Computerised Stereoscopic Measurement of the Human Retina

David George Greenwood

Image Processing Group
Department of Physics and Astronomy
University College London

Thesis for the Degree of Doctor of Philosophy
University of London
1992

Abstract

The research described herein is an investigation into the problems of obtaining useful clinical measurements from stereo photographs of the human retina through automation of the stereometric procedure by digital stereo matching and image analysis techniques. Clinical research has indicated a correlation between physical changes to the optic disc topography (the region on the retina where the optic nerve enters the eye) and the advance of eye disease such as hypertension and glaucoma. Stereoscopic photography of the human retina (or fundus, as it is called) and the subsequent measurement of the topography of the optic disc is of great potential clinical value as an aid in observing the pathogenesis of such disease, and to this end, accurate measurements of the various parameters that characterise the changing shape of the optic disc topography must be provided.

Following a survey of current clinical methods for stereoscopic measurement of the optic disc, fundus image data acquisition, stereo geometry, limitations of resolution and accuracy, and other relevant physical constraints related to fundus imaging are investigated. A survey of digital stereo matching algorithms is presented and their strengths and weaknesses are explored, specifically as they relate to the suitability of the algorithm for the fundus image data. The selection of an appropriate stereo matching algorithm is discussed, and its application to four test data sets is presented in detail.

A mathematical model of two-dimensional image formation is developed together with its corresponding auto-correlation function. In the presense of additive noise, the model is used as a tool for exploring key problems with respect to the stereo matching of fundus images. Specifically, measures for predicting correlation matching error are developed and applied. Such measures are shown to be of use in applications where the results of image correlation cannot be independently verified, and meaningful quantitative error measures are required. The application of these theoretical tools to the fundus image data indicate a systematic way to measure, assess and control cross-correlation error.

Conclusions drawn from this research point the way forward for stereo analysis of the optic disc and highlight a number of areas which will require further research. The development of a fully automated system for diagnostic evaluation of the optic disc topography is discussed in the light of the results obtained during this research.

Addenda

p.19 "[[Mikleberg86]" should read "[Mikleberg86]"

p.38 "extracted from [Donaldson64]." should read "extracted from Donaldson [Donaldson64]."

p.39 "[Gullstrand11] suggested" should read "Gullstrand [Gullstrand11] suggested"

p.47 "confirm stereo depth estimates." should be extended to
"confirm stereo depth estimates obtained from feature-based stereo matching."

p.51 "[Shields87] reported" should read "Shields [Shields87] reported"

p.52 "is that of [Algazi85]" should read "is that of Algazi [Algazi85]"

p.52 "will be to concentrated" should read "will be concentrated"

p.61 "Lens also exhibit" should read "Lenses also exhibit"

p.68 "image affects to the" should read "image affects the"

p.71 (Figure 4.5, legend)
"$h_{\text{sample}} = 3\text{mm}$" should read "$h_{\text{sample}} = 3.9\mu\text{m}$"

p.81 (Figure 4.9)
"h_{eff} 8.53 μm e.g. at $\Delta z = 10\mu\text{m}$" should read
"h_{eff} 10.27 μm e.g. at $\Delta z = 10\mu\text{m}$"
and
"D 48.33 μm e.g. points at $\Delta z = 10\mu\text{m}$" should read
"D 58.18 μm e.g. points at $\Delta z = 10\mu\text{m}$"

p.93 "in step 1, and thus" should read "in step 2, and thus"

p.122 "to fully covert" should read "to fully convert"

p.132 "For Data Set D" should read "For Data Set C"

p.167 "straight-forward" should read "straightforward"

p.168 "for small values of d and e near the correlation peak †" add footnote
† Each of the derived formulae have been experimentally verified by auto-correlating model scenes generated using equations (7.1) and (7.5) for a range of parameter values. The approximations used in deriving $C(d)$ and $C(d,e)$ were found to be valid for small d and e.

p.189 "8.1c, 8.1e, and 8.1g" should read "8.1b, 8.1e, and 8.1f"

p.193 "cross-correlation in the" should read "auto-correlation in the"

p.203 "match results was" should read "match results were"

p.209 (Figure 8.10)
"and (b) plots the rms" should read "and (d) plots the rms"

Additional Appendix inserted containing reference [Forshaw90].

Table of Contents

List of Illustrations

Acknowledgements

The author is greatly indebted to all members of the Image Processing Group, University College London. In particular I wish to thank my supervisor, Mike Forshaw, whose guidance and enthusiasm never waned throughout the course of this work. I would also like to thank Michael Duff, Terry Fountain, Shawn Javid, and Paul Otto whose insights and experience contributed to this thesis.

I am also indebted to Fred Fitzke from the Institute of Opthalmology and Roger Hitchings from Moorfields Eye Hospital, London, who provided the thesis topic, supplied data, and helped broaden my understanding of ophthalmology. I would also like to thank Gordon Douglas from the University of British Columbia who provided some of the data.

Lastly, I wish to thank my family and friends, without whose patience, moral support, and encouragement this work would not have been possible.

Chapter 1
Introduction

1.1 What is the Optic Disc ?

The Concise Oxford English Dictionary [Allen90] defines *fundus* to mean "the base of a hollow organ, the part furthest from the opening" and in Latin fundus means "bottom". Ophthalmologists use the term fundus to indicate the base or back of the human *retina*. At the base of the eye is a region where the optic nerve enters the eye called the *optic nerve head*. The join between the eye and the optic nerve forms a small surface feature called the *optic disc*.

Ophthalmologists believe that changes to the topography of the optic disc region to be an indication of advancing eye disease, such as glaucoma. Glaucoma refers to a complex of eye diseases that annually cause blindness in an estimated 50,000 people in the United States and Europe. Glaucoma is associated with an increase in intraocular pressure which causes excavation of the optic disc. Therefore, measurement of the topography of the optic disc region is fundamental to the clinical management of any patient suspected of glaucoma.

The terms *stereo* and *stereoscopic* are used throughout this thesis to refer to the visual three-dimensional effect obtained by viewing an object from two disparate positions.

1.2 Definition of the Research Problem

Clinical research has indicated a correlation between physical changes to the optic disc topography and the advance of eye disease such as hypertension and glaucoma. Stereoscopic photography of the human fundus and the subsequent measurement of the topography of the optic disc (the region where the optic nerve enters the back of the eye) is of great clinical value as an aid in observing the pathogenesis of such disease. If stereoscopic analysis is to prove valuable in the clinical assessment of optic disc disease, then it must be developed into a simple and relatively inexpensive technique that can provide accurate qualitative and quantitative measurements of the various parameters that characterise the changing shape of the optic disc topography.

The research described herein is an investigation into the problems of obtaining useful clinical measurements from stereo photographs of the optic disc, through automation of the stereometric procedure by digital stereo matching and image analysis techniques. Conclusions drawn from this research point the way forward for stereo analysis of the optic disc and highlight a number of areas which will require further research. The possibility of the development of a fully automated system for diagnostic evaluation of the optic disc topography is discussed in the light of the results obtained during this research.

1.3 Clinical Use and Application

Fundus photography and imaging are cornerstones of modern ophthalmic diagnosis. Recent developments in computer imaging technology have resulted in many improvements in the acquisition and analysis of fundus images. The potential benefits of systems which might use these new techniques are numerous. However, relatively few commercial fundus imaging systems provide the type of image analysis tools which clinicians could use to aid them in the diagnosis and treatment process.

Digital image processing has been applied widely in recent years to research involving fundus images, and measurements of the optic disc topography were a main target of this type of research. However, automated analysis and measurement of the optic nerve head is not a tool currently available to clinicians. Typically, a skilled computer operator manually marks key regions in a fundus image for subsequent (semi-) automatic measurement.

The clinical value of such a tool, and the potential benefit from the ability to evaluate changes to the optic disc topography over time, needs to be studied. For example, the accuracy of the fundus imaging system and what accuracy is required in the subsequent image analysis to provide meaningful clinical information is unknown.

In undertaking this research, it is intended that there is a move some way toward a fully automated system for optic nerve head analysis. By examining the shortcomings of the fundus imaging system, improving the image analysis techniques, and developing improved mathematical models with which to assess *in vivo* measurements, the research described herein provides a sound base from which to develop a set of useful clinical tools to aid in the process of diagnosis and treatment of optic disc disease.

1.4 Brief Overview of the Research

Chapter 2 begins by summarising the objectives of the research. In particular, the clinical objectives are presented, giving a clearer picture of the purpose of this study. The sources of fundus image data used in this research are discussed together with an examination of the clinical methods for data acquisition. The reasons for employing digital stereo matching techniques in this application are presented.

In chapter 3, a historical survey of stereoscopic fundus photography is presented, followed by an examination of the current techniques for measurement of the optic disc topography from stereoscopic image pairs. The general problems associated with clinical measurement of the optic disc are discussed.

In chapter 4, issues such as data acquisition, stereo geometry, limitations of resolution and accuracy, and other relevant physical constraints are investigated. The problems and issues associated with both non-simultaneous and simultaneous stereo photography of the optic disc are presented, together with an examination of the ideal conditions under which either method may produce useful clinical results.

In chapter 5, a survey of digital stereo matching algorithms is presented and their strengths and weaknesses are explored, specifically as they relate to the performance of the algorithm on the fundus image data. The selection of a suitable stereo algorithm is discussed, and the algorithm is examined in detail.

In chapter 6, the stereo matching algorithm selected in chapter 5 is applied to the fundus image data. The experimental methods for data acquisition, image preprocessing and filtering, application of the stereo algorithm, and subsequent post-processing are presented. Some preliminary results are given together with an analysis of the accuracy of the procedure. This leads to a review of the limitations of the stereo matching algorithm and the isolation of some of the key problems associated with this method.

In chapter 7, the use of mathematical modelling to investigate the structure of auto- and cross-correlation functions is presented. The local structure of the fundus image data is examined using well known image descriptors, leading to a functional model of image formation. Explicit formulae for the auto-correlation of the image model are derived in both one and two dimensions. It is shown that the derived formulae are useful for describing the mathematical process underlying image cross-correlation. A theoretical correlation error prediction tool is developed which provides a means by which to assess the potential accuracy of image correlation in the presence of additive noise. The utility of this error measure is demonstrated for both one- and two-dimensional image models.

In chapter 8, the experimental verification of the theoretical models developed in chapter 7 is presented. Specifically, the application of the stereo matching algorithm (selected in chapter 5) to both synthetic and real image data provides experimental evidence for confidence in the utility of the theoretical error measure. In image matching applications where ground truth verification of correlation match results is not possible, the tools developed in chapter 7 are shown to reflect the expected magnitude of matching error.

Finally, in chapter 9, some conclusions are drawn from this work to point the way forward for stereo analysis of the optic disc. A summary of the research is presented together with a review of the research objectives set out in chapter 2. The extension of this work to provide useful clinical measurements is briefly addressed, highlighting a number of areas which will require further study. A number of suggestions for further work are presented which follow on naturally from this research and which may lead to a fully automated system for diagnostic evaluation of the optic disc topography.

Chapter 2
Research Objectives

2.1 Introduction

In this chapter an overview of the research is presented and the objectives of the research are defined. The general problem of stereoscopy of the human fundus is reviewed together with a discussion of the clinical objectives, the data acquisition and sources of data, the measurement by stereo matching of stereoscopic fundus photographs, and the methods by which the data will be modelled.

Stereoscopy of the human fundus and the subsequent assessment of the topography of the optic disc (the area where the optic nerve head enters the eye) is of great clinical value as an aid in observing the pathogenesis of eye disease such as glaucoma. Research has indicated a correlation between physical changes to the optic disc topography and the advance of hypertension and glaucoma ([Kottler74], [Hitchings76], [Johnson79], [Takamoto79], [Mikelberg86]).

Interest in stereoscopic fundus examination was stimulated by both the development of the stereoscopic fundus camera by Donaldson [Donaldson64] and the application of the techniques of photogrammetry to the resulting stereo pairs ([Ffytche73], [Kottler74]). Prior to this, non-simultaneous stereo photographs were obtained by either changing patient fixation or by moving the camera base. Both methods for obtaining stereo images of the optic disc are used today, and although the method of non-simultaneous stereo photography is prone to error, there is a large volume of data that has been obtained in this manner and therefore warrants further research into the potential of fundus measurement by photogrammetric stereo matching methods.

The actual problems associated with the automatic recording of optic disc topography from stereo photographs (both simultaneous and non-simultaneous) still remain. A photogrammetrist, examing the stereoscopic pairs with opto-mechanical stereoplotting machines or using digital techniques, has no reliable method for calibrating the readings in terms of the physical geometry of the optic disc or of verifying the accuracy of the calculated topographic measurements. In recent years effort has been concentrated on improving the photogrammetric and image analysis techniques used to derive topographic information from stereo images ([Ffytche73], [Rosenthal77], [Caprioli87], [Shapiro87], [Varma87]). If stereo analysis is to prove valuable in the clinical assessment of optic disc disease then it must be developed into a simple and relatively inexpensive technique that can provide accurate, qualitative and quantitative measurements of the various parameters that characterise the changing shape of the optic disc topography ([Algazi85], [Ling86], [Whiteside86], [Mitra87]).

Figure 2.1. A typical fundus photograph printed from a 35mm transparency. Note there are monocular cues to surface topography.

In order to understand the difficulties associated with stereo analysis of the optic nerve head, some physical aspects of the problem must be examined. The constraints imposed by the camera geometry, imaging through the pupil of the eye, as well as the small extent of the optic disc area, present limitations to the accuracy to which topographic information can be extracted using photogrammetric techniques. It is important to quantify these limitations which ultimately affect decisions about the selection of stereo matching algorithms that can be used to automatically or semi-automatically compute topographic information.

2.2 Clinical Objectives

The optic disc is that position in the eye where the optic nerve enters the eye (commonly known as the blind spot) and can be viewed through the pupil by using an appropriate camera. Its position in the eye is slightly toward the nose (a left eye is shown in figure 2.2), and its shape is approximately circular with an average horizontal diameter of 1.75mm and an average vertical diameter of 1.86mm (the average diameter of an eye ball is approximately 23-25mm). The optic disc is characterised by a slight rim around its perimeter, called the neuro-retinal rim, and has the general appearance of a depression or cup (see figure 2.2). The optic cup is defined to be that portion of the optic disc beginning at a predefined depth below the neuro-retinal rim (clinically defined to be 120-150μm, see [Mikelberg84]). The optic cup can be shallow or deep, and can range from a few hundred microns to one millimeter in depth. There is often a scleral ring or bright crescent that surrounds the optic disc, which is often confused with the optic disc boundary itself. It should be noted that the physical dimensions of the optic disc are not in themselves an indication of disease, that is, a deep cup may be healthy for one individual and indicate disease in another.

The clinical value of these parameters can only be appreciated in terms of their change over time. For example, one indicator commonly used is that of cup-to-disc ratio (the diameter of the cup compared to the diameter of the disc). Any significant change to this ratio has been shown to indicate advancing glaucoma [[Mikelberg86]. Another example is a measurement of the neural-retinal rim area (a measure of the area defined by the disc and cup perimeter). This too

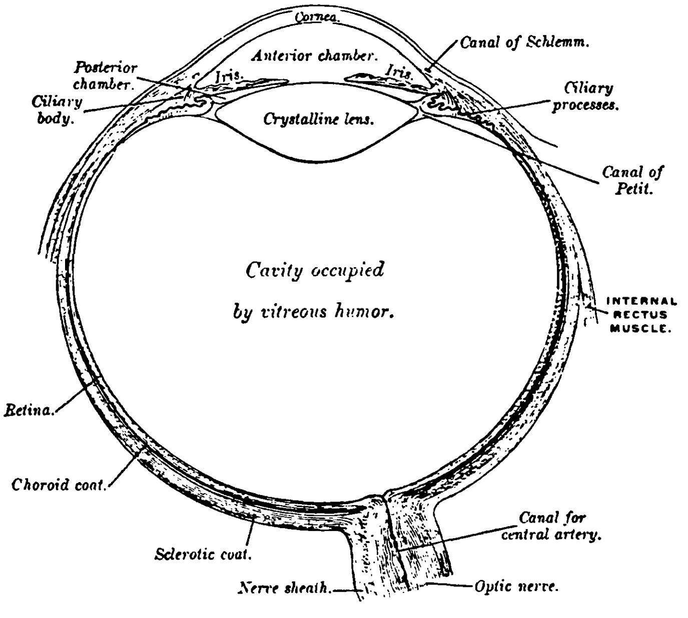


Figure 2.2. Anatomical diagram of the human eye (from [Gray74]).

has been shown to be a good indicator of advancing glaucoma, should there be a large temporal variation ([Johnson79], [Takamoto79]).

Clinical measurements of this type are typically done by hand, with a skilled clinician observing the stereo photographs through a stereo viewer, marking the appropriate regions by hand, and then entering the results into a computer for comparison. However, reproducibility studies have shown a high degree of error from subjective interpretation for this type of manual photogrammetry ([Rosenthal77], [Spaeth87], [Varma87]). Alternatively, the stereo pairs can be processed by semi-analytical stereoplotters controlled by a skilled photogrammetrist. Although the results are usually reproducible, inasmuch as photogrammetric measurements can be agreed by different operators, and are obtained to much higher accuracy, this method is very time consuming and costly. Digital photogrammetric methods are favoured because this removes the need for a skilled operator, and can often produce results far more quickly with equal accuracy. Digital methods suffer in that explicit medical knowledge used by a clinician or photogrammetrist to aid in the measurement process does not translate well into mathematical heuristics that can be applied during automated analysis ([Algazi85], [Nagin85a], [Nagin85b], [Shapiro87]).

The measurement of the optic disc parameters (e.g. disc boundary, cup diameter, neuro-retinal rim), must not exceed certain error thresholds for the results be meaningful to clinicians. It is therefore important to understand all sources of error when trying to automate the measurement process by computer.

The main clinical objective of this research is to develop a reliable method by which the parameters which characterise the physical shape of the optic disc can be extracted automatically by computer, providing a means by which changes to these parameters can be monitored.

2.3 Data Acquisition

2.3.1 Fundus Cameras and their Operation

Photographs of the fundus are obtained by special fundus cameras which are designed for optimal viewing of the retina (for example [Nordenson30], [Norton53]). The patient's head must be held in a fixed position, typically by a head rest, and the eye is positioned by means of a fixation light, at which the patient stares. This allows time for camera adjustment and focusing without much eye movement.

Although the axial length of the eye ranges from 23-25mm, the apparent focal length is approximately 17mm, due to the various effects of corneal curvature, the lens of the eye, and the aqueous nature of the organ [LeGrand80]. The fundus camera which is most commonly used gives a 30 degree field of view and a magnification of 3.4×. This gives an image distance d_i, from lens to film plane, of 57.8mm, computed from the standard magnification equation ($M = d_i/R$, where R is the object distance or range). Using a 2× magnifier, the field of view can be halved to 15 degrees, giving an image magnification of 6.8×. The aperture of the camera is often fixed at 3mm, which results in a reduced depth of field. The light source is projected into the eye away from the optical axis of the lens such that the reflected glare is stopped by the 3mm aperture. The photographs are obtained using a 35mm camera mounted along the optical path of the fundus camera, and the optic disc image occupies the central 10-15mm of a 35mm negative.

To obtain non-simultaneous stereo photographs of the optic disc, the first photograph is taken with the camera shifted as far to the left while maintaining the 3mm aperture to lie within the pupil (the pupil is typically pharmacologically dilated to 6-8mm diameter). For the second photograph the patient maintains the same viewing fixation and the camera base is moved laterally. Now the camera is shifted as far to the right while staying within the limits of the pupil (see figure 2.3). The camera will typically need to be refocused for this view.

Problems with this method are apparent. First, the amount of camera shift is not easily controlled (a standard fundus camera is not designed for such control) and will vary with respect to the operator, the diameter of the patient's pupil, and by the patient changing fixation. An unknown stereoscopic baseline in itself is not an insurmountable problem, but may also be complicated by non-lateral (non-epipolar) shifts in the second image. If the patient changes fixation, even only slightly, non-symmetric changes to the viewing position of the second photograph will result. Refocus, required when the camera is moved for the second photograph, will also cause variations of scale with respect to the first photograph (across the lens of the eye there is typically a change in power of the peripheral optics due to spherical aberrations [ElHage73], and thus refocusing is required). More serious is the fact that the stereoscopic effect is directly related to the separation, or baseline, of the two views; thus a very small baseline can eliminate the stereoscopic effect almost entirely.

Although this method of obtaining stereo photographs of the optic disc is difficult, if not impossible, to reproduce, the current technology for stereo

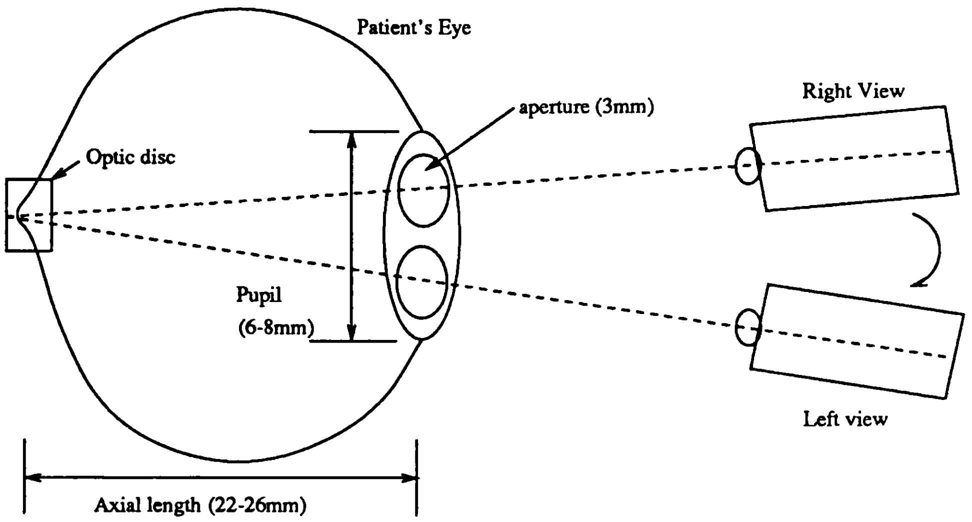


Figure 2.3. Diagram indicating the relationship of the patient's eye, pupil, and left and right camera views for the non-simultaneous stereo method.

matching can deal with a certain amount of variability in baseline, and the subtle rotation and skewing effects of non-parallel (and non-epipolar) stereo views. So, it is not until the limits of this procedure can be quantified that its apparent shortcomings can be assessed.

2.3.2 Camera Geometry

A fixed stereoscopic baseline, or at least a consistent one, can significantly improve the ability to recover depth to an acceptable accuracy. A consistent method for obtaining stereoscopic images of the optic disc would also facilitate comparison between image pairs, allowing for normalisation of results to some fixed scale. Knowledge of the camera geometry constrains the stereo matching process by providing a geometric relationship between the two image coordinate systems, and can improve the accuracy and consistency of the computed depth measurements.

Some limits on the camera geometry can be defined by re-examining the photographic procedure. If it is assumed that the patient's head does not move during the shift of the camera to obtain the second view, and that the same fixation is maintained, then the limits of camera baseline are dictated by the pupil diameter (assuming a lateral shift of the camera). If the 3mm aperture of the camera is to remain entirely within the pupil for both views, then a maximum baseline of approximately 5mm can be obtained if the pupil is dilated to 8mm. Assuming that some minimum stereoscopic baseline can be achieved, say 1mm, this provides a range of 1-5mm for the camera separation. Unfortunately, the extreme case of a 5mm baseline requires viewing through less than favourable portions of the lens of the eye. It is generally accepted that the view with least distortion is obtained through the central 6mm of the lens (with a 3mm aperture the ideal baseline is also 3mm). In practise, the patient's head will move, fixation will not be maintained, and pupil diameter will vary. This means that the stereoscopic baseline of non-simultaneous fundus photographs will vary from patient to patient, and from photographic sitting to sitting.

Hence, another objective of this research is to ascertain whether or not reliable stereoscopic measurements can be obtained from non-simultaneous stereo photographs. Given the large numbers of photographic histories that were obtained using this method, this is an important task.

2.4 Sources of Data

There are several issues that must be resolved with respect to the basic methods for obtaining stereo photographs. Although there are numerous arguments against non-simultaneous stereo photographs [Donaldson64] (as they relate to this specific application, though not in general), it must be noted that a large volume of data has been acquired in this manner and if a method existed for extracting information from this data, however crudely, it might still be of clinical value. These arguments will become apparent after examining the physical constraints more closely in chapter 4, but here it is important to note that although the recent availability of simultaneous fundus cameras provides better quality data, long clinical photographic histories of data of this type are not generally available. Medical diagnosis and treatment of diseases like glaucoma require years of clinical evaluation, and an analysis of a complete photographic history would be desirable. In the same light, it must also be said that the constraints provided by simultaneous stereo photographs of the optic disc may lead to improved measures of change. The lack of photographic histories does not preclude the advance of stereoscopic analysis of the optic disc (see [Hixson87], [Zimmerman88]).

The data used for this research were obtained from four sources.

Data Set A:

A physical model is seen as a means by which to test the stereo matching algorithm with a physical object of known dimensions. This approach has been previously employed by ophthalmologists with the development of the Zeiss Model Eye, which includes a crude but simple model of the optic disc, and has been used for purposes of calibration (see [Rosenthal80]). Although image data from such a model are not directly comparable to fundus image data, it is intended that use of such data will give an indicator of the accuracy of stereoscopic measurement. A plaster model was constructed which consists of a hemi-spherical depression in a flat plaster surface. To give the surface a pattern or texture, the model was spattered with black paint. Once the model was prepared, a workbench and vidicon camera were calibrated so that images of the model could be obtained where all the geometric parameters were known (e.g. object distance, stereoscopic baseline, camera aperture). A stereo image pair of the model was digitised using this calibrated camera arrangement, shown in figure 2.4.

Data Set B:

The Moorfields Eye Hospital, London, in conjunction with the Institute of Ophthalmology, University of London, has provided the time and facilities to photograph the author's left eye using a standard fundus camera (figure 2.5). A special apparatus called a bite-bar can be mounted to the head-rest of a standard fundus camera system, which (as near as is possible) ensures that the patient's head will be in the same position for subsequent photographs. In addition, a dial micrometer was mounted to the sliding base of the fundus camera to ensure a consistent baseline. A series of 35mm stereo photographs with a 3mm baseline was obtained. This data will provide a means for verifying the consistency of both the photographic process and the methods employed for stereo measurement. It is assumed that other methods for obtaining non-simultaneous stereo photographs would have similar physical constraints, and would therefore produce similar results.

Data Set C:

The University of British Columbia, Vancouver, Canada, with kind permission from Dr. G. R. Douglas, has provided an extensive set of stereo photographs obtained using the Donaldson Stereo Fundus Camera (discussed further in chapter 3). The data vary in quality but provide a source of simultaneous stereo photographs of a large set of patients (figure 2.6). The 35mm photographs are accompanied by patient clinical data, and therefore will be useful for estimating the accuracy of stereoscopic measurement. The Donaldson Stereo Fundus Camera has been employed in a number of studies on stereoscopic measurement of the optic disc, and thus results can be compared to those currently in the literature.

Data Set D:

The Moorfields Eye Hospital, London, in conjunction with the Institute of Ophthalmology, University of London, has been experimenting with the latest form of ophthalmic camera - the Scanning Laser Ophthalmoscope (SLO) [Plesch87]. The SLO scans a small laser beam across the retina and records the reflected light as an analogue signal (discussed further in chapter 3). This signal can be displayed on a television monitor, recorded onto video tape, or digitised using an analogue-to-digital converter. Some preliminary digitised data have been provided in order that some early indication of the benefits of this new and highly innovative camera might be seen (figure 2.7). The quality of image obtained from the SLO is far better than from traditional fundus cameras, and it is hoped that these high-quality images might lead to more accurate stereoscopic measurements. In this sense, the SLO data are intended to be used for testing purposes only.

The application of stereo matching to this comprehensive set of data will help to determine the utility of automated photogrammetric measurement of the optic disc.

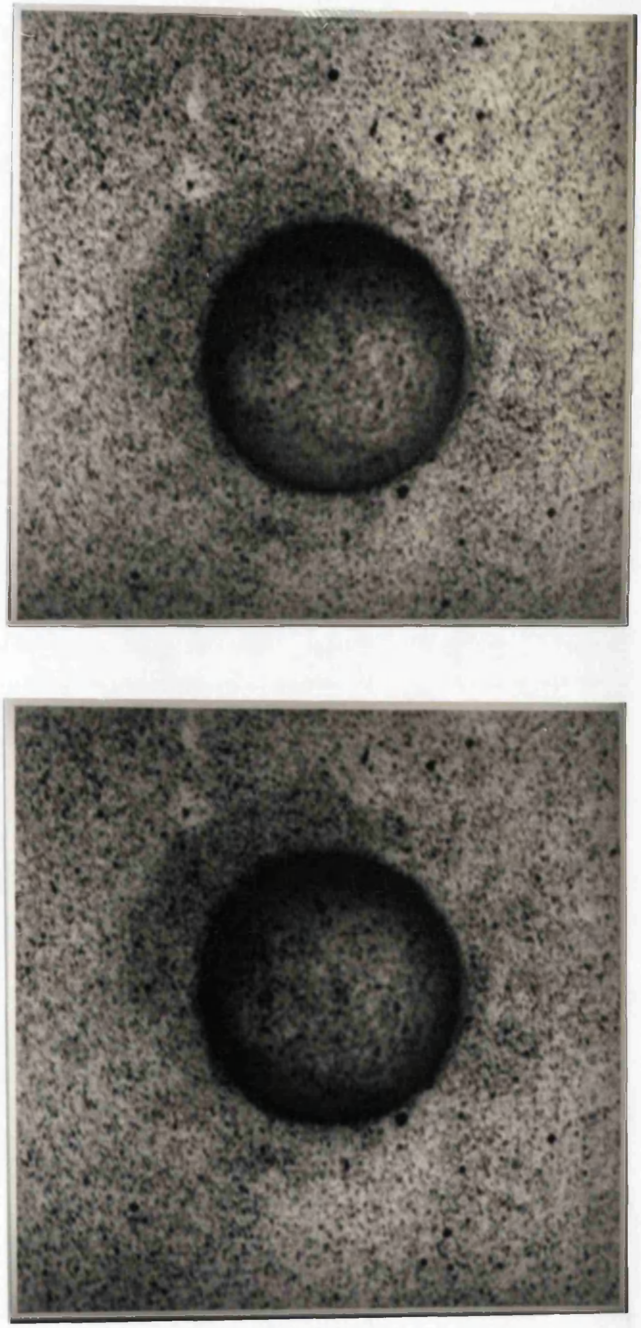


Figure 2.4. Data set A: Plaster Model stereo images.

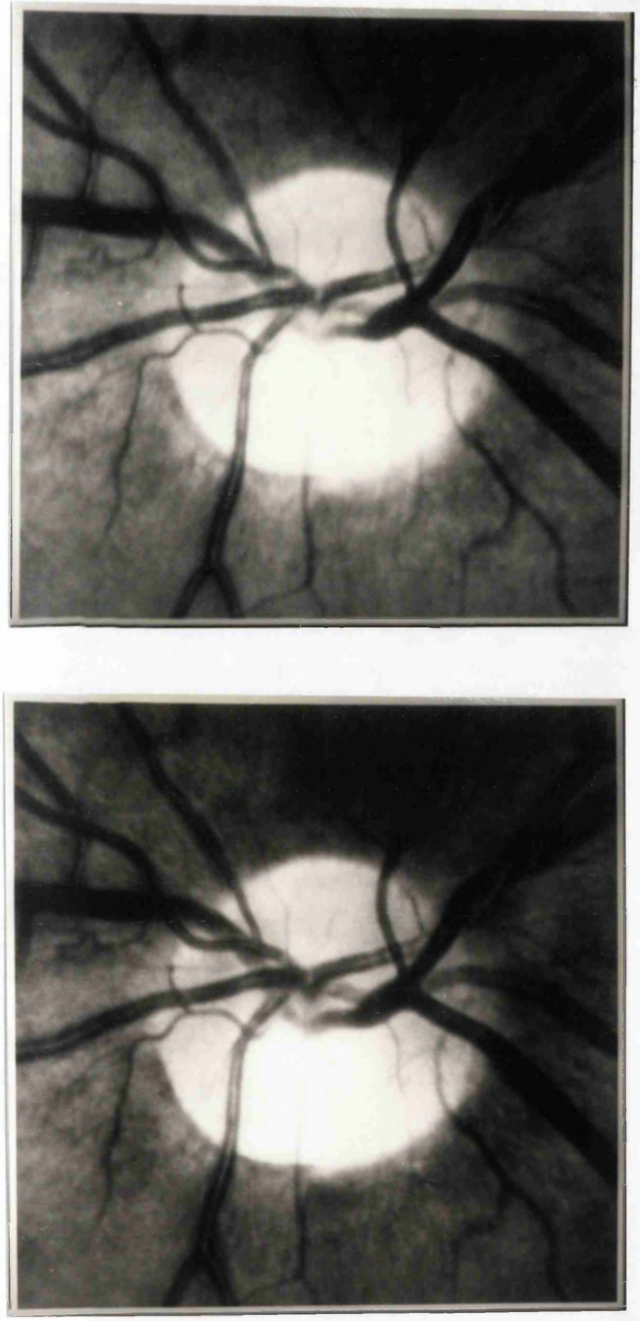


Figure 2.5. Data set B: Non-simultaneous stereo fundus images.

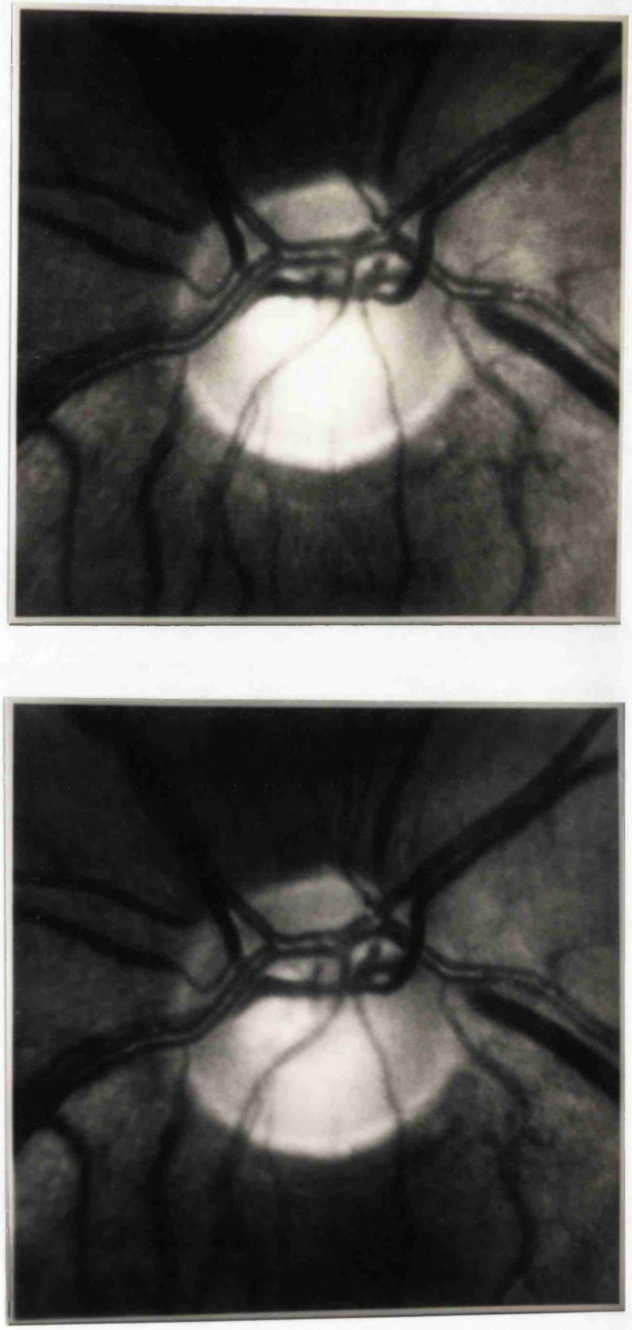


Figure 2.6. Data set C: Donaldson stereo fundus camera images.

Figure 2.7. Data set D: Scanning Laser Ophthalmoscope images.

2.5 Measurement by Stereo Matching

Clinical studies have indicated that no single measurement parameter is able to differentiate normal and glaucomatous optic cups (e.g. [Johnson79], [Mikelberg86]). Glaucomatous alterations to the optic nerve head are complex, involving changes in several geometric features in localised regions of the optic cup, including optic cup volume, area, and depth. Furthermore, measurements from individual patients are not comparable - a deep cup in one patient may be normal, whereas a deep cup in another may be an indication of hypertension or glaucoma. Thus it is important to monitor the *changes* to the optic nerve head topography over time.

There are a limited number of methods that might be employed in order to obtain topographic information about the optic disc. The size and location of the optic nerve head complicates the task. Data about this region of the eye must be obtained *in vivo*, and thus direct measurement is not possible. The eye itself is an extremely sensitive organ - precautions must be made to avoid damaging the retinal tissue and cells. Other methods of measurement, such as structured lighting [Zimmermann88], cannot generally provide the accuracy required. Hence, clinicians have traditionally relied on photographic data as a means by which to assess the condition of the optic nerve head.

Thus, it is not surprising that as early as 1909 ophthalmologists first began experimenting with stereo photographs of the optic nerve head. With the advent of the modern digital computer, photogrammetric techniques could be applied to stereo fundus photographs to automate and standardise the measurement process. However, the task is not a trivial one. The basic premise of stereoscopic measurement is that if the location of corresponding points can be found in each image, information about the three-dimensional position of each point can be calculated (given information about the camera geometry), and thus a complete three-dimensional model of the original surface can be derived from stereoscopic views of a scene. Although this can often be a trivial task for images in which the corresponding points are obvious, automatic (or even manual) stereo matching of the fundus data is difficult due to the complex visual structure of the images.

Conventional fundus photographs for the most part contain large (dark) vascular structures on a background of mottled grey texture; the retinal nerve fibre layer. The vascular structures are known to lie in and above the nerve fibre layer, and have a glossy or reflective appearance. The nerve fibre layer provides little textural information and these areas are almost homogeneous in intensity. The primary area of interest, the optic cup, is often saturated in intensity in a fundus photograph (see figure 2.1), and may provide little or no information on which to form corresponding stereo match points. A discussion of the methods used to automatically match stereo image pairs is presented in chapter 5.

Let us assume that a suitable method for automated stereo matching of the fundus data can be found. The question arises as to the accuracy required for useful clinical results. Current methodologies for measurement of the optic nerve head have obtained results with an accuracy of about 50 to 70µm in depth (Takamoto79], [Peli89]). If stereoscopic (photogrammetric) measurement is to prove viable, what resolution image must be used, to what accuracy must the stereo matching algorithm perform, and what depth accuracy can be achieved? These questions will be addressed in chapter 4.

Therefore, another objective of this research is to assess the potential accuracy of measurement from stereoscopic analysis of the optic disc, and to select a suitable stereo matching algorithm to automate the task.

2.6 Data Models

The application of stereo matching to the fundus image data presents a number of problems. As the fundus images are obtained *in vivo*, direct measurement is not possible, and therefore validating the accuracy of measurements from stereo photogrammetric methods will be difficult; this is unlike stereoscopic measurement applications where the measurements can be directly validated, such as aerial photogrammetry. The use of physical and mathematical models to represent the fundus image data may provide a means to explore key problems with respect to stereo matching. The use of such models provides a known *ground truth* from which it becomes easier to validate experimental results obtained on real data. Provided a sufficiently complex model can be developed which will simulate the fundus image data, stereo matching algorithms can be tested on the model to see how they might perform on the real fundus data. Conversely, given a stereo matching algorithm, the use of mathematical modelling may provide a means by which some key parameters might be tuned so that the algorithm will perform at its potential best for a given type of data.

The use of both physical and mathematical models will be applied in this research. The physical model, a large-scale plaster model of the optic nerve head, will be used as a *ground truth* model to "test" the stereo matching algorithm. Mathematical models will be used to explore the structure of auto- and cross- correlation functions, such as those often employed for stereo matching. This will allow predictions of the performance of a stereo matching algorithm which makes use of such correlation techniques. To predict the performance of a particular correlation algorithm from its theoretical description is extremely difficult, while at the same time it is equally difficult to extrapolate performance measures from experimental results. In chapter 7, explicit formulae which permit predictions on the performance of standard correlation algorithms (such as those used in stereo matching) in terms of a few easily-definable scene parameters are derived.

Thus, through modelling the data there are two additional research objectives: to use a physical model as a form of ground truth from which to test the performance of the stereo matching algorithm, and to develop a mathematical image model to permit valid theoretical predictions on the performance of the stereo matching algorithm.

2.7 Measuring Clinical Parameters

In conventional stereo photography of the fundus, a skilled clinician or trained observer manually processes each stereo pair, estimating parameters such as cup-to-disc ratio and cup volume. Analysis is usually qualitative or at best semi-quantitative, and is prone to a high degree of subjective error ([Rosenthal77], [Spaeth87]). At Moorfields Eye Hospital, London, clinicians observe the photographic prints through a stereo viewer and mark the appropriate regions by hand. The resulting disc and cup boundaries are then entered into a computer using a digitising tablet, and measurements such as cup volume and neuro-retinal rim area can be computed.

In some commercial optic disc analysis systems, these key topographic features can be measured automatically [Peli89]. By replacing the 35mm camera normally mounted to the fundus camera with a Vidicon or CCD video camera, images of the fundus can be obtained for immediate processing by digital computer. Digital photogrammetric methods can be applied to stereo video images to produce results far more quickly and reliably than with manual methods. However, such automated methods suffer in that expert medical knowledge used by clinicians during manual analysis does not translate well into mathematical heuristics that can be applied during automated analysis ([Algazi85], [Peli89]). Although some of these systems (reviewed in chapter 3) reconstruct the topography of the optic disc region, they often require operator intervention, especially in order to highlight the disc margin and edge of the cup. A completely automated system for measuring the disc topography has yet to emerge.

Thus, another objective in this research is to investigate methods for extracting the clinical parameters from the topographic measurements obtained from stereo matching the fundus image data. In this study, both automatic and semi-automatic methods to locate the optic disc margin will be discussed.

2.8 Summary

In this chapter a description of the research problem is presented and the following research objectives are proposed:

(1) to ascertain whether or not reliable stereoscopic measurements can be obtained from both non-simultaneous and simultaneous stereo photographs, and to assess the data already obtained for use in this study;

(2) to study the physical constraints related to stereoscopy of the human fundus, and to study any potential limitations in the measurement and analysis of the optic disc;

(3) to select a suitable stereo matching algorithm to automate the task;

(4) to determine the utility of automated photogrammetric measurement of the optic disc by application of stereo matching to a representative set of fundus data;

(5) to use a physical model as a form of ground truth from which to test the performance of the stereo matching algorithm, and to develop a mathematical model of the fundus image which will permit theoretical predictions on the performance of the stereo matching algorithm;

(6) and, to investigate methods for extracting the clinical parameters from the topographic measurements obtained from stereo matching fundus image data, providing a means by which changes to these parameters can be monitored.

The research objectives defined in this chapter will be reviewed again in chapter 9 to see how well each objective has been met, and to make suggestions for further research based on the results.

Chapter 3
Stereoscopic Fundus Photography

3.1 Introduction

In this chapter, a historical survey of stereoscopic fundus photography is presented. Current techniques for measurement of the optic disc from stereoscopic image pairs are discussed, and an analysis of the general problems associated with measurement of the optic disc is presented.

The accurate assessment of the depth of the optic disc and the extent of pathological cupping has been the goal of glaucoma researchers for many years. Any method that can measure changes of the optic disc volume would prove very helpful to the understanding of the pathogenesis of disc cupping and would also serve as a useful index of the progress of glaucoma and the response to therapy. Interest in stereoscopic fundus photography was stimulated by the development of the stereoscopic fundus camera by Donaldson and the application of the techniques of photogrammetry to the resulting image pairs [Donaldson64]. Prior to this, several authors had produced non-simultaneous stereo photographs, but this method proved too inaccurate because of the difficulty in maintaining a fixed stereoscopic base between separate exposures ([Bedell27], [Metzger27]). With the development and availability of stereoscopic fundus camera systems, simultaneous stereophotography of the optic disc eliminated many of the errors found in previous forms of measurement. A number of these systems, old and new, will be discussed.

The problems associated with the actual measurement of disc depths from stereo photographs still remain. Until recently the photogrammetrist, examining stereoscopic pairs with opto-mechanical stereoplotting machines or using digital techniques, had no reliable method for calibrating the readings in terms of the geometry of the optic disc or of verifying the accuracy of the calculated disc cup volumes. In recent years the effort has been concentrated on improving the photogrammetric and image analysis techniques used to derive topographical

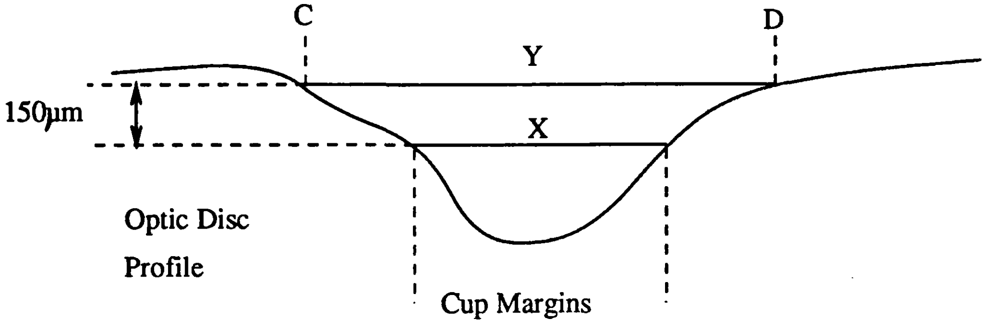


Figure 3.1. Schematic diagram of the optic disc. C and D indicate optic disc margins; the line C-D is the disc diameter; X/Y is the cup-disc ratio. The area between lines Y and X is the neuro-retinal rim area and the area below line X represents the cup volume (from [Mikelberg84]).

information from stereo images. A review of current methods for photogrammetry of the optic disc will be presented.

If photogrammetry (digital stereo analysis) of the optic disc is to prove valuable in the clinical assessment of disc cupping then it must be developed into a simple and relatively inexpensive technique that can provide accurate quantitative and qualitative measurements of the various 3-dimensional parameters that characterise the changing shape of the optic cup.

3.2 History of Non-Simultaneous Stereoscopic Fundus Photography

This early history was extracted from [Donaldson64].

Thorner was the first to publish stereoscopic photographs of the optic disc in 1909 [Thorner09]. His equipment utilised the indirect ophthalmoscopic principle. By dividing the pupillary area in half vertically, illumination could enter one half of the pupil while the other half was used for the emerging rays. Using an ophthalmoscopic lens these rays formed an inverted image which could then be photographed. The second photograph was taken by inverting the apparatus, resulting in a stereoscopic pair. The lighting for this equipment used a kerosene lamp, and the stereoscopic effect was poor.

In 1915, Nordenson first reported his camera based on principles laid down by Gullstrand [Nordenson15]. [Gullstrand11] suggested that while it was necessary that the portion of the fundus being observed must be illuminated, it was also necessary "that no part of the cornea or lens of the patient's eye shall be at once in the region of radiation of the illumination system and also in that of the observation system". Gullstrand also produced data from which Zeiss made the first aplanatic ophthalmoscopic lens. This lens eliminated many of the aberrations and thus made possible the excellent detail produced by the Zeiss-Nordenson camera first available in 1925. Although other manufacturers developed fundus cameras, it was the Zeiss-Nordenson camera which was used almost exclusively for many years.

