood surfaces obtained when the *APM* is scanned across a window centred on other image regions as in figure 3.8. Centred: (a) on a background region of an image used in training; (b) on a background region of an image not used in training, and (c) on a region containing a part of a cell from an image not used in training.

$r = n < d$ as is usually the case in image processing and computer vision, we would set the noise level:

$$\sigma^2 = \frac{1}{n-t}\langle \underline{\epsilon}^2_{PCA} \rangle_j = \frac{1}{n-t}\sigma^2_U \tag{3.53}$$

where σ^2_U is the unexplained variation. Once the noise level σ^2 has been determined either from training as above or from alternative experimental observations, a MAP estimate of the extent to which an *APM* explains a datum $\underline{x}$ may be characterised by the negative log-likelihood (*cf.* 3.46):

$$L(\underline{x}|\underline{b}) = \frac{\epsilon^2_{PCA}}{2\sigma^2} + \sum_{k=1}^{t} \frac{\underline{b}(k)^2}{2\sigma(k)^2}. \tag{3.54}$$

In a Bayesian framework, a threshold can then be set on $L(\underline{x}|\underline{b})$ in order to select data $\underline{x}$ which may be explained with confidence by the *FPM*. There is only a single decision parameter in this approach so it is important that the noise level σ^2 is set correctly so as to weight the fitting error and model likelihoods appropriately. If the noise level is not known, we could regard it as a second decision parameter and thereby effectively recover the freedom to set thresholds independently on the fitting error ${\epsilon_{PCA}}^2$ and on the model likelihood $L(\underline{b})$ given in 3.46. Using σ^2 in this manner however is clumsy and invites misinterpretation since the noise level is a parameter that should properly be set by a maximum likelihood estimate or some other expectation estimate.

3.4 Conclusions

Exploring the application of the flexible modelling approach developed by Cootes and Taylor by using the simplest patch model as an *FPM* and *APM* has been instructive and has revealed some details about the approach that are often ignored or underplayed, in particular:

(i) The extent to which the PCA approach does not (and sometimes does) depend on specific assumptions about the manner in which the data and other quantities such as the modes and PCs are statistically distributed.

(ii) The importance of using training data that is representative and uncontaminated by outliers.

(iii) How there are several ways in which the number t of significant modes retained in the model may be chosen. In general, we saw that it is difficult in principle to recommend one way in preference to the others, except possibly on the grounds of simplicity, whilst in practice it might best if two or more were to lead to similar choices for t.

(iv) There are consistency problems when the covariance matrix S used in training is not of full rank. This is often the case in image processing and computer vision applications when it is possible (and often necessary) to use a training set containing many fewer examples n than the dimensionality d of the data so that the rank r of S, $r = n < d$. In particular, these problems make it difficult consistently to estimate the noise level from the unexplained variance σ_U^2. These problems might be resolved by using probabilistic principle component analysis as described by Tipping and Bishop [180] but at the cost of losing the simple projection formula 3.12 for the PCs, $b(k)$.

(v) That flexible models are adaptive and thus when used in the active manner as a flexible template seems to be subject to the difficulties noted by Buxton and Zografos [22] and may 'hallucinate' the presence of objects of interest in background regions of an image. It is therefore particularly important when utilising such models to ensure that the active model inferred from data is reasonable, as for example, characterised by a model likelihood measure.

(vi) The importance of including an alignment step to allow for extrinsic variations that should not be incorporated in the model. In model building this leads to the need for inclusion of a Procrustes alignment procedure [32, 69]. This improves model specificity by reducing its variance. However, fitting an active model then becomes more complicated, in general requiring a non-linear optimisation, and with greater danger of over-fitting.

It is clear from the exploration that the simple flexible patch model is not adequate for the malaria application. In particular:

(i) Procrustes colour alignment should be included in order to allow for extrinsic variations in the colouration of the microscope slide images, both from image to image and, as noted in chapter 2, sections 2.2.1 and 2.6.1 and by Tek [175], within each image.

(ii) The fact that there are often parts of other RBCs in the corners of the square patches covered by the patch in the *APM* does not seem *per se* to be disastrous. The results in 3.1.5 indicate that the *APM* works fairly well even though in constructing the *FPM* patches were specifically chosen for training that did not include such effects.

(iii) However, the presence of nearby cells seems to be the major reason why RBCs are often misshapen and, with the *FPM* constructed from a training set containing mostly well-shaped (*i.e.* circular) cells, the *APM* often failed to fit well to misshapen RBCs. It is

possible that better results may have been obtained if misshapen RBCs (with the almost inevitable presence of parts of nearby cells in the corners of the patches) had been included in the training set.

(iv) In light of (ii) and (iii) above, it would seem inappropriate to use a flexible appearance model (*FAM*) and active appearance model (*AAM*) in the malaria application. The variation in RBC shape that we have seen and the work of Liu and Sclaroff reviewed in the previous chapter, section 2.6, though in a different application of microscope slide blood smear images, supports this conclusion. Though selection of RBCs and delineation of cell boundaries are not tasks requiring medical expertise, construction of a training set would then be very time consuming. Apart from a very few, insignificant preliminary experiments we were therefore unable to pursue such an approach.

(v) It seems from the results of the above exploration that our very simple flexible patch modelling based only on including healthy RBCs in the training set could be used to detect and locate both healthy and infected cells – at least unless the infection is large and fills a significant fraction of the cell as it can do in the later stages of the life cycle just before the cell is disrupted. In principle it would seem better to build separate models of healthy and infected cells but this would raise several problems, including: (a) the number of infected cells is much smaller than the number of healthy cells, so constructing a training set is more difficult; (b) at least three (or if artefacts were included four) way classification decisions would then in principle be required; (c) infections are very variable in shape and extent, especially at different stages of their life cycle; (d) identification of infected cells (and in particular the stage of their life-cycle) can be quite difficult and often requires medical expertise for a tedious and time-consuming task.

Whilst it would seem that many of both sets of issues above could be addressed by using more sophisticated flexible models, and even if necessary by going beyond PCA by using probabilistic (pPCA) and/or kernel approaches or mixture pPCA [179] three fundamental difficulties would remain. First the large amounts of time required to select suitable training data especially if expert medical knowledge is required. Second, infections vary greatly in appearance and may occur anywhere within a cell – both of which seriously complicate building a flexible model and utilising it as an active model. Third, as noted in the review chapter in the context of the possible use of Markov random field models for segmentation (section 2.2.2), using more complicated models may lead to increased systematic errors.

The ideal situation would thus seem to be to employ an unsupervised approach that relies only on very general, simple modelling. It is to such an approach that we turn in the following chapters.

Chapter 4

Discrimination – An Unsupervised Approach

We saw in the previous chapter that a model-based approach was unable to lead to a satisfactory system for the segmentation of RBCs and parasite infections. Although the patch-based flexible modelling system investigated in detail is the most simplistic of the family of flexible models, we nevertheless concluded that more sophisticated modelling approaches were also unlikely to be successful. Furthermore such models require training data that is tedious (and generally more onerous the more complicated the model) and difficult to obtain especially in a medical research application where experts' time is often a scarce resource – and different experts may not always agree (but see [193]). In the review of chapter 2, section 2.4.3, we were critical of the use of complicated models and also of the chaining of several image-processing operations as exemplified by the work of the Westminster group as such approaches are likely to be sensitive to the careful tuning of parameters (in particular thresholds) and to require back-tracking operations in an attempt to correct errors inevitably incurred in previous steps. As noted at the end of the previous chapter it would appear therefore that an approach is required that relies on a simple underlying model, utilises a small number of image processing steps and, moreover, does not require extensive training or the tuning of many image-processing parameters.

4.1 Unsupervised discrimination

As noted in chapter 2, section 2.5.1, if we ignore so-called mixed pixels and the boundary membranes of RBCs the fundamental task is to determine whether the pixels of a microscope slide image originate from the plasma background (including white blood cells and possible artefacts), the body of a RBC, or from a malarial parasite. This is a discrimination task. It is unsupervised if pixel labels are assigned according to a criterion or criteria that reflect the underlying characteristics of the problem which can be formulated and evaluated without reference to any training data that has already been labelled. There is thus no machine learning of appropriate values of thresholds or other parameters. It is also usually the case that, with the

notable and frequent exception of the number of categories or classes, C, there are few, if any, other parameters to be set manually by the user.

In image processing, an unsupervised approach often amounts to the choice of a predicate according to which discrimination is to be carried out and is most common in the clustering of pixels in a colour space. Often unsupervised algorithms are based on statistical criteria which may be parametric, as in the application of Gaussian mixture models in colour image segmentation in which case they lead both to a set of parameter estimates and a pixel labelling or segmentation, or non-parametric when a statistical criterion may be used that reflects overall properties of the pixel data. Amongst the latter is the classical Fisher discriminant according to which one seeks a labelling or classification of image pixels that maximises the ratio of a between-class variance to a within-class variance [45, 46].

4.1.1 Fisher's Discriminant and Otsu's algorithm

When applied to a one-dimensional feature space, such as pixel intensity, use of Fisher's discriminant leads to the Otsu algorithm whose implementation seems rather better known [89] than its origin [131]. We therefore briefly review this as outlined in [23] with some extra details and explanations.

Suppose we have a one-dimensional feature space, such as the intensity of the pixels in an image or the signal in a particular channel of a colour image, denoted for convenience by x. Suppose further that we wish to classify the pixels as belonging to foreground or background (say), which might be the red-blood cells and plasma respectively in the analysis of thin-film slides. According to Fisher, we should then seek to maximise the ratio of the between-class variance σ_B^2 to the within-class variance σ_W^2 by choosing a threshold, T, *i.e.*

$$\max_T \left\{ \frac{\sigma_B^2}{\sigma_W^2} \right\}. \tag{4.1}$$

Since, as we shall shortly see,

$$\sigma_T^2 = \sigma_B^2 + \sigma_W^2, \tag{4.2}$$

and σ_T^2, the total variance, is constant for a given data sample $x(i), i = 1 \cdots n$ or distribution $p(x)$:

$$\sigma_T^2 = \frac{1}{n} \sum_{i=1}^{n} (x(i) - \mu)^2 = \int dx p(x)(x - \mu)^2 \tag{4.3}$$

with mean μ, equation 4.1 is equivalent to:

$$\max_T \left\{ \frac{\sigma_B^2}{\sigma_T^2} \right\} \Leftrightarrow \max_T \left\{ \sigma_B^2 \right\}. \tag{4.4}$$

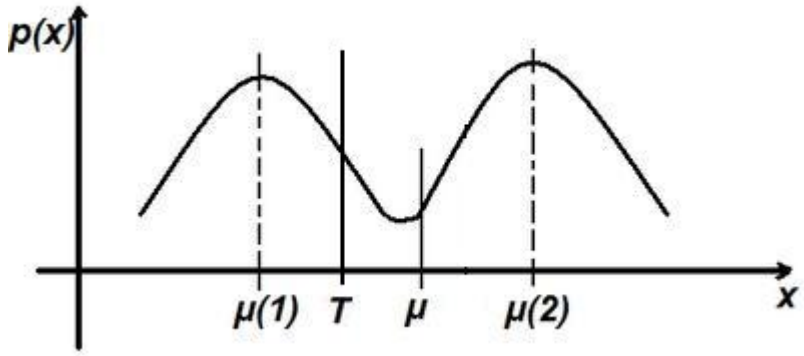


Figure 4.1: Sketch of a multi-modal histogram comprised of data from a mixture of two classes which are to be discriminated by means of a threshold T.

In practice the distribution $p(x)$ is unknown so we estimate it from the histogram $h(x)$ by assuming for n samples that

$$p(x) = \frac{1}{n} h(x). \tag{4.5}$$

Since, after thresholding, values of x above or below T will be assigned to the foreground or background classes respectively (or vice-versa) as shown in figure 4.1, it is useful to introduce the indicator density,

$$z(x|J) = p(x) \quad \text{for} \quad T_{J-1} \leq x \leq T_J \tag{4.6}$$

where $J = 1, 2$ stands for the two classes. T_1 is the threshold T, and T_0 and T_2 are the upper and lower limits of the values of the feature variable, x, often 0 and 255 in image processing applications. Note that we refer to $z(x|J)$ as an indicator density since, after thresholding, values of x in the foreground and background are distinct, whereas the conditional densities of the foreground and background, $p(x|foreground)$ and $p(x|background)$ say, may allow foreground and background values of x to overlap. Such overlap would be allowed in a *mixture model* in which we would set

$$p(x) = \sum_J p(x|J)P(J) \tag{4.7}$$

with class conditional densities $p(x|J)$ and mixing probabilities $P(J)$. With a thresholding procedure such overlap is not allowed and is, of course, a source of classification errors.

It is useful to introduce an indicator variable $Z(J)$ for each class produced by the thresholding, defined as

$$Z(J) = \int dx z(x|J). \tag{4.8}$$

$z(x|J)/Z(J)$ then behaves as a class-conditioned probability density with:

$$\sum_J Z(J) = 1 \tag{4.9}$$

which is consistent with 4.6 and 4.8. The means of the data distributed above and below threshold may then, for example, be represented as

$$\mu(J) = \frac{1}{Z(J)} \int dx z(x|J) x \tag{4.10}$$

and we may write the between-class variance as:

$$\sigma_B^2 = \sum_J Z(J)(\mu(J) - \mu)^2 \tag{4.11}$$

which is equivalent in the two-class case to the possibly more familiar:

$$\sigma_B^2 = Z(1)Z(2)(\mu(1) - \mu(2))^2. \tag{4.12}$$

The variance of class J conditioned on the thresholding, $\sigma(J)^2$ is

$$\sigma(J)^2 = \frac{1}{Z(J)} \int dx z(x|J)(x - \mu(J))^2, \tag{4.13}$$

and thus,

$$\sigma_W^2 = \sum_J Z(J)\sigma(J)^2 = \sum_J \int dx z(x|J)(x - \mu(J))^2. \tag{4.14}$$

It can now be seen that relationship 4.2 follows from equations 4.14 and 4.11 and that to maximise Fisher's discriminant 4.1 we need only maximise the between-class variance, σ_B^2 as in 4.4. Differentiation of equation 4.11 leads immediately to

$$\frac{\partial \sigma_B^2}{\partial T} = \sum_J \frac{\partial Z(J)}{\partial T}(\mu(J) - \mu)^2 + 2\sum_J Z(J)(\mu(J) - \mu)\frac{\partial \mu(J)}{\partial T} \tag{4.15}$$

from which we see that we need to evaluate the derivatives of the indicator variables $Z(J)$ for each class and of the means 4.10. Calculating these derivatives is facilitated by noting that we may write 4.6 as:

$$z(x|J) = [\vartheta(x - T_{J-1}) - \vartheta(x - T_J)]p(x) \tag{4.16}$$

where θ is the step function, zero for $x < 0$ and one for $x > 0$.

When there are two classes, it is then straightforward to show that:

$$\frac{\partial Z(1)}{\partial T} = p(T), \frac{\partial Z(2)}{\partial T} = -p(T), \tag{4.17}$$

and

$$\frac{\partial \mu(1)}{\partial T} = \frac{p(T)}{Z(1)}[T - \mu(1)], \frac{\partial \mu(2)}{\partial T} = \frac{-p(T)}{Z(2)}[T - \mu(2)]. \tag{4.18}$$

Collecting the above results and carrying out a few further manipulations then shows that differentiation of the between-class variance, σ_B^2, finally yields

$$\frac{\partial \sigma_B^2}{\partial T} = p(T)[\mu(1) - \mu(2)][2T - (\mu(1) + \mu(2))]. \tag{4.19}$$

It can be seen from 4.19 that the between-class variance, and hence for problems characterised by a single feature variable x, the Fisher discriminant 4.1 will be extremal if $p(T) = 0$, or $\mu(1) = \mu(2)$, or

$$T = \frac{1}{2}(\mu(1) + \mu(2)). \tag{4.20}$$

Obviously, if $\mu(1) = \mu(2)$, σ_B^2 is zero and is minimised. Choosing a threshold T at a "gap" in a histogram where $p(T)$ vanishes is a familiar segmentation procedure, but one that can rarely be used rigorously in practice. Of more interest is the third condition expressed by equation 4.20. Since the means $\mu(1)$ and $\mu(2)$ depend on the choice of T this is a self consistent equation for T whose iterative solution generates Otsu's algorithm. It is usually initialised by choosing a starting value for the threshold equal to the mean μ and in practice converges quickly in a few iterations $(3 - 4)$. This algorithm is particularly useful since it is a data-driven, unsupervised procedure.

Finally, we note that if the threshold T is varied freely from a low to a high value, an ROC curve may be constructed for what is effectively a type of least mean square *(MSE)* classifier. This relationship is discussed in [45] and [46] (see sections 5.8.2 in both). To obtain complete correspondence with the Otsu algorithm above, the *MSE* classifier may be chosen to:

$$\min_T \left\{ \frac{1}{n} \sum_{i=1}^{n} [y(i) - w(i)x(i) - T]^2 \right\} \tag{4.21}$$

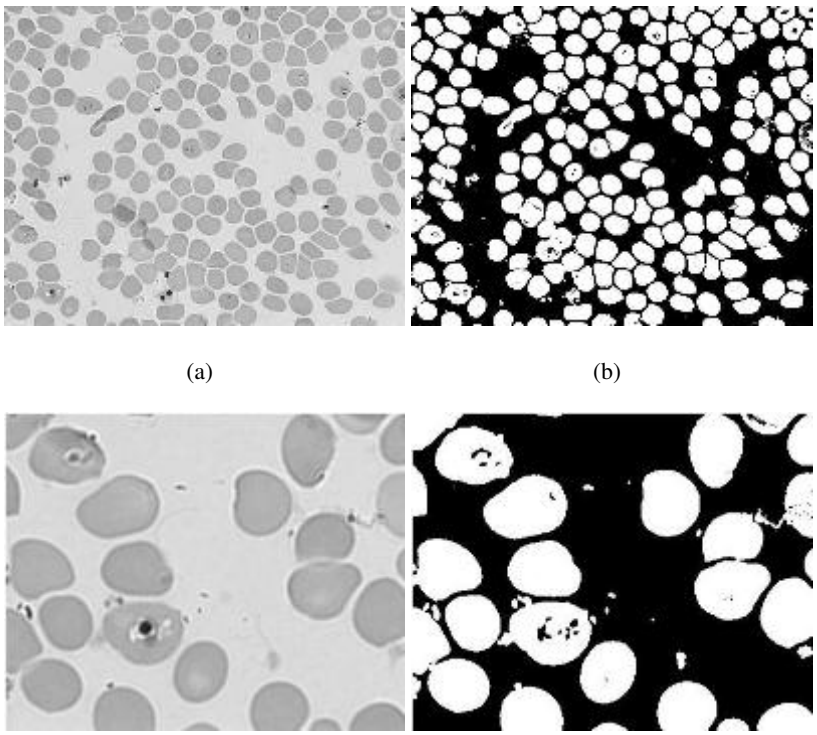
where

$$y(i) = \begin{cases} \frac{n}{n(1)} & \text{if x(i) } \in \text{ class } C(1) \\ \frac{-n}{n(2)} & \text{if x(i) } \in \text{ class } C(2) \end{cases} \tag{4.22}$$

and $n(1)$ and $n(2)$ are the actual numbers of examples in each class. Unlike the conventional *MSE* classifier [45, 46] we weight the features $x(i)$ according to:

$$w(i) = |y(i)| \tag{4.23}$$

so that the threshold T that minimises 4.21 is that defined by the Otsu algorithm 4.20 rather than simply being the sample mean μ [45, 46].



(a) (b)

(c) (d)

Figure 4.2: An example thin-film microscope slide image (a) used to illustrate use of the Otsu algorithm for segmentation (b) of RBCs from the pixel intensities. A one-ninth part of the image is shown enlarged (c) together with the corresponding segmentation (d).

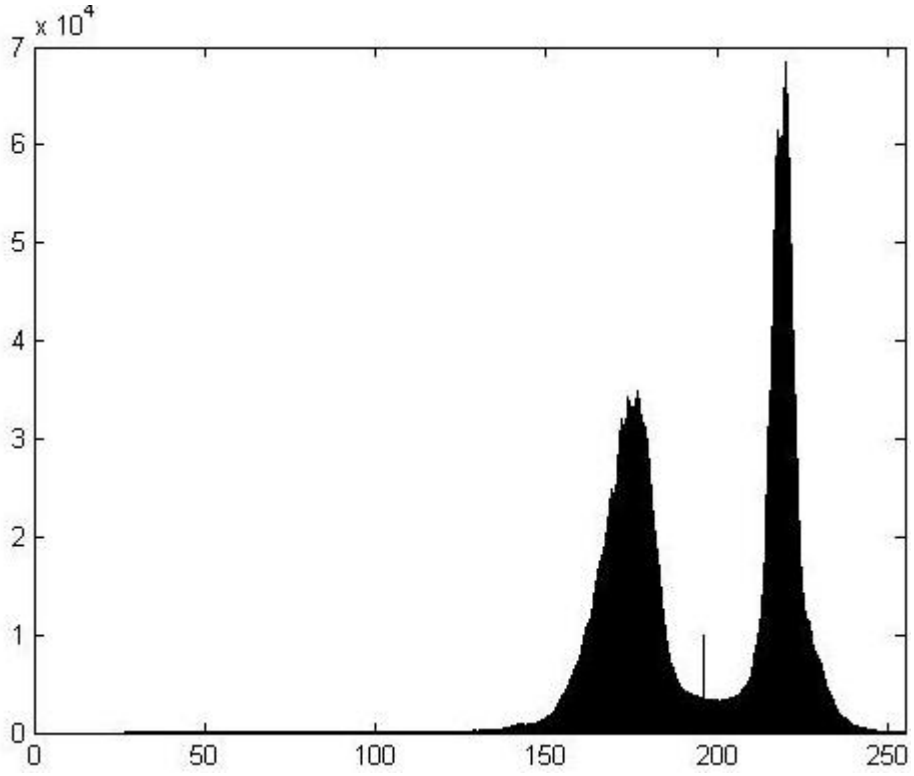


Figure 4.3: Application of the Otsu algorithm to the intensity histogram of the whole image of figure 4.2. The algorithm converged rapidly in 3 iterations to a threshold $T = 196$ as shown.

4.1.2 Intensity-based RBC segmentation

An example is shown in figure 4.3 in which the above Otsu algorithm was applied to the intensity histogram of the image displayed in figure 4.2.

In the segmentations pixels with intensities below threshold and labelled as foreground are set to white whilst those above threshold labelled as background are set to black. For the most part foreground pixels correspond to RBCs. Some of the RBCs are infected by malarial parasites which are darker than the RBCs near to which there are often small clusters of bright pixels. There are thus occasionally a few isolated pixels and some other very small groups of pixels labelled as background within some RBCs. Similarly there are some isolated pixels and small groups of pixels below threshold and set to white outside of any of the RBCs. These usually seem to be caused by artefacts in the plasma. Given that the result was obtained automatically without any manual tuning of parameters the segmentation appears to be remarkably good. This outcome was typical of a number of experiments carried out which will be discussed together with similar results obtained by using individual colour channels in the next section.

4.1.3 Thresholding single colour channels

In this section we use the Otsu algorithm for the independent thresholding of the red, green and blue channels. As for the intensity-based segmentation described above the thin-film microscope slide images, having been taken by clinical researchers who optimised the microscope and illumination settings for human inspection of the slides, were not pre-processed or standardised, for example, by histogram equalisation. Typical results obtained with the thresholds produced by the application of the Otsu algorithm are shown in figure 4.4. It is notable that there is signal in each colour channel so use of a colour transformation such as the HSV representation used in [79] may be useful (see also chapter 2, section 2.2.1).

ROC curves were calculated from comparison of the results obtained when the threshold was systematically varied from a low to a high value with a hand-segmented ground truth. A typical ROC resulting from processing the green channel is shown in figure 4.5 together with that obtained from the previous, intensity-based segmentation. The foreground (red-blood cells) is regarded as the positive class. The hand segmentation had to be carried out with great care as not only the red-blood cells are visible, but also the cell membranes, (see figure 4.2), and the latter excluded from the foreground. This was extremely time consuming so it was carried out on one complete 1300×1030 image and on one ninth of each of eight other similar images divided into 3×3 arrays of sub-images. The latter were used to provide a means of assessing performance across a set of 100 images in which human red-blood cells are infected with *P. falciparum*. These images vary considerably in their appearance (for example the mean colour varied from $(125, 124, 153)$ to $(223, 212, 236)$ and in the density of red-blood cells from 109 to 250 per image). Six examples are shown in figure 4.8. The image which was annotated in toto was similarly divided into nine sub-images to provide a means of assessing variation within an image. The 17 ROC curves produced from the green channels of these 17 image patches are shown in figure 4.6 whilst the 9 ROC curves showing the intra-image variation are displayed in figure 4.7 (a) with the 9 showing the inter-image variation at (b). It can be seen there is more inter-image variation than intra-image variation, but little variation in the results in either case. We shall return to discussion of this in the next chapter 5.

The ROC curves produced from the intensity and from the green channel of the image patch shown in figure 4.2 were shown above in figure 4.5. Similar ROC curves obtained from the red and blue channels of the *whole* image are shown superimposed on the ROC curves obtained from the intensity and from the green channel in figure 4.9.

The areas A under the ROC curves obtained from the intensity and from the red, green and blue channels of the whole image are given in table 4.1. The lower area under the ROC

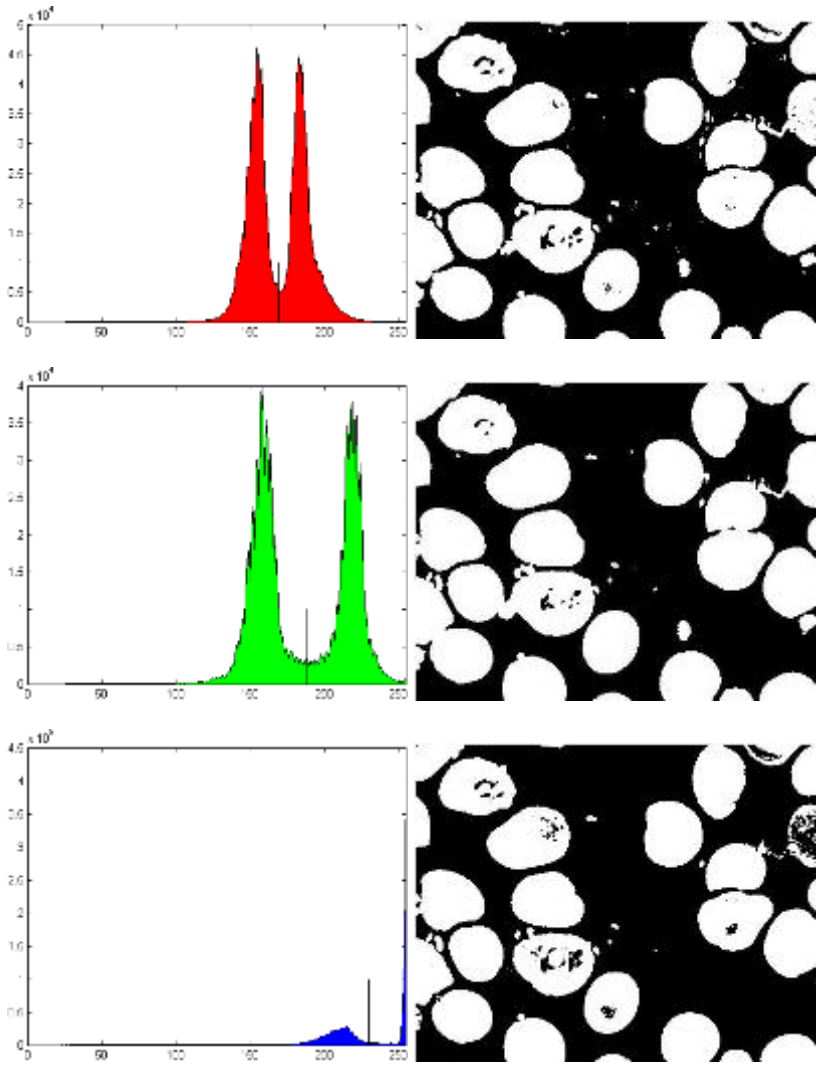


Figure 4.4: Left: histograms and thresholds obtained by application of the Otsu algorithm to the red (top), green (middle) and blue (bottom) channels of a thin-film slide image, together with right: examples of the resulting segmentations. The same portion of the image is displayed on the right as in figure 4.2 with foreground pixels below threshold set to white and those above threshold in the background black.

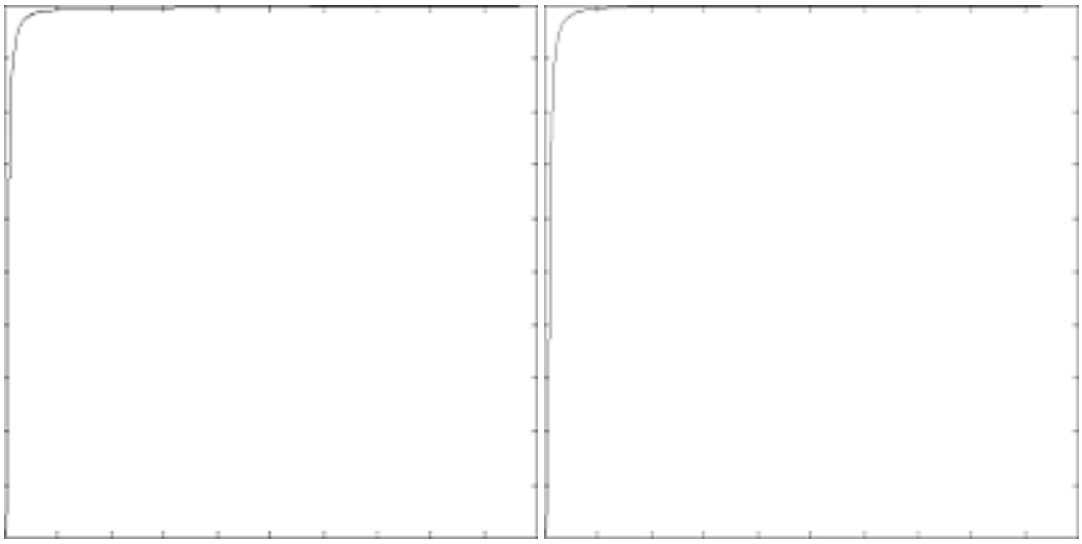


Figure 4.5: A typical ROC curve showing the variation of the true-positive rate tpr (y-axis) as a function of the false-positive rate fpr (x-axis) obtained from the intensity of the sub-image shown in figure 4.2 (left) and similarly from the green channel (right).

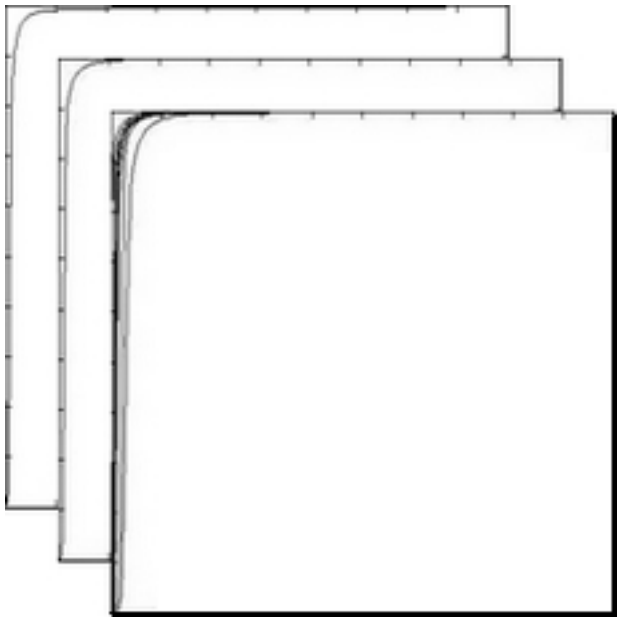
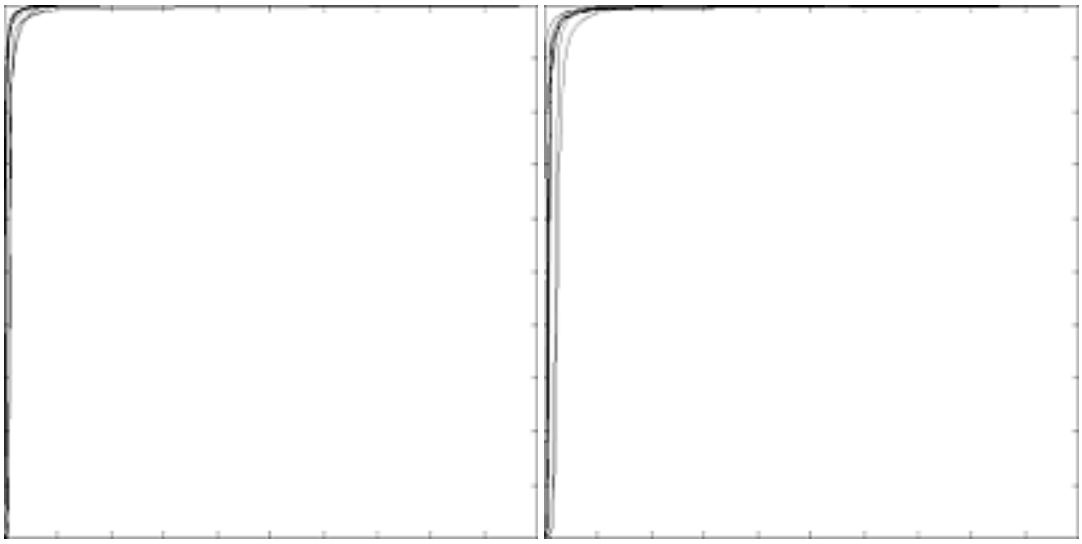


Figure 4.6: ROC curves showing the variation of the true-positive rate tpr (y-axis) as a function of the false-positive rate fpr (x-axis) obtained from the intensity of the sub-image shown in figure 4.2 (back) and from the green channel (second from the back) with the 17 curves obtained from the green channels of the sub-image patches as described in the text superimposed.



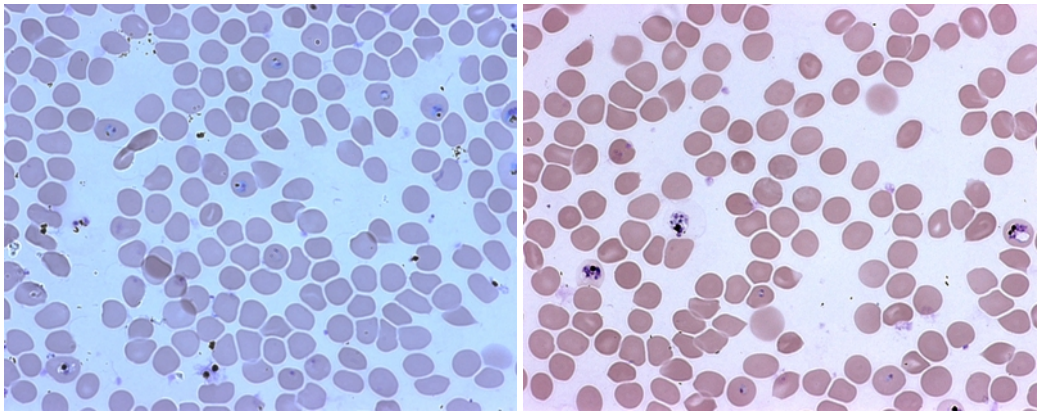
(a) (b)

Figure 4.7: The 17 ROC curves showing the variation of the true-positive rate tpr (y-axis) as a function of the false-positive rate fpr (x-axis) superimposed in figure 4.6 split into the 9 showing intra-image variation (a) and the 9 showing inter-image variation (b).

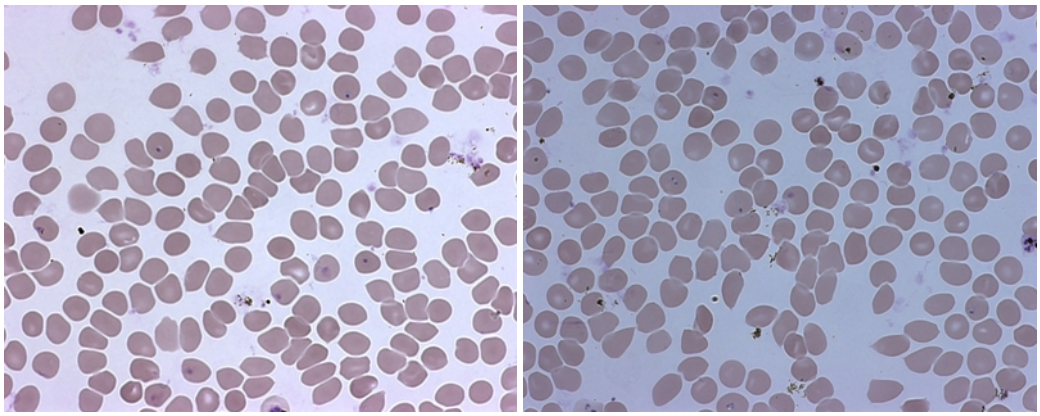
	I	R	G	B
A	0.9848	0.9760	0.9860	0.9813
tpr	0.9604	0.9523	0.9777	0.9294
fpr	0.0297	0.0345	0.0450	0.0203
ϵ	0.0693	0.0821	0.0673	0.0909

Table 4.1: Classification performance obtained from the whole image shown in figure 4.2 (a) by thresholding the intensity, red, green and blue channels denoted I, R, G, B respectively.

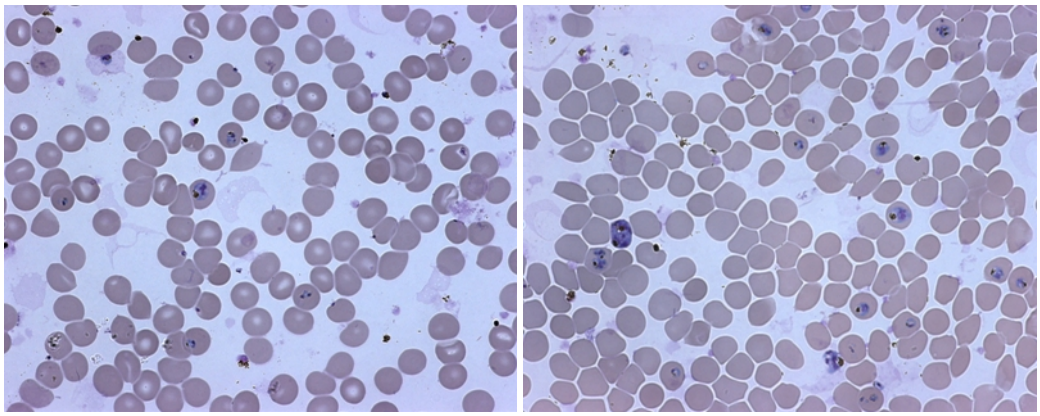
indicates the red channel is slightly less discriminating than the green or blue, whilst the green channel produces a slightly better ROC than those obtained with the blue channel or intensity. The operating points produced by the Otsu algorithm also vary. That in the green channel gives the highest true-positive rate tpr, the red channel gives a fairly good balance of the two types of error, whilst the blue channel gives a very low false positive rate fpr but its error rate $\epsilon = 1 - tpr + fpr$ is the highest and the operating point is further from the ideal of $tpr = 1$, $fpr = 0$ than those in the other colour channels. Segmentation based on the intensity is competitive with those obtained from the colour channels and, at the Otsu operating point, apart from the green channel gives a slightly lower error rate.



(a) (b)



(c) (d)



(e) (f)

Figure 4.8: Six of the 100 images of thin-film slides of human malarial infected blood including: (a) the image annotated in toto and (b) one of the eight partially annotated images, plus four others (c) – (f) chosen to illustrate the variation in colour, (c) and (d), and in RBC density, (e) and (f). Images (c) – (f) were not used in generating the ROC curves described in the text.

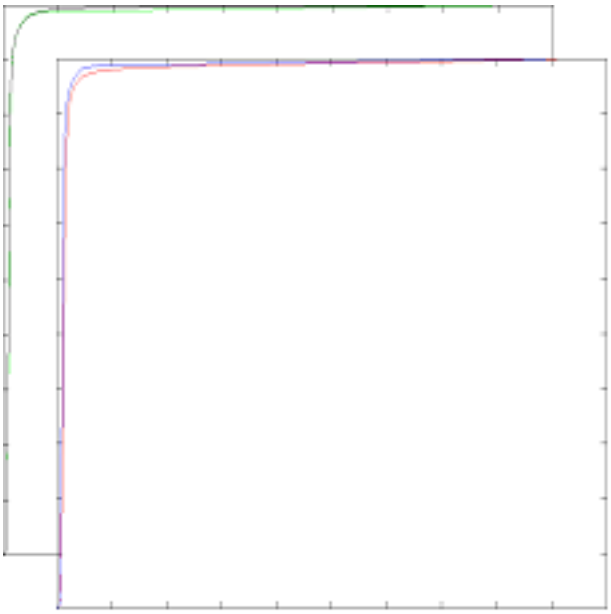


Figure 4.9: ROC curves showing the variation of the true-positive rate tpr (y-axis) as a function of the false-positive rate fpr (x-axis) obtained from the red and blue channels of the whole image shown in figure 4.2 superimposed on the ROCs obtained from the intensity and from the green channel (back). The latter are almost coincident.

4.1.4 A multi-class generalisation

In reviewing the Otsu algorithm above, we have developed a notation that enables us to generalise to a multi-class problem in which the histogram is to be segmented into C classes by use of several thresholds T_J for $J = 1 \ldots C-1$ as sketched in figure 4.10.

The derivation is given in appendix A. The result is that

$$\frac{\partial \sigma_B^2}{\partial T_J} = p(T_J)[\mu(J) - \mu(J+1)][2T_J - (\mu(J) + \mu(J+1))] \tag{4.24}$$

which may be interpreted in a similar manner to equation 4.19. In particular we obtain a multi-class version of Otsu's algorithm for the thresholds T_J from the set of $C-1$ self-consistent conditions:

$$T_J = \frac{1}{2}(\mu(J) + \mu(J+1)). \tag{4.25}$$

Initialisation may be by equally spacing the thresholds over the range $[T_0, T_C]$, where T_0 is the lower and T_C the upper limit of the values taken by x. Convergence in the multi-class case is much less certain than with a single threshold, though experiments with illustrative distributions indicate it can be a useful procedure. In practice, the available histogram data may be smoothed, for example by using a Gaussian mixture model for curve fitting and a powerful optimisation algorithm such as a particle swarm method used [202]. In fact, as pointed out in appendix B

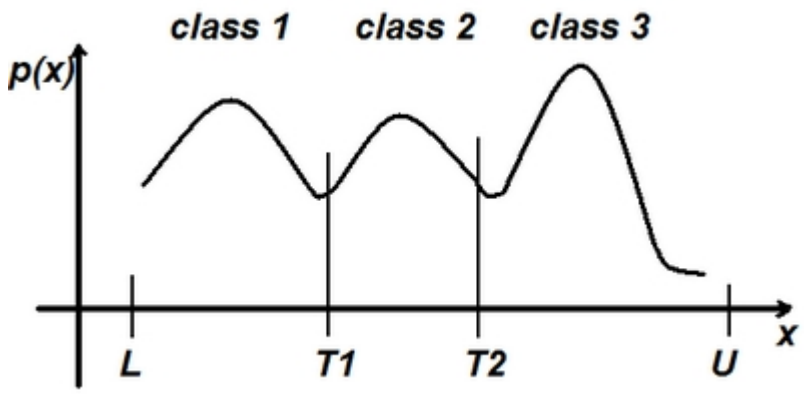


Figure 4.10: Sketch of a multi-modal histogram comprised of data from three classes which are to be discriminated by means of two thresholds T_1 and T_2.

it is straightforward to obtain formulae for the Hessian of second derivatives with respect to the thresholds. This opens up the possibility of a variety of implementations, ranging from multi-dimensional Newton-Raphson and Levenberg-Marquardt methods to conjugate gradients [141]. Alternatively, modern stochastic methods such as differential evolution and SOMA (Self Organising Migratory Algorithm) [203] could be used. Finally, it is easy to envisage the application of continuation methods and, given the close similarity to mixture models, development of EM-like algorithms.

This is an interesting generalisation but, it must be emphasized, one that is restricted to problems where there is a one-dimensional feature space x. Higher dimensional problems are much more complicated and perhaps for this reason not discussed much in standard texts [45, 46].

4.2 Using all three colour channels

The fact that there is signal in each colour channel indicates that all three should be utilised. One way would be to combine the outputs obtained separately from each colour channel in some kind of multi-classifier system. Another would be to develop an extension of Otsu's algorithm that can deal with multi-dimensional feature vectors $\underline{x}$. We consider these in turn in this section.

4.2.1 Combining outputs from the colour channels – 3 $\times$ 1D segmentation

Since there are only the three classifiers based on the red, green and blue channels to be combined the sophistication of general-purpose multi-classifier combination methods [98], such as the genetic programming techniques developed by Langdon and Buxton [105], is not needed. Nor have we implemented the EM approach developed by Warfield *et al* [193] or merging methods more familiar in image processing applications [44]. Preliminary experiments indi-

cated that the latter did not work well so, noting that there are only 162 relevant in-equivalent partitions of the eight possible outcomes of using independent thresholds on each colour channel[1], we carried out an exhaustive search for the best combinations that gave the highest value of $\sigma_B^2(R) + \sigma_B^2(G) + \sigma_B^2(B)$ across the three colour channels.

The combination of the individual colour channel classifiers was tested over 142 images – the 100 images in which human red-blood cells are infected with *P. falciparum* used in section 4.1.3 plus a further 25 similar images and 17 images of malarial infected mouse blood. The most frequent optimal combination of the colour channels was that pixels belonging to the red-blood cells were detected in all three colour channels (54 of the 142 images), with detection in at least one channel the second most frequent optimal combination (32 times), and detection in at least two channels the third (22 times).

The results for the whole image shown in figure 4.2 are summarised by the maximum realisable ROC (MRROC) curve [161] shown in figure 4.11. The values of tpr and fpr plotted are those obtained as required by the MRROC construction from the convex hull of the operating points generated by application of the best classifier combination over the whole image.

4.2.2 Multi-dimensional extensions of Fisher's discriminant

Although the results obtained by thresholding the colour channels singly or in combination are all quite good, in this approach, as in many other multi-classifier systems, the types of classifiers to be combined are defined from the outset. In particular the geometry of the decision surfaces is constrained, in this case when we threshold on the individual colour channels for each of them to be normal to the axis of a primary colour. It would seem better to retain as much freedom as possible in choosing the decision surface. This we can do, within the context of a linear model, by utilising a multi-dimensional version of the Fisher linear discriminant. Several alternatives expressed in terms of the between-class covariance S_B and within-class covariance S_W are given in the literature (see for example: [45, 43, 97, 178, 46]). The covariances are analogous to the variances σ_B^2, σ_W^2 and σ_T^2 introduced in section 4.1.1 and may be written as:

$$S_B = \sum_J \frac{n(J)}{n} (\underline{x}(i) - \underline{\mu}(J))(\underline{x}(i) - \underline{\mu}(J))^T \tag{4.26}$$

$$S_W = \sum_J \frac{n(J)}{n} S(J) \tag{4.27}$$

where

$$\underline{\mu}(J) = \frac{1}{n(J)} \sum_{i \in J} \underline{x}(i) \tag{4.28}$$

[1]There are $162 = {}^8C_1 + {}^8C_2 + {}^8C_3 + {}^8C_4$ in-equivalent partitions. Note that: ${}^8C_5 \equiv {}^8C_3$, ${}^8C_6 \equiv {}^8C_2$ and ${}^8C_7 \equiv {}^8C_1$.