In 1927, Nordenson proposed a method for taking stereoscopic fundus photographs, similar to the techniques of Thorner [Nordenson27]. At that time, others also published results which displayed stereoscopic effects ([Bedell27], [Metzger27]). The method required taking two successive exposures in order to obtain a stereoscopic pair, and most commonly the Zeiss-Nordenson camera was used. Typically, the patient simply changed fixation slightly between exposures, but others advocated shifting the camera while the patient maintained the same fixation.

Nordenson's method can produce good results in certain instances, but has several serious problems:

(1) satisfactory stereoscopic photographs cannot be obtained in patients who cannot fixate accurately.

(2) fixation is impossible to maintain between exposures in patients who cannot cooperate.

(3) the stereoscopic effect can be poor or completely lost because of changes of various factors between exposures such as focus, position of the patient's head, and operator misalignments.

(4) a dependable, consistent depth effect cannot be obtained in pictures taken subsequently as it is difficult (indeed, almost impossible) to duplicate the exact degree of camera shift (stereoscopic base), patient head position, or fixation. Even small differences have been shown to produce large variations in the resulting stereoscopic effect ([Kottler74], [Johnson79]).

3.3 History of Simultaneous Stereoscopic Fundus Photography

As early as 1930 Nordenson reported taking simultaneous stereo photographs with a specially constructed Zeiss-Nordenson camera [Nordenson30]. The stereoscopic effect was achieved by means of 2 small prisms in front of the camera lens. Undesirable distortions were caused which made the fundus image look convex rather than concave when viewed stereoscopically. The resulting images also showed poor definition and clarity. In 1953, Norton attached a 35mm camera to a Bausch and Lomb binocular ophthalmoscope to produce simultaneous stereo fundus photographs. Again, the resulting photographs lacked the detail of those taken monocularly, and the lack of adequate illumination forced exposure times which were longer than desirable [Norton53].

Finally, in 1964, Donaldson [Donaldson64] developed a fundus camera for simultaneous stereo photographs, utilising the indirect opthalmoscopic principle. A new illumination system was developed to provide high intensity, short duration flash for taking the picture, as well as low intensity rapid flashes for alignment and focusing. Since the development of the Donaldson camera numerous other systems have been built (see [Rosenthal77], [Mikelberg86], [Woon90], [Zimmermann88]), but the basic principles of simultaneous stereo fundus photography remain the same. In Donaldson's opinion, most of the problems that non-simultaneous stereo fundus photography created have effectively been eliminated with simultaneous stereo cameras. Subsequent pictures can be consistently compared since the stereoscopic base remains constant and patient fixation is no longer a problem [Donalson64].

Surprisingly, there are few published examples of developments in simultaneous stereo fundus photography. It is often the case that company confidentiality will keep useful technical reports of such systems from publication. In the following section, some of the more recent systems for simultaneous stereo fundus photography are reviewed.

3.4 Current Technologies

With increasing sophistication in biomedical photographic techniques, technicians are now able to capture a picture of the optic nerve head that was unimaginable just 20 years ago. At the same time, no matter how improved the photographs, problems associated with the photogrammetric techniques prohibit anything but a somewhat primitive analysis of the retinal condition. The emphasis has changed from image acquisition to image analysis in an effort to improve the accuracy of measurement.

Rodenstock Instruments have developed a system for measuring the optic disc topography [Bishop87]. The Rodenstock system projects a series of parallel vertical bands, alternating black and white, onto the optic nerve head (see figure 3.2). From two digitised images the angles of distortion of the bands are calculated from 1600 control points, providing topographical information about the optic disc. This system has been shown to fail on advanced glaucomatous discs, but has proved useful in the majority of cases [Mikelberg84]. The results of the optic disc measurements are produced rapidly and sequential comparisons can be made from previously stored digitised data. Theoretical depth sensitivity of 12μm is quoted, however in practice has been shown be to on the order of several tens of microns. Reliability studies have shown the Rodenstock Analyser to be an excellent diagnostic tool, yielding highly reproducible results in optic disc measurements such as cup-to-disc ratio and neural rim area, but less reliable for cup volume (see [Mikelberg86, Zimmermann88]).

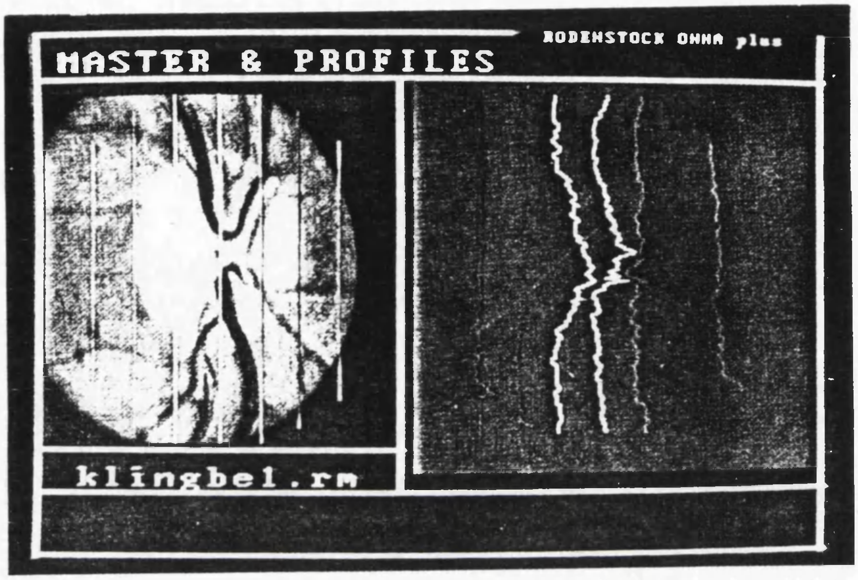


Figure 3.2. Photograph of a television image from the Optic Nerve Head Analyzer showing vertical bright stripes superimposed on an optic disc image and the corresponding contour lines (from [Peli89]).

PAR Microsystems uses a standard fundus camera with a stereoscopic attachment, interfaced with two video cameras linked to a microcomputer that digitises the stereo images simultaneously [Zimmermann88]. Images can be stored on videodisc, providing data archiving and retrieval. The images are correlated automatically by the system to produce the topographical information. The system provides no mechanism whereby subsequent images can be aligned or scaled to a previously stored image to facilitate comparisons between a time-series of images. The PAR system is in limited use and reports of its reliability and reproducibility have not as yet been published.

The Allergan Humphrey system, called the Retinal Analyser, consists of a fundus imaging system and an image analyser. The recording system can capture retinal images with 35mm or Polaroid film, or in video using either analogue or digital recording media. The menu-driven system allows recording of single images or stereo pairs. After rough alignment of the patient's eye, an internal microprocessor automatically informs the technician when all conditions are favourable for recording. Two features ensure high image quality and reliability: a new method of illumination, and a unique automatic tracking system. For better image quality a CCD array camera is scanned across the eye with a controlled flash, and the reduction in light from that of standard fundus cameras provides better image contrast. The computer tracking system ensures consistent reproducibility from one image to the next by keeping the image aligned on the eye horizontally, vertically, and axially [Zimmermann88].

The Retinal Analyser includes a colour monitor, hard disk store, an optical video disc for permanent storage, a colour printer for hardcopy output, and a light pen for menu selection and operation. The system uses image processing software to chart topographical dimensions from stereo images, and creates a grid-mapped 3-dimensional representation of the optic disc. Another program measures changes over time through image differencing. Unfortunately, there are no publications in the literature which allow for adequate assessment of this system to be made with respect to its design and operation, and there are no performance figures other than those quoted by the manufacturer.

Laser technology has been employed in a number of medical applications, and ophthalmology is no exception. Lasers are being used for various types of eye surgery, and recently have been used for funduscopic examination with the development of the Scanning Laser Ophthalmoscope (SLO) [Webb87] [Hixson87]. The SLO's illumination exposes the retina to only one ten-thousandth of the light energy required by conventional ophthalmoscopes. The need for pharmacological pupil dilation is eliminated, unless stereo images with baseline are needed, and the fluorescein doses for angiography can be greatly reduced. Videotapes of the fundus can be used to record the patient's optic deterioration or improvement from therapy.

Rodenstock Instruments is licenced to manufacture the instrument. The SLO projects a low-power 50-micron diameter laser beam on to the fundus [Plesch87], which has several clinical advantages over a standard fundus camera. Often, patients have other eye problems which can complicate examination of the optic disc. The size of the beam enables clinicians to avoid any lens opacities (such as cataracts) when imaging the optic disc, and thus visualisation of the fundus through a clouded lens can be enhanced. The system shows advantages in screening patients for early signs of glaucoma, visual field loss, and macular disease, allowing ophthalmologists to consult and review a patient's condition in real time (see also [Plesch87], [Webb87], and [Woon90]). The SLO provides confocal images allowing optical sectioning and in turn tomographic reconstruction. A comparison of stereo versus tomographic reconstruction is currently underway at the Institute of Ophthalmology. It must be stressed that the SLO has yet to be used by others for optic nerve head reconstruction from stereo images or by tomographic reconstruction. Therefore, this early example of stereo reconstruction of the optic disc will be of great interest to ophthalmologists.

3.5 Optic Disc Measurement

A history of stereoscopic fundus photography has been presented, together with a review of the current technologies for stereoscopic fundus image acquisition. The methodologies for measurement of the optic disc from stereo images are of interest, but the majority of published studies concentrate on the reproducibility of the photographic techniques, rather than providing a meaningful discussion of the techniques for measurement (automatic or manual). The limited number of articles on commercially available systems do not disclose the methods employed, the reasons for using the methods, or the problems encountered or overcome. For these reasons one must rely mainly on the reliability and reproducibility studies as an indicator of the methods currently being employed for reconstructing the optic disc topography from the stereo image pairs.

3.5.1 Photogrammetric Methods

One of the first thorough studies into the problems of photogrammetry of the optic disc was done at Moorfields Eye Hospital in collaboration with the Department of Photogrammetry, University College London, by Ffytche *et al* [Ffytche73]. Using a standard fundus camera and a device known as the Allen separator, stereo photographs of the optic nerve head were taken sequentially and then analysed in a Hilger and Watts stereo comparator. A large number of manual readings were recorded from high-confidence features in the images to define the boundary of the optic disc. The time involved in manually processing each stereo pair proved too costly to be of clinical use. Kottler [Kottler74] applied the same principles as those used by analogue stereo plotters to produce a digital photogrammetric algorithm. Using a Zeiss fundus camera and the Allen separator, stereo photographs were obtained and digitised (at 128×128 pixel resolution) for analysis. Depth sensitivity of 184μm was reported, but it was suggested that this was a restriction of the digitisation resolution and that 92μm was possible using 256x256 pixel resolution. The photogrammetric error was shown to be only 2-3% and the photographic error (focusing, alignment, radiometric differences) was 7 to 20 times greater. Processing time for an image pair was approximately 10 minutes, but the algorithm and required operator intervention were not discussed.

3.5.2 Stereo Matching Methods

Perhaps the most interesting and relevant paper to date is that of Algazi [Algazi85]. Two methods for computer analysis of the optic disc are presented. One method uses successive monocular photographs (over a period of years) which are scaled, registered, and digitised with reference to the first image in the series. These are then displayed in rapid sequence on a television monitor, and changes in the optic cup appear as localised movement. This method is similar to that of stereo chronoscopy [Goldmann77]. Stereo photographs can also be viewed by this technique, where the right image is scaled, registered, and digitised with respect to the left image. A new algorithm for digital photogrammetry is presented for extracting cup depth information. Photographs were obtained using either a Zeiss camera (non-simultaneous), a Donaldson camera, or a Topcon stereo fundus camera. The resulting colour slides were then mounted on a light table and rotational and translational adjustments were performed by means of a micrometer stage. Scaling adjustments were made by altering the video magnification. By rapid alternation between the stored, digitised left image and the live video right image, the right image could be brought into alignment by minimising the "flicker" between the two images.

Algazi attempted to overcome some of the sources of error that were apparent in the photographic process, such as misregistration, focusing, radiometric variations, photographic artifacts, and contrast range. Algazi's algorithm for stereo matching fundus images is:

(1) align and digitise the stereo pair (using the method described above);

(2) apply a radiometric correction to one of the images to bring the grey level variance of the two images into agreement;

(3) apply a highpass filter to both images to extract the high frequency information (what the author calls "skeletonisation");

(4) apply an absolute difference algorithm to both the filtered images and the grey level images to form two depth maps;

(5) using a "choice" algorithm, select a consistent depth map from the two depth maps of (4);

(6) apply a 2-d polynomial fitting algorithm to the depth map to "fill in" the missing data;

The author makes some suggestions for automating the registration process. Although a video fundus camera system which could align the images before digitisation would be desirable, it would not be useful for analysing standard fundus photographs. Some results are shown which indicate that good quantitative information about the optic disc can be obtained by this method.

A recent paper on feature-based registration of retinal images presents a new, fast technique for reliable image registration [Peli87]. A modification of the sequential similarity detection algorithm (SSD) is developed for fundus image registration. A 100 times speed-up is quoted over the conventional cross-correlation technique. This registration method might prove useful in conjunction with the stereo matching technique proposed by Algazi [Algazi85].

A more recent study of integrating stereoscopic measurement with photometric information as cues for depth is presented in [Lee91a]. The authors contend that photometric information, often termed shape from shading (see chapter 5), can be used to confirm stereo depth estimates. Some preliminary experimental results are presented for model and fundus images which show promise for this approach. Further work, involving a clinical assessment of this approach is proposed, and a new fundus camera is to be constructed which will meet the geometric and illumination requirements of both binocular and photometric stereo. The results of Lee's work will be of great interest in comparison with the research presented herein.

3.5.3 Non-Stereo Methods

A different approach to the problem was reported in Goldmann [Goldmann77]. A technique known as chronoscopy requires two photographs of the same eye be taken at different times with the axis of the fundus camera at exactly the same position, and the cornea and disc at corresponding positions on the film. If the disc surface has changed in the time period between exposures, the observer will perceive a stereo effect when viewing through a stereoscope. Goldmann discusses the efficiency in detecting both axial and frontal displacement, indicating that as the resolution of the Zeiss camera is 1.6 minutes of arc, 8μm of frontal displacement can be detected, and depth perception is approximately 7x worse than frontal change. This method is manual and requires subjective interpretation.

An alternative to stereometric analysis of the optic nerve head has been proposed by Nagin and Schwartz [Nagin85a,b]. Pallor is the region of maximal colour contrast in the optic disc. The ratio of area of pallor to optic disc size was found to be greater in ocular hypertensive patients. The retina, disc, and area of pallor correspond to 3 peaks in the intensity histogram, and the valleys between the peaks were used as a reference in a boundary tracking program. A boundary tracking technique was developed to delineate the optic disc and area of pallor. Points were tracked sequentially along a path which corresponded to one of the histogram valleys, had a steep gradient value, and was consistent with neighbouring results. Various tracking parameters were tuned empirically on 100 trial images. Red-filtered images were used for extracting the optic disc boundary, and green-filtered images were used for extracting the area of pallor. Operator intervention was required to guide the tracking process in difficult cases. The study shows the clinical significance of this type of measurement as an indicator of developing glaucoma.

A group at the Electrical Engineering Department, Texas Tech, have presented a number of papers at the Association for Research in Vision and Ophthalmology (ARVO) '86 and '87 conferences, involving microcomputer based image analysis of the fundus ([Ling86], [Whiteside86], [Mitra87]). Changes in the nerve fibre layer are detected by comparison of sequential images in the spatial and spectral frequency domain. A microcomputer-based system provides image database facilities, interactive histogram modification, pseudo-colour mapping, and image enhancement functions in a menu-driven environment. Algorithms for image registration and image differencing were designed for detecting nerve fiber layer damage. Although this work does not directly relate to optic nerve head analysis, it is part of an effort which is devoted to finding clinical indicators for developing glaucoma (see also [Peli89]).

Two papers on optic disc analysis using the PAR image analyser were presented at the ARVO '87 conference ([Spaeth87], [Varma87]). Spaeth *et al* conducted a study of optic disc vessel shift in glaucoma. The PAR image analyser was used to detect changes in the vasculature of glaucomatous discs and results were compared to the clinical observations of glaucoma experts. The PAR system was more sensitive than the clinical observation. Varma *et al* compared the results from topographic analysis of the optic disc using three different methods: clinical planimetry, the PAR image analyser, and the Rodenstock analyser. The study showed a high correlation between all three methods of analysis.

3.6 Reliability and Reproducibility

Shields [Shields89] and Peli [Peli89] give a discussion of the more recent developments in fundus imaging and measurement. New ophthalmic imaging systems, such as the Optic Nerve Head Analyser [Bishop87], the Retinal Analyser [Zimmermann88], or the Scanning Laser Ophthalmoscope (SLO) [Webb87], offer the clinician alternatives to acquisition, storage, and evaluation of fundus images. Depth resolution offered by the Optic Nerve Head Analyser varied on repeated measurements by less than 60μm. Reproducibility of repeated measurements of disc area on 10 images using the Retinal Analyser was reported to vary from 0.25 to 3.3% [Peli89]. With the Retinal Analyser system, the resolution on a model eye was reported to correspond to 20μm of depth, and the variability of depth measurements ranged from 0.7 to 25.5%. Reports on the reproducibility and reliability of measurements made using the SLO have yet to emerge. Interestingly, Peli [Peli89] concludes that "the clinical value of these instruments is still unknown, and the accuracy of these instruments and what accuracy is needed to provide meaningful clinical information are also unknown".

An interesting comparison of the equipment which was available for stereophotography of the optic disc in the late 1970's is given by Rosenthal [Rosenthal77]. A clinical comparison of results from photogrammetric analysis is presented where the images were taken from one of three systems: the Zeiss fundus camera using the Allen separator (non-simultaneous); the Donaldson stereo fundus camera; and the Zeiss fundus camera using the twin-prism separator. No discussion of the photogrammetric procedure is given regarding image resolution, algorithms, number of points sampled, etc. It was concluded that the Donaldson camera gave the best results and depth sensitivity ranged from 86 to 190µm.

A number of papers indicate the reliability of disc measurements as an indicator of glaucomatous development ([Johnson79], [Takamoto79]). Typically the Donaldson camera was used and images were processed manually using a stereoplotting device. In [Takamoto79], the stereoplotter was used to select reference points before applying a geometric transformation to align the images. After registration observations were made from 10 predetermined (high confidence) points in each image. The root-mean-square error was reported to be ±12µm frontally and ±19µm axially. It was not the intention of the study to reconstruct topographic information, but rather to determine the best photographic aperture and magnification that would give the most consistent results. A similar study was reported in [Rosenthal80] in which a Zeiss model eye was photographed using the Donaldson camera. Experiments were conducted to determine the impact of photographic and ocular variables on the photogrammetric estimates of a known cup depth. It was shown that the photogrammetric procedure tolerates refocusing, repositioning, and realignment of the camera, but underestimation and overestimation of cup depth was noted to be a function of the refractive error of the lens.

The availability of the Rodenstock Analyzer brought forth a number of reproducibility studies using the system in a clinical environment (eg. [Mikelberg84], [Mikelberg86], [Bishop87], [Caprioli87], [Kolli87], [Miller87], [Prince87], [Shapiro87]). The Rodenstock Analyzer (discussed earlier) calculates the depth of 1600 points with the help of microcomputer. In [Mikelberg84] (refer to figure 3.1) a description of the techniques for extracting clinical measurements from the resulting depth map is presented. The optic disc boundary is determined with the help of a human operator who selects four points at the horizontal and vertical disc margins to which the computer fits an ellipse (not all discs are elliptical!). The cup margin is then defined by subtracting 150μm (a clinically accepted value) from the defined disc, and the disc-cup ratio is computed by measuring the distance from the centre of the disc along a given meridian to a point matching the cup depth. This is repeated superiorly and inferiorly in a span of 50 degrees. The mean of the measurements defines the vertical disc-cup ratio. Neuro-retinal rim area and cup volume can be measured in a similar fashion. This study shows the potential of the Rodenstock Analyzer in the clinical environment.

[Shields87] reported a reproducibility study using the Rodenstock Analyzer. The key result was that reproducibility was found to be poor for cup volume, but good for disc-cup ratio and neural rim area. Subjective comparisons implied good reproducibility of the measurements and established the Rodenstock Analyzer as a clinically viable diagnostic tool.

3.7 Discussion and Analysis

A wide range of papers dealing with the general problem of optic nerve head analysis have been reviewed. The majority are presented from a medical viewpoint and therefore lack a great deal of technical detail about the image processing, image analysis, and photogrammetric algorithm development. It is clear from the literature that simultaneous stereo photographs are essential for reproducible results, and the predominating equipment was either the Donaldson fundus camera or the Rodenstock Analyzer. It is evident that fundus cameras correct for some of the refractive errors of the lens of the eye, but in an effort to obtain a larger stereo baseline, simultaneous stereo cameras are viewing through perhaps less than favourable portions of the lens, causing distortions which cannot be quantified. Moreover, even with the vast number of reproducibility studies in the literature, a standard measure by which to quantify changes of the optic disc has yet to emerge. The problem of analysis of the optic nerve head appears to be in its infancy when compared to other types of ophthalmologic diagnosis.

Perhaps the most interesting paper is that of [Algazi85], as it describes in some detail the digital stereo matching algorithm employed in reconstructing topographic information from stereo fundus images. Although some impressive results are shown, there are several problems with the complete algorithm. However, the paper presents a reasonable solution to the problem, with minor weaknesses in the image processing. With the automation of the alignment process (as in [Peli87]) it may be possible to automate the entire measurement process.

Lee and Brady have shown that by integrating binocular stereo with photometric stereo, reliable estimates of disc topography can be obtained [Lee91a]. The authors propose a new camera which will meet both the geometric requirements for binocular stereo and the illumination requirements for photometric stereo. Briefly, photometric stereo is a method by which the three-dimensional shape of an object is inferred from local scene irradiance. If the reflectance properties of the surface are known, then local measures of surface gradient can be obtained (see chapter 5 for more details). The authors indicate that the future direction of their work will be to concentrated on the image acquisition (with the construction of a new fundus camera) and not on the image analysis. Their study, although of great interest, does not greatly affect the research presented here.

3.8 Summary

It may be useful to review some of the major stumbling blocks for stereo analysis of the optic disc photographs, as extracted from the literature.

(1) Image registration - both non-simultaneous and simultaneous stereo photographs require some registration involving scaling, rotation, and translation.

(2) Intensity variations - radiometric differences between the stereo pair require some method of adjusting the brightness and contrast of the two images to improve cross-correlation measures.

(3) Focusing error - stereo fundus photographs, regardless of the type of camera, are often slightly out of focus, and some type of deconvolution may be necessary to correct for focusing errors.

(4) Lack of image detail - as noted previously, regions of non-uniform, low contrast texture predominate in the areas between the vascular structure. The area of interest, the optic disc, is often overexposed in the data collected and available here, providing little or no detail of the optic cup, and this will present problems for stereo matching algorithms.

(5) Arbitrary optical distortions - photographs are taken through the lens of the eye with all its subtle asymmetric optical distortions. Hence, obtaining reliable and reproducible disparity measurements is made more difficult.

(6) Clinical constraints - the ultimate goal of this research would be to provide a clinically valuable diagnostic tool. Generally, this means that a procedure must be applicable to standard fundus photographs; it is desirable that historical data still of clinical significance be examined, although aware that new data would be collected using the simultaneous method.

(7) Quantitative measures - as noted earlier, a standard measure for the quantification of optic disc change would be a desirable goal, and any progress toward such a measure would be of great value to clinicians.

The problem of topographic analysis of the optic nerve head from stereo photographs is by no means solved, even by commercially available systems. There are a lot of interesting aspects to this problem which are addressed in the literature. Although commercial systems are now becoming available which may solve some of the current problems, information about such systems is not easily obtained. It is hoped that this brief historical survey of stereoscopic fundus photography has given some insight into the difficulties that may arise in the application of digital stereo matching to the fundus data.

Chapter 4
Physical Constraints in Stereoscopic Fundus Imaging

4.1 Introduction

In this chapter a discussion of the physical constraints in stereoscopy of the human fundus is presented. Issues such as image acquisition, camera geometry, image resolution, and other physical constraints are investigated.

In order to understand the difficulties associated with stereo analysis of the optic disc the various physical constraints of the problem must be examined. The constraints imposed by the camera geometry, imaging through the pupil of the eye, as well as the small extent of the optic disc area, all present limitations to the accuracy to which stereo information can be extracted. It is important to quantify these limitations, which ultimately define the performance of stereo matching algorithms that can be used to automatically or semi-automatically compute topographic information about the optic nerve head.

There are many parameters which can affect the accuracy of any measurement of the optic disc. These include the size of the optic disc, the viewing distance, and the stereoscopic baseline, which define depth resolution; eye movements, motion blur, and optical distortions, such as defocus, which produce an optical degradation of the image; illumination changes, such as contrast and spatial lighting variations, between exposures; and parameters associated with any electronic components. Changes in any of these parameters complicate correlation of the stereo pairs. It is important to define the relationship between these parameters in order to assess the limitations of stereoscopic depth recovery.

4.2 Clinically Significant Change

It is natural to ask and to have the answers to several key questions before investigating the physical parameters which affect the accuracy of stereoscopic measurement of the optic disc:

1) What is the scale of change to the topography of the optic disc that clinicians are looking for ?
2) Are there known and accepted relationships between, for example, changes in optic disc diameter and changes in optic cup depth ?
3) What constitutes a clinically 'significant' change ?

The answers to these questions are, unfortunately, not quantitative. Even today, clinicians have yet to make a quantitative selection of the aspects of structural change that will be the most useful in detecting and following the progression of glaucomatous damage [Shields89].

As the size and shape of the optic nerve head in any individual is governed by many physiological factors, *absolute* measurements are not of much use for diagnosis. To differentiate the pathological from the physiological features, the patient's eyes must be examined periodically. If on successive examinations there are changes in the size of the cup or the diameter of the disc, then disease might be diagnosed. The diagnosis, however, is still very much *subjective.* Thus, the answer to the first question is not known.

The second question addresses the problem of how best to interpret and utilise any measurement data which is available to clinicians. There are at best *ad hoc* or rule-of-thumb interpretations of the types of optic disc measurements available today. Research work presented in the literature concerns itself mainly with the study of the reproducibility of a given measurement technique, and not with the relationships between accepted measurements and their clinical significance. Clinicians have yet to establish criteria to define the disc topography or other aspects of the optic nerve head that will allow for a clear distinction between normal optic nerve heads and those with developing glaucomatous damage [Shields89].

To answer the third question it is necessary not only to measure change, but also to associate with that change an understanding of its clinical significance. Throughout the literature (see chapter 3) 'significant' change has been defined only through a subjective interpretation of optic disc measurement. Worse, the accuracy of measurement that is needed to provide meaningful clinical information is still unknown [Peli89]. Because automated measurement of the optic disc topography is still a relatively new technique, there are few studies which provide sufficient long term results from which clinicians can begin to turn a heuristic or subjective interpretation into a more quantitative analysis of the data.

Although the scale of change that can be attributed clinically 'significant' is not readily available, some estimate of the magnitude of change can be derived from the data presented in the literature. For example, Algazi [Algazi85] presents a comparison of computed topography for the same patient, with one set of data obtained in 1974 and the other in 1978. Although the cross section of depth is given on a relative scale, it is shown that an approximate 20% change in the depth of the cup had taken place, and that this patient is known to have glaucomatous damage. In another example, Takamoto *et al* [Takamoto79] conclude that changes in optic cup width of ± 2 percent and changes in optic cup depth of ± 6 percent can be detected using standard photogrammetric techniques. These percentages correspond to changes in width of approximately 30μm and changes in depth on the order of 40μm. Takamoto does not, however, indicate how much change is clinically significant.

It is clear that there is no quantitative assessment of the scale of change that can be taken to be an indication of advancing glaucoma. However, it is still important to quantify the physical constraints in stereoscopic measurement of the optic disc so that lower bounds can be placed on any new clinical assessments that may emerge.

4.3 Stereo Image Geometry

In this section the concepts of stereoscopic imaging are introduced, the terminology is defined, and the physical constraints specific to stereoscopic fundus photography are discussed.

4.3.1 Introduction

Figure 4.1 represents two optical systems, O_1 and O_2, each with a focal length f in a configuration suitable for stereoscopic imaging. The optical axes of both systems are parallel with a separation B and lie in the X–Z plane (the Y-axis is perpendicular to the page). The coordinate systems X_1–Y_1–Z_1 and X_2–Y_2–Z_2 are defined for these two optical systems respectively. Given a point P with coordinates (x, y, z) situated in front of these two systems, the image of P is located at the horizontal coordinates x_1 and x_2 in each system respectively. The relative displacement of x_1 and x_2 with respect to their coordinate systems provides stereoscopic information about the three-dimensional coordinate of the point P.

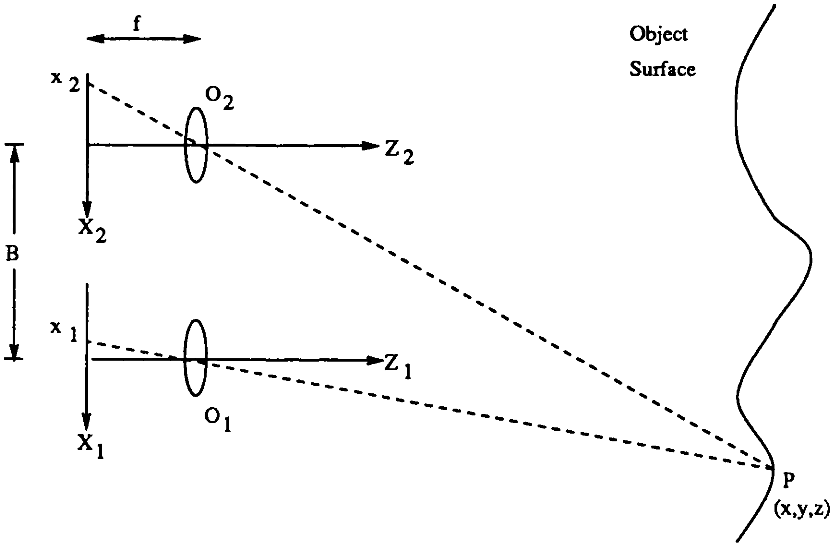


Figure 4.1. The basic geometry of a stereo imaging system.

The geometric relationship between the coordinate systems X_1–Y_1–Z_1 and X_2–Y_2–Z_2 can be defined as:

$$X_1 = X_2 + B \ , \quad Y_1 = Y_2 \ , \quad \text{and} \quad Z_1 = Z_2 \ . \tag{4.1}$$

If a three-dimensional coordinate system is defined relative to O_2, which has its origin at the point $(X_2, Y_2, Z_2) = (0, 0, 0)$, then from similar triangles in figure 4.1, a relationship can be defined between the image coordinates x_1 and x_2, and an object point (x, y, z) as

$$x = \frac{z \, x_2}{f} \tag{4.2}$$

and

$$x = \frac{z \, x_1}{f} + B \ , \tag{4.3}$$

which can be equated to yield

$$x_2 = x_1 + \frac{f \, B}{z} \ . \tag{4.4}$$

Equation (4.4) provides a constraint on the relationship between two points in the image planes. For an image point at x_2, its corresponding image point at x_1 can be inferred if the geometric variables z, f and B are known. Often, there is a known range over which z may vary, which places a constraint on the possible value of x_1 given x_2. In systems which can be calibrated such that f and B are known and z has a known range, the process of stereo matching (locating the corresponding points x_1 and x_2) can be greatly simplified by utilising the constraint specified by equation (4.4). Note that in this coordinate system $Y_1 = Y_2$. This is often termed *epipolar* stereo to reflect the fact that the two optical systems are vertically aligned and that there is no stereoscopic shift in the Y axis.

Equation (4.4) can be re-arranged as

$$z = \frac{f \, B}{x_2 - x_1} \ . \tag{4.5}$$

Equation (4.5) now gives a direct relationship between the distance to the object point P and the difference in the image coordinate positions x_1 and x_2. To determine the three-dimensional coordinate (x, y, z) of the object point P, image points x_1 and x_2 must be determined. Given x_1 and x_2, z can be calculated using equation (4.5) (assuming f and B are known). The x-coordinate can be obtained by substituting z into either equation (4.2) or (4.3). The y-coordinate, in this case, is relative to Y_1. This procedure allows us to calculate the x-, y-, and z-coordinates of every point on an object that images to a point in both systems O_1 and O_2.

4.3.2 Geometric Optics

Figure 4.2 diagrams the geometric relationships in fundus imaging, and will be referred to in sections 4.4 and 4.5. In this section the various parameters which describe the imaging process will be defined, and their use in defining the depth resolution of a stereoscopic system is presented in the following two sections.

Figure 4.2 indicates how an image of the optic disc is obtained in a typical fundus imaging configuration. A positive lens is placed in front of the eye and focused to form an image of the retina on the image plane. The point K represents an object point in the plane of best focus, which forms an image point k on the image plane. A nearby object point J which lies a distance Δz from the plane of best focus is imaged to a spot j on the image plane. The diameter of the spot image j increases as Δz increases, due to blurring. The effects of blurring are discussed in section 4.4. Note that if the diameter of the image spot j increases to the point where it overlaps with the image point k, then j becomes indistinguishable from k. The minimum distance at which two object points are still distinguishable in the image plane is termed the effective lateral resolution, or h_{eff}. The parameters which will determine the lateral resolution are discussed in section 4.4.

Behind the imaging lens lies a stop or aperture which controls the amount of light reaching the image plane. The choice of the size of aperture yields a tradeoff between spherical aberrations and diffraction effects. Spherical aberrations are caused when a beam of light parallel to the axis passes through a lens with

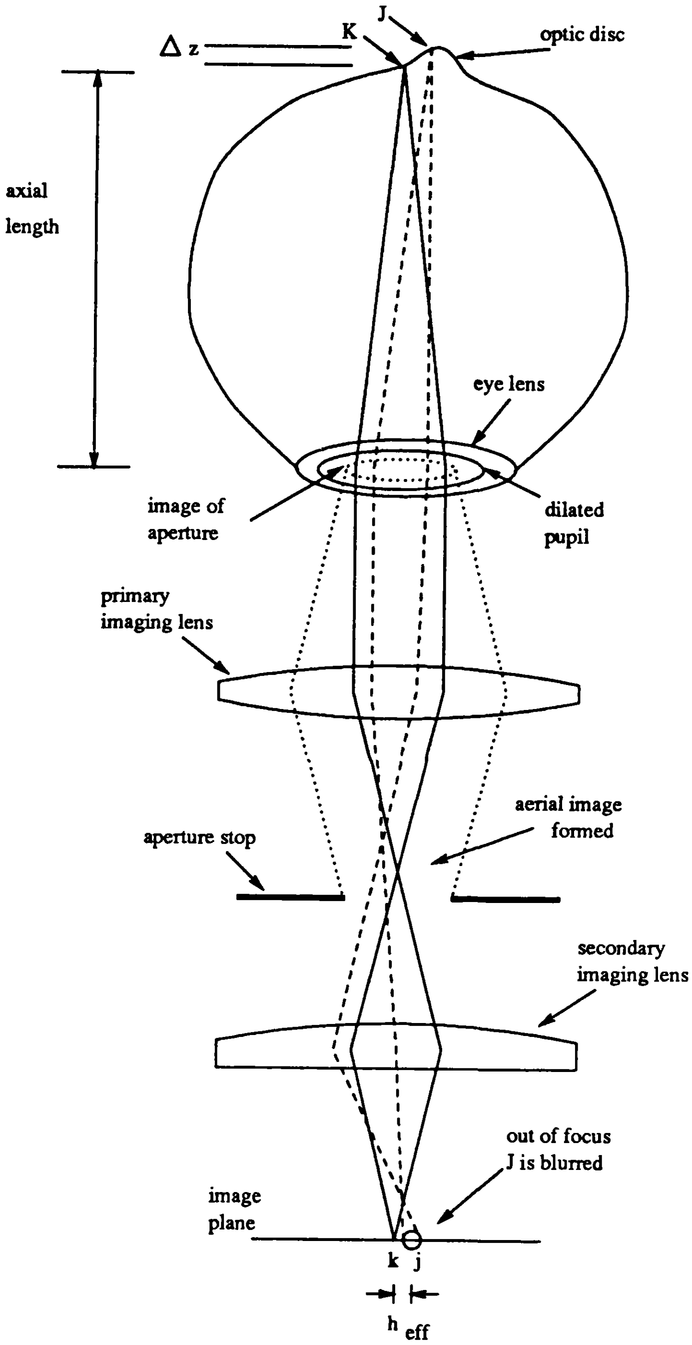


Figure 4.2. The geometric relationships in fundus imaging.

spherical surfaces, and rays passing through the outer zones of the lens come to a focus nearer the lens than the rays through the central zone. Consequently, the image is unsharp wherever the image plane is placed. This failure of the lens to form a point image of a point object is termed spherical aberration. Spherical aberration can be reduced with the use of a stop or aperture, which blocks the rays from the outer zones of the lens. Lens also exhibit other aberrations, but they can be neglected in the present context.

Diffraction is the name given to the phenomenon which occurs when light passes through any aperture. Even with an ideal aberration-free lens, the image of a point object is not a point, but a circle of light of finite diameter. The diameter of the circle is called the *Airy disc*. Although it is desirable to decrease the aperture to reduce spherical aberrations, diffraction effects increase as aperture decreases. As shown in section 4.4, depth of focus improves as the diameter of the aperture decreases. The aperture will form an aerial image on the cornea of the eye, and the aperture must fit within the eye's pupil: indeed, <u>two</u> such aerial images of the aperture must fit within the pupil if adequate stereoscopic imaging is to occur. The pupil is normally pharmacologically dilated, and therefore the size of the aperture is limited to the size of the pupil (typically 1-3mm undilated, 6-8mm dilated).

The axial length of the eye can vary from 22 to 26mm, though the apparent length of the eye, due to the focal power of the eye's lens and various other factors such as corneal curvature and the aqueous nature of the eye, is assumed to have a focal distance of approximately 17mm [Littmann82]. Light rays from a point on the retina which lies at the focal distance (e.g. the point K in figure 4.2) will pass through the lens of the eye and emerge as parallel rays. Points on the retina not at the focal distance (e.g. J in figure 4.2) will produce either diverging or converging rays from the lens of the eye depending on whether they are in front of or behind the focal plane respectively.

4.3.3 Physical Constraints

Figure 4.2 indicates how a single image of the optic disc can be formed on an image plane given a suitable optical system. In a stereoscopic imaging system, such as that diagrammed in figure 4.1, two such images must be taken at a finite separation or baseline B to produce a stereoscopic effect. However, the finite diameter of the pupil will constrain both the aperture a and the stereoscopic baseline B, and an optimal aperture-baseline relationship must be given.

As a published design of a simultaneous stereoscopic fundus camera, the Donaldson camera [Donaldson64] serves as a suitable example through which to examine the problem of obtaining two disparate images of the optic disc through a pupil of finite diameter (recall from chapter 2 that the pupil is typically pharmacologically dilated to 6-8mm in diameter).

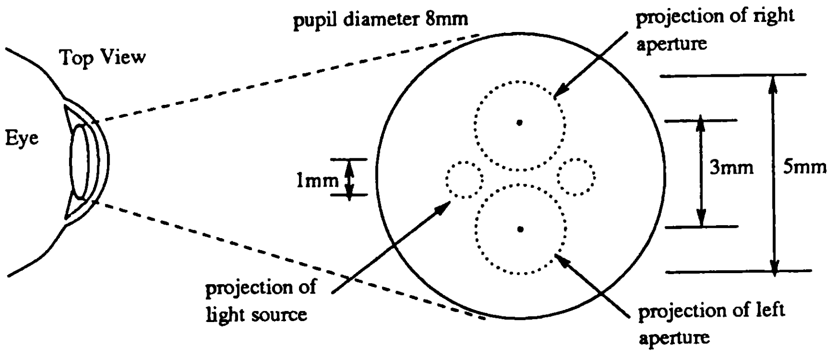


Figure 4.3. The relationship between aperture, illumination source, and pupil diameter with the Donaldson stereoscopic fundus camera.

In both non-simultaneous and simultaneous stereoscopic fundus photography, the combined diameters of two apertures together with an illumination source should remain entirely within the diameter of the pupil. In the first

Donaldson camera, for example, the apertures are fixed at 5mm with a baseline separation of 9mm. With a magnification ratio of 3:1, the dimensions of the apertures when imaged onto the pupil are 1.7mm, with a combined width of less than 5mm. With limited space for the illumination source, two 3mm light sources are imaged onto the pupil at the 12 o'clock and 6 o'clock positions. Figure 4.3 shows the relationship between dual aperture, illumination source, and pupil diameter.

It is desirable that the aperture of the fundus camera and the light source (see section 4.6) should remain entirely within the dilated pupil. This places a severe constraint on the size of camera aperture that can be used effectively. When two imaging systems are used, the baseline should be as large as possible so as to create the maximum stereoscopic effect. To maintain a large stereoscopic baseline while keeping both apertures within the pupil places a constraint on any stereoscopic system for imaging the optic disc. The optimal aperture-baseline relationship can be given as

$$a = P - B \tag{4.6}$$

where a is the diameter of the aperture, P is the diameter of the pupil, and B is the stereoscopic baseline. The effect of increasing the baseline B on stereoscopic depth resolution is discussed in section 4.5.

4.4 Lateral Resolution

In this section the various parameters which limit the effective lateral resolution of an imaging system are described. The lateral resolution will directly affect the depth resolution of a stereoscopic system, discussed in section 4.5, and it is therefore important to quantify the many factors which may limit lateral resolution in stereoscopic fundus imaging.

4.4.1 Optical Diffraction

Optical diffraction occurs because light consists of waves which are scattered by the edges of lenses or optical stops. It is therefore not possible, even with perfect optical systems, to image infinitesimally small details.

Assuming a circularly symmetrical, diffraction-limited imaging system, the diameter of an image point is given by the diameter between the first minima of its Airy disc and may be defined as:

$$C \approx 1.22 \frac{\lambda R}{a} \tag{4.7}$$

where,

C : diameter of the focussed image of a point object

λ : wavelength of visible light (400 - 700nm)

R : range or object distance (for the human eye, average value 17mm)

a : aperture diameter (in fundus cameras, 2mm to 5mm)

For example, using typical values for λ=500nm (yellow-green light), R=17mm, and a=3mm, the diameter of an aberration-free image point of an object in the plane of best focus will be approximately

$$C \approx 1.22 \frac{(500\text{nm})(17\text{mm})}{(3\text{mm})} \approx 3.5\mu\text{m}$$

This gives an indication of the best lateral resolution possible in the plane of best focus in the absence of any other limiting factors on lateral resolution (see [Jenkins51], [Morton84]). The effect of diffraction on lateral resolution is denoted h_{diff}.

4.4.2 Defocus Effects

It is only possible to focus an optical system sharply on one plane in space at a time; if the rim of the optic disc is sharply in focus then the bottom of the optic cup will be partly out of focus, and *vice versa*. As a result, objects which lie away from the plane of best focus will be blurred when imaged onto the image plane (e.g. the object point J in figure 4.2). From similar triangles in an object-lens-image diagram (see figure 4.2), the diameter of the blur circle on the image plane when an object point is Δz µm from the in-focus distance at range R can be approximated [Morton84] by:

$$b = \frac{a \ \Delta z}{R} \tag{4.8}$$

where

b : lateral diameter of blur circle
a : aperture diameter
Δz : out-of-focus object distance from R
R : object distance at the plane of best focus

For example, the aperture a of the fundus camera may be set to 3mm, the range R is approximated to 17mm, and values for Δz range from 100-500µm for typical cup depths. The diameter of the blur circle of an image point when the object point lies, for example, $\Delta z = 100\mu m$ from the plane of best focus would then be

$$b = \frac{(3\text{mm})(100\mu\text{m})}{(17\text{mm})} = 17.65\mu\text{m}$$

The effect of defocus on lateral resolution is denoted h_{defoc}.

4.4.3 Eye Movement

It takes a finite time to form an image of the optic disc on film placed at the image plane. During that time, the eye being examined will undergo involuntary movements such as saccades, drift, and tremor. Such eye movements will blur the image and will cause distortion. The amount of motion blur can be estimated from the typical values of about 10 arc-minutes/second for eye drifts, and about 6 arc-minutes/second for saccades [Pirenne67]. With photographic exposure times on the order 1/250 second, for example, there will be less than 2µm root-mean-square movement of the retina. For this reason the effect of eye movement on lateral resolution for the data used in this research can be neglected.

In general, when the image is formed on a medium other than film (e.g. sampled directly using a video camera or other recording device), the effects of eye movement may not be neglected. For completeness, then, the effect of eye movement on lateral resolution is denoted h_{move}.

4.4.4 Discrete Sampling

The average horizontal diameter of the optic disc is approximately 1.8mm, and the average vertical diameter is approximately 1.9mm. Therefore, a typical image of the optic disc will cover a $2\times 2\text{mm}^2$ area of the retina. If a discrete image of the optic disc is to be formed, then a series of discrete samples must be taken across the imaged area. On a 35mm photograph, for example, the discrete samples are composed of grains of silver halide crystals, and the resolution of the film is on the order of 500 grains per millimeter (i.e. 1000 samples horizontally across the optic disc at unit magnification). If the optic disc image is scanned, for example, using a monochrome Vidicon camera to form an analogue signal, and the signal is then digitised to form a discrete image, then the resolution is limited to that of the digitiser used. Digitisers with a resolution of 512×512 picture elements (pixels) are readily available, though higher resolution devices could be used if required. More generally, the lateral resolution of a sampled image is given by

$$h_{\text{sample}} = \frac{(\text{width of the area imaged})}{(\#\text{ of discrete samples})} \ . \tag{4.9}$$

For example, if a Vidicon camera is used with a digitiser at a sampling resolution of 512 pixels, the lateral resolution due to sampling would be

$$h_{\text{sample}} = \frac{2\text{mm}}{512\text{ samples}} = 3.9\mu\text{m}/\text{sample} \ .$$

This example is given because the digitiser available for use on the data for this research has a resolution of 512×512 pixels.

When digitising an image from 35mm film, it should be noted that as image resolution increases the lateral resolution is limited to the film grain resolution.

In fundus camera systems which are directly connected to a digitiser, or which provide analogue video output (e.g. the Scanning Laser Ophthalmoscope - see chapter 3), the digital resolution of the device is a constraint.

The effect of sampling on lateral image resolution is denoted h_{sample}.

4.4.5 Optical Defects of the Eye

The human eye is not a perfect imaging system; the eye itself will degrade the lateral resolution of the image due to the varying defects of the eye lens, the curvature of the cornea, and the axial length of the eye. For example, the focal power of the eye (the eye's own magnification) is determined by the focal power of the lens of the eye and the curvature of the cornea at the viewing position. It is known that the focal power of the eye can vary by as much as ± 2 diopters across the cornea [ElHage73]. When viewing the eye stereoscopically, both views will therefore not be subject to the same focal power (magnification) or refractive error.

Littmann [Littmann82] has shown that even with a value for the axial length of the eye (determined by ultra-sound) and a measure of the corneal curvature (determined from keratometer readings), that only an estimate of the focal power of the eye can be made. Therefore, an average value for the focal power of the eye is most often used. The average or "idealised" eye is assumed to have a focal length of 17mm and no refractive error. Given that the optical defects for an individual's eye are difficult to determine, the degradation of lateral resolution due to the optical defects of the eye will be ignored.

4.4.6 Total Lateral Resolution

To a first approximation the lateral resolution contributions are taken to be additive:

$$h_{\text{eff}} \approx h_{\text{diff}} + h_{\text{defoc}} + h_{\text{move}} + h_{\text{sample}} \ . \tag{4.10}$$

Equation (4.10) gives an approximation for the effective lateral resolution of a single image of the optic disc. Although it is only an approximation, any other possible sources of degradation (e.g. spherical aberration) in lateral resolution can be neglected. The term *effective* lateral resolution is used to imply that there may be a more accurate means by which to determine the *total* lateral resolution.

For example, suppose an imaging system is used with an aperture a of 3mm, the pupil is dilated to 8mm, the focal length of the eye is taken as the average (R = 17mm), and the stereoscopic baseline B, given by equation (4.6), is 5mm. A digitiser with a sampling resolution of 512×512 pixels is used to form an image. For points in the image which lie in the plane of best focus,

$$h_{\mathrm{diff}} = 3.5\mu\mathrm{m} \ ,$$

$$h_{\mathrm{defoc}} = 0 \quad \text{(i.e. the plane of best focus)} \ ,$$

$$h_{\mathrm{move}} = 0\mu\mathrm{m} \quad \text{(little or no eye movement)} \ ,$$

and

$$h_{\mathrm{sample}} = 3.9\mu\mathrm{m}/\mathrm{sample} \ .$$

Thus the effective lateral resolution would be

$$h_{\mathrm{eff}} \approx 7.4\mu\mathrm{m} \ .$$

Similarly, for object points which lie, for example, 100μm from the plane of best focus, there is now a degradation in lateral resolution due to blurring, and thus

$$h_{\mathrm{defoc}} = 17.65\mu\mathrm{m} \ ,$$

which now yields

$$h_{\mathrm{eff}} \approx 25\mu\mathrm{m} \ .$$

When two images are taken stereoscopically, the effective lateral resolution of each image affects to the resolvable depth of the stereoscopic system. In the following section the depth resolution of a stereoscopic system is discussed.

4.5 Depth Resolution

In a stereoscopic system the accuracy to which the depth of an object can be measured is determined by the accuracy to which that object can be located in each of the stereoscopic images, and the separation or baseline between the two images (refer to figure 4.4).

The depth resolution of a stereoscopic system is, to a first approximation, defined by

$$D \approx \frac{h_{\mathrm{eff}}\, R}{B}\,, \tag{4.11}$$

where

D : resolvable depth increment at range R

h_{eff} : effective lateral resolution

R : range or object distance

B : camera separation or stereoscopic baseline

Figure 4.4 indicates the relationship between depth resolution, effective lateral resolution, object distance and camera baseline. The effective lateral resolution and the baseline separation have a large effect on the depth resolution of a stereoscopic system, as seen from equation (4.11). In this application, the object distance, R, is relatively constant and is beyond our control. Too small a baseline, for example, can degrade the depth resolution to the point where no accurate depth measurements can be made. Similarly, if the effective lateral resolution is poor (e.g. due to excessive blurring) the depth resolution will be degraded.

For example, if the effective lateral resolution in the plane of best focus (as shown in section 4.4) is 7.4μm, the object distance is 17mm, and the stereoscopic baseline is 5mm, then the depth resolution will be

$$D = \frac{(7.4\mu\mathrm{m})(17\mathrm{mm})}{(5\mathrm{mm})} = 25\mu\mathrm{m}\,.$$

If the lateral resolution is degraded, due to blurring, to 25μm, then the depth resolution would become

$$D = \frac{(25\mu\mathrm{m})(17\mathrm{mm})}{(5\mathrm{mm})} = 85\mu\mathrm{m}\,.$$

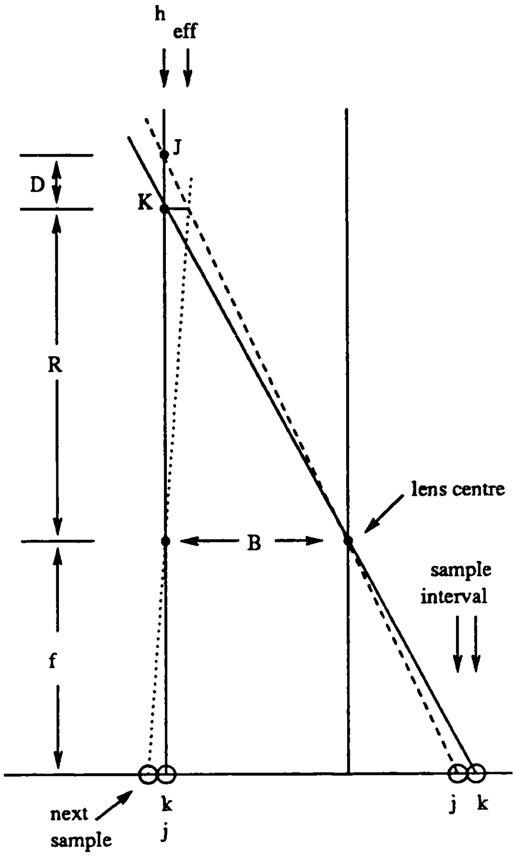


Figure 4.4. Diagram showing the relationship between D, h_{eff}, R, and B from equation (4.11). As h_{eff} decreases, the image point j becomes more visibly separate from k, improving the resolvable depth D between J and K.

In this example, subtle changes in depth on the surface of an object less than 85µm can not be detected, at least for small object details.

In the previous section the various parameters which contribute to the effective lateral resolution were discussed. Figures 4.5 and 4.6 plot a number of these parameters versus depth resolution, indicating the sensitivity of depth resolution to diffraction effects, optical defocus, sampling effects and stereoscopic baseline.

From figure 4.5 it is clear that the ability to measure depth accurately decreases as one moves from the plane of best focus. The depth resolution

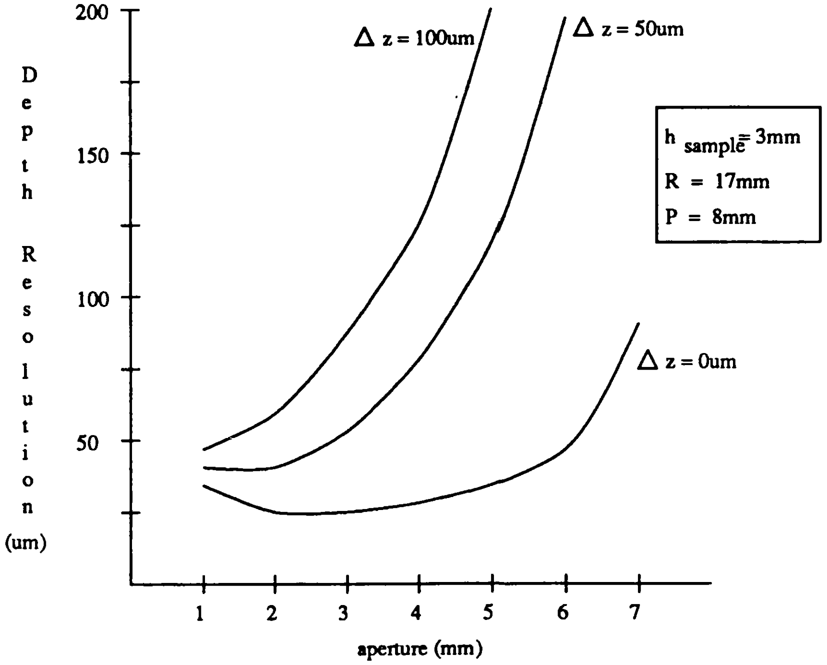


Figure 4.5. Graph of resolvable depth versus aperture size in a stereoscopic system, assuming a stereoscopic baseline determined by equation (4.6), with a pupil diameter of 8mm. Both h_{diff} and h_{defoc} are affected by a change in aperture.

deteriorates (increases) in a non-linear fashion, and is dependent on the depth of field (the aperture of the camera) and the object distance. As the object distance is nearly constant (i.e. 17mm), the amount of defocus will be proportional to the distance from the plane of best focus. With cup depths ranging from 100 to 500μm (or larger) one can expect the largest degradation in lateral resolution to be due to optical defocus.

From figure 4.6 it is clear that the depth resolution of a stereoscopic system is not as sensitive to the sampling resolution as it is to the amount of optical defocus. With the types of digitisers available today, fairly high resolution digitisation (e.g. 1024×1024 pixel resolution) of the image should be feasible. The above discussion applies to depth discrimination between an in-focus point at

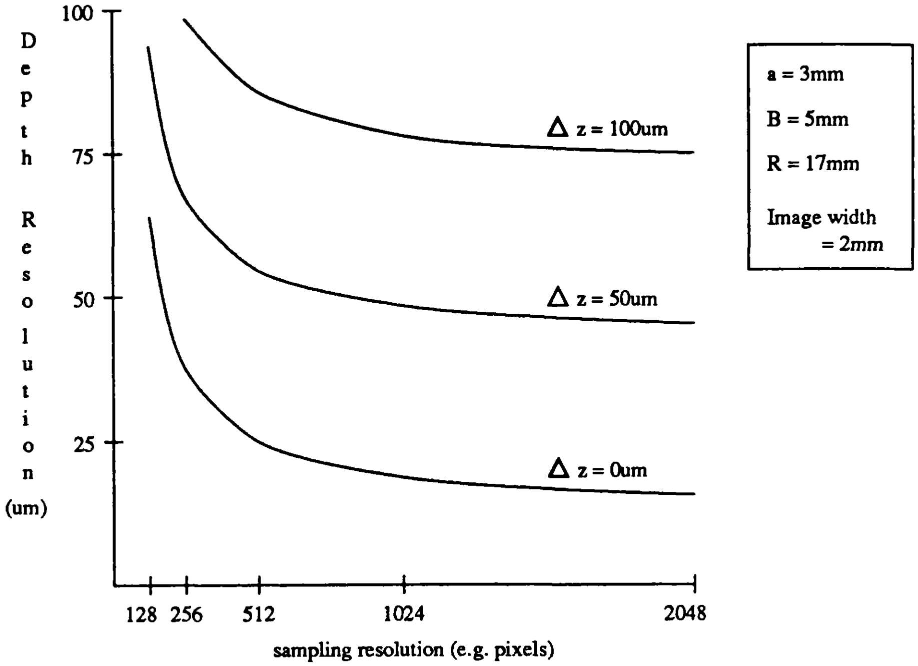


Figure 4.6. Graph of resolvable depth versus sampling resolution in a stereoscopic system.

range R, and an out-of-focus point at range $R + \Delta z$. In practice, as is shown in chapter 6, stereo matching which uses information from more than one image point at a time (i.e. template matching) can perform much better than the predicted depth resolution shown in figures 4.5 and 4.6. This point will be addressed more fully in chapter 6 in the presentation of the stereo matching results. In general, the surface of the optic disc is assumed to vary continuously in small increments of depth relative to the sampling resolution. Thus, large discontinuous changes in depth on the order of 100μm, as given in the example, are unlikely to occur.

4.6 Illumination

Illumination plays an important role in the depth recovery process. Donaldson [Donaldson64] discussed the problems associated with illuminating the retina with sufficient light to provide good quality photographs while maintaining light levels which would not damage the retinal cells. The development of illumination systems in fundus cameras has been a trade-off between light level and flash duration. The anatomy of the living eye places severe restrictions on the way the fundus may be illuminated and viewed. As the fundus must be illuminated and viewed through the same small aperture (the pupil), the viewing system must provide a means for avoiding reflections from the optical surfaces of the eye [Woon90]. Typically, Gullstrand's principle [Gulstrand11] is applied in the design of fundus cameras so that the paths of the illuminating and viewing rays are kept separate at the pupil. In a standard fundus camera, for example, the light used to illuminate the fundus passes through an aperture with its centre occluded, and the illuminating light passes through an outer annulus at the pupil (see figure 4.7). Hence, the illumination beam is kept away from the center of the cornea.

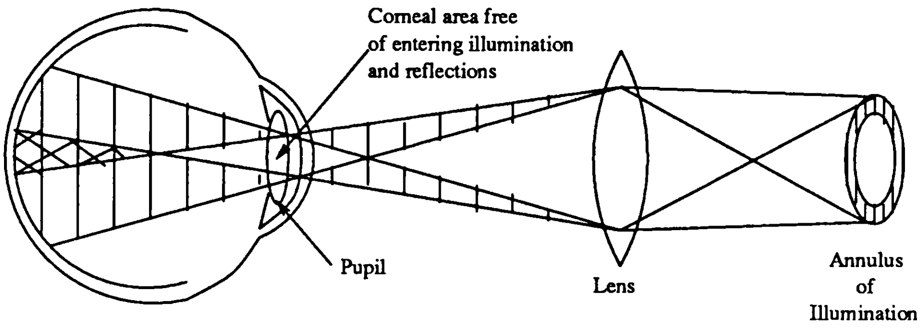


Figure 4.7. Diagram showing a typical fundus camera illumination system following Gullstrand's principle.

Relatively high levels of retinal irradiance are required by fundus cameras because of their low light-gathering efficiency (i.e. small aperture). If the illuminating and viewing areas of the pupil could be interchanged, or were it not a requirement to adhere to the Gullstrand principle, then a much higher proportion of the incident light could be collected to image the fundus. The efficiency of utilisation of light would be increased and lower levels of illumination could be employed [Woon90].

If the illumination is poor, the effective lateral resolution (h_{eff} described in section 4.4.6) will deteriorate due to the low image contrast (image points must be *visibly* separate in order to be observed stereoscopically). Variations of illumination between exposures can also complicate the stereo matching process, which then must cope with the radiometric shift between the two images. For the data and fundus camera which are discussed in this thesis, it will be assumed that these effects are minimal and can be compensated for by applying a simple radiometric correction to one of the images. Image pairs with large lighting variations, perhaps caused by artifacts in the patient's lens, are not considered. Most stereo matching algorithms are capable of dealing with moderate changes of illumination between the images (see chapter 5).

4.7 Relative Orientation

In the absence of any measurable constraints on the camera geometry in, for example, non-simultaneous stereo photography, some information about the camera positions can be obtained by means of a photogrammetric procedure known as analytic relative orientation [Wolf83]. It might be used as a method for determining the orientation parameters of the two camera positions when such information is not readily available. If an estimate of the displacement of camera O_1 relative to camera O_2 can be found (see figure 4.1), then a geometric constraint on conjugate image points exists, given by equation (4.4).

Analytic relative orientation is a procedure by which the relative camera orientation parameters are expressed for the first (right) view in terms of the second (left) view. In other words, if the left view is considered to define a coordinate system for observing an object, and the left camera position is at (0, 0, 0) in that system, then the position of the right view in terms of this coordinate system is required. The right view will, of course, define its own coordinate system, so it is important to express the position of the right view as a transformation from the coordinate system of the left view. This transformation will involve translational as well as rotational parameters.

Note that these positional parameters are only relative; that is, camera 1 relative to camera 2. The procedure is founded in the basic photogrammetric equation of collinearity, where the perspective origin or viewing position, the image or photograph point, and the object point must all lie on the same line (see figure 4.1). If two views of the same object are obtained, as in stereo, then the line given by image point, lens centre and object point defined for the first camera must intersect the similar line defined for the second camera at the object point. Indeed, it is this relationship from which equation (4.4) is derived.

Typically, the positional parameters of the second camera relative to the first camera to be determined are:

X_1, Y_1, Z_1 : camera baseline or translation
ω : rotation about the x-axis
ϕ : rotation about the y-axis
κ : rotation about the z-axis

and each object point will have positional parameters:

P_x, P_y, P_z : object coordinates

The collinearity condition yields two equations from each image, one in x and one in y, giving four equations for each object point. Since there are six unknown positional parameters for the camera and three positional parameters for each object, by using at least six object points (ie. a description of the absolute location in space of the object points) a solution can be obtained by the least-squares method. Details of the least-squares method for obtaining a solution will not be discussed here: it is a fairly standard photogrammetric exercise (see [Wolf83]).

To set up the solution matrix for the set of collinearity equations, initial estimates of the parameters are needed. The translational parameters can be initialised as X_1 = approximately the known baseline, $Y_1 = 0$, and $Z_1 = 0$ (this simple configuration is diagramed in figure 4.1). These estimates can be scaled from the photographs or computed from the basic parallax equation, given by equation (4.5). The rotational parameters are typically set to 0 initially, since a parallel view is assumed. However, the solution is very sensitive to the initial estimates for the object positions. The estimates for P_x and P_y can be calculated from P_z.

It can be shown that using the parallax equation to compute initial estimates for the object positions is not sufficient. However, even given good initial estimates for the object coordinates, the highest source of error lies in the accuracy of the match point coordinates from the two photographs or images. As shown in section 4.5, image points which lie in the plane of best focus have a finite accuracy laterally and in depth. If all match points lie in the plane of best focus then a small error is expected in defining a match between the image pairs. This will affect the initial estimates for the object positions, and in turn the accuracy of the resulting orientation parameters. As it is nearly impossible to select all match points to lie in the plane of best focus (giving a fixed and quantifiable error), match points will be chosen with varying amounts of error in their disparity, which will have unsatisfactory effects on the stability of the solution.

It can also be shown that several of the orientation parameters (eg. stereoscopic baseline vs. camera tilt) are not linearly independent, thus making a unique solution difficult to obtain using the least-squares method. In other words, the least-squares solution to a system of equations (in this case equations specifying the orientation of the cameras with respect to one another) cannot be unique if any of the variables are linearly dependent. Constraints on the solution can be applied, but the solution is still not guaranteed to be unique ([Golub83], [Ralston85]).

Analytic relative orientation requires further study before it can be used effectively as a means to calculate the position of the right camera view relative to the left camera view in non-simultaneous stereoscopic fundus photography.

4.8 Non-simultaneous Stereo Imaging

A number of factors that determine the ability to recover depth information from stereo photographs have been discussed. Perhaps the most important parameter is the recoverable depth D from equation (4.11). This equation is derived from similar triangles with a camera geometry that assumes two parallel views of the object. However, if the change of camera position from camera 1 to camera 2 (right view to left view) is other than simply a lateral shift, or if the second camera is oriented with its axis at an angle to that of the first view, then the equation for determining the resolvable depth becomes more complicated (hence the need to address the procedure of relative orientation in the previous section).

If the photographic procedure requires non-simultaneous image acquisition (as for Data Set B, discussed in chapter 2), then the amount of camera shift (baseline) varies with each stereo pair. The fundus camera design allows for shifting and rotating the camera position in order to allow for arbitrary angles for viewing the retina. The operator may, through unnecessary adjustment of the camera alignment and focus, minimise the stereoscopic effect. In addition, the movement of the patient's eye between exposures (moving the camera and alignment for the second photograph can take up to a minute, during which time the patient must remain fixated) adds an unknown amount of shift to the stereoscopic baseline.

4.8.1 Depth Resolution

One set of non-simultaneous stereoscopic fundus data was available for study in this research; namely Data Set B. The 35mm photographs for Data Set B were obtained using a standard fundus camera, as outlined in chapter 2. Figure 4.8 lists the values of the geometric parameters that determine effective lateral resolution, and gives an indication of the expected depth resolution for

Data Set B. The digitiser available for this research has a sampling resolution of 512×512 pixels.

Parameter	Value	Comment
a	3 mm	set by clinician
B	3 mm	set by clinician
R	17 mm	(average eye) standard value
h_{diff}	3.45 μm	determined by aperture
h_{defoc}	1.76 μm	e.g. at $\Delta z = 10\mu m$
h_{defoc}	8.82 μm	e.g. at $\Delta z = 50\mu m$
h_{defoc}	17.64 μm	e.g. at $\Delta z = 100\mu m$
h_{move}	0 μm	assuming no eye movement
h_{sample}	3.90 μm	(512 pixel resolution) variable
h_{eff}	7.35 μm	at $\Delta z = 0$ μm
h_{eff}	9.15 μm	e.g. at $\Delta z = 10$ μm
h_{eff}	16.17 μm	e.g. at $\Delta z = 50$ μm
h_{eff}	24.99 μm	e.g. at $\Delta z = 100$ μm
D	41.68 μm	points in focus
D	51.85 μm	e.g. points at $\Delta z = 10\mu m$
D	91.66 μm	e.g. points at $\Delta z = 50\mu m$
D	141.64 μm	e.g. points at $\Delta z = 100\mu m$

Figure 4.8. Geometric parameters for Data Set B.

4.8.2 Discussion

The non-simultaneous photographic method for obtaining stereo pairs cannot provide a *consistent* stereoscopic baseline. This lack of a consistent stereoscopic baseline not only complicates the stereo matching or depth recovery process, but also will make accurate comparisons between successive depth measurements difficult ([Algazi85], [Donaldson64]). A stereo pair taken at time t will produce different depth values from that of a pair taken at time $t+1$, if the stereoscopic baseline is different at time $t+1$. If the baseline is known for both stereoscopic pairs, this difference can be taken into account. When the baseline is not known, accurate comparisons of time-series stereoscopic data is not possible.

The ideal conditions for obtaining non-simultaneous stereo photographs of the optic disc would include:

- the patient's head does not move, and fixation is maintained;
- the camera is shifted laterally to obtain the second view, minimising rotation effects, and maintaining a nearly co-planar geometry;
- camera refocus is not significant for the second view;
- a consistent stereoscopic baseline is used.

These conditions cannot be met for several reasons: 1) satisfactory non-simultaneous stereoscopic photographs cannot be obtained in patients who cannot fixate accurately, and fixation is impossible to maintain between exposures in patients who cannot cooperate, 2) the stereoscopic effect can be poor or completely lost because of changes of various factors between exposures such as focus, position of the patient's head, and operator misalignments, and 3) a dependable, consistent depth effect cannot be obtained in photographs taken subsequently as it is difficult (if not impossible) to duplicate the exact degree of camera shift (stereoscopic base), patient head position, or fixation.

The ideal conditions for obtaining stereoscopic photographs of the optic disc for time-series analysis via the non-simultaneous method, in short, cannot be obtained. However, this does not preclude the use of non-simultaneous stereoscopic photographs for clinical analysis if some of the sources of error can be minimised or quantified. For example, in the acquisition of Data Set B (see chapter 2), the patient fixation was carefully controlled (using a bite-bar), there was no refocusing for the second view, and the stereoscopic baseline was carefully monitored, minimising the error inherent in the method.

4.9 Simultaneous Stereo Imaging

There are a number of benefits to be gained by utilising a fundus camera which can obtain simultaneous stereo images. The issues about camera geometry are shifted from consistent baseline, refocus, and patient movement in the case of non-simultaneous imaging, to a more detailed analysis of the physical and optical constraints in the case of simultaneous imaging. The first, and best known, stereoscopic fundus camera is reported by Donaldson [Donaldson64], in which many of the key physical and optical problems associated with simultaneous stereo imaging of the fundus are solved.

4.9.1 Depth Resolution

Several sets of simultaneous stereoscopic fundus data were available for study in this research, an example of which is Data Set C. The 35mm photographs for Data Set C were obtained using the Donaldson stereoscopic fundus camera, as outlined in chapter 2. Figure 4.9 lists the values of the geometric parameters that determine effective lateral resolution, and gives an indication of the expected depth resolution for Data Set C. The digitiser available for this research has a sampling resolution of 512×512 pixels.

Parameter	Value	Comment
a	2 mm	fixed in the camera
B	3 mm	fixed in the camera
R	17 mm	(average eye) standard value
h_{diff}	5.18 μm	determined by aperture
h_{defoc}	1.17 μm	e.g. at Δz = 10μm
h_{defoc}	5.88 μm	e.g. at Δz = 50μm
h_{defoc}	11.76 μm	e.g. at Δz = 100μm
h_{move}	0 μm	assuming no eye movement
h_{sample}	3.90 μm	(512 pixel resolution) variable
h_{eff}	9.09 μm	at Δz = 0 μm
h_{eff}	8.53 μm	e.g. at Δz = 10 μm
h_{eff}	14.97 μm	e.g. at Δz = 50 μm
h_{eff}	20.85 μm	e.g. at Δz = 100 μm
D	51.51 μm	points in focus
D	48.33 μm	e.g. points at Δz = 10μm
D	84.83 μm	e.g. points at Δz = 50μm
D	118.15 μm	e.g. points at Δz = 100μm

Figure 4.9. Geometric parameters for Data Set C.

4.9.2 Discussion

From figures 4.8 and 4.9 it is clear that the *expected* depth resolution for the simultaneous stereo photographs is not much better than the expected depth resolution for the non-simultaneous stereo photographs. This is not surprising, since both methods are similarly constrained. However, the consistent stereoscopic baseline provided by simultaneous stereo will facilitate comparisons between successive depth measurements. This is important for the clinical monitoring of the progression of eye disease such as glaucoma. In addition, since both left and right photographs are taken simultaneously, patient movement and change of fixation is not a problem, and camera refocus is not required. With a significant reduction in the possible sources of systematic error using simultaneous stereoscopic fundus photography, it is likely that the depth measurements taken from Data Set C will have greater consistency than the depth measurements from Data Set B.

In summary, simultaneous stereoscopic fundus imaging holds much promise, and it is unfortunate that only a limited number of commercially available fundus cameras can provide this type of imaging (see chapter 3). As technical advances improve the means by which fundus images can be obtained (e.g. the Scanning Laser Ophthalmoscope [Woon90]), further work will be required to investigate the potential benefits of simultaneous stereoscopic fundus imaging.

4.10 Summary

In order to understand the difficulties associated with stereo photography of the optic disc, various physical constraints were examined. The parameters which ultimately define the accuracy to which topographic measurements of the optic disc can be made include the size of the optic disc, the stereoscopic baseline, eye movements, motion blur, optical defocus, and illumination variations, all of which have been discussed. The geometric constraints imposed on the camera viewing positions are determined by the camera aperture and pupil diameter. Two methods for obtaining stereoscopic views of the optic disc have been discussed.

Non-simultaneous stereoscopic fundus images are obtained using a standard fundus camera, recorded on film, and then subsequently digitised. The limitation in image resolution is (perhaps) not the film grain, but rather the pixel resolution of the digitiser. Simultaneous stereoscopic fundus cameras similarly are limited in image resolution primarily by the digitiser. However, some commercially available stereoscopic fundus cameras provide both analogue and digital output, in addition to standard film camera mounts [Zimmermann89]. With these more modern fundus cameras, the maximum image resolution has been restricted by the manufacturer. The resolution of the digitiser available for this research is

512×512 pixels.

The most problematic parameter in non-simultaneous stereo fundus photography, and perhaps the most important, is the camera separation or baseline. The basic premise of stereoscopy is that the baseline is sufficient to allow for accurate triangulation of an object position, and without sufficient baseline the ability to recover depth information is restricted. There are several parameters which are beyond the control of the clinician taking non-simultaneous stereo photographs; these include patient movement and change of fixation. It is known that some patients cannot fixate accurately or for a long time, and a number of patients simply cannot cooperate to allow for positioning and refocusing for the second view. For the non-simultaneous stereo photographs used in this research, Data Set B, the expected depth resolution is between 40 and 190μm.

In simultaneous stereo fundus photography, optical defocus, image resolution, and image contrast are of key importance. With a fixed camera baseline, attention moves to the potential improvements in reliable depth measurements. However, there is a trade-off between aperture diameter and stereoscopic baseline in order to fit centrally within the dilated pupil. Only a limited number of commercially available fundus cameras can provide this type of imaging (see chapter 3), and for this research, only a small set of simultaneous stereoscopic fundus photographs could be obtained. For the simultaneous Donaldson stereo photographs used in this research, Data Set C, the expected depth resolution is between 50 and 150μm.

In conclusion, with the technology available today, one might construct an ideal simultaneous stereo fundus camera with a minimum depth resolution as follows:

Pupil diameter

Although this will vary from patient to patient, the pupil would ideally be dilated to a maximum of 8mm.

Aperture

There is a limit to the amount of light that can be safely used to illuminate the eye without damaging vision. This, in turn, directly affects the minimum size of aperture that can be used, since sufficient light to form an image (dependent on the detector) must pass through the aperture. The ideal camera aperture should be both small to minimise defocus effects, and large to minimise diffraction effects. However, the magnitude of defocus effects on lateral resolution are large compared to diffraction effects. To be consistent with what is known to be a safe level of illumination (see, for example, [Woon91]), and to minimise both defocus and diffraction effects, the aperture will be set at a value of 2mm.

Baseline

The camera baseline should be as large as possible while maintaining both views centrally within the pupil. The ideal camera baseline, from equation (4.6), is therefore 6mm.

Lateral resolution

For an ideal system, the image will be directly digitised, and therefore the sampling resolution is not the limiting factor in lateral resolution. It is feasible to have a sampling resolution of 4096×4096 pixels. Given the above parameters, the effective lateral resolution will be $h_{eff} \approx 5\mu m$.

Depth resolution

The minimum resolvable depth of the ideal system in the plane of best focus would be $D \approx 16\mu m$. For points which lie 100μm from the plane of best focus, the depth resolution would be $D \approx 50\mu m$.

Chapter 5
Stereo Matching Algorithms

5.1 Introduction

A brief overview of the basic methods for matching stereoscopic images (i.e. matching a pixel in one image to its corresponding pixel in the other image, thus determining the relative disparity at that position) is given here to introduce the general concepts involved in digital stereo matching. This chapter will then address the applicability of various stereo matching methods to the fundus image data. Specifically, several current algorithms based on these methods will be reviewed and evaluated, ultimately leading to the selection of a suitable stereo matching algorithm for the fundus image data.

Perhaps the most common method for stereo matching tries to identify those points in both images that are projections of distinct, precisely positioned local features on surfaces, such as edges, spots, corners, and other identifiable surface patterns. This will be termed *feature-based* stereo; see, for example [Marr76], [Marr79], [Grimson81], [Grimson82], [Nishihara84], [Pollard85], [Pollard87]. Feature-based stereo, represented in figure 5.1, is founded in the assumption that two views of the same scene will have identical projections of local features, provided the baseline is not excessive, and that such features occur at surface discontinuities, thus demarcating a surface boundary. Matching is then performed between these *feature* primitives. This method performs well when there is sufficient structure in the scene to allow a reasonably dense feature map of uniquely identifiable features, thus avoiding a resulting disparity map whose data are too sparse to permit estimation of the original surface of the scene. This method also relies on the use of a suitable operator to extract the desired features, and the accuracy of the match is thus dependent on the operator's ability to locate features accurately.

A second method for stereo matching makes use of the original intensity information in the image pair and utilises a correlation or difference function to determine a measure of similarity between the two images, as shown in figure 5.2. For a given pixel or position in the left image, a small region surrounding the pixel (a matching window) is defined. The position of maximal similarity in the right image is determined by scanning a correlation "window" over a search area which encompasses the maximum possible disparity. A "match" is typically defined between the pixels that lie in the centre of each region, and odd sized windows are often used for simplicity. This method will be termed *area-based* stereo; see, for example [Levine73], [Barnard80], [Sutton83], [Gruen85], [Rosenholm87a], [Medioni85]. Area-based stereo is highly sensitive to the amount of information content present in the region to be matched, and can be parameterised by such properties as the contrast and variance of the intensity values. The accuracy of this method is dependent on the size of the matching window that surrounds the desired pixel, but in most cases it can produce more accurate results than those which are obtained using feature-based methods, due to the supporting effect of neighbouring pixels within the matching window.

There are a number of other methods in the literature that have shown some promise for matching stereo images. These methods are not generally used to obtain fine-grained (accurate) disparity information, but rather yield a coarse-grained global estimate of disparity. The technique of using *optical flow*, or instantaneous flow field, assigns to every point in an image field a two-dimensional "velocity" which is a measure of the point's motion from one image field to the next in a sequence of images ([Ullman79], [Horn81]). Approximations to instantaneous flow can be computed from a sequence of discrete images, such as stereo images. Scheuing and Nieman [Scheuing86], for example, have applied this technique to stereo images and obtained an accuracy of about 95% in disparity globally across the image.

Another technique is based upon the concept of distorting the entire right-hand image until it "matches" the left-hand image. A second or third-order functional must be defined to model the various distortions between the two images. This technique is generally termed *elastic* or rubber-sheet matching and has been

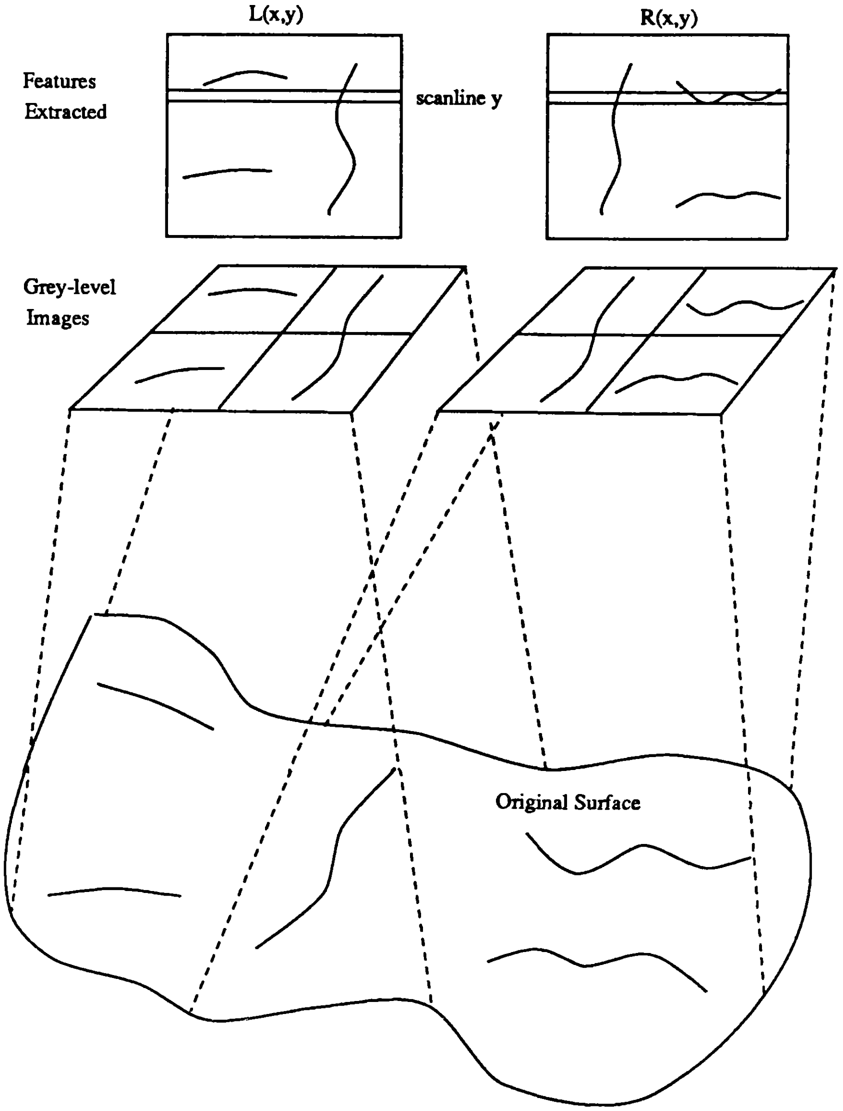


Figure 5.1. Representation of feature-based stereo matching, whereby features are extracted from the original images and are matched along corresponding scanlines.

applied, for example, to matching a set of data to a known model [Bajcsy89]. Elastic matching can be applied to stereo matching by considering the left-hand image to be the template to which the right-hand image data must be matched [Goshtasby87]. Another class of surface measurement method is termed *shape-*

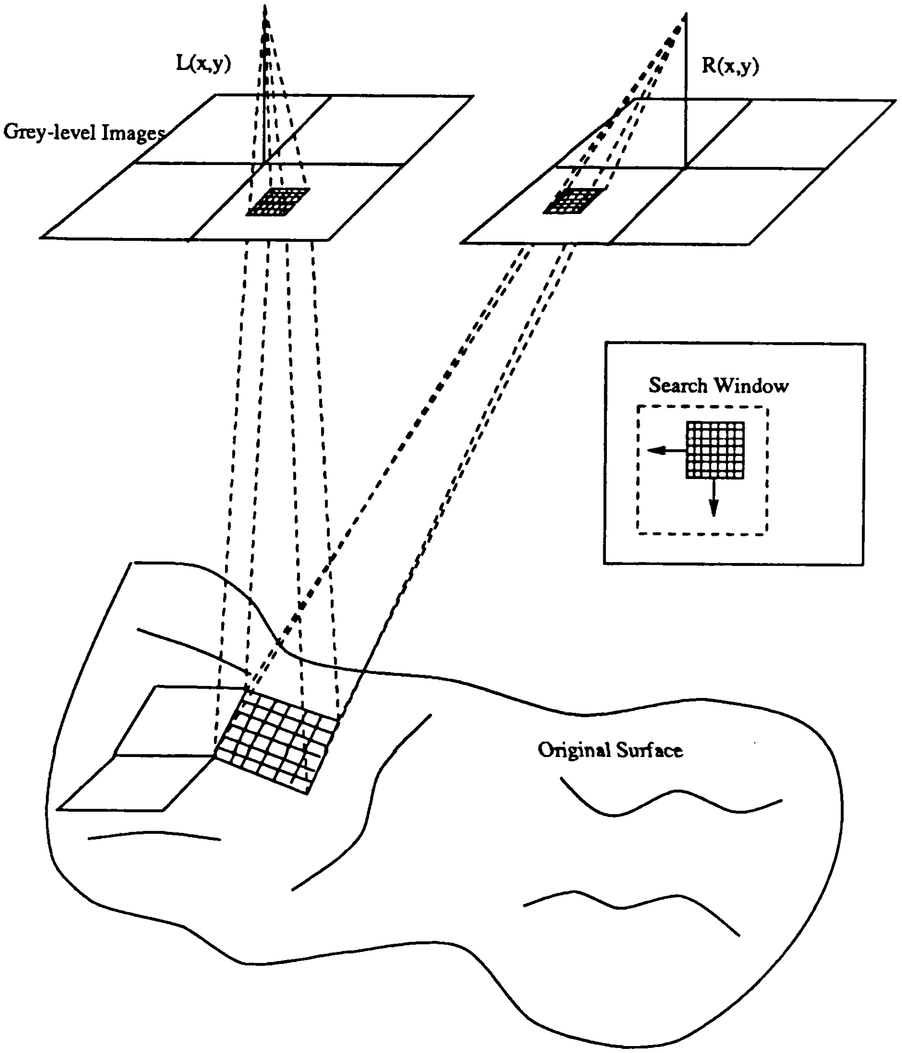


Figure 5.2. Representation of area-based stereo matching, whereby image patches are extracted from the original images and are matched using a similarity measure. The left-hand template patch is scanned across the right-hand image and the similarity measure is applied to determine the position of best match.

from techniques. The premise in shape-from techniques is that the visual cues in the images such as texture orientation, shading information, or photometric information (eg. polarised light) can be used to obtain measurements of surface shape (see, for example [Brzakovic86], [Lee84], [Woodham78]). One benefit of all these methods is that *a priori* information about any physical constraints can be included in the matching process. As the images are matched globally, any of the above methods could be used as the first component in a hierarchical approach to stereo matching.

It should be noted that feature-based (and often area-based) stereo matching algorithms typically require that the images are aligned prior to processing; that is, the images should be registered to establish a common line along the y-axis, such that disparity is constrained to the x-axis (and thus modeling image distortion is also similarly constrained). This constraint is not unreasonable, as it reduces the dimensionality of the search for determining a match, and can be obtained by simply using a stereo camera system that has been calibrated to provide (nearly) epipolar views (see chapter 4), or by careful image alignment, perhaps from known matched points in the images (eg. fiducial marks). Due to camera distortions, perspective distortions and the effects of noise, no two views will be *exactly* epipolar, and therefore errors in one-dimensional matching algorithms may be high. Some extensions to both matching methods have been developed to cope with two-dimensional search areas, and can be used when registration cannot be obtained prior to matching (see [Goshtasby85], [Ohta85], [Gruen86], [Lloyd85,86], [Eastman87], [Rosenholm87a]).

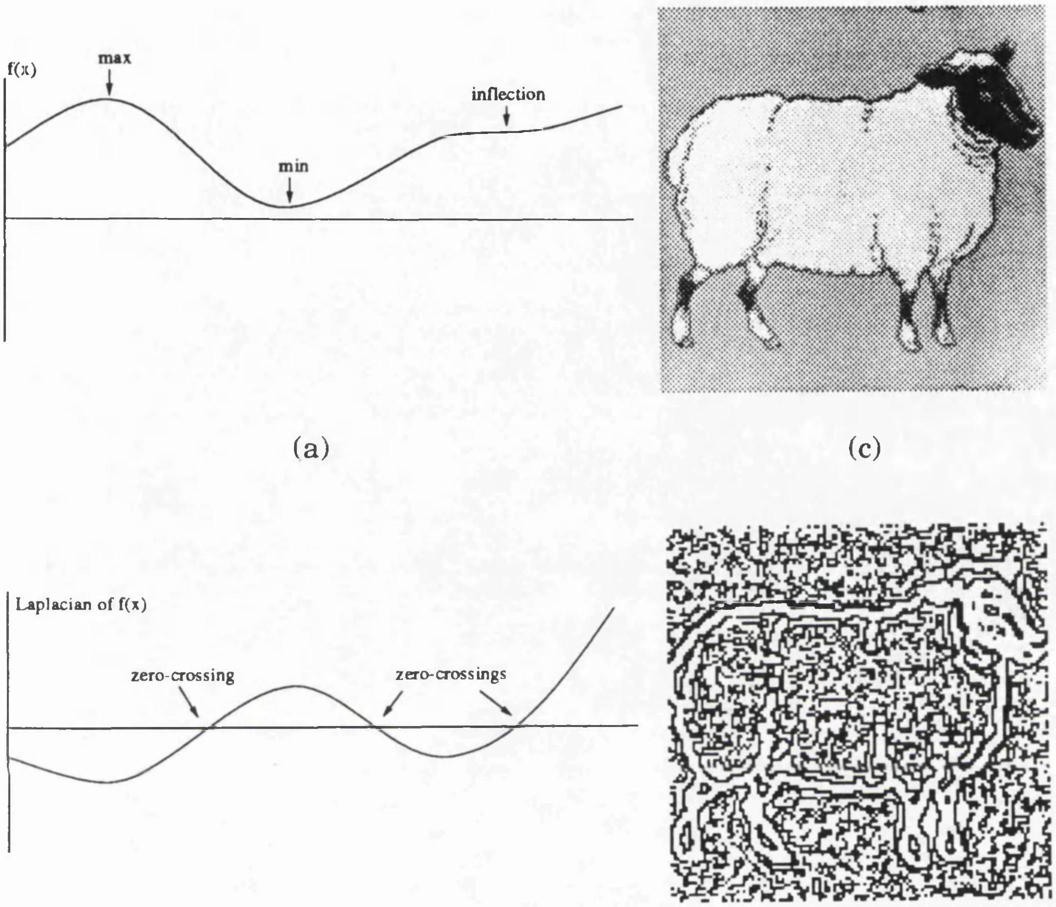
5.2 Feature-based Stereo Matching

5.2.1 Early Foundations in Computational Stereo

In 1976, Marr and Poggio [Marr76] proposed a computational theory of the stereo process for the human visual system, founded in the early work of Julesz with random-dot stereograms [Julesz60]. Julesz demonstrated that two images, apparently consisting of random dots when viewed monocularly, may be fused to form patterns separated in depth when viewed stereoscopically. This is important in that the problem of human stereopsis reduces to that of obtaining primitive descriptions of points to be matched from the images (in this case the primitives were simply dots), and solving the correspondence problem for these points.

Marr and Poggio proposed an algorithm [Marr79] which solves the stereoscopic correspondence problem. It contains five main steps: (i) the left and right images are each filtered with masks of four sizes that increase in diameter, and the structure of these masks can be approximated by the difference of two Gaussians with a size ratio of 1:1.75, or more commonly the Laplacian of a Gaussian function; (ii) zero-crossings in the filtered images are found along horizontal scan lines (see figure 5.3); (iii) for each mask size, matching takes place between zero-crossings of the same sign (the same directional sign transition) and roughly the same orientation, for a range of disparities up to the width of the mask's central region along corresponding scanlines; (iv) the output of the large masks can be used to constrain local disparity range, aiding the smaller masks to come into correspondence. The matching process proceeds from large disparities at low resolution to small disparities at a high resolution (as the diameter of the filter mask decreases); (v) correspondence is achieved in this coarse-to-fine manner, and results are stored in a dynamic buffer termed the 2-1/2-dimensional sketch [Marr79].

This early foundation brings about two concepts which are fundamental to the feature-based method of stereo matching; the properties of *uniqueness* and *continuity*. Uniqueness is the assumption that an object feature corresponds to something that has a unique physical position, and thus each feature from each image can only be assigned at most one disparity value (ie. each feature in the left image has only one corresponding feature in the right image). Continuity is the assumption that disparity varies smoothly almost everywhere, where only a small fraction of the area of the image is composed of boundaries that correspond to discontinuities in depth. Regions in the image where there is little change in



(a) (c)

(b) (d)

Figure 5.3. Examples of applying the Marr-Hildreth zero-crossing operator to a one- and two-dimensional scene. (a) one-dimensional $f(x)$, (b) Marr-Hildreth operator applied to (a) with detected zero-crossings, (c) an example image, and (d) the detected zero-crossings of (c).

intensity correspond to surfaces, whereas sharp changes in intensity usually correspond to a surface marking (texture) or a change in surface position (edges).

5.2.2 The Marr-Poggio-Grimson Algorithm

The early work of Marr and Poggio was followed with the implementation of the computational model for human stereo vision by Grimson [Grimson81]. To test and support the Marr-Poggio model, Grimson designed and implemented a computer algorithm based on the model and tested his algorithm on a wide range of stereo images. Grimson used the Laplacian of a Gaussian (LoG) filter (of varying sizes) to filter the images, followed by the detection of zero-crossings to mark locations of significant changes in the intensity function of the original images (see figure 5.3), and obtained two feature-point images where each point (zero-crossing) is defined by its position, contrast sign (the change in sign across the zero-crossing), and local orientation (determined from the gradient of the zero-crossing contour).

Given a set of feature-point images obtained from applying filters of varying sizes, the matching proceeds in a coarse-to-fine manner as outlined by Marr and Poggio (see [Marr79]). The matching process (refer to figure 5.4) consists of six steps for each size of filter:

1. obtain zero crossing descriptions for the current filter size;
2. align the two zero crossing descriptions by the current estimate of disparity (initially zero for the largest filter);
3. given the location of a zero crossing in one image, partition the region about the same location in the other image (along the corresponding scanline) into three pools, which form the areas to be searched for a possible matching zero crossing. These pools span a disparity range equal to twice the width of the central region of the LoG filter. There are two large convergent and divergent pools and a smaller one lying centrally between them;
4. assign a match to the zero crossing, using the criteria that the zero crossings come from convolutions with the same filter size, have the same sign, and have roughly the same orientation;

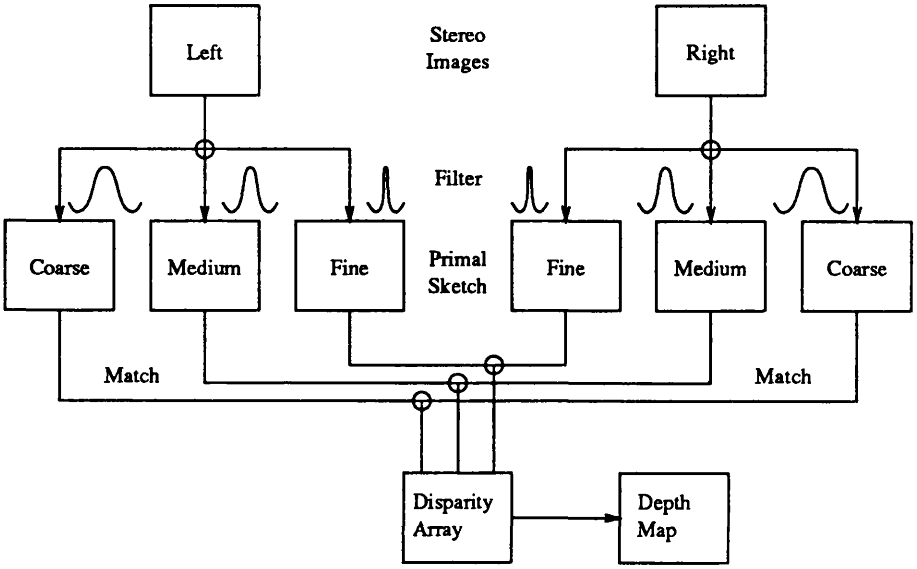


Figure 5.4. Representation of the Marr-Poggio-Grimson algorithm. The stereo images are first filtered using the LoG operator, and then zero-crossings are extracted to form the primal sketches. The stereo correspondence problem is solved by matching tokens in the left and right primal sketches. The disparity array is built up in a coarse-to-fine manner, where the coarse channels control registration and help bring the fine channels into alignment. Depth information is determined from the disparity estimates.