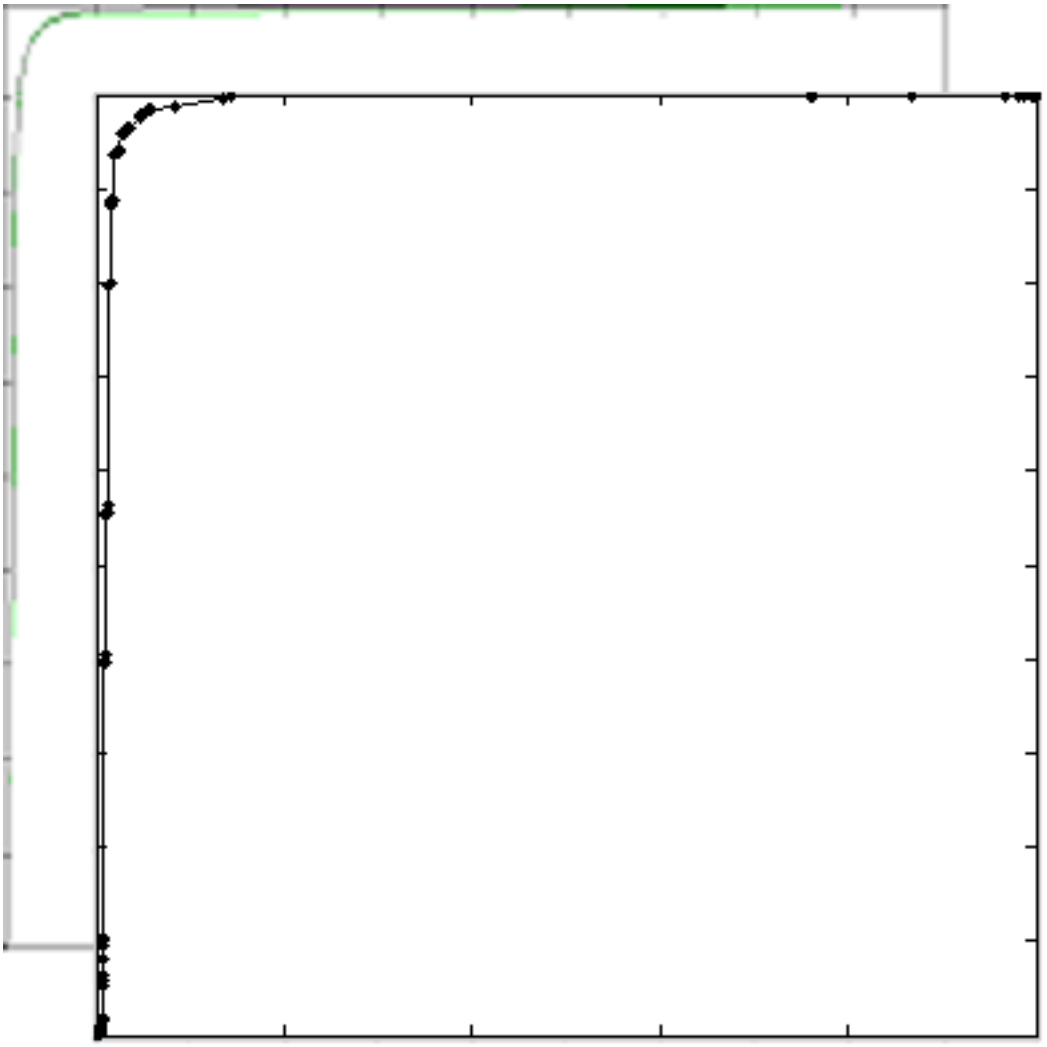


Figure 4.11: The MRROC showing the variation of the true-positive rate tpr (y-axis) as a function of the false-positive rate fpr (x-axis) generated as described in the text shown superimposed on the ROC curves for the intensity-based segmentation and that obtained from the green channel as in figure 4.9.

are the class-conditioned means,

$$S(J) = \frac{1}{n(J)} \sum_{i \in J} (\underline{x}(i) - \underline{\mu}(J))(\underline{x}(i) - \underline{\mu}(J))^T \tag{4.29}$$

are the class-conditioned covariances, and $n(J)$ the number of samples in class J. The class-conditioned means sum to the sample mean $\underline{\mu}$ in the usual manner:

$$\underline{\mu} = \sum_J \frac{n(J)}{n} \underline{\mu}(J), \tag{4.30}$$

the total covariance

$$S_T = \frac{1}{n} \sum_{i=1}^{n} (\underline{x}(i) - \underline{\mu})(\underline{x}(i) - \underline{\mu})^T \tag{4.31}$$

and, corresponding to 4.2:

$$S_B + S_W = S_T. \tag{4.32}$$

The generalised forms of the Fisher discriminant expressed in terms of S_B and S_W include three criteria which, following [43] (section 6.6) we shall write as:

$$D_2 = \frac{tr(A^T S_B A)}{tr(A^T S_W A)} \tag{4.33}$$

$$D_3 = tr\left\{(A^T S_W A)^{-1}(A^T S_B A)\right\} \tag{4.34}$$

$$D_4 = \frac{det\left|A^T S_B A\right|}{det\left|A^T S_W A\right|} \tag{4.35}$$

where A is a rectangular matrix whose d-dimensional columns define the feature space generated by optimising the various discriminants. There is only one criterion based on determinants because it follows from the properties of determinants that the last, 4.35, is equivalent to

$$D_4 = det\left|(A^T S_W A)^{-1}(A^T S_B A)\right|. \tag{4.36}$$

Both types of criterion measure aspects of the 'size' of the between-class scatter or covariance in comparison to that of the within-class scatter or covariance. In particular, we note that the determinant criterion D_4 compares the squares of their volumes in the feature space defined by the columns of the matrix A [46].

Other forms leading to equivalent results may be found when the total covariance S_T is used instead of one of the other matrices. In particular, Devijver and Kittler [43] discuss a fourth criterion

$$D_1 = tr\left\{A^T(S_B + S_W)A\right\} \tag{4.37}$$

which may be re-written ultimately in terms of the total covariance S_T as:

$$D_1 = tr\left\{A^T S_T A\right\} = tr\left\{(AA^T)S_T\right\}. \tag{4.38}$$

This criterion must be supplemented by other constraints, such as normalisation of the columns of A, in order to avoid the trivial solution that D_1 is zero as discussed in [43] section 6.6.1 (and see also appendix D where, for completeness, we summarise some aspects of this old and now rather inaccessible literature[2]). It is also worth noting that 4.38 does not simplify any further when the feature space defined by the columns of A is of lower dimension than d, the dimension of the original feature space $\underline{x}$ so that AA^T is a projection operator. This reminds us that for similar reasons, 4.33, 4.34 and 4.35 also do not in general simplify.

4.2.3 Generalised eigensolutions

Criterion D_2 may be optimised quite easily by differentiation with respect to the components $a_i(k)$ of the columns $\underline{a}(k)$ of A (appendix C) leading to a generalised matrix eigenequation:

$$S_B\underline{a}(k) = \lambda(k)S_W\underline{a}(k) \tag{4.39}$$

and the conclusion that

$$\max\{D_2\} = \frac{tr\left\{\underline{a}(k)^T S_B\underline{a}(k)\right\}}{tr\left\{\underline{a}(k)S_W\underline{a}(k)\right\}} = \lambda(k). \tag{4.40}$$

It thus follows from the definition of D_2 that the generalised eigenvalues $\lambda(k)$ of 4.39 are positive. Since S_B is of rank r_B there are r_B eigensolutions which we may thus label in non-increasing order with:

$$\lambda(1) \geq \lambda(2) \geq \cdots \lambda(r_B) > 0. \tag{4.41}$$

Maximisation of D_2 thus generates a hierarchy of feature vectors $\underline{a}(k)$ where $\underline{a}(1)$ defines the most discriminating direction in the original feature space $\underline{x}$, $\underline{a}(2)$ defines the next most discriminating direction and so on. This linear discriminant analysis or LDA is reminiscent of PCA but it must be remembered that it is designed to serve the very different purpose of *discrimination* as opposed to *data description*. Mathematically, we note that the feature vectors $\underline{a}(k)$ may not completely span the feature space $\underline{x}$ since $r_B \leq C-1$. Also, they are not in general orthogonal as may be seen by writing 4.39 as:

$$S_W^{-1}S_B\underline{a}(k) = \lambda(k)\underline{a}(k) \tag{4.42}$$

and noting that the matrix ${S_W}^{-1}S_B$ in general is not real symmetric. Apart from this, 4.42 is of limited use as in general it will generate $d-r_B$ spurious eigenvalues equal to zero and, since ${S_W}^{-1}S_B$ will not in general be normal[3], it cannot be guaranteed to possess a complete set of d eigenvectors.

[2]The book [43] is long out of print and other sources tend to give much less detail.

[3]A normal matrix M satisfies $MM^T - M^TM = 0$.

Differentiation of D_3 is more complicated (appendix C) but nonetheless straightforward and leads to the same generalised matrix eigenequation 4.39 as above, but in matrix form

$$S_B A = S_W A \Lambda \tag{4.43}$$

where Λ is the $r_B \times r_B$ diagonal matrix of the generalised eigenvalues $\lambda(k)$, and the conclusion that:

$$\max\{D_3\} = tr(\Lambda) = \sum_{k=1}^{r_B} \lambda(k). \tag{4.44}$$

In this case the whole of the space of new discriminating features defined by the $\underline{a}(k)$ is generated immediately though we may of course use as many or as few as we choose.

Differentiation of D_4 is described by [45] (page 120) as "not easy" and by [46] (page 123) as "tricky". The trick required [43] (appendix B.3, pages 437-439) is that the derivative of the determinant of a matrix, M say, with respect to its matrix elements, may be written as:

$$\frac{\partial |M|}{\partial M} = adj(M) = |M| M^{-1}. \tag{4.45}$$

Given 4.45 the differentiation is still complicated (appendix C) but leads eventually once again to the generalised eigenequation 4.43 and the conclusion that:

$$\max\{D_4\} = |\Lambda| = \prod_{k=1}^{r_B} \lambda(k). \tag{4.46}$$

4.2.4 A multi-dimensional extension of Otsu's algorithm – Otsu 3D

For a two-class problem, when S_B is rank one, A is a vector and the complications of the previous two sections disappear. Moreover, given the special form of the between-class covariance corresponding to 4.12

$$S_B = \frac{n(1)n(2)}{n^2} (\underline{\mu}(1) - \underline{\mu}(2))(\underline{\mu}(1) - \underline{\mu}(2))^T \tag{4.47}$$

it is apparent that solution of the generalised eigenequation 4.39 leads to a single vector:

$$\underline{a} \propto S_W^{-1} (\underline{\mu}(2) - \underline{\mu}(1)) \tag{4.48}$$

which defines the most discriminating direction in the feature space $\underline{x}$.

To locate the optimal position of the decision surface, we also need to know how close to locate it to $\underline{\mu}(1)$ or $\underline{\mu}(2)$. This is essentially the one-dimensional problem of finding the optimal thresholding of the histogram of feature vectors projected onto $\underline{a}$ which may be solved by use of the Otsu algorithm.

For a two-class problem we thus have a very simple generalisation of the Otsu algorithm in which we iteratively re-compute the most discriminating direction $\underline{a}$ and the threshold defining the location of the decision plane. To initialise the algorithm, we may carry out PCA to determine the direction of greatest variance in the data and take this as our first approximation to $\underline{a}$, project the data onto this direction and apply Otsu's algorithm [23]. The resulting segmentation into two classes is used to construct the class-conditioned means $\underline{\mu}(1)$ and $\underline{\mu}(2)$ and the covariance matrix S_W for the next iteration and so on.

It is easy to envisage several ways in which the algorithm may be refined – for example, by an alternative initialisation and by carrying out an 'early' recalculation of the discriminating direction $\underline{a}$ after each just one iteration of the Otsu thresholding procedure. Here we instead follow [23] and turn to application to the RBC segmentation problem, pausing first only to note that, though generalisation to more than two classes by using the multi-thresholding version of Otsu's algorithm discussed in section 4.1.4 is straightforward, it seems unfortunately to be difficult to devise a generalisation that could utilise several discriminating vectors $\underline{a}(k)$.

4.2.5 Application to colour image segmentation

Application of the above multi-dimensional extension of Otsu's algorithm to segmentation of the red-blood cells in images of thin-film slides is straightforward and leads to very pleasing results such as those illustrated in figures 4.12 and 4.13. Putative RBC pixels displayed as the white foreground in the figures were selected automatically as the minority class. Visualisation of the vector $\underline{a}$ (computed from the whole image) as iteration proceeds indicates considerable change from the direction of principal variance used to initialise the process. This arises because there are more pixels belonging to the blood plasma background than to the foreground of the red-blood cells, resulting in the distribution in colour space being weighted towards the top of the pixel distribution and the initial decision surface being similarly displaced towards this point as shown in figure 4.14. This configuration produces an imperfect segmentation as illustrated in the figure.

As iteration proceeds (figures 4.15 and 4.16) the decision surface swings around until it is slicing the distribution of pixel colours more longitudinally than laterally as shown in figure 4.16. This was somewhat surprising, but even though the segmentation is poor at this stage, the orientation of the decision surface continues to change and the procedure finally produces a very good segmentation as depicted in figure 4.14. Convergence of the direction of $\underline{a}$ and of the segmentation is quite rapid in 15 iterations.

Once the direction of vector $\underline{a}$ and the segmentation have settled down, it is easy to produce

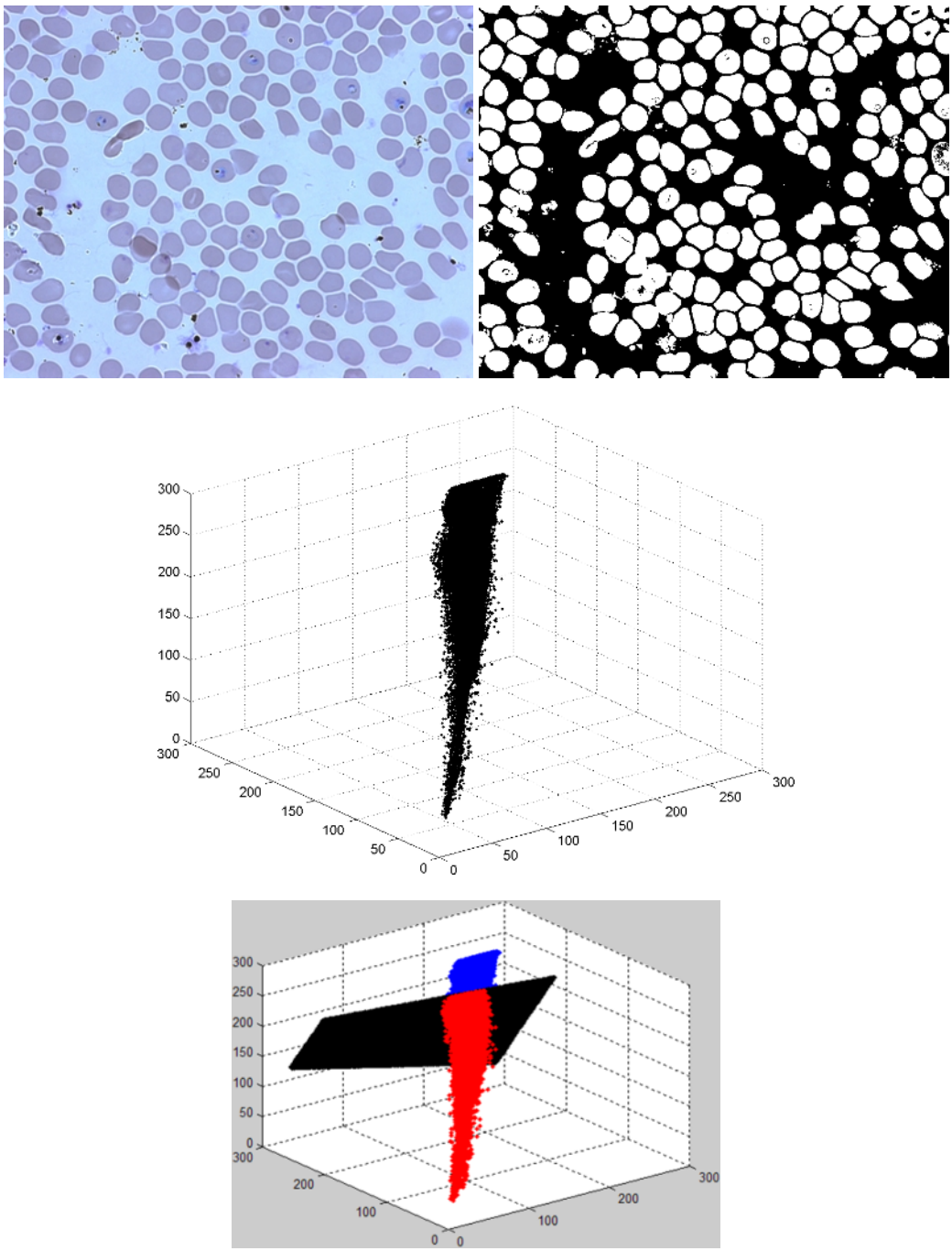


Figure 4.12: The colour image shown originally in monochrome in figure 4.2 (top) alongside the final segmentation produced by applying the Otsu 3D algorithm. Its pixel distribution is shown by occupancy of the RGB triplets in colour space (middle) with (bottom) the final decision surface separating foreground, putative RBC pixels (red) from the background (blue).

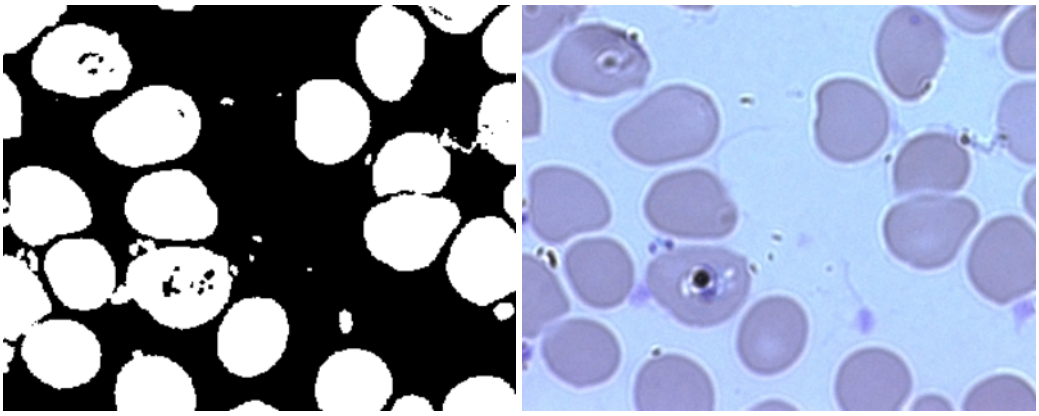


Figure 4.13: Detail from figure 4.12 showing (left) the segmentation produced by applying the Otsu 3D algorithm to the bottom-left portion of the colour image (right) displayed in monochrome in figure 4.2.

a final ROC curve for segmentation of the image as shown in figure 4.17. The area under the ROC curve, operating point, and error rate are given in table 4.2. The results are very similar to those obtained from combination of the outputs of the separate thresholding on the three colour channels and also similar to the best obtained with a single colour channel (green) shown in table 4.1. The computation time is small (2 secs for processing a 1300×1030 image on a Pentium M 1.8 GHz processor using a Matlab implementation. This is only a small multiple of the time required for the one-dimensional thresholding, unlike the search for the optimal combination described in section 4.2 which was computationally intensive and took 57 secs on the same machine.

To assess how well the various algorithms cope with variability within and between images we used the subdivision of one image into nine regions as described in section 4.1.3 and similar regions extracted from eight other images. Areas under the ROC curves, the operating points selected by the Otsu algorithms, and corresponding error rates were calculated. There was little variation within the selected image shown in figure 4.2 that was analysed in detail but, as we might expect from the fact the mean colour of the images varied by almost a factor of two, greater variation between images. Results, averaged over the selected regions from the nine images are summarized in the last column of table 4.3. On average, performance of the $3D$ extension of the Otsu algorithm is usually a little better than was obtained on the image shown in figure 4.2, in particular with a higher value of tpr. The error rates are a little lower and the two types of error somewhat better balanced. Evidently the image (figure 4.2) analysed in detail was somewhat atypical.

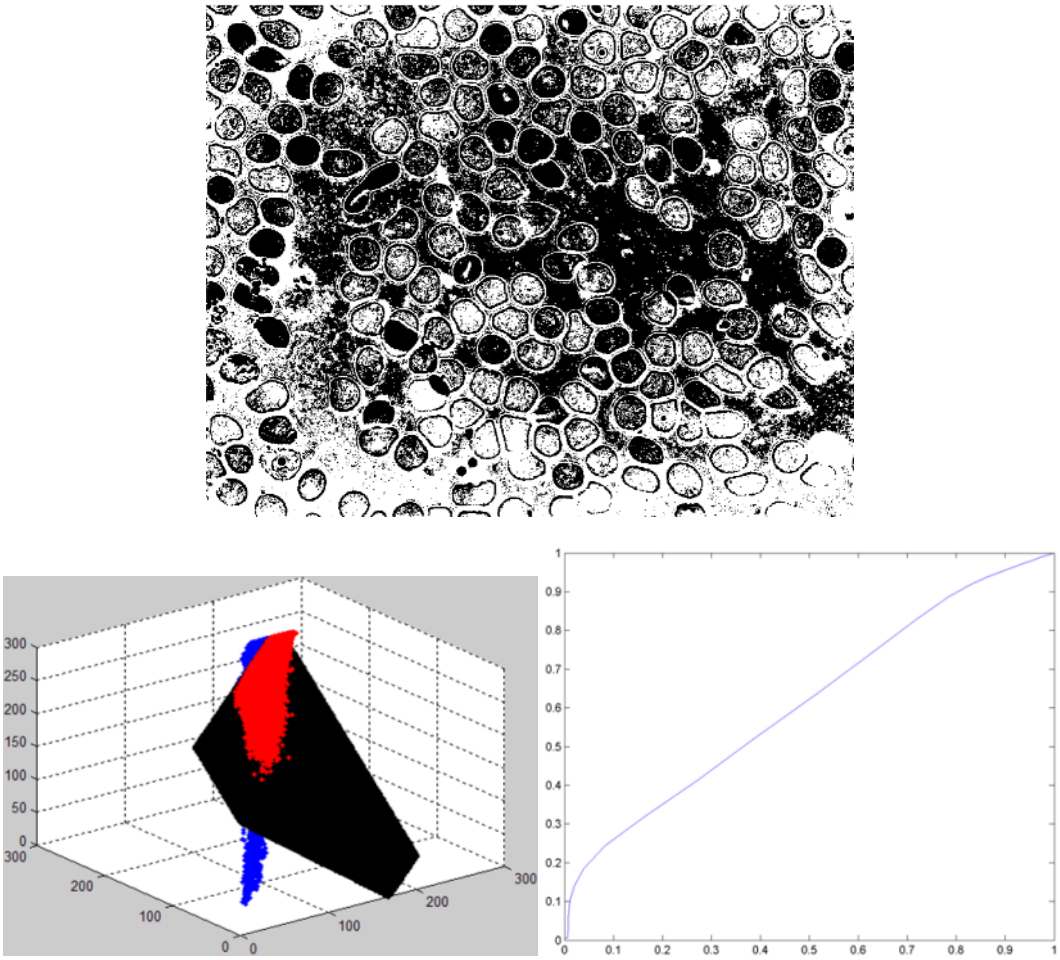


Figure 4.14: Bottom left: the initial decision surface in 3D obtained from the first iteration initialised by PCA. The decision surface is superimposed on the pixel distribution in RGB colour space shown by occupancy of the RGB triplets as in figure 4.12 but in this case in red for pixels belonging to the background and blue for pixels belonging to foreground, *i.e.* supposedly to RBCs. Top: the corresponding segmentation. Bottom right: the ROC curve obtained by varying the threshold on the distribution projected onto the initial feature vector $\underline{a}$ obtained from PCA.

	3×1D	3D
A	0.9858	0.9863
tpr	0.9574	0.9669
fpr	0.0324	0.0345
ϵ	0.0750	0.0676

Table 4.2: Performance of the system obtained by combination of the separately thresholded colour channels ($3\times$1D) and of the multi-dimensional extension of the Otsu algorithm (3D).

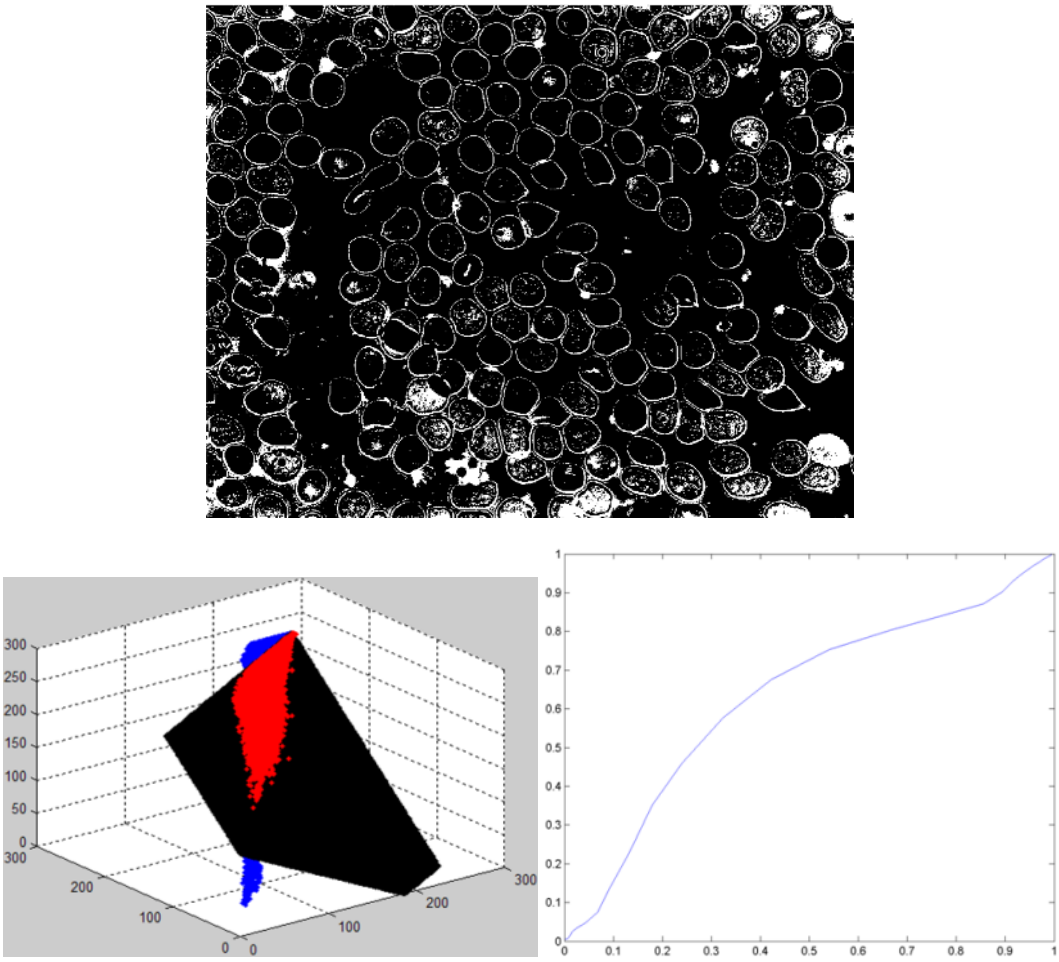


Figure 4.15: As in figure 4.14 but showing an intermediate segmentation obtained after 5 iterations (top) together with its decision surface (bottom left) and the ROC curve produced at this stage (bottom right).

	R	G	B	3×1D	3D
A	0.9880	0.9917	0.9929	0.9928	0.9930
tpr	0.9700	0.9808	0.9740	0.9768	0.9777
fpr	0.0345	0.0364	0.0270	0.0327	0.0274
ϵ	0.0646	0.0556	0.0530	0.0560	0.0497

Table 4.3: Performance measures evaluated by averaging over the seventeen selected one-ninth regions from nine different images as described in the text for thresholding algorithms applied to each of the separate colour channels (R,G,B); for the optimal combination of these algorithms ($3 \times 1D$); and for the multi-dimensional extension of Otsu's algorithm ($3D$).

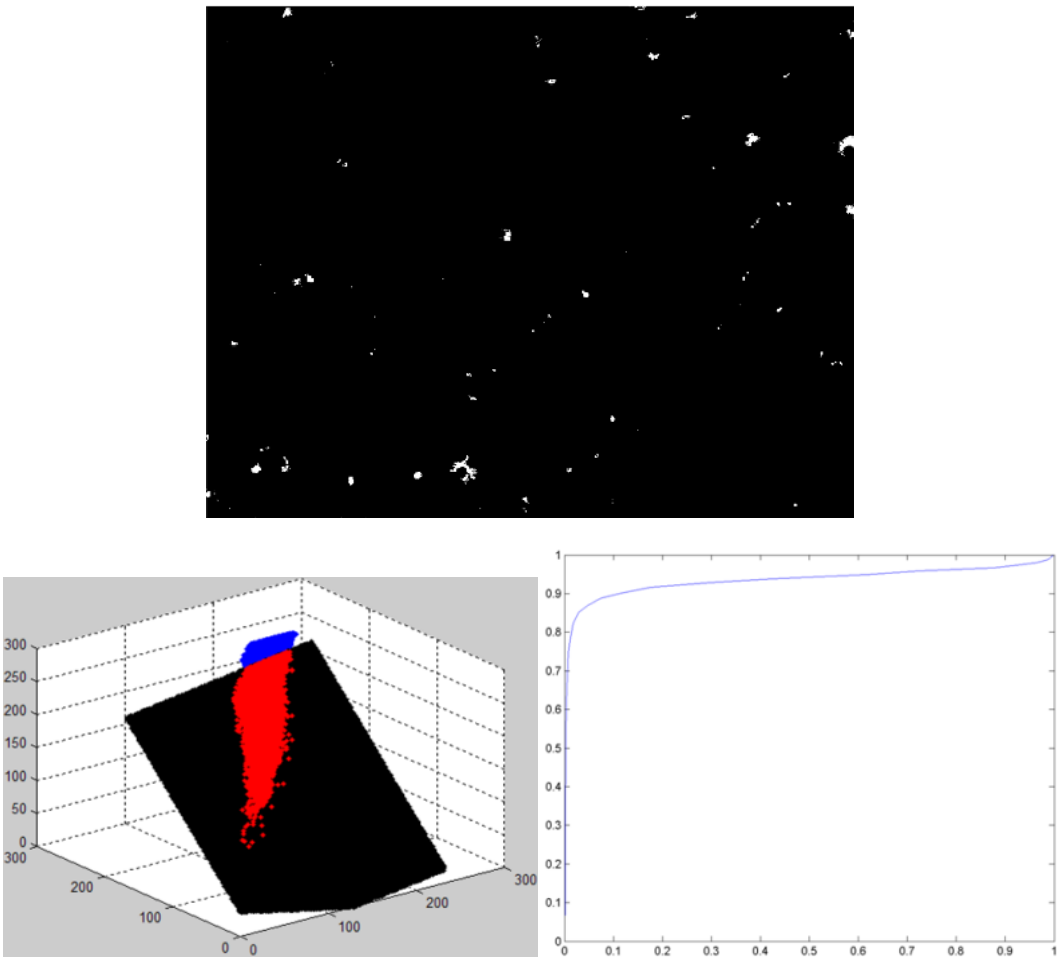


Figure 4.16: As in figure 4.15 but after 10 iterations when the orientation of the decision surface is changing most rapidly. The final decision surface is shown in figure 4.12.

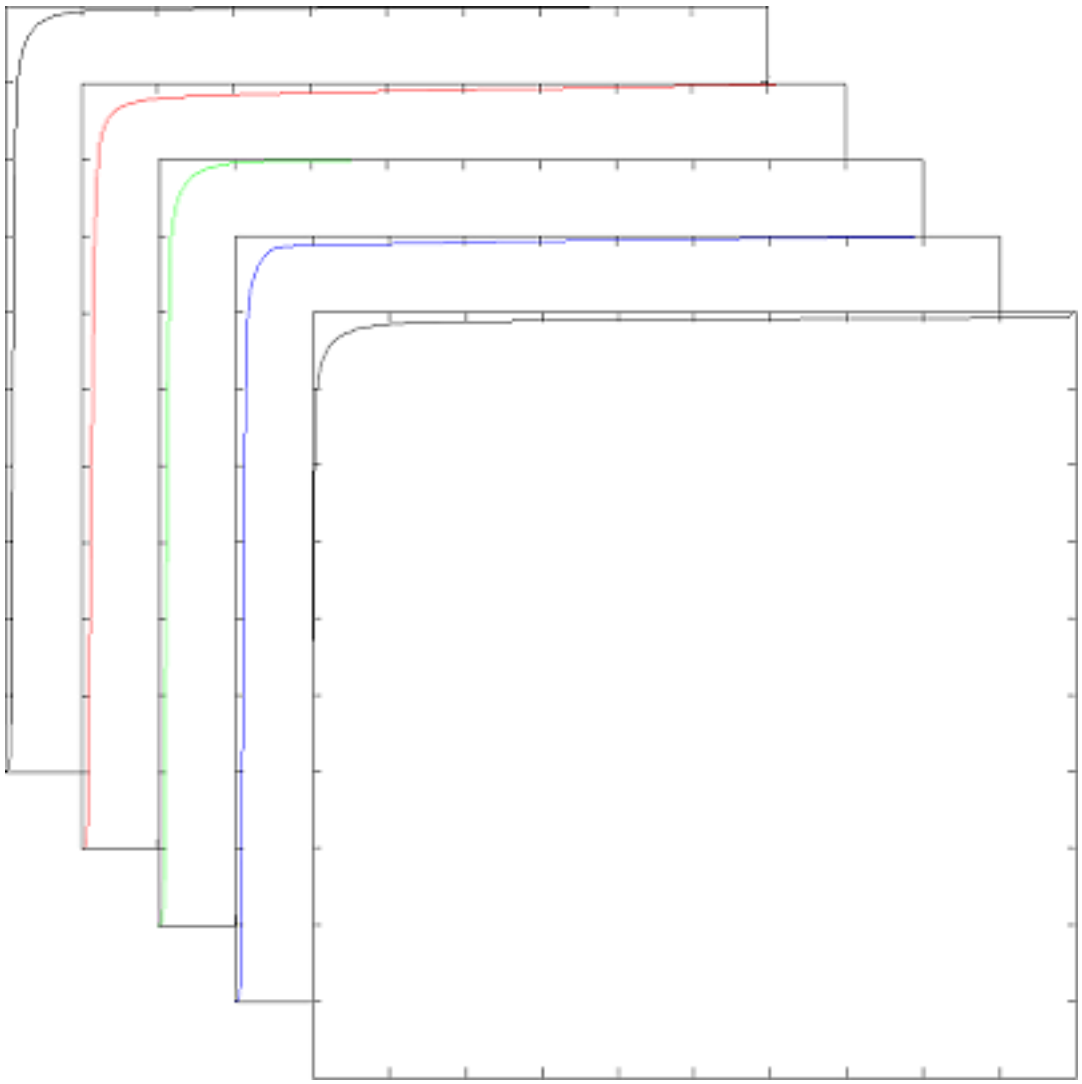


Figure 4.17: The ROC curve showing the variation of the true-positive rate tpr (y-axis) as a function of the false-positive rate fpr (x-axis) obtained by varying the threshold on the distribution projected onto the final feature vector $\underline{a}$ produced on convergence of the 3D Otsu algorithm applied to the whole colour image in figure 4.12 (front-most) superimposed on the ROC curves shown in figure 4.9 obtained by application of the Otsu algorithm in 1D to the intensity (back), green, red and blue (front) channels, respectively.

4.2.6 Sequential application to a multi-class problem

We mentioned in section 4.1.3 that, as can be seen from figure 4.2, careful inspection of the images reveals that the red-blood cell membrane is visible as distinct from the body of the cell and the background plasma. Also, some of the red blood cells are infected by malarial parasites and display structures within them where the malarial DNA has been Giemsa stained a dark-blue colour. These structures have a complicated and somewhat variable shape which, as discussed in the review chapter 2 section 2.3.1, to the expert observer indicates the type of infection and stage of the parasite's life-cycle [175, 176].

Unfortunately, as noted in section 4.2.4, we don't have a multi-class version of our 3D extension of the Otsu algorithm which might enable us to produce an algorithm that segmented these structures as well as pixels belonging to the body of the RBCs. An alternative that would seem well-suited to detection of parasite infections whose spatial structure can be very variable but which must be inside a RBC is sequential, hierarchical or recursive application of the multi-dimensional extension of the Otsu algorithm. This will be discussed in chapter 6. Here we limit ourselves to a preliminary observation illustrating that sequential application of the multi-dimensional extension of the Otsu algorithm described above can produce very interesting results, in particular a segmentation of the RBC membrane. Having used the multi-dimensional extension of the Otsu algorithm to produce a colour image segmentation of the image shown in figure 1.3 in chapter 1, we re-ran the process once on the background class and once on the foreground class and then once again on each of the two segmentation classes obtained from the foreground class to produce six final segmentation classes. Merging the minority class from the segmentation of the background with one of the four leaf classes obtained from the foreground in a tree-structured classifier combination as sketched in figure 4.18 produced a very good segmentation of the red-blood cell membrane (figure 4.19, top). However, such a pleasing result was rarely obtained with other images and seems to be sensitive to the image colour, contrast and content (RBC density). For example, carrying out this procedure on the image shown in figure 4.2 produced the rather mixed, difficult to interpret and somewhat disappointing results also shown in figure 4.19 (bottom).

4.3 Summary, discussion and conclusions

Having seen in the previous chapter that model-based approaches were unlikely to meet our needs, in this chapter the search for alternatives focused on some well-established image processing and pattern recognition techniques that depend only on very general underlying models and, since they are unsupervised, do not require training.

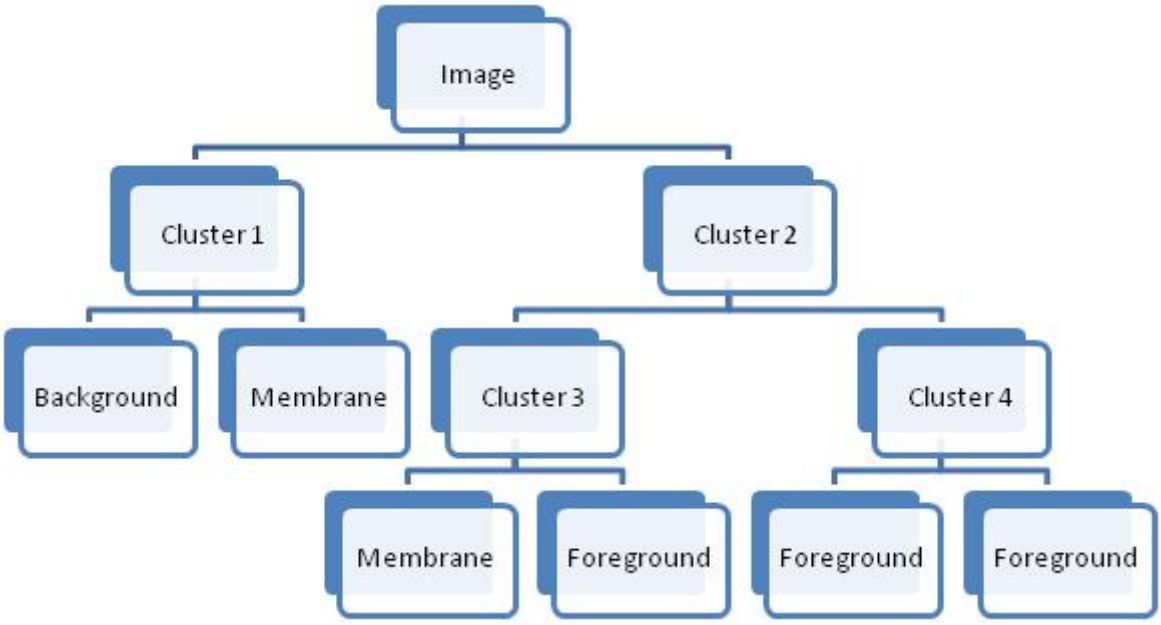


Figure 4.18: Tree structure illustrating sequential application of the Otsu 3D algorithm and merging of the outputs which detected the cell membranes shown in figure 4.19.

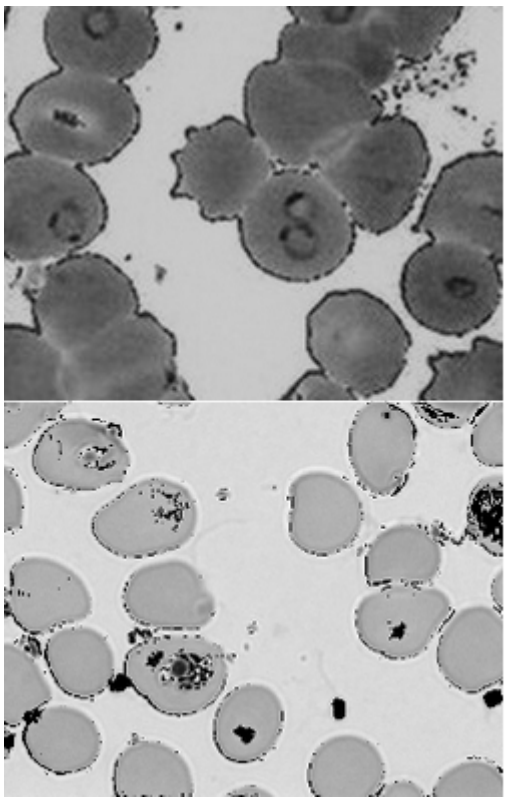


Figure 4.19: Segmentation of the red-blood cells, cell boundaries and background plasma obtained by sequential application of the Otsu 3D algorithm, shown superimposed on the grey-level image (top). A similar procedure on the image shown in figure 4.2 produced a rather different result (bottom).

(i) The Otsu algorithm for thresholding an image from its histogram was reviewed as an unsupervised procedure, with emphasis on its relationship to Fisher's discriminant expressed as maximisation of the ratio of the between-class variance to the within-class variance, σ_B^2/σ_W^2, equation 4.1. Since this relationship is often not mentioned even in image processing and computer vision texts that discuss the algorithm (section 3.8.2 of [164] and chapter 7 of [135] are exceptions – though neither gives the derivation as presented here) a derivation of Otsu's algorithm by maximisation of the Fisher discriminant was given in section 4.1.1 following the treatment in [23] with further detail in appendix A. Discussion included a multi-class (*i.e.* a multi-threshold) version of the algorithm.

(ii) This very simple algorithm was then used for the identification of pixels belonging to the red blood cells in images of thin-film microscope slides of laboratory samples of malarial infected blood. Segmentation of images from a set of 142 1300×1030 24-bit RGB images was illustrated by using the intensity and also carried out by independently thresholding each of the colour channels. The latter revealed that each colour channel carried information useful for segmentation of foreground pixels belonging to RBCs from the background plasma.

(iii) It was noted in section 4.1.1 that maximising the Fisher discriminant could also be regarded as a way of selecting the threshold in a kind of least mean square (*MSE*) classifier. Results were thus characterised by constructing an ROC curve for this classifier by varying the threshold T and computing the area under the curve in addition to noting the tpr, fpr and error rates ϵ at the operating point (*i.e.* at the threshold) obtained from the Otsu algorithm. Though there were some differences, all measures indicated that quite good segmentations could be obtained whichever of the red, green or blue colour channels or intensity was used.

(iv) Results from all three colour channels were combined by selecting the combination which, at their Otsu thresholds, maximized the sum of the between-class variance obtained from each individual colour channel. Maximum realisable ROC curves (MRROC) were then used to assess performance of the resulting classifier combination together with, as above, the appropriate true-positive, false-positive and error rates. Using all three colour channels in this manner in general led to slightly improved performance though the particular combination delivering the best performance varied from image to image.

(v) In order to relax constraints on the form of the decision surfaces when combining outputs from the colour channels as above, multi-dimensional versions of the Fisher discriminant used in the literature for the generation of optimally discriminating features in LDA were used to propose a multi-dimensional extension of the Otsu algorithm for the discrimination of two classes. This algorithm was applied to the same set of thin-film colour images and performance analysed in a similar manner to the above.

(vi) Performance evaluation required laborious labelling of the pixels by hand using the tools described in previous chapters. Just one 1300×1030 image was thus annotated in full. It was divided into a 3×3 grid of sub-images in order to provide a means of characterising variation within an image, whilst one-ninth sub-regions of eight other images were similarly annotated in order to study variation between images. There was little variation in performance within or across images. All algorithms performed well with average error rates of less than 6.5% with the multi-dimensional algorithms just superior to the best of the results obtained on a single channel.

(vii) We were unable to develop a multi-class version of the multi-dimensional extension of Otsu's algorithm. However, it was noted that sequential application of the multi-dimensional extension of the Otsu algorithm produced interesting results, as illustrated for example by detection in one image of the boundary membranes separating RBCs from the blood plasma. Though this particular result was difficult to reproduce on other images, it suggests that sequential application of the Otsu 3D algorithm may be useful in other contexts, in particular the detection of malarial parasites within RBCs.

Useful though the multi-dimension extension to Otsu's algorithm thus appears to be, we have yet to study it theoretically. For example: we cannot be sure it converges, though it appears to do so in practice, and to do so quickly. Also, though there was little variation in performance of the algorithm across the set of images investigated it is not clear why this was so and whether optimizing the colour representation or standardizing the images might lead to improved performance. We thus need a deeper understanding as to why the technique works so well. These issues will be explored in the next chapter. Here we note that a particular strength of the approach is that it is a data driven, unsupervised technique and that it does not assume any particular structure of the image objects. This is very attractive as the red-blood cells do not assume perfect geometrical shapes, such as circles or ellipses, and the parasite stains exhibit a complicated and somewhat variable shape. The extent to which the parasite shape can be modelled remains unknown and we note once again our belief that it is better to have a very

simple model than to use a more complicated but incorrect or inadequate model.

Finally, we note that the algorithms and systems developed must be thoroughly tested and evaluated if they are to find medical application, for example in research laboratories where reliable, accurate automatic determination of the degree of parasitemia would be very useful and would free medical researchers from a tedious, routine, but necessary task. The techniques required are mostly in place, including ROC and MRROC analyses and classifier combination techniques. If desired, ROC methods could be extended to multi-class problems as described in section 2.4.3 though this is not a paramount or urgent requirement for determination of the parasitemia given that the malarial parasites of interest for the calculation are only to be found inside RBCs and the required classifications may be carried out sequentially. Of great importance, however, is the provision of well-characterised or standardised data. This is a time consuming and labour intensive task. We have built a small database of annotated thin-film slide images of malarial infected blood for the research presented in this thesis. For a world-wide problem as important as malaria, however, and in particular for the development of techniques which could aid the detection and diagnosis of malarial infections in the field [176] it would be beneficial if a centralised, freely available database were prepared.

In summary, this chapter describes an important first step in the development of a system for the analysis of thin-film microscope slides for the determination of malarial parasitemia, – a multi-dimensional extension to Otsu's algorithm that works well for the segmentation of RBCs. There is also scope to apply the algorithm sequentially for application to the detection of parasite infections inside the RBC. The latter, as indicated above, will be discussed in chapter 6 together with the determination of the degree of parasitemia. First, in the next chapter, we turn to a more detailed analysis of the algorithm itself.