5. disambiguate any multiple or ambiguous matches using information about unambiguous matches in the local neighbourhood;
6. store the estimated disparity for the zero crossing.

As stated above, the matching proceeds in a coarse-to-fine manner starting with the largest filter size and initial disparity estimates of zero. Disparity estimates from the previous (larger) filter can be used for alignment (local registration) in step 1, and thus local convergence is obtained through successive iterations with decreasing filter size (see also [Grimson85]).

5.3 Other Feature-based Methods

Numerous other feature-based methods have grown out of the early work of Marr & Poggio. Efforts were made to improve the local convergence of the matcher and to devise improved methods for disambiguating matches. Work by [Arnold80], [Nishihara84], [Grimson85], and others, explored modification of the types of filtering and the features which were extracted. For example, Medioni and Nevatia [Medioni85] used straight line segments as opposed to individual zero crossings as primitives for matching. This method performs well for images with a large number of near vertical linear features.

Pollard, Mayhew, and Frisby [Pollard85] developed a computational model based on the disparity gradient, where edge primitives are matched only in the presence of local support for which the disparity gradient lies within a given band (see also [Lloyd85], [Trevidi85], and [Pollard87]). They found that by restricting matches to lie within a disparity gradient limit, disambiguation of false matches could more easily be obtained.

Eastman and Waxman [Eastman87] among others ([Olsen86], [March88]) have applied analytic models of disparity to constrain potential feature point matches and obtain an overall surface reconstruction as a disparity functional, not just as estimates of disparity at isolated feature points. Ohta and Kanade [Ohta85] and [Lloyd86,87] have extended the search for corresponding matches to two dimensions by using dynamic programming heuristics to control the search both along corresponding scan-lines and among scan-lines, obtaining correspondence between edge delineated intervals. This proves useful when the epipolar constraint cannot be met, and some vertical disparity remains.

The fundamental principles underlying all these methods has remained unchanged: a spatial filter is applied to the images in order to extract those features which are suitable for matching; feature primitives in both the left and right stereo image are extracted to form feature maps; matching rules are applied to each feature in the left image in order to locate a matching feature in the right image (often involving a disambiguation or constraining process); and the resulting disparity is recorded. This process is the premise on which all feature-based methods are founded.

5.4 Area-based Stereo Matching

5.4.1 Foundations

The filtering processes used to extract primitives in feature-based stereo enhance (and de-emphasise) certain features in the image, resulting in a sparse set of features, and thus a sparse set of disparity measurements. Often, it is desirable to match and process the images in their original format: a spatial array of grey levels. Area-based stereo matching uses the original intensity information in the two images and utilises a correlation function to determine a measure of the similarity between the two images. For each small region in one image (centered about a specific location), the position of maximal similarity in the other image is determined by scanning a correlation "window" over a search area and applying the correlation measure at each position. A match is defined at the pixel positions that lie in the centre of the region or window.

Area-based matching is highly sensitive to the amount of image contrast or variance of the intensity values. For example, to match a region which lay across a boundary in the image where one half is almost entirely white, the other half almost entirely black, the corresponding position in the second (similar) image would be difficult to locate. The correlation measure would give a high score to any number of positions along the boundary (a correlation ridge, with no clearly defined maximum). This highlights one of the key parameters in area-based correlation: the size of the correlation window. If the window is too small, there may be insufficient data to locate the match. If the window is too large, there may be too many data, resulting in poor localisation and accuracy.

As with feature-based stereo matching, area-based stereo matching algorithms typically require that the images are aligned prior to processing (i.e. the images should be registered to establish the epipolar lines along the y-axis).

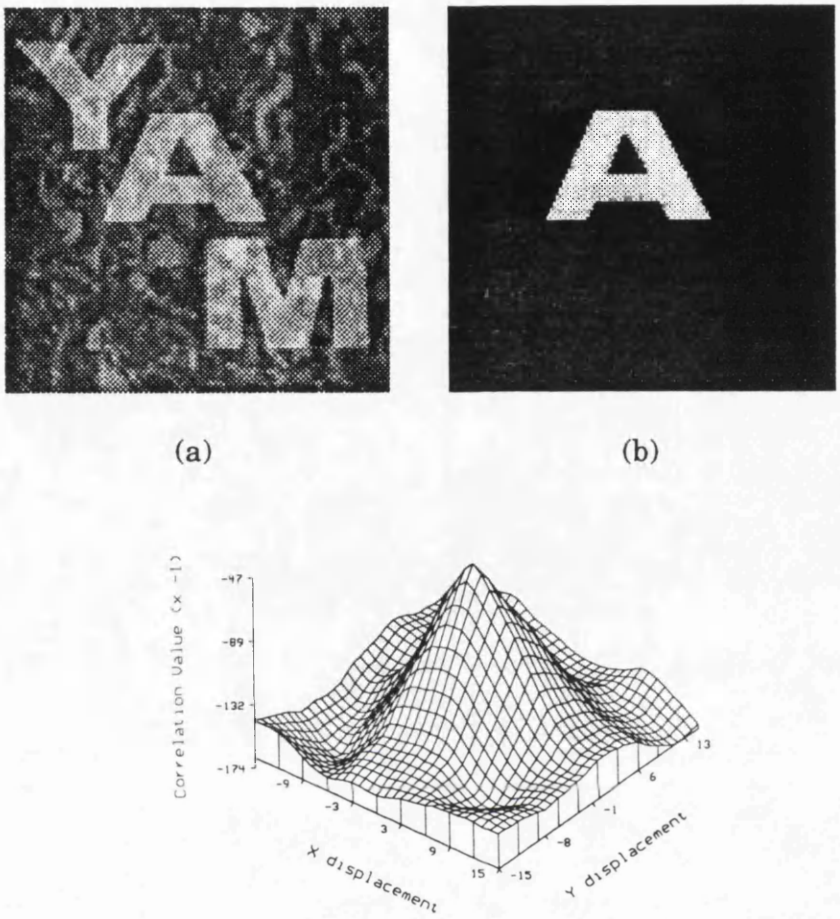
This constraint is not unreasonable as it reduces the dimensionality of the search area for determining a match. In the above example, if the search is constrained to one dimension, the correlation measure would have little difficulty locating a match. One-dimensional searches along corresponding scanlines can provided a fast and effective method of stereo correlation. Area-based matching, unlike feature-based matching, is more easily extended to cope with two dimensional search areas, and is an attractive approach in applications where registration is difficult.

5.4.2 Template Matching

Template matching forms the basic principle behind area-based stereo matching. Template matching is a simple filtering method for detecting a particular object or feature in an image. If the attributes or appearance of an object are known, it can be detected in an image by using a template or subimage that looks just like the image of the object. A measure of similarity is computed to reflect how well the image data match the template data for each possible template location. The position which gives the maximal match score is selected as the location of the object in the image (assuming the sought-for object is located *somewhere* in the image).

There are a number of standard similarity measures that can be used to locate the position in the image at which the template yields the maximal match score. Many similarity functions have been proposed for specific applications (see, for example [Peli87], [Algazi85], [Badique88], and [Bajcsy89]). One such similarity measure between an image function $f(x)$ and a template $t(x)$ is the Euclidean distance squared, which is denoted $D(d)$, often called the minimum-error, minimum-square error or least-squares algorithm [Lee88]. The one-dimensional correlation measure is considered for simplicity, where the two-dimensional equivalent is readily inferred. The one-dimensional form of $D(d)$ is given by

$$D(d) = \sum_{x=-M/2}^{M/2} [\, f(x+d) - t(x) \,]^2 \tag{5.1}$$



(a) (b)

(c)

Figure 5.5. Example of a scene (a), a small template (b), and (c) the correlation surface that results from using the similarity measure in equation (5.1).

Here M represents the size of the template subimage, and d the offset or displacement of the template in the image relative to some starting position. If the image at position d is an exact match of the template, then $D(d) = 0$ and would otherwise be > 0. This function is often expanded to extract what is called the cross-correlation coefficient between f and t, given by

$$C(d) = \sum_{x=-M/2}^{M/2} f(x-d)\, t(x) \tag{5.2}$$

This expression is the most frequently used algorithm for template matching, where the image and template data values are superimposed at the offset d, multiplied together and the products are added. The resulting sum forms an entry in a correlation array at the position corresponding to the offset d. The template is scanned across the image for a valid set of offsets. Figure 5.5 shows an example of a two-dimensional scene, a small template, and the correlation surface that results from using the similarity measure in equation (5.1). The popularity of this approach manifests itself in that careful management of computed products at a given offset d allows for only a small number of new products to be computed at the neighbouring offset $d + 1$. This is often called sequential similarity detection (SSD) (see [Barnea72]).

There may be several copies or near-copies of the template in the image, in which case there will be several minima (5.1) or maxima (5.2). Scaled or normalised versions of both these similarity measures exist in the literature (see, for example [Pratt78], [Rosenfeld82], and [Griffin90]), but the underlying principles are the same. Auto- and cross-correlation algorithms are discussed in more detail in chapter 7.

It is often noted that the practical advantage of the correlation measure D over C is that it is less affected by a large offset in the average intensity value (the DC shift) of the image compared to that of the template, and it is far less sensitive to intermittent bright spots or noise in the image. The correlation measure C can be modified to avoid the effect of DC bias by calculating the mean for the image and template and removing it (see, for example [Keating85] and [Ballard82]). However, although the normalised correlation measure is less sensitive to the local properties of the template and image, it is sensitive to the signal-to-noise content of the image. High uncorrelated noise in the template or image decreases the value of the correlation. The strengths and weaknesses of both similarity measures are debatable, but it is usually the case that one method is better suited to a particular application.

5.4.3 Cross-Correlation

The method of template matching assumes an idealised template sub-image which is to be scanned across a less than ideal (perhaps noisy) image to locate a match position. The concepts of template matching can be applied to stereo matching by considering the region about a position in the left-hand image to be the template sub-image, and the entire right-hand image to be the scene or image to be scanned. In this case the template sub-image has no fixed size; the appropriate size of sub-image to be used as the template must be chosen. If the noise and signal characteristics of the data are known, the window (or *patch size* as it is often called) may be optimised by using that information together with some simple statistical information about the grey-level structure of the region about the point in question. However, such estimates about optimal patch size do not consider the effects on systematic, non-statistical error such as imaging distortions (perspective distortion), rotation and scale differences between the images, all of which are commonplace in stereoscopic image pairs.

The process of cross-correlation can be computationally very expensive, especially when used for stereo matching where a large number of points need to be matched. For points in the left-hand image at which a disparity measure is required, it is often advantageous to ensure that there is enough “information” in the template or patch, to make sure that the image contains enough signal to give a reliable correlation measure. If the image region is relatively uniform in intensity (slowly varying signal), the resulting correlation measure will be flat with no clearly defined minimum (or maximum). Moravec [Moravec77] and Barnard and Thompson [Barnard80] apply an inexpensive “interest” operator to the template region before the relatively expensive correlation, in order to determine the signal content of the template sub-image. Levine, O’Handley and Yagi [Levine73] use an adaptive correlation window, the size of which varies inversely with the variance of the region surrounding the point.

Cross-correlation is generally a good similarity criterion for area-based stereo matching. A number of techniques can be applied to increase the sensitivity of the similarity measure to avoid such problems as, for example, the effects of noise or distortion where the correct match is not the one of maximum similarity ([Levine73], [Barnard80]). Several strategies can be used to limit the search area in a cross-correlation approach. Studies in stereopsis often use a fixed camera model to constrain the search to one dimension (e.g. [Levine73], [Nevatia76]). Alternatively a coarse-to-fine control strategy can be used; a coarse search is used to approximately locate the matching points, followed by successively finer searches to more accurately locate them ([Hoff86], [Xu87]). Finally and perhaps most importantly, cross-correlation formulae are easily extended to cope with the additional parameters that arise in stereo matching applications, such as perspective distortion, rotation and scale changes (see, for example [Bajcsy82], [Castan85], [Gruen85], and [Rosenholm87a]). The additional complexity introduced by multi-parameter searching is often most easily handled by the minimum-square error or least-squares algorithm, as outlined in 5.4.2.

5.5 Other Stereo Matching Methods

There are a number of other methods for stereo matching images in the literature that are of interest. Briefly mentioned in section 5.1, two methods in particular, namely *optical flow* and *elastic matching*, may be applicable to the fundus image data and therefore will be reviewed.

Optical flow, or instantaneous flow field, attempts to assign to every point in an image field a two-dimensional velocity vector which is a measure of each point's "motion" from one image to the next in a sequence of images (see [Horn81]). This motion vector, in the case of stereo images, can be used to represent the disparity at each image point. Scheuing and Niemann [Scheuing86] have developed an approach to depth determination which uses stereo images and a combination of the iterative computation of optical flow with a local correspondence criterion. In a time sequence of images, one may imagine the intensity value at every image point to have an associated velocity vector. The field of velocity vectors is called *optical flow*. If the left and right stereo images are considered as time sequence images one unit of time apart, an iterative equation for computing disparity can be derived from the optical flow equations. Schueng and Niemann have shown that accurate results are obtainable, and that results can be computed in a reasonable amount of time using this approach (see also [Ballard82], [Nagel83]).

Elastic matching is a template matching technique whereby an entire image is distorted or stretched (hence *elastic*) until it "matches" the template or model image. Bajcsy [Bajcsy82,89] has applied this technique to matching deformed images to a known model image for comparison and analysis. A model of the physical system that simulates the manual registration process is used, and by applying external forces the shape of the model can be changed so that the model data become more similar to the image data. The model image is distorted until an equilibrium state is achieved between the external forces and the (elastic) internal resting forces. An equilibrium state of the system corresponds to a local minimum of the total energy in the system, and in this case can be measured as the difference between the cost due to similarity between the model and the image, and the cost due to the deformation of the model image. A suitable cost function is when the cost of similarity is comparable with the cost of deformation, as this allows the model to be elastic enough to achieve similarity but rigid enough to avoid very small local distortions. Often, a second or third-order functional must be defined to represent complex distortions between the model and the image. Goshtasby [Goshtasby87] employed a piece-wise cubic mapping function in the application of image registration. This technique could be applied to stereo matching by considering the left-hand image to be the model to which the right-hand image must be deformed and matched.

In summary, both the optical flow and elastic matching techniques could be used on the fundus image data. One benefit of these methods is that *a priori* information about any physical constraints can be included in the matching process, and thereby improving the accuracy of the resulting disparity measurements. Although these methods are not generally used to obtain fine-grained disparity information, such methods could be used in a hierarchical approach to stereo matching to provide an initial coarse-grained global estimate of disparity.

5.6 Selecting An Algorithm for the Fundus Data

5.6.1 Problems and Issues

The selection of a stereo algorithm is ultimately based on the information content in the images and this factor is often difficult to quantify [Toriwaki78]. The fundus images for the most part contain large (black) vascular structures on a background of mottled grey texture, the retinal nerve fibre layer. The vascular structures are known to lie on and within the nerve fibre layer, and have a glossy or reflective appearance. The nerve fibre provides some texture information, but these areas are typically low in contrast. Oddly the primary area of interest, the optic cup, is often saturated in intensity in a fundus photograph, and would provide little or no information.

Feature-based stereo cannot rely on information in the nerve fibre to provide *distinct* edges or features, and would therefore rely mainly on the edges along the veins. However, the images are initially not registered (aligned), and manual alignment of the photographic data sets, for example, is impractical. Two-dimensional searches along edge contours would add considerable complexity to an algorithm in order to obtain accurate results.

Area-based stereo is not without problems as well. Correlation, as pointed out above, will have great difficulties in areas of little image contrast and along boundaries between such regions. The fundus images have little contrast in areas of nerve fibre layer and on the veins themselves, and present boundary problems near and along the edges of the veins. However, the ability of area-based methods to be easily extended to 2-dimensional searches, as well as the ability to cope with local affine distortions, makes this an attractive method provided sufficient image information is present to enable accurate correlation.

It is clear that feature-based stereo could provide some disparity information about the veins, given that the images could be registered to reduce the dimensionality of the matching process. Area-based stereo may provide some useful disparity information in the large nerve fibre layer areas, given that regions of poor image contrast and/or variance are avoided or excluded. Extensions to the area-based stereo method can allow for arbitrary rotation and translation between stereo views. Not without merit are the methods of optical flow and elastic matching, which could be used to obtain an initial (coarse) estimate of the surface shape, to be "fine-tuned" by measurements made from both feature-based and/or area-based matching. A hierarchical combination of methods could provide a dense disparity map for analysis.

When implementing an algorithm on a computer it is important to consider the computational complexity of the algorithm and its ease of implementation. These factors are important in determining the feasibility of implementation and the amount of processing time the algorithm will require to reach a solution. Often, by adding some simple constraints, the complexity of the problem can be reduced, thereby providing greater ease of implementation and reduced computation time. Present methods for stereo photography of the fundus typically provide only a few constraints to aid in the stereo matching process, thereby increasing the complexity of the algorithm, and decreasing the probability of reaching an acceptable solution.

5.6.2 Selecting an Algorithm

After considering all the possible stereo matching algorithms discussed herein, a choice of algorithm for the fundus image data will be made, based on the following observations:

- The images contain a lot of small-scale texture (though often poor in contrast), but large-scale features (ie. veins) are relatively sparse - therefore the algorithm must be able to cope with the large retinal nerve fibre layer regions in the images. Area-based algorithms are known to perform reasonably well even in regions which are nearly homogeneous (using large patches);

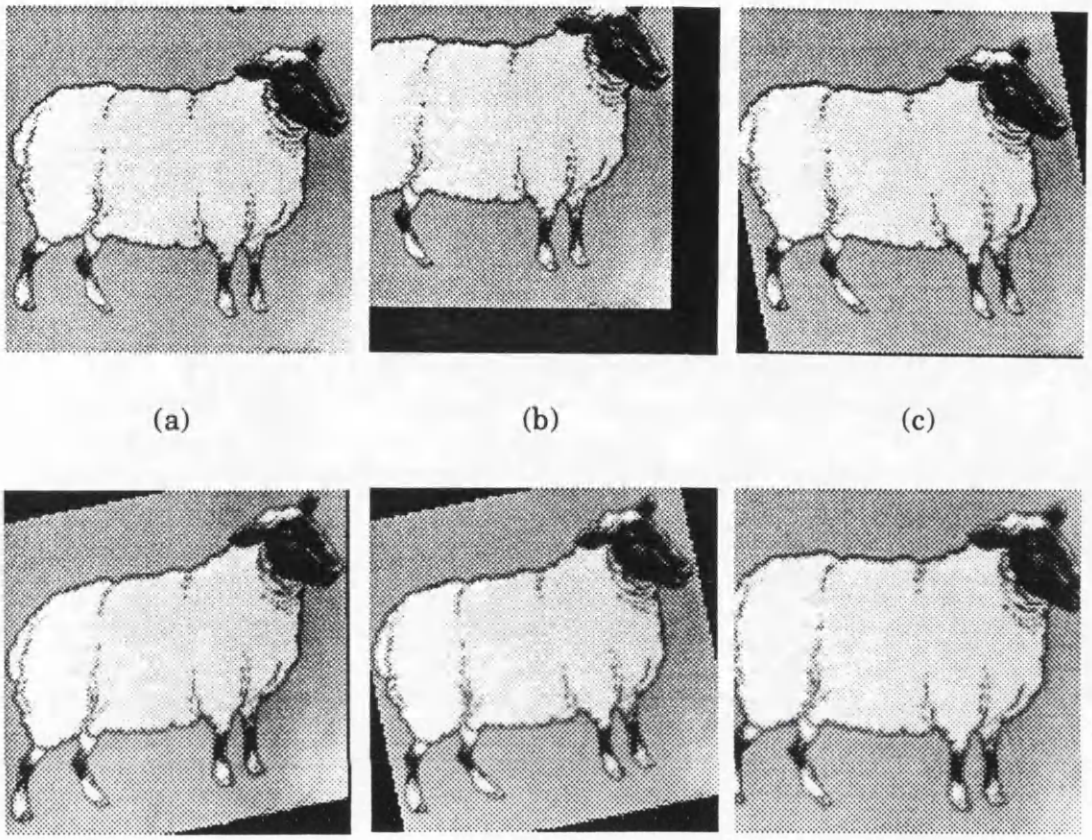
- There are often significant distortions between corresponding image patches, and the ability to compensate for geometric and radiometric parameters in the matching process is important;
- Obtaining accurate image alignment prior to processing is relatively difficult, and therefore algorithms which can cope with two-dimensional searches are preferred;
- A relatively dense set of disparity measurements is required to accurately reconstruct the entire optic disc region; feature-based algorithms cannot meet this requirement, since typically only a sparse set of features is extracted for matching;
- The accuracy of the disparity measurements is important; the accuracy of disparity measurements using feature-based algorithms is dependent on the locational accuracy of the feature detector, whereas area-based algorithms can produce very accurate answers (often sub-pixel) since they compare the pixel data directly;
- The computation time for the entire matching process must be considered; the algorithm must be made to run at an acceptable speed, given the number of image points that must be matched.

In considering all of the above criteria, an area-based algorithm is selected. The area-based algorithm chosen is one described by Gruen [Gruen85] because it is claimed to achieve very high accuracy [Gruen86], and because it incorporates geometric and radiometric distortions in the matching process. Rosenholm [Rosenholm87b] has applied this algorithm to both low and high contrast regions of aerial stereo photographs, and reported better than half a pixel accuracy. Otto and Chau [Otto89] have also applied this algorithm to satellite digital stereo images (SPOT images) and obtained an average accuracy of 0.3 pixel. Interestingly, portions of these SPOT images, and indeed some of the images selected by Rosenholm, have a similar appearance to the retinal nerve fibre layer regions of the fundus images. The good to excellent results obtained on the SPOT images offers promise for the application of Gruen's algorithm to the fundus data. There is increasing support in the literature for algorithms of this type. In the next section Gruen's algorithm is described in some detail.

5.6.3 A Least-Squares Correlation Algorithm

Gruen's algorithm is a least-squares correlation algorithm. The algorithm minimises the sum-of-squared-differences between two image patches, the minimisation being over a set of parameters specifying how the patches are allowed to be distorted to obtain a match between the images. Gruen allows for both affine transformations between image coordinates and radiometric distortions between the grey-levels (see figure 5.6). The least-squares process is essentially a multi-parameter optimisation problem, which can be solved iteratively by making initial estimates of the parameters, linearising a set of homogeneous equations which model the changes to the parameters which will move in the direction of a local minimum, and iterating the procedure to converge on a local minimum. Although the correlation algorithm has been shown to accurately match points between two images, the initial starting values for the parameters need to be fairly close to the solution, typically within two or three pixels.

It is impractical to manually or automatically select estimates of the starting parameters for all intended match points. However, by exploiting the continuity of the surface being viewed, points for which a match has already been obtained can be used as a basis for predicting the initial parameters for neighbouring points. Otto and Chau [Otto89] developed a *region growing* algorithm which requires only a few initial match estimates on which Gruen's algorithm is run. From these initial *seed* points, the location of approximate matches for neighbouring points can be predicted, to which Gruen's algorithm can be applied. By iterating this process, a complete set of matched points can be *grown* out from an initial (small) set of seed points. The seed points can either be selected manually, taken from obvious matching features or marks in the images, or can be



(a) (b) (c)

(d) (e) (f)

Figure 5.6. The types of distortions which can be modelled by Gruen's algorithm. (a) original image, (b) translation, (c) x-shear, (d) y-shear, (e) rotation, and (f) scaling (radiometric distortion is not shown).

generated automatically using, for example, a high-confidence, sparse stereo matcher such as the Barnard and Thompson algorithm [Barnard80]. Otto and Chau found that typically only two or three seed points are required (some points may fail to converge on a local minimum and must be discarded).

The basic version of Gruen's algorithm (as described in [Gruen85]) works by matching image patches (typically in the range of 15x15 to 30x30 pixels) which are centered about the positions in the images at which a match is to be calculated. The left-hand patch is extracted from the left-hand image and remains constant. The right-hand patch is a distorted and resampled sub-window from the right-hand image, based on an affine transformation and a radiometric adjustment (applied to each pixel in the patch). Figure 5.7 represents the spatial relationships of the parameters.

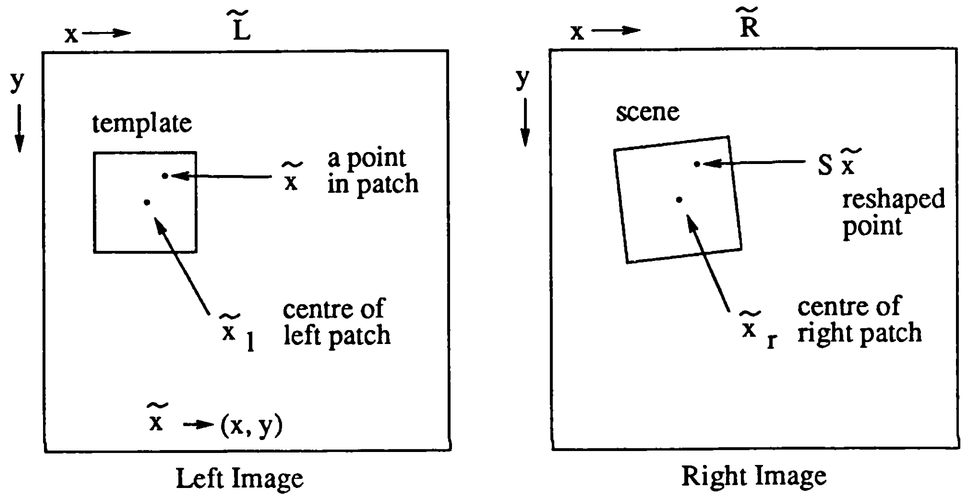


Figure 5.7. The spatial relationships in Gruen's algorithm.

Let L denote the left-hand image, R denote the right-hand image, $\tilde{x}_l$ be the position (x_l, y_l) in the left-hand image which is to be matched, and $\tilde{x}_r$ be the initial starting position (x_r, y_r) in the right-hand image. The value at a point $\tilde{x}$ in the left-hand image patch is given by

$$L(\tilde{x}_l + \tilde{x}) \tag{5.3}$$

and the corresponding point in the right-hand image patch is given by

$$R(\tilde{x}_r + S\tilde{x}) + r_s \tag{5.4}$$

where

$r_s \equiv$ radiometric shift parameter

$S \equiv$ the transformation or shaping matrix

$$\begin{bmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{bmatrix} .$$

The shaping matrix corresponds to an affine transformation of the points in the right-hand patch. This geometric distortion is equivalent to assuming that the viewed surface is approximately planar within the region visible through the patch. For small patch sizes, this approximation is not unreasonable. The objective is to minimise the L_2 norm of the error ε given as

$$\varepsilon(\tilde{x}) = L(\tilde{x}_l + \tilde{x}) - \left[R(\tilde{x}_r + S\tilde{x}) + r_s \right], \text{ for all } \tilde{x} \in \text{patch} . \tag{5.5}$$

In order to apply the least-squares algorithm, ε be must linearised with respect to all the parameters. Linearising $R(\tilde{x}_r + S\tilde{x})$ with respect to $\tilde{x}_r$, S, and r_s the following is obtained,

$$\begin{aligned} R(\tilde{x}_r^0 + S^0\tilde{x}) &+ \frac{\partial R}{\partial x}\bigg|_{(\tilde{x}_r^0 + S^0\tilde{x})} \left[\Delta x_r + \Delta s_{11}x + \Delta s_{12}y \right] \\ &+ \frac{\partial R}{\partial y}\bigg|_{(\tilde{x}_r^0 + S^0\tilde{x})} \left[\Delta y_r + \Delta s_{21}x + \Delta s_{22}y \right] , \end{aligned} \tag{5.6}$$

where $\tilde{x}_r^0$ and S^0 are the initial estimates of the parameters. Completing the linearisation and denoting $\frac{\partial R}{\partial x}$ as R_x and $\frac{\partial R}{\partial y}$ as R_y, ε becomes

$$\begin{aligned} \varepsilon(\tilde{x}) \approx{} & L(\tilde{x}_l + \tilde{x}) - R(\tilde{x}_r^0 + S^0\tilde{x}) - r_s \\ & - R_x(\tilde{x}_r^0 + S^0\tilde{x})\,(\Delta x_r + \Delta s_{11}x + \Delta s_{12}y) \\ & - R_y(\tilde{x}_r^0 + S^0\tilde{x})\,(\Delta y_r + \Delta s_{21}x + \Delta s_{22}y) . \end{aligned} \tag{5.7}$$

This can be rewritten as a set of homogeneous equations in vector form as

$$\varepsilon(\tilde{x}) \approx (\tilde{L} - \tilde{R}) - \begin{bmatrix} R_x & xR_x & yR_x & R_y & xR_y & yR_y & 1 \end{bmatrix} \begin{bmatrix} \Delta x_r \\ \Delta s_{11} \\ \Delta s_{12} \\ \Delta y_r \\ \Delta s_{21} \\ \Delta s_{22} \\ r_s \end{bmatrix} \qquad (5.8)$$

for all $\tilde{x} \in$ patch .

The values for L, R, R_x, and R_y are taken from the positions $\tilde{x}$ in the sampled patches, forming a set of over-constrained linear equations. Letting $\tilde{l} = (\tilde{L} - \tilde{R})$, $A = (R_x \; xR_x \; \cdots \; 1)$, and $\tilde{\alpha} = (\Delta x_r \; \Delta s_{11} \; \cdots \; r_s)^T$, the more familiar form of the equation is then obtained:

$$\tilde{\varepsilon} = \tilde{l} - A\tilde{\alpha} \, . \qquad (5.9)$$

Equation (5.9) can be solved iteratively using standard least-squares techniques (see, for example [Golub83], [Ralston85]). Through each successive application of equation (5.9), the estimates of the parameters are adjusted by the Δ's given by $\tilde{\alpha}$. When the changes to the parameters are smaller than a predetermined threshold, the iterations terminate and the solution is found. In this manner, the algorithm *tracks* or homes in on the position in the right-hand image ($\tilde{x}_r$) for which the left-hand position ($\tilde{x}_l$) has a match, for the given patch size. The importance of an initial estimate of the parameters which is near the correct solution is now apparent. If the starting estimate is too far from the correct solution, the algorithm may *track* toward an incorrect, but locally minimal, solution. Detecting false matches will be addressed in the application of Gruen's algorithm to the fundus data in Chapter 6.

One interesting aspect of the algorithm is that, through linearisation of equation 5.5, it becomes necessary to take the first derivative about each point in the right-hand patch. Consequently, the algorithm is making use of edge or high-contrast feature information in much the same way that feature-based algorithms do (many feature-based algorithm employ a first derivative operator such as [Canny83a,b] to extract the primitives for matching). The use of first derivatives elevates Gruen's algorithm above simple template matching, and shows the hybrid feature-based/area-based nature of the algorithm. Further research is required to investigate this property more thoroughly.

Finally, note that if the parameters applied to the right-hand patch require that samples be taken from sub-pixel locations in the right-hand image during resampling, the data must be interpolated from the original discrete set of image values. Gruen suggests using bilinear interpolation for sub-pixel estimates of image values. Such interpolation is only meaningful if some sort of correlation exists between pixels values. Fortunately it is often the case that significant correlation exists over a few pixels, perhaps due to image blurring or signal degradation. In the application of Gruen's algorithm, which is described in chapter 6, it is assumed that a meaningful estimate of sub-pixel image values can be made between discrete samples in the fundus image data using bilinear interpolation.

5.7 Summary

This chapter has addressed the applicability of various stereo matching methods to the fundus image data. Stereo matching algorithms including feature-based matching, area-based matching (template matching and cross-correlation), optical flow techniques, and elastic matching have been reviewed and evaluated. The potential problems for stereo matching algorithms specific to the fundus image data have been addressed, leading to a set of criteria on which to base the selection of an algorithm. The algorithm must cope with large regions of small-scale texture, large geometric and radiometric distortions between the two images, two-dimensional searches, and provide a dense set of accurate disparity measurements. In considering the relevant algorithms in the literature, an area-based algorithm based on the least-squares matching technique met all the criteria. Gruen's algorithm [Gruen85] and its extensions [Otto89] are described in detail. In the following chapter, the application of Gruen's algorithm to the fundus image data will be reviewed.

Chapter 6
Application of Stereo Matching

6.1 Introduction

In this chapter the application of a stereo matching algorithm to the fundus image data is presented. The implementation of the Gruen's [Gruen85] stereo matching algorithm, selected for use on the fundus image data in chapter 5, is discussed, and a pseudo-code description of the algorithm is given. A brief analysis of the computational requirements of the algorithm is also presented.

Next, the experimental methods are presented, outlining the processes of data acquisition, image pre-processing and filtering, stereo matching, and post-processing and subsequent representation of the disparity data. The experimental results from the application of these methods to the Data Sets A, B, C and D (discussed in chapter 2) are presented and discussed.

Finally, an analysis of the results, addressing such issues as matching error and accuracy, leads to a discussion on the limitations of stereo matching algorithm on the fundus image data. The isolation of several key issues associated with not only the stereo matching algorithm but also the fundus image data, points the way forward for the development of useful mathematical models in chapter 7.

6.2 Implementation of Gruen's Algorithm

In this section the implementation of Gruen's [Gruen85] stereo matching algorithm is presented. The mathematical process underlying Gruen's algorithm is discussed in chapter 5, section 6.3, and here we turn our attention to the computer implementation of the algorithm. A pseudo-code description of the algorithm is presented, together with an analysis of the computational requirements.

6.2.1 Overview of the Algorithm

An overview of Gruen's stereo matching algorithm is given here to show how the mathematical description presented in chapter 5 is translated into a process which can be implemented on a computer. Gruen's algorithm is a least-squares correlation algorithm which attempts to minimise the sum of the squares of the differences between two image patches, the minimisation being over a set of parameters specifying how the patches can be transformed to obtain a match. In the implementation of Gruen's algorithm presented here, the parameters are not just translational $(\Delta x, \Delta y)$, which would specify the location of the match, but include the set of affine transformation parameters which allow for translation, scaling, shearing, and rotation of one patch with respect to the other (see chapter 5, figure 5.5). Two radiometric parameters are also used to adjust for differences in illumination (radiometric) offset (r_s) and multiplicative contrast (r_m) changes between the image patches. The addition of the multiplicative parameter changes the original equation (5.4) to allow for an improved model of radiometric difference; thus, $R(r_m \tilde{x}_r + r_m S\tilde{x}) + r_s$.

The least-squares process is a multi-parameter optimisation problem which can be solved iteratively. Recall equation (5.8) from chapter 5, which embodies the mathematical process of Gruen's algorithm in the standard least-squares form. A slight modification of equation (5.8) yields a form suitable for computer implementation.

$$\varepsilon(\tilde{x}) \approx \tilde{L} - \begin{bmatrix} r_m R_x & x r_m R_x & y r_m R_x & r_m R_y & x r_m R_y & y r_m R_y & 1 & R \end{bmatrix} \begin{bmatrix} \Delta x_r \\ \Delta s_{11} \\ \Delta s_{12} \\ \Delta y_r \\ \Delta s_{21} \\ \Delta s_{22} \\ r_s \\ r_m \end{bmatrix} . \qquad (6.1)$$

for all $\tilde{x} \in$ patch ,

where the values of L and R are taken from the vector position $\tilde{x}$ in the left and right image patches with respect to the centre of patch coordinate positions $\tilde{x}_l$

and $\tilde{x}_r$ respectively. The values R_x and R_y represent the first derivatives of R at the position $(\tilde{x}_r + \tilde{x})$. The parameters s_{11}, s_{12}, s_{21}, and s_{22} form the affine transformation matrix, and the parameters r_s and r_m represent the radiometric offset and contrast respectively. Note that the values for $\tilde{L}$ need only be evaluated once, and are constant for subsequent iterations. The collection of ε's for all $\tilde{x}$ in the patch forms a set of over-constrained linear equations which can be solved using matrix methods. Expressing equation (6.1) in this form, where $l = \tilde{L}$, $A = (r_m R_x \; x r_m R_x \; \cdots \; R)$, and $\alpha = (\Delta x_r \; \Delta s_{11} \; \cdots \; r_m)^T$ for all $\tilde{x}$ in the patch, the matching equation is written as

$$\varepsilon = l - A\alpha . \tag{6.2}$$

Equation (6.2) can be solved iteratively by using standard least-squares techniques ([Golub83], [Ralston85]), where the solution which minimises the error ε is obtained. One of the reasons for linearising the problem in this way, is that equation (6.2) is in the form of a Gauss-Markov estimation model. With some assumptions about the covariance of the errors in the final estimates of the parameters, this model gives a direct estimate of the standard deviations of the parameter estimates [Gruen85]. This “error measure” is discussed further in 6.5.2. Through each successive application of equation (6.2), the Δ's of each parameter are adjusted. When the changes to the parameters are less than a specified threshold, the solution is found. The iterative process essentially *tracks* or homes in on the solution from a given set of starting parameters. Thus, it is important for the initial estimates of the parameters to be near the correct solution, to avoid *tracking* toward an incorrect, but locally minimal solution.

To solve equation (6.2) numerically, we multiply both A and l by A^T (A transpose), and take the inverse of A^TA. Thus,

$$A \; \alpha = l \; , \tag{6.3}$$

$$(A^TA) \, \alpha = A^T \, l \; ,$$

and

$$\alpha = (A^TA)^{-1} A^T \, l \; .$$

To compute the inverse $(A^TA)^{-1}$ Gauss-Jordan elimination is used. Golub [Golub83] gives a more detailed discussion.

Initial estimates of the affine and radiometric parameters are typically

$$\begin{bmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \; r_s = 0 \,, \text{ and } \; r_m = 1 \,. \tag{6.4}$$

These starting values are respectively, the identity matrix (no affine transformation), no radiometric offset, and a radiometric contrast ratio of unity. The initial values for the translational parameters x_r and y_r must be either pre-calculated or estimated from a nearby match.

As it is impractical to manually select starting estimates of the translational parameters x_r and y_r for all intended match points, Otto and Chau [Otto89] have proposed a method by which these parameters can be estimated from points for which a match has already been obtained. This *region-growing* technique is also used in the implementation of Gruen's algorithm presented here. First, a small set of *seed* points are manually or automatically selected. Second, Gruen's algorithm is run on patches about these seed points to obtain a set of matched points. From this initial set of matched points, the location of approximate matches for neighbouring points can be predicted, to which least-squares matching can be applied. As this process is iterated, a complete set of matched points is *grown* out from an initial (small) set of seed points. Otto and Chau found that typically only two or three seed points are required [Otto89]. For this implementation, the seed points are selected manually from obvious matching features or marks in the images, such as bifurcations in the retinal vasculature.

6.2.2 Pseudo-code Description

The Gruen stereo matching algorithm [Gruen85] with the region-growing extensions proposed by Otto and Chau [Otto89] is described in pseudo-code in figure 6.1. Refer to the pseudo-code description for the following discussion.

```
function Gruen
    sample a patch from the left image about the point $\tilde{x}_l$
    iterate
        resample a patch from the right image about the point $\tilde{x}_r$
        calculate the first derivatives for each point in the patch
        construct the design matrix $A$ and the vector $l$
        solve equation (6.2) to obtain $\Delta$'s for each parameter
        update the parameters
    until $\Delta$'s < threshold or maximum iterations exceeded
initialise
    input a left and right pair of stereo images
    input 1 or more approximate (seed) matches between the images
    initialise the termination conditions (thresholds and maximum iterations)
    initialise a list of matches $Q$ to be empty

    for each approximate (seed) match $(\tilde{x}_l, \tilde{x}_r)$
        call function Gruen
        if Gruen converges on a solution
            store this match in $Q$
    next seed
matching procedure
    while $Q$ is not empty
        select a match $q$ from $Q$
        remove $q$ from $Q$ and print $q$
        for each "neighbour" $n$ of $q$
            if $n$ is not already matched
                predict $\tilde{x}_r$ for $n$ from $q$
                call function Gruen
                if Gruen converges on a solution
                    store this match in $Q$
        next "neighbour"
```

Figure 6.1. A pseudo-code description of the stereo matching algorithm

In the stereo matching algorithm, the matching procedure continues until the list Q is empty, and thus all potential matches that can be reached from the initial seed points are attempted. The definition of "neighbour" is a parameter which can be variable, depending on the desired output resolution. Typically, "neighbours" are defined to be points in both the x and y directions which are a given distance away (say 5 pixels). Such an implementation of the region-growing algorithm results in a grid-like coverage of the left-hand image, where the grid spacing is specified by the "neighbour" distance.

In the function Gruen, the left-hand image need only be sampled at discrete image values, as this is the target image to which the right-hand template must be matched. The process of resampling the right-hand image, however, will be at coordinate positions which are not discrete values as $\tilde{x}_r$ varies over each iteration. To obtain sub-pixel values to form the right-hand patch, bilinear interpolation of nearby whole-numbered pixels is used. It is assumed that meaningful sub-pixel values can be estimated from the discrete image values using bilinear interpolation. The implications of this assumption are addressed in chapter 7.

One final note is the question of convergence. A ceiling on the maximum number of iterations is set in the function Gruen due to the fact that the algorithm may fail to converge on a local minimum. Such failures are not stored in the list of matches Q. Further, for those points which do converge to a solution within the maximum number of iterations, a measure of "goodness" can be assigned to the match. Gruen [Gruen85] proposes that the largest eigenvalue of the 2×2 matrix composed of the co-variances of the translational parameters x_r and y_r (the stereo disparity), correlates well with magnitude of the error associated with the match. Otto and Chau [Otto89] employ this measure as a means by which to order the list of matches Q. Such an ordering results in match predictions which are taken from the best match first. Match predictions which fail to converge may be tried again from any one of their other three "neighbour" points (each of which will provide a slightly different prediction).

6.2.3 Analysis of Computational Load

In this section the computational requirements of the Gruen-Otto stereo matching algorithm are addressed. The amount of computer processing time is of importance when considering time constraints in a clinical environment. Therefore, we will decompose the algorithm into the types of basic operations that are likely to be executed by a microprocessor; additions, multiplications, and square roots. Let us assume that we wish to match a square patch of size M by M in the left-hand image to a patch in the right-hand image. Further, let us assume that we will allow for n parameters (e.g. in equation (6.1) we have $n = 8$). Otto and Chau [Otto89] have determined the number of each type of operation for the main components of the algorithm, resulting in the following estimate of computational load for one iteration of equation (6.2):

$$\begin{aligned}\text{Computation operations} &= \left(\frac{n^2}{2}+n+7\right) M^2 \text{ multiplications}\\ &+ \left(\frac{n^2}{2}+n+13\right) M^2 \text{ additions}\\ &+ \left(\frac{n^3}{6}+2n^2\right) \text{ floating-point operations}\\ &+ n \text{ square roots} + \text{housekeeping}\end{aligned}$$

The time to execute an addition, multiplication, or floating point operation will vary depending on the speed of the computer hardware used. If we assume that each type of operation will take equal time t, we can evaluate a rough estimate of the total computation time. For an image of size N × N, where a match is desired at every p-th pixel, and an average of i iterations is required for each match, the total compute time required can be estimated as:

$$\text{Total compute time} = \frac{t\ i\ N^2}{p^2}\left[(n^2+2n+20)\,M^2 + \left(\frac{n^3}{6}+2n^2+n\right)\right]$$

Using typical values for M and n (e.g. $M = 30$ and $n = 8$), and assuming the average time of each operation is 1µs, we get a rough estimate of about 45ms per iteration of equation (6.2). For a typical image of size 512 × 512 pixels, and assuming that we wish to obtain a match for every (say) fifth pixel (not including pixels near the edge of the image, where a full patch cannot be sampled), we must match 100 × 100 image points. If, on average, 3 or 4 iterations are required to reach a solution (Otto and Chau have shown that by using the estimated starting parameters obtained from the region-growing method, convergence is usually obtained in a few iterations), then we can estimate a total compute time of about 25 minutes for the entire matching process. When special purpose computer hardware is used, or if calculations are computed in integer format whenever possible, this time might be reduced significantly.

6.3 Experimental Methods

In this section the experimental methods for the complete process of stereo matching the plaster model and fundus image data are presented. This is not intended to be an exhaustive survey of methods for data acquisition, image preprocessing, and post-processing, but rather to outline the methods employed in this research, and the limitations imposed by the form in which the data are originally available.

6.3.1 Data Acquisition and Pre-processing

There are four sources of data for this research, as discussed in chapter 2. Data Set A is a plaster model designed to test the accuracy of the stereo matching algorithm on a physical object of known dimensions, whose shape is a first-order approximation to the optic disc region. The images for Data Set A were digitised from a video camera observing the model from two separated positions (calibrated to simulate simultaneous stereo images, although the same camera is used in both positions). Data Sets B and C are provided as colour transparencies which must be digitised from a suitable video source. Data Set D, on the other hand, is provided as digital data, which was quantised into 256 grey levels from the analog output signal of the Scanning Laser Ophthalmoscope (see chapter 3) at a resolution of 256 × 256 pixels (this resolution is a limitation of the equipment available at the Institute of Ophthalmology). Thus, we will concern ourselves with the methods and equipment used to digitise Data Sets A, B and C.

The method employed to obtain each digitised image for Data Sets B and C consists of four main steps:

(1) mount the transparencies on a light table and cover the area surrounding the slide with matte black paper;

(2) adjust the rack-mounted vidicon camera to a height suitable for viewing the central 6mm of the slide (the area of the optic disc on the slide) and focus the image;

(3) use the CRS-4000 framestore to digitise a set of 16, 512 × 512 pixel resolution images;

(4) process the image set to reduce the effect of inherent electrical line noise in the system.

For step (1), a large matte black sheet of cardboard was prepared with a small rectangular region, the size of 35mm transparency, cut out of the centre. In step (2), due the scale of the region of the transparency, a RCA TC1002 video camera was used with a Tarcus 4207 (f2.8/28mm) macro lens, to provide sufficient magnification. The lens aperture was set to f5.6, which, although visually did not appear to yield high image contrast, reduced the amount of signal saturation in the resulting image. In the fundus images, the area surrounding the optic disc contains dark vascular structures and the central region is often bright by comparison. Poor choice of aperture can result in an image where the vascular structures are 0 in grey value, and the central region is saturated to 255 in grey value.

In order to minimise the effects of electrical line noise and to improve the signal, in step (3) a series of 16 frames were digitised at a resolution of 512 × 512. In step (4), the resulting images were then point-wise averaged to obtain the final image for the Data Set. The number of frames required varies with any given system, and it has been determined experimentally that a minimum of 10 images will remove 95% of additive electrical line noise in the CRS-4000 system used.

It is prudent to consider whether any further pre-processing is desirable at this stage. There are a large number of image enhancement techniques in the literature (e.g. noise filtering [Castleman79], bandpass filtering [Bracewell65], histogram modification [Ballard82], and spatial filtering [Gonzalez87]), but there is little quantitative justification for their use in this application. Gruen has shown empirically that the application of spatial filters prior to image matching can improve the match result, but the choice of filter is application (image) dependent [Gruen86]. A better understanding of the potential source and magnitude of any image distortion or noise is required before any further pre-processing of the digitised fundus images can be undertaken methodically.

6.3.2 Applying Stereo Matching

Given a digitised and pre-processed pair of stereo images, a number of algorithmic parameters must be set in order to begin the Gruen-Otto matching process. The actions required for parameter determination include:

(1) select a small set of seed match points to begin the matching process;

(2) select the match window size (the size of the image patch surrounding a given match point);

(3) select the termination criteria or thresholds on the changes to the parameters, below which the match is considered to be correct (i.e. converged);

(4) select region-growing parameters defining the matching boundaries and points to be attempted within those boundaries (i.e. the *neighbour* spacing).

For each Data Set, a small number of *seed* match points were selected manually by visual inspection. Given that the seed points themselves were processed using the algorithm, only approximate matches were required. The selection of the match window size shall be deferred for the moment. The termination criteria (the test for convergence) are a set of parameter specific thresholds, below which a given parameter is considered to have converged to an acceptable solution. When the changes to all the parameters were below their corresponding thresholds, the match solution was accepted. These thresholds were set such that any further iterations of the algorithm would not yield any appreciable difference in the location of the match result. The region-growing parameters were simply the set of boundary points within which the region-growing could take place. The *neighbour* distance was specified to define the regular grid-like coverage of the left-hand image.

The selection of the match window size is non-trivial. A number of researchers have investigated methods for automatically selecting an optimal window size given *a priori* information about the signal and noise ([Barnea72], [Moravec77], [Shirai87]). With the fundus image data, it has only been possible to make a qualitative assessment of the signal content versus the noise content. In this preliminary application of the stereo matching algorithm, the match window size was assigned by examining the image spectrum to determine the frequency distribution (i.e. how many frequency components for a given window size), the grey-level statistics to determine signal content, and the convergence of the least-squares algorithm on a few trial match windows. As is demonstrated in chapter 7, a more quantitative relationship can be made between window size, signal content and noise. Initially, it was sufficient that the window size was large enough such that the matching algorithm can converge to a solution.

6.3.3 Post-processing and Data Representation

Following the matching procedure, the resulting image point matches can be interpreted in a number of different ways. Generally, when converting from pixel disparities to meaningful measures of depth, the computed geometric and radiometric parameters which define the transformation of the right-hand image patch to the left-hand image patch are not taken into consideration, and only the resulting match coordinates are utilised.

Recall from chapter 4 equations (4.3) and (4.5), which specify the relationship between match points in the left and right-hand images, given information about the stereo image geometry. In Data Set A, the plaster model, the stereo image geometry is known, and thus a meaningful transformation from pixel disparity to real-world depth values can be made through the application of these equations. Therefore, the estimates of disparity for the plaster model will be expressed in real-world units.

Conversion of disparity estimates to real-world values for Data Sets B and C is complicated by a lack of information not only about camera geometry, but also about the optical properties of the subject's eye. In order to fully covert the disparity estimates to real-world depth estimates, values for the stereoscopic baseline, the refractive error of the eye, the axial length of the eye, and the curvature of the cornea (keratometer reading) are required. One complication is that the focal power of the eye (the eye's own magnification) is determined by the focal power of the lens and the curvature of the cornea at the viewing position. The focal power varies across the cornea by as much as ±2 diopters [ElHage73]. This variation introduces a refractive error. When viewing the eye stereoscopically, both views will not be subject to the same focal power (magnification) or refractive error.

To overcome the difficulties of measuring the eye's individual optical properties, an average value for the focal power of the eye is often used [Littmann82]. Littmann has shown that even with a value for the axial length of the eye and a measure of the corneal curvature, only an estimate of the focal power of the eye can be made. The average or idealised eye, then, is assumed to have a focal length of 17mm and no refractive error. However, in order to make such an estimate, the fundus camera must be positioned such that the nodal point (focal point) of the camera lens is congugate to the nodal point of the lens of the eye [LeGrand80]. If the operator does not align the nodal point of the camera with that of the eye (independent of focus), then the perceived magnification will not correspond to the estimate of 17mm as the focal length of the eye. In other words, operator alignment error can introduce an unknown optical magnification in the system.

Therefore, for Data Sets B and C, where a value for the stereoscopic baseline is known but little is known about the optical parameters of the patient's eye, the disparity values will not be fully converted to real-world depth estimates. However, if an idealised eye is assumed, a relative scale (in microns) can be assigned to the resulting disparity estimates. The scale associated with Data Sets B and C will be discussed with the results. Although the conversion to real-world depth values involves a simple scaling, in order to avoid making unfounded claims of matching accuracy the topography of the optic disc region will be represented by plotting the disparity values alone.

The presentation of 3-dimensional information in a format which best highlights those aspects of the data which are of interest is non-trivial. Indeed, a great deal of effort is often required to present 3-dimensional data adequately. In order to highlight many of the interesting aspects of the stereo matching data, three methods were chosen for displaying the match results:

(1) a grey scale (depth-shaded) image of disparity values (depth values);

(2) a 1-dimensional cross-section of disparity (depth) through the surface near points of interest (compared with idealised depth values where applicable);

(3) a 2-dimensional contour plot of the disparity values (depth values).

In the following section, the results from the application of the Gruen-Otto stereo matching algorithm to Data Sets A, B, C, and D are presented in this form.

6.4 Experimental Results

In this section the experimental results from the application of the Gruen-Otto stereo matching algorithm to the four data sets are presented and discussed. The results for each Data Set are reviewed in turn, and the common stereo matching problems encountered are summarised in the following section.

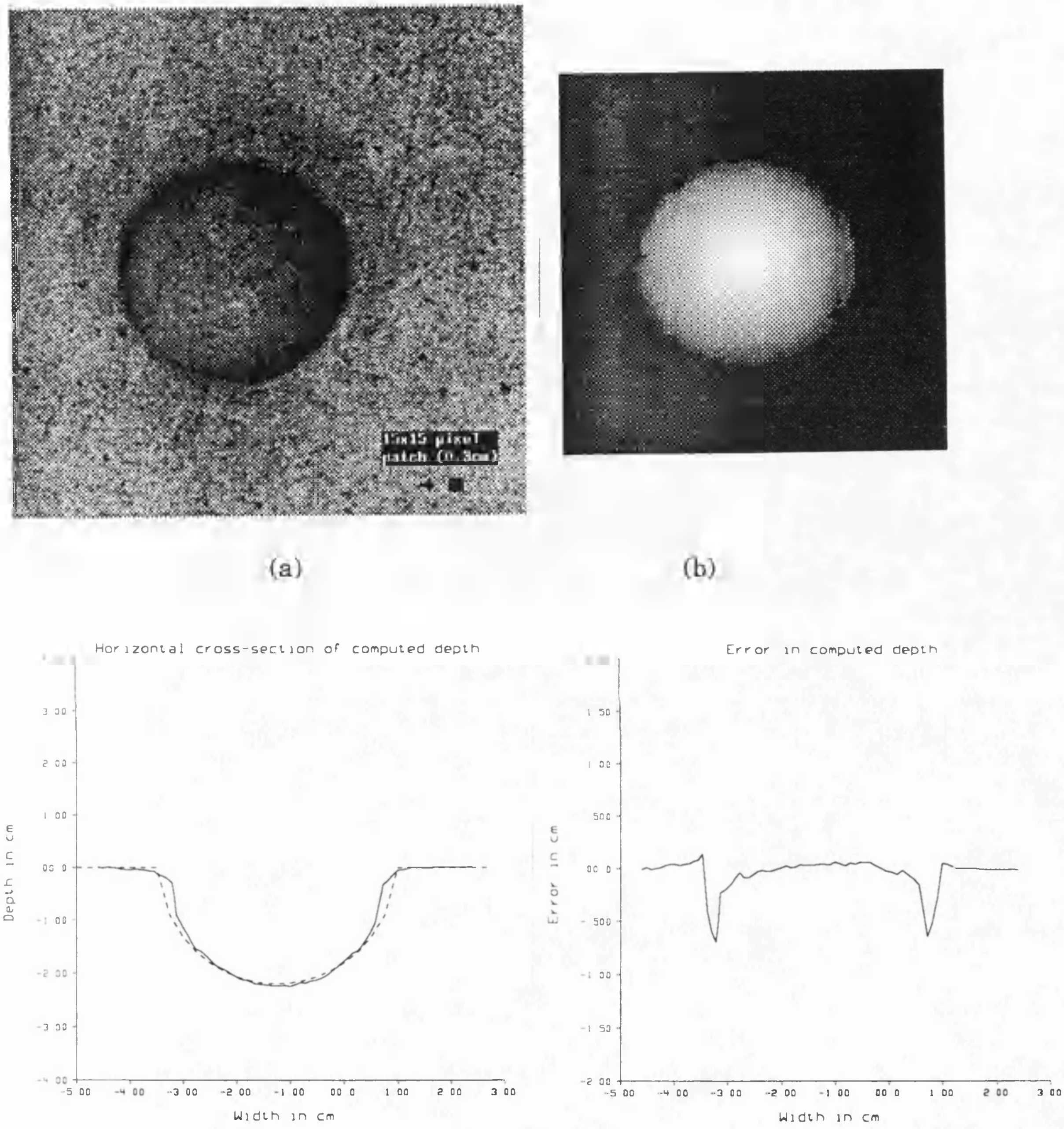
6.4.1 Data Set A: Plaster Model Image Pair

A stereo pair of 512×512 pixel resolution images of the plaster model, Data Set A, were obtained using the methods outlined in section 6.3, where a sequence of 16 images were digitised and averaged for both the left and right view. Two seed match points were manually selected; one lying centrally in the depression, and the other in the surrounding flat region. The match window size was set to 15×15 pixels (corresponding to 0.3cm), and the region-growing neighbourhood set to a spacing of 5 pixels. Due to the large stereoscopic shift in the right image as compared to the left image, only a central region of 400×400 pixels could potentially be matched. The physical depth of the hemisphere is 2.2cm. The (x, y) parameter thresholds for the least-squares iterations were set to 0.01 pixel displacements; that is, for a Δx and Δy less than 0.01, the match was considered converged. Similar thresholds were set for the affine parameters.

The output from the stereo matching program is a list of corresponding coordinates in the left and right-hand images. These disparity values are then converted using the geometric relationships shown in chapter 4, given by equations (4.3) and (4.5). As the geometric parameters defining the stereoscopic relationship between the left and right-hand images are known, the estimates of disparity (in pixels) are easily converted into estimates of real-world depth (in cm). The converted results of the stereo matching are shown in figures 6.2 and 6.3.

In figure 6.2a, the original left-hand image is shown with a sample pixel patch of size 15×15 pixels superimposed in order to show the relative scale of the mask window size. Figure 6.2b shows a depth shaded image of the estimate depth values (over the matchable 400×400 image region), where the brighter values represent increased depth. Figures 6.2c and 6.2d show a horizontal cross-section of estimated depth (solid line) compared to an idealised cross-section (dashed line), and the error in computed versus ideal depth, respectively. Although the computed depth does not match the expected depth exactly, it is understood that the plaster model will not quite have such an ideal shape (due to unknown factors such as shrinkage during drying). Moreover, the largest errors near the rim of the cup depression indicate, as is the case in the real plaster model shape, that the rim has a slight curvature; that is, the plaster model does not have the ideal right-angle entrance to the cup depression.

Of interest is the mean error in estimated depth. As the stereo matching algorithm gives sub-pixel estimates of depth, it is not surprising that in regions of the image with little depth change (as in the bottom of the cup) the actual depth estimate is highly accurate. The mean error across the entire matched region is less than 0.25mm in magnitude, and more specifically, the mean error of the horizontal cross-section shown in figure 6.2c is 0.51mm. Figure 6.3 shows a contour plot of the estimated depth, where each contour line represents a 1mm interval in depth.



(a) (b)

(c) (d)

Figure 6.2. Stereo match results of Data set A, the plaster model stereo images, showing (a) original left-hand image, (b) grey-scale (depth-shaded) image of converted disparity values, (c) horizontal cross-section of computed depth (solid line) and the idealised cross-section (dashed line), and (d) the error in computed depth.

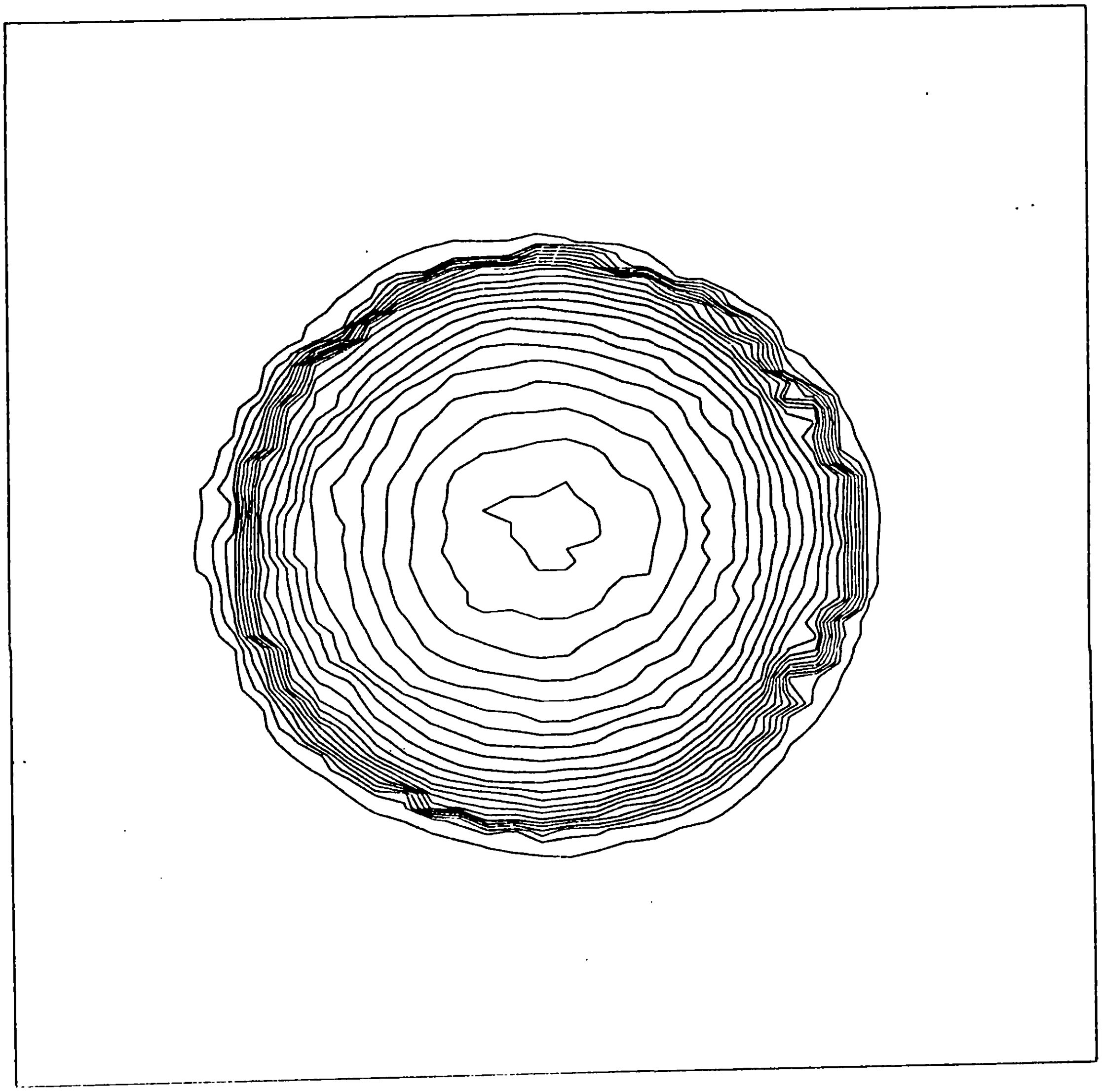


Figure 6.3 Contour plot of Data Set A stereo match results. The contour lines represent 1mm intervals in depth.

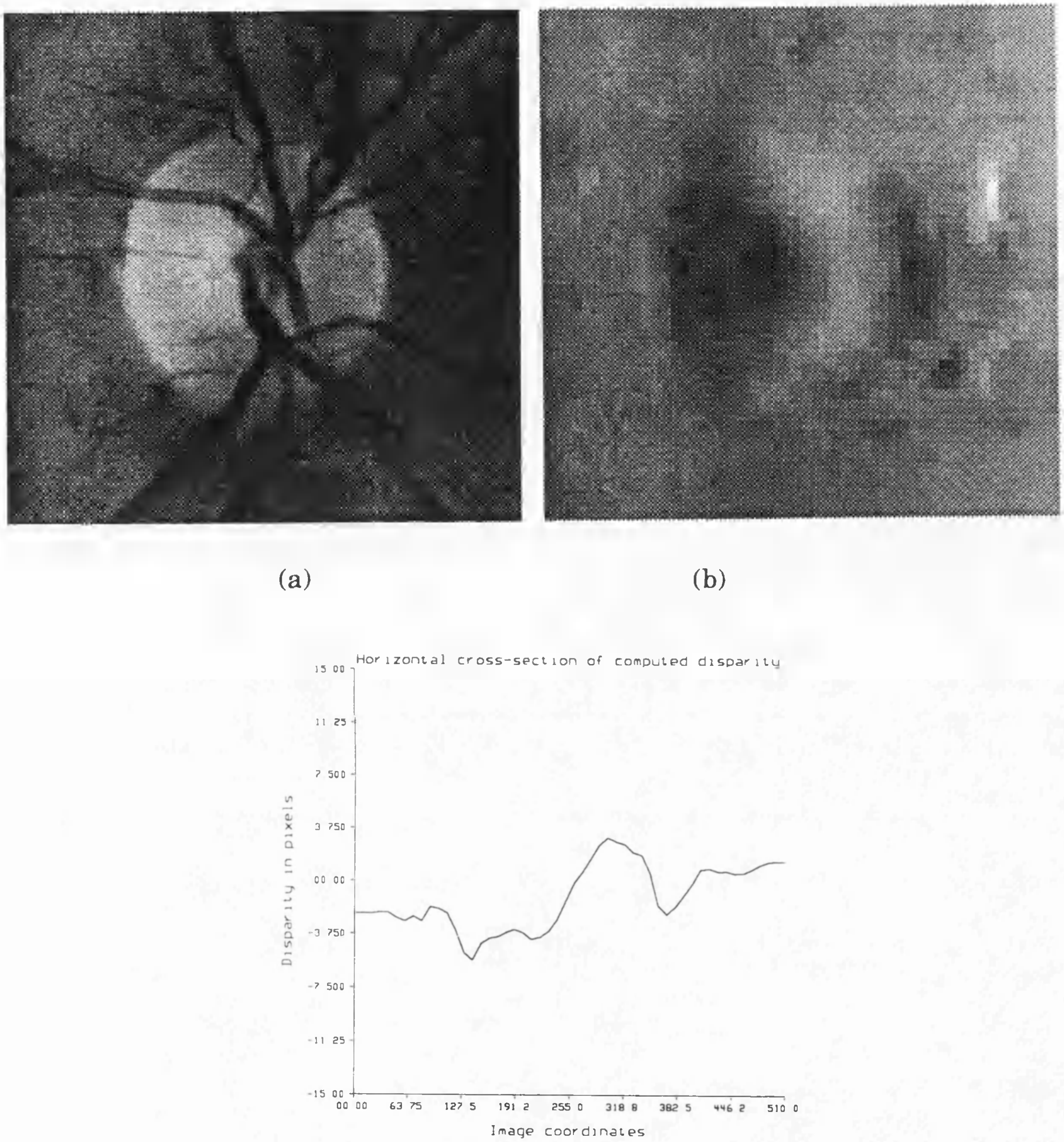
6.4.2 Data Set B: Non-simultaneous Fundus Image Pair

A stereo pair of 512×512 pixel resolution images of the standard fundus camera data, Data Set B, were obtained using the methods outlined in section 6.3. One seed match point was selected manually on an arterial bifurcation near the centre of the optic disc. Due to a slight shift in the right-hand image, a potential match area of only 450×450 pixels results. The match window size was set to 35×35 pixels, with a neighbourhood spacing of 10 pixels. The total number of points to be matched was 2025. Due to the size of the match window, and the number of least-squares iterations required to converge to a solution for each match point, a neighbour spacing of 10 pixels, rather than 5 pixels, was used to reduce the overall computation time. As each potential match point was solved independently, using a grid coverage of 5 pixels did not significantly affect the results, but would have increased the computation time by a factor of four.

As discussed in chapter 2, the data were obtained from a standard fundus camera using the method of non-simultaneous fundus photography. A special apparatus called a bite-bar was mounted to the head mount of a fundus camera system, so as to ensure (as near as possible under such conditions) that the patient's head would remain in the same position for both photographs. The main potential source of error from the patient, then, is eye movement, and a fixation light is used to provide a visual target to minimise this error. The fundus camera itself was carefully shifted laterally 3mm to obtain the second view (without refocusing). This is considered to be the "best case" non-simultaneous stereo fundus image data.

As discussed in section 6.3.3, conversion of the disparity esitimates to real-world values is complicated by a lack of information about the optical properties of the patient's eye (e.g. spherical abberations). However, the idealised eye with a focal length of 17mm was assumed in order that a relative scaling can be assigned to the resulting disparity values. By applying equation (4.5) from chapter 4, a relative scale between units of pixel disparity and units of real depth (microns) can be found. The stereoscopic baseline was assumed to be 3mm (neglecting any corneal refraction or spherical aberations). The lateral magnification of the fundus camera is also known, and is 3.4×. By measuring a small area on the film transparency and dividing by the number of pixels covering that area, a lateral scale for 1 pixel can be approximated (neglecting diffraction and defocus effects). For Data Set B, the lateral scale for 1 pixel at unit magnification was approximately 7μm. Applying equation (4.5), the scale for 1 pixel disparity is 40μm (in the plane of best focus). Therefore, the horizontal cross-section of disparity shown in figure 6.4c corresponds to approximately 400μm from the highest to the lowest point.

Figures 6.4 and 6.5 show the results of applying the stereo matching algorithm to Data Set B. In figure 6.4a, the original left-hand image is shown. Figure 6.4b shows a depth shaded image of the disparity estimates, where darker values indicate increased depth. Regions of the image lying outside the matchable 450×450 pixel area are extrapolated from match results (for display purposes only). Interestingly, the central area clearly shows the raised vascular structure often visible in stereoscopic optic disc photographs (this is caused by supporting tissue which elevates the vessels near the optic disc). Figure 6.4c is a horizontal cross-section of computed disparity through the centre of the optic disc area (scanline 260). Finally, figure 6.5 shows a contour plot of the computed disparity, where each contour line represents a 1 pixel (approximately 40μm) disparity interval. One can infer from this figure a possible boundary for the optic disc, but marking the true location is is still very much a subjective task.



(a) (b)

(c)

Figure 6.4. Stereo match results of Data Set B, the non-simultaneous stereo fundus images, showing (a) original left-hand image, (b) grey-scale (depth-shaded) image of disparity values, and (c) horizontal cross-section of computed disparity through the line $y = 260$.

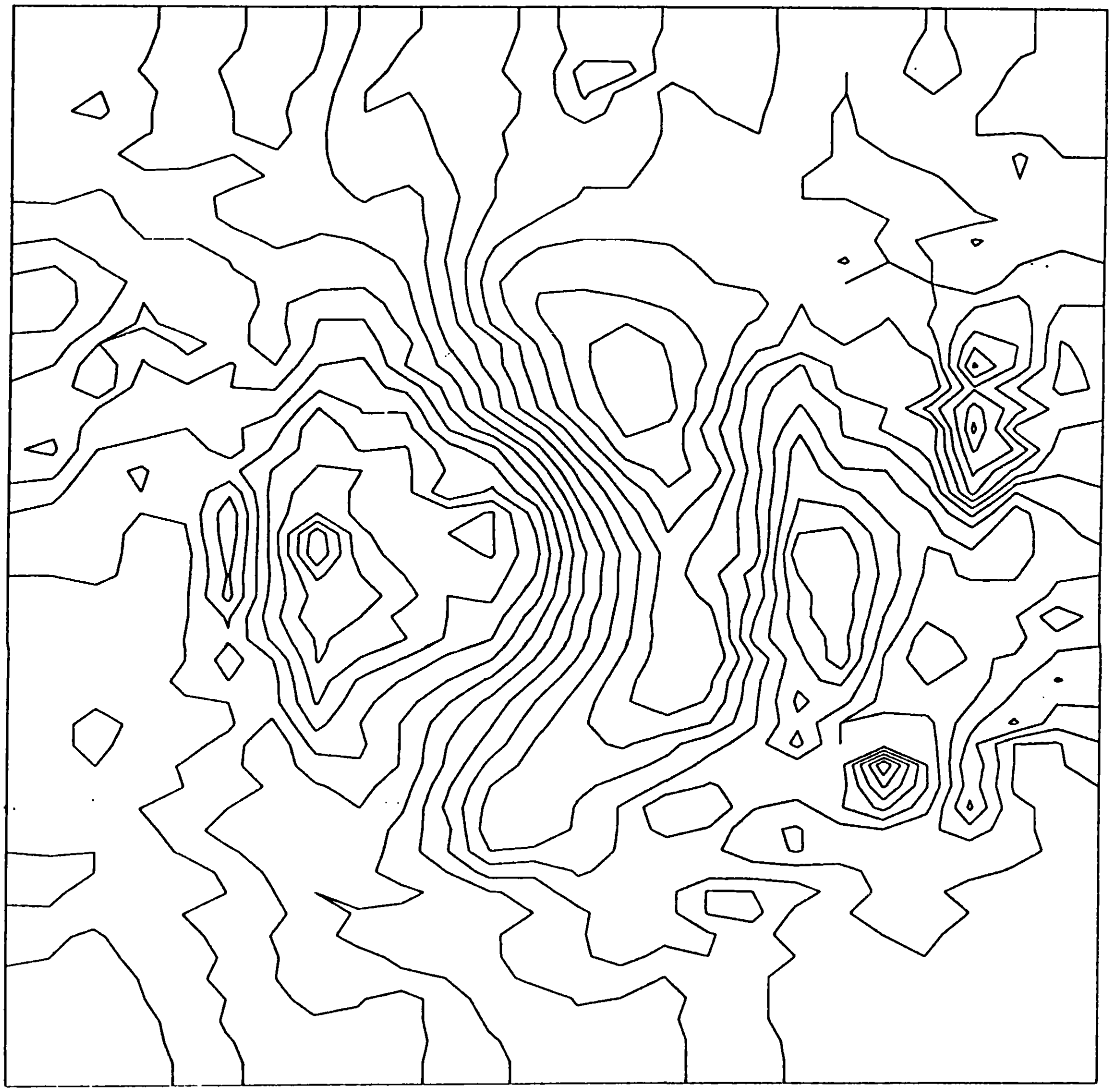


Figure 6.5 Contour plot of Data Set B stereo match results. The contour lines represent 1 pixel disparity intervals.

6.4.3 Data Set C: Donaldson Fundus Image Pair

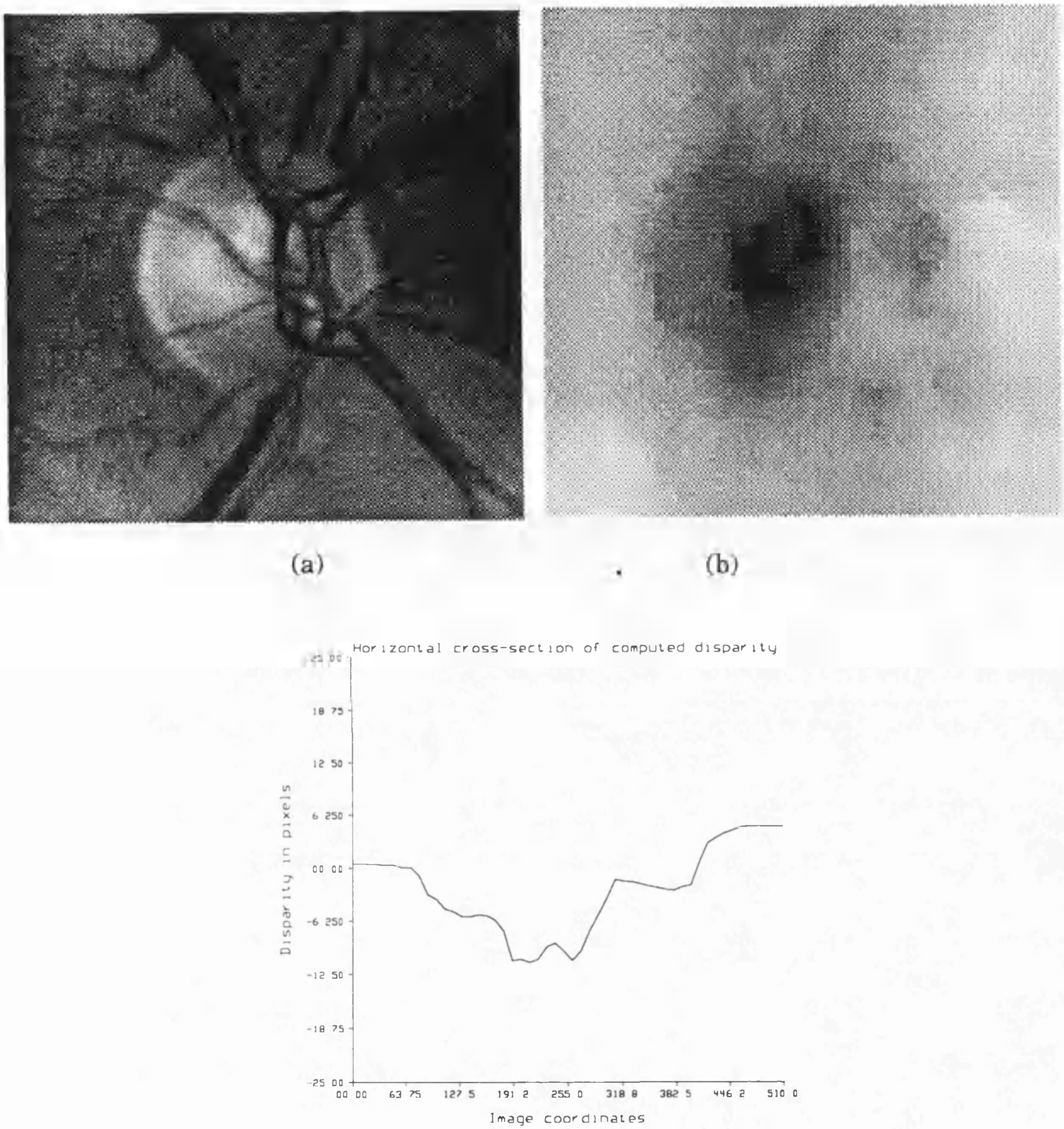
A stereo pair of 512×512 pixel resolution images of the Donaldson stereo fundus camera data, Data Set C, were obtained using the methods outlined in section 6.3. One seed match point was selected manually near the centre of optic disc. The match window size was set to 35×35 pixels, with a neighbourhood spacing of 10 pixels. The potentially matchable area ranges over 480×480 pixels, and thus the total number of points to be matched is 2300. As discussed for Data Set B, a neighbour spacing of 10 pixels was selected to minimise the computation time without affecting the results.

A description of the Donaldson stereo fundus camera was given in chapter 3, and a discussion of the physical constraints associated with obtaining photographs follows in chapter 4. The photographic image pair used here has a stereoscopic baseline of 3mm. One potential source of error from the patient is eye movement, causing motion blur during exposure. Further, as noted in chapter 4, obtaining two views through a small pupil requires that each view passes through less than favourable portions of the lens of the eye. This may cause unknown distortions in the stereo images not due to the stereoscopic effect.

As discussed for Data Set B, a relative scale between units of pixel disparity and units of real depth (microns) can be found by applying equation (4.5). For Data Set D, the stereoscopic baseline is known, and is 3mm. The lateral magnification on the film is also known, and is 3×. By measuring a small area on the film transparency, dividing by the number of pixels covering that area, and further dividing by the magnification, a lateral scale for 1 pixel at unit magnification was found to be approximately 8μm (ignoring defocus and diffraction effects). By applying equation (4.5), and assuming an idealised eye with a focal length of 17mm, the scale for 1 pixel disparity is found to be approximately 45μm (in the plane of best focus). The horizontal cross-section of disparity shown in figure 6.5c corresponds to approximately 650μm from the highest to the lowest point. If the optic cup begins at a depth of 150μm below the disc rim (see chapter 2), then cup in figure 6.5c has a measured depth of approximately

220μm.

Figures 6.6 and 6.7 show the results of applying the stereo matching algorithm to Data Set C. In figure 6.6a, the original left-hand image is shown. Figure 6.6b shows a depth shaded image of the disparity estimates, where darker values indicate increased depth. Regions of the image lying outside the matchable 480×480 pixel area are extrapolated from match results (for display purposes only). Figure 6.6c is a horizontal cross-section of computed disparity through the centre of the optic disc area (scanline 250). Finally, figure 6.7 shows a contour plot of the computed disparity, where each contour line represents a 1 pixel disparity interval.



(a) (b)

(c)

Figure 6.6. Stereo match results of Data Set C, the Donaldson stereo fundus images, showing (a) original left-hand image, (b) grey-scale (depth-shaded) image of disparity values, and (c) horizontal cross-section of computed disparity through the line $y = 250$.

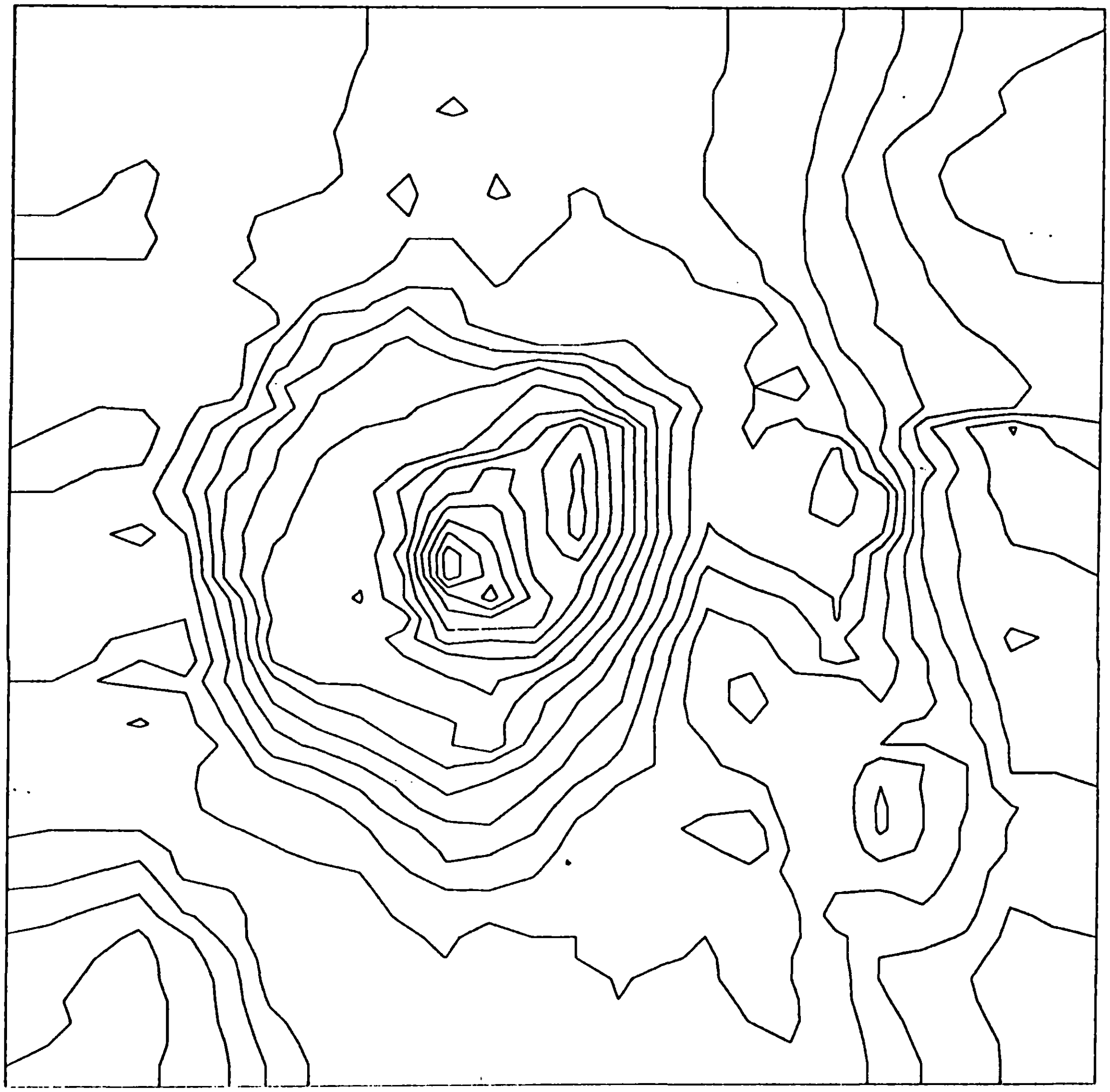


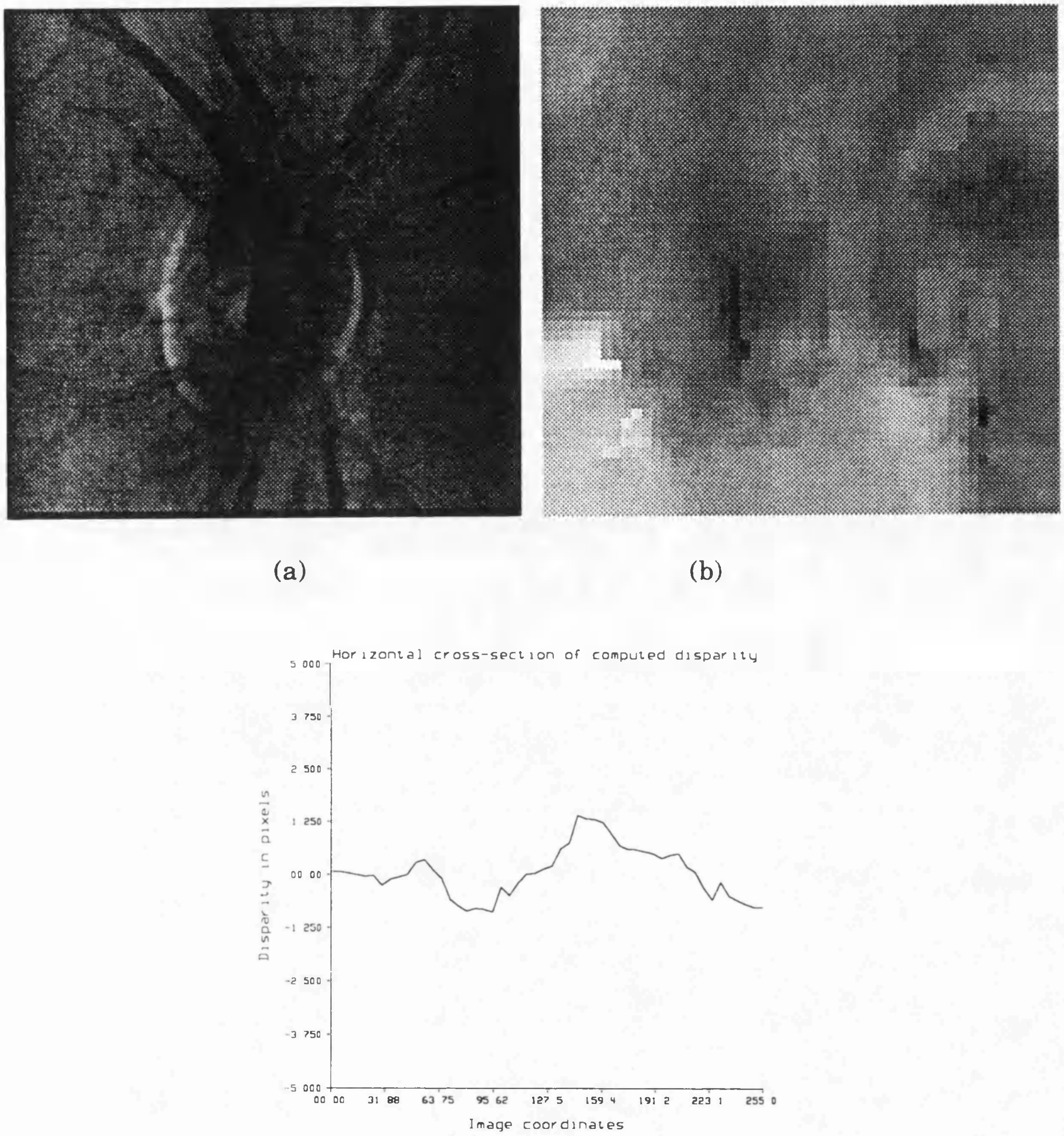
Figure 6.7 Contour plot of Data Set C stereo match results. The contour lines represent 1 pixel disparity intervals.

6.4.4 Data Set D: SLO Image Pair

A stereo pair of 256×256 pixel resolution images of the Scanning Laser Ophthalmoscope (SLO) data, Data Set D, were obtained by direct digitisation of the analogue output signal from the SLO (via an Acorn ARM computer and a Wild Vision V10 frame grabber), and were provided "as is" by the Institute of Ophthalmology, University of London. Experimentally, the images are obtained in a similar manner to Data Set B, where the camera is shifted laterally to obtain the second view. For this data, the stereoscopic baseline is 2mm. The SLO represents a relatively new technology in ophthalmic imaging, and therefore this data has been used to provide an early indication of the benefits of this new and highly innovative camera. The results obtained here are simply an indication of the potential of this fundus image source.

One seed match point was selected manually near the centre of optic disc. The match window size was set to 20×20 pixels, with a neighbourhood spacing of 5 pixels. This relatively large window size is required to overcome the high levels of electrical line noise present in both images. One of the disadvantages of the SLO data is that although it is visually of high contrast in the centre of the disc area, there is a lot of noise in the system. Also note that the images visually have a different appearance to that of the fundus photographs, due to the imaging model of the SLO (see chapter 3, section 4), and are low in resolution.

Figures 6.8 and 6.9 show the results of applying the stereo matching algorithm to Data Set D. In figure 6.8a, the original left-hand image is shown. Figure 6.8b shows a depth shaded image of the disparity estimates, where darker values indicate increased depth. Figure 6.6c is a horizontal cross-section of computed disparity through the centre of the optic disc area (scanline 130). Finally, figure 6.9 shows a contour plot of the computed disparity, where each contour line represents a 1 pixel disparity interval.



(a)

(b)

(c)

Figure 6.8. Stereo match results of Data Set D, the SLO stereo fundus images, showing (a) original left-hand image, (b) grey-scale (depth-shaded) image of disparity values, and (c) horizontal cross-section of computed disparity through the line $y = 130$.

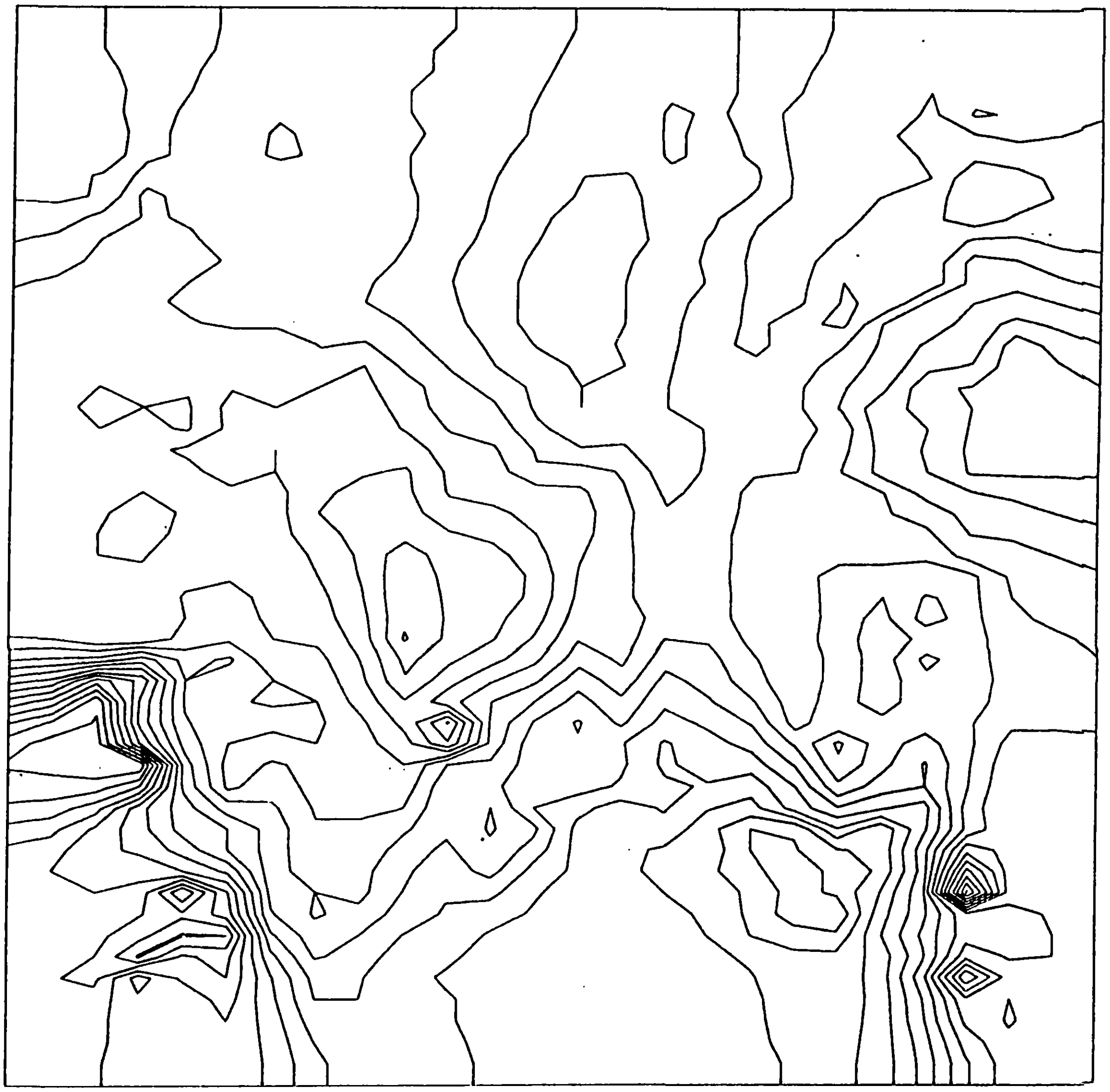


Figure 6.9 Contour plot of Data Set D stereo match results. The contour lines represent 1 pixel disparity intervals.

6.5 Analysis of Results

In this section a summary of the results presented in section 4 is given, highlighting the stereo matching problems encountered with each data set. An analysis of matching accuracy is presented and known methods for measuring matching error are briefly discussed. Finally, some of the common problems encountered in the application of the Gruen-Otto stereo matching algorithm to the fundus image data are discussed.

6.5.1 Overview

In order to test the stereo matching algorithm, a physical model is used. This approach has been previously employed by Rosenthal using the Zeiss Model Eye, as a means by which to verify clinical and photogrammetric estimates of estimated optic cup depth [Rosenthal80]. The model is a simple plaster cast with a hemispherical depression. Although image data from such a model are not directly comparable with fundus image data, they serve as a useful test-bed for the Gruen-Otto stereo matching algorithm. The results from stereo matching the plaster model image data show that accurate measurements of disparity (depth) can be obtained when the image contains sufficient signal for robust cross-correlation, and when the stereoscopic baseline is sufficiently large. The mean matching error is less than 0.2 pixels displacement, corresponding to 0.25mm in depth.

For the fundus image data, the results indicate that a dense set of disparity values can be obtained across the optic disc region in the fundus images, and that "plausible" estimates of depth can be derived. Although several additional parameters are required to yield real-world estimates of depth for the fundus image data, plotting the disparity values themselves provides a good indication of the potential sources of error in the stereo matching process. The results from matching Data Set B, the non-simultaneous stereo fundus images, indicate a topography which is biologically plausible. One interesting point is that with the presence of a large raised vascular structure which bisects the optic disc vertically (see figure 6.5), it becomes difficult to infer, even manually, the location of the optic disc boundary. As these data were obtained with a 3mm stereoscopic baseline, qualitatively they provide acceptable results.