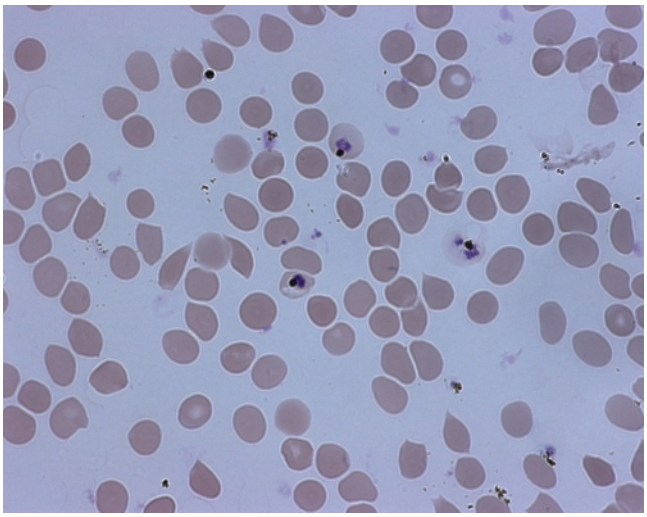
Chapter 5

Analysis of the Otsu 3D algorithm

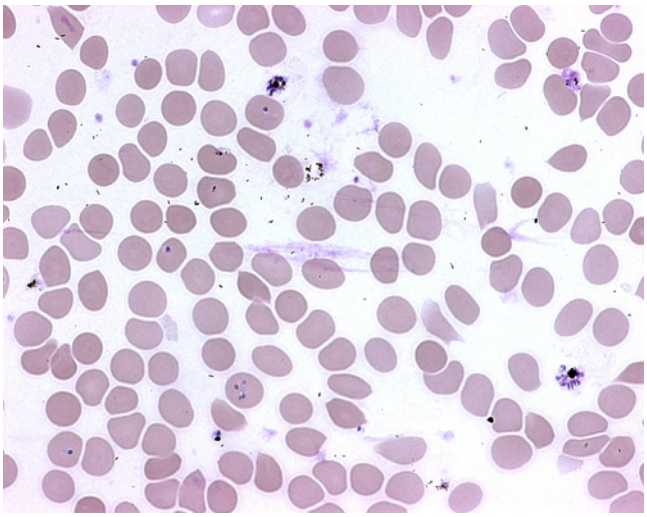
In the previous chapter we developed a multi-dimensional extension to Otsu's algorithm, which we called the Otsu 3D algorithm, and explored its application to the segmentation of RBCs in microscope images of malarial infected thin-film blood slides. The algorithm proved quite effective. In this chapter our aim is to obtain a deeper understanding of the algorithm *per se*, its behaviour, and why it proved effective for RBC segmentation. In particular, in the following we consider the algorithm's relationship to clustering techniques, its convergence and its invariance properties, and how image data variability might be characterised. Pending implementation of such a quantitative approach, the three images shown in figure 5.1 will largely be used in this and the following chapter as illustrative examples.

5.1 Clustering

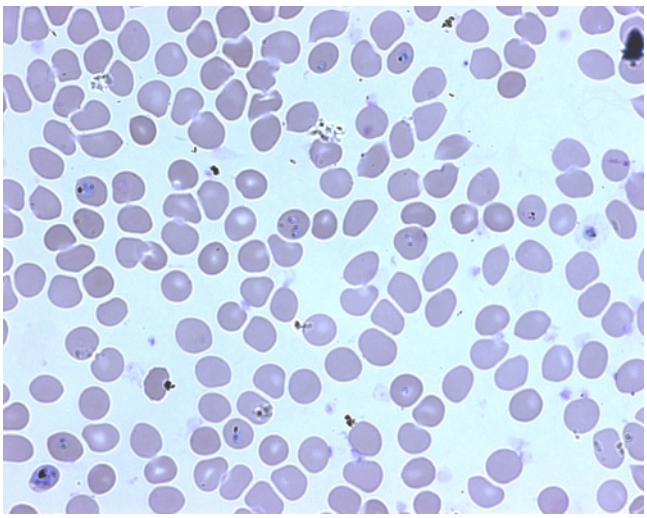
In pattern recognition, clustering is often discussed together with unsupervised learning approaches or as synonymous with such approaches (see respectively for example: [45, 46] and [178] chapters 11-16). In addition there are many whole books devoted to 'Clustering' and 'Cluster Analysis' (see for example [47] which is in its fifth edition) and an extensive Wikipedia page [195] containing many links to other pages and to applications. The Wikipedia page says "there are possibly over 100 published clustering algorithms" many of which it describes under various headings such as: "Connectivity based clustering (hierarchical clustering), Centroid-based clustering, Distribution-based clustering, Density-based clustering" and "Recent developments". There is also a section on "Evaluation of clustering results" (which the page notes "sometimes is referred to as cluster validation") that contains many clustering quality measures. Some quality measures are described as 'indices' to be used in an 'internal evaluation' when, in the spirit of unsupervised learning there is no ground truth labelling of the clusters, but many are based on quantities such as true-positive and false-positive rates familiar from the characterisation of classifier performance which can only be used in an 'external evaluation' when



[1]



[4]



[45]

Figure 5.1: Three images, respectively numbers 1 (top), 4 (middle) and 45 (bottom) used as illustrative examples.

ground truth data is available.

To discuss all of this literature would take us too far afield. It is still growing, in particular in areas such as robust clustering when the signal data may be contaminated by outliers and when estimation of the appropriate number of clusters to use is required (see for example [80] and [81]). In the context of our work, the important distinctions are between *heuristic* approaches and those based on *statistical models* which, as usual, may be either parametric or non-parametric. Parametric models include Gaussian mixture models which, as noted in section 4.1 are much used in colour image segmentation, whilst non-parametric statistical partitioning methods are often based on minimisation of the trace or determinant of a pooled within-cluster (or class) scatter or covariance matrix such as S_W. According to [63] methods based on S_W may be regarded either as heuristically motivated or as maximum likelihood estimators of well-defined statistical models.

5.1.1 C-means clustering

Known widely as *k-means clustering*[1] the aim in this approach is to assign data $\underline{x}(i), i = 1 \cdots n$ to classes or clusters $J = 1 \cdots C$ so as to:

$$\min \left\{ D_E = \sum_{J=1}^{C} \sum_{i \in J} [\underline{x}(i) - \underline{\mu}(J)]^2 \right\} \tag{5.1}$$

where, as usual, $\underline{\mu}(J)$ are the class-conditioned or 'cluster' means:

$$\underline{\mu}(J) = \frac{1}{n(J)} \sum_{i \in J} \underline{x}(i). \tag{5.2}$$

It is evident that this criterion is based on a Euclidean measure of the distance within each cluster and self-consistently assigns each datum $\underline{x}(i)$ to the cluster whose mean is nearest to $\underline{x}(i)$. Not only might we consider such a particular measure based on a Euclidean distance better replaced by one using Mahanalobis distances based on class-conditioned matrices $S(J)$ but, in cluster analysis in general, the whole issue of the use and type of a distance or *dissimilarity* measure, of a *similarity* measure, or of more general *connectivity* measures arises. Duda, Hart and Stork [46], section 10.6 contains a useful discussion of such questions which are also well exemplified by the organisation of the Wikipedia pages ([195]).

Since there are approximately $C^n/C!$ ways of partitioning n examples into C such subsets[2] the minimisation 5.1 is *NP-hard* and approximate methods are required in practice. The usual approach is to adjust the cluster membership iteratively which, as [46] describes, may be

[1] For consistency, we shall refer to this as *C-means* clustering as our aim is to produce C clusters corresponding to C classes; *cf.* [46] section 10.4.3, page 526.

[2] See for example [46], section 10.8, who also give the exact number as the solution of problem 18 to chapter 10.

implemented either in batch or sequential mode. The basic procedure is a greedy, 'hill-climbing' algorithm and usually converges quite quickly with a complexity $O(ndC)$ per iteration. Finding the optimal clustering is not guaranteed so in practice the algorithm is often run multiple times from random initialisations. However, given the large number of distinct clusterings, this may often fail to find the optimum or even a consensus as to the best estimate of the optimum especially if clusters in the data are not very distinct or obvious. Alternatively, solutions with $C-1$ clusters may be used to initialise the procedure in a hierarchical manner ([46, 117]) or one of the hierarchical clustering algorithms used to provide a starting point though these may have a higher complexity [46]. The hierarchical algorithms may involve either 'merging' or 'splitting' clusters and are very like those used for segmentation in image processing discussed in section 2.3.3. The properties of the iterative algorithms have been widely discussed in the literature, in particular in signal processing where the common version is often known as Lloyd's algorithm or the LBG algorithm for vector quantisation [115, 108] and may be found in probabilistic, soft or fuzzy versions (see for example [178], chapter 14 and references therein). In pattern recognition it is sometimes known as the 'isodata' algorithm [43].

5.1.2 The Otsu algorithm and C-means clustering in 1D

Criterion 5.1 is just the within-class variance:

$$\sigma_W^2 = tr(S_W) \tag{5.3}$$

so it is easily seen that minimisation 5.1 is equivalent to maximising the between-class variance $\sigma_B^2 = tr(S_B)$. We should thus expect C-means clustering to be closely related to the Otsu algorithms discussed in the previous chapter – and in fact in a one-dimensional feature space it is immediately obvious that they are designed to optimise equivalent criteria. Moreover, in a one-dimensional feature space the Otsu algorithm with thresholds

$$T_J = \frac{1}{2}(\mu(J) + \mu(J+1)) \tag{5.4}$$

is precisely equivalent to self-consistently assigning each datum $x(i)$ to the class or cluster defined by the nearest mean $\mu(J)$. We thus see that the C-class Otsu algorithm and a C-means algorithm are carrying out the same processes and, from the same initialisation, will produce identical results.

5.1.3 The Otsu algorithm and C-means clustering in 3D

When the feature space is multi-dimensional our Otsu 3D algorithm has only been developed for the two class case so comparison with a 2-means algorithm is all that is required. As indicated in

the Wikipedia pages [195] and noted above, the C-means algorithm separates data into Voronoi-cells centred on each of the cluster (or class) means $\underline{\mu}(J)$. The cluster boundary produced by the 2-means algorithm is thus always perpendicular to the line joining the cluster means, $\underline{\mu}(2) - \underline{\mu}(1)$. However, this is *not* the case for our Otsu 3D algorithm where the boundary is normal to the optimal feature vector

$$\underline{a} \propto S_W^{-1}(\underline{\mu}(2) - \underline{\mu}(1)). \tag{5.5}$$

Furthermore , its location is determined in the second stage of our 3D algorithm by application of the one-dimensional Otsu algorithm to the histogram of the data projected onto this direction. As we have seen, the second step is equivalent to using a 2-means algorithm on the histogram so differences between the two algorithms (2-means and our Otsu 3D) will arise from the effect of the matrix inverse ${S_W}^{-1}$. Unless S_W is isotropic, the two algorithms will thus produce different outcomes even when both converge to their desired optima.

5.1.4 Effect of the within-class covariance

To characterise the effect of the within-class covariance matrix S_W obtained on convergence of the Otsu 3D algorithm for the segmentation of the images shown in figure 5.1 the acute angle α between the optimal feature vector $\underline{a}$ and $\underline{\mu}(2) - \underline{\mu}(1)$ was calculated:

$$\cos(\alpha) = \frac{(|\underline{\mu}(2) - \underline{\mu}(1))^T S_W^{-1}(\underline{\mu}(2) - \underline{\mu}(1))|}{|\underline{\mu}(2) - \underline{\mu}(1)||S_W^{-1}(\underline{\mu}(2) - \underline{\mu}(1))|}. \tag{5.6}$$

Results are given in the first three rows of table 5.1. The angle α shows how much the factor S_W^{-1} in 5.5 makes the normal to the decision surface deviate from the direction $\underline{\mu}(2) - \underline{\mu}(1)$ it would have taken in a C-means clustering algorithm given these class labels. It can be seen that this deviation varies considerably, from around $30\,\mathrm{deg}$ in images 1 and 45 to just over $60\,\mathrm{deg}$ in image 4, but is never insignificant.

The eigenvalues and eigenvectors of the within-class covariance S_W for each of these images are also shown in table 5.1 together with the acute angles each makes with $\underline{\mu}(2) - \underline{\mu}(1)$. It can be seen that S_W is highly anisotropic in each case, in particular for image 4 where the principal eigenvalue accounts for 98% of the variance but only around $70 - 80\%$ for the other two images. The within-class covariance is also oriented quite similarly at an acute angle of less than $30\,\mathrm{deg}$ and often much smaller with respect to $\underline{\mu}(2) - \underline{\mu}(1)$ in each case although for image 1 the principle axis is more closely parallel to $\underline{\mu}(2) - \underline{\mu}(1)$ than for the other two images.

It was also noticed that the three vectors defined by the principal axis of S_W, $\underline{a}$ and $\Delta\underline{\mu} = \underline{\mu}(2) - \underline{\mu}(1)$ were in each case almost co-planar, the magnitude of the scalar triple products of the corresponding unit vectors in no case exceeding 0.03. The acute angles between $\underline{a}$ and $\Delta\underline{\mu}$

	Image 1		Image 4		Image 45	
$\underline{\mu}(2)-\underline{\mu}(1)$	-29.701 -48.472 -58.744		-53.768 -77.965 -63.993		-49.829 -73.906 -43.488	
$\underline{\hat{a}}$	0.0472 -0.5254 -0.8495		0.1484 0.0029 -0.9889		0.0472 -0.7229 -0.6894	
Angle α deg	25.2		61.2		35.2	
Eigenvalues	92060 19404 5446		1682017 18138 15836		265597 65383 36634	
Eigensolution	Eigenvector	Angle θ	Eigenvector	Angle θ	Eigenvector	Angle θ
First	0.4730 0.5587 0.6813	6.9	-0.6344 -0.7644 -0.1159	27.8	0.7282 0.6238 0.2838	17.2
Second	0.8725 -0.4047 -0.2739	83.1	-0.7657 0.6003 0.2311	79.7	0.6770 -0.5906 -0.4391	73.0
Third	-0.1227 -0.7240 0.6788	89.2	0.1071 -0.2353 0.9660	64.5	0.1063 -0.5119 0.8525	87.4

Table 5.1: First three rows: values of $\underline{\mu}(2)-\underline{\mu}(1)$, the optimal feature vector $\underline{\hat{a}}$ and the acute angle α in degrees between them for each of three illustrative images, numbers: 1, 4 and 45 shown in figure 5.1. Eigenvalues of the within-class covariance matrix and their eigenvectors are shown in the remainder of the table together with the acute angles θ between each eigenvector and $\underline{\mu}(2)-\underline{\mu}(1)$.

and between the principal eigenvector, $\underline{v}(1)$ say, of the within-class covariance and $\Delta\underline{\mu}$ may thus be added to give, at least for these three illustrative images, to within less than $1\,\mathrm{deg}$ the angles between $\underline{a}$ and the principal eigenvector $\underline{v}(1)$.

Using the spectral decomposition of S_W it may be shown that

$$\underline{a} \propto \sum_{k=1}^{3} \left[\frac{\underline{v}(k)^T \Delta\underline{\mu}}{\lambda(k)} \right] \underline{v}(k) \tag{5.7}$$

where $\lambda(k)$ and $\underline{v}(k)$ are the eigenvalues and eigenvectors of S_W respectively with the principal eigenvector belonging to the largest eigenvalue labelled as $\underline{v}(1)$. It then follows from 5.7 that

the scalar triple product[3]

$$\{\underline{v}(1), \hat{\underline{a}}, \Delta\underline{\mu}\} = \frac{(\underline{v}(2)^T \Delta\underline{\mu})(\underline{v}(3)^T \Delta\underline{\mu})[\frac{1}{\lambda(2)} - \frac{1}{\lambda(3)}]}{|\Delta\underline{\mu}| \sqrt{\sum_{k=1}^{3} [\frac{\underline{v}(k)^T \Delta\underline{\mu}}{\lambda(k)}]^2}}. \tag{5.8}$$

The three vectors would thus be coplanar if $\lambda(2) \approxeq \lambda(3)$ but, it can be seen from table 5.1 that this is not the case. If these vectors were to be approximately coplanar in general, there would evidently be some interplay between the values of the inner products and eigenvalues in 5.8 which would make it so. We have not been able to ascertain what this might be so the approximate co-planarity of the vectors remains a somewhat surprising empirical observation, supported only by the evidence from these three illustrative images.

5.2 Convergence of the Otsu algorithms

It is well-known as we noted in section 5.1.1 that the C-means algorithms converge to a local optimum. The behaviour of the Otsu algorithm in one-dimension will be similar and thus at best we may expect our multi-dimensional extension, Otsu 3D, also to converge to a local optimum. For all the examples of segmentation of malaria thin-film slide images into two classes of pixels belonging to RBCs and to background that we have run, neither the original one-dimensional Otsu algorithm (applied either to the intensity or to a single colour channel) nor our multi-dimensional extension, the Otsu 3D algorithm, ever seemed to produce unsatisfactory results. It is known that the original Otsu algorithm in a one-dimensional feature space "breaks down when the two populations are very unequal" ([135] page 282)[4] so it would appear that in all our examples, the numbers of pixels in the background and belonging to RBCs were not too unequal – in fact, as usual for blood thin-film slide images, the ratio was often $\sim 40 - 50\%$.

5.2.1 The basin of attraction of the Otsu algorithm

The closeness of the fraction of RBC pixels in an image to 50% and the clear bimodality of the histograms shown in chapter 4 indicate from the remarks of Petrou and Bosdogianni [135] not only that the Otsu algorithm should converge but why it produces such good segmentations of the RBCs. In addition, the form of the histograms shows that we may expect initialisation of the algorithm with the threshold $T = \mu$ to be quite close to the final converged value of T but to understand more fully the robustness of the algorithm and why it produced good results for *every* example on which it was run, we need to investigate its basin of attraction.

[3] Provided the eigenvectors of S_W form a right-handed set; if they are left-handed the sign of the result changes.

[4] Similarly, we note that, for example according to the Wikipedia article [195], ideally the C-means algorithm "assumes equal-sized clusters".

Since the feature space is one-dimensional it is easy to search the entire range $0 - 255$ for the optimal value of the threshold T which maximises the Fisher discriminant[5], or equivalently, the between-class variance σ_B^2. Doing so shows that the between-class variance, and thus the Fisher discriminant, each have a single maximum (figure 5.2) and confirms that in this case the optimal threshold for intensity-based segmentation is $T = 155$ and that the Otsu algorithm will converge to this threshold.

These results are typical (see for example further results in section 6.3.1) and indicate why, for intensity-based segmentation of RBC pixels, the Otsu algorithm always converged and produced good results.

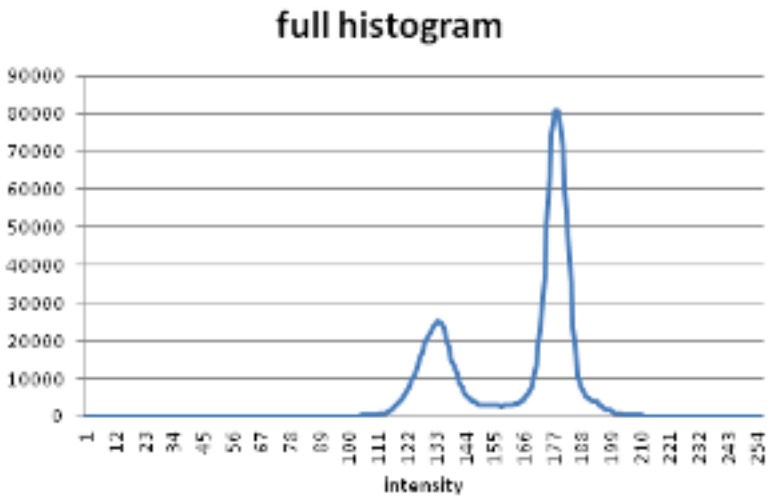
5.2.2 The basins of attraction of the Otsu 3D algorithm

Characterising the behaviour of our extension of Otsu's algorithm is more difficult because the higher dimensionality complicates both complete search of the parameter space and easy visualisation of what the algorithm does. In the previous chapter (section 4.12 and figures 4.14 - 4.16) we saw an example of how the algorithm converged from initialisation via PCA but to characterise the basins of attraction of the algorithm would need similar results for every possible initialisation. This is not feasible, but having seen how the 1D Otsu algorithm used in the second stage of each iteration of the 3D algorithm behaves, we focus on using the direction of the vector $\underline{a}$ in the first stage of the algorithm to characterise the behaviour of our Otsu 3D algorithm.

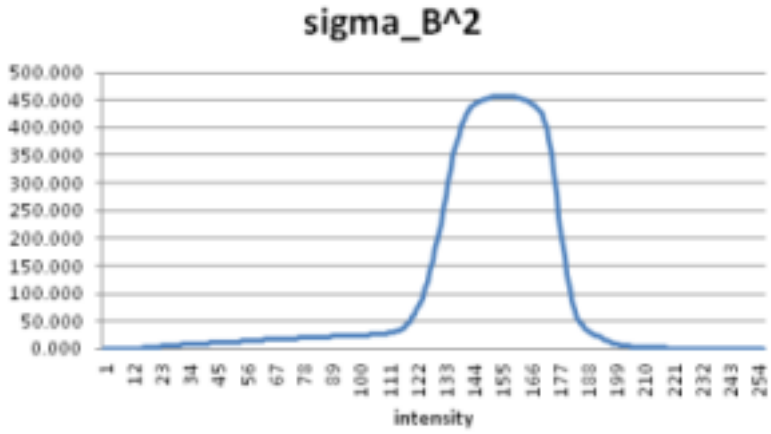
Using the image shown in chapter 4 figure 4.12 as a typical example, the Otsu 3D algorithm was run 1000 times using directions of the feature vector $\hat{\underline{a}}$ on the surface of the unit sphere [121] to initialise the algorithm rather than PCA as described in section 4.2.4. In each case we checked whether the algorithm converged successfully (S) to a good segmentation (figure 5.3 (a)) like that shown in chapter 4, figure 4.12. We also recorded when initialisation of the Otsu 3D algorithm converged to a poor segmentation (U) and when it failed (F) by reaching a set limit of 100 on the number of updates of $\hat{\underline{a}}$. Results are summarised in figure 5.3 (b). It can be seen that most initialisations led to successful segmentations, that there are two bands on the unit sphere where the initial choice of $\hat{\underline{a}}$ led to an unsuccessful segmentation and a few choices, lying on narrow stripes on the surface of the unit sphere, for which the algorithm failed to converge.

Strictly, the fact that the between-class variance for an intensity-based segmentation of

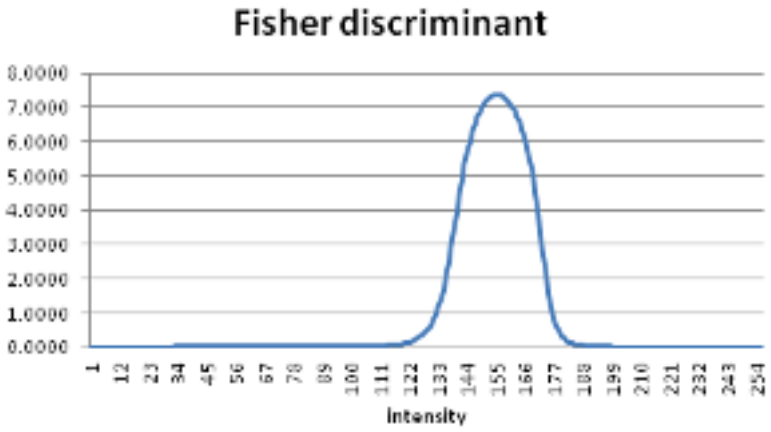
[5]There are remarks in [135] pages 281-282 and [164] page 87 about such sequential searching, checking that maxima of σ_B^2 have been found and other issues, including the use of efficient recursive algorithms (see appendix E), but all we need here is the between-class variance as a function of T.



(a)



(b)



(c)

Figure 5.2: The intensity histogram (a) for image number 1 shown in figure 5.1, the between-class variance σ_B^2 (b) as a function of the threshold T, and (c) the corresponding Fisher discriminant. The histogram has two, well-separated distinct peaks whilst both the between-class variance and Fisher discriminant each have a unique, well-defined maximum – the maximum in the latter being somewhat more prounced.

RBC pixels is well-behaved with a single maximum does not suffice to ensure that the 1D Otsu stage of the algorithm will always converge. Figures 4.12 and 4.14 - 4.16 suggest that the initial and intermediate stages of the Otsu 3D algorithm may produce distributions of the one-dimensional discriminating feature that are not similarly well-behaved with a single maximum. As a double-check we therefore re-ran the whole experiment with an exhaustive search for the threshold used in each application of the 1D Otsu algorithm. This double-check produced no significant difference when the segmentation was successful.

5.3 Invariance properties

It is noted in Duda, Hart and Stork (Chapter 3, section 3.8.3, p124, [46], and in the earlier [45], Chapter 4, section 4.11, p121) that solution for the matrix of discriminating directions, A, derived from the various Fisher discriminant criteria is invariant to transformations which include "rotation and scaling of the axes in various ways" which are "linear transformations from a $(C-1)$ dimensional space to a $(C-1)$ dimensional space". Of greater interest, however, are transformations of the d-dimensional space of the original data which Duda, Hart and Stork only discuss in the context of clustering when the feature matrix A is redundant ([46], chapter 10, p546 and p589, problem 28). In particular in the malarial application such transformations can represent colour transformations reflecting variability in the image data as discussed in chapter 2, section 2.2.1.

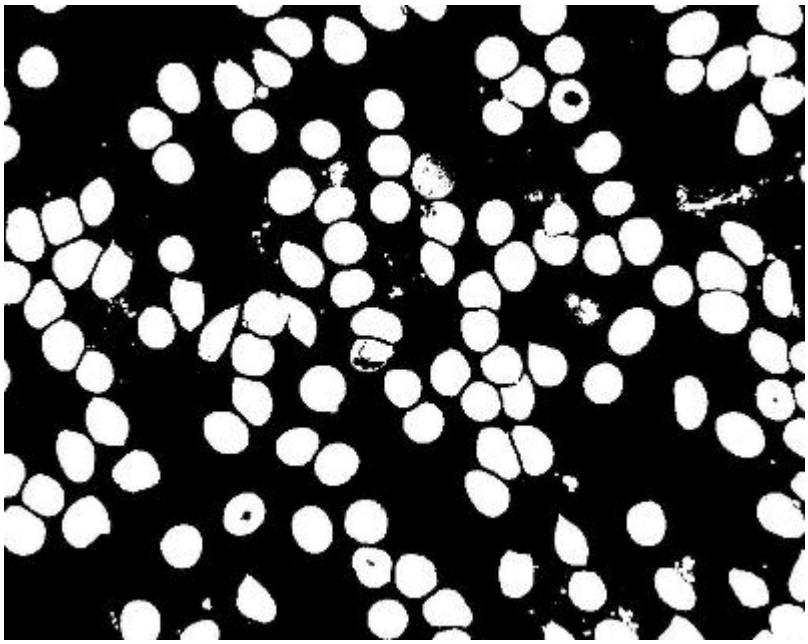
All the discrimination criteria discussed in chapter 4, section 4.2.2: the three relative criteria, D_2, D_3, D_4 and the absolute criterion D_1, are based on matrix expressions of the form

$$A^T S A = P \tag{5.9}$$

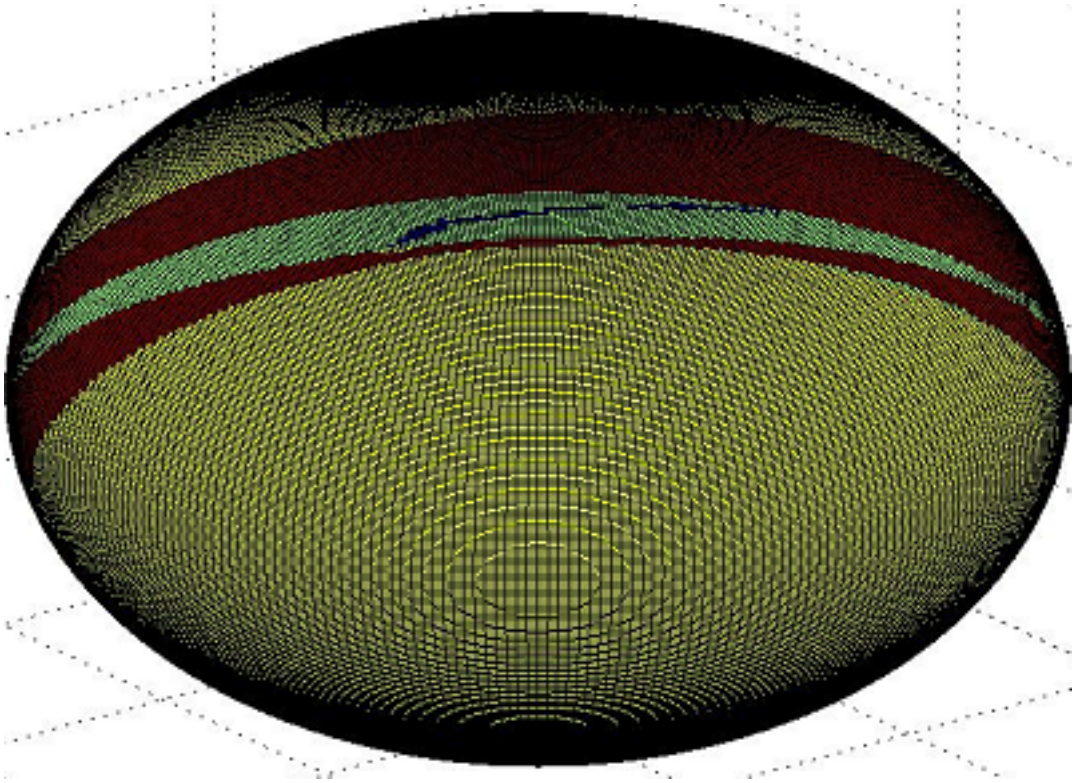
(say) where S is a $d \times d$ covariance matrix of some type and A is a $d \times r$ dimensional feature matrix. Non-linear transformations of the colour space are sometimes used, especially for improving the display of image data, but these won't lead to straightforward transformation of the covariance matrices, so we restrict attention to affine transformations of the form

$$\underline{x}^{'}(i) = M\underline{x}(i) + \underline{b} \tag{5.10}$$

with some non-singular matrix M and an additive constant or colour bias vector $\underline{b}$. In the malaria application i labels the pixels within an image or sub-image region. The affine transformation $(M, \underline{b})$ may have from one to twelve degrees of freedom depending on what restrictions are placed on the matrix M and the vector $\underline{b}$. Special forms of interest include: for M the unit matrix, isotropic and diagonal matrices; and for $\underline{b}$ the zero vector (leading to linear transformations), and constant vectors of the form $\underline{b} = b(1,1,1)^T$ leading to isotropic biasing of the



(a)



(b)

Figure 5.3: (a) A good segmentation of RBC pixels. (b) Feature vectors $\underline{\hat{a}}$ used to test convergence of the Otsu 3D algorithm as described in the text. Those which converged to a successful segmentation (S) are shown as points in green on the surface of the unit sphere whilst those which produced a poor segmentation (U) are shown in red and those for which the algorithm failed to converge in 100 iterations (F) are shown blue.

colours. There are thus $4 \times 3 - 1 = 11$ different types of affine transformation of the colour space which may be regarded in general as not contrived. They include isotropic and anisotropic scaling of the colour axes, with or without corresponding colour biases or off-sets.

Since the mean colour of on an image or image region transforms in the same manner as the $\underline{x}(i)$ it follows that the total covariance matrix S_T transforms as:

$$S_T' = M S_T M^T \Leftrightarrow S_T = M^{-1} S_T' (M^{-1})^T. \tag{5.11}$$

Furthermore, since the (actual) class membership of a pixel is unaffected by the transformation it is easy to see that the class-conditioned means $\underline{\mu}(J)$ also transform according to 5.9 and thus that the class-conditioned covariance matrices $S(J)$, the within-class covariance S_W and the between-class covariance S_B will all transform as indicated in 5.11. Any expression of the form 5.9 will therefore be invariant with:

$$P' = A'^T S' A' = P \tag{5.12}$$

if the feature matrix transforms according to:

$$A' = (M^{-1})^T A \Leftrightarrow A = M^T A'. \tag{5.13}$$

5.3.1 Invariance of the Fisher discriminant criteria

It follows immediately from the above that the Fisher criteria D_2, D_3, D_4 are invariant under non-singular affine transformations of the data. This can also be seen from the fact that optimisation of the criteria leads to the same generalised eigenequation 4.39

$$S_B \underline{a}(k) = \lambda(k) S_W \underline{a}(k) \quad k = 1 \cdots r_B. \tag{5.14}$$

and that the optimum may in each case be written in terms of the eigenvalues $\lambda(k)$ (see for example, appendix D) which are invariant under transformation of the covariance matrices in 5.14. Furthermore, we see from 5.14 that the feature vectors themselves may be transformed as

$$\underline{a}'(k) \propto (M^{-1})^T \underline{a}(k) \Leftrightarrow \underline{a}(k) \propto M^T \underline{a}'(k) \tag{5.15}$$

which is consistent with 5.13. The proportionality in 5.15 highlights the fact that only the directions of the feature vectors $\underline{a}(k)$ are important.

It is notable that, unless M is orthogonal or proportional to an orthogonal matrix, the feature vectors $\underline{a}(k)$ do *not* transform in a manner precisely similar to transformation of the data $\underline{x}(i)$ 5.10. To understand this we recall the two-class case where the Fisher feature vector

$\underline{a}$ is given explicitly by 5.5. This shows that, owing to the presence of the ${S_W}^{-1}$, the feature vector must in general transform as:

$$\underline{a} \propto M^T \underline{a}^{'} \tag{5.16}$$

consistent with 5.15.

5.3.2 Effect of a transformation on the Otsu algorithm

Whilst it is pleasing to see that the Fisher criteria are invariant under affine transformations of the data 5.10, this only indicates similar invariance of a segmentation of an image in which the class labels of the pixels $\underline{x}(i)$ are fixed. Invariance of the algorithms and of the results they produce requires that we demonstrate:

(i) each iteration of the algorithm transforms in an appropriate manner,

(ii) as does its initialisation.

We consider first a one-dimensional feature space in which the data $x(i)$ are subject to

$$x^{'}(i) = mx(i) + b \tag{5.17}$$

which may represent a scaling or gain change m and a translation b, a shift or bias, of a single colour attribute or of the image intensity. On using 5.17 to change variables in the derivation of the Otsu algorithm developed in chapter 4, sections 4.1.1, 4.1.4 and appendix A we find that, provided the thresholds are transformed according to 5.17 with

$$T_J^{'} = mT_J + b \tag{5.18}$$

that the indicator variable transforms as

$$Z(J) = Z(J) \quad J = 1 \cdots C \tag{5.19}$$

and thus that the class-conditioned means defined by the thresholds T_J as in 4.10 or 4.28 also transform according to 5.17:

$$\mu^{'}(J) = m\mu(J) + b \quad J = 1 \cdots C. \tag{5.20}$$

It then follows that:

$$T_J^{'} = \frac{1}{2}[\mu^{'}(J) + \mu^{'}(J+1)] \quad J = 1 \cdots C-1, \tag{5.21}$$

and we may conclude that the iterative step in the one-dimensional Otsu algorithm is invariant under an affine transformation of the data $x(i)$. Provided it is initialised in a manner that transforms the initial T_J according to 5.18 or the means $\mu(J)$ according to 5.20 the whole algorithm

will be invariant. In particular, the two-class Otsu algorithm is invariant since in this case the threshold T is initially chosen to be the mean μ which we already know transforms in the same manner as the data $x(i)$ according to 5.17.

Finally we note that, since all variances transform as:

$$\sigma^{'2} = m^2\sigma^2, \tag{5.22}$$

the maximium value of the between-class variance σ_B^2 is scaled but its location is invariant. In one-dimension not only is the location of the maximum of the Fisher discriminant $\max\{\sigma_B^2/\sigma_W^2\}$ invariant, but so also is the discriminant itself.

5.3.3 Effect of a transformation on the extended Otsu 3D algorithm

Each iteration of our multi-dimensional extension of the Otsu algorithm is comprised of two steps:

(i) update of the direction of the Fisher feature vector $\underline{a}$ according to 4.48

$$\underline{a} \propto S_W^{-1}(\underline{\mu}(2) - \underline{\mu}(1)), \tag{5.23}$$

(ii) projection of each datum $\underline{x}(i)$ onto $\underline{a}$ to define a one-dimensional feature datum:

$$x(i) = \hat{\underline{a}}^T\underline{x}(i), \tag{5.24}$$

and subsequent application of the one-dimensional Otsu algorithm to segment the data $\underline{x}(i)$ according whether the $x(i)$ is greater than or less than a threshold T.

From the transformations of the data $\underline{x}(i)$ 5.10 and of the feature vector $\underline{a}$ 5.16 it follows that we may say:

$$\hat{\underline{a}}^{'} = m(M^{-1})^T\hat{\underline{a}}, \tag{5.25}$$

where m is a scaling dependent on $\underline{a}$ and M (or equivalently on $\underline{a}$ and $\underline{a}^{'}$). We can then say that the projected variable $x(i)$ undergoes an affine transformation like 5.17 in which the scaling factor m is as in 5.25 and the additive constant:

$$b^{'} = m\hat{\underline{a}}^T M^{-1}\underline{b}. \tag{5.26}$$

Thus, we ultimately see that both steps in the iteration of our multi-dimensional extension of the Otsu algorithm transform in an invariant manner and that, given an initialisation that is similarly invariant, the whole algorithm would be invariant.

Unfortunately, initially setting $\underline{a} = \underline{p}(1)$ where $\underline{p}(1)$ is the mode corresponding to the largest variance in a PCA of the total covariance S_T is *not* in general invariant. Under an affine

transformation of the data 5.10, PCA modes $\underline{p}(k)$ may only be invariant if the matrix M is orthogonal or proportional to an orthogonal matrix since:

$$M^{-1}S_T'(M^{-1})^T\underline{p}(k) = \sigma^2(k)\underline{p}(k) \Rightarrow S_T'(M^{-1})^T\underline{p}(k) = \sigma^2(k)M\underline{p}(k). \tag{5.27}$$

Since there is no obvious alternative to a deterministic initialisation of $\underline{a}$ via PCA that would transform in the required manner, we have to concede that our multi-dimensional extension of the Otsu algorithm is *not* invariant under a general affine transformation 5.10. It is invariant when M is orthogonal or proportional to an orthogonal matrix but such transformations are, in an imaging application such as our malaria problem, very restrictive and of little use as they exclude for example independent rescaling of each colour channel.

All is not lost, however, for the final segmentation produced by our algorithm. We have seen previously in section 5.2.2 that the basin of attraction in which the direction of the initial feature may be chosen so that the algorithm converges to a successful segmentation is large. Provided the PCA mode $\underline{p}'(1)$ of the transformed total covariance matrix S_T' lies in the same basin of attraction as $\underline{p}(1)$, the corresponding mode of variation of S_T, the algorithm should result in the same (or, depending on the termination condition used, very nearly the same) segmentation.

Finally, we note that conventional C-means clustering which minimises a criterion such as 5.1 based on a Euclidean distance is *not* in general invariant under an affine transformation of the data since:

$$D_E = \sum_{J=1}^{C}\sum_{i\in J}(\underline{x}'(i) - \underline{\mu}'(J))^T(M^{-1})^TM^{-1}(\underline{x}'(i) - \underline{\mu}'(J)) \tag{5.28}$$

which is neither equal nor proportional to D_E' unless M is orthogonal or proportional to an orthogonal matrix. As Duda, Hart and Stork [46] state (section 10.6.1, page 539) a clustering "should be invariant to transformations natural to the problem" so a better choice for image segmentation problems might be to cluster by minimising a criterion like ([46] section 10.7.3, page 545):

$$D_W = det|S_W| \tag{5.29}$$

which transforms as

$$D_W' = (det|M|)^2D_W \tag{5.30}$$

and therefore may lead to clusterings or segmentations that are invariant ([46] problem 27, page 589). A clustering based on the between-class covariance and maximising $det|S_B|$ would be similarly invariant – as would clustering based on the ratios of such quantities mentioned in section 5.1 which are completely invariant under such affine transformations.

Alternatively, as also mentioned in section 5.1, clustering may be based on a criterion such as:

$$D_T = \sum_{J=1}^{C} \sum_{i \in J} (\underline{x}(i) - \underline{\mu}(J))^T S_T^{-1} (\underline{x}(i) - \underline{\mu}(J)) \tag{5.31}$$

based on a Mahalanobis distance. It is easy to show that D_T, and all similar criteria based on Mahalanobis distances involving for example, the within-class covariance S_W or the class-conditioned covariances $S(J)$, may produce clusterings with the required invariance to affine transformations. Duda, Hart and Stork (problem 30, page 590) show how an algorithm to iteratively minimise D_T might be developed, but even if this (or some other convenient iterative algorithm were used) there remains the problem of finding a way of initialising the process that has the required invariance. As this difficulty is common to all such clustering algorithms and since a random initialisation would merely bring us back to characterisation of the basins of attraction of the clustering algorithms, it precludes the possibility of using any of them as a useful, invariant way of initialising our extension of the Otsu algorithm.

5.4 Characterisation of data variability

The analysis in the previous sections of the invariance properties of our extension of Otsu's algorithm and its convergence indicates that it would be useful to have a way of characterising the variability of our thin-film slide image data and, in particular, of determining to what extent the variability may be represented by an affine transformation of the colour space. The obvious way to characterise the difference between two images $\underline{x}(i)$ and $\underline{x}^{'}(i)$ is via a mean squared error measure:

$$\epsilon^2 = \frac{1}{n} \sum_{i=1}^{n} |\underline{x}^{'}(i) - \underline{x}(i)|^2, \tag{5.32}$$

or, if an affine colour transformation 5.10 is included to align the images, via:

$$\epsilon^2(M, \underline{b}) = \min_{M, \underline{b}} \left\{ \frac{1}{n} \sum_{i=1}^{n} |\underline{x}^{'}(i) - M\underline{x}(i) - \underline{b}|^2 \right\}. \tag{5.33}$$

However, the number of RBCs and their locations vary from image to image, so the pixels i in images $\underline{x}(i)$ and $\underline{x}^{'}(i)$ do not necessarily correspond. One way forward would be to restrict the comparison to background regions in common between the images [175], but this doesn't use the data efficiently and, moreover, is dependent on a sufficiently good segmentation of both images being available. An alternative which removes dependence on the locations of the RBCs is to compare the colour distributions $p^{'}$ and p, say, of the images. Such a comparison still depends on the number of RBCs in each image, but this is a variability with which our algorithm has to cope.

In practice, the distributions $p^{'}$ and p will be estimated empirically from histogram data so the above amounts to histogram alignment. This is a topic that we mentioned briefly in our review, section 2.2.1 but it is worth considering a little more deeply in principle.

Statisticians have developed a variety of ways of testing for the equivalence of two distributions, the best known of which for the comparison of two empirical distributions is probably the Kolmogorov-Smirnov test [198]. This is usually applied to univariate data and depends on the cumulative probability distributions which, being integral measures, are not very sensitive to the data distributions. From our purposes, it seems better to compare the probability density functions $p^{'}$ and p directly. A squared-error, Euclidean metric is an obvious possibility:

$$\epsilon_{PF}^2 = \int d^d\underline{x}[p^{'}(\underline{x}) - p(\underline{x})]^2, \tag{5.34}$$

but the discussion of probabilistic distance and dependence measures in Devijver and Kittler (see [43] chapter 7, in particular sections 7.2 and 7.3) suggests a wide variety of other possibilities, including: Chernov, Bhattacharyya, divergence, Lissack-Fu, Kolmogorov and entropy or information measures (see also [118]). These and others may be found on various webpages, including Wikipedia articles. The squared error measure 5.34 appears in [43] associated with the name "Patrick-Fisher" – hence our notation above. We also note that it depends on the overlap or intersection of the distributions as used in a method for recognizing objects from their colour [173] and that 5.34 may be generalised to allow for a colour transformation by setting:

$$\epsilon_{PF}^2(M, \underline{b}) = \min_{M,\underline{b}} \left\{ \int d^d\underline{x}[p^{'}(\underline{x}^{'}) - p(\underline{x})]^2 \right\}. \tag{5.35}$$

Though 5.35 can be differentiated in order to develop equations defining the transformation which minimises 5.35, the resulting equations are non-linear and of little practical use as the required derivatives of $p^{'}(\underline{x}^{'})$ are not available when the distributions are only known empirically [21]. Moreover, the Patrick-Fisher probabilistic distance is not dimensionless and not bounded above, so there is no natural scale on which a particular value of ϵ_{PF}^2 or $\epsilon_{PF}^2(M, \underline{b})$ may be interpreted. What we might, following [43], call the Kolmogorov distances:

$$\epsilon_K = \int d^d\underline{x}|p^{'}(\underline{x}) - p(\underline{x})| \tag{5.36}$$

and

$$\epsilon_K(M, \underline{b}) = \min_{M,\underline{b}} \left\{ \int d^d\underline{x}|p^{'}(\underline{x}^{'}) - p(\underline{x})| \right\}, \tag{5.37}$$

are free of these defects. In particular we note that ϵ_K and $\epsilon_K(M, \underline{b})$ are dimensionless and that it follows from the triangle inequality that $0 < \epsilon_K < 2$ with the upper bound achieved when the distributions p and $p^{'}$ do not overlap, and similarly $0 < \epsilon_K(M, \underline{b}) < 2$ with the upper bound attained when the distributions $p(x)$ and $p^{'}(x^{'})$ of the transformed image do not overlap.

5.5 Conclusions

It the previous chapter we saw that our multi-dimensional extension of Otsu's algorithm seemed to be a useful tool in 3D for a colour-based segmentation of the RBCs in thin-film microscope slide images. However, at the close of the chapter we indicated that we had little theoretical understanding as to how the algorithm worked: how it converged, when it could be expected to converge, and to what kind of solutions, and why it seemed equally effective across the range of our image data in which the image colour and number of RBCs varied. In this chapter we have attempted to obtain a deeper understanding of the algorithm and to answer such questions. To this end we first:

(i) Briefly reviewed some standard clustering procedures, in particular those based on statistical, covariance measures. In particular, C-means (better known as k-means) clustering was discussed and it was shown that, when the feature space $x(i)$ is one-dimensional as, for example, when pixel intensity or data from a single colour channel is used to segment an image, C-means and the Otsu algorithm were based on optimisation of the same criterion – minimisation of the within-class variance σ_W^2 or equivalently maximisation of the between-class variance σ_B^2.

(ii) It was noted that common implementations of C-means proceed by iteratively changing class membership according to a greedy algorithm whilst the Otsu algorithm is based on an iterative procedure derived by direct differentiation of σ_B^2 with respect to the discriminating thresholds. In addition it was noted that C-means is usually initialised from a random clustering or segmentation whereas the Otsu algorithm, at least in the two-class case, has a deterministic initialisation. For RBC segmentation the algorithms should thus give similar, and in principle barring discretisation effects, identical results when initialised in the same way.

(iii) In higher dimensions d it was shown that clustering algorithms do not attempt to optimise a criterion equivalent to the various forms of the Fisher discriminant. In particular, the Fisher discriminant is based on matrix expressions of the form $A^T S A$ in which a $d \times d$ covariance matrix S is pre- and post-multiplied by a $d \times r$ rectangular matrix A which selects the r most discriminating feature directions in the data space $\underline{x}$. In one-dimension, A is nugatory hence the equivalence noted above, but this is no longer the case in higher dimensions as there is nothing equivalent to A in the clustering criteria.

(iv) In particular, when there are two-classes, the only case for which we have been able to

develop our multi-dimensional extension of Otsu's algorithm, there is a single most discriminating feature direction, $\underline{a} \propto S_W^{-1}(\underline{\mu}(2) - \underline{\mu}(1))$ so, it was noted, the most important differences between our algorithm and a 2-means algorithm arise from the inverse of the within-class covariance S_W. If S_W is highly anisotropic the discriminating surface between the two classes or clusters will not be perpendicular to the line joining the respective class means $\underline{\mu}(2) - \underline{\mu}(1)$. Analysis of the S_W produced in some of the RBC segmentations confirmed the anisotropy and thus that a colour-based segmentation via a 2-means algorithm would not be the same as that produced by our Otsu 3D algorithm. It was noted that although the direction of $\underline{\mu}(2) - \underline{\mu}(1)$ and its magnitude varied somewhat from image to image reflecting variation in colour and brightness, the principal axis of the within-class covariance always tended to be parallel to $\underline{\mu}(2) - \underline{\mu}(1)$ though again with some variation from image to image.