The results from stereo matching Data Set C, the Donaldson (simultaneous) stereo fundus images, are (qualitatively) an improvement over Data Set B. This is not a surprising result, given the consistent stereoscopic baseline of 3mm, simultaneous image acquisition, and a reduction in error caused by patient movement. Simultaneous stereo fundus images, obtained with a relatively large baseline, provide what must be considered "ideal" stereoscopic data. The larger stereoscopic baseline (and fixed camera geometry) qualitatively gives improved results over Data Set B, and it is much easier to match depth map features to anatomical features.

In the horizontal cross-section of disparity shown in figure 6.6c, a small plateau is visible before the contour drops into what appears to be the optic cup. It is interesting to compare the estimated depth of this plateau (and thus the estimated depth of the top of the cup) with the clinically accepted range of 120 to 150μm, which is taken to indicate the depth of the optic cup relative to the position of the optic disc (see chapter 2). Using the scale value of 45μm of depth to 1 unit of pixel disparity (from section 6.4.3), the plataeu (or cup entrance) lies at a measured depth of approximately 200μm. In figure 6.7, an elliptical relief is clearly visible in the contours, and represents a possible boundary for the optic disc. However, it must be made clear that such shapes and topographical features are not directly comparable among data sets from different patients; it may simply be the case that this particular data set has a deeper optic cup than that of Data Set B. Although figures 6.6 and 6.7 show results which are more anatomically plausible than the results shown in figures 6.4 and 6.5, there is no indication that the results are more accurate due to simultaneous stereo image acquisition.

Finally, the results from stereo matching Data Set D, the Scanning Laser Ophthalmoscope stereo fundus images, are presented only to provide an indication of the potential benefit that this new and highly innovative camera may provide. Figure 6.9 shows that some anatomical features are recognisable and that a slight elliptical relief can be detected. This is a promising, if not surprising, result given the low resolution, very small stereoscopic baseline (2mm), and high noise content present in the images. It is clear that with future improvements to the SLO system, the potential exists for accurate stereoscopic measurement of the optic disc.

6.5.2 Error Analysis and Accuracy

One difficulty in the stereo matching process is the detection of gross mismatches, or blunders. It is possible that a point match will have converged to a locally minimal solution, but that the solution is not correct. The presence of such blunders is visible in, for example, figure 6.4b, where a bright region in the right-hand side of the disparity map is due to a region-growing error. An error in one match point can spread to neighbouring matches since starting seed values for neighbour match points are estimated from points which have already been matched. What is required, then, is some indicator of the "goodness" of the estimated match. With such a measure, blunders could be detected during the matching and region-growing process, thereby limiting the overall matching error.

The linearised equation (6.2) can be expressed in the form of a Gauss-Markov estimation model, by making some simple assumptions about the distribution of the errors (see [Gruen85]). By formulating the correlation procedure mathematically in this way, the least-squares model leads to an unbiased minimum variance estimator. The precision of the estimated parameters is expressed by the co-variance matrix

$$K = \sigma_\varepsilon^2 \, Q = \sigma_\varepsilon^2 \, (A^T A)^{-1} \tag{6.5}$$

following the notation in section 6.2.1, where

A is the matrix of observations,

$\varepsilon = l - A\alpha$ is the residual vector,

$\sigma_\varepsilon^2 = \frac{1}{r}\varepsilon^T\varepsilon$ is the variance of the residuals,

n is the number of observations,

u is the number of transformation parameters,

$r = n - u$ is the redundancy, and

Q is the cofactor matrix.

Of particular interest are the specific standard deviations

$$\sigma_{\Delta x} = \sigma_\varepsilon (q_{\Delta x})^{1/2}$$

$$\sigma_{\Delta y} = \sigma_\varepsilon (q_{\Delta y})^{1/2} . \tag{6.6}$$

of the translation parameters Δx and Δy, because they describe the precision of the match location. Here, q_j indicates the jth diagonal element of Q. Gruen proposes that the size of the standard deviations of the translation parameters and the correlations between themselves (the eigenvalues of a sub-matrix of the elements of K) may give useful information concerning the stability of the solution [Gruen85]. Further, Otto and Chau [Otto89] use the largest eigenvalue of the 2 × 2 sub-matrix formed from Q for the parameters Δx and Δy to give an indication of the accuracy of the match location. Currently this measure is used to prioritise the region-growing list, but not as a direct indicator of the accuracy of the match.

The results from Data Set A, for which real error estimates are known, provide a means by which to verify the use of the Otto and Chau [Otto89] eigenvalue measure as an indicator of the accuracy of disparity estimates. The values for the eigenvalues obtained for matched points from Data Set A do correlate well with the observed error in the match; that is, a larger eigenvalue corresponds to a larger real error in estimated depth. Figure 6.10 plots the (scaled) error measure together with the observed error for the horizontal cross-section of computed depth shown in figures 6.2c and 6.2d.

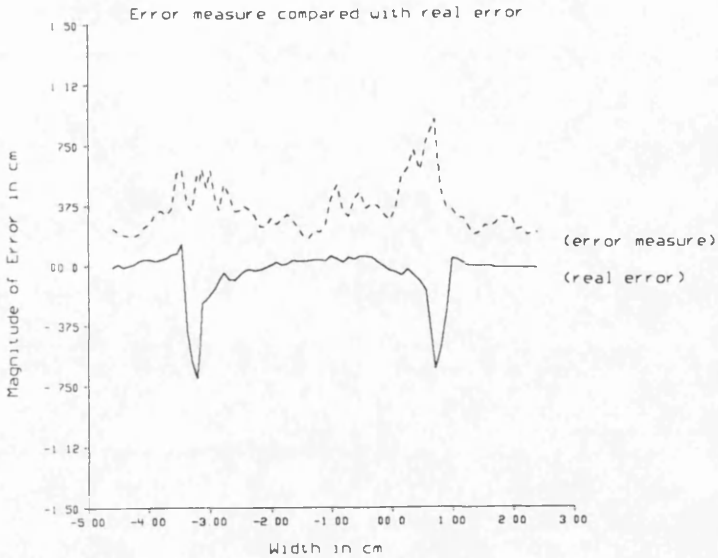


Figure 6.10. Plot of the largest eigenvalue measure (see text) and the observed error for matched points in the horizontal cross-section from Data Set A (figure 6.2). The error estimator has been scaled for illustrative purposes.

Although the eigenvalue measure does correlate well with the magnitude of the error in estimated depth, the derivation of the (co)variances depends on assumptions about the image and noise statistics which are not easily generalisable. Using this measure does not provide a quantitative means by which to threshold matching errors so as to discard matching blunders; the relative accuracy of a match can be expressed, but not the absolute accuracy. In chapter 7 a more rigorous treatment of the problem of estimating matching accuracy in the presence of additive noise is presented, with the specific aim of resolving this problem.

6.5.3 Isolation of key problems

We conclude the analysis of the results from stereo matching by isolating some of the key problems common to all four data sets. Several of these problems are addressed in chapters 7 and 8, and are simply listed here as warning signs, pitfalls, and dilemmas encountered during the experimental stage. In chapter 5, a least-squares stereo matching algorithm is chosen because it overcomes some of the obvious difficulties of correlating the fundus images. However, the algorithm has some hidden problems:

Window Size

The selection of the mask window size, the size of the region surrounding the point to be matched, is ad hoc at best. If *a priori* information about the signal and noise characteristics are available, a mask window size can be more appropriately selected (see [Barnea72], [Secilla88]). Once selected, the window size is fixed for all attempted matches. One may ask, can the mask window size be made adaptive, so as to improve (locally) an individual match. The results obtained for the fundus image data, where the signal content varies across the image, may well be improved by some method for dynamically determining mask window size on a match point by match point basis.

Parameter Ranking

The selection of the set of parameters which model the transformation of the right-hand image window to the left-hand image window is not a straightforward task. Here, the set of affine parameters allowing for planar geometric distortions are modelled. The question arises as to the individual weighting assigned to each parameter; that is, should all the parameters be weighted equally (in the least-squares sense)? In applications where some detail about the camera geometry is known, constraints on some parameters can be added in the form of a weighting matrix (see [Rosenholm87a]).

Alignment

One of the prime considerations in selecting the Gruen-Otto stereo matching algorithm is that no image alignment prior to processing is required. It is assumed that due to the parameter tracking property of the least-squares approach that y and Θ alignment prior to processing would be unnecessary (and would result in resampling the right-hand image twice - once for alignment, again during matching). There is the mathematical danger that the inclusion of y-related parameters has over-parameterised the problem and increased the likelihood of error (see [Keating85]).

Image Pre-filtering

In his 1986 paper, Gruen discusses the use of image pre-filtering to improve the match results (and the convergence of the minimisation). However, as noted earlier, there is little quantitative justification for the use of filtering in this application. Rosenholm has also indicated that the use of image pre-filtering can improve the convergence of the algorithm, provided there is sufficient signal content in the images [Rosenholm87b]. Can such image pre-filtering (median, average, edge enhancement) be meaningfully applied, and will its use significantly improve the accuracy of the results?

Detecting Errors

As discussed in detail above, the detection of match errors (blunders, false local minima) is far from simple. Gruen has expressed the linearised equations in the form of a Gauss-Markov estimation model, which allows for direct estimates of the (co)variances of the parameters [Golub83], as demonstrated by figure 6.10. Currently this measure is used to prioritise the region-growing list, but not as a direct indicator of the accuracy of the match. Can the minimum variance estimator measure be quantitatively verified and subsequently used as a means by which to filter inaccurate matches, thus improving the effectiveness of the region-growing strategy as well as the accuracy of the resulting disparity estimates?

6.6 Summary

In this chapter the computer implementation of the Gruen-Otto stereo matching algorithm is presented together with a brief analysis of the computational requirements. The experimental methods applied to four test data sets are discussed, including data acquisition, image pre-processing, stereo matching, and post-processing and analysis. The results from applying the stereo matching algorithm to each data set are reviewed. An analysis of the results highlights a number of important issues with respect to determining matching accuracy. Finally, several of the key problems associated with the stereo matching algorithm and the fundus image data are discussed.

In summary, a number of points related to the overall problem of stereoscopic fundus imaging have become clear. The importance of data standardisation becomes apparent, and a more consistent data format must be adopted by clinicians if the methods employed here are to be reliable. Similarly, the importance of stereoscopic baseline, discussed in chapter 4, is well demonstrated by comparing the results from Data Set B (3mm baseline, non-simultaneous), Data Set C (3mm baseline, simultaneous), and Data Set D (2mm baseline, simultaneous). The results from Data Set A indicate that the algorithm can work well in favourable conditions, and thus the importance of image quality (resolution, illumination and contrast, low noise) also becomes apparent. In conclusion, it is reassuring that the fundus image results represent biologically plausible estimates of optic disc topography, but locating the optic disc remains a subjective task.

Chapter 7
Modelling the Data

7.1 Introduction

In this chapter, the use of mathematical models to explore the structure of auto- and cross-correlation functions is presented. Stereo data for which no known ground truth can be obtained, present problems for conducting error analysis on correlation measures. This has prompted the need to explore the behaviour of correlation algorithms, and more specifically the shape of the correlation function over which the algorithm is to be applied. Consider a sampled two-dimensional function $f(x,y)$ which is to be examined to see if a small portion of the sampled data matches a reference pattern $g(x,y)$. The process of searching $f(x,y)$ to locate that portion which best matches $g(x,y)$ is called *template matching*. Algorithms for organising the search process to find the best match are based on the concepts of *correlation*. The term *correlation algorithm* is used synonymously with the term *search process* in this chapter. Where the term correlation is to refer to the actual mathematical process, this will be made explicit.

Correlation algorithms, such as the stereo matching algorithms discussed in chapter 5, manage the process of moving the template across the scene and making some measure of the quality of the match at any given position. The correlation measure can be thought of as a two-dimensional function or surface on which, ideally, a well defined peak exists representing the location in $f(x,y)$ where a near-copy of $g(x,y)$ is to be found. When $g(x,y)$ is identical to some part of $f(x,y)$, the process is termed *auto-correlation*. When $g(x,y)$ differs from $f(x,y)$, for example two different views of the same scene as with stereo, the process is termed *cross-correlation*. Often, the ability to find the correct match of $g(x,y)$ in $f(x,y)$ is complicated by the structure of the functions themselves, or by deficiencies in the correlation algorithm itself. False correlation maxima may be present in the correlation surface, and they may be larger than the maximum of the correct match. Noisy data may distort or obscure the correlation peak, or the structure of the scene and/or template may be such that the shape of the correlation surface near the maximum is excessively broad, making it difficult to *accurately* locate the best match position. The issue of *accuracy* has to be addressed.

It would seem important to be able to predict the performance of a given correlation algorithm for specific scene types, though surprisingly, relatively few theoretical predictions have been published. Such analyses have usually been confined to general discussions of unspecified scene structures (see [DeSoete84], [Dvornychenko83], [Parthasarathy86], [Sclove81], [Therien84]), or at the other extreme, to experimental work on specific scene types, which is predominately empirical in nature (see [Gruen85], [Caelli88], [Chan88], [Lee90], [Rosenholm87a,b]). To predict the performance of a particular correlation algorithm from its theoretical description is extremely difficult, while at the same time it is equally difficult to extrapolate performance measures from a limited set of published experimental results. The goal here is to derive explicit formulae which permit predictions of the performance of standard correlation algorithms (such as that used in stereo matching) in terms of a few easily-definable scene parameters.

To begin, the various parameters which characterise the fundus images are studied in section 7.2. This leads to the development of a functional model of local scene structure which exhibits the same characteristics as the real fundus image data. With a suitable model of scene formation, in section 7.3 the classical autocorrelation formulae for a scene defined by this model are derived. The formulae are extended to sum of squared difference correlation in section 7.4, as this reflects the nature of the stereo correlation algorithm selected in chapter 5. In section 7.5, the addition of gaussian noise to the scene model and its effects on the correlation formulae are discussed. The corruption of the correlation function by additive noise leads to a set of formulae for predicting the likelihood of a correlation error for a given set of scene parameters. The auto-correlation formulae are then extended to cross-correlation as discussed in section 7.6. The derived formulae are shown to provide a useful set of tools for predicting correlation performance on specific scene types.

7.2 Image Models

In this section, a study of the various parameters which characterise the fundus images is presented. This will allow us to devise a model of local scene structure which exhibits the same characteristics as the real fundus image data. With a suitable model for scene formation, the mathematical structure underlying image correlation can be explored.

7.2.1 Local Models of Image Structure

To examine the various parameters which characterise the fundus image data, well known image models which classify images by grey-level distribution statistics [Pratt78], stochastic texture models [Conners80], and Fourier spectral content [Bracewell65], are reviewed with the aim of determining a suitable model of local image structure. It is shown, in due course, that a functional image model not only provides a flexible model for a large class of image types, but also allows us to examine the functional shape of the image and its corresponding auto-correlation surface. Modelling images by a function of a few easily definable parameters will enable theoretical predictions about the variability of the correlation of such images with respect to the parameters themselves. Predictions about convergence, accuracy, and probability of false correlation maxima (or minima) can be developed from the functional model.

It is important to understand, qualitatively, the types of parameters which might affect the result of a correlation algorithm. This will give us a set of criteria by which to assess a particular image model. Correlation, as discussed in 7.1, is the process of searching an image $f(x,y)$ to see if a portion of the data matches a reference pattern $g(x,y)$. Note that the size or extent of $g(x,y)$ is not defined. If, for example, the process of image alignment is required, $g(x,y)$

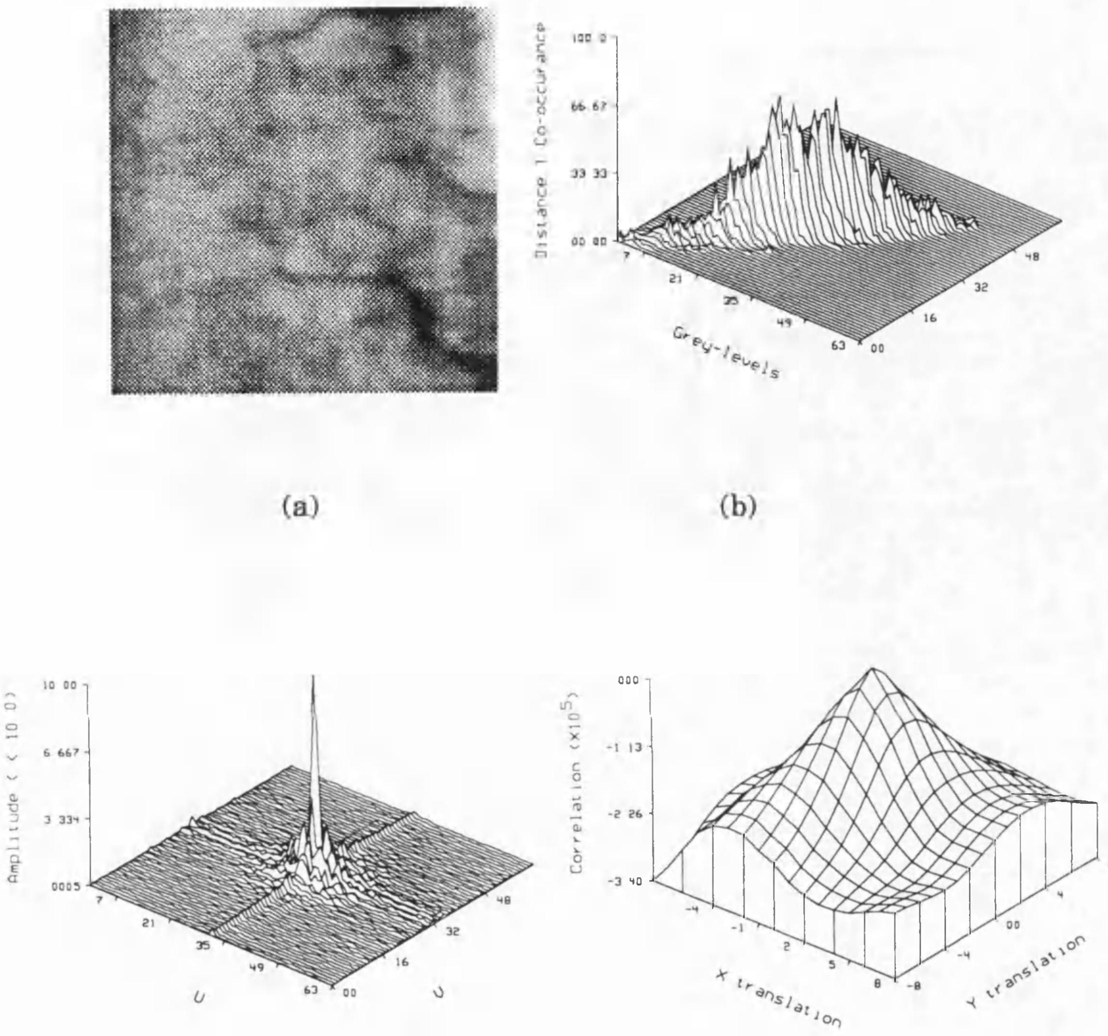
will be near-equal in size to $f(x, y)$. If, on the other hand, one wished to identify a particular object in $f(x, y)$, $g(x, y)$ may only cover a small portion of the total area of $f(x, y)$. The question arises as to how large a portion of $f(x, y)$ must be modelled. The answer for our particular application, that of stereo correlation, is that $f(x, y)$ need only be modelled locally over a small range. Indeed the large scale structure of $f(x, y)$ may be quite different from that seen locally over a small portion of the image, but if $g(x, y)$ is to be matched at this local scale, then the large scale structure will not greatly affect the correlation result. Hence, the model will be confined to local image structure.

The second question which arises is how "image structure" is defined. Briefly, image structure can be defined to be the pattern or texture in a local region of an image. An image may contain several different patterns or textures but it is assumed that at the scale of $g(x, y)$ a unique pattern can be recognised. Although the pattern or texture in a local region of $f(x, y)$ can be modelled, it is also important to vary such parameters as image contrast (variance) and brightness, which may affect the correlation result. Orientation and scale of the pattern must be considered as well; the model should contain spatial information about the pattern if it is to be applicable to a larger class of images. To summarise, an image model is required which closely represents the image structure in a local region of $f(x, y)$ that can be tuned by parameters such as orientation, scale, contrast and brightness.

Naively, then, a portion of a fundus image is examined using some well known image descriptors. Figure 7.1 shows a 64 × 64 pixel region of a fundus image together with its grey-level co-occurrence statistics, its Fourier spectrum, and its auto-correlation function. The grey-level distribution statistics, as proposed by Haralick (see [Pratt78]), offer a method of examing the relationships between neighbouring pixels. These relationships loosely define the structure or texture in the image by giving a measure of the transition from one grey-value to another in the local neighbourhood of a given pixel. Note that these relationships are not unique to a given image, nor to a given class of images, since an infinite number of patterns or textures can exhibit similar distributions. Also note that the overall spatial information (orientation and scale) about the plăcement of individual grey-values is not retained. It therefore becomes difficult to infer the original image pattern from its grey-level statistics. As these measures are also non-isomorphic, they are not suitable as a means by which to model arbitrary image textures, and thus they will not be considered further.

Another approach to studying texture in the image follows the work of a number of researchers in stochastic image modelling (for example [Wong77], [Yokoyama78] and [Thomason87]). The premise is that natural textures can be modelled on the basis of a Gauss-Markov random field, and that modelling textures in this way enables the study and classification of natural textures ([Gagalowicz81], [Kaneko82]). A Gauss-Markov random field is an array of Gaussian distributed random variables which are governed by some conditional probability requirements which define the value of an image point. This approach has proved useful in synthesising natural textures from measured statistical properties of a natural image scene. Although for a given set of random variables the process of image formation is deterministic, the model itself is not easily generalised. For each pattern or texture which is to be modelled, a unique set of conditional probability requirements must be determined. This violates our criteria that the image model be easily generalisable and flexible. The complexity of the stochastic model approach makes it unattractive even for a limited set of image types. Extending the model to tunable parameters such as spatial orientation, contrast, and brightness is not straight-forward. Stochastic image modelling is attractive in that texture fields can be generated with specified statistical properties, but the mathematical complexity of the model makes it unattractive as a tool for exploring the mathematical structure underlying image correlation.

In Figure 7.1c, the Fourier spectrum of the fundus image region is shown. Apart from the large DC component, the shape of the remaining frequency components exhibit a rather simple shape or form. A large body of literature exists



(a) (b)

(c) (d)

Figure 7.1. A fundus image region and various image descriptors. (a) region of fundus image, (b) distance 1 grey-level co-occurrence matrix at 0^o, (c) Fourier spectrum, and (d) the 2-dimensional auto-correlation surface.

on the application of Fourier theory to image filtering, image analysis, and image correlation (see e.g. [Haykin86], [Castleman79], and [Rosenfeld82]).

Interestingly, a number of mathematical properties define the relationship between an image and its auto-correlation function, which can be expressed in terms of its Fourier spectrum [Bracewell65]. If a simple mathematical description of the shape of an image spectrum could be derived, meaningful information about its auto-correlation function would be readily available. As will be shown in section 7.2.2, a functional description of the shape of the Fourier spectrum can be derived which meets the desired criteria for an image formation model. Here a discussion of why this might provide a good model is appropriate.

The Fourier spectrum exhibits a number of properties which give the flexibility desired in an image model. First, the Fourier transform is an isomorphic transform, providing a simple means by which to translate between the functional (spectral) description and the image (spatial) domain. Second, the spatial relationships between features in the image are retained in the Fourier spectrum, and thus the model could accommodate changes in spatial orientation and scale of the resulting image texture. Third, changes to image parameters such as contrast (variance) and brightness are simply a function of the amplitude of the frequency components and thus are simply modelled by a scalar multiplier. Finally, the model can be expressed in an easily manageable mathematical form, whose parameters correspond directly to the types of visual image parameters one might wish to vary. This makes the model very general, and a wide range of image textures can easily be modelled. In the following section a model of local image structure is developed which is based on summing frequency components in the Fourier domain.

In summary, the local structure of the fundus image data has been examined with the aim of determining the parameters which characterise the pattern or texture present in the images. A set of criteria are given with which to assess any proposed image model. Well known image descriptors such as grey-level co-occurrence statistics, Gauss-Markov stochastic processes, and Fourier spectra are examined to evaluate whether they might provide a useful mathematical model for image formation which meets the desired criteria. With such a model the mathematical structure underlying image correlation can be explored. It is determined that a functional description based on the Fourier spectrum can meet the criteria for a flexible model of local scene structure.

7.2.2 A Functional Model of Local Scene Structure

In this section, a functional model of local scene structure which well approximates the scene structure of the fundus image data is derived. As discussed above, the fundus image data exhibit properties which can be described by the distribution of component frequencies in the image spectrum. Image spectra often exhibit a characteristic functional shape in amplitude and phase, where the phase component can be considered, but is not limited to, a random distribution between 0 and 2π.

As a functional approximation to the shape or structure of the amplitude and phase components of the spectrum of the fundus image, a model of image formation is proposed where image frequency components are summed to form discrete samples in the spatial domain. Initially the model is considered in one dimension for simplicity, but is extended to two dimensions thereafter.

The discrete image formation function $f(x)$ is defined as

$$f(x) = \sum_{n=0}^{n=n_c} a(n) \cos\left[\frac{2\pi\, n\, x}{N} + \Phi(n)\right] , \tag{7.1}$$

where

$a(n)$ defines the amplitude at frequency n, up to the cutoff frequency n_c,

n_c defines the cutoff frequency for the spectrum ($n_c < \frac{N}{2}$),

$\Phi(n)$ defines the phase at frequency n,

and N defines the image width.

The flexibility of this simple model becomes apparent through the selection of a suitable amplitude function $a(n)$. A suitable function $a(n)$ might be given by

$$a_1(0) = B \tag{7.2}$$

$$a_1(n) = b\ (1 - |\frac{n}{n_c}|), \quad 1 \le n \le n_c\ ,$$

which defines a linearly decreasing amplitude function with the phase component, $\Phi(n)$, considered to be randomly distributed between 0 and 2π. The DC component B in $a_1(0)$ allows for modelling varying background levels against a relatively low-contrast scene structure. The parameter b is a scalar which allows for varying the dynamic range of the individual frequency amplitudes; that is, the frequency components are scaled between 0 and b, except at $a_1(0)$. Some other typical image spectra can be modelled by, for example,

$$a_2(0) = B \tag{7.3}$$

$$a_2(n) = b, \quad 1 \le n \le n_c\ ,$$

and

$$a_3(0) = B \tag{7.4}$$

$$a_3(n) = \frac{b}{n}, \quad 1 \le n \le n_c\ .$$

The amplitude function $a_2(n)$ defines a spectrum which is flat and of amplitude b out to the cut-off frequency n_c, except for the DC component which has amplitude B. The amplitude function $a_3(n)$ defines a spectrum which follows a $1/F$ law over the range 1 to n_c, also with DC component B. The phases of the individual components are assumed to be randomly distributed between 0 an 2π. Figure 7.2 plots the shape of these example amplitude functions and the corresponding 1-dimensional scenes generated using equation (7.1).

To extend the model to the discrete 2-dimensional image $f(x, y)$, frequency components are simply modelled in two dimensions. Hence, $f(x, y)$ can be written as

$$f(x, y) = \sum_{s=0}^{s=n_c} \sum_{r=-n_c}^{r=n_c} a(r, s) \cos\left[\frac{2\pi}{N}(rx + sy) + \Phi(r, s)\right]. \tag{7.5}$$

Functions for $a(r, s)$ defining the amplitude of each frequency component can be obtained by extending the above 1-dimensional amplitude functions to their circularly symmetric or separable 2-dimensional equivalents.

linearly decreasing (circularly symmetrical):

$$a_{1c}(0, 0) = B \tag{7.6}$$

$$a_{1c}(r, s) = b \left[1 - \frac{\sqrt{r^2 + s^2}}{n_c}\right]$$

linearly decreasing (separable-axes):

$$a_{1s}(0, 0) = B \tag{7.7}$$

$$a_{1s}(r, s) = b \left[1 - \frac{|r|}{n_c}\right] \left[1 - \frac{|s|}{n_c}\right]$$

flat:

$$a_2(0, 0) = B \tag{7.8}$$

$$a_2(r, s) = b$$

$1/F$:

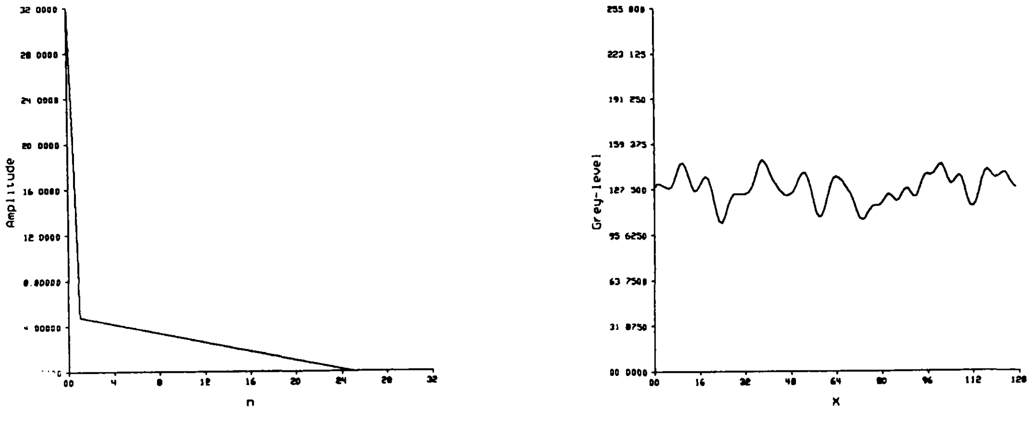
$$a_3(0, 0) = B \tag{7.9}$$

$$a_3(r, s) = \frac{b}{r\, s}$$

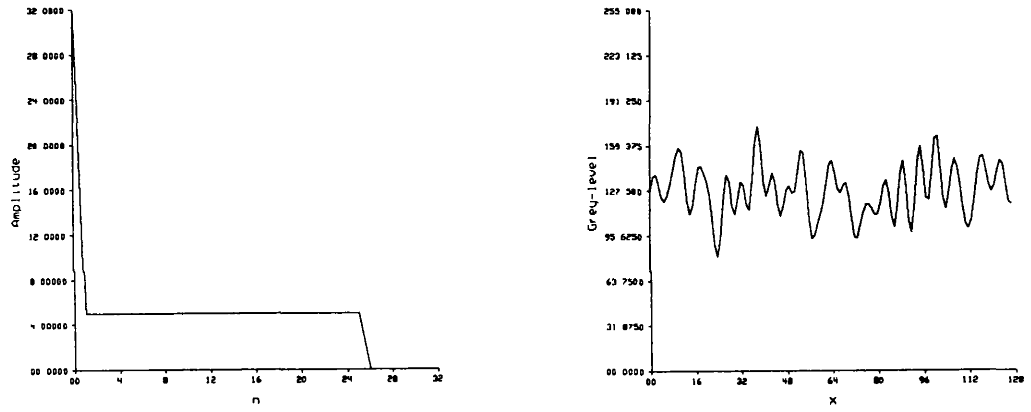
where

$$-n_c \le r \le n_c, \quad 0 \le s \le n_c \ .$$

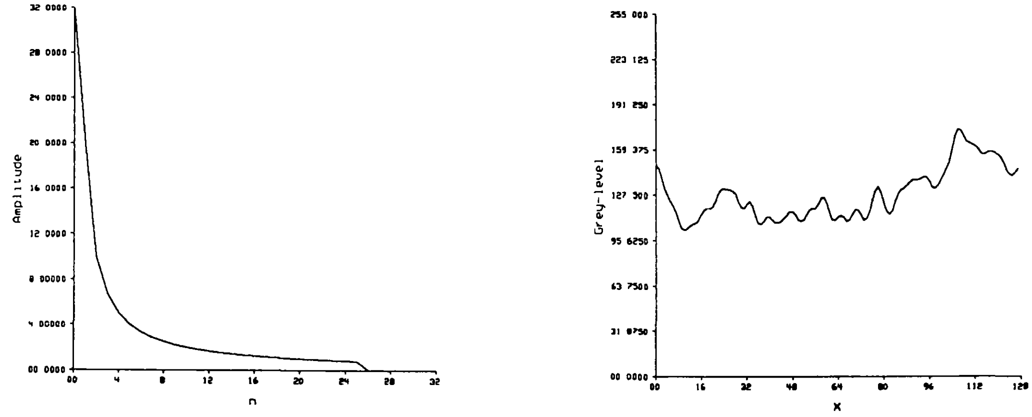
Equation (7.5) defines a 2-dimensional model for the formation of the discrete image $f(x, y)$. Note however that the summation over r extends to both positive and negative frequencies to create asymmetry in the scene structure, whereas the summation over s need only extend over the positive frequencies (extending the range of s would simply result in an overall doubling of the scene amplitude). One can ignore, for mathematical simplicity, the fact that the frequency components for which $r < 0$ and $s = 0$ should not strictly be included. The phase



a(*n*) Linear *f*(*x*)



a(*n*) Flat *f*(*x*)



a(*n*) 1/F *f*(*x*)

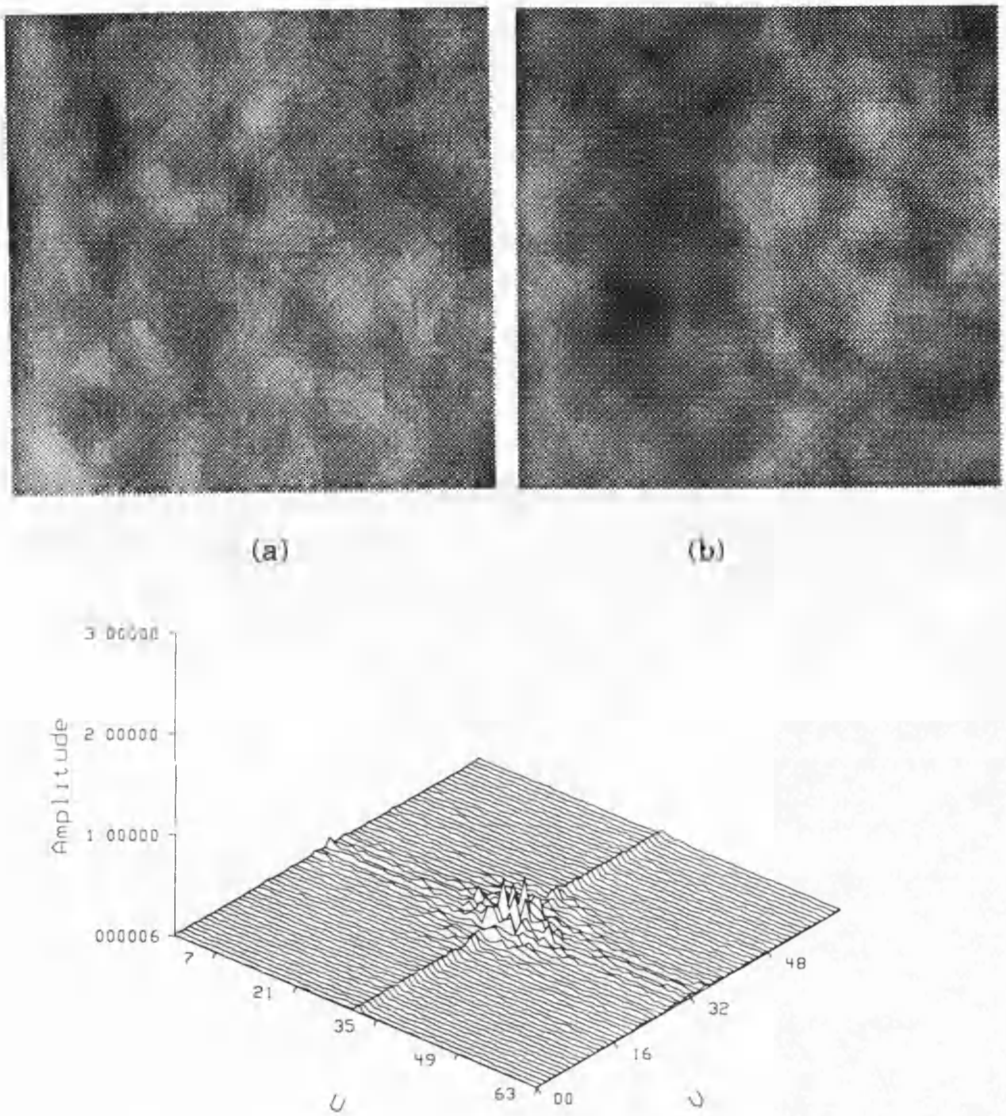
Figure 7.2. Some typical amplitude functions and corresponding values of $f(x)$ generated from equation (7.1).

components are again, without loss of generality, considered to be randomly distributed between 0 and 2π.

The functional image model defined by equation (7.5) provides a good qualitative approximation to the fundus image data, but a formal quantitative measurement of the suitability of this model is difficult to provide. As a first-order measure, figure 7.3 shows a region of a fundus image and a synthetic region generated using the 2-dimensional model of frequency distribution given by equation (7.9). Also shown is the absolute difference between the model image and the real fundus image spectrum. Despite small differences in the low order frequencies, there is good agreement between the spectrum of functional image model and the real fundus image. More generally, to give a visual (qualitative) feel for the flexibility of the functional model, figure 7.4 shows some images generated using the linearly decreasing, flat, and $1/F$ spectral models given here.

The aim is to develop a functional model of local scene structure to enable generation of synthetic images which closely models the fundus image data. Equation (7.5) defines a discrete 2-dimensional image formation function suitable for the task. However, the limitations of this model must be made explicit. Summing cosine components at discrete intervals in x and y may introduce undesirable aliasing artifacts in the resultant image. Further, not all real image spectra exhibit amplitude functions which are easily functionally defined, nor is it true that phase components will in general be random. Although the example amplitude functions have been defined as separable in r and s, this will not in general be the case. The amplitude function $a_3(r, s)$ provides the a good approximation to the fundus image data in local regions of predominantly nerve fibre.

The assumption of separability can be used to simplify the derivation of the correlation function (section 7.3), but one must be aware of the limitations of this assumption (see [Oppenheim75], [Therien84], and [Bracewell85]). Real image spectra often exhibit symmetric properties; however, a good separable approximation for $a(r, s)$ can typically be found for most classes of image. No restriction has been placed on the phase function $\Phi(r, s)$; a good functional approximation



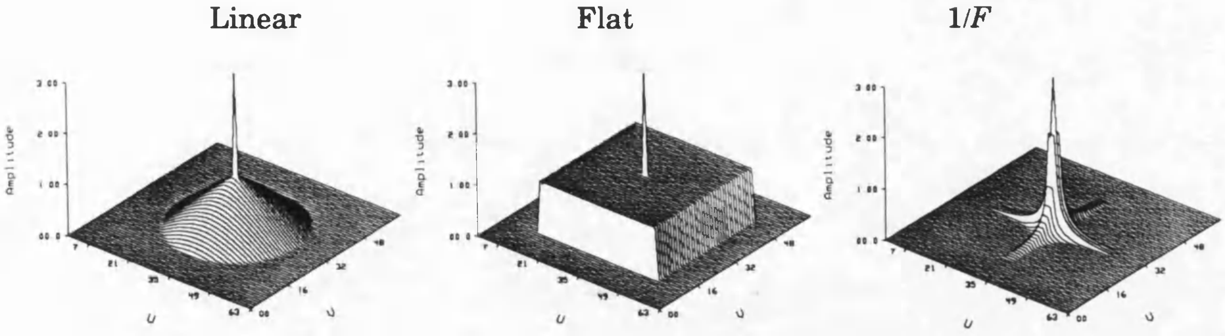
(a)

(b)

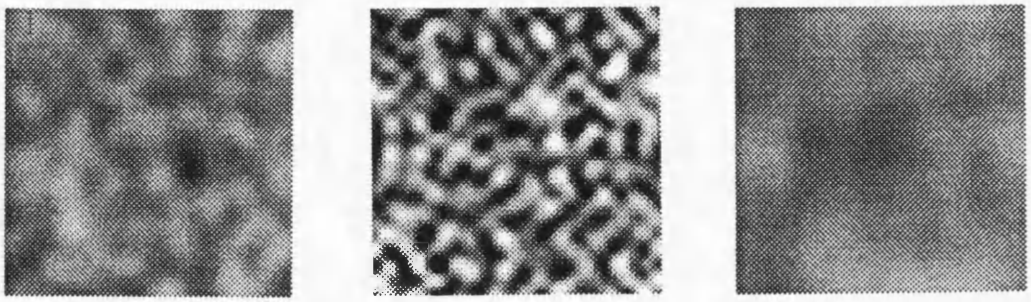
(c)

Figure 7.3. (a) real and (b) modelled region of a fundus image, together with (c) the absolute difference between the real fundus image spectrum and the model image spectrum given by equation (7.9).

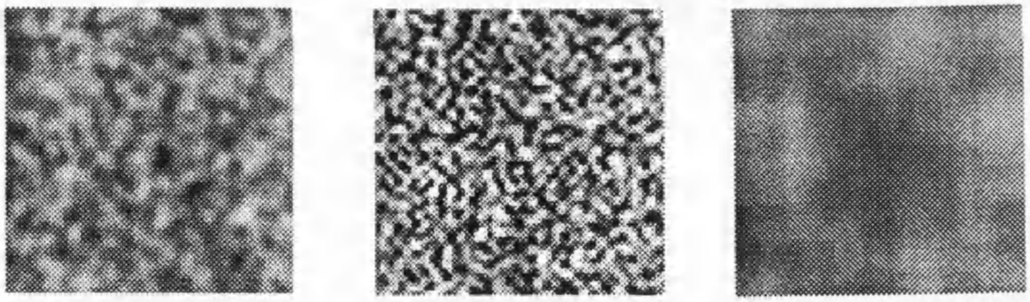
to $\Phi(r, s)$ can also typically be found for a given class of image. In the case of the fundus image data, $a(r, s)$ is well modelled by equation (7.9), and $\Phi(r, s)$ is well modelled by a random distribution of phase between 0 and 2π. Figure 7.4



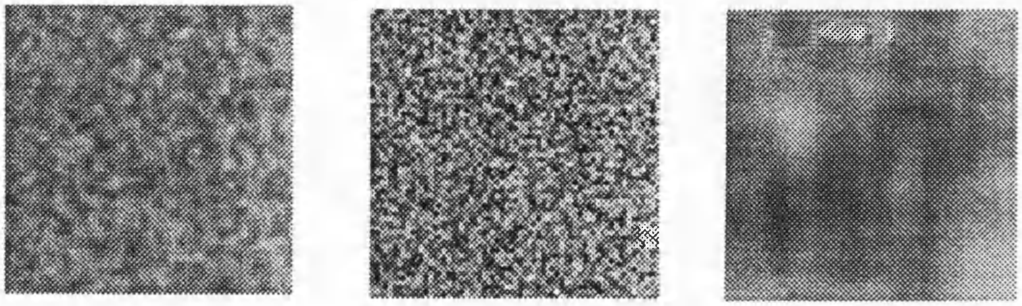
amplitude function



$n_c = 10, b = 5$



$n_c = 20, b = 3$



$n_c = 30, b = 1$

Figure 7.4. Some typical (separable) two-dimensional amplitude functions and corresponding values of $f(x, y)$ generated from equation (7.5), where $B = 140$, $N = 64$, and b and n_c vary as listed.

plots the shape of these example 2-dimensional amplitude functions and shows the corresponding 2-dimensional scenes generated using equation (7.5).

7.3 Auto-Correlation of the Functional Model

Presented here is the derivation of the auto-correlation of the functional image model developed in section 7.2. Describing the shape of the auto-correlation function near the origin (that is, near the peak in the correlation and thus near the correct solution), will give a direct indicator of the effects such parameters as correlation window size, scene spectral content (a parameter of the functional model), and added noise will have on the ability to accurately locate the auto-correlation peak. This then will be extended to cross-correlation for a better understanding of matching accuracy obtained by the stereo cross-correlation algorithm used to match regions of the fundus images.

7.3.1 The One-Dimensional Auto-Correlation Function

The following derivation of the one-dimensional auto-correlation of the function $f(x)$ defined by equation (7.1) is given not only because it is of value in the general signal processing field, but also as a tool for the derivation of the two-dimensional auto-correlation formulae.

The classical auto-correlation of a function $f(x)$ is given by

$$C(d) = \int_{-M/2}^{M/2} f(x)\, f(x+d)\, dx \ , \tag{7.10}$$

where M is the width of the correlation window. Substituting $f(x)$ defined by equation (7.1), this expands to

$$C(d) = \int_{-M/2}^{M/2} \sum_{n=0}^{n=n_c} a(n) \cos(pnx + \Phi) \sum_{m=0}^{m=n_c} a(m) \cos(pm(x+d) + \Phi)\, dx \tag{7.11}$$

where $p = 2\pi/N$ and Φ abbreviates the random phase component $\Phi(n)$. Decomposing the double summation inside the integral and rearranging terms, the expression becomes

$$C(d) = \int_{-M/2}^{M/2} \sum_{n=0}^{n=n_c} a^2(n) \cos(pnx + \Phi) \cos(pn(x+d) + \Phi)\, dx \tag{7.12}$$

$$+ \int_{-M/2}^{M/2} \sum_{m=0}^{m=n_c} a(m) \sum_{n=0}^{n=n_c} a(n) \cos(pmx + \Phi) \cos(pnx + \Phi)\, dx \ .$$

The first set of terms are those for which $n = m$, and hence the amplitude components are combined (the phase components in this case of auto-correlation are also identical). The expectation value for this set of terms will be of order $O(M\ n_c\ \langle a^2 \rangle)$; the cos()cos() term in the first integral will, on average, integrate to a value of approximately $M/2$ for small values of (npd), and the subsequent summation over n will yield $n_c \langle a^2 \rangle$. The second set of terms are those for which $n \neq m$. The expectation value for this set of terms will be of order $O(n_c \langle |a| \rangle)$. If the phase terms are assumed to be randomly distributed, then the cos()cos() terms will integrate to an average value of order $O(1)$. The summations over n and m will yield a quantity whose average expectation value will be of order $O(n_c \langle |a| \rangle)$. Thus, $C(d)$ can be approximated by neglecting the second set of terms in comparison of magnitude with the first. Hence,

$$C(d) \approx \int_{-M/2}^{M/2} \sum_{n=0}^{n=n_c} a^2(n) \cos(pnx + \Phi) \cos(np(x+d) + \Phi)\, dx \ . \tag{7.13}$$

$C(d)$ is simplified further by expanding $\cos(a+b)$ to $\cos(a)\cos(b) - \sin(a)\sin(b)$. By making the approximation that the resulting $\cos^2()$ terms yield an average expectation value of 1/2 over the symmetric integral, and the sin()cos() terms yield an average expectation value of 0 (the integral is odd!), the following expression for $C(d)$ is obtained; note that the approximation for the $\cos^2()$ term at $n = 0$ is not valid, and a value of unity must be retained. Hence,

$$C(d) \approx M\ a^2(0) + M/2 \sum_{n=1}^{n=n_c} a^2(n) \cos(npd)\, . \tag{7.14}$$

Note here that the auto-correlation function $C(d)$ is now independent of the phase component Φ in $f(x)$. In Fourier theory, the Autocorrelation Theorem states that: the auto-correlation function of a signal is the Fourier transform of its Power Spectrum [Bracewell65]. The unique feature of this theorem is that information about the phase of $F(s)$ is entirely missing from $|F(s)|^2$, the Power Spectrum. The auto-correlation function $C(d)$ in equation (7.14) correspondingly contains no information about the phase of the Fourier components (the frequency components) of $f(x)$. Although the phase components of $f(x)$ are assumed to be randomly distributed between 0 and 2π, any functional model for $\Phi(n)$ would suffice, since the resulting correlation function $C(d)$ is independent of phase.

The expression for $C(d)$ given by equation (7.14) is a good approximation for large values of M and for relatively busy scenes (i.e. large n_c). As M and/or n_c become small, the above formulae should be treated, not as an exact expression for the auto-correlation of any individual realisation of $f(x)$, in the limiting case where M approaches N and n_c approaches $N/2$, but as the expectation value of the auto-correlation of $f(x)$ for a large number of averaged trials.

The sampled function $f(x)$ given by equation (7.1) has thus far been treated as a discrete realisation of a continuous function composed of sinusoidal components characterised by $a(n)$. It is often convenient to treat the sampled function $f(x)$ as if it were continuous. This can be justified by considering that the scene structure is highly correlated from sample to sample (see [Sclove81] and [Pratt78, p132ff]). Image blurring and signal degradation along the signal processing chain which ultimately produces the (real) sampled $f(x)$ yields an inter-sample correlation which is generally high. For the functional image model given by equation (7.1), cosine components are sampled at discrete intervals which are inherently highly correlated from point to point, provided the discrete intervals are relatively small. In the present case, it is of interest to investigate the shape of the auto-correlation function for sub-pixel displacements, and some approximation must therefore be made for the data between individual samples. Sub-pixel correlation searches can only be made by interpolating values between discrete samples, and it is assumed that a meaningful interpolation can be made between discrete samples of $f(x)$. Hereafter, it is assumed that $f(x)$ can be approximated as a continuous function.

Under this assumption, the summation in equation (7.14) is replaced with an integral; swapping between summations and integrals will thus be carried out where it appears convenient to do so, and a similar argument of justification applies. Thus, $C(d)$ can be rewritten as

$$C(d) \approx M\ a^2(0) + M/2 \int_{n=1}^{n=n_c} a^2(n) \cos(npd)\, dx\ . \tag{7.15}$$

Provided the function $a^2(n)$ defining the amplitudes of the frequency components in $f(x)$ can be integrated, a simple analytic expression for $C(d)$ can be obtained. The specific scene structures defining $a(n)$ given by equations (7.2), (7.3), and (7.4) are examined in this context.

Linearly-decreasing Spectrum

For the specific scene structure defined by $a_1(n)$ in equation (7.2), $C(d)$ can be readily derived. Given the approximations used in obtaining a manageable form for $C(d)$ above, only the form of $C(d)$ in the region of small d near the correlation maximum is considered. Thus, $\cos(x)$ will be approximated by its second order series expansion $1 - x^2/2$. Substituting $a_1(n)$ into equation (7.15) and using this approximation, after integration the following approximate expression for $C(d)$ results;

$$C(d) = M\ B^2 + \frac{M\ b^2\ n_c}{6} - \frac{M\ b^2\ \pi^2\ d^2\ {n_c}^3}{30\ N^2}\ . \tag{7.16}$$

Flat Spectrum

The expression for $C(d)$ for the case of a flat image spectrum $a_2(n)$ defined by equation (7.3) can be derived in a straightforward manner. Substituting $a_2(n)$ into equation (7.15), the following is obtained after integration,

$$C(d) = M\ B^2 + \frac{M\ b^2}{2\,p\ d} \left[\sin(n_c pd) - \sin(pd)\right]\ . \tag{7.17}$$

Again, the primary interest is an expression for $C(d)$ which is valid for small values of d, although one must be aware that the approximation becomes poorer away from the correlation peak. The sin() terms are approximated by a power series expansion to obtain the following approximate expression (see [Forshaw90]);

$$C(d) = M\ B^2 + \frac{M\ b^2\ n_c}{2} - \frac{M\ b^2\ \pi^2\ d^2\ n_c{}^3}{3\ N^2}\ . \tag{7.18}$$

<u>1/*F* Spectrum</u>

The derivation of $C(d)$ for the $1/F$ spectrum defined by equation (7.4) can be obtained following a number of simple approximations. Reference [Forshaw90] gives a detailed derivation and the resulting expression is provided.

$$C(d) = M\ B^2 + \frac{M\ b^2}{2} - \frac{M\ b^2\ \pi^2\ d^2\ n_c}{N^2} \tag{7.19}$$

Clearly, the derived expressions of $C(d)$ for the specific scene structures have almost the same form, varying in dependence on the cut-off frequency n_c. This dependence will prove interesting when examining the theoretical versus experimental accuracy of the derived formulae in chapter 8.

7.3.2 The Two-Dimensional Auto-Correlation Function

The following derivation of the two-dimensional auto-correlation function $C(d, e)$ builds from the methods used to derive the one-dimensional formulae. The classical auto-correlation function $C(d, e)$ of $f(x, y)$ is

$$C(d, e) = \int_{-M/2}^{M/2} \int_{-M/2}^{M/2} f(x, y)\ f(x+d, y+e)\ dx\ dy\ . \tag{7.20}$$

Recall the two-dimensional image model $f(x, y)$ given by equation (7.5). Substituting into equation (7.20) above,

$$C(d, e) = \int_{-M/2}^{M/2} \int_{-M/2}^{M/2} \sum_{s=0}^{s=n_c} \sum_{r=-n_c}^{r=n_c} a(r, s) \cos(rpx + spy + \Phi) \tag{7.21}$$

$$\sum_{v=0}^{v=n_c} \sum_{u=-n_c}^{u=n_c} a(u, v) \cos(up(x+d) + vp(y+e) + \Phi)\ dx\ dy\ ,$$

where d and e are the correlation displacements in x and y. Here r, s, u, and v are the spatial frequencies of $f(x, y)$, $p = 2\pi/N$ as before, and Φ abbreviates the two-dimensional phase components $\Phi(r, s)$. The full derivation of the two-dimensional auto-correlation function to a reduced form is given in [Forshaw90], and the resulting formula is provided.

$$C(d, e) = M^2 a^2(0, 0) + \frac{M^2}{2} \int_{-M/2}^{M/2} \int_{-M/2}^{M/2} a^2(r, s) \cos(rpd + spe)\, dx\, dy \quad (7.22)$$

As for the one-dimensional case, the auto-correlation function is now independent of $\Phi(r, s)$, the phase components of $f(x, y)$. Notice also that the formula takes a similar form to the one-dimensional $C(d)$, modified only by the now two-dimensional correlation window M^2, and the extra dimension in the frequency components $a(r, s)$.

Provided the function $a^2(r, s)$ defining the amplitudes of the frequency components in $f(x, y)$ can be integrated, a simple analytical expression for $C(d, e)$ can be obtained for a given scene type. Therefore, the form of $C(d, e)$ for the specific two-dimensional scene spectra described in 7.2 will be examined, whose corresponding one-dimensional formulae were examined in 7.3.1.

Linearly Decreasing Spectrum

The expression for $C(d, e)$ for the specific scene structure defined by both the circularly symmetric $a_{1c}(r, s)$ and the separable $a_{1s}(r, s)$ linearly decreasing spectra given by equations (7.6) and (7.7) is derived. Given the approximations used in deriving a manageable form for $C(d, e)$, only the form of $C(d, e)$ in the region of small d and e near the correlation maximum is considered. Thus, $\cos(x)$ will be approximated by its second order series expansion $(1 - x^2/2)$. First, the separable form of $a_{1s}(r, s)$, in following on from the one-dimensional derivations, is examined. Although the separable forms of $a(r, s)$ will, in general, produce simpler expressions for $C(d, e)$, and although $C(d, e)$ derived from the separable form experimentally agrees quite well with $C(d, e)$ derived from the circularly symmetric form (at least for the linearly-decreasing spectra), it may be unwieldy if not impossible to make a separable approximation to an experimentally determined radially symmetric frequency distribution. It is for this reason the derivation for the circularly symmetric form, $a_{1c}(r, s)$, for the linearly decreasing spectrum is also included.

One can derive $C(d, e)$ for the separable $a(r, s)$ given by equation (7.7) in a straight-forward manner by substituting $a_{1s}(r, s)$ into equation (7.22) and using the series expansion of $\cos(x)$. After integration, the following approximate expression for $C(d, e)$ is obtained (see [Forshaw90]);

$$C(d, e) = M^2 B^2 + \frac{M^2 b^2 n_c^2}{9} - \frac{M^2 b^2 \pi^2 n_c^4 (d^2 + e^2)}{45 N^2} . \tag{7.23}$$

To derive $C(d, e)$ for the circularly symmetric $a_{1c}(r, s)$ given by equation (7.6), $a_{1c}(r, s)$ is redefined as follows

$$a'_{1c}(q) = 1 - \frac{|q|}{n_c},\ q^2 = r^2 + s^2,\ r = p\cos(\Theta),\ s = p\sin(\Theta) . \tag{7.24}$$

Further, $C(d, e)$ given by equation (7.22) is re-formulated as

$$C(d', 0) = M^2 a'^{2}_{1c}(0) + \frac{M^2}{2} \int_{-M/2}^{M/2} \int_{-M/2}^{M/2} a'^{2}_{1c}(q) \cos(r'pd')\, dr'\, ds' , \tag{7.25}$$

where d' is the radial displacement $\sqrt{d^2 + e^2}$, and r' and s' are the newly oriented axes of the frequency components along d'. The angular displacement from the new axis r' is expressed as Θ, and this converts fully to (q, Θ). Dropping the prime for notational simplicity,

$$C(d) = M^2 a^{2}_{1c}(0) + \frac{M^2}{2} \int_{-M/2}^{M/2} \int_{-M/2}^{M/2} a^{2}_{1c}(q) \cos(qpd\ \cos(\Theta))\, q\ dq\ d\Theta . \tag{7.26}$$

Note that the limits of integration have changed to reflect the new coordinate axes. Following the series approximation of $\cos(x)$, subsequent integration yields the desired expression,

$$C(d, e) = M^2 B^2 + \frac{M^2 \pi b^2 n_c^2}{24} - \frac{M^2 b^2 \pi^3 n_c^4 (d^2 + e^2)}{120 N^2} , \tag{7.27}$$

where the original coordinates in d and e have been returned. Interestingly, this expression for $C(d, e)$ differs approximately by a factor of π from the separable equivalent. Hence, for small d and/or e, we might expect the circularly symmetric formula to be well approximated by the separable axes formula.

Flat Spectrum

The expression for $C(d, e)$ for the case of a flat image spectrum defined by equation (7.8) can be derived in a straightforward manner. Substituting $a_2(r, s)$ into equation (7.22) we obtain, after integration,

$$C(d, e) = M^2 B^2 + \frac{M^2 b^2}{p^2 d^2} \sin(n_c pd) \sin(n_c pe) . \tag{7.28}$$

Again, one is primarily interested in an expression for $C(d, e)$ which is valid for small values of d and e, aware that the approximation becomes poorer away from the correlation peak. Thus, as with the one-dimensional case, expanding the $\sin(x)$ terms yields the following expression,

$$C(d, e) \approx M^2 (B^2 + b^2 n_c^2) - \frac{2 M^2 b^2 \pi^2 n_c^4 (d^2 + e^2)}{3 N^2} . \tag{7.29}$$

$1/F$ Spectrum

To derive $C(d, e)$ for the $1/F$ spectrum defined by equation (7.9), a number of approximations for small d and e near the correlation peak must be applied (this follows on directly from the one-dimensional derivation). The following expression is obtained (see [Forshaw90]):

$$C(d, e) = M^2 (B^2 + b^2) - \frac{2 M^2 b^2 \pi^2 n_c (d^2 + e^2)}{N^2} . \tag{7.30}$$

7.3.3 A Review of Derived Correlation Formulae

A theoretical model for the two-dimensional auto-correlation of the synthetic image function $f(x, y)$ has been derived. The theoretical formulae for $C(d, e)$ for the specific scene structures provide a good approximation for small values of d and e near the correlation peak. Knowledge of the underlying shape of

$C(d, e)$ will allow predictions on the ability of a given correlation algorithm to accurately locate the correlation maximum, and thus the correct match.

The following tables give a summary of the one- and two-dimensional correlation formulae derived.

One-Dimensional Autocorrelation Formulae

amplitude $a(n)$	auto-correlation $C(d)$
$b(1 - \frac{n}{n_c})$	$M\,B^2 + \frac{M\,b^2\,n_c}{6} - \frac{M\,b^2\,\pi^2\,d^2\,{n_c}^3}{30\,N^2}$
b	$M\,B^2 + \frac{M\,b^2\,n_c}{2} - \frac{M\,b^2\,\pi^2\,d^2\,{n_c}^3}{3\,N^2}$
$\frac{b}{n}$	$M\,B^2 + \frac{M\,b^2}{2} - \frac{M\,b^2\,\pi^2\,d^2\,n_c}{N^2}$

Figure 7.5. Table of derived one-dimensional auto-correlation functions.

Two-Dimensional Autocorrelation Formulae

amplitude $a(r, s)$	auto-correlation $C(d, e)$
$b(1 - \frac{r}{n_c})(1 - \frac{s}{n_c})$	$M^2\,B^2 + \frac{M^2\,b^2\,n_c^2}{9} - \frac{M^2\,b^2\,\pi^2\,n_c^4\,(d^2 + e^2)}{45\,N^2}$
$b(1 - \frac{\sqrt{r^2 + s^2}}{n_c})$	$M^2\,B^2 + \frac{M^2\,\pi\,b^2\,n_c^2}{24} - \frac{M^2\,b^2\,\pi^3\,n_c^4\,(d^2 + e^2)}{120\,N^2}$
b	$M^2\,(B^2 + b^2 n_c^2) - \frac{2\,M^2\,b^2\,\pi^2\,n_c^4\,(d^2 + e^2)}{3\,N^2}$
$\frac{b}{rs}$	$M^2\,(B^2 + b^2) - \frac{2\,M^2\,b^2\,\pi^2\,n_c\,(d^2 + e^2)}{N^2}$

Figure 7.6. Table of derived two-dimensional auto-correlation functions.

7.4 Mean Square Difference Correlation

In chapter 5 a stereo cross-correlation algorithm, based on the sum of squared differences or mean square difference (MSD) correlation measure, is selected for use on the fundus image data. To better understand the behaviour of the stereo algorithm, the classical auto-correlation functions developed in section 7.3 are extended to MSD correlation. The goal of this section is to derive expressions for the sum of squared difference auto-correlation of the one-dimensional and two-dimensional image models $f(x)$ and $f(x, y)$ developed in section 7.2.

Note that whereas the classical auto-correlation formulae $C(d)$ and $C(d, e)$ produce a correlation maximum at d and e equal zero, the MSD auto-correlation formulae will produce a correlation minimum. The MSD correlation is often termed the minimum-square error or least-squares algorithm. Often, scaled or normalised versions of these formulae are used to produce more manageable correlation values (see [Rosenfeld82], [Pratt78], [Griffin90]). The general form for MSD auto-correlation is

$$MSD(d) = \sum_{-M/2}^{M/2} \left[f(x) - f(x+d) \right]^2 \tag{7.31}$$

for the one-dimensional case and

$$MSD(d, e) = \sum_{-M/2}^{M/2} \sum_{-M/2}^{M/2} \left[f(x, y) - f(x+d, y+e) \right]^2 \tag{7.32}$$

is the two-dimensional form. The advantage to using MSD correlation over classical correlation is that it shows less sensitivity to the large DC component (the average scene value), and this will become apparent in the following derivations.

7.4.1 One-Dimensional MSD Correlation

The one-dimensional sum of squared difference auto-correlation formula given above can be expanded to

$$MSD(d) = \sum_{-M/2}^{M/2} \left[f(x)^2 + f(x+d)^2 - 2f(x)f(x+d) \right] . \tag{7.33}$$

The first two terms of equation (7.33) can be considered the average value of $f(x)$ over the window M, which in many cases can be approximated by the DC term of the image. Here such an approximation is reasonable, given the definition of $f(x)$ given by equation (7.1). The third term of equation (7.33) can be recognised as the classical auto-correlation of $f(x)$ given by $C(d)$ in equation (7.15), multiplied by the constant –2. Thus, $MSD(d)$ will be approximated as a function of $C(d)$, and written as

$$MSD(d) \approx 2C(0) - 2C(d) . \tag{7.34}$$

Using this approximation, the MSD auto-correlation formulae for the specific one-dimensional scene spectra defined in section 7.2 are obtained directly from their corresponding expressions for $C(d)$.

Linearly-Decreasing Spectrum

$$MSD(d) \approx \frac{M\ \pi^2\, b^2\, d^2\, n_c^3}{15\, N^2} \tag{7.35}$$

Flat Spectrum

$$MSD(d) \approx \frac{2\, M\ \pi^2\, b^2\, d^2\, n_c^3}{3\, N^2} \tag{7.36}$$

1/*F* Spectrum

$$MSD(d) \approx \frac{2\, M\ \pi^2\, b^2\, d^2\, n_c}{N^2} \tag{7.37}$$

7.4.2 Two-Dimensional MSD Correlation

As with the one-dimensional sum of squared difference auto-correlation formula, the two-dimensional formula can be expanded to

$$MSD(d, e) = \sum_{-M/2}^{M/2} \sum_{-M/2}^{M/2} f(x, y)^2 + f(x+d, y+e)^2 \tag{7.38}$$

$$- 2f(x, y) f(x+d, y+e) .$$

The first two terms are nearly constant and can be approximated by the DC term for a scene defined by $f(x, y)$ given by equation (7.5). Hence, $MSD(d, e)$ will be written as a function of $C(d, e)$ as

$$MSD(d, e) \approx 2C(0, 0) - 2C(d, e) . \tag{7.39}$$

Using this approximation, the MSD auto-correlation formulae for the specific two-dimensional scene spectra defined in section 7.2 are obtained directly from their corresponding expressions for $C(d, e)$.

Linearly-Decreasing Spectrum

(separable)

$$MSD(d, e) \approx \frac{2\, M^2\, \pi^2\, b^2\, n_c^4\, (d^2 + e^2)}{45\, N^2} \tag{7.40}$$

and

(symmetric)

$$MSD(d, e) \approx \frac{M^2\, \pi^3\, b^2\, n_c^4\, (d^2 + e^2)}{60\, N^2} \tag{7.41}$$

Flat Spectrum

$$MSD(d, e) \approx \frac{4\, M^2\, \pi^2\, b^2\, n_c^4\, (d^2 + e^2)}{3\, N^2} \tag{7.42}$$

1/*F* Spectrum

$$MSD(d, e) \approx \frac{4\, M^2\, \pi^2\, b^2\, n_c\, (d^2 + e^2)}{N^2} \tag{7.43}$$

By extending the MSD correlation formulae to the specific two-dimensional scene spectra above, a description of the correlation function which is equivalent to that used in the least-squares stereo cross-correlation algorithm is obtained (for linear displacements in x and y). This is important for several reasons; (1) experiments involving the theoretical model give a direct insight into the performance of the stereo algorithm, (2) the theoretical model can be easily extended to other scene types, (3) it can be applied to other forms of correlation formulae (not just stereo cross-correlation - for example, image alignment), and (4) most importantly, the model can be extended to incorporate the presence of additive noise in the scene, which is all too common in real world applications.

7.5 Cross-Correlation

In this section the extension of the one- and two-dimensional auto-correlation formulae to cross-correlation is discussed. Although the formulae derived in sections 7.3 and 7.4 are strictly only meaningful for auto-correlation, a number of simplifying assumptions allows one to make use of the formulae for the analysis of cross-correlation. The extension of the formulae to true cross-correlation of two image regions is extremely complex. In this section it is shown that the mathematical process underlying cross-correlation (particularly of stereo images) is related to the concepts of auto-correlation, and that the auto-correlation formulae will be useful for predicting the performance of image cross-correlation.

It is important to recall from section 7.2 that it is only of interest to assess the ability to accurately correlate a small portion of the image in relatively stable parts of the scene; that is, there is an assumption that the region to be matched is uniform in *texture*. In stereo imagery, there may be portions of one image that are obscured or occluded in the other, and indeed there may be a large variation in scene texture across the image. However, it is assumed that the region to be matched does not encompass portions which are obscured, and does not encompass the boundary between two vastly different scene textures.

7.5.1 Types of Cross-Correlation

There is a wide range of image processing tasks where cross-correlation might be required. The following list is an attempt to give some idea of the spectrum of applications for cross-correlation. It is not intended to be a definitive classification, but rather to show the range of complexity in correlating images, from simple auto-correlation to multi-dimensional cross-correlation.

- *object identification or location*: correlation of a noise-free template with a noise-free or noisy scene (x, y).

Examples: Robotics, assembly line inspection.

- *two-dimensional image alignment*: correlation of a noise-free or noisy template with a noise-free or noisy scene (x, y, Θ).

Examples: Serial-section alignment, voxel processing.

- *multi-dimensional image alignment*: correlation of a noise-free or noisy template with a noise-free or noisy scene (x, y, Θ, scale, etc.).

Examples: Stereo matching (fundus image data), elastic matching.

- *partial or temporal matching*: correlation of a noise-free or noisy complete template with a noise-free or noisy partial or time varying scene (x, y, Θ, scale, etc.).

Examples: Robotics, bin picking.

- *multi-dimensional best-match cross-correlation* where the template and scene are not identical (x, y, Θ, scale, etc.).

Examples: Close-range (large-baseline) photogrammetry, karyotyping.

The types of cross-correlation listed above must be examined individually to determine the extent to which auto-correlation underlies the matching process. Clearly, *object identification or location* can be considered as auto-correlation when some portion of the scene is identical to the template. With or without additive noise, this is simply an exhaustive search process, and the auto-correlation formulae can be used to help make predictions on the expected accuracy.

Another common cross-correlation application is that of *image alignment.* Here, one can consider the scene to be a near-copy of the template, where both the template and scene may be corrupted by additive noise. To the extent that the scene and template are not identical, this is strictly cross-correlation. However, in many applications of image alignment (e.g. serial section alignment), the scene and template differ only slightly. This difference could be treated as simply another form of noise, and therefore the auto-correlation formulae may still prove useful for making some estimate of the expected error in applications of image alignment.

As discussed above, image alignment often only takes place over two or three dimensions (i.e. x, y, Θ). When the matching process is extended to higher dimensions, but the scene and template (when transformed) are still near-copies, the assertions made above can be applied to *multi-dimensional image alignment.* Although the correlation formulae derived in this chapter only encompass the two dimensions x and y, it is often the case that measurements of accuracy are restricted to a subset of a multi-parameter space, perhaps for simplicity or, as is the case in stereo matching, because the relative weight of one parameter dominates over several others. For these reasons, the auto-correlation formulae may prove useful in such applications as stereo matching, where the estimated parameters Δx and Δy are of prime importance.

The final two classifications of cross-correlation, namely *partial or temporal matching* and *multi-dimensional best match cross-correlation* are clearly beyond the scope of the auto-correlation formulae. Partial matching of obscured or occluded scenes with a whole template is difficult to analyse given limited information about which subset of data samples are available in the scene, and the cause or structure of the occlusion. When the template and scene are not identical, and the best global match of the template across the scene is required, then the auto-correlation formulae are not applicable. Further development of the formulae is required for such applications of cross-correlation.

7.5.2 Solving Cross-Correlation

The types of cross-correlation discussed above can be further analysed by looking at the approaches to calculating a cross-correlation minimum for a given application. Recall from chapter 5, in the context of stereo matching, that there are a number of ways in which to make some measure of the goodness of match between that template and the scene. For *object identification or location* and indeed for *two-dimensional image alignment* the complexity of the process can range from simple image differencing, to pre-filtering followed by a sum of squared difference measure. In the case of *multi-dimensional image alignment* a number of suitable methods apply. For the application of stereo matching (over several parameters) matrix methods can prove useful, where the template is transformed over the multi-parameter space, iterating in the direction of steepest gradient, until a minimum is found. When the application is considered *best-match cross-correlation*, there may be multiple minima. The choice of a starting location in the scene can be critical for convergence to a valid minimum.

It was natural to extend the auto-correlation formulae to sum of squared difference correlation in section 7.4. The primary aim of this chapter is to investigate the means by which to make some estimate of matching accuracy in the application of stereo matching. The addition of noise to the derived correlation formulae will be investigated in the following section, leading to a theoretical relationship between cross-correlation error and the various parameters of the functional image model.

7.6 Noise and Correlation

In this section the functional image model is extended to incorporate additive Gaussian noise to investigate the effects of such noise on the theoretical performance of cross-correlation formulae. An analytical expression for predicting alignment accuracy is described. This work builds closely on prior work presented in [McGillem76], [Mostafavi78], [Ryan80], and [Bogler86]. It is assumed here that the cross-correlator takes the form of the sum of squared difference or mean square difference (MSD) operator. Furthermore, the presence of noise in the image model is simplified to include only zero-mean, random Gaussian additive noise. To simplify the notation, all functional relationships are expressed in one dimension. Extension to two dimensions is direct and does not alter the analysis.

7.6.1 Terminology

The variables and expressions used in this section are listed below with a brief description. Their definition is given in the text where appropriate.

f	the underlying continuous scene
$f_i(x)$	the ith observed (discrete) image of f $(i = 1,2)$
$n_i(x)$	observed Gaussian distributed additive noise in image i
$C(d)$	underlying scene auto-correlation function
$MSD(d)$	observed MSD cross-correlation of f_1 and f_2 at position d
$\eta(d)$	noise component of $MSD(d)$ at position d
$< \; >$	expected value of the quantity between the brackets
σ_n^2	variance of the total noise
$\hat{d}$	estimated positional shift
$\bar{d}$	true positional shift
σ_d^2	variance of the error in location, $<(\hat{d} - \bar{d})^2>$.

7.6.2 Additive Noise

There are many sources of additive noise that can affect alignment accuracy in cross-correlation. The noise may be due to either film-grain noise, shot noise (photon noise) in a photo-electric sensor, or a direct consequence of the process of image acquisition, sensing or recording. A further possible source of noise is that caused by the correlation process itself, due to the interpolation necessary to produce intensity values at other than at whole numbered pixel coordinates. The process of acquiring a discrete image function necessarily involves the averaging of small regions of image density data. Even when an image region is not corrupted by noise, obtaining two independent digital representations will not usually result in identical pixel arrays. Fundus image stereo cross-correlation, where each data set is digitised from two film transparencies, where the video camera system may add electrical noise, and where the correlator will itself cause "interpolation" noise, is a good example of the presence of systematic additive noise. The combined effects of all sources of noise can be adequately modelled by zero-mean, Gaussian distributed noise. The reason for the use of the Gaussian distribution is that it leads to a closed form analytical solution in estimating error variance.

Consider the 1-dimensional observed image function $f_i(x)$ as the sum of the noise-free signal $f(x)$ and zero mean, random Gaussian distributed noise $n_i(x)$ with variance σ_n^2.

$$f_i(x) = f(x) + n_i(x)\,, \quad i = 1,2\,. \tag{7.44}$$

Due to the additive noise, the MSD cross-correlation of $f_1(x)$ with $f_2(x)$ will introduce errors in estimating the location of the true minimum. Note that the addition of two independent noise functions to $f_1(x)$ and $f_2(x)$ respectively (e.g. the image and template are both corrupted by independent additive noise) can be modelled by one cumulative noise function added to $f_2(x)$ alone (i.e. $n = n_1 + n_2$). The expression for the MSD cross-correlation of $f_1(x)$ with $f_2(x)$ over M samples is given by

$$MSD(d) = \int_{-M/2}^{M/2} (f_1(x) - f_2(x+d))^2 \, dx\,. \tag{7.45}$$

The expression $MSD(d)$ can be decomposed into the sum of two functions, written as

$$MSD(d) = \langle MSD(d) \rangle + \eta(d)\,, \tag{7.46}$$

where the first function, $\langle MSD(d) \rangle$, denotes the average or expected value (the value of $MSD(d)$ when the noise is averaged out), and the second function, $\eta(d)$, represents the noise in the system which acts to perturb the true minimum [Bogler86]. Consider $\langle MSD(d)\rangle$ to represent the unbiased version of the true correlation function $MSD(d)$ that yields a minimum at the correct value of d. The noise component in $MSD(d)$ is signal dependent, and may perturb the correlation function sufficiently to give rise to a false minimum for a given value of d. The question arises as to the accuracy to which the correct value of d can be estimated in the presense of noise.

7.6.3 The Cramer-Rao Bound

The accuracy of correlation processors has been extensively studied in the context of radar and sonar target range determination, and it has been shown by the method of Cramer-Rao [Whalen71] that there is a lower bound on the accuracy of *any* correlation process in the presence of additive noise. In the context of image cross-correlation, the Cramer-Rao bound is a lower bound on the variance of any unbiased estimate of the displacement d (see, for example, [Ryan81]). The Cramer-Rao bound is given by

$$\sigma_{\hat{d}}^2 = \langle(\hat{d} - \bar{d})^2\rangle \geq \frac{\frac{N_o}{2}}{(2\pi)^2 \int\limits_{-\infty}^{\infty} u^2 \, |F(u)|^2 \, du} , \tag{7.47}$$

where $\frac{N_o}{2}$ represents the spectral density of the additive noise, and $|F(u)|^2$ is the Fourier power spectrum of $f(x)$. The Cramer-Rao inequality specifies a lower bound on the displacement error variance, but it does not guarantee that this bound can be reached or even closely approximated. Equation (7.47) implies that the accuracy of displacement estimation (regardless of the correlator applied) varies inversely with signal bandwidth (the spectral density of $f(x)$).

The spectral density of the noise is equivalent to the variance of the noise when the sampled noise $n_i(x)$ is bandlimited to one half-cycle per pixel. Furthermore, from the Fourier moment theorem ([Bracewell65, p.172]) one can write

$$(2\pi)^2 \int\limits_{-\infty}^{\infty} u^2 \, |F(u)|^2 \, du = - \frac{\partial^2}{\partial d^2} C(d) \, |_{d=0} \, . \tag{7.48}$$

Thus, a relationship is found between the Cramer-Rao bound and the auto-correlation function. Substituting equation (7.48) into (7.47) above, one can write

$$\sigma_d^2 \approx \frac{\sigma_n^2}{C''(d)\,|_{d=0}} \, . \tag{7.49}$$

Although the correlation variance estimate in equation (7.49) is not a bound, it can be interpreted as a measure of the shape or curvature at the peak of the autocorrelation function at the origin, and it relates the shape of the underlying auto-correlation function to the theoretical cross-correlation accuracy.

7.6.4 Correlation Error Variance

Alternatively, a measure of the positional uncertainty can be found directly by calculating the variance of the error in the estimate of d (see [McGillem76] and [Bogler86]). The one-dimensional form is developed here, but is easily extended to the two-dimensional equivalent. The development of the expression for the variance of registration error begins (following [Bogler86]) with a second-order Taylor series expansion about the match position:

$$\begin{aligned} MSD(d) &= \langle MSD(\bar{d})\rangle + (d-\bar{d})\langle MSD'(\bar{d})\rangle \\ &\quad + \frac{(d-\bar{d})^2}{2}\langle MSD''(\bar{d})\rangle + \eta(d)\, . \end{aligned} \tag{7.50}$$

Note that since $\langle MSD(d)\rangle$ is maximised at $d = \bar{d}$, the necessary condition of a maximum is that $\langle MSD'(\bar{d})\rangle = 0$. Further, the maximum is found by differentiating with respect to d, and equating the resultant expression to zero. One can now find the variance of registration error by squaring and taking expectations, to arrive at

$$\sigma_d^2 = \langle(\hat{d} - \bar{d})^2\rangle = \frac{\mathrm{var}\ \eta'(\bar{d})}{[\langle MSD''(\bar{d})\rangle]^2} \ , \tag{7.51}$$

where

$$\eta'(\bar{d}) = \frac{\partial}{\partial d}\eta(d)\, ,$$

var η' is the variance on η', and

$$MSD''(d) = \frac{\partial^2}{\partial d^2} MSD(d) .$$

Bogler has shown that expressions for both the numerator and denominator can be derived [Bogler86]. The resulting (approximate) expression for the variance of the match location in d (in the notation presented here) is given by

$$\sigma_d^2 = \langle (\hat{d} - \bar{d})^2 \rangle = \frac{\sigma_n^2}{C''(d)\,|_{d=0}} . \tag{7.52}$$

Not surprisingly, equation (7.52) is identical to the approximation of the Cramer-Rao bound given by equation (7.49). When the signal-to-noise ratio is high (i.e. the spectral density of $f(x)$ is large compared to the spectral density of the noise), one might expect the correlation variance estimate given by equation (7.52) to converge to the Cramer-Rao bound. For reference, equation (7.52) will be termed the Cramer-Rao bound estimate (CRBE).

7.6.5 Using the Estimated Variance

In this section the utility of the displacement variance measure is demonstrated by examining the Cramer-Rao bound estimate (CRBE) in the context of the auto-correlation formulae derived in section 7.3 and 7.4. As discussed in section 7.5, the auto-correlation formulae can be used in the context of cross-correlation for a number of application areas. With the development of the CRBE, the importance of knowing the underlying auto-correlation function becomes apparent. Using the functional image model developed in section 7.2, a direct relationship between the parameters which describe the image spectral content, the image auto-correlation function, and the CRBE can now be given. When the image contains additive Gaussian noise, the CRBE provides an analytical expression for predicting correlation accuracy.

To demonstrate the use of the CRBE in one dimension, recall the image model given by equation (7.1) and a suitable amplitude function given by, for example, the linearly decreasing spectrum in equation (7.2). The auto-correlation function of this image is given explicitly by equation (7.16) as

$$C(d) = M\ B^2 + \frac{M\ b^2\ n_c}{6} - \frac{M\ b^2\ \pi^2\ d^2\ n_c{}^3}{30\ N^2}\ .$$

Taking the second derivative of $C(d)$, the desired form is obtained;

$$C''(d) = \frac{M\ b^2\ \pi^2\ n_c{}^3}{15\ N^2}\ . \tag{7.53}$$

Substituting this expression into the CRBE, the resulting estimate on the variance of the correlation error is

$$\sigma_d^2 = \frac{15\ N^2\ \sigma_n^2}{M\ b^2\ \pi^2\ n_c{}^3}\ . \tag{7.54}$$

Thus, in the one dimensional case, if the amplitude of the frequency components of $f(x)$ can be modelled with a suitable function, following the method described in section 7.3 leads to an analytical expression for the auto-correlation function $C(d)$. The second derivative of the auto-correlation function can be used in the CRBE to provide a measure of correlation error in the presence of additive Gaussian noise.

To demonstrate the use of the CRBE in two dimensions, the one-dimensional CRBE measure given by equation (7.49) must be extended to two dimensions. The extension of equation (7.49) is based on the assumption that the correlation error is independent in x and y; that is, the error in estimating d can be determined independently of the error in estimating e. In two dimensions the Cramer-Rao bound is given by

$$\sigma_d^2 \geq \frac{\frac{N_o}{2}}{(2\pi)^2 \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} u^2\ v^2\ |F(u,v)|^2\ du\ dv}\ . \tag{7.55}$$

Note that the bound is only expressed for the error in d. Following the approximations for the one-dimensional case, where the Fourier power spectrum is separable in u and v, the second (partial) derivative of the auto-correlation function (with respect to d) can be substituted in the denominator. Thus,

$$\sigma_d^2 \approx \frac{\sigma_n^2}{C''(d,e)|_{d=0,\, e=0}} . \tag{7.56}$$

A similar expression can be derived for the error in the y direction by simple transposition of coordinates.

Recall the two-dimensional image model $f(x,y)$ given by equation (7.5). With a suitable amplitude function given by, for example, the $1/f$ spectrum defined by equation (7.9), an expression for the auto-correlation function can be derived. Equation (7.30) gives the two-dimensional auto-correlation function for the model $1/f$ spectrum as

$$C(d,e) = M^2\,(B^2 + b^2) - \frac{2\,M^2\,b^2\,\pi^2\,n_c\,(d^2 + e^2)}{N^2} .$$

Taking the second derivative of $C(d,e)$, the desired form is obtained;

$$C''(d,e) = \frac{\partial^2}{\partial d^2} C(d,e) = -\frac{4\,M^2\,b^2\,\pi^2\,n_c}{N^2} . \tag{7.57}$$

Substituting this expression into the two-dimensional CRBE, the resulting estimate on the variance of the correlation error is given as

$$\sigma_d^2 \approx \frac{N^2\,\sigma_n^2}{4\,M^2\,b^2\,\pi^2\,n_c} . \tag{7.58}$$

Thus, for two-dimensional cross-correlation, if the amplitude of the frequency components of $f(x,y)$ can be modelled with a suitable function, following the method described in section 7.3 leads to an analytical expression for the two-dimensional auto-correlation function $C(d,e)$. The second derivative of the auto-correlation function can be used in the CRBE to provide a measure of cross-correlation error in the presence of additive Gaussian noise. In chapter 8, the verification of the CRBE as a useful error measure will be investigated in the context of stereo cross-correlation.

7.7 Summary

The use of mathematical modelling to investigate the structure of auto- and cross-correlation functions has been presented. The local structure of the fundus image data has been examined using well known image descriptors, and this led to a functional description of image formation based on modelling the spectral components. An image model is developed which characterises the image structure of not only the fundus data, but of discrete image data in general. This model is based on summing sinusoidal frequency components with varying amplitude and phase. Explicit formulae for the auto-correlation of the image model, given some example amplitude functions, are derived in both one and two dimensions. These formulae are then extended to mean square difference auto-correlation. Subsequently, it is shown that the derived auto-correlation formulae are useful for describing the fundamental process underlying cross-correlation, for a large class of applications. Finally, a description of the Cramer-Rao bound provides a means by which to assess the potential accuracy of cross-correlation when the scene contains additive Gaussian noise. The utility of this measure is demonstrated on both one- and two-dimensional image models developed in this chapter.

The formulae derived in this chapter are not intended to *accurately* define the performance of cross-correlation algorithms. The image model is too simplistic to provide accurate measures about scene parameters versus correlation parameters. Rather, it is hoped that the formulae will provide some insight into the *expected* performance of a specific correlation algorithm, given information about the spectral properties of the scene. The ability to make *simple* predictions about the performance of standard cross-correlation algorithms is essential for investigating alternative methods for correlating complex scenes. With the tools to make such simple predictions, one might envisage a process by which we adaptively assess the expected performance of a given algorithm over the variety of textures in a given scene, altering correlation parameters to compensate for expected error rates. Although no quantitative evidence is provided to indicate that the auto-correlation formulae will be useful for analysing cross-correlation, in the next chapter some experimental results are given which suggest that we can make valid predictions about the expected accuracy in applications of cross-correlation using the tools developed in this chapter.

Chapter 8
Model Verification and Analysis

8.1 Introduction

In this chapter the image model, correlation formulae, and error prediction formulae developed in chapter 7 are tested and applied. There are many experimental methods which could be used to verify the utility of the image model together with the corresponding error prediction formulae. In the present context, it is sufficient to demonstrate that the formulae are of use with the fundus image data, with the implication that, after further development, there is a broader range of application. Time did not permit a full study of the utility of the formulae, although the experimental results presented in this chapter indicate that the image model and the error prediction formulae are useful in their current form.

The verification of the utility of the image model and the error prediction formulae are to be taken in the context of the cross-correlation algorithm employed for the fundus image data; the least-squares matching algorithm. The question is, whether valid predictions about the accuracy of cross-correlation can be assessed, in the absence of ground truth information. In this chapter the image model developed in chapter 7 is shown to be applicable to both Data Set A (the plaster model) and the fundus image data. Further, with an appropriate image model, the corresponding error prediction formulae are shown to be valid. And finally, the least-squares error estimator (discussed in chapter 6, section 6.5), which is a product of the least-squares technique, is also assessed as a match error measure.

The ways in which the formulae were tested and applied were:

1) synthetic data, generated using the image model, were used to verify matching accuracy in the absence and presence of additive noise;
2) plaster model data (Data Set A), for which ground truth was known, were used to assess the utility of the image model and the corresponding error

prediction formula for a given class of image;

3) fundus data, from Data Sets B and C, under the assumption that the error prediction formulae are valid, were used to test the applicability of the image model, the validity of the error prediction formulae, and to assess the least-squares error estimator;

By applying Gruen's algorithm to synthetic data, we could test for any inherent error in the algorithm in the absence of noise (e.g. due to interpolation error) as well as the utility of the Cramer-Rao bound estimate (CRBE) error measure in the presence of noise. Further, we could also verify the least-squares error estimator, demonstrated in chapter 6 to be an *a posteriori* measure of match error. Experiments involving synthetic data are presented in sections 8.2 and 8.3.

The plaster model data, Data Set A, are a set of data for which ground truth is known, and thus, as shown in chapter 6, the real matching error could be assessed. In this instance the full range of techniques developed in chapter 7 were applied: estimating the shape of the image spectrum, employing the image model by using a suitable amplitude function, defining the specific correlation formulae for the model image, and finally using the CRBE to make estimates of the expected error. This is presented in section 8.4.

Use of the error prediction formulae with the real fundus image data was complicated by the lack of ground truth from which to verify predictions with experiments. However, if the formulae are assumed to be useful (as is demonstrated in sections 8.2, 8.3, and 8.4), both the image model and the error prediction formulae could be used to provide estimates of matching accuracy. The suitability of the image model to represent a range of regions in a typical fundus image was also assessed. Experiments involving regions of a real fundus image and the application of the image model and the CRBE error prediction formulae are presented in section 8.5.

Other real-world data (e.g. landsat images) might be used to demonstrate the utility of the formulae for other classes of image or other application areas (e.g. image alignment), but time did not permit this. Although it would be of general interest to extend the use of the formulae to other image application areas, time did not permit this. Further work is required to develop the image model beyond the scope of the fundus image application, and to verify the utility of the CRBE for other application areas and image domains. Initial application of the formulae to the fundus image data indicates that valid predictions about matching accuracy can be made, but with certain limitations and restrictions. The limitations of the formulae are discussed in section 8.6.

8.2 Application to Synthetic Data in the Absence of Noise

In this section the application of the stereo matching algorithm, Gruen's algorithm, to synthetic data in the absence of noise is assessed. This will be important for several reasons. Firstly, both real and synthetic data are discrete, and matching to sub-pixel accuracy requires interpolation of whole valued pixels. It was therefore important to determine matching error due to the necessity for interpolation. In order to test this a model image was transformed (geometrically and radiometrically) by a known amount to form a second (sampled) image. By using Gruen's algorithm to match selected points - in this context Gruen's algorithm is used for template matching - and comparing the expected result with that obtained experimentally, the matching error due to interpolation could be assessed. As Gruen's algorithm (as implemented here) is intended to cope with a wide range of geometric and radiometric differences between two images, it was interesting to measure the matching error in the absence of any other source of additive noise.

Another reason for applying Gruen's algorithm to synthetic data in the absence of additive noise (although sampling can be thought of as a type of noise) was to assess the least-squares error estimator (LSEE) presented in chapter 6, which gives an *a posteriori* estimate of the accuracy of the match based on the largest eigenvalue of the cofactor matrix for the parameters Δx and Δy. The LSEE of each match point is shown experimentally for Data Set A to be a reasonable indicator of match precision (see chapter 6, section 6.5). It was therefore important to attempt to quantify the utility of the LSEE as an error analysis tool.

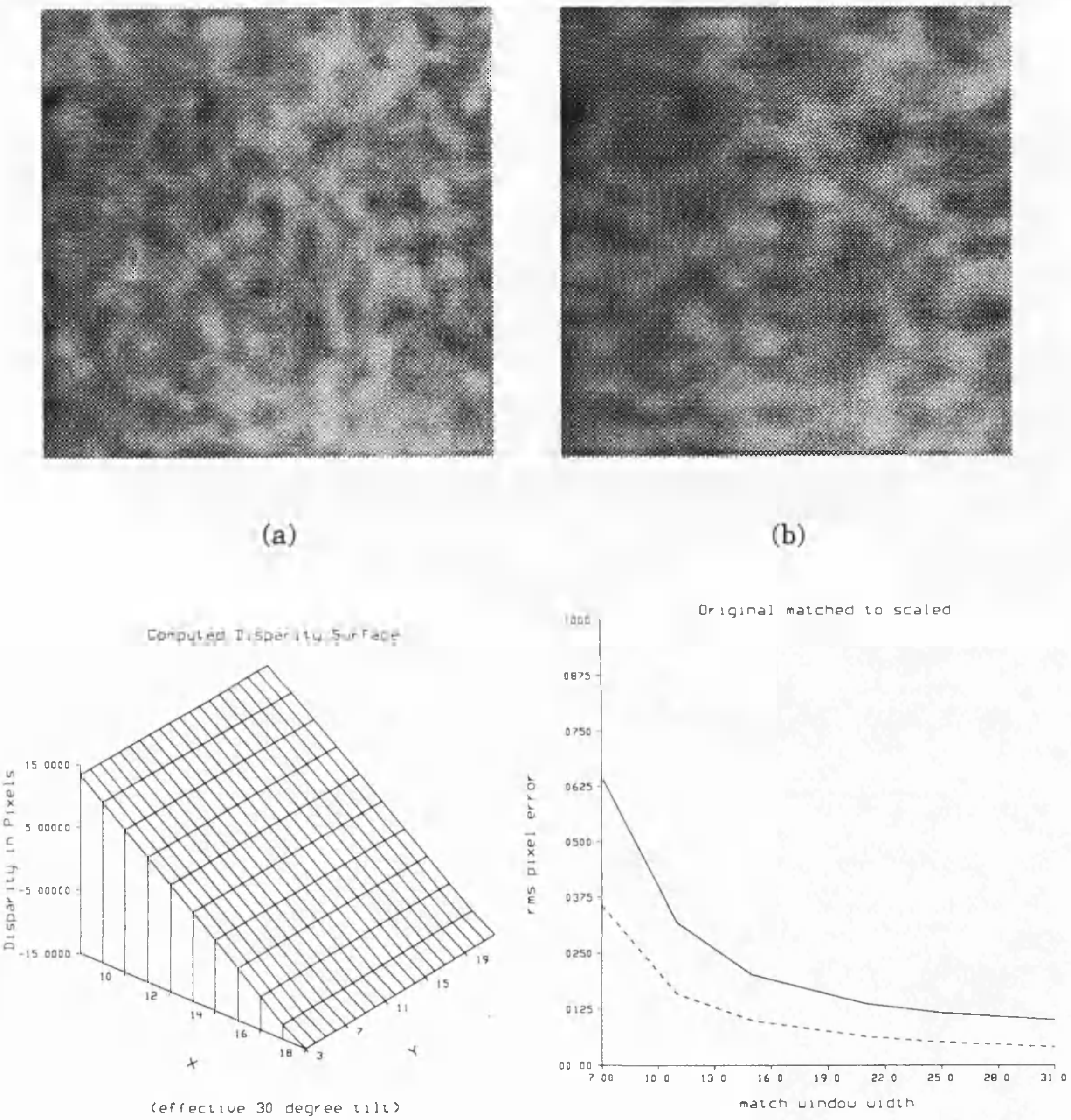
A model image was generated using the two-dimensional image formation formula developed in chapter 7, given by equation 7.5. The separable, linearly decreasing amplitude function, given by equation 7.7, was used to model the image spectrum and the resulting synthetic image is shown in figure 8.1a (the image parameters are $N = 128$, $n_c = 20$, $b = 2$, and $B = 130$). Three test images were derived from the original image by applying a scaling in the x axis (corresponding to a 30° tilt, if the images are taken to be a stereoscopic pair), a rotation of 15°, and a radiometric contrast stretch, which are shown in figures 8.1c, 8.1e, and 8.1g respectively.

In order to obtain a statistical measure of matching error, Gruen's algorithm was applied to the original image and the transformed image for all possible match points over a grid-like coverage of the original image, with a spacing of 5 pixels. The root mean square (r.m.s.) pixel error for all resulting match points was then obtained by comparing the expected and experimental match result for the x coordinate position. In addition, the mean LSEE was calculated from all resulting match points. This procedure was repeated for a varying match window size over a range from 7×7 to 31×31 pixels, and for each test image in turn. Figures 8.1d, 8.1g, and 8.1h plot both the r.m.s. pixel error and the mean LSEE measure against increasing match window size for each of the three test images respectively. Figure 8.1c shows the disparity surface obtained from matching the original and scaled images for a match window size of 15×15 pixels.

The resulting plots indicate several important properties of the matching process. As expected, the least-squares matching algorithm is not immune to interpolation error, even for a large number of iterations and low parameter thresholds. A small match window size can yield significant match errors, depending on the severity of the geometric and/or radiometric distortions. As the match window size increases, the r.m.s. pixel error approaches 1/100th of a pixel.