Issues concerning the convergence of the Otsu algorithms were addressed next, in particular by exploring the basins of attraction of the algorithms in the two-class case.

(v) In one-dimension the basin of attraction of the Otsu algorithm could be studied by an exhaustive search of all possible initialisations, essentially by evaluating the between-class variance σ_B^2 as a function of the threshold T. It was found that σ_B^2 and thus similarly the Fisher discriminant σ_B^2/σ_W^2 typically had a single maximum. The basin of attraction from which all initialisations of the algorithm converged to this maximum thus covered the complete range of intensities and the algorithm never failed to produce to a good segmentation of the RBC pixels.

(vi) In order to characterise the behaviour of our Otsu 3D algorithm, we took advantage of the fact that each iteration comprised two stages: application of the Otsu algorithm to the data projected onto the single dimension of the current estimate of the feature direction $\underline{a}$ followed by an update of $\underline{a}$. Given behaviour of the Otsu algorithm as in (v) above and the fact that the 3D algorithm commences by selecting an initial $\underline{a}$, our Otsu 3D algorithm was characterised by studying how different choices for the initial direction of the feature vector affected its behaviour. It was shown that a large fraction of the choices for the initial direction of $\underline{a}$ covering most of the unit sphere led to successful segmentation of the RBCs in a typical thin-film microscope slide image. Since the segmentation was successful when the direction of $\underline{a}$ was chosen deterministically from the first mode $\underline{p}(1)$ of a PCA analysis, evidently $\underline{p}(1)$ lay within this basin of attraction.

(vii) In order to help understand why the Otsu algorithms worked well on a variety of data,

in particular on images of noticeably different colours, we studied the invariance properties of the criteria, solutions and of the algorithms derived therefrom. In one dimension, the criteria, solutions and algorithm were all found to be invariant under affine transformations of the feature space. In higher dimensions, the criteria and generalised eigensolutions are invariant provided the feature matrix A itself transformed appropriately according to 5.13. It was shown explicitly that the discriminating feature direction $\underline{\hat{a}}$ transformed in the required manner. The iterative stages of our Otsu 3D algorithm were thus invariant but initialisation via the PCA analysis was found only to be invariant provided the transformation matrix M was proportional to an orthogonal matrix. This strongly constrains the form of the colour transformations for which our algorithm is completely invariant and unfortunately excludes the most common types for which M is diagonal but not isotropic. However, it was noted that successful segmentation may still be obtained under such transformations (and more general ones) if the PCA initialisation remains within the required basin of attraction of the algorithm.

Finally, with these properties of the algorithms in mind, a method was discussed for characterising the variability of the image data.

(viii) It was argued that a probabilistic distance measure should be used as such measures, unlike image alignment, do not depend on availability of a successful segmentation. It was argued that a Kolmogorov distance between the probability density functions (*pdfs*) of two datasets should be used for this purpose as it is a dimensionless number in the range $0 - 2$. It was shown how the Kolmogorov distance may be generalised to include not only the direct comparison of the distribution of one dataset with another but also to allow first for an affine transformation of one dataset in order better to align it with the other.

In summary, the analysis described in this chapter gives us confidence in the application of the original Otsu algorithm and our extended Otsu 3D algorithm to the segmentation of RBCs in images of blood, thin-film slides. It is thus appropriate to press on and, in the next chapter, to explore how these algorithms might further be used to count the total number of RBCs, to explore how they might be used to distinguish healthy and infected RBCs and thus ultimately, via the number of cells infected with parasites, to determine the parasitemia of a blood sample.

Chapter 6

Counting RBCs, Detecting Infections and Estimating Parasitemia

Having a system that can automatically detect RBCs in an image of a thin-film slide is only of mild interest and, unless possibly it were used as a tool to aid a human observer, on its own not of much practical significance. Of greater interest would be a system which could automatically detect RBCs, count them and determine whether the cells were infected or not. As noted in the review, chapter 2 such a system would be useful in laboratory work to relieve staff of a repetitive and time consuming task in determining the degree of infection or parasitemia of a sample. In the field, an automatic system could be used for similar purposes, or even to determine the type of malaria infection, though we recall that for detecting whether a patient is infected thick-film slides have greater sensitivity (section 2.3) and that determining the type of parasite infecting a RBC may also require recognizing the stage of its life-cycle (section 2.3.1) and [175]). In this chapter we explore to what extent the discrimination approach adopted in this thesis might be used to count RBCs and report some preliminary work on similarly determining whether RBCs are infected or healthy. We begin with the former.

6.1 Counting RBCs

Review of the literature (chapter 2, section 2.3.4) indicates that there are two fundamentally different ways of counting RBCs. One approach (*cf.* [165]) would be to make a statistical estimate of the number of RBCs from the number of RBC pixels which, following the results reported in chapter 4 we would take from the segmentation produced by either the 1D Otsu algorithms or by our extended Otsu 3D algorithm.

The alternative is to separate RBC objects and to count the number of cells they represent. In addition to distracting artefacts in the plasma the main complication in this case, if we avoid problems of over segmentation (*cf.* section 2.3.4), is that RBCs which are close to each other

or touching may present as a single object comprised of two or more cells. Previous authors have worked hard to overcome such problems with Tek, in particular [175], employing a battery of techniques including granulometry, a minimum area watershed transformation and a radon transform to obtain objects having the expected RBC radius. The advantage of counting individual RBCs is that, if they can be identified as healthy or infected, the counts may be used to determine the parasitemia. There are several reasons why a statistical, pixel-based approach cannot be extended to estimating the number of infected RBCs. First, our preliminary work to be discussed below (sections 6.2 and 6.3) indicates that attempts to segment parasite pixels are somewhat variable and of unknown quality. Secondly, there tends to be false positives in the background and also possibly many false negatives where stained pixels should have been detected inside infected RBCs. Finally, the proportion of such errors in comparison to the number of true-positives is unknown and variable from image to image, owing to *inter alia* the stage of the life cycle of the parasite infection. This destroys any hope of a applying a statistical correction to counts of infected pixels for estimating the number of infected RBCs. However, it leaves open the possibility that if infected RBCs can be identified the number of them might still be estimated statistically from the total number of pixels belonging to the corresponding image objects.

6.1.1 Further use of the Otsu algorithm

Although Tek's approach to RBC counting seems to work well within the context of the overall system he developed, it is not well-suited to our approach and seems overly complicated. As we have seen, our segmentation of RBC pixels is very good, and running a straightforward connected components algorithm on the segmentation, isolates many individual RBCs as shown in figure 6.1. However, it can be seen that some objects represent clusters of two or occasionally three RBCs. In figure 6.1 (c) we show a histogram of object sizes. This has a very simple peak structure which suggests that an Otsu algorithm might be used to separate and label the objects into single RBCs, doublets, triplets *etc.*. Unfortunately, the extent feature histogrammed in figure 6.1 isn't very suitable for this purpose. Not only are there too few doublet and triplet RBC clusters to produce easily discernible peaks but, as we shall see below in contrast to the object area A, there is no simple, ideal relationship between the extents of objects representing singlet, doublet and multiplet RBC clusters. However, given a suitable one-dimensional feature space the first thought would be to use the algorithm with $C - 1$ thresholds or the equivalent C-means algorithm but there are two complications.

First, we do not know how many RBCs are in the largest contiguous clusters – and there

could be artefacts which produce objects larger than the biggest cluster of adjoining RBCs. White blood cells caused problems of this kind in the work carried out at the University of Westminster as mentioned in section 2.3.1. Such artefacts would be expected to be lighter than the blood plasma and therefore should not be segmented amongst the putative RBC pixels in our system but irrespective of this the number of classes C remains unknown.

Secondly, the area of a pair of contiguous RBCs is expected to be approximately twice that of a single cell, that of a triplet three-times *etc.* so there ought to be some kind of constraint on the corresponding means $\mu(J)$ produced by such a segmentation. If the cluster comprised of objects with the smallest areas has mean $\mu(0)$ (this class corresponds to small artefacts since RBCs near the image boundary and only partially within the field of view are excluded by rejecting objects wholly within a border 81 pixels wide), that comprised of singlet RBCs has mean $\mu(1)$, *etc.* and that comprised of the largest objects has mean[1] $\mu(C-1)$ and if the relationship between the areas of singlet RBC objects, doublets, triplets *etc.* were ideal, such a constraint might be expressed as:

$$\mu(J) = J\mu(1) \text{ for } J = 2 \cdots C-1. \tag{6.1}$$

One could try to develop a version of the multi-threshold Otsu algorithm to accommodate such a constraint. For example, if we assume that the largest clusters are doublets the process of minimising:

$$\min_{T_1,T_2} \left\{ \sigma_W^2 + \lambda[2\mu(1) - \mu(2)] \right\} \tag{6.2}$$

is at first no more complicated than the derivation in appendix A especially as, with appropriate adjustment of the class labels J, we can use the fact that

$$\frac{\partial \sigma_W^2}{\partial T_J} = -\frac{\partial \sigma_B^2}{\partial T_J}. \tag{6.3}$$

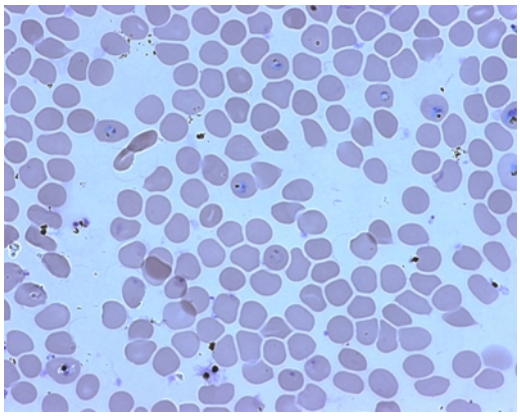
The result is:

$$\begin{aligned} &p(T_1)[\mu(1) - \mu(0)][2T_1 - (\mu(1) + \mu(0))] + 2\lambda \frac{p(T_1)}{Z(1)}[\mu(1) - T_1] = 0 \\ &p(T_2)[\mu(2) - \mu(1)][2T_2 - (\mu(2) + \mu(1))] \\ &\qquad + \lambda p(T_2) \left\{ \frac{2[T_2 - \mu(1)]}{Z(1)} - \frac{[\mu(2) - T_2]}{Z(2)} \right\} = 0 \end{aligned} \tag{6.4}$$

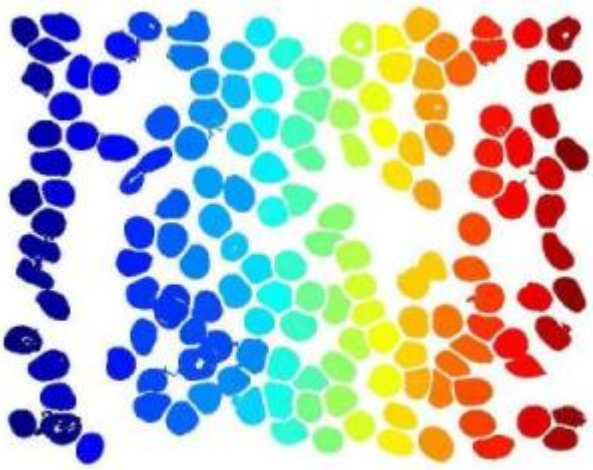
in which $Z(0), Z(1)$ and $Z(2)$ indicate the fractional class memberships like the variables introduced in 4.8 section 4.1.1. Equations 6.4 are to be solved simultaneously with the constraint

$$\mu(2) = 2\mu(1). \tag{6.5}$$

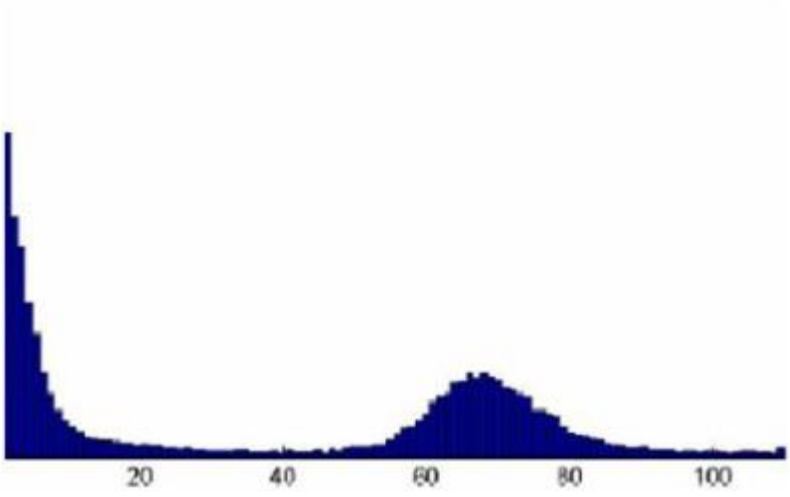
[1]The class labels are chosen here so that the class of objects representing a contiguous cluster of c RBCs has mean $\mu(c)$ rather than being labelled as in section 4.1.4 or in appendix A.



(a)



(b)



(c)

Figure 6.1: The thin-film slide image from figure 4.12 (a) with below (b) a colour-coding of the objects obtained as 4-connected components from the segmentation shown in figure 4.12 by means of the Otsu 3D algorithm. A histogram of the object sizes obtained as the extremal x and y extents of each object is shown bottom (c). Objects in the periphery of the image are excluded so the peak at the left represents small artefacts.

The Lagrange multiplier λ may be eliminated from 6.4, but the result is complicated, explicitly bilinear in T_1, T_2, and does not seem to lead to anything with the simplicity of an Otsu algorithm. This does not seem a promising approach.

Furthermore, it may be argued that applying constraint 6.1 in an exact manner is not appropriate. We might therefore favour imposing 6.5 as an approximate, soft constraint via a penalty term and (say):

$$\min_{T_1,T_2} \left\{ \sigma_W^2 + \lambda[2\mu(1) - \mu(2)]^2 \right\}. \tag{6.6}$$

However, even though the resulting equations are explicitly linear in T_1, T_2 and are easily solved for the thresholds in an 'Otsu-like' form, we do not advocate this approach as λ is then an additional parameter whose value has to be specified.

6.1.2 Recursive use of the Otsu algorithm with checks

A simpler approach is to use the Otsu algorithm recursively and to use relationship 6.1 to check whether the results are reasonable. We thus propose that:

(0) The area histogram is segmented into three classes by recursive application of the 1D Otsu algorithm with:

(i) a first application of the algorithm defining a threshold T_1 with small artefacts with area $A < T_1$ with mean $\mu(0)$ and larger objects with area $A > T_1$ with mean $\mu(A > T_1)$,

(ii) a second application of the Otsu algorithm to the histogram above T_1 defining a threshold T_2 with putative singlet RBC image objects with $T_2 > A > T_1$ with mean $\mu(1)$ and putative multiple RBC image objects with area $A > T_2$ with mean $\mu(2)$ that are expected to be mostly RBC doublets and some occasional RBC triplets and perhaps, very occasionally, larger clusters of four or more adjoining cells.

As usual, according to step (i) above:

$$T_1 = \frac{1}{2}[\mu(0) + \mu(A > T_1)] \text{ and } T_2 = \frac{1}{2}[\mu(1) + \mu(2)]. \tag{6.7}$$

Ideally:

$$\mu(0) \approxeq 0 \text{ and } \mu(2) \approxeq 2\mu(1) \tag{6.8}$$

with, since these objects are supposedly RBC doublets with occasional RBC triplets or larger clusters of adjoining cells, the expectation that:

$$\mu(2) \geq 2\mu(1). \tag{6.9}$$

However, if the thresholds are not correct we may have $\mu(2) < 2\mu(1)$ which is most likely to happen if T_2 is incorrect. Finally, we note that it follows from 6.7 and 6.8 that:

$$T_2 \approxeq 3T_1 \text{ and } \mu(1) \approxeq (T_1 + T_2)/2. \tag{6.10}$$

The image object area spectra have substantial gaps between the small artefacts, single RBC objects, and the clusters of adjoining RBCs where $p(A) = 0$ as illustrated in figure 6.2. The thresholds T_1 and T_2 are expected to lie in these gaps. Furthermore, in the ideal case where $\mu(0) = 0$ and $\mu(2) = 2\mu(1)$ and there are few doublet and no triplet or larger clusters of adjoining RBCs, $T_1 \approxeq 0.5\mu(1)$ and $T_2 \approxeq 1.5\mu(1)$. We thus further propose that an image object of area A is labelled as an $s = 1, 2, 3....$ singlet, doublet or triplet RBC *etc.* according to

$$s = round(A/\mu(1)). \tag{6.11}$$

The second relationship in 6.10 could also be used to label objects as singlet, doublet, triplet RBCs *etc.* according to $s = round(2A/(T_1 + T_2))$ but, given the way the area data is distributed, the mean $\mu(1)$ is likely to be a less sensitive estimate than the thresholds T_1 and especially T_2 so 6.11 is preferred.

The procedure described above is illustrated for each of the three images shown in figure 5.1. First, in figures 6.3 and 6.4 we show the result of a recursive application of the Otsu algorithm to histograms of the areas of image objects to reject small artefacts (class 0) and to label singlet (class 1) and doublet (class 2) RBCs.

Labelling of the image objects in the examples above seems to be quite good but there is scope to make the process much more rigorous and to check that it is reliable.

6.1.3 Reliability of RBC cluster labelling and counts

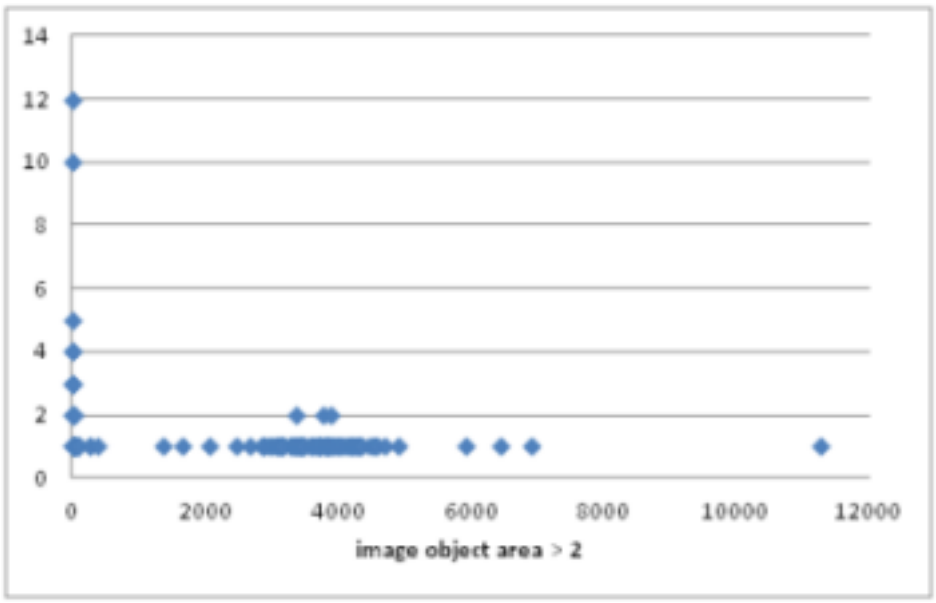
The discussion in the previous section provides a number of tests which may be carried out to check whether the procedures described above work properly in practice and are likely to produce accurate cell counts. First, we may test the following numerical relationships which follow from equations 6.7 – 6.10:

$$0 < t_1 = \mu(0)/\mu(1) \ll 1, \tag{6.12}$$

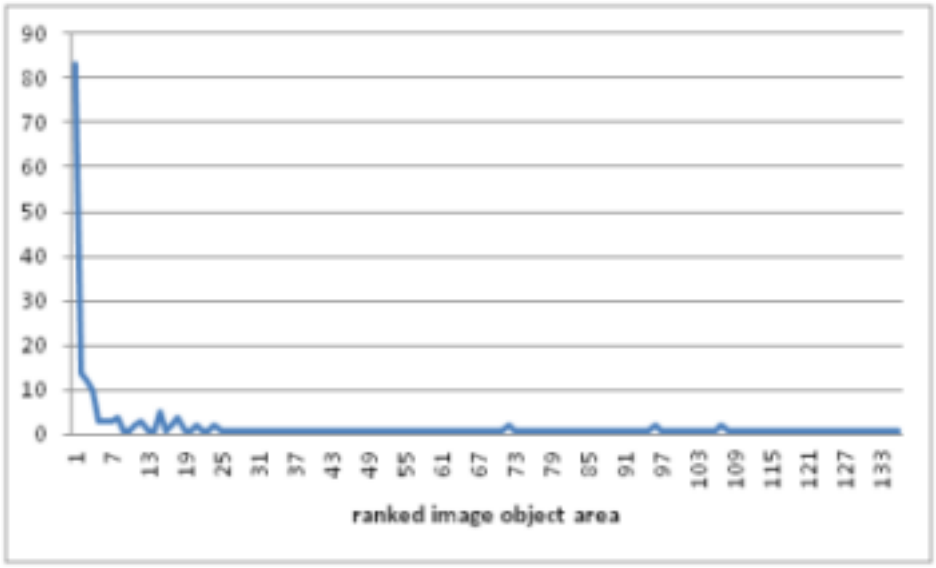
$$0 < t_2 = (\mu(2)/2\mu(1)) - 1 \ll 1, \tag{6.13}$$

$$0 < t_3 < |3T_1/T_2 - 1| \ll 1. \tag{6.14}$$

Ratio t_1 can never be negative and we expect it will almost always be small. It is therefore not a very strong test. However, if it were found that t_2 were negative and $|t_2|$ were *not* $\ll 1$ this would be a very strong indication that the procedure is not working properly.

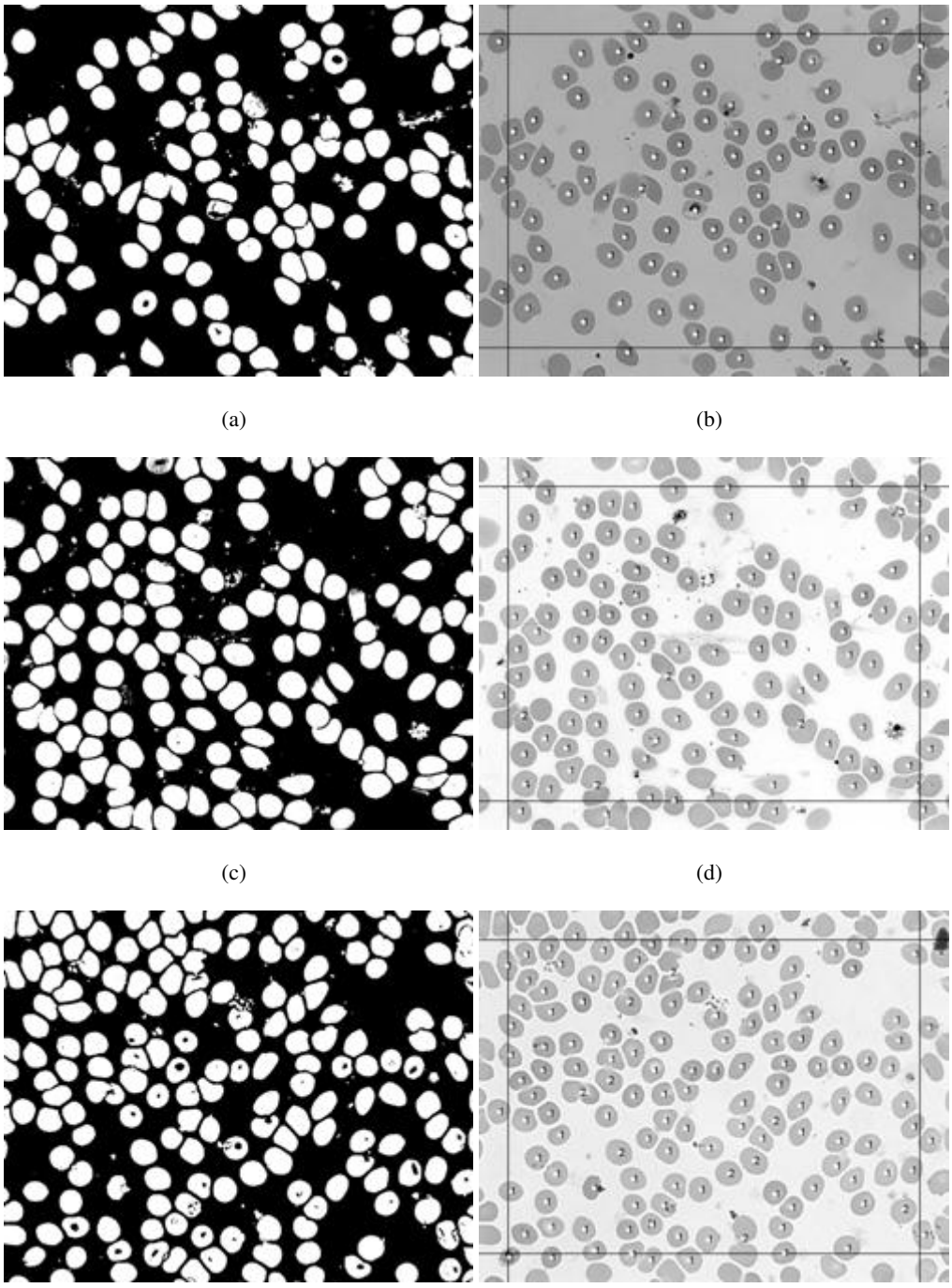


(a)



(b)

Figure 6.2: Histograms of the areas of the objects segmented in image number 1 of figure 5.1 as putative RBCs using the Otsu algorithm on the intensity. In (a) the histogram is depicted by showing as data points the heights of the histogram bins for objects of area $A > 2$ so as to allow the cluster of data points for objects representing singlet RBCs around $A = 3500$ to be visible together with the much smaller clusters for doublets and triplets around $A = 7000$ and above $A = 10500$ respectively. In (b) all the histogram bin values are plotted as a graph as a function of the rank of non-empty bins. In this representation, peaks for singlet, doublet and triplet clusters of RBCs are barely visible, but the peak near $A = 0$ for objects representing small artefacts is clear. Empty bins are ignored so gaps in the histogram where $p(A) = 0$ do not appear in this representation.

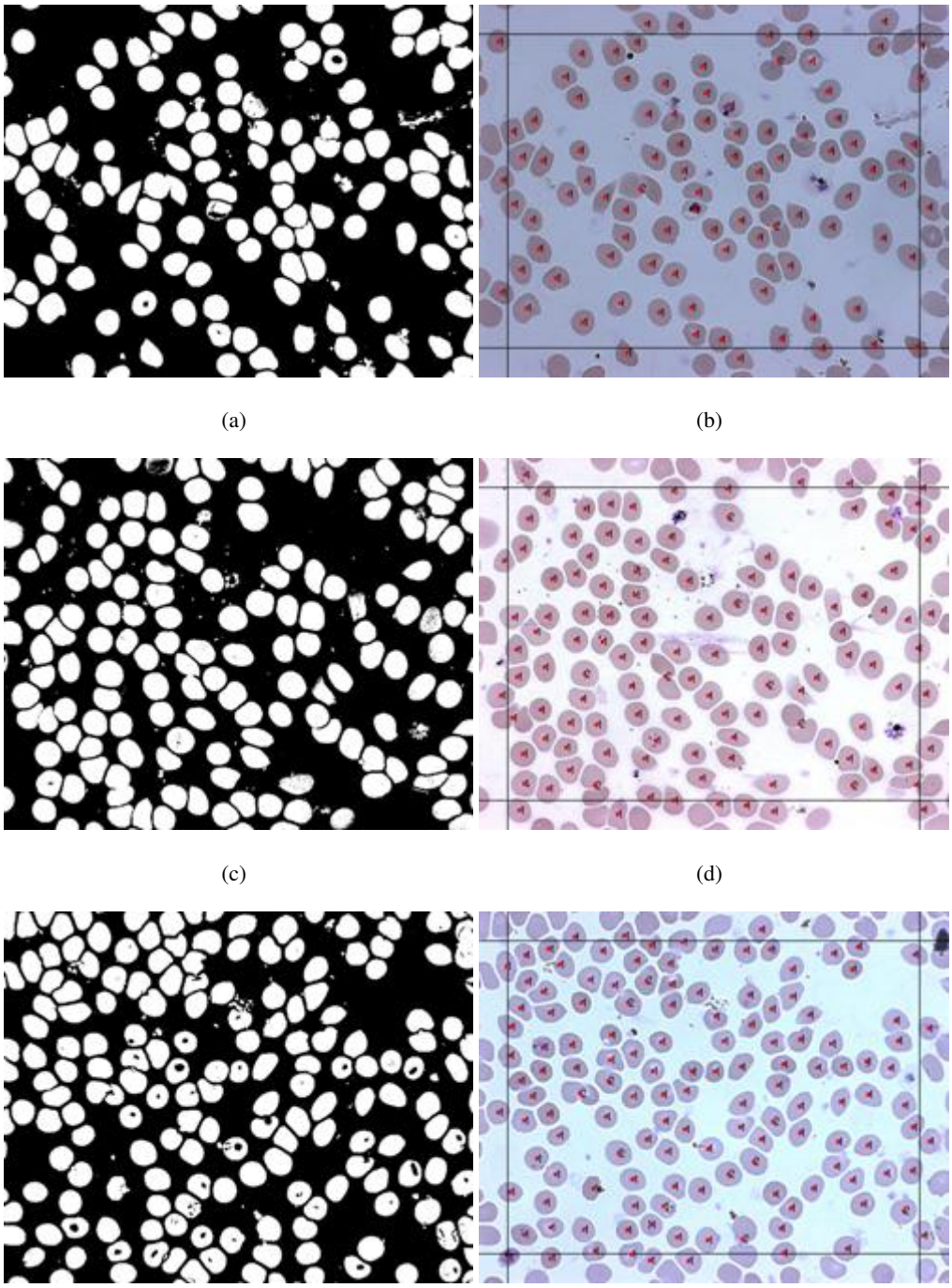


(a) (b)

(c) (d)

(e) (f)

Figure 6.3: Intensity-based segmentations of the three images from figure 5.1 showing for each image 1 (top), image 4 (middle) and image 45 (bottom): left – the segmentation of pixels putatively belonging to RBCs, and right – the corresponding 4-connected image objects labelled as singlet, or doublet RBC clusters by a second recursive application of an Otsu algorithm to the histogram of object areas as described in the text. Objects in the periphery of the image which could include only partially visible RBCs are excluded as shown by the borders images on the right.



(a) (b)

(c) (d)

(e) (f)

Figure 6.4: Left – colour-based segmentations of the three images obtained by means of our Otsu 3D algorithm and, right – the corresponding labelling of 4-connected objects as singlet or doublet RBC clusters as in figure 6.3.

Second, we can check that the labelling s of image objects as singlet RBCs, doublets, triplets *etc.* is consistent with the thresholds T_1 and T_2. Ideally:

$$s = 0 \,\forall \text{ objects with } A < T_1, \tag{6.15}$$

$$s = 1 \,\forall \text{ objects with } T_1 < A < T_2, \tag{6.16}$$

$$s \geq 2 \,\forall \text{ objects with } T_2 < A. \tag{6.17}$$

Since $\mu(0) \approxeq 0$ the first relationship is very unlikely not to be satisfied. The others however are more stringent tests, in particular as to whether the second threshold T_2 is erroneous. Thus, if O is the total number of putative RBC image objects present (*i.e.* the number with $A > T_1$) and we count the number O_{IL} of image objects for which *any* of 6.15 do *not* hold and thus may be said to be *inconsistently labelled*, we have the following additional labelling check:

$$t_L = O_{IL}/O \ll 1. \tag{6.18}$$

Finally, in the spirit of the MalariaCount system developed by Sio *et al* [165] we may also use the average area of a RBC $\mu(1)$ to estimate the number of RBCs in an image statistically according to:

$$Nstatistical = Npixels(A > T_1)/\mu(1) \tag{6.19}$$

where $Npixels(A > T_1)$ is the number of pixels belonging to objects with area $A > T_1$. This may then be compared with the number of RBCs obtained by counting the cell clusters

$$Ncount = \sum_{\substack{\text{image} \\ \text{objects}}} s, \tag{6.20}$$

and a test ratio:

$$t_N = 2(Ncount - Nstatistical)/(Ncount + Nstatistical) \tag{6.21}$$

formed[2]. As indicated above borders are used to exclude partially visible objects near the periphery of the image in all estimates of the number of RBCs.

To illustrate these tests we apply them to the interpretation of the three images shown in figure 5.1. Table 6.1 summarises the characteristics of these images interpreted via intensity and colour-based segmentation of putative RBC objects whilst the results of the above tests and statistical estimation of the numbers of RBCs are given in table 6.2. The labellings of the image objects used for the cell counts in the table were shown in figures 6.3 and 6.4 for the intensity-based and colour-based segmentations respectively.

[2]The test ratio is defined in this manner since neither $Nstatistical$ nor $Ncount$ is necessarily equal to the ground-truth number of RBCs, N (say). Note also that O is *not* necessarily equal to the actual (ground truth) number of RBCs N nor to either of the estimates $Nstatistical$ or $Ncount$.

	Image #1		Image #4		Image #45	
Image attributes used	Intensity	Colour	Intensity	Colour	Intensity	Colour
Segmentation algorithm	1D Otsu	Otsu 3D	1D Otsu	Otsu 3D	1D Otsu	Otsu 3D
Segmentation characteristics						
#pixels belonging to image objects	371633.0	372922.5	481871.0	463600.0	459166.0	456004.0
#(image objects)	276	255	584	668	280	262
μ	1346.5	1462.4	825.1	694.0	1639.9	1740.5
largest object area	11277.5	11307.5	11469.5	6993.5	8294.5	8194.5
Image object characteristics						
T_1	1922	1939	2070	1930	1712	1690
#pixels($A < T_1$)	5656.0	5551.0	7544.5	5293.5	3017.0	2645.5
#(image objects$\|A < T - 1$)	180	160	469	549	146	127
#pixels($A > T_1$)	365977.0	367371.5	474326.5	458306.5	456149.0	453358.5
#(image objects$\|A > T_1$)	96	95	115	119	134	135
$\mu(0)$	31.4	34.7	16.1	9.6	20.7	20.8
$\mu(A > T_1)$	3812.3	3843.8	4124.6	3851.3	3404.1	3358.2
T_2	5648	5468	5844	4070	3469	3430
#pixels($T_1 < A < T_2$)	335374.0	331084.0	411595.5	291904.5	239569.0	243323.0
#(image objects$\|T_1 < A < T_2$)	92	90	107	83	78	80
#pixels($A > T_2$)	30603.0	36287.5	62731.0	166402.0	216580.0	210035.5
#(image objects$\|A > T_2$)	4	5	8	36	56	55
$\mu(1)$	3645.4	3678.7	3846.7	3516.9	3071.4	3041.5
$\mu(2)$	7650.8	7257.5	7841.4	4622.3	3867.5	3818.8

Table 6.1: Recursive application of the Otsu algorithm to the 4-connected objects obtained from the three images: numbers 1, 4 and 45 shown in figure 5.1 interpreted via intensity and colour-based segmentation of putative RBC objects.

It can be seen from the first four rows of data in the table that the characteristics of the intensity and colour-based segmentations are similar for images 1 and 45 but somewhat different for image 4. Comparison of the two segmentation methods is not included in the tests developed above, but it is useful to note this difference for image 4 and that, as the values of the mean and maximum areas of the image objects indicate, the difference is *not* confined to spurious small objects. For example: close comparison of figures 6.3 (c) and 6.4 (c) reveals three image objects in the top, right corner of the latter which are connected as a single, large object in the former and two RBCs connected by a thin 'peninsula' of segmented pixels near the bottom, middle of the former which are separate in the latter. This may be seen more clearly in figure 6.5 which shows for image 4 at (a) the pixels in the intensity-based segmentation which are not included

in the colour-based segmentation and at (b), the opposite.

If desired, a measure of the similarity of the two segmentations could be calculated [84] as:

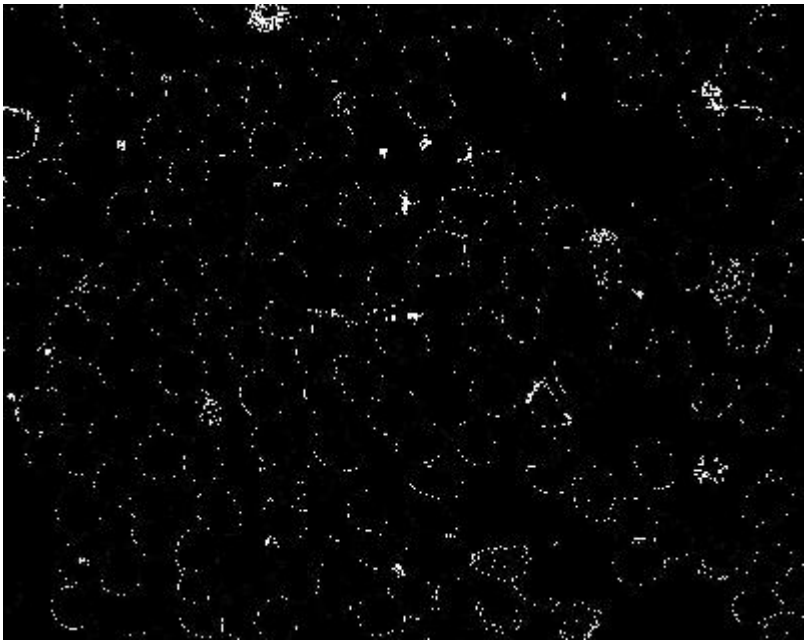
$$t_{IC} = \frac{2 \times |S_I \cup S_C - S_I \cap S_C|}{|S_I \cup S_C + S_I \cap S_C|} \tag{6.22}$$

where S_I is the intensity-based segmentation and S_C the colour-based segmentation. Even without computing t_{IC} we may note that according to the data in table 6.1 the corresponding fractional differences in the number of pixels in the intensity n_I and colour-based segmentations n_C are also quite different with $2(n_I - n_C)/(n_I + n_C) = 0.0386$ for image 4 but only -0.0035 and 0.0069 for images 1 and 45 respectively. This difference in the two segmentations of image 4 indicates that there seems to be something difficult about this image that affects much of the rest of the data in the table, in particular quantities dependent on the upper object area threshold T_2. The values inferred from the intensity and colour-based segmentations for this threshold differ far more for image 4 than for the other two: $2(T_{1I} - T_{1C})/(T_{1I} + T_{1C}) = 0.358$ for image 4 using a similar relative measure of comparison as against 0.032 and 0.011 for the other two images respectively.

	Image #1		Image #4		Image #45	
Image attributes used	Intensity	Colour	Intensity	Colour	Intensity	Colour
Segmentation algorithm	1D Otsu	Otsu 3D	1D Otsu	Otsu 3D	1D Otsu	Otsu 3D
Segmentation characteristics						
Numerical checks:						
$t_1 = \mu(0)/\mu(1)$	0.009	0.009	0.004	0.003	0.007	0.007
$t_2 = (\mu(2)/(2\mu(1)) - 1$	0.049	-0.014	0.019	-0.343	-0.370	-0.372
$t_3 = \|3 * (T_1/T_2) - 1\|$	0.021	0.064	0.063	0.423	0.481	0.478
Cell cluster labelling check						
# cells	101	101	124	124	138	138
# inconsistent object labels	0	0	0	31	53	53
t_L	0.000	0.000	0.000	0.250	0.384	0.384
# cells inferred from # pixels						
#pixels$(A > T_1)/\mu(A > T_1)$	96.0	95.6	115.0	119.0	134.0	135.0
#pixels$(A > T_1)/\mu(1)$	100.4	99.9	123.3	130.3	148.5	149.1
t_N	0.006	0.011	0.006	-0.050	-0.073	-0.077

Table 6.2: Data for the interpretation of the three images shown in figure 5.1 via intensity and colour-based segmentation of putative RBC objects shown in figures 6.3 and 6.4.

Turning now to the five test quantities t_1, t_2, t_3, t_L and t_N introduced above that are ap-



(a)

(b)

Figure 6.5: (a) Pixels included in the intensity-based segmentation of image 4 shown in figure 6.3 (c) but not in the colour-based segmentation shown in figure 6.4 (c). (b) The opposite, pixels included in the colour-based segmentation but not in the intensity-based segmentation. As the numbers in the third row of table 6.1 indicate, for this image the intensity-based segmentation is considerably more generous than the colour-based segmentation.

plied separately to each segmented image we note from table 6.2 that all of these are small for both segmentations of image 1 and that they are all similarly small for the intensity-based segmentation of image 4. The tests thus give no reason to doubt either interpretation of image 1, nor that of image 4 via the intensity-based segmentation. For image 45 and the colour-based interpretation of image 4, however, only t_1 is small with all four other test parameters taking absolute values approximately a factor of 10 or more larger than they did in the previous cases. It can only be concluded that these interpretations are not reliable and, since t_1 was acceptably small, that the source of error is likely to be associated with the determination of the threshold T_2.

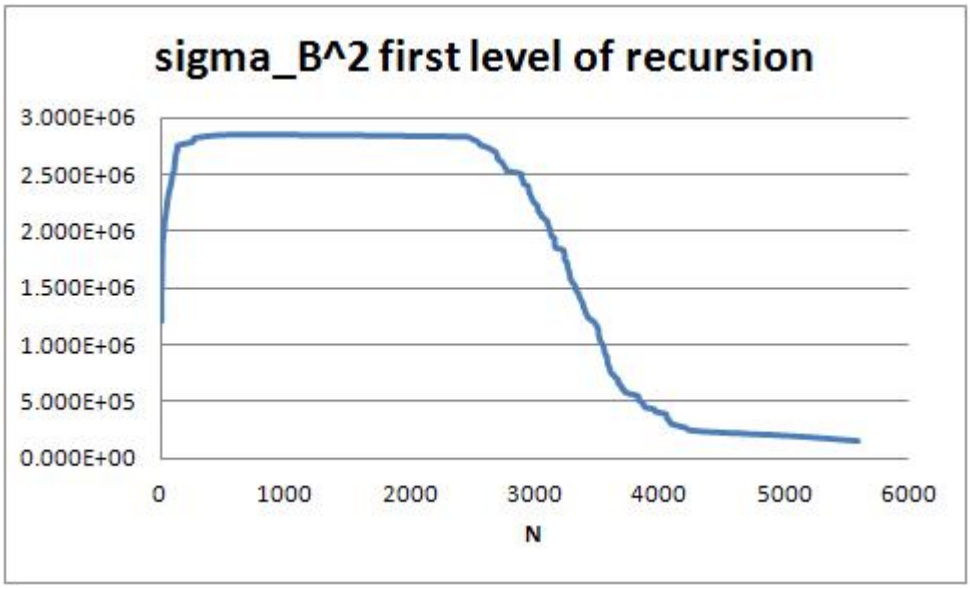
6.1.4 Analysis of between-class variances

Analysis of the behaviour of the between-class variances σ_B^2 as a function of the thresholds confirms the above. First we note that, as illustrated in figures 6.6 and 6.7 respectively for the intensity and colour-based segmentations of image 45, at the first level of recursion when the lower threshold T_1 is being determined in order to eliminate spurious, small objects from the segmentations, σ_B^2 has a broad plateau in the lower half of the range of image object areas A. The threshold T_1 determined by the Otsu algorithm is located within this plateau but to see this as a well-defined maximum of the between-class variance it is necessary to plot σ_B^2 as a function of the rank of the object areas A. This mapping is monotonic and does not create new extrema but distorts the ordinate so that it is clear where the maximum lies.

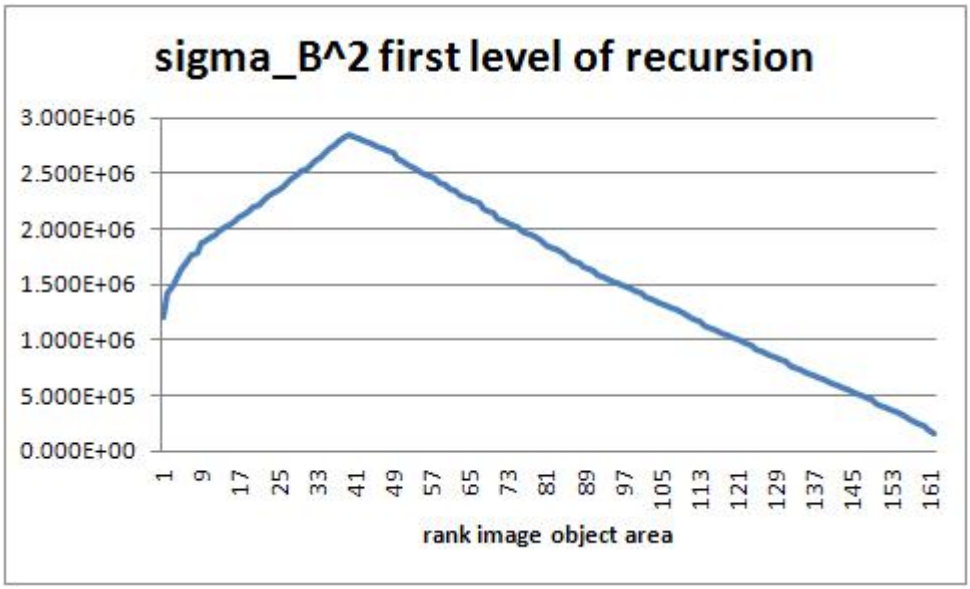
The behaviour of this between-class variance is similar for the other two images. Using the Otsu algorithm to determine T_1 thus works well and small, spurious objects may be successfully eliminated from the segmentations.

In section 6.1.2 the Otsu algorithm was then used again on remaining objects with areas $A > T_1$ to determine the second threshold T_2 and the means $\mu(1)$ and $\mu(2)$ of putative singlet and (mostly) doublet RBCs. In this case, the between-class variances for image 45 show multiple extrema which are rather more pronounced when the colour-based segmentation is used (figure 6.9) than when the intensity-based segmentation is used (figures 6.8) . For this image, the upper thresholds $T_2 = 3469$ and 3430 (Table 6.1) determined by the Otsu algorithms correspond to the local maxima of the between-class variances at low values of the thresholds – best (but even then, only just) discernible in figure 6.8 (b) and figure 6.9 (b) between histogram bin ranks $70 - 71$ and $77 - 78$ respectively.

It is clear that in this case the behaviour of the between-class variance is more complicated than it was at the first level when T_1 was being determined. Inspection of the numerical values

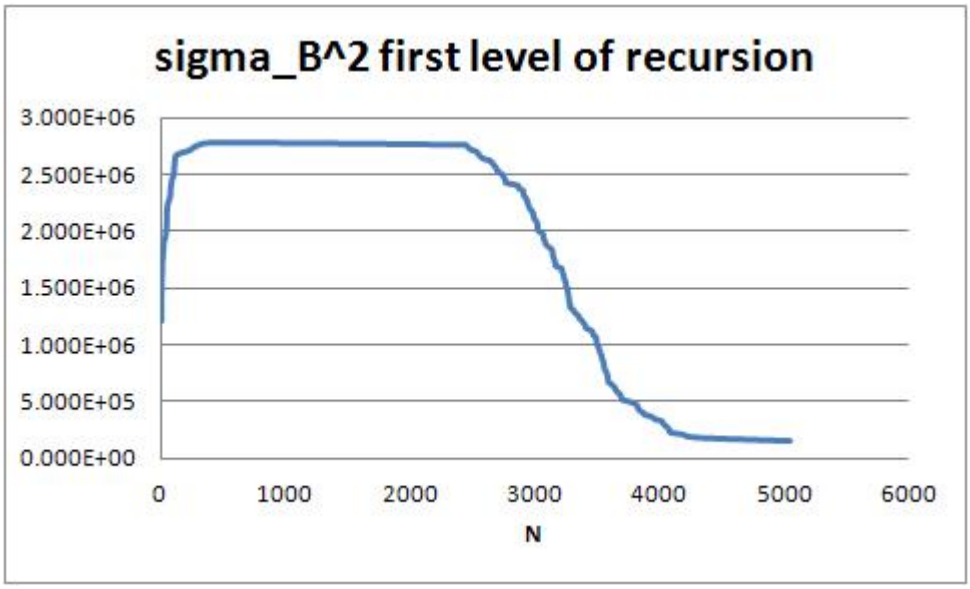


(a)

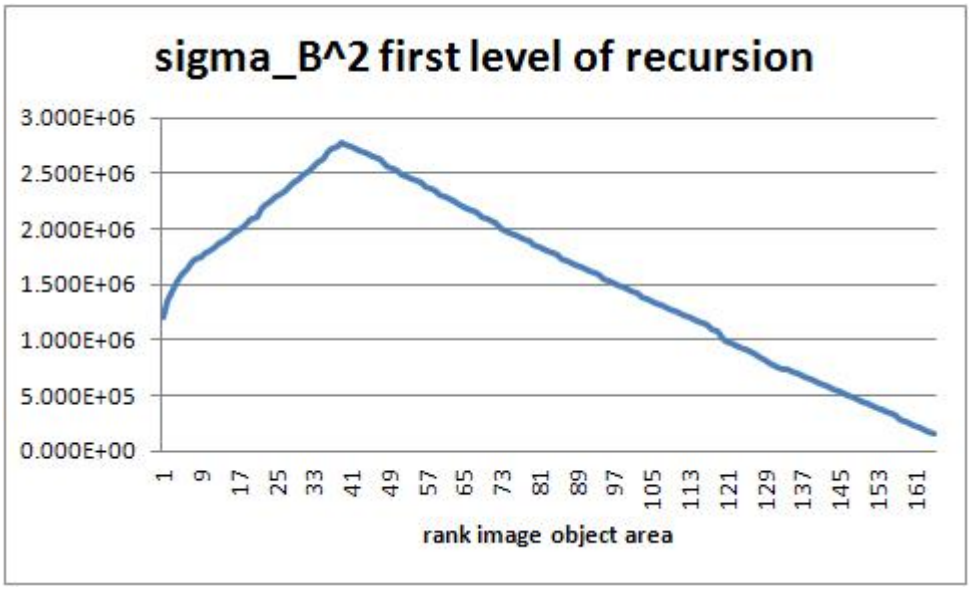


(b)

Figure 6.6: Between-class variances of the object areas at the first level for image 45 where the lower threshold T_1 is being determined using the intensity-based segmentation of the image. σ_B^2 is shown as a function of the threshold, top (a); and with the threshold mapped onto the rank of the image object area bins in the histograms, bottom (b).

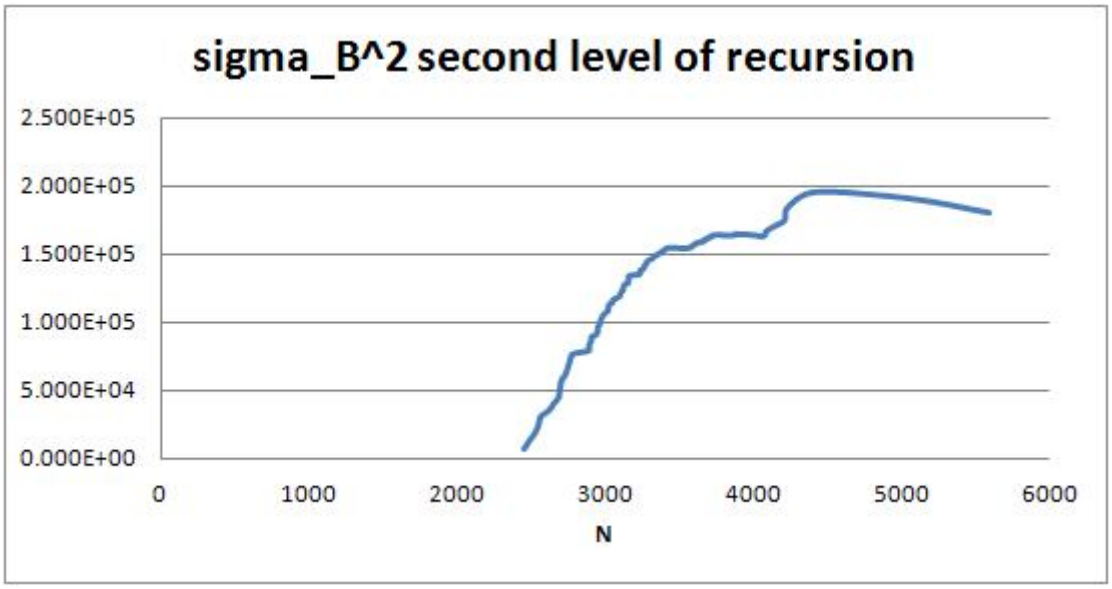


(a)

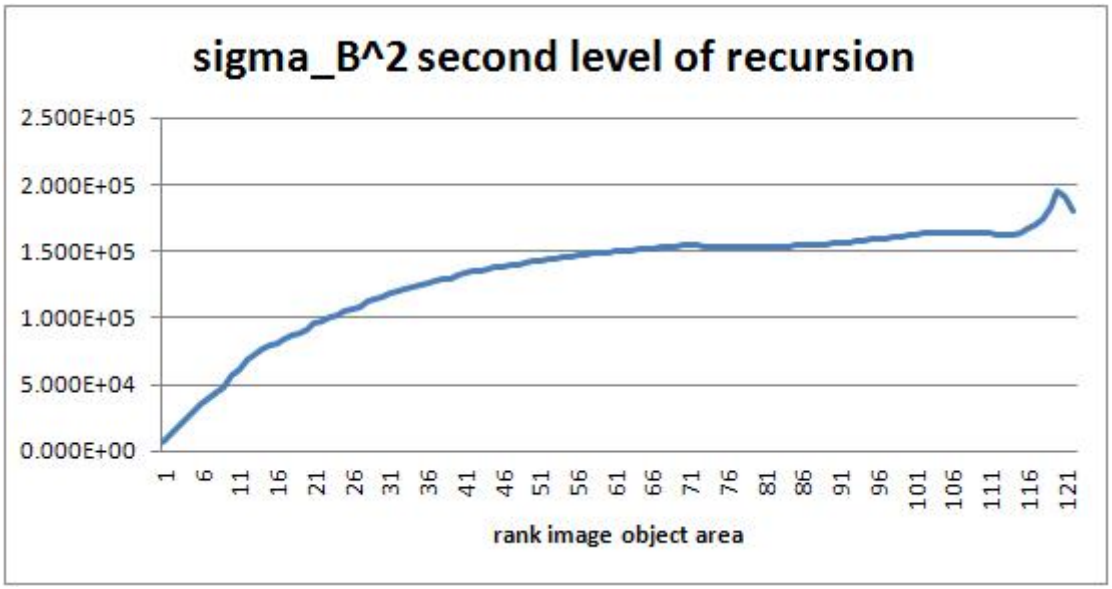


(b)

Figure 6.7: Between-class variances of the object areas at the first level for image 45 where the lower threshold T_1 is being determined using the colour-based segmentation of the image. σ_B^2 is shown as a function of the threshold, top (a); and with the threshold mapped onto the rank of the image object area bins in the histograms, bottom (b).

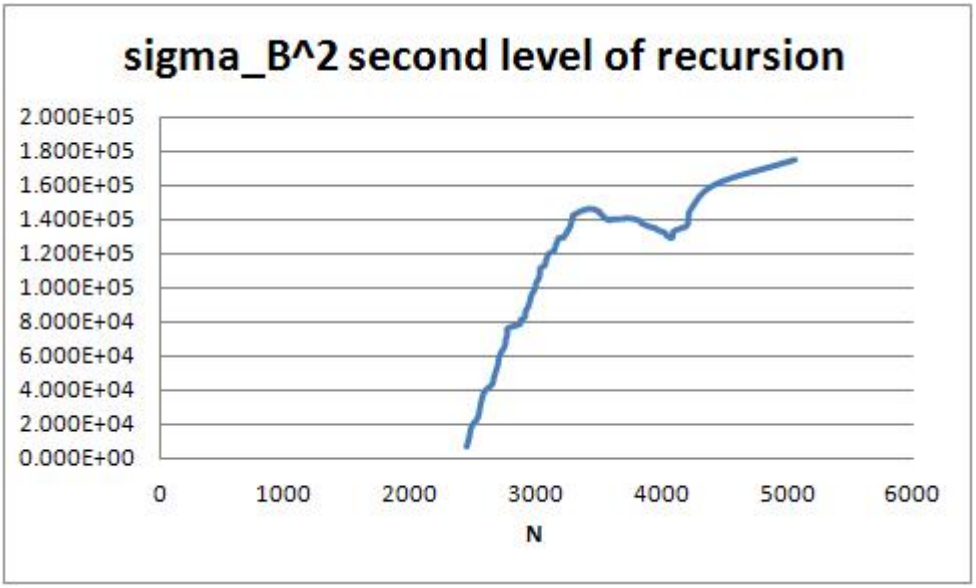


(a)

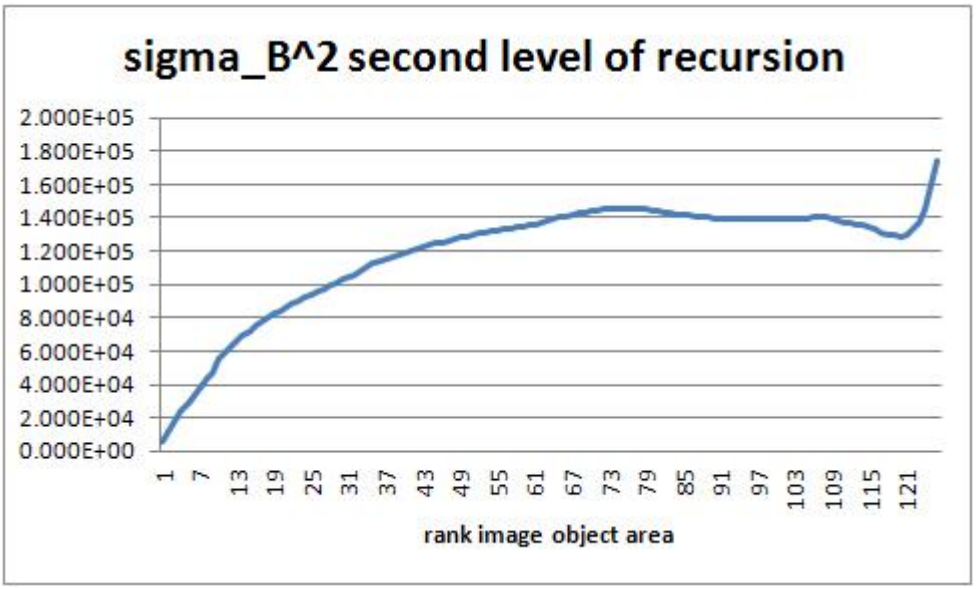


(b)

Figure 6.8: Between-class variances of the object areas at the second level for the intensity-based segmentation of image 45 after removal of spurious, small objects. σ_B^2 is shown, top (a), as a function of the threshold for $T_2 > T_1$ and, bottom (b), with the threshold mapped onto the rank of the image object areas for $A > T_1$.



(a)



(b)

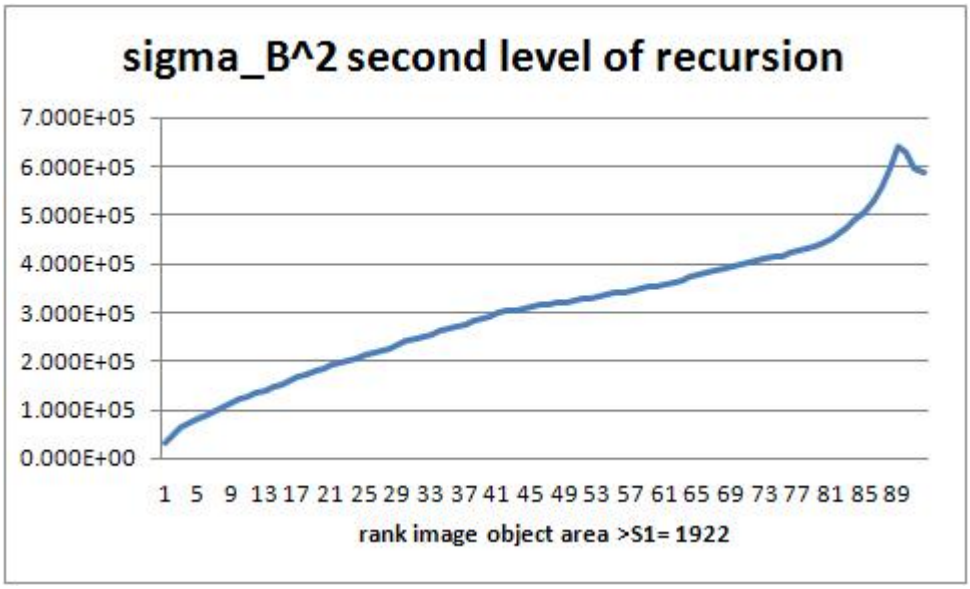
Figure 6.9: Between-class variances of the object areas at the second level for image 45 after removal of spurious, small objects. The second, upper threshold T_2 is being determined for objects segmented using the colour-based segmentation of the image. σ_B^2 is shown, top (a), as a function of the threshold for $T_2 > T_1$ and, bottom (b), with the threshold mapped onto the rank of the image object areas $A > T_1$.

reveals that in both the intensity and colour-based interpretations of image 45 at this second level of recursion σ_B^2 has at least five extrema as summarised in table 6.3. In both cases these extrema are alternately maxima and minima with, in figure 6.9 (b) the final and highest maximum at the end of the range. The three intermediate extrema are weak and, indeed, there may be other pairs of very weak extrema which we ignore in regions where σ_B^2 is flattish – hence we said "at least five extrema" above. In both cases the Otsu algorithm produces a threshold (table 6.1) corresponding to (in fact probably owing to the sparseness of the data just below) the first maximum at the lowest value of A in table 6.3. However, it can be seen from table 6.3 that the highest maximum of the between-class variances occurs at the extrema at the largest values of the object area A given. Detailed inspection of the object area histograms reveals that, again because of the sparseness of the data, these correspond to the areas of almost the largest image objects found in the segmentations (see table 6.1 for comparison). In contrast, for image 1 where both the intensity and colour-based interpretations appear to be successful, the between-class variances at the second level are much simpler with only a single maximum in each case near the largest values of the object areas A (figure 6.10).

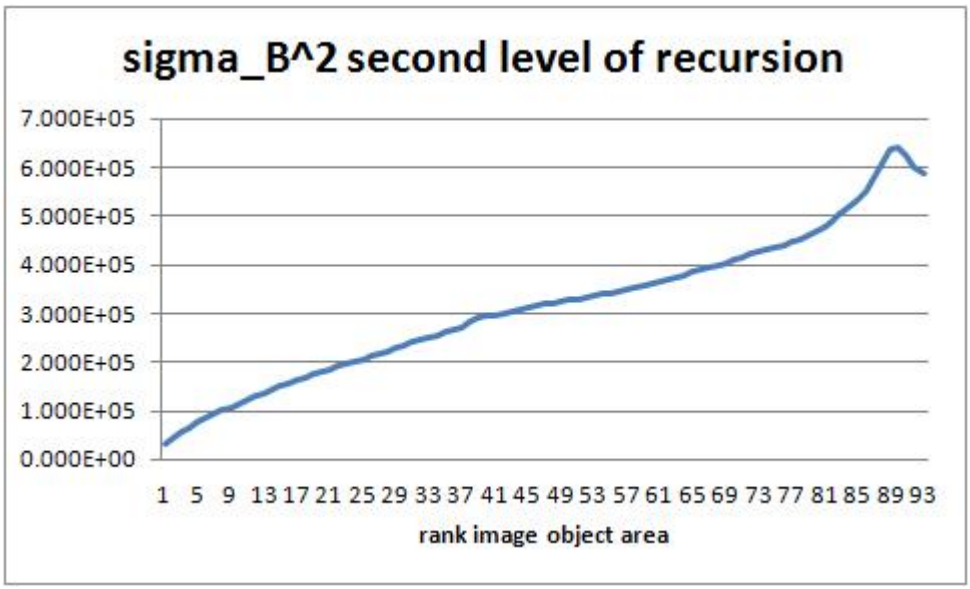
Intensity-based			Colour-based		
rank(A)	A	$\sigma_B^2 \times 10^{-5}$	rank(A)	A	$\sigma_B^2 \times 10^{-5}$
71	3482	1.540	78	3470	1.459
76	3513	1.535	95	3586	1.392
103	3770	1.635	107	3702	1.404
114	4060	1.622	120	4074	1.288
120	4424	1.950	126	5051	1.745

Table 6.3: Extrema of the between-class variances of image object areas at the second level of recursion for the intensity and colour-based interpretations of image 45.

As the tests summarised in table 6.2 suggest, the colour-based interpretation of image 4 generates a between-class covariance which has multiple extrema similar to those found for image 45. Placing the boundary between the RBC singlet and multiplet clusters near the largest maximum of the between-class variances for the colour-based interpretation of image 4 and both interpretations of image 45 leads to the results summarised in table 6.4 with the first three test parameters, t_1, t_2, t_3 small and cell counts and number estimates consistent to within approximately 1%. Admittedly the values of t_2 and t_3 for the colour-based segmentation of image 4 are larger than the largest of those in the first three columns of table 6.2 by approximately a factor of two, but it can be seen by comparison with table 6.2 that the cell counts and number



(a)



(b)

Figure 6.10: Similar results to those shown in figures 6.8 and 6.9 for image 1 when the behaviour of the between-class variances is simpler with only a single peak when either: (a) the intensity-based segmentation of RBCs is used, or (b) the colour-based segmentation is used.

estimates are the same as those obtained from the intensity-based interpretation.

	Image #4	Image #45	
Image attributes used	Colour	Intensity	Colour
Segmentation algorithm	Otsu 3D	1D Otsu	Otsu 3D
Image object characteristics	T_2 from $max(\sigma_B^2)$		
T_2	5005.2	4829.6	4965.8
#pixels($T_1 < N < T_2$)	425696.0	437183.5	440113.5
#(image objects$\|T_1 < N < T_2$)	114	131	133
#pixels($N > T_2$)	37731	18965.5	13245.0
#(image objects$\|N > T_2$)	6	3	2
$\mu(1)$	3721.9	3337.3	3309.1
$\mu(2)$	6288.5	6321.8	6622.5
Numerical checks			
$t_1 = \mu(0)/\mu(1)$	0.003	0.006	0.006
$t_2 = (\mu(2)/(2\mu(1)) - 1$	-0.155	-0.053	0.001
$t_3 = \|3 * (T_1/T_2) - 1\|$	0.157	0.063	0.021
Cell cluster labelling check			
# cells	124	137	137
# inconsistent object labels	1	0	0
t_L	0.008	0.000	0.000
# cells inferred from #pixels			
#pixels($N > T_1$)/$\mu(1)$	123.1	136.7	137.0
t_N	0.007	0.002	0.000

Table 6.4: Interpretation of images 4 and 45 at the second level when the between-class variances have multiple extrema and the thresholds T_2 are set to correspond to the largest maxima of σ_B^2.

6.1.5 Accuracy

In the previous section, two parameters t_L and t_N were computed to test the consistency of the cell counting process and the extent to which it agreed with a simple statistical estimate of the number of RBCs. These are *internal* characterisations of the algorithm's performance (*cf.* [195]) unlike characterisation of the algorithm's accuracy which requires the *external* input of the actual number of RBCs or *ground truth*, N, say. The latter is usually obtained manually,

if possible via the intervention of a human expert who in this case would ideally be a medical researcher. Unfortunately, as we remarked in chapter 2, medical researchers' time is at a premium and best not spent on the tedious task of counting hundreds of RBCs in many samples. Indeed, the main aim of this research was to develop a system capable of performing tedious cell counting tasks automatically and thereby saving much time in obtaining the parasitemia of a sample. In counting the number of RBCs visible in an image, one has to ignore cells that are partially visible near the periphery of the image, distinguish RBCs from artefacts in the plasma, and decide how many cells may be in a cluster of adjoining RBCs. These tasks are fairly straightforward though the last can sometimes be a little tricky when cells overlap or are rather misshapen. Even so, and in particular with the help of the software tools described in chapter 3, section 3.2.1 counting the number of RBCs in a sample is a relatively straightforward task that does not require the full expertise of a medical researcher. To illustrate this, counting the RBCs was therefore carried out on the three images shown in figure 5.1 by the author as 'developer' after a period of training and supervision by medical collaborators. Such ground-truth counts are given in table 6.5 where they are compared with the *reliable* algorithm counts taken from tables 6.2 and 6.4 – *i.e.* from the first three columns of the former and the three columns of the latter.

	Image #1		Image #4		Image #45	
Image attributes used	Intensity	Colour	Intensity	Colour	Intensity	Colour
Segmentation algorithm	1D Otsu	Otsu 3D	1D Otsu	Otsu 3D	1D Otsu	Otsu 3D
user count # cells	100		120		137	
# cells	101	101	124	124	137	137
#pixels$(A > T_1)/\mu(1)$	100.4	99.9	123.3	123.1	136.7	137.0

Table 6.5: Reliable cell counts for images 1, 4 and 45 compared to counts made by the author as described in the text. In this table whenever there were multiple maxima in the between-class variance for setting the threshold T_2 for classification of image objects according to their areas, T_2 was set to correspond to the largest maxima of σ_B^2.

It can be seen that, with the exception of image 4 where there is an over-estimation of 4 cells representing a 3% error, these 'reliable' cell counts are not only internally consistent, but also accurate to within 1%. Given that the counts are in error by a few $\%$ only for image 4 it is useful to recall that for this image there were difficulties in producing intensity and colour-based segmentations of pixels belonging to RBCs consistent with each other (*cf.* remarks in section 6.1.3) as well as in producing reliable cell counts (section 6.1.4). Finally, it may be noted

that the reliable RBC counts in the last three columns of table 6.5 for image 45 and the colour-based interpretation of image 4 do not differ much (by less than 1%) from the initial counts (# cells) in the corresponding last three columns of table 6.2 but that the statistical estimates of the number of cells (#pixels$(A > T_1)/\mu(1)$) do differ significantly (by up to 9% for image 45). Both types of cell numbers for the first three columns are of course identical.

6.2 Parasite detection

Determining whether RBCs are infected or healthy requires at least a three-way classification of the pixels in a thin-film microscope slide image – whether they correspond to the background (plasma or possibly an artefact or WBC); to a healthy part of a RBC; or to a stained parasite. The only simplifications are that, for determination of the degree of infection or parasitemia, only parasites within RBCs are of interest (*i.e.* before the cell is disrupted, section 1.2) and that there is almost never more than one parasite within an infected cell. Classification of pixels as belonging to a healthy part of a RBC or to a parasite infection may thus be made as a binary decision on each of the pixels which have already been segmented from the background in a hierarchical architecture (section 4.2.6) as an alternative to attempting the multi-class discrimination directly which is not likely to be a viable approach.

Typically, although some $40 - 50\%$ of the pixels in an image may belong to RBCs only a few $\%$ of the RBCs may be infected and, unless almost mature, a parasite may only occupy a small fraction of the area of a RBC. The number of pixels in an image belonging to stained parasite infections may thus be very small compared to the $1.3M$ pixels in an image – say $1/1000th$ given the remarks above which would make the infected pixels a very small minority class indeed. Detecting such pixels and subsequently determining whether they lie inside a RBC and that the cell is infected and reliably counting the small number of such cells are thus all quite difficult tasks. Only preliminary work has been carried out and satisfactory solutions that would, for example, enable an accurate estimate of the parasitemia to be made have yet to be found. Nevertheless, in the remainder of this chapter we describe this preliminary work as it may provide a useful guide to further research.

6.2.1 A structural approach?

In addition to their propensity to favour balanced classes unsupervised, multi-class algorithms are very general. They therefore make little or no use of any regularities or structure in the problem. In image processing and computer vision, the aim is often to utilise spatial structure but we have seen that, in particular for the parasite infections, it can be quite difficult to encapsulate such structure. However, when the feature space is one-dimensional and the input is a

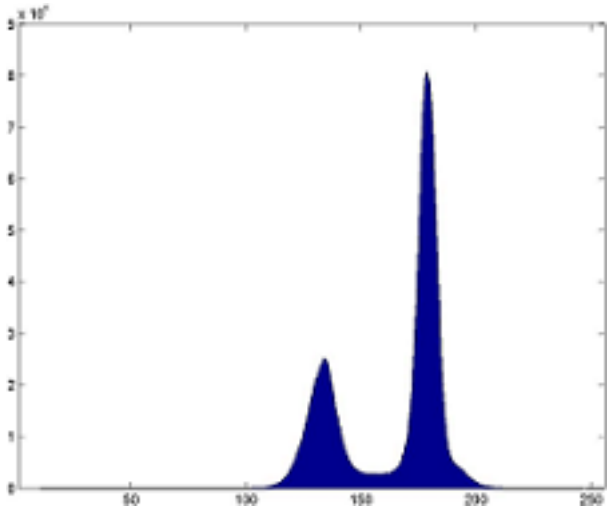
histogram, some structure of the histogram may be characterised and utilised in a quite general manner. For example, inspection of the intensity histograms indicates that pixels corresponding to stained parasite infections are deep in the lower tail of the histogram.

Consider image number 1 shown in figure 5.1 (a). The Otsu algorithm applied to the intensity histogram of this image shown in figure 6.11 (a) may be used to segment pixels belonging to RBCs (b) with a threshold $T = 155$. Background pixels above this threshold may be removed and the peak in the remainder of the histogram easily found automatically – in this case at $I_P = 134$. If we then reflect the threshold T about I_P and retain only the tail of the histogram below $I_P - (T - I_P) = 113$ we are left (figure 6.11 (c)) with a cluster which appears (see figure 6.12 (a)) to include pixels that may belong to parasite infections but also some others. A further application of the Otsu algorithm to this histogram produces a threshold of $T = 73$ and leaves only pixels very deep in the tail of the histogram that seem, in addition to some artefacts in the background plasma, to be predominantly stained parasite pixels as shown in figure 6.12 (b) and (c).

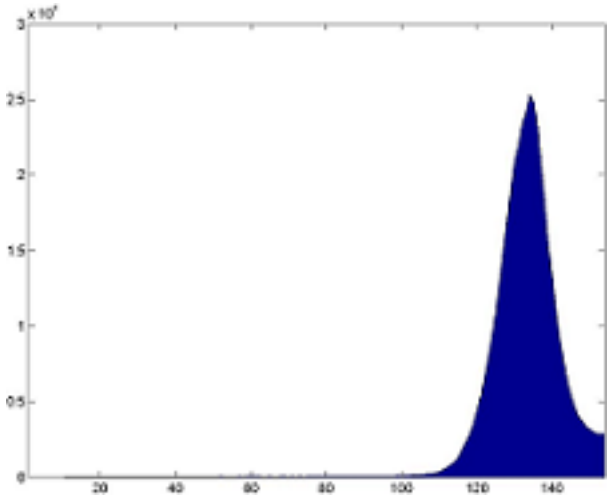
This example shows that there are approximately 905000 pixels in the background plasma and ~ 433800 pixels in the healthy parts of the RBCs but only ~ 7200 pixels corresponding to putative stained parasite pixels. Some of these appear to be artefacts but the numbers[3] nevertheless confirm and emphasize the smallness of the parasite pixel class and the difficulty of segmenting it.

It would also seem from the above that such an approach might work more generally since to determine the parasitemia it is only necessary to detect malaria infections within RBCs and thence to decide whether a cell is infected or not. A crude segmentation of parasite pixels within a RBC might thus suffice, rather than the detailed segmentation that would be required for determining the size, shape and other characteristics of a parasite in order to classify the type of infection and stage of its life cycle [175, 176]. Furthermore, since the Giemsa stain of the parasite is dark blue, one could also envisage processing individual colour channels in a similar manner and possibly combining information from two or more colour channels. However, the approach is rather *ad hoc* – there is no reason to expect the histogram to have the symmetry proposed – and thus potentially fickle and will only be used for comparison with other methods to be discussed below based on recursive application of Otsu algorithms.

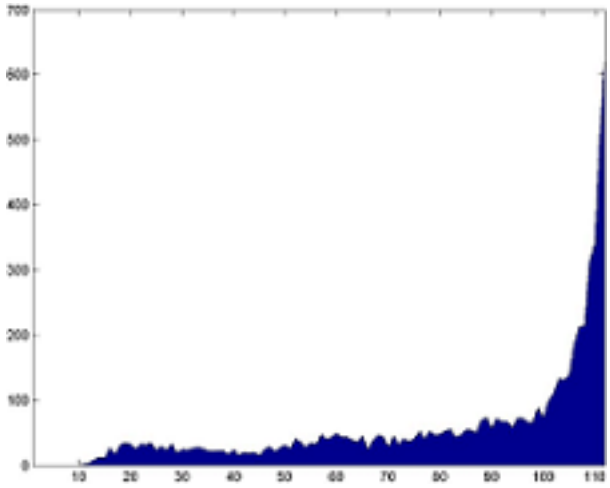
[3]The number of stained pixels detected is $\sim 0.5\%$ of the image pixels and $\sim 1.7\%$ of putative RBC pixels.



(a)

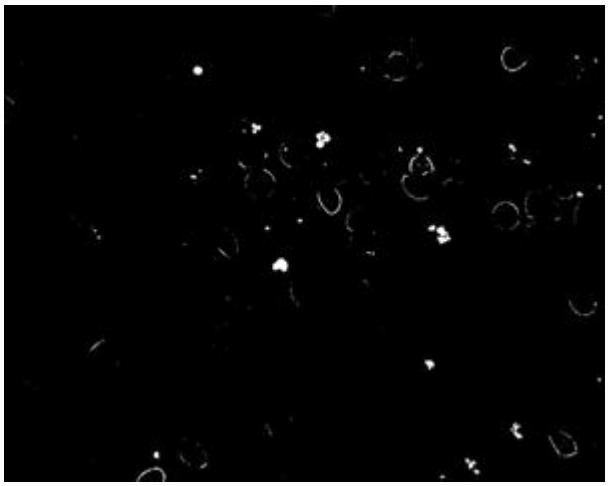


(b)

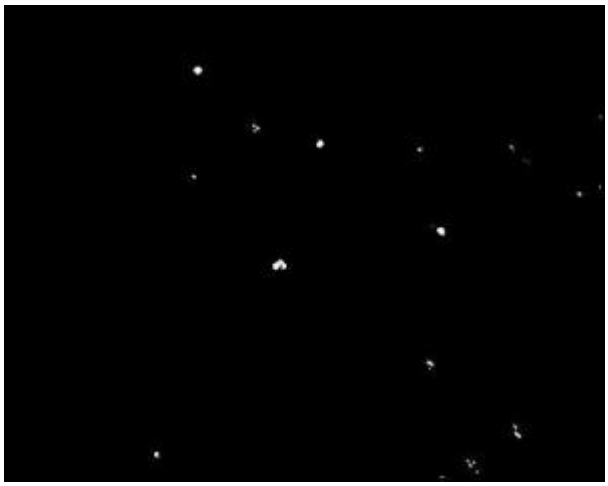


(c)

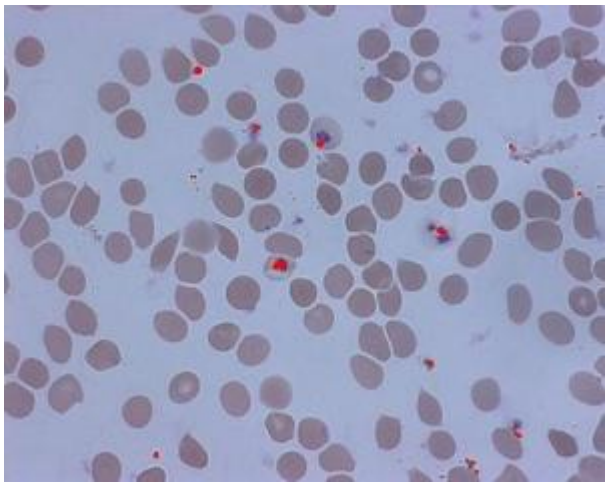
Figure 6.11: Intensity histogram (a) of image number 1 shown in figure 5.1 (a) for which the intensity-based RBC pixel segmentation was shown in figure 6.3. The peak (b) of the lower part of the histogram at $I = 134$ and (c) the tail at values of $I < 113$ selected from structural considerations as described in the text.



(a)



(b)



(c)

Figure 6.12: Results of the 'structural segmentation': (a) pixels segmented from the tail of the histogram selected in figure 6.11(c); (b) pixels remaining with intensity below $I = 73$ after a further application of the Otsu algorithm; (c) as in (b) but superimposed on the original image shown in 5.1.

6.3 Recursive application of the Otsu algorithms

For the purpose of determining the parasitemia only parasite infections within RBCs are of interest. Recursive application of the Otsu algorithms in a hierarchical, multi-classifier architecture would thus seem an approach worth exploring – and one that would help overcome the difficulties caused by the fact that parasite pixels usually belong to a very small, minority cluster. The three images shown in figure 5.1 are used to illustrate the approach.

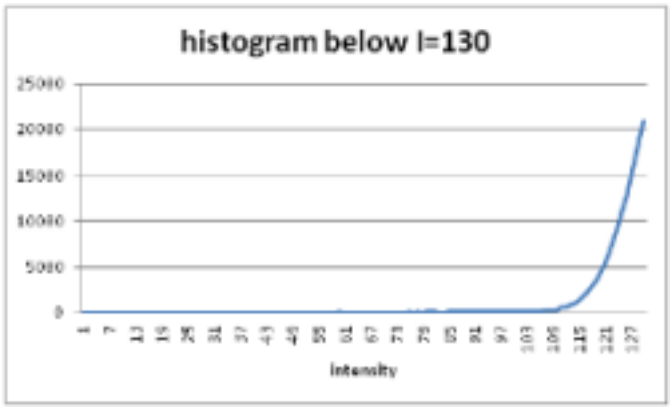
6.3.1 Recursive application of the 1D Otsu algorithm

We consider first recursive application of the Otsu algorithm to the image intensity histogram as illustrated in figure 6.13. The first Otsu threshold $T = 155$ is of course just what we have seen previously for this image corresponding to a value in the dip between the background and RBC peaks, but the second at 130 is not far from the peak in the RBC pixels at $I_P = 134$ identified previously in section 6.2.1 whilst the final threshold at 92 is somewhat higher than the final threshold of 73 determined structurally.

A point to note from figure 6.13 is that the Otsu algorithm partitions the histogram into two classes even when the histogram is mono-modal. This is because, regarded as a function of the threshold T, the between-class variance σ_B^2 has at least one peak at some intermediate value of T as shown in figure 6.14. In fact a little algebra shows this is true even when the histogram is constant and there is no structure whatsoever to suggest a segmentation into two or more classes and one may also recall that the C-means algorithm similarly always partitions a dataset in C clusters whatever its structure may be. In some respects the C-means algorithm should therefore perhaps better be regarded as providing C representatives of a dataset as in vector quantisation [115, 108].

The main point, however, is that the between-class variance σ_B^2 is bimodal at the second level of recursive application of the Otsu algorithm. It is not obvious what characteristics of the histogram shown in figure 6.13 (a) produce this bimodality, but we note that at the third level of recursion σ_B^2 is again monomodal though with a broad, more structured peak than at the first level. We have previously seen bimodality – and indeed multimodality – of the between-class variance in section 6.1.4 but in the context of classifying image objects according to their area when there were gaps in the feature histograms. In contrast, there are no gaps in the intensity histograms here. Results for the three images shown in figure 5.1 are shown in figures 6.15, 6.16 and 6.17. The final segmentation (c) for the first image, number 1, is quite similar to the structural segmentation shown in figure 6.12 (b) but, owing to the higher threshold, rather more generous with a few more pixels selected.

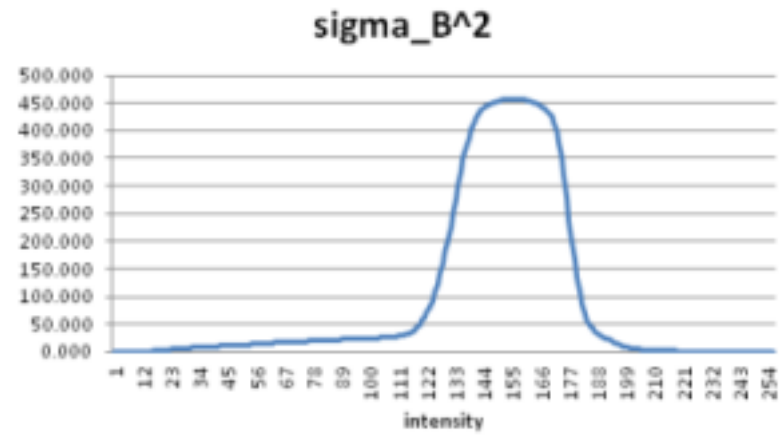
(a)



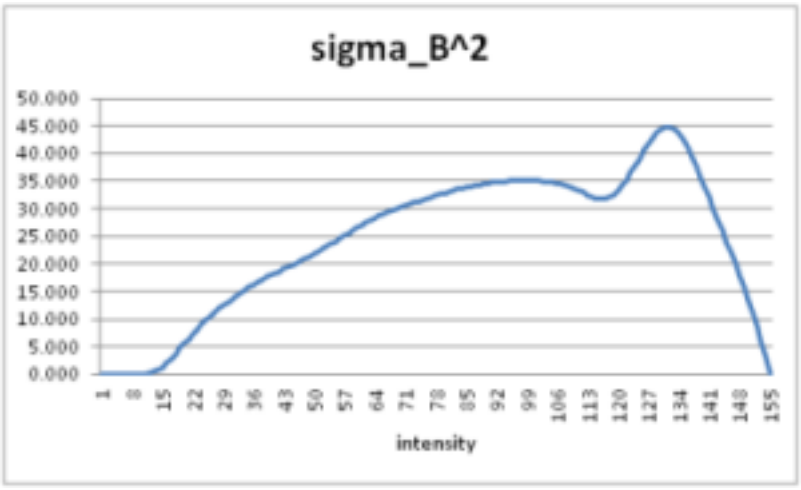
(b)

(c)

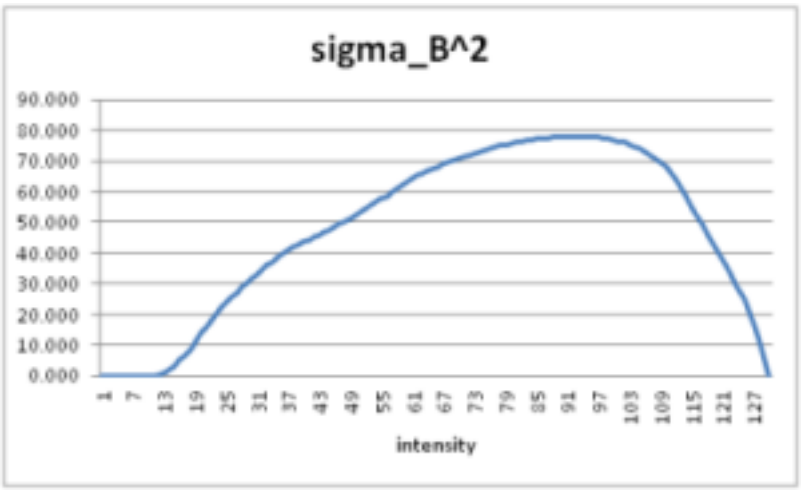
Figure 6.13: (a) the part of the intensity histogram below the first Otsu threshold of $T = 155$ for image 1 whose full histogram was shown in figure 6.11 (a); (b) the part of the histogram below the second threshold at 130 obtained when the Otsu algorithm is applied to the histogram in (a), and (c) the tail of the histogram below a threshold of 92 obtained from histogram (b) for comparison with figure 6.11 (c).



(a)

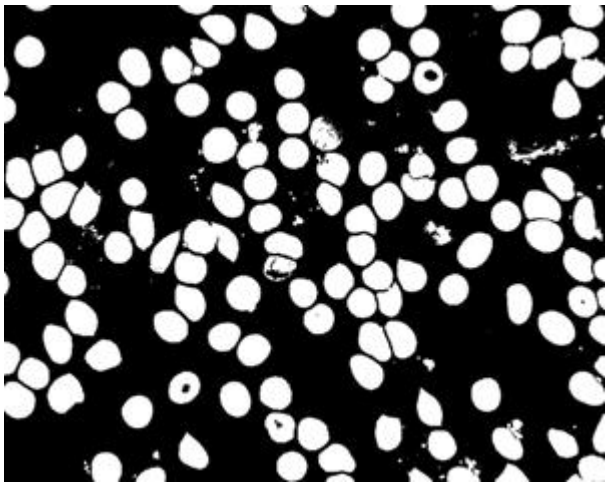


(b)

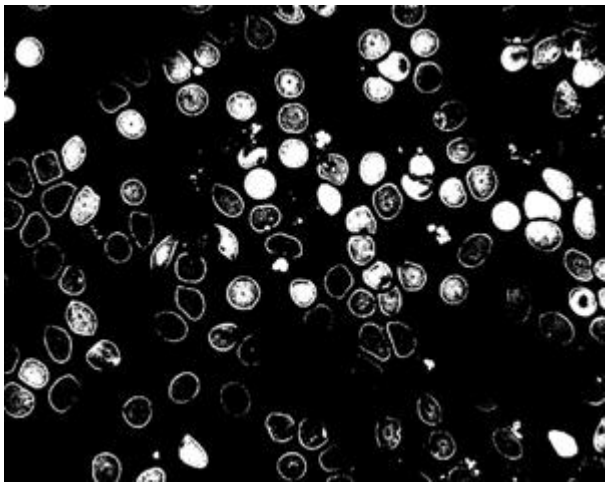


(c)

Figure 6.14: The between-class variance σ_B^2 as a function of the threshold T: (a) for the image intensity histogram of figure 6.11 (a) which has two distinctive peaks; (b) for the histogram below the threshold of $T = 155$ shown in figure 6.13 (a); and (c) for the histogram below $T = 130$ shown in figure 6.13 (b). In neither figure 6.13 (a) nor (b) is there structure in the histograms to indicate that they should be partitioned into two classes.



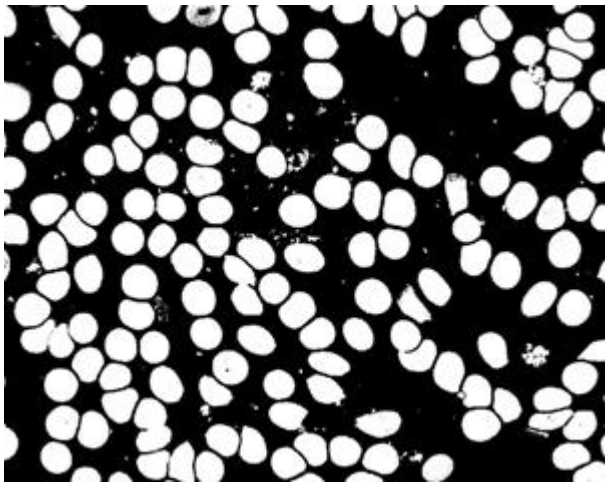
(a)



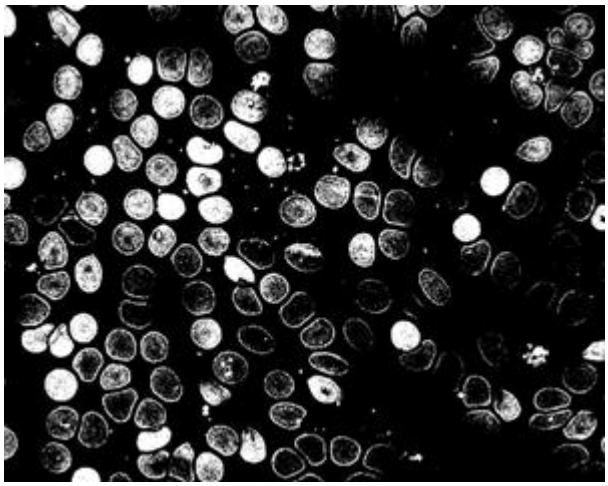
(b)

(c)

Figure 6.15: Recursive intensity-based segmentation of image number 1 shown in figure 5.1 using the Otsu algorithm: (a) the first application to produce a segmentation of RBC pixels from the background; (b) a second application, and (c) the third application.



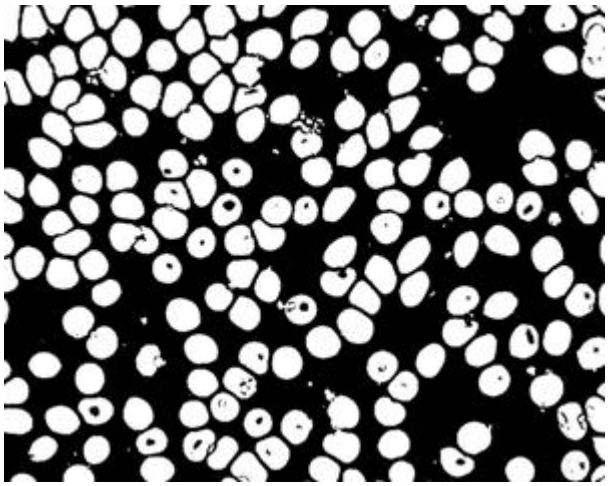
(a)



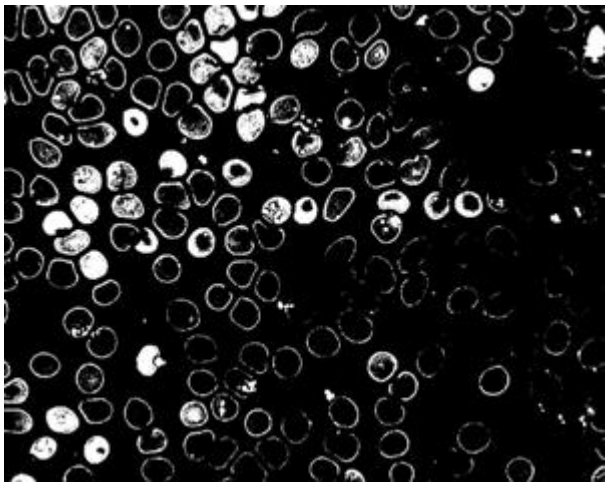
(b)

(c)

Figure 6.16: Recursive intensity-based segmentation of image number 4 shown in figure 5.1 using the Otsu algorithm: (a) the first application to produce a segmentation of RBC pixels from the background; (b) a second application, and (c) the third application.



(a)



(b)

(c)

Figure 6.17: Recursive intensity-based segmentation of image number 45 shown in figure 5.1 using the Otsu algorithm: (a) the first application to produce a segmentation of RBC pixels from the background; (b) a second application, and (c) the third application.

6.3.2 Recursive application of the Otsu 3D algorithm

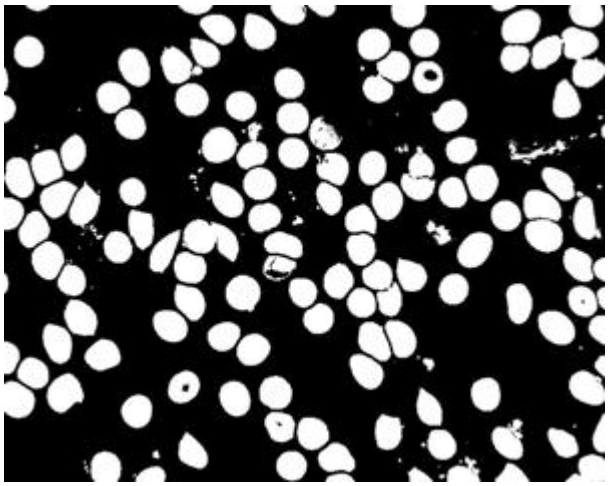
The extended Otsu 3D algorithm developed in Chapter 4, section 4.2.4 can be applied recursively in a similar manner. Segmentations produced by three-fold recursive application of the algorithm to the images shown in figure 5.1 are shown in figures 6.18, 6.19 and 6.20. It can be seen from comparison with figures 6.15, 6.16 and 6.17 respectively that in this colour-based segmentation rather more pixels tend to be selected at the second level of recursion than when the intensity was used.

The only slight difference in the implementation of recursive use of the Otsu 3D algorithm in comparison to that of the above recursive application of the one-dimensional Otsu algorithm itself is that selecting which of the two classes partitioned at each step should be used in the next step is a little more complicated. We can no longer in principle simply take the class which is below threshold, say, as when the intensity and 1D Otsu algorithm were used above. At each application of the Otsu 3D algorithm used to produce the results shown in figures 6.18 – 6.20 the minority class was nevertheless selected as described in section 4.2.5. The results seem not unreasonable, but it should be borne in mind that it is much more uncertain when selecting the minority class at the higher levels of recursion that the cluster containing the parasite pixels will be the minority class than it was at the first level that RBC pixels would be the minority class.

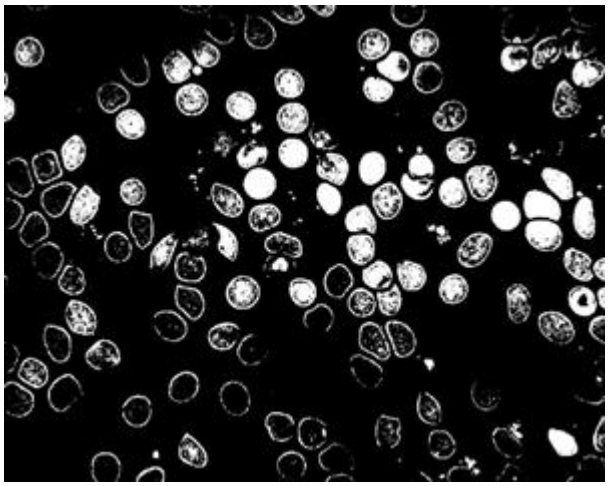
Since stained parasite pixels are dark and few in number, three alternative possibilities come to mind in addition to (i) selecting the minority class: (ii) selecting the class for which the mean $\underline{\mu}(J)$ is closer to the origin of the colour space, and (iii) selecting the class from the side of the decision surface which contains the origin. Finally, since the stained parasite pixels are dark blue a refinement of the above considerations would be: (iv) to take the class for which $\hat{\underline{\mu}}(J).\hat{\underline{b}}$ is the larger where $\hat{\underline{b}}$ is the unit vector $(001)^T$ along the blue colour axis. We note that the second and third criteria above would be the same if the decision surface were the perpendicular bisector of $\underline{\mu}(2) - \underline{\mu}(1)$. Although none of these alternatives were implemented in the preliminary work reported here, in practice all three of these criteria could be used with a warning to be flagged and (say) the last to take precedence in the event of disagreement.

6.4 Degree of infection

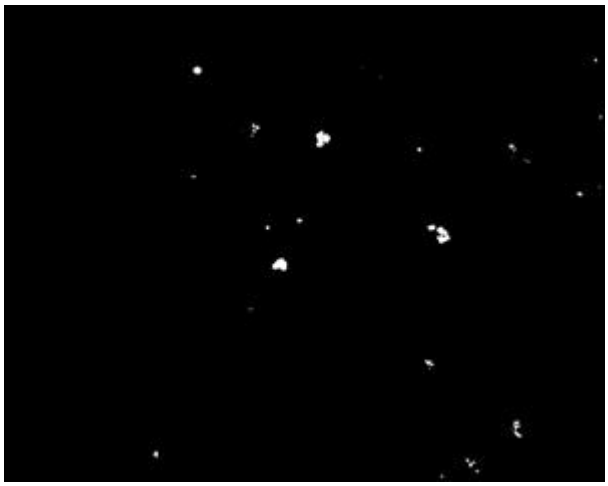
At this stage no detailed, systematic evaluation of the approaches discussed in the previous sections are available. However, though they are somewhat imperfect the preliminary results reported in sections 6.2 and 6.3 indicate that it might be possible to use unsupervised algorithms based on the Fisher discriminant to detect not only pixels belonging to RBCs, but also to determine whether some of them belong to stained parasite infections. This suggests that it is



(a)

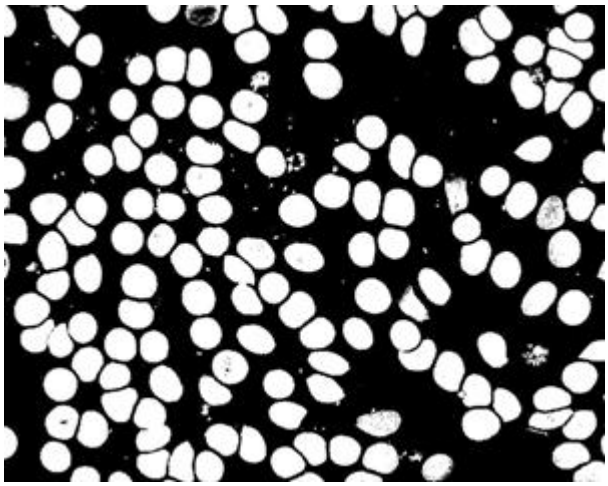


(b)

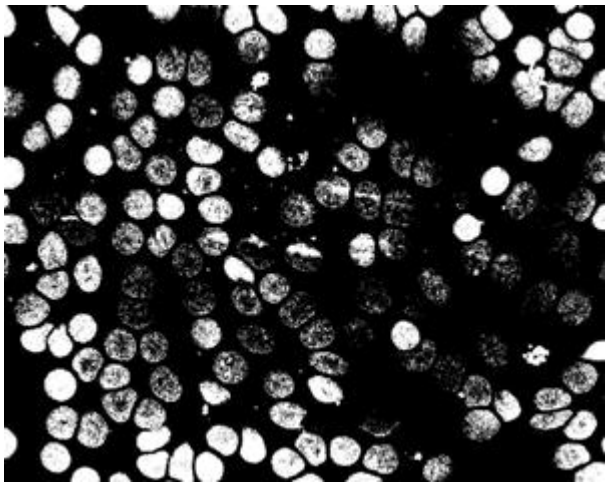


(c)

Figure 6.18: Recursive segmentation of image number 1 shown in figure 5.1 using our extended Otsu 3D algorithm: (a) the first application to produce a segmentation of RBC pixels from the background; (b) a second application, and (c) the third application.



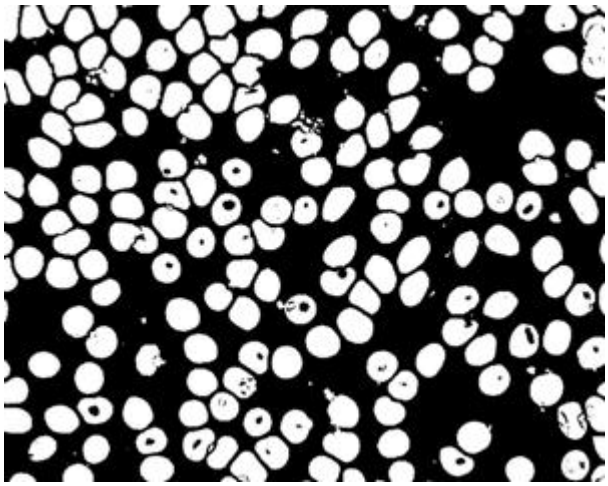
(a)



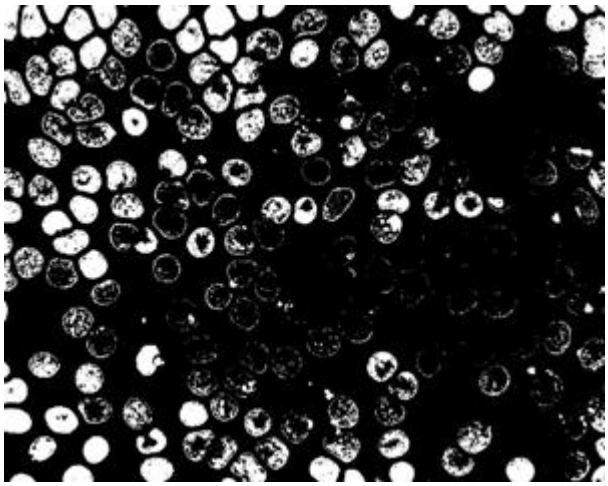
(b)

(c)

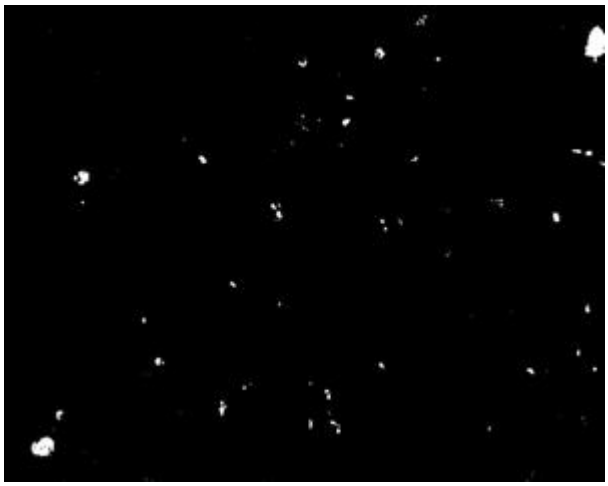
Figure 6.19: Recursive segmentation of image number 4 shown in figure 5.1 using our extended Otsu 3D algorithm: (a) the first application to produce a segmentation of RBC pixels from the background; (b) a second application, and (c) the third application.



(a)



(b)



(c)

Figure 6.20: Recursive segmentation of image number 45 shown in figure 5.1 using our extended Otsu 3D algorithm: (a) the first application to produce a segmentation of RBC pixels from the background; (b) a second application, and (c) the third application.

worth exploring how we might determine whether each RBC is healthy or infected and whether we could subsequently adequately determine the degree of infection or parasitemia.

Before doing so, we note that, in the procedures for detecting parasite pixels, perhaps we should use the outcome of the counting procedures established in section 6.1 to restrict the pixels selected from the first application of an Otsu algorithm to belong only to image objects representing singlet or larger clusters of adjacent RBCs. This would eliminate pixels belonging to small artefacts but it is not obvious whether doing so would improve final results for segmentation of parasite pixels. Making such a selection would be straightforward when discriminants based on the intensity are used since stained parasite pixels are even darker than pixels belonging to healthy parts of RBCs. Small, brighter artefacts inside RBCs are thus not a problem and application of a mask derived from the counting procedures should suffice. Unfortunately, because of the way orientation of the decision surface is affected by S_W^{-1}, a similar approach might not suffice when the Otsu 3D algorithm is used to segment pixels in the colour space. In general, a procedure based on identifying pixels contained within the boundary of a RBC might thus be required – which could, if desired, be implemented so as to include as "within a RBC" any of the small, bright artefacts mentioned above that are not currently classified as RBC pixels.

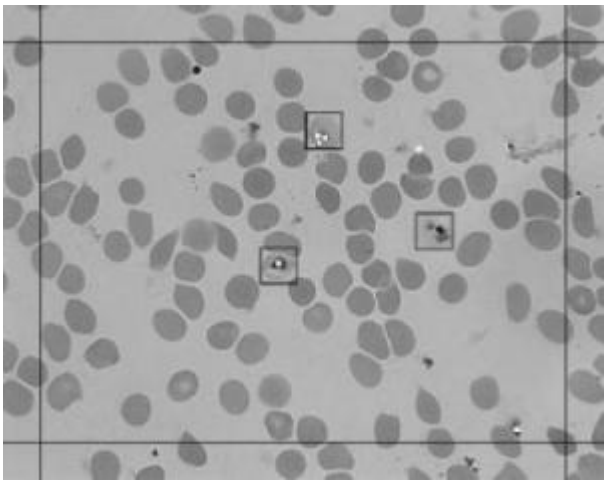
6.4.1 Determining whether a RBC is infected

In principle, we could say that any RBC which contained a parasite pixel was 'infected' but this would be very susceptible to error owing to 'false-positives', *i.e.* whenever there were pixels identified incorrectly as belonging to parasite infections[4]. For this reason and remembering that parasite infections should be within RBCs, after some experimentation we chose to designate a RBC object as infected according to one or the other of the rules below:

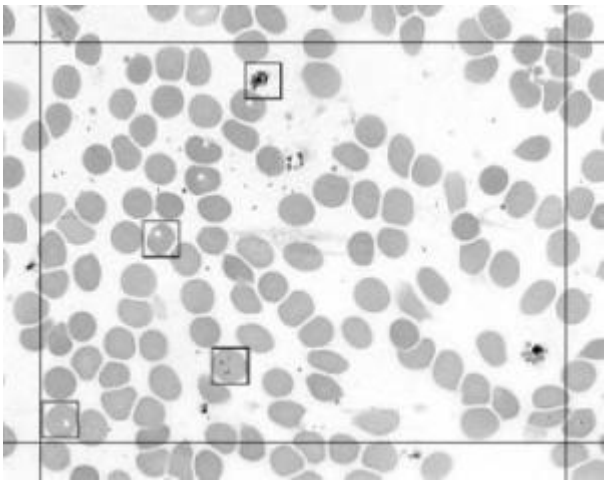
- there was a 4-connected cluster of at least 15 putative parasite pixels within a proximity box of ± 36 pixels of the centre of a RBC,
- there was a 4-connected cluster of at least 25 putative parasite pixels within a proximity box of ± 10 pixels of the centre of a RBC.

The former was first used for detection of parasite pixels by recursive application of the Otsu algorithm to the image intensities (section 6.3.1) as shown in figure 6.21. Clusters of parasite pixels selected by the procedure are marked by small, white squares, whilst cells identified by medical experts as infected are enclosed in square boxes.

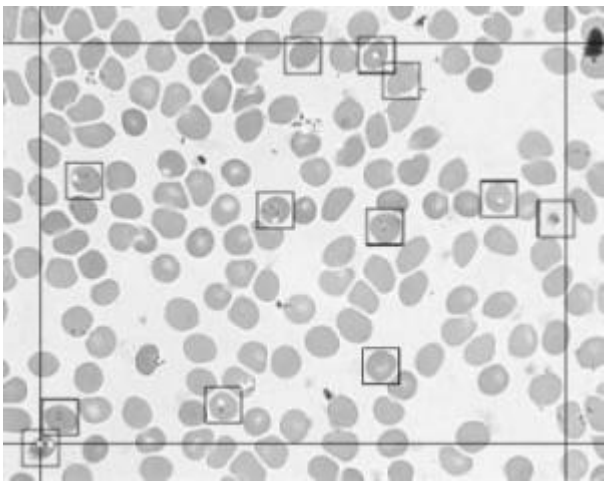
[4] *cf.* the approach adopted by Ross *et al* [151] discussed in the review for determining whether a sample was from an infected patient.



(a)



(b)



(c)

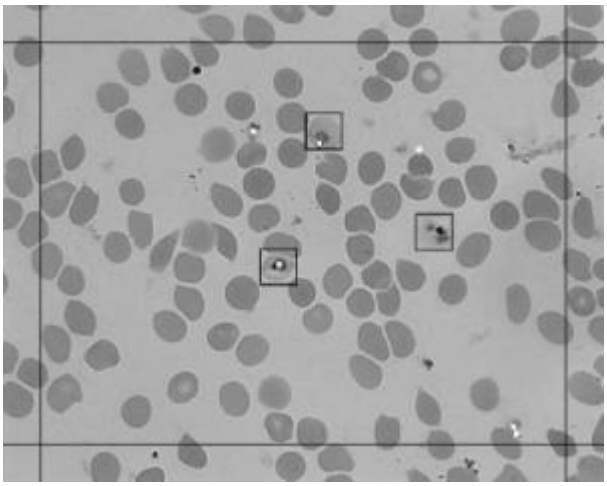
Figure 6.21: Results illustrating recursive application of the Otsu algorithm to the intensity for the identification of infected RBCs. Significant clusters of 15 or more 4-connected parasite pixels within ± 36 pixels of the centre of a RBC are indicated by small, white squares with cells identified as infected by medical experts enclosed within boxes: (a) for image number 1, (b) image 4, and (c) image 45.

We show in figure 6.22 a similar result obtained via the structural approach described in section 6.2.1 which utilises a lower threshold (73 *cf.* 92 for image number 1, 94 *cf.* 118 and 116 *cf.* 123 for images 4 and 45 respectively) and thus produces slightly fewer putative parasite pixels.

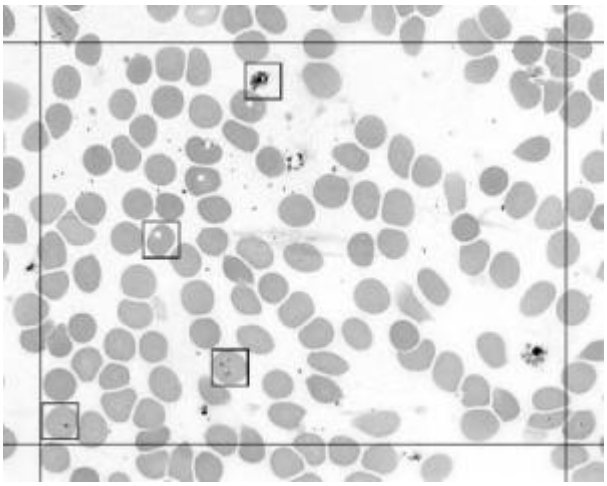
Recursive application of our Otsu 3D algorithm to colour data seems to be the most generous of our procedures for segmenting parasite pixels. Thus, when this approach (section 6.3.2) was used clusters of at least 25 4-connected putative parasite pixels within a proximity box of ± 10 pixels of the centre of a RBC as in the second rule above were selected as putative parasite infections as shown by the small red squares in figure 6.23. For comparison, application of the above "25 pixel cluster" criterion to segmentations produced by recursive application of the Otsu algorithm to the image intensities produced the results shown in figure 6.24.

None of the above approaches seems very satisfactory. Not only do we have to overcome the difficulty of detecting the very small minority class of infected pixels and that the stained pixels may be very faint (see for example, figure 6.23 (c)), but also that there may be few infected cells within an image. Furthermore, in using the pixels segmented as putatively belonging to parasite infections to identify infected RBCs we have to contend with the fact that some of these pixels may be located very near to, or what appears at first sight to be just outside, the boundary of a RBC. Medical experts tend to reject these as indicating infected cells (though they occasionally may make mistakes as figure 6.23 (b) appears to show!). It was to reject such pixels that the criteria for proximity to a RBC centre were introduced in the above. Without such criteria, the algorithms would produce many false-positive RBCs labelled as infected when they were in fact healthy. Finally, we note that close inspection of the images reveals that these pixels are frequently just inside protuberances on the boundaries of the image objects segmented as RBCs. These protuberances are included within the areas of cells in the counting processes and thus would not be removed at that stage, even if the robust counting procedure were used prior to attempts to segment parasite pixels.

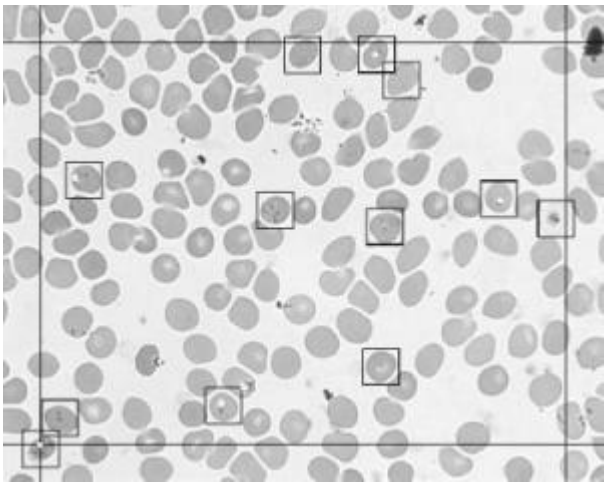
Though more research is evidently required to develop a way of detecting pixels belonging to stained parasite infections and of reliably labelling RBCs which are infected, we attempted to count the numbers of infected cells within all of the images in our database of 100. This preliminary work using the structural approach was carried out before the robust and reliable counting methods described in sections 6.1.2, 6.1.3 and 6.1.4 were developed. Multiple 'hits' of clusters of parasite pixels within a RBC (as can be seen for example in figure 6.21 (a) and figure 6.23 (c) were ignored as, during the parasite life-cycle stages of interest for determining the parasitemia, a cell almost never contains more than one parasite. Identification by medical



(a)

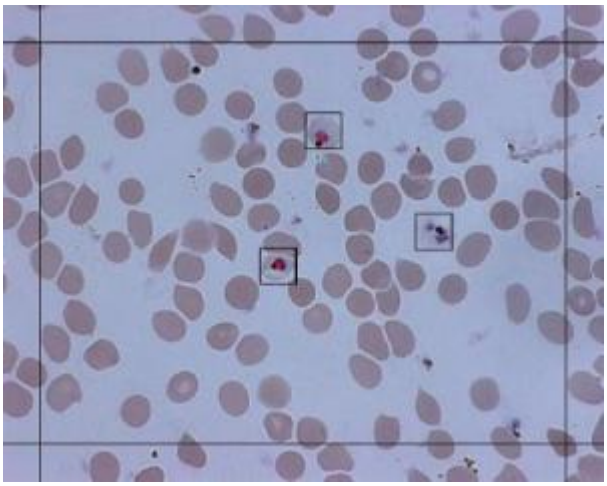


(b)

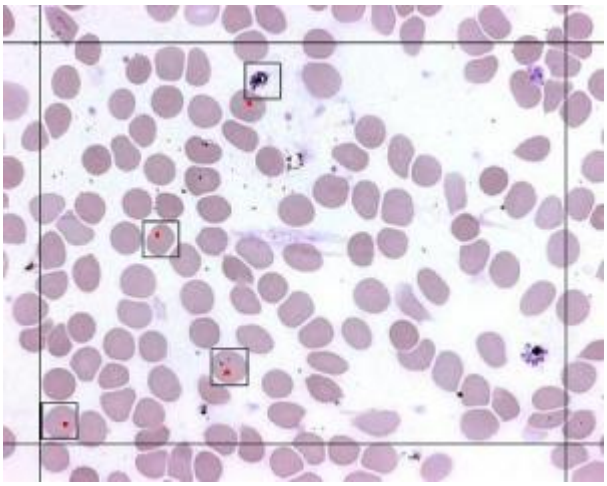


(c)

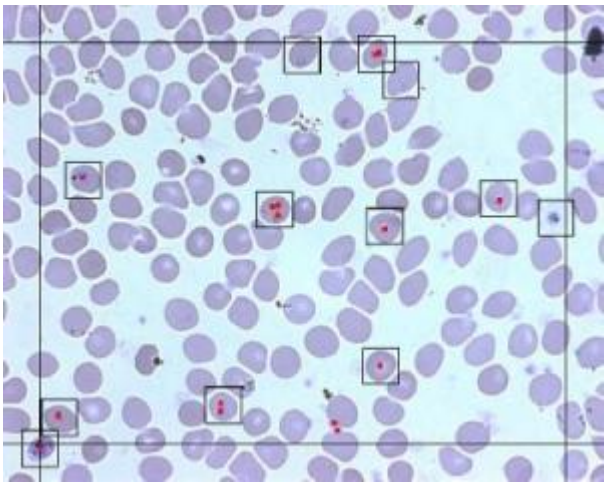
Figure 6.22: Results from using the structural approach to segment parasite pixels for identification of infected RBCs. As in figure 6.21 significant clusters of 15 or more 4-connected parasite pixels ± 36 pixels of the centre of a RBC are indicated by small, white squares with cells identified as infected by medical experts enclosed within boxes: (a) for image number 1, (b) image 4, and (c) image 45.



(a)

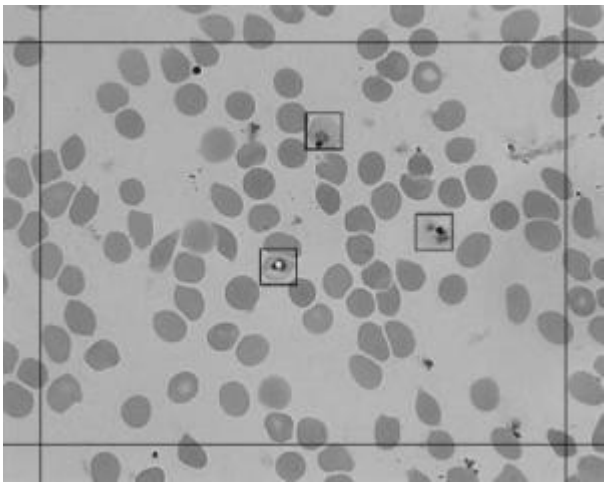


(b)

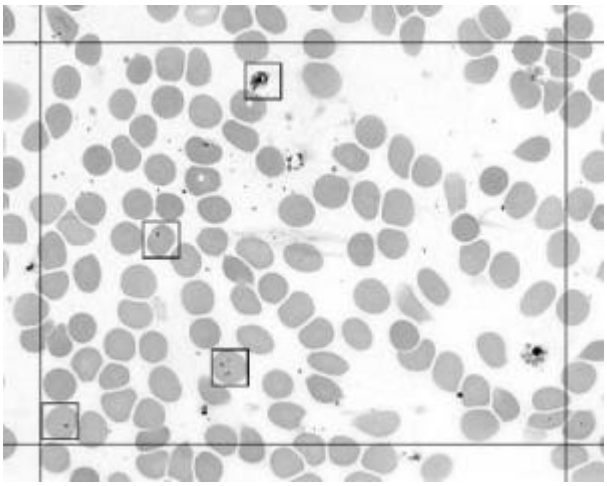


(c)

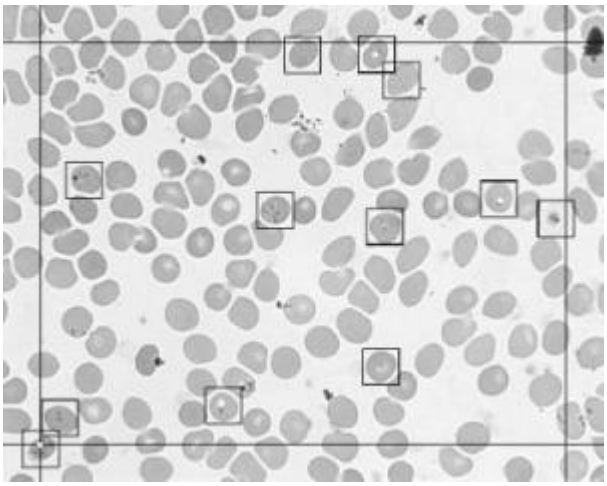
Figure 6.23: Illustrative results using recursive application of the Otsu 3D algorithm to colour data and marking significant clusters of 25 or more 4-connected pixels within ± 10 pixels of the centre of a RBC by small, red squares: (a) image number 1, (b) image 4, and (c) image 45. As usual, cells identified as infected by medical experts enclosed within boxes.



(a)



(b)



(c)

Figure 6.24: Illustrative results similar to those in 6.21 obtained by recursive application of the Otsu algorithm, but with significant clusters of 25 or more 4-connected pixels within ± 10 pixels of the centre of a RBC to indicate infected cells: (a) image number 1, (b) image 4, and (c) image 45.

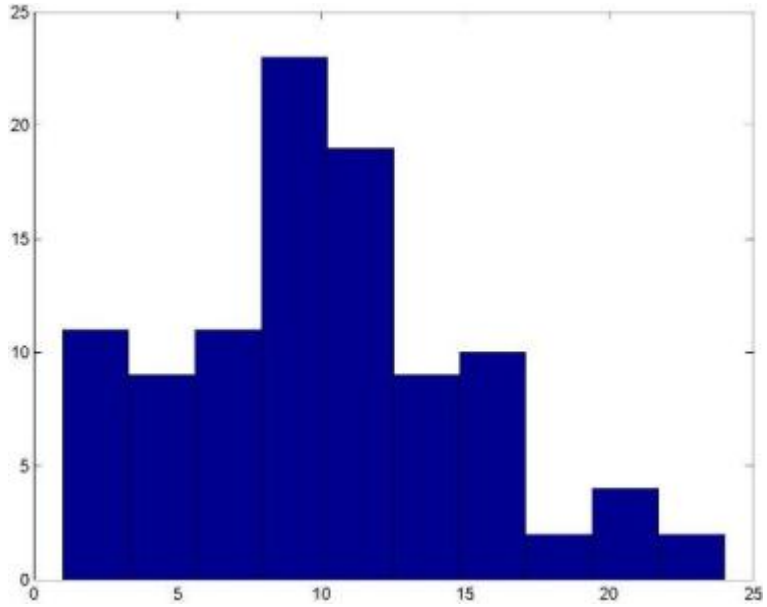


Figure 6.25: A histogram, consisting of ten bins, of the numbers of infected RBCs found in each of the 100 images in our database.

experts of infected cells in the 100 images in our database showed that, as the three sample images, numbers 1, 4 and 45 illustrate, there were not many infected RBCs in each image – in most cases less than 20 or so (Figure 6.25). Comparison of the computer counts of the number of infected RBCs in each image as described above with counts of the number of cells identified as infected by the experts is summarised in figure 6.26 and showed that this automatic counting tended to underestimate the number of infected cells. Given the small numbers of infected cells in each of the images it can be seen that these very preliminary computer counts of the number of infected RBCs are not sufficiently accurate to warrant calculation of the parasitemia. This was even though the ordinate in figure 6.26 is the net count in which false-positives and false-negatives may cancel. The error would undoubtedly have cast the results in a poorer light. In future research on the counting of infected RBCs it is important that detailed, separate consideration of the numbers of true-positives and false-positives be included.

Given the unsatisfactory and ad-hoc nature of the current preliminary work it was, as noted above, not worthwhile determining the degree of infection or parasitemia. In addition to improving the algorithm used, we note that a reliable and accurate determination of the parasitemia will also almost certainly require use of many images to provide a much more extensive sampling of the thin-film slide and thus greater numbers of infected cells for the estimation.

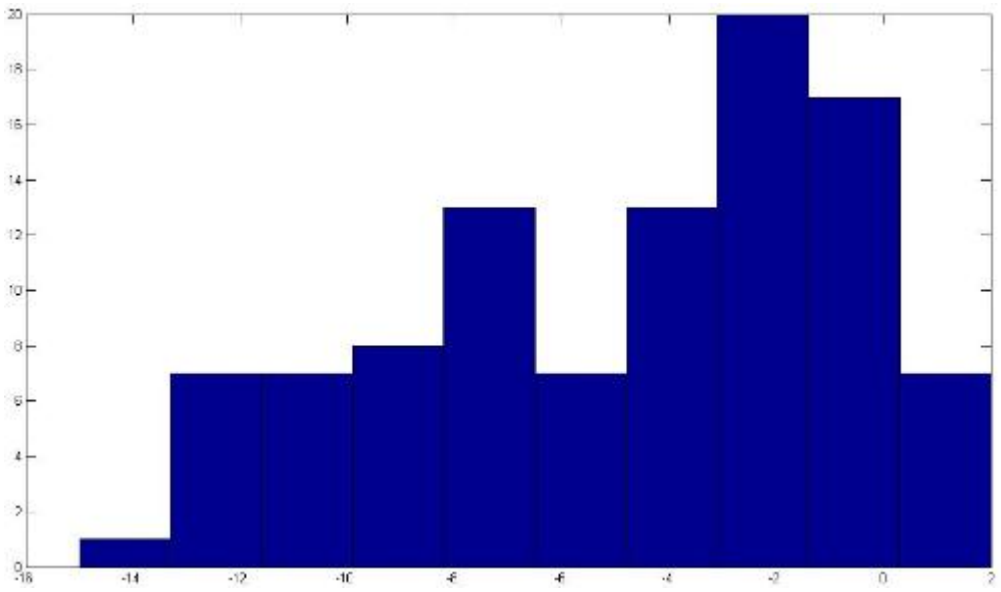


Figure 6.26: A histogram, also consisting of ten bins, showing that in preliminary work computer counts of infected RBCs tended to be under-estimates when compared to the counts made by medical experts.

6.5 Summary, discussion and conclusions

This chapter has been focussed on determining the number of RBCs in an image and on detecting whether they are healthy or infected with a malaria parasite. Successful implementation of these two operations and their application to a large number of images taken from a single thin-film slide would enable the degree of infection or parasitemia to be determined by machine. The work presented, however, is preliminary, relying for the most part on experiments carried out on three example images from our dataset of 100: numbers 1, 4 and 45 chosen to illustrate variations in the colour of the images, in the number of RBCs present, and in the number of cells which were infected.

Motivated by the distribution of the object sizes obtained from the segmentation procedures developed in chapter 4 we first considered how unsupervised algorithms based on the Fisher discriminant might be further used to determine the number of RBCs in an image:

(i) Classification of image objects according to their areas either as small artefacts to be ignored, or as clusters of singlet, doublet, triplet (*etc.*) RBCs was considered with particular attention to the fact that the area of doublet cell clusters should be approximately twice that of singlets, triplets three times, *etc.*. It was argued that it was better to employ such facts as an *internal* check on the classification of the image objects rather than to employ them as constraints on the optimisation of a discriminant.

(ii) It was thus decided to use the Otsu algorithm recursively to classify the image objects with a first application to removal of small artefacts and a second to classification of putative RBC objects as singlets or larger clusters of mostly adjoining doublet cells with perhaps a few triplets and very occasionally larger multiplets. This was developed not only to provide a means of counting RBCs from classification of individual RBC clusters but also to provide a statistical estimate of the number of RBCs from their average area as determined by the classification process. Furthermore, it led to development of three numerical parameters t_1, t_2 and t_3 whose smallness could be used as *internal* checks on the performance of the classification process together with a fourth parameter t_L whose smallness would indicate consistent labelling of individual RBC objects. A fifth parameter t_5 was developed to quantify the difference between the RBC count and the statistical estimate of the number of RBCs.

(iii) The first application of the Otsu algorithm to eliminate small artefacts was found to be straightforward with the first parameter t_1 as expected always small, but large values of all the other parameters, in particular $|t_2|$ which was expected to be most sensitive and informative, indicated problems with classification of RBC clusters for both segmentations of image 45 and for the colour-based segmentation of image 4.

(iv) Analysis showed that, whereas for both segmentations of image 1 and the intensity-based segmentation of image 4 the between-class variances of the putative RBC cluster object areas were simple and the Otsu algorithm could be expected to work, for both segmentations of image 45 and the colour-based segmentation of image 4 it was multimodal with the largest variance occurring at the largest value of the object areas at which it was extremal. Using this extremum to set the threshold for classification of image objects as singlet or (mostly) doublet RBCs then gave consistent results with low values of all five parameters t_1, t_2, t_3, t_L and t_N. Unlike image intensity histograms which are usually 'dense', it was noted that histograms of image object areas have many 'gaps'. If we recall equation 4.24 it appears that these gaps may be facilitating the multimodality of the between-class variance of the object areas though, as the example of image 1 showed, they do not always give rise to multiple extrema.

(v) After training by medical experts the number of RBCs in the selected images were counted by the author as developer/user and compared with computer-based RBC counts and statistical estimates obtained by the 'reliable' procedures summarised in (iv) above in which all the internal performance characterisation parameters were satisfactorily small.

Results for images 1 and 45 were almost perfect with errors of at most one cell in ~ 100 or more whilst those for image 4 were in error by $\sim 3\%$.

(vi) No attempt has been made quantitatively to compare the intensity-based and colour-based segmentations of RBCs though it was noted these seemed to be potentially most dissimilar for image 4, the image for which the final RBC counts in (v) above were most in error. It would seem that systematic comparison of the two RBC segmentation procedures described in chapter 4 – one based on the image intensity, the other on its colour – could be used to provide a further *internal* measure of confidence.

It thus seems that we have found a reliable and accurate way of using unsupervised algorithms based on the Fisher discriminant to determine the number of RBCs produced by an image segmentation procedure. The method requires a full, systematic evaluation for which it would be best to use a recursive algorithm to find the largest maximum of the variance via an exhaustive search. Such algorithms have been discussed previously in the literature (see for example the text books [135] and [164]). The algebra underpinning a recursive algorithm that would provide an efficient evaluation of σ_B^2 for an exhaustive search for the highest maximum of the between-class variance when there are many gaps in the histogram was presented in appendix E. Such an implementation would certainly be required were this approach to be used in practice.

In the following part of this chapter (sections 6.2, 6.3 and 6.4), we briefly explored the use of the type of unsupervised algorithms considered in this thesis for detecting pixels representing parasite infections and how such pixels might be used in deciding whether a RBC is healthy or infected.

(vii) It was noted that stained parasite pixels have a distinctive dark-blue colour; that usually only a small fraction of the RBCs are infected; and that, unless close to the end of its life-cycle within a cell, parasites occupy only a small fraction of a RBC. Stained parasite pixels are thus a very small, minority class comprised of $\sim 1\%$ of the pixels in an image or less and therefore difficult to detect reliably. Multiclass-algorithms were thus ruled out. The Otsu algorithm applied to image intensity data and our Otsu 3D extension applied to colour data were therefore used recursively. In each case, the smallness of the infected pixel class meant that the algorithms had to be applied three times: once to segment RBC pixels – then twice more, each time selecting the minority class, to segment putative parasite pixels. Some alternative ways of selecting the appropriate class at each recursive application of the Otsu 3D algorithm to colour data were discussed.

(viii) In addition, for comparison with recursive application of the Otsu algorithms, it was noted that structure of the RBC pixel peak in the intensity histogram could be used if followed by further applications of the Otsu algorithm to select very dark, stained parasite pixels from deep in the tail of the histogram.

(ix) It was not expected that any of the above techniques would produce an accurate segmentation of parasite pixels but an approximate segmentation that, when combined with appropriate rules for designating clusters of putative parasite pixels within RBCs as indicative of infection, would be sufficient for deciding whether a RBC was healthy or infected. Two rules setting different thresholds on the size of the cluster of pixels segmented as representing a parasite infection and on its proximity to the centre of a RBC were proposed.

(x) The three images, numbers 1, 4 and 45, were again used as explorative examples and illustrative results presented for each of the three ways of segmenting stained parasite pixels and subsequently using the two rules proposed in (ix) above to classify RBCs as healthy or infected. Cells classified as infected were compared with those designated as infected by expert medical researchers.

(xi) With such preliminary results for only three images, each containing only some $\sim 3-10$ infected RBCs, it was not possible to characterise the accuracy of the detection of infected RBCs but it was noted that the structural approach was more stringent – *i.e.* in any given image produced fewer putative stained parasite pixels – than recursive application of the Otsu algorithm to the image intensities, which itself tended to be more stringent than recursive application of the Otsu 3D algorithm to the image colour attributes. It could also be seen that whichever of the above approaches was used there were instances of both false negatives (infected cells classified as healthy) and of false positives (healthy cells classified as infected). It was concluded therefore that none of the methods explored was satisfactory and that none merited calculation of the degree of parasitemia.

(xii) In spite of the above, a net count of the number of infected cells was carried out for each of the 100 images in our database using the 'structural' segmentation of stained parasite pixels via location of the peak in the intensity histogram. Comparison with counts obtained by expert medical researchers showed that this computational approach, which was the most stringent of the methods explored for segmenting parasite pixels, systematically under-estimated the number of infected RBCs. Since false positives and

false negatives cancel in such a net count the results do not fully characterise performance and are optimistic.

The preliminary results discussed in this chapter for detecting pixels representing parasite infections and of classifying RBCs as infected or healthy are interesting but inconclusive. They suggest that hierarchical application of unsupervised techniques based on optimisation of Fisher discriminants might be made to work but evidently considerable research is needed in order to develop such approaches further and to evaluate the extent to which they could be used to meet the requirements of medical researchers for reliable computer counts of the number of infected RBCs and ultimately determination of the parasitemia of the blood film on a slide. This will be discussed in the concluding chapter.

Chapter 7

Summary, Conclusions and Discussion

In this chapter we briefly summarise the work reported in this thesis and the main conclusions to be drawn from it, followed by a discussion of their significance and the need/opportunities for further research.

7.1 Summary and Conclusions

Our work has focused on the automatic processing of images of thin-film microscope slides of malarial infected blood in order to provide counts of the RBCs, to detect and label cells which were infected with a malaria parasite, and if possible thence to calculate the degree of infection or parasitemia.

(0) This is a very different aim from the detection of malaria infections in the field, addressed for example by Ross *et al.* [151] for which thick-film slides are probably more sensitive [175], and the identification of the type of infection and, necessarily concomitantly, the stage of the parasite life cycle within a cell, as studied most comprehensively by Tek [175], the last of a series of several doctorate students with the group at the University of Westminster in the 2000s whose work was summarised in a paper in the Malaria Journal in 2009 [176].

7.1.1 Background

The main conclusions to be drawn from the review and background work were:

(i) Around the time this thesis commenced, the work of Sio *et al.* [165], which used traditional image processing and neural network techniques together with a statistical method for estimating cell counts, had focused similarly on counting RBCs and determining the parasitemia. It was noted that neural network techniques are widely used in the processing of images of biological samples, including in particular images of blood films and

other thin-film sections, but given the lack of a reliable, large database of images annotated by medical experts in the malaria application, our aim was to see to what extent simpler, unsupervised methods could be utilised to similar ends.

(ii) In the late 1990s and early 2000s, Liu and Sclaroff [111, 112, 110, 113, 114] used flexible models in a different application involving the analysis of blood, thin-film microscope slide images for misshapen RBCs indicative of various medical conditions. In the malaria application unless there is some other medical condition simultaneously present, RBCs remain ideally round. The application of flexible models to our task was thus reviewed and explored in chapter 3 via the simplest modelling approach, the flexible patch model (*FPM*), as building such a model of a RBC required only cells selected by the author as a 'semi-expert' user irrespective of whether they were healthy or infected. Unfortunately, application of the model in its active version, the *APM*, revealed that it was likely to be difficult to detect RBCs with acceptable false positive and false negative rates and, once again given the absence of a reliable, large database of annotated images that could be used as training data, the flexible modelling approach was abandoned in favour of unsupervised methods.

(iii) The simplest unsupervised criterion, maximisation of a Fisher discriminant, was chosen as, in image processing it leads to the very simple Otsu algorithm for segmentation of an image from its grey-level, intensity histogram. The Fisher discriminant has a long history in pattern recognition though much of the theoretical detail, including its relationship to the Otsu algorithm, is now buried in old and often neglected papers, including that by Otsu [131], or in old or unfashionable and often inaccessible texts such as [45, 97, 178, 135, 164, 46] and, in particular [43] which is long out of print. Considerable detail of this background was therefore given, where possible in mathematical appendices.

7.1.2 Application of Otsu algorithms to the segmentation of RBCs

Chapters 4 and 5 were devoted to the application and extension of Otsu algorithms to the segmentation of RBCs in thin-film, microscope slide images. This is an almost ideal application of these techniques as it is a simple two-class discrimination problem with the foreground class of RBC pixels typically representing $\sim 40 - 50\%$ of the image data. The main conclusion drawn from these chapters were:

(iv) Application of the Otsu algorithm to histograms of image intensities was successful for segmentation of pixels belonging to RBCs, irrespective of whether they were healthy or

infected with a malaria parasite. In addition to the almost ideal weight of the foreground RBC pixel class above, it was noted that the RBC and background peaks in the intensity histograms were well-separated and that the between-class variance σ_B^2, and hence similarly the Fisher discriminant, were mono-modal. The Otsu algorithm was thus guaranteed to converge to a unique threshold – again an ideal situation.

(v) A multi-dimensional extension of Otsu's algorithm was developed by alternate iteration of a projection of pixel colours onto the most discriminating direction $\hat{\underline{a}}$ in the colour space defined by a maximisation of a Fisher discriminant and subsequent selection of a threshold by application of Otsu's algorithm to the histogram of the projected data. Application of this procedure, initialised via PCA of the total covariance matrix S_T of the pixel colours and called here the Otsu 3D algorithm (section 4.2.4), to the segmentation of RBC pixels converged in $\sim 10 - 15$ iterations. The algorithm appeared to produce a successful segmentation of RBC pixels for every image studied.

(vi) Evaluated together with the Otsu algorithm (iv) applied to the image intensities and an application of the Otsu algorithm to individual colour channel data and in combination, our extended Otsu 3D algorithm was found to be just superior (section 4.2.5). This result, based on the evaluation of 17 $1/9$th sub-images from the 100 thin-film slide images made available to us by medical experts at the National Institute for Medical Research (NIMR), was published in [23]. ROC curves and the MRROC construction were used for these comparisons based on a detailed, laborious labelling by the author as a 'semi-expert' user of the pixels belonging to RBCs in the aforementioned 17 sub-images.

(vii) When a one-dimensional feature such as pixel intensity is used the Otsu algorithm is equivalent to an appropriately initialised C-means clustering algorithm and, in our case for distinguishing RBC pixels from pixels belonging to the background plasma, specifically to a 2-means algorithm. This equivalence, however, no longer holds for our extended Otsu 3D algorithm as the orientation of the decision surface perpendicular to the final discriminating vector $\underline{a} \propto S_W^{-1}[\underline{\mu}(2) - \underline{\mu}(1)]$ produced on convergence is in general no longer normal to $\underline{\mu}(2) - \underline{\mu}(1)$ as it would be for a 2-means clustering with the same sample labels. Analysis of three illustrative example images showed that S_W is very anisotropic and that its effect on the orientation of the decision surface is significant.

(viii) Behaviour of the extended Otsu 3D algorithm during iteration was analysed via evolution of the discriminating direction $\hat{\underline{a}}$ for an illustrative example image. It was found that $\hat{\underline{a}}$

had a large basin of attraction to a successful segmentation that covered most, but not quite all of the unit sphere. There was only a small band of initial directions $\underline{\hat{a}}$ on the unit sphere for which the final segmentation was poor and an even smaller stripe of initial directions from which the algorithm failed to converge in 100 iterations.

(ix) Since the colour of the images in the dataset provided by the NIMR varied somewhat the extent to which the algorithms would be invariant under affine transformations of the colour space was analysed. The Otsu algorithm was found to be completely invariant, but only the iterative steps of our extended Otsu 3D algorithm were invariant. Initialisation via PCA of the total covariance matrix S_T is invariant only if the colour transformation matrix is proportional to an orthogonal matrix – unfortunately ruling out the common colour transformations defined by independent shifts and scaling (gains) in each colour channel. However this lack of invariance of the PCA initialisation of the algorithm is mitigated by the large basin of attraction (viii) above.

(x) Ways of comparing the variations in colour between different images (or sub-images) were discussed, both with and without application of affine colour transformations to improve colour alignment. It was argued that a measure, named the Kolmogorov criterion by analogy with a measure in the old pattern recognition textbook [43], might be preferred as it is dimensionless and restricted to the range $[0 - 2]$ but it was not implemented.

7.1.3 Counting RBCs

Chapter 6 was concerned first with the counting of RBCs:

(xi) Exploration of the size distribution of image objects obtained from the segmentation of RBC pixels suggested that the Otsu algorithm might be further applied, in a recursive manner, in order to classify the objects which, if they were individual RBCs would be wholly visible within the field of view, as small artefacts to be ignored, singlet RBCs and clusters of two or more adjoining RBCs. Object area was selected as the discriminating feature because of the simple relationship between the areas of singlet RBCs, doublets, triplets *etc.*. Building these relationships as hard constraints into the algorithm would have destroyed its simplicity whilst utilising them as soft constraints would have introduced an extra parameter.

(xii) The relationships between the object areas were therefore used to provide *internal* checks on the performance of object size classifications and, together with an additional test based on comparison with a statistical estimate of the number of RBCs, on the reliability

of the automated counts of the number of cells wholly within the field of view. Study of the three illustrative images selected from our database of 100 showed that application of the Otsu algorithm in this manner was *not* in general successful because of multi-modality of the between-class variances, in particular at the second level of recursion when singlet RBCs were being distinguished from larger clusters of adjoining cells.

(xiii) Location of the largest maxima of the between-class variances by exhaustive search was shown however to be successful and to lead to object classifications and RBC counts which, for all three example images and for both intensity-based and colour-based segmentations of RBC pixels produced respectively by means of the Otsu algorithm and our extended Otsu 3D algorithm, passed all tests and produced completely consistent, reliable results[1].

(xiv) For two of these example images, the RBC counts were also accurate, to within less than 1% (*i.e.* to within one cell) in comparison to manual counts made by the author as a 'semi-expert' user. For the remaining example image (4 in our database) the automated counts overestimated the number of RBCs by $\sim 3\%$ (*i.e.* by 4 cells). In all cases, similar counts to within one cell, were obtained both by counting the individual image objects classified as RBC singlets, doublets *etc.* and via statistical estimates based on the average area of singlet RBC objects. These consistent results were obtained when either an intensity-based segmentation of RBC pixels produced by the Otsu algorithm or a colour-based segmentation produced by the Otsu 3D algorithm was used.

7.1.4 Parasite detection

The last part of chapter 6, sections 6.2 – 6.4, described some preliminary work on the use of unsupervised, Fisher-Otsu algorithms ultimately to label RBCs as infected or healthy.

(xv) Attempts to segment pixels belonging to parasite infections were, unfortunately, not sufficiently successful to allow reliable identification of infected RBCs even though, since we are not interested in the type of infection or stage of the life cycle within a RBC, a rough segmentation of clusters of pixels representing parasite infections within RBCs would suffice to determine whether a cell was infected or healthy.

(xvi) Segmentation of pixels representing parasite infections within RBCs is difficult because

[1] Somewhat surprisingly, cell counts obtained (table 6.2) from inconsistent classifications of putative RBC clusters when the Otsu algorithm was used at both levels of recursion as in (xii) above were found to be closely similar to the 'reliable' cell counts (table 6.5) obtained from the consistent classification described here.

such pixels are a very small minority class, usually $< 1\%$ of the pixels in an image and $< 2\%$ of the pixels belonging to RBCs themselves, owing to the small fraction of RBCs infected (typically itself at most $\sim$ a few $\%$ in laboratory samples) and, unless the parasite is near the end of its life cycle within a RBC, the small size of the parasite body within a cell. Three-fold recursive applications of both the intensity-based (Otsu) and colour-based (Otsu 3D) algorithms were thus used to segment such pixels. A method which tried to exploit the approximate symmetry of the RBC peak in the intensity histogram to select pixels prior to a second application of the Otsu algorithm was explored for comparison (section 6.2.1).

(xvi) Given these difficulties and the small number of infected RBCs ($\sim 3 - 15$) identified in each image by medical experts from the NIMR it was not possible to reduce the number of false positives (healthy RBCs classified as infected) and false negatives (infected RBCs classified as healthy) to a sufficiently low level to permit a meaningful calculation of the parasitemia. Use of several images from different parts of a microscope slide to raise the number of infected cells to a reasonable number is evidently required together with further research, either to develop further the kind of unsupervised approach presented in this thesis or alternatives (see sections 7.2.2 and 7.3 below), hopefully to deliver a more accurate classification of RBCs as infected or healthy.

7.2 Discussion

Though a much more extensive and systematic evaluation is required before any of the methods explored in this thesis could be used in the laboratory it would seem that we have, as noted above, overall two successes: segmentation of RBC pixels and counting of RBCs and, owing to the preliminary nature of the work carried out, one inconclusive outcome: the labelling of RBCs as infected or healthy.

7.2.1 Segmentation and counting of RBCs

First, *re* the successes of using unsupervised Fisher-Otsu algorithms for the segmentation of RBC pixels and the counting of RBCs wholly visible within an image:

(i) When the feature space is one dimensional as in the segmentation of RBC pixels from their intensity or the classification of image objects according to their area, the between-class variance should be maximised via a recursive implementation of an exhaustive search in preference to using the Otsu algorithm. Since the Otsu algorithm typically converges within a few iterations the computational load of the exhaustive search is likely to

be greater but, with modern computers, not punishingly so. For counting RBC objects such an implementation is essential[2], but in general the extra computational effort would be well worth the gain in reliability even when the Otsu algorithm appears, as in the RBC pixel segmentation, to be working well.

(ii) Though it appears never to have failed for any of the images to which it was applied, convergence of the Otsu 3D algorithm used in the colour-based segmentation of RBC pixels is more problematic. Unfortunately it is more difficult to envisage an efficient, feasible exhaustive search procedure that could be used in this case. In the three-dimensional colour space, the decision plane may be represented by each of its intercepts on the red, green and blue colour axes. A crude exhaustive search for the decision plane which would optimise a Fisher discriminant would, for the usual 24-bit colour imagery, thus require the order of 16 million calculations of the discriminant. However, even for our 1.3 mega-pixel images, most of the bins in a colour histogram would be empty, so an efficient representation should be used. These are commonly found in image retrieval systems (see for example chapters 8 and 6 of [164]) and may, for example, be based on a multi-resolution binary tree [9] or some kind of transform (including PCA [182][3]) used to compress the representation. It would also seem that calculations of the discriminant could be made recursive so, without going into details, it would seem there is scope for an efficient, exhaustive search procedure to be developed for the Otsu 3D algorithm should application needs warrant it.

(iii) Comparison of the intensity-based and colour-based segmentations of RBC pixels was mentioned in section 6.1.3. The quantitative calculation of the similarity of the segmentations produced by these two approaches (or any other) has not been carried out, although it would be straightforward to do so. Though the two methods are not entirely independent, close similarity of the two segmentations would give confidence in the results produced by either and would seem well-worth implementing as a routine check in any application. Figure 7.1 left, (a), (c) and (e) shows RBC pixels included in the intensity-based segmentations of figure 6.3 but not in the colour-based segmentations of figure 6.4 whilst at right (b), (d) and (f) it shows the converse. It can be seen that the more generous segmentation includes more pixels around the boundaries of RBCs and that, with the exception of image 1, the intensity-based-based segmentation tends to be

[2]But see figures 6.6 and 6.7 and point (iv) below for a warning that care and, in practice, use of the Otsu algorithm, would be required in precisely locating maxima of σ_B^2 in gaps in the area histograms.

[3]This thesis also discusses various ways of comparing colour distributions – a matter also mentioned in [164].

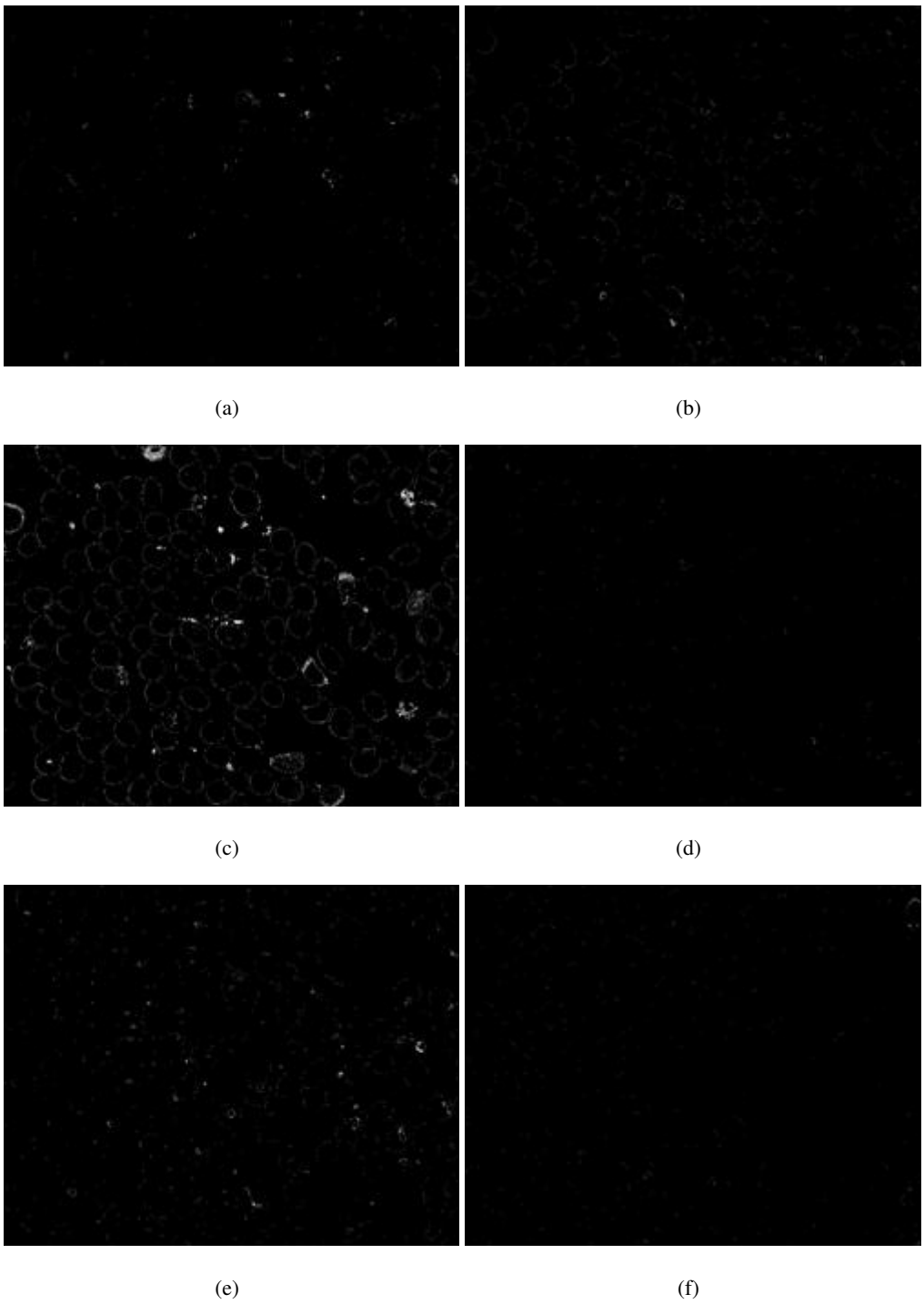
the more generous in accord with the figures given in the first row of table 6.1. More striking is the much greater difference in the two segmentations of image 4 visible in figure 7.1 (c) where there are many pixels in the intensity-based segmentation that are not included in the colour-based segmentation. However without implementing exhaustive search procedures such as those discussed in (i) and (ii) above and some means of assessing the reliability of each segmentation, perhaps such as those discussed in (iv) below, we do not know which may be inadequate.

(iv) Furthermore, reliability measures of each segmentation method could themselves be constructed. For example, given an exhaustive search of the between-class variance in one-dimension which yielded several maxima, $\sigma_B^2(k)$ say, labelled such that $\sigma_B^2(1) \geq \sigma_B^2(2) \geq \cdots \geq \sigma_B^2(K)$, a reliability measure taking values in the range 0 (completely unreliable when there are many equal maxima) to 1 (completely reliable when there is a single, unique maximum or when there is one very prominent maximum of value $\sigma_B^2(1)$ in comparison to which the remaining $K-1$ maxima are negligible) would be:

$$\begin{aligned} R &= \frac{\sigma_B^2(1) + \sum_{k=1}^{K}(\sigma_B^2(1) - \sigma_B^2(k))}{K\sum_{k=1}^{K}\sigma_B^2(k)} \\ &= \frac{1}{K}\left\{\frac{(K+1)\sigma_B^2(1)}{\sum_{k=1}^{K}\sigma_B^2(k)} - 1\right\}. \end{aligned} \tag{7.1}$$

This measure (and perhaps even just a simple count of the number of maxima, K) could be applied to both an intensity-based segmentation of RBC pixels (where we would find $R = 1$) and to the classification of image objects by area as small artefacts, singlet RBCs, or doublets. In the latter, the graphs in figures 6.6 and 6.7 indicate that, at the first level of recursion, we might again find $R = 1$ yet be given little warning[4] of the large plateaux in σ_B^2. A local curvature measure would be required in order to flag such situations. At the second level of recursion for classification of the image objects from their areas, the values of the three maxima in table 6.3 would more usefully give $R = 0.174$ and $R = 0.172$ for the classification of the RBC objects obtained from the intensity-based and colour-based segmentations respectively. Whether these are large enough to give confidence in the image object classification requires further research and evaluation over a wide range of images but we recall that in addition one can also (and should) use the parameters already discussed in section 6.1.3 as checks on the RBC object classification and counting processes.

[4]Recall the footnote to item (i) above.



(a) (b)

(c) (d)

(e) (f)

Figure 7.1: Left: RBC pixels included in the intensity-based segmentations of images 1 (a), 4 (c) and 45 (e) but not included in the colour-based segmentations, and right: the converse.

If a search procedure were used as indicated in (ii) above, thought not quite so straightforward a similar reliability measure to that defined in 7.1 could in principle be constructed for the colour-based segmentation of RBC pixels. As in the one-dimensional case, it would also need to be supplemented by local curvature measures in order to flag the occurrence of ridges and plateau regions in σ_B^2.

(v) The colour-based segmentation offers another way of potentially characterising the reliability of the segmentation via stability of projection of colour data onto the discriminating direction $\underline{\hat{a}}$. For example, one could use the total covariance matrix S_T to check stability of projection of all the pixel colour data onto the discriminating direction, and similarly use the between and within-class covariances S_B and S_W respectively and, if desired, the individual class-conditioned covariances $S(1)$ and $S(2)$. A suitable measure in the range $[0-1]$ may be constructed from an eigen-decomposition of each of the matrices:

$$P = \frac{\sum_{k=1}^{3} |(\underline{\hat{a}}^T \underline{v}(k))\lambda(k)|}{\sum_{k=1}^{3} \lambda(k)} \tag{7.2}$$

where $\lambda(k)$ are the eigenvalues of the matrix labelled in descending order and $\underline{v}(k)$ are the corresponding eigenvectors. The between-class covariance S_B is rank one and has only one non-zero eigenvalue with eigenvector $\underline{v}(1) \propto (\underline{\mu}(2) - \underline{\mu}(1))$ so 7.2 in this case reduces to:

$$P = cos(\alpha) \tag{7.3}$$

where α is the acute angle between $\underline{a}$ and $\underline{\mu}(2) - \underline{\mu}(1)$ whose values for the three illustrative images shown in figure 5.1 were given in table 5.1. It can be seen from the table that for image 4 the value of α is quite different from the values $\sim 25^\circ$ and $\sim 35^\circ$ taken for images 1 and 45 respectively. Again further research is required to establish the extent to which such *internal* performance indicators might be useful.

(vi) Were use of the kind of *internal* performance parameters discussed above and elsewhere in the main body of the thesis to indicate potentially unsatisfactory or unreliable results in the segmentation of RBC pixels, for example in applications to a much larger range of imagery than the 100 in our small database, one further opportunity remains open as noted in section 2.2.1 via a kind of *transfer learning*. Since the fraction of RBC pixels in a correctly taken image of a thin-film slide should always be in the range $\sim 30-50\%$ such failure is likely to be caused by variation in image colour. To overcome this, we could try to use the techniques discussed in section 5.4 to adjust, by histogram or colour histogram alignment, the colour of such troublesome images to be more like the colours of images

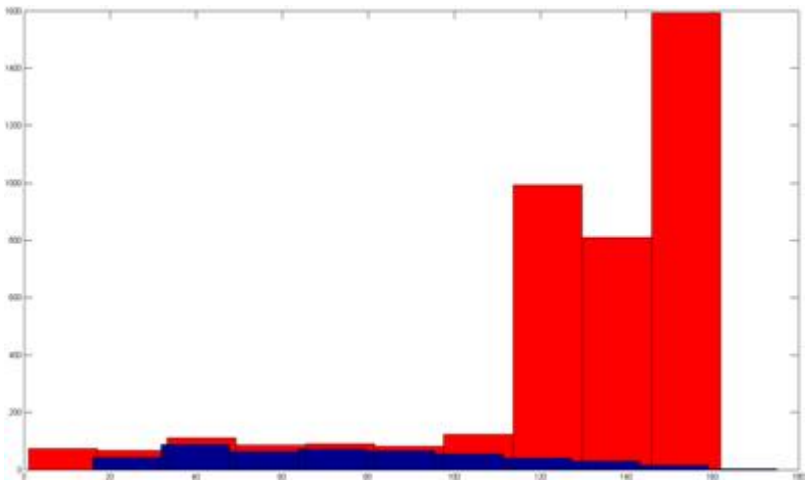
which had been successfully processed. Such techniques have been used in the context of an inspection application [163] and histogram alignment algorithms developed for that purpose [162]. Again, however, further research is required in order to see if such colour alignment would be effective and whether new algorithms would need to be developed in order to carry it out, for example, in accordance with the criteria discussed in section 5.4.

7.2.2 Parasite detection and calculation of parasitemia

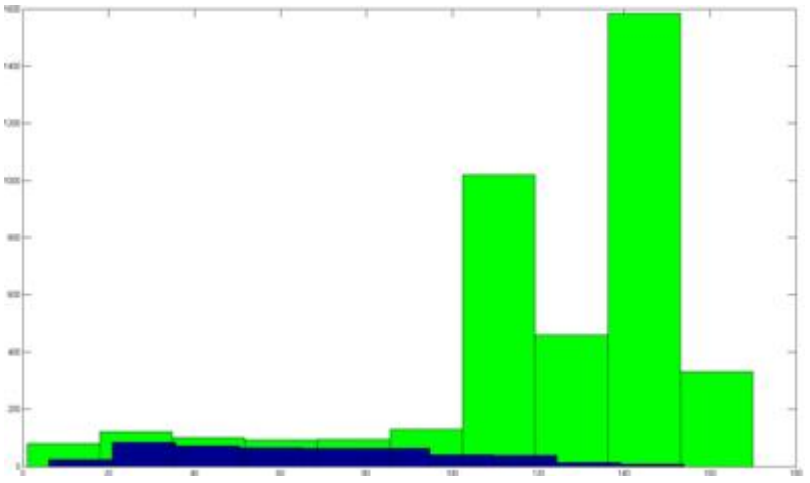
It is surprising that few (if any, except perhaps [145] – see section 7.3 below) authors have commented on the difficulty of reliably detecting pixels representing parasite infections within RBCs when such pixels usually form a small minority class, relative to both the total number of pixels in an image and to the number of RBC pixels. Within the context of ROC analysis Flach [57] discussed the problem of class imbalances and the trade-off of decision costs and values although particular attention was not drawn to the difficulty of developing reliable classification systems in such circumstances.

(i) The colour and configuration of Giemsa stained parasites seems quite distinctive to an expert medical observer but their colour is not sufficiently distinctive (figure 7.2) to enable such pixels easily to be detected individually. It is possible that the crude discretisation of these histograms is misleading but, given this data, it is a little surprising that it is possible to detect parasite pixels at all from intensity data. Projection of the pixel colours on the green-blue and red-blue colour planes as shown in figure 7.3 indicates a more optimistic scenario for detecting such pixels in the colour space especially if, as for our extended Otsu 3D algorithm, the orientation of the decision surface is determined by the procedure and is not fixed a priori relative to the colour axes. Finally, we note that these histograms indicate that there might not be much to be gained by using only pixels within reliably counted RBCs as input to parasite pixel detection rather than using directly the pixels segmented as belonging to RBCs – *i.e.* as in sections 6.2 and 6.3 by applying the Otsu algorithms recursively to the whole image.

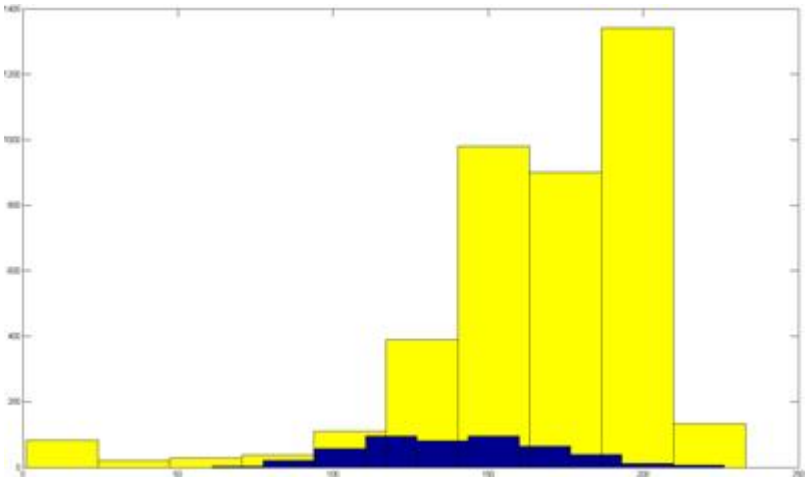
(ii) In light of the data presented in figures 7.2 and 7.3 it may be asking too much to expect an unsupervised algorithm, especially if the predicate or criterion is a function only of individual pixel attributes, to work well for the segmentation of parasite pixels. A first step away from such a paradigm might be to use pixel data labelled as originating from healthy or from infected parts of RBCs to infer a suitable choice for the decision plane. Using such data to estimate $\underline{\mu}(1), \underline{\mu}(2)$ and the class-conditioned covariances $S(1), S(2)$ and hence the within-class covariance S_W, the vector $\hat{\underline{a}}$ and threshold T on this dis-



(a)

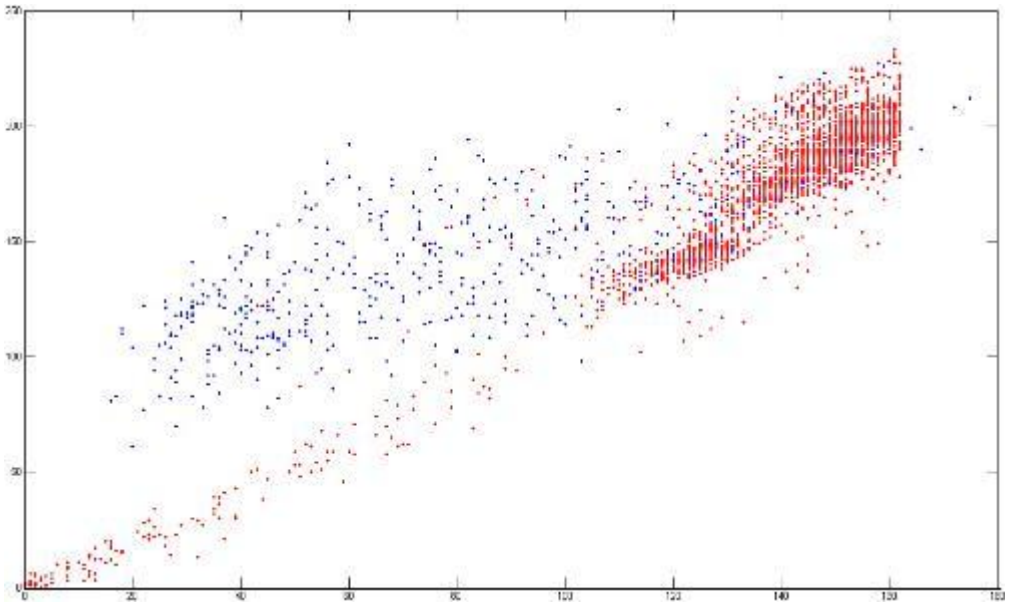


(b)

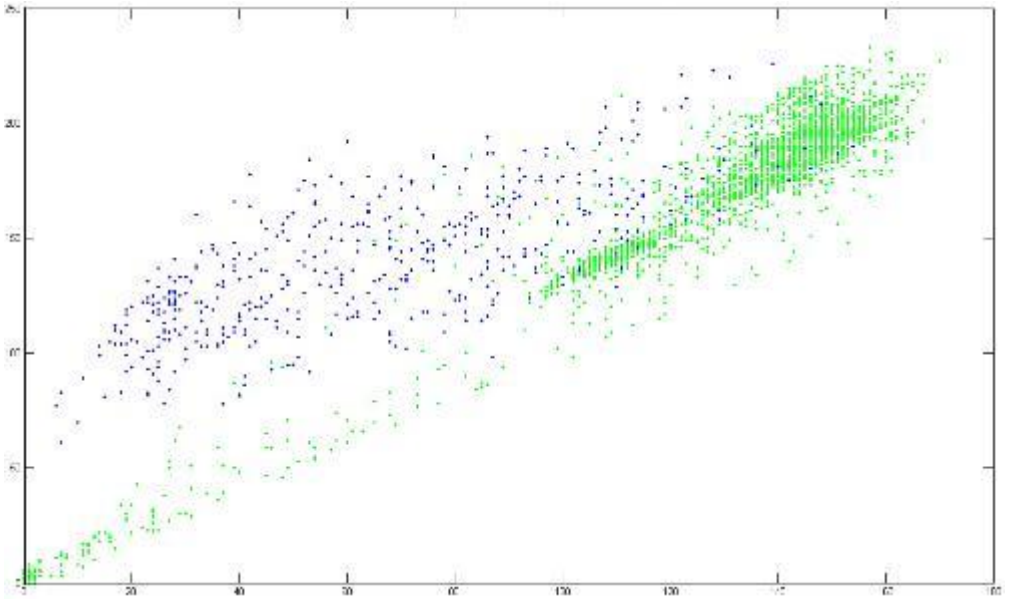


(c)

Figure 7.2: Histograms of pixels from healthy parts of a number of RBCs identified by medical experts as infected and from the stained parasite infections within the cells: (a) in the red colour channel, (b) the green channel, and (c) in the blue channel. Counts of pixels belonging to stained parasites are represented in blue throughout with the counts of pixels from healthy parts of cells shown in a contrasting colour: red for the red channel, green for the green channel, but in yellow for the blue channel.



(a)



(b)

Figure 7.3: Colours of pixels from a number of infected RBCs projected: (a) on the RB colour plane and (b) on the GB colour plane. In each case, as in figure 7.2 pixels belonging to parasite infections are depicted in blue with pixels from healthy parts of cells depicted in a contrasting colour.

criminating direction would construct in a supervised manner a type of least-mean square classifier corresponding to the Fisher discriminant as noted in section 4.1.1. Alternatively, such data could be used to train a modern capacity controlled neural network [13, 48] or support-vector-machine [190, 18] which, with suitable kernel functions [39] would enable the decision surface in RGB colour space to be non-planar, as figure 7.3 suggests it may need to be.

(iii) It was noted in (i) above that medical experts use both the colour and appearance of clusters of pixels to determine whether they belong to the image of a stained parasite. When determining the severity of infection they further check that the stained object is properly within a RBC in order to decide that a cell is infected[5]. Since medical researchers frequently count the number of infected RBCs in the course of their work – a task much less laborious than counting the number of RBCs – there are opportunities, especially if some simple software tools were available and appropriate protocols established, to create a large database of thin-film slide images in which infected RBCs have been labelled by experts.

(iv) For example, one of the standard methods of quantifying the severity of infection is the number of parasites per unit volume of blood (*e.g.*, usually expressed as thousands per μl). Usually, this relies on estimating the blood volume from finding a certain number, typically a few hundred, WBCs which is then converted via a standard of 8000 WBCs per μl to give an estimate of the density of infected RBCs. Since WBCs comprise only $\sim 1\%$ of the blood volume and are ~ 50 to 100 times rarer than RBCs [199], observation of a few hundred WBCs requires study of a similar number of images taken from many different fields of a thin-film slide. Using the several hundred fields, or tiles as they are sometimes known, required would typically lead to identification of a few 100 to ~ 1000 infected RBCs from the slide. Repeated over many slides, this should enable large databases of malaria infected RBCs to be built up, both from laboratory work where the type and stage of parasite life-cycle could be part of the database records and perhaps also from field studies where typically there may be within one sample from an individual patient parasites at different stages of the life-cycle.

(v) Moreover, if desired and perhaps with the aid of techniques such as those devel-

[5] Recall *e.g.* from the background section 1.2 that the degree of infection or parasitemia is defined from the number of number of parasites in asexual stage malaria when the parasite infection is within a RBC. This is consistent with the remarks in section 6.4.1.

oped in this thesis to isolate the RBCs, labelling could be carried through to the individual pixel level with clusters of pixels belonging to the stained parasites within the infected cells identified by technicians or 'semi-experts', or perhaps even 'crowd-sourced' [116, 124, 122]. Such data, capturing the skill of medical staff in using the appearance and location of stained parasite infections within RBCs to identify cells that are infected would, in particular, embody the spatial relationships between pixels representing each stained parasite. It would thus be ideal for training supervised systems either to segment parasite pixels, detect clusters of stained parasite pixels within a RBC, or to classify cells directly as infected or healthy. As noted several times within this thesis and in [23] provision of a large database of standardised, labelled data would be a big step forward in research on the automatic processing of malarial infected blood films.

It is notable that Mavandadi *et al.* [124, 123] consider how the decisions of individual gamers should be combined by MAP (maximum *a posterior*) estimation or an *EM* algorithm. The same decision logic may of course be applied to a group of experts (who are the gamers in some versions) and also to the combination of computational decisions as described earlier by Warfield *et al.* [193]. Details of the mathematics are given in the supplementary material to [124] and, in particular in [123] which describes the optimal combination of decisions in three categories: healthy, infected and uncertain. The former also describes the local colour peak histogram features and adaptive boosting algorithm used in their computational system whilst the latter [123] also describes a probabilistic framework for proceeding from decisions on individual RBCs to a slide or patient level diagnosis. This fills a gap left by [151] (section 2.5.2). The paper therefore makes two important contributions.

7.3 Epilogue

We finish with an update on some recent papers on computational analysis of microscope images of malaria blood smears together with some closing remarks on the distinctiveness of the research described in this thesis. Also mentioned are some modern alternatives that might have been utilised instead of the Fisher discriminant.

Recent important papers Several important papers [106, 144, 62, 145, 140, 192, 109] have recently been published in leading journals. Several of these utilised available image analysis programmes such as ImageJ [150] ([62]), Matlab ([145, 109] – the latter as an annotation tool [187] or bespoke commercial scanning hardware and processing software [140]. By 2009 the first three of the above, [106, 144] for the analysis of thin-film images and [62] for thick-films,

all reported semi-automated systems capable of accurately quantifying the parasitemia with correlation coefficients $R \geq 0.96$ in comparison to results obtained by medical experts. According to Frean [61] in 2010 "trained technologists readily recognise and count erythrocytes (the numerator) in an adequately stained thin smear, (but) estimating the denominator (total number of erythrocytes) of the required fraction is the source of most error in estimating the parasitemia". The tedium of manual total RBC counts indicates that this is probably still true and thus the importance of automating total RBC counts. Three of the above subsequent publications have continued to focus on thin-films [145, 192, 109] whilst in the other [140] the processing of both thin and thick-films is considered. With the exception of the two papers by Frean, all of the above are multi-author papers indicative of the large, multidisciplinary, collaborative teams needed to address these problems to the standards set by the WHO and to collect process and, where required for evaluation, annotate many hundreds of images. Four points may be noted from these papers:

(i) Detection and identification of malaria infection seem to be more prominent than its quantification with the reliable estimate of parasitemia particularly difficult. For example: provisos such as counting only well-defined singlet and well-rounded RBCs enabled [109] to obtain accurate results with again $R = 0.97$ in comparison with medical experts, whilst [140] failed to reach the lowest level of WHO 'expert' reader standard with a somewhat lower $R = 0.91$. Purwar *et al.* [145] using an active contour algorithm [25] to detect RBC boundaries obtained counts of better than 1% accuracy per image. Unfortunately their small dataset prevented them from counting more than a few infected RBCs so it is not possible to draw reliable, quantitative conclusions from their counts of infected RBCs. It is interesting however to note that they are critical of the large training sets required by supervised machine learning techniques and use an adaptation of K-means clustering for the detection of stained, parasite pixels in which clusters are probabilistically re-weighted. Walliander *et al.* [192] obtained RBC counts with an overall accuracy of 0.21% by considering three categories of RBCs – 'RoundCells' and less-round 'FoundCells' detected via a Hough transform and misshapen cells called 'ApproxCells' which were counted statistically from comparison of their areas with the average area of cells in the two former categories.

(ii) Thresholding is frequently used in the early stages of the image analysis in many of the above especially for segmentation of RBCs: *e.g.* [106, 140, 192, 109]. Variants of an old thresholding algorithm due to Zack [201] which utilises a geometric construction similar

to the 'scree-graph' discussed in section 3.2.4 are often used as described by [106, 192] and [109]. It is perforce a one-dimensional algorithm applied to the image intensity, to one of the colour channels (*e.g.* the green channel), or to a difference of the blue and green channels [106]. For automatic detection of parasite infections comparison[6] with a database of 'parameters' [140] and supervised, machine learning techniques utilising support vector machines operating on a range of local image features [109] are used, in accord with our remarks in section 7.2.2 above – though not of course with the critique of such supervised methods by [145] and with what we ourselves had hoped to achieve.

(iii) Several of the counting methods employed have utilised size distribution histograms or statistical properties of image object size distributions (see *e.g.* [106, 145, 140, 192, 109]) but none has integrated the kinds of constraints on the relationship between the sizes of singlet and doublet, triplet or larger clusters of adjoining RBCs in the way we have. It is notable that [106] found it necessary to deal with clusters of large numbers of adjoining RBCs (sometimes 10 or more) and strongly advocated the analysis of whole images rather than a grid of RBCs excised from the microscope image in some manner as in many of the other papers.

(iv) Most of these works rely on trained technicians to acquire digital thin-film slide images using conventional microscopes. Prescott *et al.* [140] describe the use of two types of automatic scanner in the bespoke (WHT – World Health Technology) system but, even then, with an operator required to identify autofocus points. It thus seems that acquisition of digital microscope slide images is a potential bottle-neck especially at a cost and utilising equipment suitable for use in the tropics. It is somewhat surprising not to have encountered research on such matters but this may be because it is reported elsewhere, for example: in the literatures on the development of so-called lensless systems, micro-technology lab-on-chip systems, and of mobile phone applications.

Malarial image analysis papers As for papers in the image processing/computer vision literature, there are many, for example the 112 articles reviewed by Das *et al* [40] whose production, including the papers in leading medical journals cited above, seems to have peaked in 2012 – 2014. There is little to add to this review which includes eleven tabulations ranging from the characteristics of different types of malaria parasite (cf. the discussion in Tek's thesis [175]) to the performance of various parasite count methodologies, the latter showing that the results

[6] No more detailed description of this bespoke approach by a commercial organisation, WHT – World Health Technologies, is given.

obtained by Sio *et al.* in 2006-7 with their MalariaCount system [165] are still competitive. The review also shows that most of this recent work has been on the computational analysis of thin-film images, often for diagnosis of malaria, with supervised machine learning techniques mainly used to identify parasite infections. Unfortunately the paper by Mavandadi *et al* [123] with its important contributions to decision fusion and patient level diagnosis seems to be one of the few that were overlooked.

Alternative approaches It is also worth noting that the unsupervised approach adopted in this thesis could have been based on alternative criteria, in particular on the use of image entropies, which like the Fisher discriminant are also a statistical property of the distribution of pixel attributes. This includes the non-additive[7] entropies introduced by Tsallis which have become popular in some areas of medical image analysis [41, 138] and registration [2]. As for the Fisher discriminant in one-dimension, entropy criteria may be used for multi-thresholding though care must be taken efficiently to find a good enough 'global' optimum (see for example [158]). It seems that the choice of criterion, including selection of the q parameter in the Tsallis entropies to reflect the degree of short-range order in an image [41], should be based on that which works best for an application and, as noted in section 5.2, optimisation of Fisher discriminants seems eminently well-suited to our problem – at least for RBC segmentation!

Finally, we note that it may not be necessary to choose one particular criterion, or even having chosen a criterion, to select to which feature or features it should be applied. We saw in section 4.2.1 that one could exhaustively search for the combination of Otsu thresholds on each of the individual colour channels which optimised the sum of the between-class variances across the three channels. With a multi-level, hierarchical system or some other kind of more complicated architecture, the number of possible combinations would be so great as to preclude an exhaustive search. The evolutionary technique developed by Oechsle [128] could be useful in such cases and it would then seem that the approach could be adapted to produce a system optimised, or perhaps even *Pareto* optimised (section 2.5.1), over combinations of many criteria. Better still, if sufficient training data were available, it might be possible to evolve a system for the analysis of microscope images of malarial blood smears, ultimately optimised for counting cells, identifying healthy and infected cells, estimating parasitemia or whatever end result is required for a particular application.

[7] But *not* non-extensive [184].

Appendix A

Derivation from Fisher's discriminant

We wish to maximize the between-class covariance, σ_B^2, but this time with respect to each of the $C-1$ thresholds T_J , $J = 1 \cdots C-1$ [recall $T_0 = L$ and $T_C = U$ are fixed], *i.e.* we want to set:

$$\frac{\partial \sigma_B^2}{\partial T_J} = 0 \qquad J = 1 \cdots C-1. \tag{A.1}$$

$$\text{Since} \quad \sigma_B^2 = \sum_{I=1}^{C} Z(I)(\mu(I) - \mu)^2 \tag{A.2}$$

$$\Rightarrow \frac{\partial \sigma_B^2}{\partial T_J} = \sum_{I=1}^{C} \frac{\partial Z(I)}{\partial T_J}(\mu(I) - \mu)^2 + 2\sum_{I=1}^{C} Z(I)(\mu(I) - \mu)\frac{\partial \mu(I)}{\partial T_J}. \tag{A.3}$$

We first evaluate using 4.16:

$$\frac{\partial Z(I)}{\partial T_J} = \frac{\partial}{\partial T_J}\int dx\; z(x|I) = \frac{\partial}{\partial T_J}\int dx[\vartheta(x - T_{I-1}) - \vartheta(x - T_I)]p(x)$$

$$\Rightarrow \frac{\partial Z(I)}{\partial T_J} = -p(T_{I-1})\delta_{I-1,J} + p(T_I)\delta_{I,J} \tag{A.4}$$

which remains true $\forall I = 1 \cdots C$ and $\forall J = 1 \cdots C-1$.

In the 2-class case with $T_1 = T$, A.5 leads as required to equations 4.17.

We note that A.5 may be written as:

$$\frac{\partial Z(I)}{\partial T_J} = p(T_J)[\delta_{I,J} - \delta_{I-1,J}]. \tag{A.5}$$

We also have to evaluate

$$\begin{aligned}
\frac{\partial \mu(I)}{\partial T_J} &= \frac{1}{Z(I)}\int dx \frac{\partial z(x|I)x}{\partial T_J} - \frac{1}{Z(I)^2}\frac{\partial Z(I)}{\partial T_J}\int dx\; z(x|I)x \\
&= \frac{1}{Z(I)}[-T_J p(T_J)\delta_{I-1,J} + T_J p(T_J)\delta_{I,J}] - \frac{1}{Z(I)}[-p(T_J)\delta_{I-1,J} + p(T_J)\delta_{I,J}]\mu(I) \\
&= \frac{p(T_J)}{Z(I)}T_J[\delta_{I,J} - \delta_{I-1,J}] - \frac{p(T_J)\mu(I)}{Z(I)}[\delta_{I,J} - \delta_{I-1,J}]
\end{aligned}$$

$$\Rightarrow \frac{\partial \mu(I)}{\partial T_J} = \frac{p(T_J)}{Z(I)}[T_J - \mu(I)][\delta_{I,J} - \delta_{I-1,J}] \tag{A.6}$$

which also remains true $\forall I = 1 \cdots C$, $J = 1 \cdots C - 1$.

Again, we note that in the 2-class case with $T_1 = T$ A.6 leads to equations 4.18.

Thus, on substituting A.5 and A.6 in A.3 we obtain:

$$\begin{aligned}
\frac{\partial \sigma_B^2}{\partial T_J} &= \sum_{I=1}^{C} p(T_J)[\delta_{I,J} - \delta_{I-1,J}][\mu(I) - \mu]^2 \\
&\quad + 2\sum_{I=1}^{C} Z(I)[\mu(I) - \mu]\frac{p(T_J)}{Z(I)}[T_J - \mu(I)][\delta_{I,J} - \delta_{I-1,J}] \\
&= p(T_J)\{[\mu(J) - \mu]^2 - [\mu(J+1) - \mu]^2\} + 2p(T_J)\{[\mu(J) - \mu][T_J - \mu(J)] \\
&- [\mu(J+1) - \mu][T_J - \mu(J+1)]\} \\
&= p(T_J)[\mu(J)^2 - 2\mu\mu(J) - \mu(J+1)^2 + 2\mu\mu(J+1) + 2\mu(J)T_J - 2\mu T_J - 2\mu(J)^2 \\
&\quad + 2\mu\mu(J) - 2\mu(J+1)T_J + 2\mu T_J + 2\mu(J+1)^2 - 2\mu\mu(J+1)] \\
&= p(T_J)[\mu(J+1)^2 - \mu(J)^2 + 2\mu(J)T_J - 2\mu(J+1)T_J] \\
&\Rightarrow \frac{\partial \sigma_B^2}{\partial T_J} = p(T_J)[\mu(J) - \mu(J+1)][2T_J - (\mu(J) + \mu(J+1))]. \qquad \text{(A.7)}
\end{aligned}$$

In the 2-class case with $T_1 = T$ A.7 immediately reduces to 4.19:

$$\frac{\partial \sigma_B^2}{\partial T} = p(T)[\mu(1) - \mu(2)][2T - (\mu(1) + \mu(2))]. \qquad \text{(A.8)}$$

A.7 holds $\forall J = 1 \cdots C - 1$.

We finally note that, if required it is straightforward similarly to calculate the second derivative from A.7.

Appendix B

Multi-threshold segmentation

On the assumption that $\mu(J) \neq \mu(J+1)$ for any two adjacent classes, A.7 implies that the J^{th} threshold T_J should be chosen either such that

$$\begin{aligned} & p(T_J) = 0 \\ \text{OR} \quad & T_J = \frac{1}{2}[\mu(J) + \mu(J+1)]. \end{aligned} \tag{B.1}$$

If there are no clear breaks in the histogram we have to iterate each of the thresholds according to

$$T_J = \frac{1}{2}[\mu(J) + \mu(J+1)] \quad \forall J = 1 \cdots C-1 \tag{B.2}$$

or find a solution by means of some optimisation or search procedure.

The important point to note is that each class conditional mean

$$\mu(J) = \frac{1}{Z(J)} \int dx \; z(x|J) \; x = \frac{1}{Z(J)} \int dx \; x \; p(x)[\vartheta(x - T_{J-1}) - \vartheta(x - T_J)] \tag{B.3}$$

depends on T_J and on T_{J-1}.

Several possibilities are now open to us:

(I) Solve the set of $(C-1)$ equations B.2 by a generalisation of Otsu's algorithm. The main issue in this case is initialisation within a basin of attraction that converges to the desired optimal solution.

(II) Solve the set of equations B.2 by employing standard means such as a multi-dimensional Newton-Raphson or Levenberg-Marquardt method.

(III) Solve the set of equations B.2 by a continuation method in which one starts with an easy problem with (say) C well separated classes when $\lambda = 0$ with distribution $p_e(x)$ and an obvious solution set $T_J(0)$ and tracks the thresholds $T_J(\lambda)$ which segment

$$p_\lambda(x) = (1-\lambda)p_e(x) + \lambda p(x) \tag{B.4}$$

to the solutions $T_J = T_J(1)$ of the desired problem with $p_1(x) = p(x)$.

(IV) Directly solve the optimisation problem to

$$\max_{T_J}\{\sigma_B^2\} \tag{B.5}$$

using any convenient, sufficiently powerful optimisation procedure. Since the derivatives $\frac{\partial\sigma_B^2}{\partial T_J}$ are available and by further differentiation of A.7 the second derivatives, the following possibilities spring to mind:

(i) simple gradient ascent (with or without a momentum / memory term *etc.*)

(ii) a conjugate gradient method

(iii) a multi-dimensional Newton-Raphson method (*cf.* (II) above)

(iv) a Levenberg-Marquardt method (again *cf.* (II) above)

(v) any other convenient method described in the literature (*e.g.* in [141] or available in Matlab or (say) Mathematica – *e.g.* trust region methods).

In addition methods which do not require derivatives could be used, including:

(vi) a simplex method

(vii) stochastic methods such as: stochastic gradient, simulated annealing, genetic algorithms, and the more modern differential evolution and SOMA (self organizing migrating algorithm).

The stochastic methods may be important as it is likely that there are multiple local minima of σ_B^2 when there are multiple thresholds [202]).

It is also worth noting that the solution set is constrained, *i.e.* by virtue of the segmentation model:

$$L = T_0 < T_J < T_C = U \tag{B.6}$$

$$\text{and in general} \quad T_{J-1} < T_J < T_{J+1} \quad J = 1 \cdots C-1 \tag{B.7}$$

which includes B.6 as a special case with fixed $T_0 = L$ and $T_C = U$. Note that B.7 also excludes combinatorial replication of otherwise identical solutions which differ only in the choice of class labels.

It is also important to note that, in particular, the derivative based algorithms (gradient, Newton-Raphson, *etc.*) could easily violate the constraints B.7. Imposition of B.7 as a feasible solution set at each iteration might therefore be necessary. The constraints should also help any search method by reducing the solution search space.

(V) Finally, we note that B.2 and B.3 are reminiscent of an EM algorithm. In fact, the standard way of solving the classification by use of a linear mixture model would be via an EM algorithm. There is a close similarity between a linear mixture model in which we would let (say):

$$p(x) = \sum_{I=1}^{C} p(I)p(x|I) \tag{B.8}$$

and the segmentation model in which we say:

$$p(x) = \sum_{I=1}^{C} z(x|I) = \sum_{I=1}^{C} Z(I) \left(\frac{z(x|I)}{Z(I)} \right) \tag{B.9}$$

with, as already noted $z(x|I)/Z(I)$ having the properties of a probability density. Apart from the fact that the classes do not overlap in the segmentation model, the major difference is that parametric models of the class-conditional distributions (*e.g.* Gaussians) are usually used in the linear mixture model B.8 whereas the segmentation model B.9 is non-parametric.

Appendix C

Optimisation of Fisher's discriminants

Using the summation convention in the notation we have adopted whereby components in the original feature space are labelled by suffices whilst components along the directions of the feature vectors generated are labelled in brackets, differentiation of D_2 (4.33)

$$D_2 = \frac{tr(A^T S_B A)}{tr(A^T S_W A)} = \frac{a_i(k) S_{Bij} a_j(k)}{a_{i'}(k') S_{Wi'j'} a_{j'}(k')} \tag{C.1}$$

is straightforward:

$$\frac{\partial D_2}{\partial a_l(m)} = \frac{1}{[tr(A^T S_W A)]^2} \left\{ \begin{array}{l} tr(A^T S_W A)[\delta_{il}\delta(km) S_{Bij} a_j(k) + a_i(k) S_{Bij} \delta_{jl} \delta(km)] \\ - tr(A^T S_B A)[\delta_{il}\delta(km) S_{Wij} a_j(k) + a_i(k) S_{Bij} \delta_{jl} \delta(km)] \end{array} \right\}. \tag{C.2}$$

Since the covariance matrices are symmetric, setting the derivatives to zero leads to:

$$tr(A^T S_B A) S_{Blj} a_j(m) = tr(A^T S_W A) S_{Wlj} a_j(m)) \tag{C.3}$$

which has the form of a generalised eigen-equation:

$$S_B \underline{a}(m) = D_2(m) S_W \underline{a}(m) \tag{C.4}$$

in which the eigensolution chosen determines the value of D_2.

To differentiate D_3 (4.34) it is helpful to introduce $r_B \times r_B$ symmetric matrices:

$$P = A^T S_B A, \quad Q = A^T S_W A \tag{C.5}$$

so that

$$\begin{aligned} \frac{\partial D_3}{\partial a_l(m)} &= \frac{\partial}{\partial a_l(m)} \left\{ Q^{-1}(uv) P(vu) \right\} \\ &= \frac{\partial Q^{-1}(uv)}{\partial a_l(m)} P(vu) + Q^{-1}(uv) \frac{\partial P(vu)}{\partial a_l(m)}. \end{aligned} \tag{C.6}$$

The second term is straightforward to evaluate and, since the matrices are symmetric, yields

$$Q^{-1}(uv) \frac{\partial P(vu)}{\partial a_l(m)} = 2 S_{Blj} a_j(u) Q^{-1}(um) \tag{C.7}$$

which may be written in matrix form as $2S_B AQ^{-1}$.

The first term requires us to evaluate the derivative of the inverse matrix Q^{-1}:

$$\frac{\partial Q^{-1}}{\partial a_l(m)} = -Q^{-1}\frac{\partial Q}{\partial a_l(m)}Q^{-1} \tag{C.8}$$

which, since all the matrices are symmetric leads to:

$$\frac{\partial Q^{-1}(uv)}{\partial a_l(m)}P(vu) = -2S_{Wlj}a_j(s)Q^{-1}(sv)P(vu)Q^{-1}(um) \tag{C.9}$$

which in turn may be written in matrix form as $-2S_W AQ^{-1}PQ^{-1}$.

Setting the derivatives to zero thus leads to:

$$S_B AQ^{-1} = S_W AQ^{-1}PQ^{-1} \tag{C.10}$$

as given by [43] section 6.6.3, page 252.

Finally, we note that if we have the eigen-equation 4.43, $S_B A = S_W A\Lambda$, it follows that $P = Q\Lambda$ and thus C.10 is satisfied.

As noted in the text, differentiation of D_4 (4.35) requires the trick (4.45) for differentiating the determinant of a matrix with respect to its elements. Utilising the matrix P introduced in (C.5) it is straightforward to show that

$$\frac{\partial |P|}{\partial a_l(m)} = \left[\delta_{il}\delta(um)S_{Bij}a_j(v) + a_i(u)S_{Bij}\delta_{jl}\delta(vm)\right]|P|P^{-1}(uv) \tag{C.11}$$

and thus, since the covariance matrix is symmetric that, in matrix form:

$$\frac{\partial |P|}{\partial A} = 2|P|S_B AP^{-1}. \tag{C.12}$$

Proceeding in a similar manner using the other matrix Q introduced in (C.5) it follows that:

$$\frac{\partial |Q|^{-1}}{\partial A} = -\frac{2|Q|}{|Q|^2}QS_W AQ^{-1}. \tag{C.13}$$

At the optimal D_4 we thus see that:

$$S_B AP^{-1} = S_W AQ^{-1} \tag{C.14}$$

which again is satisfied if we have the eigen-equation 4.43, $S_B A = S_W A\Lambda$.

Appendix D

Absolute trace criteria

Three absolute trace criteria may be constructed from the three familiar covariance matrices, S_T, S_B and S_W:

$$\left.\begin{aligned} D_1 \equiv D_T &= tr(A^T S_T A) \\ D_B &= tr(A^T S_B A) \\ D_W &= tr(A^T S_W A) \end{aligned}\right\} \tag{D.1}$$

where we have noted that D_T is the same as criterion D_1 mentioned in section 4.2.2. Each of these criteria is unbounded and must be used with some constraint or constraints on the feature matrix A. We shall consider a single normalisation constraint expressed in the general form:

$$\underline{a}^T(k) W \underline{a}(k) = \alpha(k) \tag{D.2}$$

with a weight matrix W and normalisations $\alpha(k)$ to be specified.

The criteria are each of the form $tr(A^T S A)$ and optimisation leads to generalised eigen-equations of the form:

$$SA = WA\Xi. \tag{D.3}$$

The obvious choices are for W to be the identity matrix I to introduce a conventional Euclidean distance constraint or one of the covariance matrices other than the matrix S itself appearing in D.3 to constrain a within-class distance ($W = S_W$), a between-class distance ($W = S_B$), or a sample data distance ($W = S_T$).

The unit matrix normalisation with the $\alpha(k) = 1$ and the total covariance matrix leads to PCA whilst the six non-trivial cases in which W is chosen to be a covariance matrix all lead to LDA and may be summarised as in table D.1 with

$$D(opt) = \sum_{k=1}^{r_B} \alpha(h)\xi(k). \tag{D.4}$$

S	S_B	S_T	S_W
W	S_W , S_T	S_B , S_W	S_T , S_B
$\xi(k)$	$\lambda(k)$, $\frac{\lambda(k)}{1+\lambda(k)}$	$\frac{\lambda(k)+1}{\lambda(k)}$, $\lambda(k)+1$	$\frac{1}{\lambda(k)+1}$, $\frac{1}{\lambda(k)}$

Table D.1: Outcomes for the non-trivial choices of using a covariance matrix as the weight matrix W given the choice of covariance matrix S in D.3.

Appendix E

Recursive Calculations

It is well known in statistics that means and variances may be calculated recursively, in particular when data is gathered sequentially over time (see for example [65]). A similar approach may be used when the Fisher discriminant (or equivalently when the feature space is one-dimensional, either of the between-class or within-class variances) is being opitimised.

Recall that, in our probabilistic development of the Otsu algorithm we introduced the indicator densities which for the two-class case were;

$$z(x|1) = \begin{cases} p(x) & \text{for} \quad T \leq x \\ 0 & \text{for} \quad T \geq x \end{cases} \qquad z(x|2) = \begin{cases} 0 & \text{for} \quad T \leq x \\ p(x) & \text{for} \quad T \geq x \end{cases}$$

so that

$$\mu(1) = \frac{1}{Z(1)} \int_{-\infty}^{T} dx p(x) x \qquad \mu(2) = \frac{1}{Z(2)} \int_{T}^{\infty} dx p(x) x \tag{E.1}$$

where

$$Z(1) = \int dx z(x|1) = \int_{-\infty}^{T} dx p(x) \quad Z(2) = \int dx z(x|2) = \int_{T}^{\infty} dx p(x). \tag{E.2}$$

Class-conditioned variances $\sigma^2(1)$ and $\sigma^2(2)$ may be similarly defined but are not required here since $\max_T \left\{\sigma_B^2\right\}$ suffices and the between-class variance depends only on the class-conditioned means (recall equation 4.11 and, in particular in the two-class case, 4.12). If we have a discrete, finite dataset with feature values $x(i)$ over the range $[L, U]$ characterised by a histogram $h(x)$, the first of equations E.2 would, for example, be replaced by:

$$Z(1|T) = \sum_{x=L}^{T} h(x) = h(L) + \cdots + h(T) \tag{E.3}$$

where, since we will be systematically changing the value of the threshold, it is convenient to make the dependence of $Z(1)$ on T explicit. Analogous to the second of equations E.2, we then also have:

$$Z(2|T) = \sum_{x=T+\Delta T(T)}^{U} h(x) = h(T + \Delta T(T)) + \cdots + h(U) \tag{E.4}$$

where $\Delta T(T)$ is the increment in x from $x = T$ to the next non-empty histogram bin $h(T + \Delta T(T))$ after $h(T)$.

It follows from E.3 and E.4 that, if the threshold is incremented:

$$\begin{aligned} Z(1|T + \Delta T(T)) &= Z(1|T) + h(T + \Delta T(T)) \\ Z(2|T + \Delta T(T)) &= Z(2|T) - h(T + \Delta T(T)). \end{aligned} \tag{E.5}$$

Similarly, in the discrete case using the same notation as above to indicate explicitly that the threshold is T:

$$\begin{aligned} \mu(1|T) &= \frac{1}{Z(1|T)} \sum_{x=L}^{T} xh(x) \\ \mu(2|T) &= \frac{1}{Z(2|T)} \sum_{x=T+\Delta T(T)}^{U} xh(x). \end{aligned} \tag{E.6}$$

and:

$$\sigma_B^2(T) = Z(1|T)Z(2|T)[\mu(1|T) - \mu(2|T)]^2. \tag{E.7}$$

It can be seen from equations E.6 and E.7 that the normalisation of the histogram, which would usually be either to 1 or to n, the number of samples, is immaterial.

It follows from equations E.6 that:

$$\begin{aligned} \mu(1|T + \Delta T(T)) &= \frac{Z(1|T)\mu(1|T) + (T + \Delta T(T))h(T + \Delta T(T))}{Z(1|T + \Delta T(T))} \\ \mu(2|T + \Delta T(T)) &= \frac{Z(2|T)\mu(2|T) - (T + \Delta T(T))h(T + \Delta T(T))}{Z(2|T + \Delta T(T))}. \end{aligned} \tag{E.8}$$

Similar equations result if the threshold T were decremented, but with the $+$ and $-$ signs in the numerators on the right hand sides interchanged.

Equations E.8 and E.5 are generalisations to variable $\Delta T(T)$ of familiar equations for the recursive computation of the mean of a data sequence $x(i), i = 1 \ldots n$ and, in the present specific context of equations (3.32) and (3.31) respectively in [164] (Chapter 3, section 3.8, page 88). When the feature x is the image intensity, the increment ΔT would usually be equal to one and therefore constant. Although gaps in the intensity histogram are rare they may occur and, indeed, as we have seen are one of the classic conditions for choice of a threshold to segment completely separated classes. More importantly, we have seen in section 6.1 that gaps in image object area histograms are much likely. This motivated the development of the more general recursive equations in this appendix.

In order to carry out the above recursive calculations, they must be initialised. Inspection

of E.3 and E.4 reveals that, if the histogram is normalised to n:

$$\begin{aligned} Z(1|L) &= h(L) & Z(2|L) &= n - h(L), \\ Z(1|U) &= n & Z(2|U) &= 0. \end{aligned} \tag{E.9}$$

These equations satisfy the requirement that;

$$Z(1) + Z(2) = \sum_{x=L}^{U} h(x) = n \tag{E.10}$$

which is true whatever the value of T as may be readily verified from E.4.

With the same normalisation of the histogram to n it follows from E.6that:

$$\begin{aligned} \mu(1|L) &= L & \mu(2|L) &= \frac{n\mu - Lh(L)}{n - h(L)} \\ \mu(1|U) &= \mu \qquad \text{and that} & \mu(2|U) &\text{ is indeterminate.} \end{aligned} \tag{E.11}$$

where, as usual μ is the mean. The above satisfy a constraint similar to E.10:

$$Z(1)\mu(1) + Z(2)\mu(2) = \sum_{x=L}^{U} xh(x) = n\mu, \tag{E.12}$$

which is again true for any T – as may be verified from E.6.

According to the indeterminacy of E.12 we may set $\mu(2|U) = 0$ but the initialisations of $Z(2|L)$ and $\mu(2|L)$ are rather inelegant and equations E.9 – E.12 lack symmetry. The simplicity and symmetry may be restored by setting:

$$\begin{aligned} Z(1|L - \Delta L) &= 0 & Z(2|L - \Delta L) &= n \\ Z(1|U) &= n & Z(2|U) &= 0 \\ \mu(1|L - \Delta L) &= 0 & \mu(2|L - \Delta L) &= \mu \\ \mu(1|U) &= \mu & \mu(2|U) &= 0. \end{aligned} \tag{E.13}$$

and carrying out the recursive calculations with T incremented from an initial hypothetical $L - \Delta L$. Alternatively, similar recursive relationships to E.5 and E.8 (with appropriate sign changes) may be derived were the threshold T decremented and calculations initialised from $T = U$ as in E.13 above.

Bibliography

[1] T. Abe, M. Yamaguchi, Y. Murakami, N. Ohyama, and Y. Yagi. Color correction of pathological images for different staining-condition slides. In *Enterprise Networking and Computing in Healthcare Industry*, pages 218–223, 2004.

[2] Henrique Amaral-Silva, Lauro Wichert-Ana, Luiz Otavio Murta, Larissa Romualdo-Suzuki, Emerson Itikawa, Geraldo Filho Bussato, and Paulo Azevedo-Marques. The superiority of Tsallis entropy over traditional cost functions for brain MRI and SPECT registration. *Entropy*, 16(3):1632–1651, 2014.

[3] M. Anastasio, M. Kupinski, and R. Nishikawa. Optimization and froc analysis of rule-based detection schemes using a multiobjective approach. *IEEE Transactions on Medical Imaging*, 17(6):1089–93, 1998.

[4] Annalisa Barla, Francesca Odone, and Alessandro Verri. Old fashioned state-of-the-art image classification. In *Proceedings of 12th International Conference on Image Analysis and Processing*, pages 566–571. IEEE Computer Society, September 2003.

[5] K. Barnard, V. Cardei, and B. Funt. A comparison of computational color constancy algorithms. I: Methodology and experiments with synthesized data. *IEEE Transactions on Image Processing*, 11(9):972–984, September 2002.

[6] Kobus Barnard, Lindsay Martin, Adam Coath, and Brian Funt. A comparison of computational color constancy algorithms - part ii: Experiments with image data. *IEEE Transactions on Image Processing*, 11(9):985–996, 2002.

[7] R. Beale and T. Jackson. *Neural Computing: An Introduction*. IoP Publishing Ltd., Adam Hilger, 1990.

[8] P. N. Belhumeur, J. Hespanha, and D. J. Kriegman. Eigenfaces versus Fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720, July 1997.

[9] David Bellot and Pierre Bessière. Approximate discrete probability distribution representation using a multi-resolution binary tree. *CoRR*, abs/cs/0605055, 2006.

[10] J. Ross Beveridge, Joey Griffith, Ralf R. Kohler, Allen R. Hanson, and Edward M. Riseman. Segmenting images using localized histograms and region merging. *International Journal of Computer Vision*, 2(3):311–347, 1989.

[11] C. M. Bishop. Variational principal components. In *Ninth International Conference on Artificial Neural Networks*, volume 1, pages 509–514, 1999.

[12] C. M. Bishop and J. Winn. Non-linear Bayesian image modelling. In *Proceedings Sixth European Conference on Computer Vision*, volume 1, pages 3–17. Springer, 2000.

[13] Christopher M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, Inc., New York, NY, USA, 1995.

[14] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

[15] Andrew Blake, Pushmeet Kohli, and Carsten Rother. *Markov Random Fields for Vision and Image Processing*. The MIT Press, 2011.

[16] R. Brunelli and T. Poggio. Face recognition: features versus templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(10):1042–1052, 1993.

[17] R. Burbidge, M. Trotter, B. Buxton, and S. Holden. Drug design by machine learning: support vector machines for pharmaceutical data analysis. *Computers & Chemistry*, 26(1):5–14, 2001.

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

[19] B. F. Buxton. Image processing course notes.

[20] B. F. Buxton. Image modelling. Lecture delivered at the British Machine Vision Association Summer School on Computer Vision, 2006.

[21] B. F. Buxton. Private communication. 2014.

[22] Bernard Buxton and Vasileios Zografos. Flexible template and model matching using intensity. In *Proceedings of Digital Image Computing: Techniques and Applications*, page 64. IEEE Computer Society, December 2005. `http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1587666`.

[23] Bernard F. Buxton, Houari Abdallahi, Delmiro Fernandez-Reyes, and William Jarra. Development of an extension of the Otsu algorithm for multidimensional image segmentation of thin-film blood slides. In *International Conference on Computing: Theory and Applications*, pages 552–562, Los Alamitos, CA, USA, 2007. IEEE Computer Society.

[24] J. F. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698, 1986.

[25] T. F. Chan and L. A. Vese. Active contours without edges. *IEEE Transactions on Image Processing*, 10(2):266–277, Feb 2001.

[26] R. Chellappa, P. Sinha, and P. J. Phillips. Face recognition by computers and humans. *IEEE Computer*, 43(2):46–55, 2010.

[27] Paul B. Chou and Christopher M. Brown. The theory and practice of Bayesian image labeling. *International Journal of Computer Vision*, 4(3):185–210, 1990.

[28] Dorin Comaniciu and Peter Meer. Robust analysis of feature spaces: Color image segmentation. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 750–755, June 1997.

[29] T. F. Cootes. Eigenfaces and eigenpatches. PowerPoint presentation retrieved from personal webpages 2003.

[30] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):681–685, 2001.

[31] T. F. Cootes, G. J. Page, C. B. Jackson, and C. J. Taylor. Statistical grey-level models for object location and identification. In D. Pycock, editor, *Proceedings of 6th British Machine Vision Conference*, pages 533–542. BMVA Press, September 1995.

[32] T. F. Cootes and C. J. Taylor. Statistical models of appearance for computer vision. `http://www.isbe.man.ac.uk/~bim/Models/app_models.pdf`.

[33] T. F. Cootes and C. J. Taylor. Modelling object appearance using the grey-level surface. In Edwin R. Hancock, editor, *Proceedings of the 5th British Machine Vision Conference*, pages 479–488. BMVA Press, September 1994.

[34] T. F. Cootes and C. J. Taylor. Statistical models of appearance for computer vision. Technical report, University of Manchester, 2004.

[35] T. F. Cootes, G. V. Wheeler, K. N. Walker, and C. J. Taylor. View-based active appearance models. *Image and Vision Computing*, 20:657–664, 2002.

[36] D. P. A. Corney, B. F. Buxton, W. B. Langdon, and D. T. Jones. Biorat: Extracting biological information from full-length papers. *Bioinformatics*, 20(17):3206–3213, 2004.

[37] Microsoft Corporation. Microsoft encarta encyclopedia, 2002.

[38] N. P. Costen, T. F. Cootes, G. J. Edwards, and C. J. Taylor. Automatic extraction of the face identity-subspace. *Image and Vision Computing*, 20:319–329, 2002.

[39] Nello Cristianini and John Shawe-Taylor. *An introduction to support vector machines and other kernel-based learning methods*. 2000.

[40] D. K. Das, R. Mukherjee, and C. Chakraborty. Computational microscopic imaging for malaria parasite detection: a systematic review. *Journal of Microscopy*, 260(1):1–19, 2015.

[41] M. Portes de Albuquerque, I. A. Esquef, A. R. Gesualdi Mello, and M. Portes de Albuquerque. Image thresholding using Tsallis entropy. *Pattern Recognition Letters*, 25(9):1059–1065, 2004.

[42] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39(1):1–38, 1977.

[43] Pierre A. Devijver and Josef Kittler. *Pattern Recognition: A Statistical Approach*. Prentice Hall International, 1982.

[44] Yingzi Du, Chein-I Chang, and Paul David Thouin. Unsupervised approach to color video thresholding. *Optical Engineering*, 43(2):282–289, February 2004.

[45] R. O. Duda and P. E. Hart. *Pattern Classification and Scene Analysis*. John Wiley & Sons, Inc., 605 Third Avenue, New York, 10158-0012, 1973.

[46] Richard O. Duda, Peter E. Hart, and David G. Stork. *Pattern Classification Second Edition*. John Wiley & Sons, Inc., 605 Third Avenue, New York, 10158-0012, 2001.

[47] Brian S. Everitt, Sabine Landau, Morven Leese, and Daniel Stahl. *Cluster Analysis*. Wiley Series in Probability and Statistics, 2011.

[48] Theodoros Evgeniou, Massimiliano Pontil, and Tomaso Poggio. Regularization networks and support vector machines. *Advances in Computational Mathematics*, 13(1):1–50, 2000.

[49] Tom Fawcett. ROC graphs: Notes and practical considerations for researchers. Technical report, HP Laboratories, 2004.

[50] Dr Delmiro Fernandez-Reyes. Private communication.

[51] J. M. Ferryman, A. D. Worrall, G. D. Sullivan, and K. D. Baker. A generic deformable model for vehicle recognition. In D. Pycock, editor, *Proceedings of 6th British Machine Vision Conference*, pages 127–136. BMVA Press, September 1995.

[52] J. E. Fieldsend and R. M. Everson. Formulation and comparison of multi-class ROC surfaces. In *Proceedings of the 2nd ROC Analysis in Machine Learning Workshop*, International Conference on Machine Learning, pages 41–48, August 2005.

[53] G. D. Finlayson, B. Schiele, and J. Crowley. Comprehensive colour image normalisation. In Hans Burkhardt and Bernd Neumann, editors, *Proceedings of the 5th European Conference on Computer Vision*, volume 1 of *Lecture Notes in Computer Science*, pages 475–490. Springer, June 1998.

[54] R. A. Fisher. The statistical utilization of multiple measurements. *Annals of Eugenics*, 8:376–386, 1938.

[55] R. B. Fisher. CVonline: The evolving, distributed, non-proprietary, on-line compendium of computer vision. http://homepages.inf.ed.ac.uk/rbf/CVonline/.

[56] P. A. Flach and S. Wu. Repairing concavities in ROC curves. In *Proceedings of 2003 UK Workshop on Computational Intelligence*, pages 38–44. University of Bristol, August 2003.

[57] Peter A. Flach. The geometry of ROC space: understanding machine learning metrics through ROC isometrics. In *Proceedings of the Twentieth International Conference on Machine Learning*, pages 194–201. AAAI Press, January 2003.

[58] National Center for Infectious Diseases. Division of Parasitic Diseases.

[59] L. A. Forbes and B. A. Draper. Inconsistencies in edge detector evaluation. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 398–404, June 2000.

[60] D. A. Forsyth and J. Ponce. *Computer Vision A Modern Approach*. Prentice Hall, Upper Saddle River, 2003.

[61] John Frean. Microscopic determination of malaria parasite load: role of image analysis. *Microscopy: Science, Technology, Applications and Education*, 3:862–866, 2010.

[62] John A. Frean. Reliable enumeration of malaria parasites in thick blood films using digital image analysis. *Malaria Journal*, 8(218), 2009.

[63] Maria Teresa Gallegos and Gunter Ritter. A robust method for cluster analysis. http://arxiv.org/pdf/math/0504513.pdf.

[64] Antonio Garrido and Nicolas Pérez de la Blanca. Applying deformable templates for cell image segmentation. *Pattern Recognition*, 33(5):821–832, 2000.

[65] Arthur Gelb. *Applied Optimal Estimation*. MIT Press, 1974.

[66] Herbert M. Gilles and David A. Warrell. *Bruce-Chwatt's Essential Malariology Third Edition*. 1993.

[67] G. H. Golub and C. F. van Loan. *Matrix Computations*. The Johns Hopkins University Press, 3rd edition, 1996.

[68] R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Addison-Wesley, Reading Massachusetts, 1992.

[69] J. C. Gower. Generalized procrustes analysis. *Psychometrika*, 40:33–31, 1975.

[70] U. Grenander and A. Srivastava. Probability models for clutter in natural images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(4):424–429, April 2001.

[71] Ningning Guo, Libo Zeng, and Qiongshui Wu. A method based on multispectral imaging technique for white blood cell segmentation. *Computing in Biology and Medicine*, 37(1):70–76, January 2007.

[72] Steven Haker, William M. Wells III, Simon K. Warfield, Ion-Florin Talos, Jui G. Bhagwat, Daniel Goldberg-Zimring, Asim Mian, Lucila Ohno-Machado, and Kelly H. Zou. Combining classifiers using their receiver operating characteristics and maximum likelihood estimation. In James S. Duncan and Guido Gerig, editors, *Medical Image Computing and Computer-Assisted Intervention MICCAI 2005*, volume 3749 of *Lecture Notes in Computer Science*, pages 506–514. Springer Berlin Heidelberg, 2005.

[73] Silvia Halim, Timo R. Bretschneider, Yikun Li, Peter R. Preiser, and Claudia Kuss. Estimating malaria parasitaemia from blood smear images. In *ICARCV '06. 9th International Conference on Control, Automation, Robotics and Vision*, pages 1–6, December 2006.

[74] D. J. Hand and R. J. Till. A simple generalisation of the area under the ROC curve for multiple class classification problems. *Machine Learning*, 45:171–186, 2001.

[75] J. A. Hanley and B. J. McNeil. The meaning and use of the area under a receiver operating characteristic (roc) curve. *Radiology*, 143:29–36, 1982.

[76] J. A. Hanley and B. J. McNeil. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*, 148:839–843, 1983.

[77] R. M. Haralick and L. G. Shapiro. Image segmentation techniques. *Computer Vision, Graphics, and Image Processing*, 29:100–132, 1985.

[78] D. O. Hebb. *The organization of behavior : a neuropsychological theory*. John Wiley And Sons, New York, 1949.

[79] Heiko Hengen, Susanne Spoor, and Madhukar Pandit. Analysis of blood and bone marrow smears using digital image processing techniques. *SPIE Medical Imaging*, 4684:624–635, February 2002.

[80] Christian Hennig. `http://www.homepages.ucl.ac.uk/~ucakche/`.

[81] Christian Hennig. Measurement of quality in cluster analysis. `http://www.homepages.ucl.ac.uk/~ucakche/presentations/cqualitybolognahennig.pdf`.

[82] A. Hill, C.J. Taylor, and T. Cootes. Object recognition by flexible template matching using genetic algorithms. In G. Sandini, editor, *Proceedings of Second European Conference on Computer Vision*, volume 588 of *Lecture Notes in Computer Science*, pages 852–856. Springer Berlin Heidelberg, May 1992.

[83] Ming-Kuei Hu. Visual pattern recognition by moment invariants. *IEEE Transactions on Information Theory*, 8(2):179–187, February 1962.

[84] Song Hu and B. F. Buxton. Using temporal coherence for gait pose estimation from a monocular camera view. In William Clocksin, Andrew Fitzgibbon, and Philip Torr, editors, *Proceedings of British Machine Vision Conference*, volume 1, pages 449–458. BMVA Press, September 2005.

[85] Jing Huang, S. Ravi Kumar, Mandar Mitra, Wei-Jing Zhu, and Ramin Zabih. Spatial color indexing and applications. *International Journal of Computer Vision*, 35(3):245–268, December 1999.

[86] S. Ioffe. Probabilistic linear discriminant analysis. In Ales Leonardis, Horst Bischof, and Axel Pinz, editors, *Proceedings of the 9th European Conference on Computer Vision - Volume Part IV*, volume 3954 of *Lecture Notes in Computer Science*, pages 531–542, May 2006.

[87] Anil K. Jain. *Fundamentals of Digital Image Processing*. Prentice Hall, Englewood Cliffs, New Jersey, 1989.

[88] Anil K. Jain and Aditya Vailaya. Image retrieval using color and shape. *Pattern Recognition*, 29:1233–1244, 1996.

[89] Ramesh Jain, Rangachar Kasturi, and Brian G. Schunck. *Machine Vision*. NcGraw-Hill, New York, 1995.

[90] Tianzi Jiang and Faguo Yang. An evolutionary tabu search for cell image segmentation. *IEEE Transactions on Systems, Man and Cybernetics*, 32(5):675–678, 2002.

[91] I. T. Jolliffe. Discarding variables in a principal component analysis I: Artificial data. *Applied statistics*, 21:160–173, 1972.

[92] I. T. Jolliffe. *Principal Component Analysis Second edition*. Springer, New York, 2002.

[93] Tapas Kanungo, Byron Dom, Wayne Niblack, and David Steele. A fast algorithm for MDL-based multi-band image segmentation. In *Proceedings of IEEE Computer Vision and Pattern Recognition Conference*, pages 609–616, June 1994.

[94] Tapas Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu. An efficient k-means clustering algorithm: analysis and implementation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):881–892, Jul 2002.

[95] Z. Karel. Contrast limited adaptive histogram equalization. *Graphics Gems*, 4:474485, 1994.

[96] M. Kirby and L. Sirovich. Application of the Karhunen-Loeve procedure for the characterization of human faces. *IEEE Transactions on Pattern analysis and Machine Intelligence*, 12(1):103108, 1990.

[97] Josef Kittler. Feature selection and extraction. *Handbook of Pattern Recognition and Image Processing*, pages 59–83, 1986.

[98] Josef Kittler and Fabio Roli. Multiple classifier systems. volume 1857 of *Lecture Notes in Computer Science*. Springer, 2000.

[99] Ron Kohavi. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of International Joint Conference on Artificial Intelligence*, volume 2, pages 1137–1145, August 1995.

[100] D. Koller, K. Daniilidis, and H. H. Nagel. Model-based object tracking in monocular image sequences of road traffic scenes. *International Journal of Computer Vision*, pages 257–281, 1993.

[101] S. Kumar, S. H. Ong, S. Ranganath, T. C. Ong, and F. T. Chew. A rule-based approach for robust clump splitting. *Pattern Recognition*, 39:1088–1098, 2005.

[102] Matthew A. Kupinski and Mark A. Anastasio. Multiobjective genetic optimization of diagnostic classifiers with implications for generating receiver operating characteristic curves. *IEEE Transactions on Medical Imaging*, 18:675–685, 1999.

[103] Thomas Landgrebe and Robert P. W. Duin. Combining accuracy and prior sensitivity for classifier design under prior uncertainty. In *Structural, Syntactic, and Statistical Pattern Recognition*, pages 512–521, 2006.

[104] W. B. Langdon, S. J. Barrett, and B. F. Buxton. Genetic programming for combining neural networks for drug discovery. In Rajkumar Roy, Mario Köppen, Seppo Ovaska, Takeshi Furuhashi, and Frank Hoffmann, editors, *Soft Computing and Industry*, pages 597–608. Springer London, 2002.

[105] W. B. Langdon and B. F. Buxton. Genetic programming for improved receiver operating characteristics. In Josef Kittler and Fabio Roli, editors, *Second International Conference on Multiple Classifier System*, volume 2096 of *LNCS*, pages 68–77, Cambridge, July 2001. Springer Verlag.

[106] Minh-Tam Le, Timo R Bretschneider, Claudia Kuss, and Peter R Preiser. A novel semi-automatic image processing approach to determine plasmodium falciparum parasitemia in Giemsa-stained thin blood smears. *BMC Cell Biology*, 9(15), 2008.

[107] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521:436–444, 2015.

[108] Y. Linde, A. Buzo, and R. M. Gray. An algorithm for vector quantizer design. *IEEE Transactions on Communications*, 28:84–95, 1980.

[109] Nina Linder, Riku Turkki, Margarita Walliander, Andreas Mårtensson, Vinod Diwan, Esa Rahtu, Matti Pietikäinen, Mikael Lundin, and Johan Lundin. A malaria diagnostic tool based on computer vision screening and visualization of Plasmodium falciparum candidate areas in digitized blood smears. *PLoS ONE*, 9(8):e104855, 08 2014.

[110] Lifeng Liu and S. Sclaroff. Index trees for efficient deformable shape-based retrieval. In *IEEE Workshop on Content-based Access of Image and Video Libraries*, pages 83–87, 2000.

[111] Lifeng Liu and Stan Sclaroff. Color region grouping and shape recognition with deformable models. Technical report, Boston University, 1997.

[112] Lifeng Liu and Stan Sclaroff. Deformable shape detection and description via model-based region grouping. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23:475–489, 1999.

[113] Lifeng Liu and Stan Sclaroff. Region segmentation via deformable model-guided split and merge. In *Proceedings of International Conference on Computer Vision*, pages 98–104, July 2001.

[114] Lifeng Liu and Stan Sclaroff. Retrieval by shape population: An index tree approach. Technical report, Boston University, 2001.

[115] S. P. Lloyd. Least squares quantisation in PCM. *IEEE Transactions on Information Theory*, 28(2):129–137, 1982.

[116] Miguel Angel Luengo-Oroz, Asier Arranz, and John Frean. Crowdsourcing malaria parasite quantification: An online game for analyzing images of infected thick blood smears. *J Med Internet Res*, 14(6):e167, November 2012.

[117] S. P. Luttrell. Hierarchical vector quantisation. In *IEE Proceedings*, volume 136, pages 405–413, 1989.

[118] David J. C. MacKay. *Information Theory, Inference & Learning Algorithms*. Cambridge University Press, New York, NY, USA, 2002.

[119] J. D. Maguire, E. R. Lederman, M. J. Barcus, W. A. P. O'Meara, R. G. Jordon, S. Duong, S. Muth, P. Sismadi, M. J. Bangs, W. Roy Prescot, J. K. Baird, and C. Wongsrichanalai. Production and validation of durable, high quality standardized malaria microscopy slides for teaching, testing and quality assurance during an era of declining diagnostic proficiency. *Malaria Journal*, 5:92, 2006.

[120] J. Matas, R. Marik, and J. Kittler. Colour-based object recognition under spectrally non-uniform illumination. *Image and Vision Computing*, 13(9):663–669, 1995.

[121] MathWorld. Sphere point picking. http://mathworld.wolfram.com/SpherePointPicking.html.

[122] Sam Mavandadi, Stoyan Dimitrov, Steve Feng, Frank Yu, Uzair Sikora, Oguzhan Yaglidere, Swati Padmanabhan, Karin Nielsen, and Aydogan Ozcan. Distributed medical image analysis and diagnosis through crowd-sourced games: A malaria case study. *PLoS ONE*, 7(5):e37245, 05 2012.

[123] Sam Mavandadi, Steve Feng, Frank Yu, Stoyan Dimitrov, Karin Nielsen-Saines, William R. Prescott, and Aydogan Ozcan. A mathematical framework for combining decisions of multiple experts toward accurate and remote diagnosis of malaria using telemicroscopy. *PLoS ONE*, 7(10):e46192, 10 2012.

[124] Sam Mavandadi, Steve Feng, Frank Yu, Stoyan Dimitrov, Richard Yu, and Aydogan Ozcan. Biogames: A platform for crowd-sourced biomedical image analysis and telediagnosis. *Games for Health Journal: Research, Development, and Clinical Applications*, 1(5):373–376, 2012.

[125] A. Meijster and M. Wilkinson. Fast computation of morphological area pattern spectra. In *Proceedings of International Conference on Image Processing*, volume 3, pages 668–671. IEEE Signal Processing Society, 2001.

[126] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K. Muller. Fisher discriminant analysis with kernels. In Yu-Hen Hu, Jan Larson, Elizabeth Wilson, and Scott Douglas, editors, *Neural Networks for Signal Processing IX, Proceedings of IEEE Signal Processing Society Workshop*, pages 41–48, August 1999.

[127] Douglas Mossman. Three-way ROCs. *Medical Decision Making*, 19(1):78–89, January 1999.

[128] Olly Oechsle. *Towards the Automatic Construction of Machine Vision Systems using Genetic Programming*. PhD thesis, School of Computer Science and Electronic Engineering, University of Essex, September 2009.

[129] G. Ongun, U. Halici, K. Leblebicioglu, V. Atalay, M. Beksac, and S. Beksac. Feature extraction and classification of blood cells for an automated differential blood count system. In *Proceedings of International Joint Conference on Neural Networks*, volume 4, pages 2461–2466, July 2001.

[130] Godfrey C. Onwubolu and B. V. Babu. *New Optimisation Techniques in Engineering*, volume 141 of *Studies in Fuzziness and Soft Computing*. Springer Berlin Heidelberg, 2004.

[131] Nobuyuki Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man and Cybernetics*, 9(1):62–66, 1979.

[132] Oliver Partida. Classification of malaria infected blood cells using a mixture of Gaussians and morphology. Master's thesis, UCL, September 2005.

[133] James A. Paterson and Andrew W. Fitzgibbon. 3D head tracking using non-linear optimization. In Richard Harvey and Andrew Bangham, editors, *Proceedings of British Machine Vision Conference*, pages 609–618. BMVA Press, 2003.

[134] Jim Pearce. Mice and men: Making the most of our similarities. Technical report, Oak Ridge National Laboratory Review, 1994.

[135] M. Petrou and P. Bosdogianni. *Image Processing: The Fundamentals*. John Wiley & Sons, New York, 1999.

[136] Dzung L. Pham, Chenyang Xu, and Jerry L. Prince. Current methods in medical image segmentation. *Annu. Rev. Biomed. End.*, 2:315–337, 2000.

[137] M. Pontil and A. Verri. Support vector machines for 3-d object recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20:637–646, 1998.

[138] Diniz PR, Murta-Junior LO, Brum DG, de Arajo DB, and Santos AC. Brain tissue segmentation using q-entropy in multiple sclerosis magnetic resonance images. *Brazilian Journal of Medical and Biological Research*, 43(1):77–84, 2010.

[139] W. K. Pratt. *Digital Image Processing Second edition*. John Wiley & Sons, Inc., New York, 1991.

[140] William R. Prescott, Robert G. Jordan, Martin P. Grobusch, Vernon M. Chinchilli, Immo Kleinschmidt, Joseph Borovsky, Mark Plaskow, Miguel Torrez, Maximo Mico, and Christopher Schwabe. Performance of a malaria microscopy image analysis slide reading device. *Malaria Journal*, 11(155), 2012.

[141] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannerty. *Numerical Recipes in C: The Art of Scientific Computing (2nd Edition)*. Cambridge University Press, New York, 1992.

[142] S. J. D. Prince. *Computer Vision Models, Learning and Inference*. Cambridge University Press, 2012.

[143] S. J. D. Prince and J. H. Elder. Probabilistic linear discriminant analysis for inferences about identity. In *Proceedings of IEEE 11th International Conference on Computer Vision*, pages 1–8, October 2007.

[144] Owen Proudfoot, Nathan Drew, Anja Scholzen, Sue Xiang, and Magdalena Plebanski. Investigation of a novel approach to scoring Giemsa-stained malaria-infected thin blood films. *Malaria Journal*, 7(1):62, 2008.

[145] Yashasvi Purwar, Sirish L Shah, Gwen Clarke, Areej Almugairi, and Atis Muehlenbachs. Automated and unsupervised detection of malarial parasites in microscopic images. *Malaria Journal*, 10(1):364, 2011.

[146] K. N. R. M. Rao. *Application of mathematical morphology to biomedical image processing*. PhD thesis, University of Westminster, 2004.

[147] K. N. R. M. Rao and A. Dempster. Area-granulometry: an improved estimator of size distribution of image objects. *IEE Electronics Letters*, 37:50–951, 2001.

[148] K. N. R. M. Rao and A. Dempster. Modification on distance transform to avoid over-segmentation and under-segmentation. In Mislav Grgic, editor, *International Symposium on Video/Image Processing and Multimedia Communications*, pages 295–301, June 2002.

[149] K. N. R. M. Rao, A. G. Dempster, B. Jarra, and S. Khan. Automatic scanning of malaria infected blood slide images using mathematical morphology. *IEE Seminar Medical Applications of Signal Processing*, October 2002.

[150] W. S. Rasband. Image j. `http://rsb.info.nih.gov/ij/`.

[151] Nicholas E. Ross, Charles J. Pritchard, David M. Rubin, and Adriano G. Duse. Automated image processing method for the diagnosis and classification of malaria on thin blood smears. *Med Biol Eng Comput*, 44:427–436, 2006.

[152] S. Roweis. EM algorithms for PCA and SPCA. *Neural Information Processing Systems*, 10:626–632, 1997.

[153] Cecilia Di Ruberto, Andrew Dempster, Shahid Khan, and Bill Jarra. Automatic thresholding of infected blood images using granulometry and regional extrema. In *Proceedings of International Conference on Pattern Recognition*, volume 3, pages 441–444, 2000.

[154] Cecilia Di Ruberto, Andrew Dempster, Shahid Khan, and Bill Jarra. Segmentation of blood images using morphological operators. In *Proceedings of International Conference on Pattern Recognition*, volume 3, pages 397–400, 2000.

[155] Cecilia Di Ruberto, Andrew Dempster, Shahid Khan, and Bill Jarra. Morphological image processing for evaluating malaria disease. In Carlo Arcelli, Luigi P. Cordella, and Gabriella Sanniti di Baja, editors, *Visual Form*, volume 2059 of *Lecture Notes in Computer Science*, pages 739–748. Springer Berlin Heidelberg, 2001.

[156] Cecilia Di Ruberto, Andrew Dempster, Shahid Khan, and Bill Jarra. Analysis of infected blood cell images using morphological operators. *Image and Vision Computing*, 20(2):133–146, February 2002.

[157] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning representations by back-propagating errors. *Nature*, 323:533–536, 1986.

[158] P.D. Sathya and R. Kayalvizhi. Optimum multilevel image thresholding based on Tsallis entropy method with bacterial foraging algorithm. *International Journal of Computer Science Issues*, 7(5):1694–0814, 2010.

[159] Robert J. Schalkoff. *Pattern Recognition: Statistical, Structural and Neural Approaches*. John Wiley & Sons, Inc., 1991.

[160] Bernhard Schlkopf, Alexander Smola, and Klaus-Robert Müller. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10(5):1299–1319, July 1998.

[161] M. J. J. Scott, M. Niranjan, and R. W. Prager. Realisable classifiers: Improving operating performance on variable cost problems. In Mark Nixon and John Carter, editors, *Proceedings of the 9th British Machine Vision Conference*, pages 306–315, September 1998.

[162] C. R. Senanayake and Daniel C. Alexander. Colour transfer by feature based histogram registration. In *Proceedings of British Machine Vision Conference*, pages 1–10. BMVA Press, September 2007.

[163] Chris Senanayake. *Reducing sensitivity of high precision machine-vision inspection systems for detecting foreign or defective bodies/objects on factory conveyor belts*. PhD thesis, UCL, 2009.

[164] L. G. Shapiro and G. C. Stockman. *Computer Vision*. Prentice Hall, 2001.

[165] Selena W. S. Sio, Weiling Sun, Saravana Kumar, Wong Zeng Bin, Soon Shan Tan, Sim Heng Ong, Haruhisa Kikuchi, Yoshiteru Oshima, and Kevin S. W. Tan. Malariacount: An image analysis-based program for the accurate determination of parasitemia. *Journal of Microbiological Methods*, 68(1):1118, 2007.

[166] K. Sirlantzis, M. C. Fairhurst, and M. S. Hoque. Genetic algorithms for multi-classifier system configuration: A case study in character recognition. In Josef Kittler and Fabio Roli, editors, *Multiple Classifier Systems*, volume 2096 of *Lecture Notes in Computer Science*, pages 99–108. Springer Berlin Heidelberg, 2001.

[167] L. Sirovich and M. Kirby. Low-dimensional procedure for the characterization of human faces. *Journal of the Optical Society of America*, 4(3):519524, 1987.

[168] Milan Sonka, Vaclav Hlavac, and Roger Boyle. *Image Processing, Analysis, and Machine Vision*. Chapman & Hall Computing, 1993.

[169] B. Southall, B. F. Buxton, J. A. Marchant, and T. Hague. On the performance characterisation of segmentation algorithms: A case study. In David Vernon, editor, *European Conference on Computer Vision*, volume 1843 of *Lecture Notes in Computer Science*, pages 351–365. Springer Berlin Heidelberg, 2000.

[170] M. Spencer. *Fundamentals of Light Microscopy*. UK: Cambridge University Press, 1982.

[171] R. Storn and K. Price. Differential evolution: a simple and efficient adaptive scheme for global optimization over continuous spaces. Technical report, International Computer Science Institute, Berkeley, 1995.

[172] J. Sullivan, A. Blake, M. Isard, and J. MacCormick. Bayesian object localisation in images. *International Journal of Computer Vision*, 44(2):111–136, 2001.

[173] M. J. Swain and D. H. Ballard. Color indexing. *International Journal of Computer Vision*, 7(1):11–32, 1991.

[174] J. A. Swets. Form of empirical ROCs in discrimination and diagnostic tasks: Implications for theory and measurement of performance. *Psychological Bulletin*, 99(2):181–198, March 1986.

[175] F. Boray Tek. *Computerised Diagnosis of Malaria*. PhD thesis, University of Westminster, September 2007.

[176] F. Boray Tek, Andrew Dempster, and Izzet Kale. Computer vision for microscopy diagnosis of malaria. *Malaria Journal*, 8(1):153, 2009.

[177] F. Boray Tek, Andrew G. Dempster, and Izzet Kale. Malaria parasite detection in peripheral blood images. In *Medical Image Understanding and Analysis*, 2006.

[178] Sergios Theodoridis and Konstantinos Koutroumbas. *Pattern Recognition*. Academic Press, 1999.

[179] M. E. Tipping and C. M. Bishop. Mixtures of probabilistic principal component analyzers. *Neural Computation*, 11(2):443–482, 1999.

[180] M. E. Tipping and C. M. Bishop. Probabilistic principal component analysis. *Royal Statistical Society*, 61(3):611–622, 1999.

[181] W. Trager and J. B. Jensen. Human malaria parasites in continuous culture. *Science*, 193:673–675, August 1976.

[182] Linh Viet Tran. *Efficient Image Retrieval with Statistical Color Descriptors*. PhD thesis, Linköping University, 2003.

[183] E. Trucco and A. Verri. *Introductory Techniques for 3-D Computer Vision*. Prentice Hall, Upper Saddle River, 1998.

[184] Constantino Tsallis. The nonadditive entropy Sq and its applications in physics and elsewhere: Some remarks. *Entropy*, 13(10):1765–1804, 2011.

[185] M. Turk and A. Pentland. Eigenfaces for recognition. *Cognitive Neuroscience*, 3(1):71–86, 1991.

[186] M. Turk and A. Pentland. Face recognition using eigenfaces. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, page 586591, June 1991.

[187] Riku Turkki, Margarita Walliander, Ville Ojansivu, Nina Linder, Mikael Lundin, and Johan Lundin. An open-source, MATLAB based annotation tool for virtual slides. *Diagnostic Pathology*, 8(1), 2013.

[188] E. M. Tzanakou, H. Sheikh, and B. Zhu. Neural networks and blood cell identification. *Journal of Medical Systems*, 21:201–210, 1997.

[189] D. M. Ushizima, A. C. Lorena, and A. Carvalho. Support vector machines applied to white blood cell recognition. In Nadia Nedjah, Luiza M. Mourelle, Marley M. B. R. Vellasco, Ajith Abraham, and Mario Köppen, editors, *Proceedings of International Conference on Hybrid Intelligent Systems*, pages 379–384. IEEE Computer Society, November 2005.

[190] Vladimir Naumovich Vapnik. *Statistical learning theory*. 1998.

[191] V. V. Vinod and Hiroshi Murase. Object location using complementary color features: histogram and DCT. In *Proceedings of International Conference on Pattern Recognition*, volume 1, pages 554–559, August 1996.

[192] Margarita Walliander, Riku Turkki, Nina Linder, Mikael Lundin, Juho Konsti, Ville Ojansivu, Taru Meri, Ville Holmberg, and Johan Lundin. Automated segmentation of blood cells in Giemsa stained digitized thin blood films. *Diagnostic Pathology*, 8(1), 2013.

[193] Simon K. Warfield, Kelly H. Zou, and William M. Wells III. Simultaneous truth and performance level estimation (staple): an algorithm for the validation of image segmentation. *IEEE Transactions on Medical Imaging*, 23(7):903–921, July 2004.

[194] A. Wetzel, P. Feineigle, and J. Gilbertson. Design of a high-speed slide imaging system for pathology. In *Proceedings of IEEE International Symposium on Biomedical Imaging*, pages 305–308, July 2002.

[195] Wikipedia. Cluster analysis. `http://en.wikipedia.org/wiki/Cluster_analysis`.

[196] Wikipedia. Eigenface. `http://en.wikipedia.org/wiki/Eigenface`.

[197] Wikipedia. Gini coefficient. `http://en.wikipedia.org/wiki/Gini_coefficient#Relation_to_other_statistical_measures`.

[198] Wikipedia. Kolmogorov-Smirnov test. `http://en.wikipedia.org/wiki/Kolmogorov-Smirnov_test`.

[199] Wikipedia. Red blood cell. `http://en.wikipedia.org/wiki/Red_blood_cell`.

[200] Y. Yagi and J. R. Gilbertson. Digital imaging in pathology: the case for standardization. *Telemedicine and Telecare*, 11:109–116, 2005.

[201] G. W. Zack, W. E. Rogers, and S. A. Latt. Automatic measurement of sister chromatid exchange frequency. *Journal of Histochemistry and Cytochemistry*, 25(7):741–753, 1977.

[202] E. Zahara, S. K. S. Fan, and D. M. Tsai. Optimal multi-thresholding using a hybrid optimization approach. *Pattern Recognition Letters*, 26(8):1082–1095, June 2005.

[203] I. Zelinka. SOMA - self organizing migrating algorithm. In G. C. Onwubolu and B. V. Babu, editors, *New Optimization Techniques in Engineering*, volume 141 of *Studies in Fuzziness and Soft Computing*, pages 167–217. Springer Berlin Heidelberg, 2004.

[204] I. Zelinka and L. Nolle. Plasma reactor optimizing using differential evolution. *Differential Evolution: A Practical Approach to Global Optimization*, pages 499–512, 2005.

[205] Y. W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Computing Surveys*, 35(4):399–458, 2003.

[206] V. Zografos and Bernard F. Buxton. Pose-invariant 3D object recognition using linear combination of 2D views and evolutionary optimisation. In *Proceedings of International Conference on Computing: Theory and Applications*, pages 645–649, Los Alamitos, CA, USA, 2007. IEEE Computer Society.

[207] Vasileios Zografos. *Pose-invariant, model-based object recognition, using linear combination of views and Bayesian statistics*. PhD thesis, Department of Computer Science, University College London, 2009.