This is an acceptable error when considering all the factors that contribute to numerical precision, such as rounding error. More interestingly, the LSEE measure agrees well in magnitude with the experimental r.m.s. pixel error for all three test images, despite the absence of additive noise. This shows promise as a direct measure of match precision and might be used to prune wildly inaccurate matches.

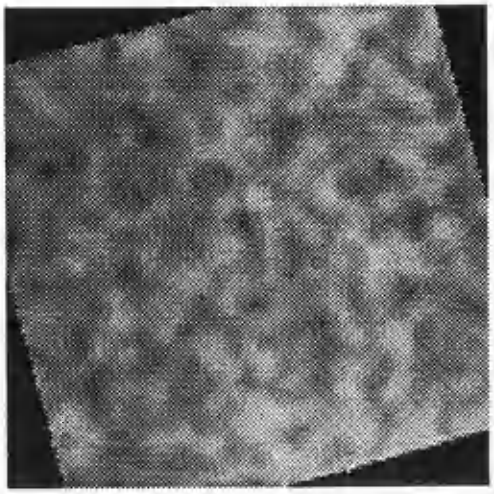
Quite surprisingly, the r.m.s. pixel error for the radiometric test image (figure 8.1h) was unexpectedly high, given that no interpolation was required for the correct match. The error may be due to the fact that all affine and radiometric parameters were weighted equally (in the least-squares sense) and thus the correct solution was not obtained directly. Another possible cause of the error was over-parameterisation; that is, there were too many parameters to compensate for such a simple (in this case radiometric) set of distortions. The weighting of parameters has been discussed by, for example, Rosenholm [Rosenholm87a], and requires further investigation in the context of this research.



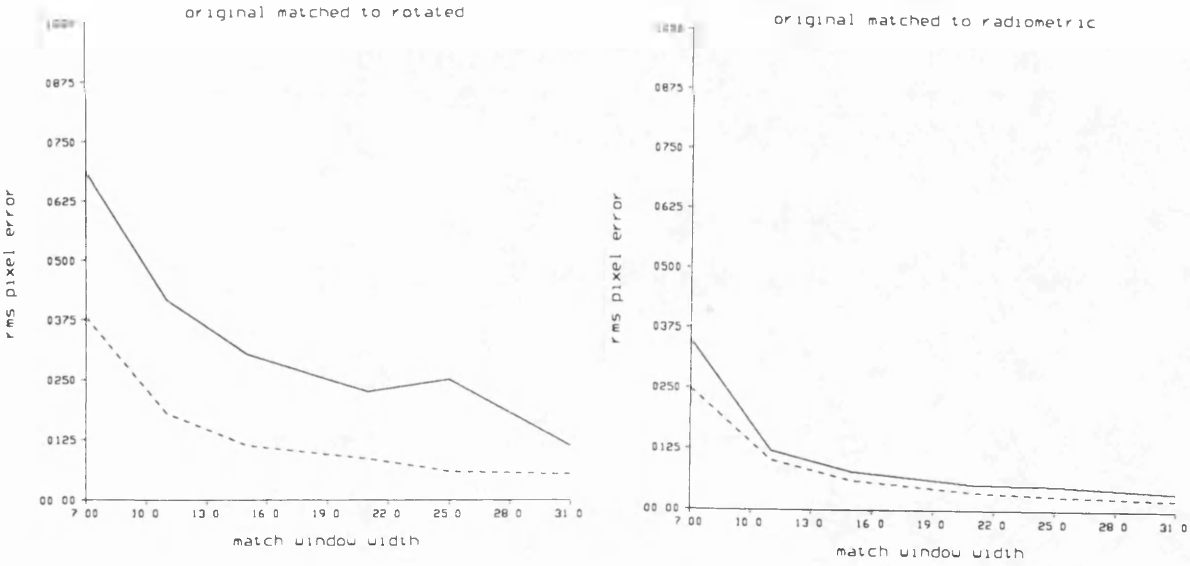
(c)

(d)

Figure 8.1. Application of Gruen's algorithm to synthetic data in the absence of noise. (a) original model image, (b) original image scaled in x (representing a 30° tilt), (c) disparity surface of (a) matched to (b), and (d) the rms pixel error versus match window size (the solid line represents the experimental error and the dashed line represents the mean LSEE measure). (continued ...)



(e) (f)



(g) (h)

Figure 8.1. (cont.) (e) original image rotated by 15°, (f) original image with radiometric variation (contrast stretch), (g) and (h) the rms pixel error versus match window size for (e) and (f) respectively. In each plot the solid line represents the experimental error and the dashed line represents the mean LSEE measure.

8.3 Application to Synthetic Data in the Presence of Noise

In this section the application of Gruen's algorithm to synthetic data in the presence of noise is assessed. In the previous section, the performance of Gruen's algorithm in the absence of additive noise was shown to be highly accurate for a range of geometric and radiometric differences between two images. However, the error prediction formulae developed in chapter 7 are valid only for cross-correlation in the presence of additive noise, and it will therefore be important to validate the theoretical Cramer-Rao Bound Estimate (CRBE) on synthetic data. In order to assess the utility of the CRBE, a series of synthetic images was generated which were based on the same underlying spectral model; that is, the amplitude function required by the image formation formula (equation (7.5)) was the same for all images in the series. The images were not identical in the spatial domain, however, since the phase components of their spectra were random.

To each image in the series, zero mean, Gaussian distributed random noise was added to generate an image pair, where the standard deviation of the noise was known. By using Gruen's algorithm to match selected points for each image pair and comparing the expected match result with that obtained experimentally, statistics on matching error could be obtained. As each image was generated from the model, with a known amplitude function, a corresponding formula for the CRBE could be obtained, as shown in chapter 7. The CRBE formula provided a prediction of the variance (and standard deviation) of the experimental error. The r.m.s. pixel error associated with all match points for each pair in the image series could then be compared to the predicted error given by the CRBE. As with the application of Gruen's algorithm to synthetic data in the absence of noise, it was also useful to assess the utility of the least-squares error estimator (LSEE) as an *a posteriori* measure of match precision in the presence of noise.

In order to give an unbiased validation of the error prediction formulae, two image series were generated using the two dimensional image model, where each series was based on a different amplitude function; the first on the separable, linearly decreasing amplitude function given by equation (7.7), and the second on the $1/f$ amplitude function given by equation (7.9). Each series consisted of 100 images, and for each image pair, 16 candidate match points were attempted, resulting in 1600 potential match results in the series. The candidate points were located such that their surrounding match windows had minimal overlap with match windows of neighbouring candidate points, to avoid systematic error. For each image series two parameters were varied. First, the amount of additive noise was varied from 1 to 25 standard deviations while maintaining a fixed match window size of 21×21 pixels, and second, the match window size was varied from 7×7 to 31×31 pixels for a fixed amount of additive noise representing a signal-to-noise ratio of 1:1. Thus for each image series two plots were obtained; one which plots experimental error versus increasing noise, and one which plots experimental error versus increasing match window size. Together with the experimental error, both the CRBE and mean LSEE were also plotted. (The reader is reminded that adding noise of standard deviation σ_n to one image only is equivalent to adding noise of standard deviation $\sigma_n/\sqrt{2}$ to each image).

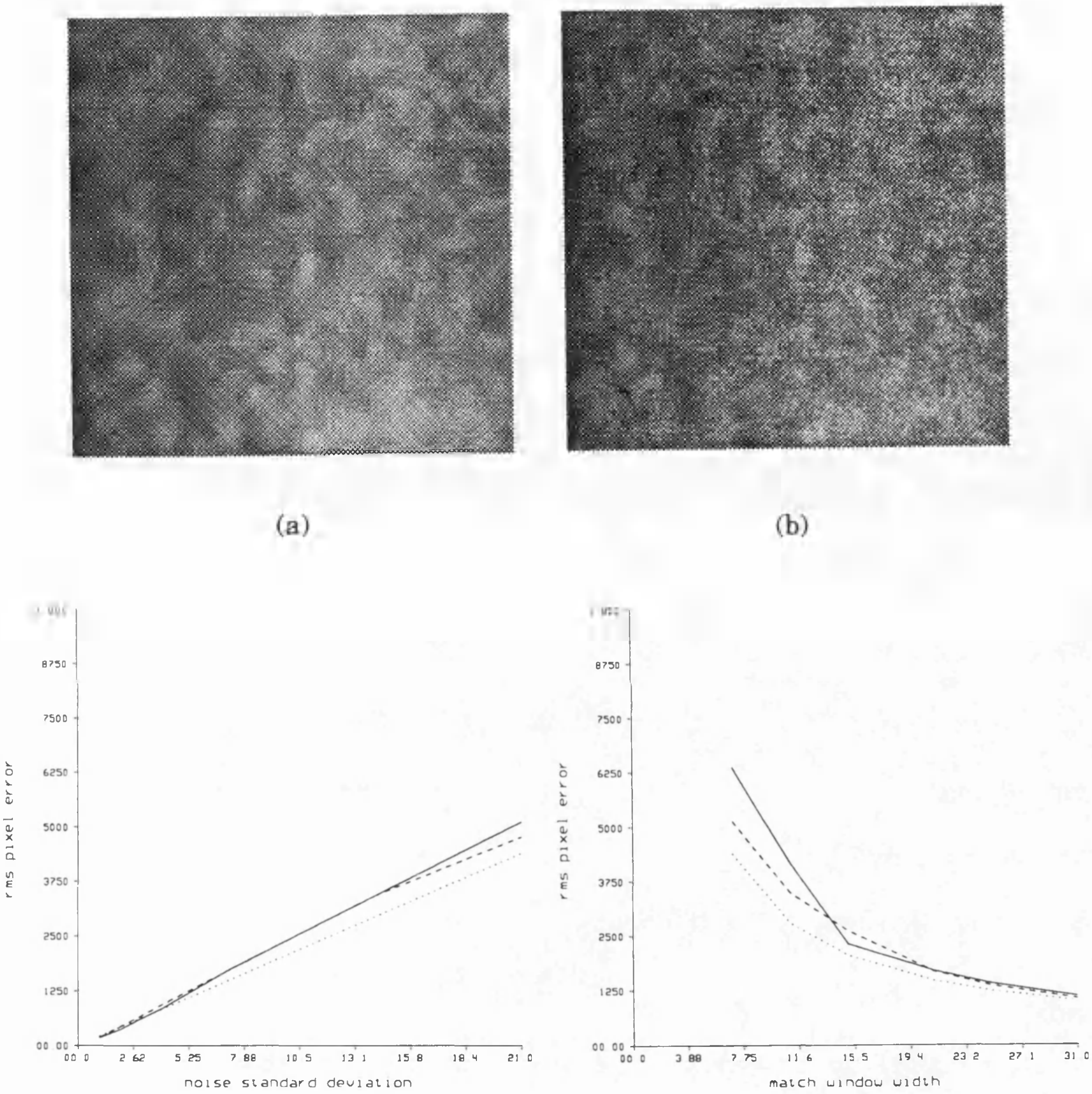
The first image series was generated using a separable, linearly decreasing amplitude function, and one of the 128×128 images in the series is shown in figure 8.2a. The model parameters were $N = 128$, $n_c = 21$, $b = 1$, and $B = 130$. The standard deviation of the model image, σ_f, was measured to be 7 grey levels. Figure 8.2b is the result of adding Gaussian distributed, random noise to the image in figure 8.2a, where the standard deviation of the noise, σ_n, was also 7 grey levels (and thus the signal-to-noise ratio, SNR $= \sigma_f/\sigma_n$, is 1:1). Figure 8.2c and 8.2d plot the experimental r.m.s. pixel error, the CRBE, and the mean LSEE versus increasing additive noise and increasing match window size respectively. The CRBE for the variance of the correlation error for the images in the first series is given by

$$\sigma_d^2 \approx \frac{60\, N^2\, \sigma_n^2}{4\, M^2\, b^2\, \pi^3\, n_c^4} \quad . \tag{8.1}$$

The second image series was generated using a 1/*F* amplitude function, and one of the 128×128 images in the series is shown in figure 8.3a. The model parameters were $N = 128$, $n_c = 16$, $b = 6$, and $B = 130$. The standard deviation of the model image, σ_f, was measured to be 12 grey levels. Note that this image is visually quite different from that in figure 8.2a. Figure 8.3b is the result of adding Gaussian distributed, random noise to the image of figure 8.3a, where the standard deviation of the noise, σ_n, was also 12 grey levels (SNR 1:1). Figure 8.3c and 8.3d plot the experimental r.m.s. pixel error, the CRBE, and the mean LSEE versus increasing additive noise and increasing match window size respectively. The CRBE for the variance of the correlation error for the images in the second series is given by

$$\sigma_d^2 \approx \frac{N^2 \sigma_n^2}{4 M^2 b^2 \pi^2 n_c} \quad . \tag{8.2}$$

For the experimental results plotted in figures 8.2c and 8.2d for the first image series and figures 8.3c and 8.3d for the second image series, the solid line represents the experimental r.m.s. pixel error, the dashed line represents the mean LSEE measure, and the dotted line represents the theoretical CRBE prediction for the given image model, given by equations 8.1 and 8.2. The results indicate that the experimental error agrees quite well with the theoretical prediction given by the CRBE, although the agreement deteriorates for large amounts of additive noise (SNR < 1:2) and small match window sizes ($M < 10$). Further, the *a posteriori* estimate of match error, the LSEE, also agrees well with both the experimental error and the CRBE, although this measure appears to be an over optimistic estimate of match error, as demonstrated in, for example, figure 8.3c. There is increasing experimental evidence that the magnitude of the LSEE can be used a direct indicator of match precision.


(a) (b)

(c) (d)

Figure 8.2. Comparison of theoretical and experimental matching error in the presence of noise - example #1. (a) original model image (linearly decreasing spectrum), (b) image (a) with additive noise (SNR 1:1), (c) plot of the rms pixel error versus increasing noise, and (d) the rms pixel error versus match window size. In each plot the solid line represents the experimental error, the dashed line represents the mean LSEE measure, and the dotted line represents the theoretical CRBE error measure for the given model.

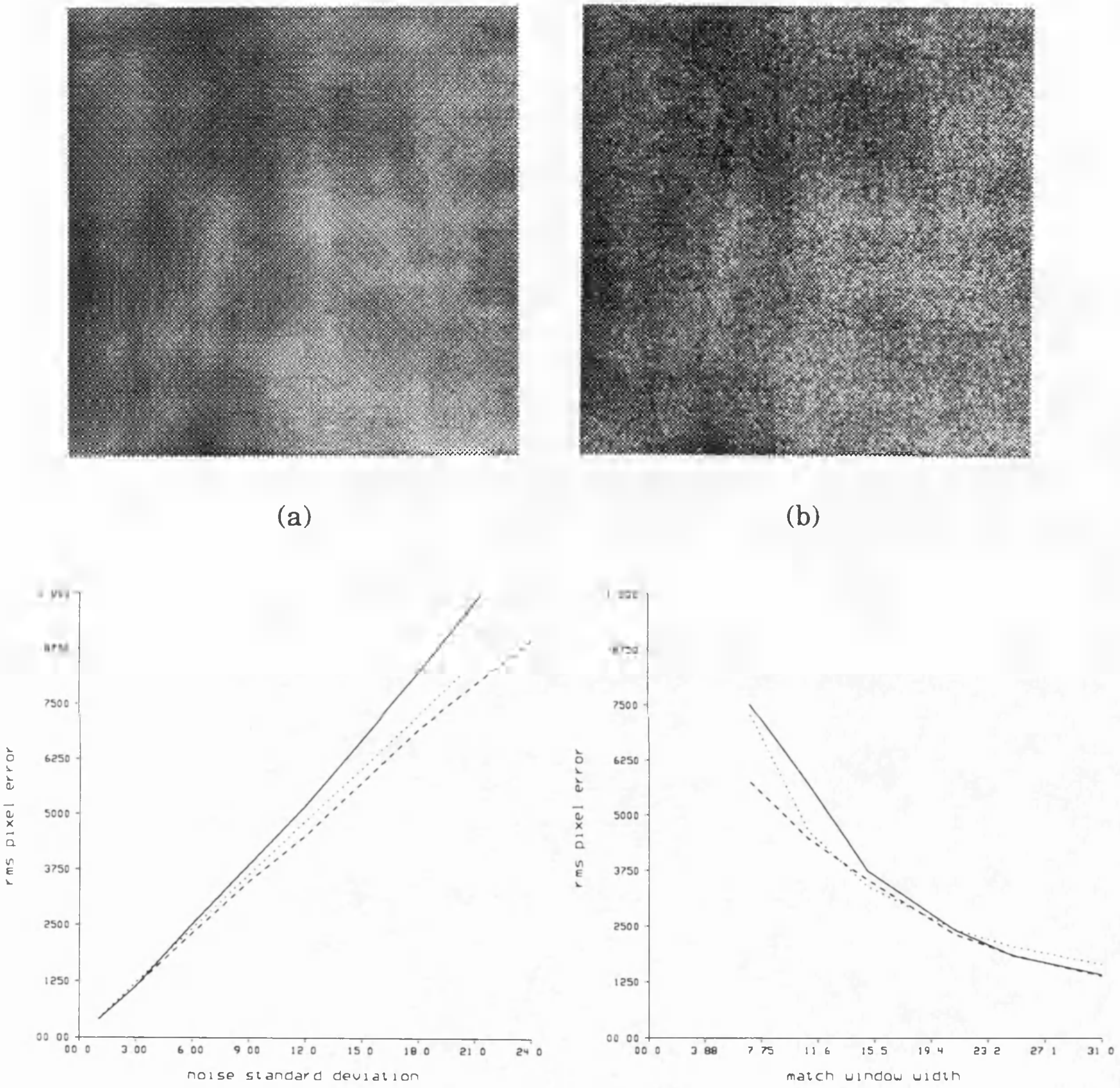

(a) (b) (c) (d)

Figure 8.3. Comparison of theoretical and experimental matching error in the presence of noise - example #2. (a) original model image (1/F spectrum), (b) image (a) with additive noise (SNR 1:1), (c) plot of the rms pixel error versus increasing noise, and (d) the rms pixel error versus match window size. In each plot the solid line represents the experimental error, the dashed line represents the mean LSEE measure, and the dotted line represents the theoretical CRBE error measure for the given model.

8.4 Application to Model Data

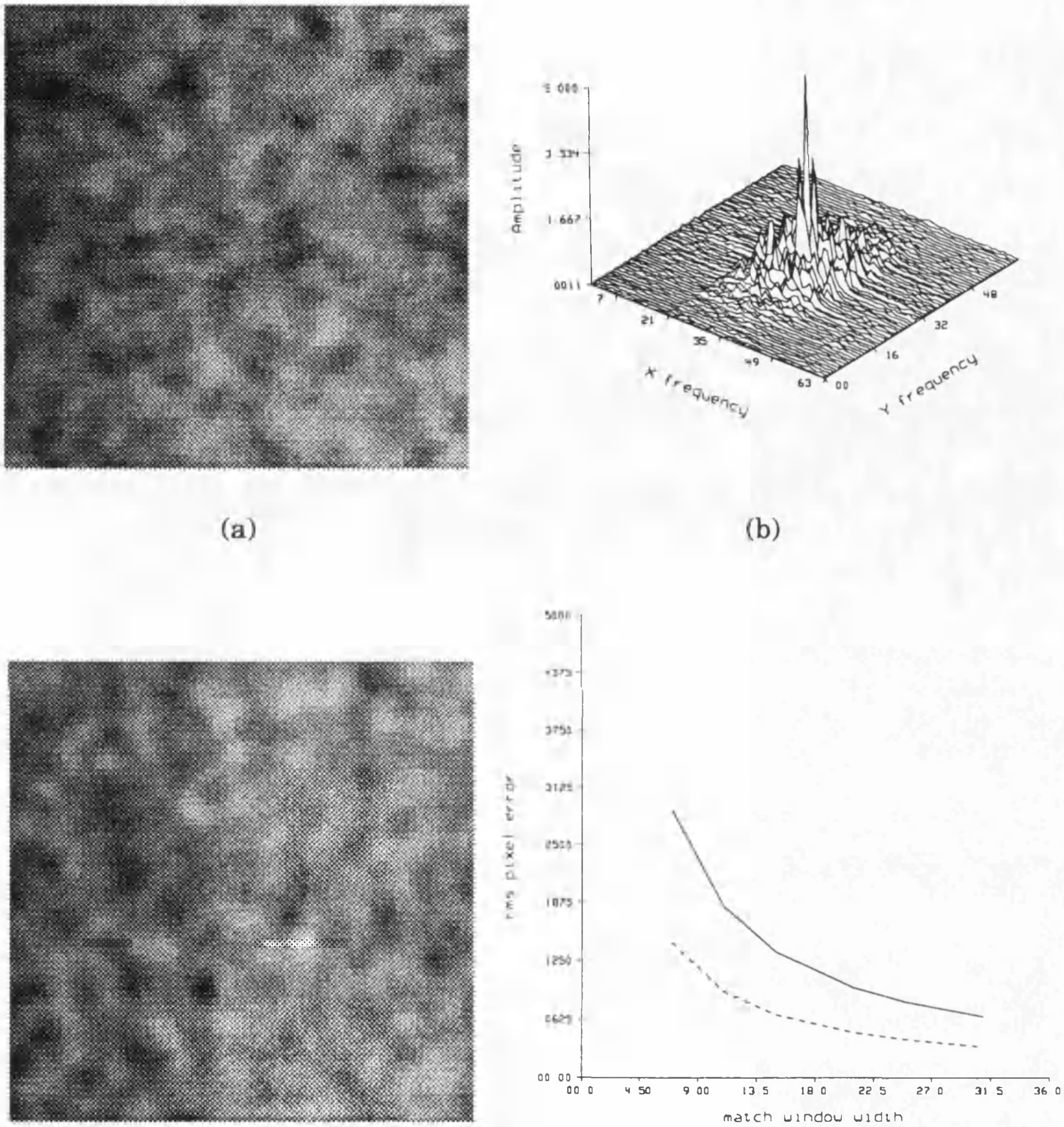
The plaster model data, Data Set A, were the only data used in this research for which ground truth values were known and thus real matching errors could be measured. In this section the tools for error prediction developed in chapter 7 are applied to this real world data. In the previous two sections the image model and the corresponding error prediction formulae have been validated, and the indication is, at least for synthetic data, that the formulae were useful for *a priori* estimates of matching accuracy. It was therefore of interest to see whether the same formulae could prove useful when applied to real world data such as Data Set A. In this section, the utility of the Cramer-Rao bound estimate (CRBE) is assessed for real data, with the added benefit that the ground truth is known for this data and previous experimental results were available.

To apply the image model developed in chapter 7, a 64×64 pixel region from the left-hand image from Data Set A was first extracted, and this image region is shown in figure 8.4a. Recall that the image model is only intended to model the local texture in an image, and thus only a small 64×64 region was extracted. The procedure was to determine the characteristic shape of the amplitude components of the Fourier spectrum, and then select an appropriate amplitude function for use with the image model (equation (7.5)). The grey level statistics of the image region were also measured; the mean grey value was 120, and the standard deviation, σ_f, was 16 grey levels. The Fourier spectrum of the image region is shown in figure 8.4b, for which a separable, linearly decreasing amplitude function (given by equation 7.7) was chosen as a suitable functional model. The parameters of the model were set to $N = 64$, $n_c = 16$, $b = 3$, and $B = 120$, and the resulting synthetic model image is shown in figure 8.4c. There is a visual similarity between the synthetic model image and the real image.

With an image model based on the separable, linearly decreasing amplitude function, the expression for the auto-correlation formula of the model could be derived, and is given by equation (7.23). This expression could then be used with the CRBE to make predictions about matching error. For example, given some estimate of the amount of noise present in the image (from sources such as electrical noise), a suitable match window size can be selected to yield an average match error below a desired threshold. Further, having previously selected a match window size of 15×15 pixels for the application of Gruen's algorithm to Data Set A in chapter 6, one can now ask what the expected matching error would be for a moderate amount of additive noise. In this case, such predictions can be compared directly with experimental results already obtained.

Figure 8.4d plots the predicted r.m.s. pixel error versus increasing match window size for the model image in the presence of noise. Two levels of noise are shown: the solid line represents additive noise yielding a signal-to-noise ratio (SNR) of 2:1, and the dashed line represents SNR 1:1. If the signal-to-noise ratio in the real image is taken to be 2:1 (which is not unlikely), the r.m.s. matching error is predicted to be 0.0671 pixels, using a match window size of 15×15 pixels.

From the application of Gruen's algorithm to synthetic data in the absence of noise it was shown that in the presence of geometric distortions, the matching error could be as high as 0.07 rms pixel, depending on the image model, match window size, and type of distortion. If the effects of matching error to due geometric corrections (interpolation error, numerical accuracy) and the predicted matching error due to additive noise are combined, the matching error could be as high as 0.15 r.m.s. pixels. This estimate agrees well with the mean experimental error of 0.194 pixels obtained in the application of Gruen's algorithm to Data Set A (presented in chapter 6, section 6.5.1). If the component sources of matching error are not cumulative, then either the prediction is over optimistic (which is the more likely) or the estimate of the amount of additive noise is too low. Despite this discrepancy, the application of the image model and the Cramer-Rao bound estimate to the plaster model data gave some indication that the error prediction formulae were of use with real image data.



(a) (b)

(c) (d)

Figure 8.4. Application of the image model and the matching error prediction theory to Data Set A. (a) a 64 × 64 pixel region of the plaster model image, (b) the Fourier spectrum of (a), (c) synthetic model image (see text), and (d) plot of the predicted rms pixel error versus match window size for the model image in the presence of additive noise; the solid line represents SNR 2:1 and the dashed line represents SNR 1:1.

8.5 Application to Fundus Data

In this section the application of the image model and error prediction formulae to the fundus data are studied, specifically to assess the utility of the formulae when applied to real data (the fundus image data), for which no ground truth were known. The driving inspiration for development of the error prediction tools was primarily to enable analysis of the matching error in the application of Gruen's algorithm to the fundus image data. In a sense, this section is the culmination of the development of the image model and error prediction formulae in chapter 7, and the experimental validation of the formulae in the earlier sections of this chapter. The suitability of the image model to represent a range of regions in a typical fundus image will be assessed. Experiments involving selected regions of a real fundus image and the application of the image model and CRBE formulae indicate that valid predictions about matching accuracy can be made.

Use of the error prediction formulae with the real fundus image data was complicated by the lack of ground truth from which to verify predictions with experiments. In the absence of ground truth measurements for the fundus image data, one could argue that it would be difficult to apply the image model and the CRBE in a meaningful way. It is argued here that there is strong experimental evidence for confidence in the error prediction formulae, and the least-squares error estimator has been shown to accurately reflect the magnitude of the accuracy of the match (as is demonstrated in sections 8.2, 8.3, and 8.4), and therefore both *a priori* (the CRBE) and *a posteriori* (the LSEE) analysis of fundus data disparity estimates obtained from stereo cross-correlation are possible.

To assess both the image model and error prediction formulae developed in chapter 7 for the fundus image data, three representative regions from both Data Sets B and C were studied. The results for the image regions from Data Set B are presented in detail, whereas the results for Data Set C are only intended to give additional experimental evidence that the theoretical formulae can be

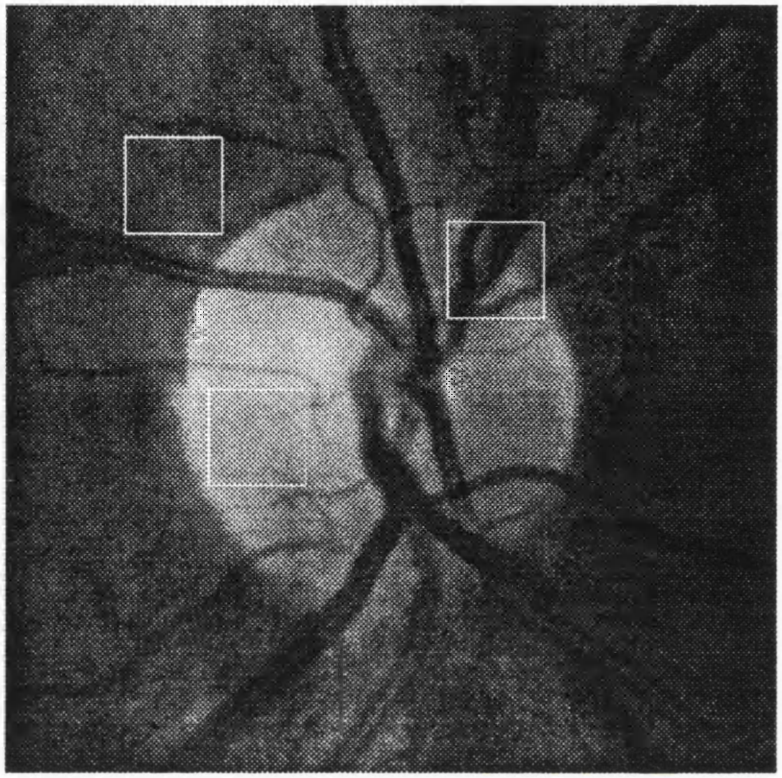


Figure 8.5. The left-view fundus image from Data Set B, indicating three regions selected for experimentation.

meaningfully applied to real image data.

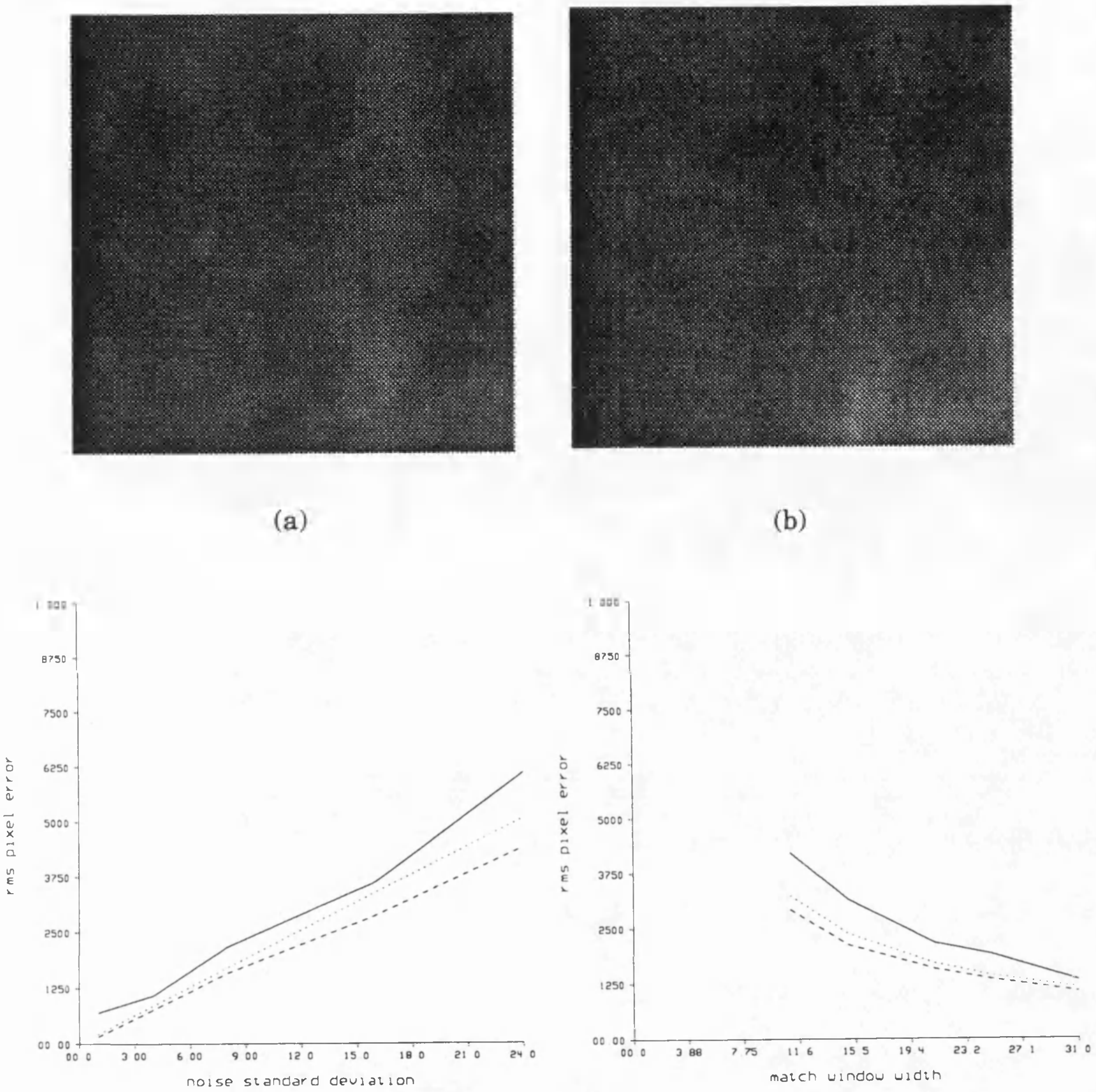
Figure 8.5 shows the left-hand image from Data Set B with the three selected regions highlighted. Each region was 64×64 pixels square; one located in the nerve fibre area surrounding the optic disc, the second located centrally in the optic cup area, and the third located in an area of retinal vessels. For each image region, the characteristic shape of the Fourier spectrum was modelled using a suitable amplitude function, and a synthetic image was generated using the image model (equation (7.5)). Unlike the earlier experiments, the model image (and its derived correlation function) was used only as a tool for error prediction, and the actual experimental match results were obtained from the real fundus image regions.

The predicted matching error estimated by the CRBE for each model image was compared to the experimental error obtained from the application of Gruen's algorithm to the corresponding real image region in the presence of additive noise. As with the application of Gruen's algorithm to synthetic data in the presence of noise, presented in section 8.3, match results was obtained for both a varied amount of additive noise with a fixed match window size and for a varied match window size with a fixed amount of additive noise. Zero mean, Gaussian distributed random noise was added to each image region to produce an image pair for matching. Gruen's algorithm was then applied to match points densely across the image pair. With a few starting seed match points, the region growing technique was used with Gruen's algorithm to match the every fifth pixel across the entire image. This was repeated by generating a new image with added noise, and thus a new image pair, and matching the entire image again. By repeating the entire process 10 times for each image region, where typically 64 match results are obtained for each image pair, a statistical measure of about 500 to 1000 match points was obtained for each experiment, and for each image region.

The nerve fibre image region was shown in figure 8.6a together with the model image in figure 8.6b. The model image for the nerve fibre region is based on a circularly symmetric, linearly decreasing amplitude function, where the model parameters were $N = 64$, $n_c = 10$, $b = 2$, and $B = 90$. The grey level statistics of the nerve fibre image region were known, with a mean value of 90 and a standard deviation of 8 grey levels. Figure 8.6c plots the experimental r.m.s. pixel error versus increasing additive noise, where the standard deviation of the noise ranged from 1 to 24 grey levels, and the match window size was constant at 21×21 pixels. Figure 8.6d plots the experimental r.m.s. pixel error versus increasing match window size, where the window size varied from 11×11 to 31×31 pixels and the standard deviation of the noise was constant at 8 grey levels (SNR 1:1).

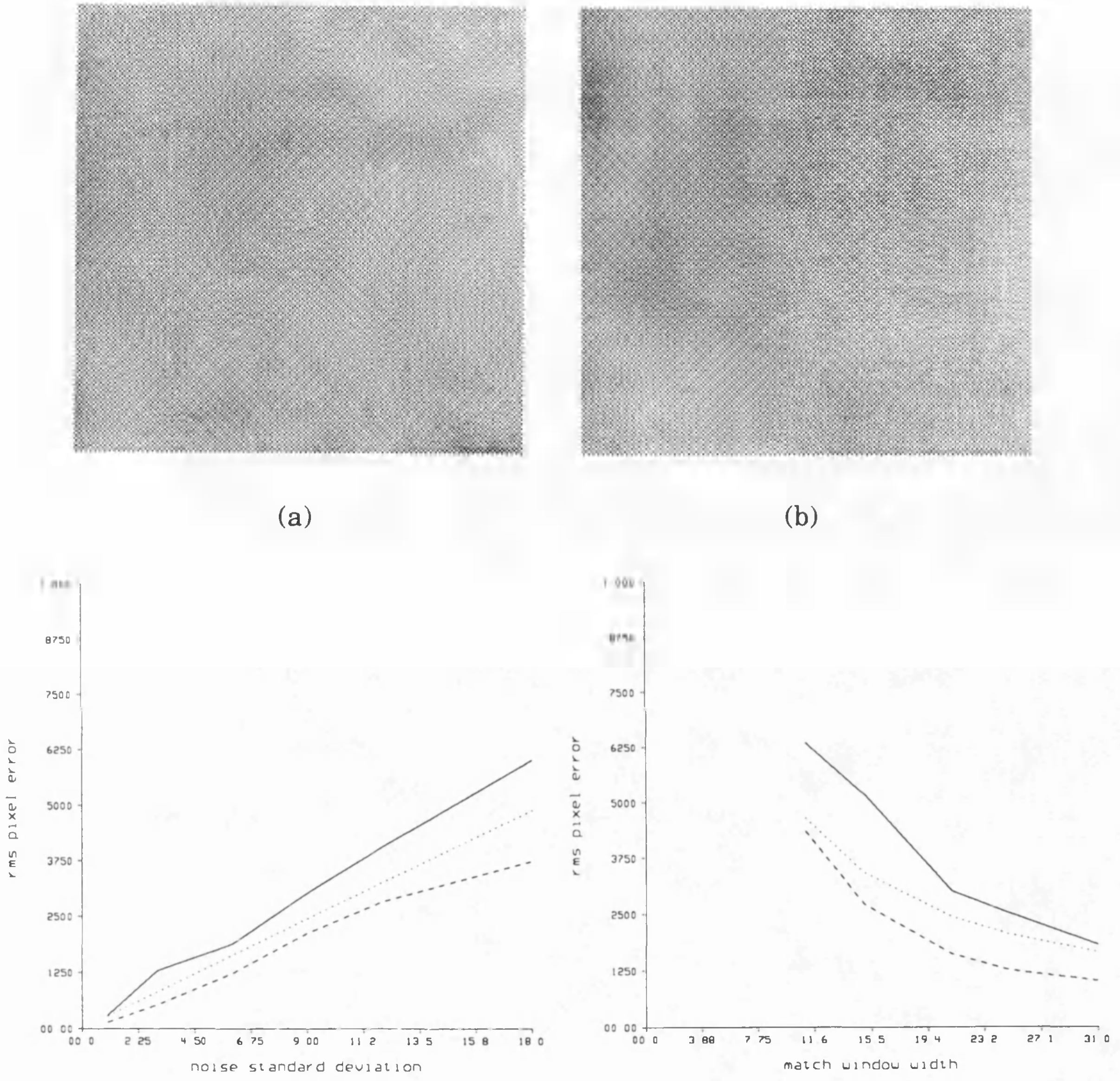
The optic cup image region is shown in figure 8.7a together with the model image in figure 8.7b. The model image for the optic cup region was based on a 1/F amplitude function, where the model parameters were $N = 64$, $n_c = 20$, $b = 4$, and $B = 188$. The grey level statistics of the optic cup image region were known, with a mean value of 188 and a standard deviation of 9 grey levels. Figure 8.7c plots the experimental r.m.s. pixel error versus increasing additive noise, where the standard deviation of the noise ranged from 1 to 18 grey levels, and the match window size was constant at 21×21 pixels. Figure 8.7d plots the experimental r.m.s. pixel error versus increasing match window size, where the window size varied from 11×11 to 31×31 pixels and the standard deviation of the noise was constant at 9 grey levels (SNR 1:1).

The retinal vessel image region is shown in figure 8.8a together with the model image in figure 8.8b. The model image for the retinal vessel region was based on a 1/F amplitude function, where the model parameters were $N = 64$, $n_c = 4$, $b = 14$, and $B = 97$. The grey level statistics of the retinal vessel image region were known, with a mean value of 97 and a standard deviation of 24 grey levels. Figure 8.8c plots the experimental r.m.s. pixel error versus increasing additive noise, where the standard deviation of the noise ranged from 1 to 24 grey levels, and the match window size was constant at 21×21 pixels. Figure 8.8d plots the experimental r.m.s. pixel error versus increasing match window size, where the window size varied from 11×11 to 31×31 pixels and the standard deviation of the noise was constant at 12 grey levels (SNR 2:1 - matching error for SNR 1:1 is too high, and a sufficient number of match results could not be obtained for the smaller window sizes).



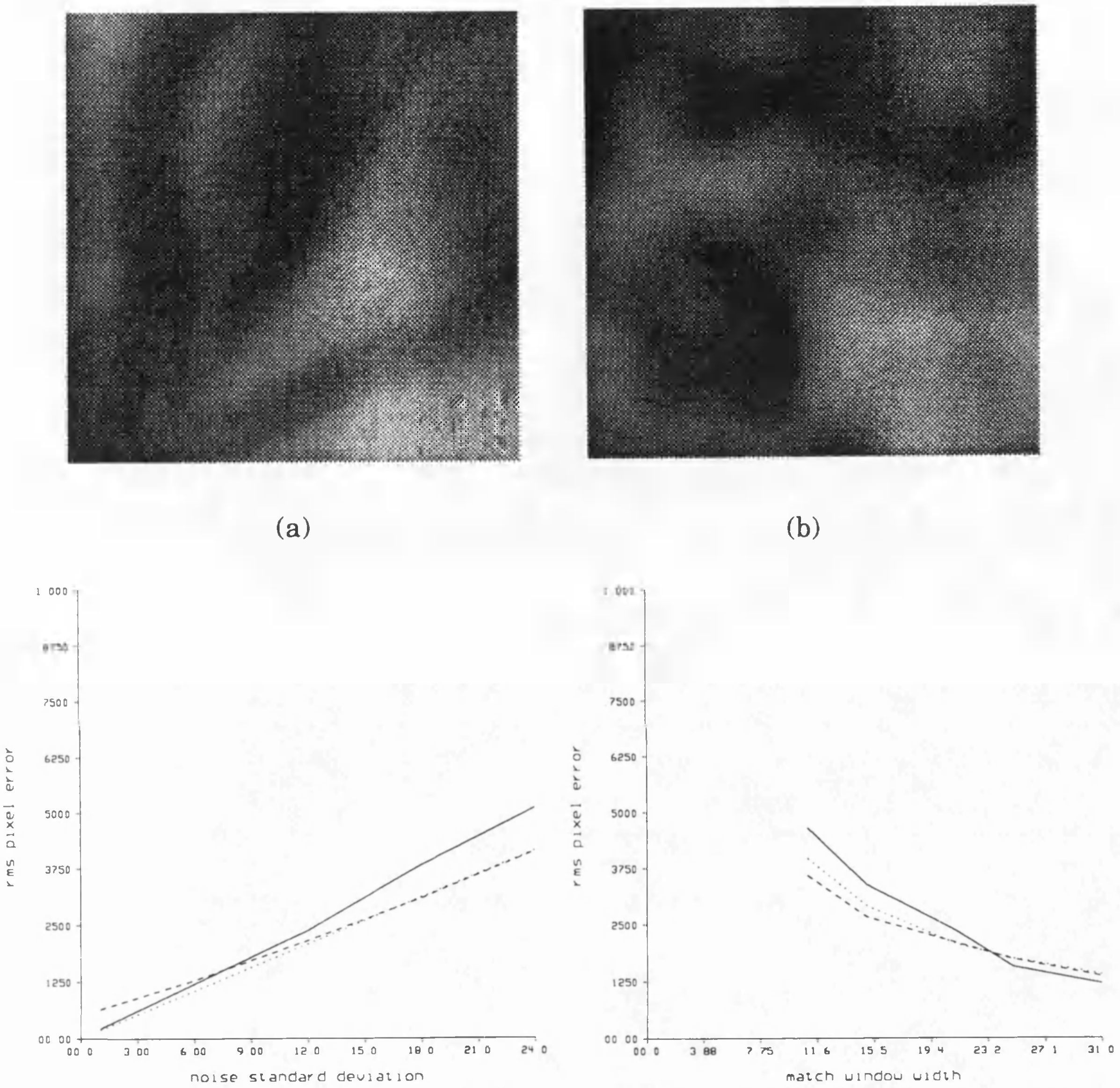
(c) (d)

Figure 8.6. Nerve fibre region - comparison of theoretical and experimental matching error in the presence of noise. (a) 64 × 64 fundus image region (see figure 8.5), (b) generated model image representing (a), (c) plot of the rms pixel error versus increasing noise, and (d) the rms pixel error versus match window size. In each plot the solid line represents the experimental error, the dashed line represents the mean LSEE measure, and the dotted line represents the theoretical CRBE error measure based on the image model.


(a) (b)

(c) (d)

Figure 8.7. Optic cup region - comparison of theoretical and experimental matching error in the presence of noise. (a) 64 × 64 fundus image region (see figure 8.5), (b) generated model image representing (a), (c) plot of the rms pixel error versus increasing noise, and (d) the rms pixel error versus match window size. In each plot the solid line represents the experimental error, the dashed line represents the mean LSEE measure, and the dotted line represents the theoretical CRBE error measure based on the image model.



(a) (b)

(c) (d)

Figure 8.8. Retinal vessel region - comparison of theoretical and experimental matching error in the presence of noise. (a) 64 × 64 fundus image region (see figure 8.5), (b) generated model image representing (a), (c) plot of the rms pixel error versus increasing noise, and (d) the rms pixel error versus match window size. In each plot the solid line represents the experimental error, the dashed line represents the mean LSEE measure, and the dotted line represents the theoretical CRBE error measure based on the image model.

Figure 8.9 shows the left-hand image from Data Set C with the three selected regions highlighted. Each region was 64×64 pixels square, and as with Data Set B, was located in a representative area of the fundus image. For each image region, the characteristic shape of the Fourier spectrum was modelled using a suitable amplitude function. The predicted matching error given by the CRBE for each model image was compared to the experimental match error obtained from the application of Gruen's algorithm in a like manner to that outlined for the image regions from Data Set B.

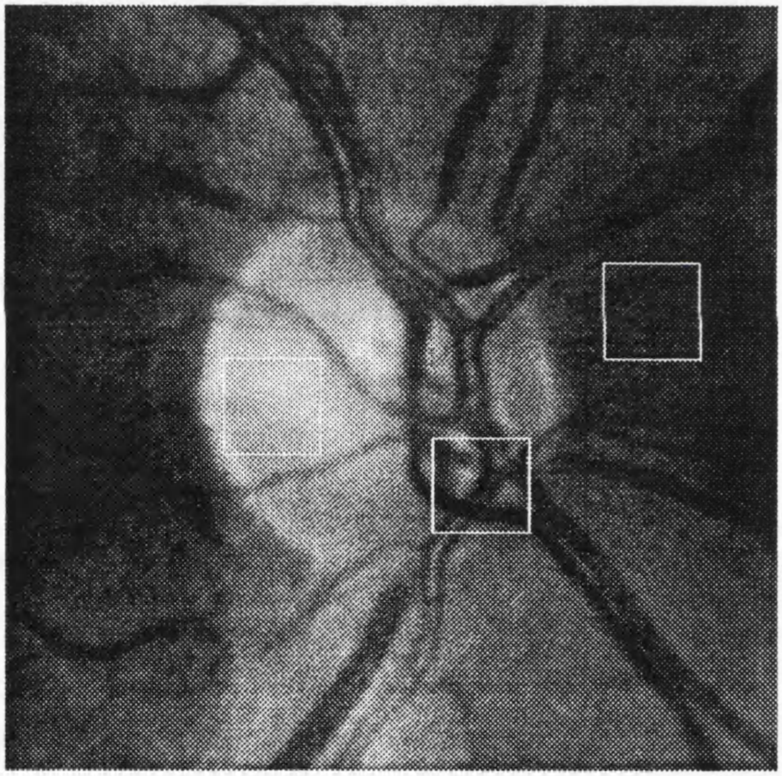
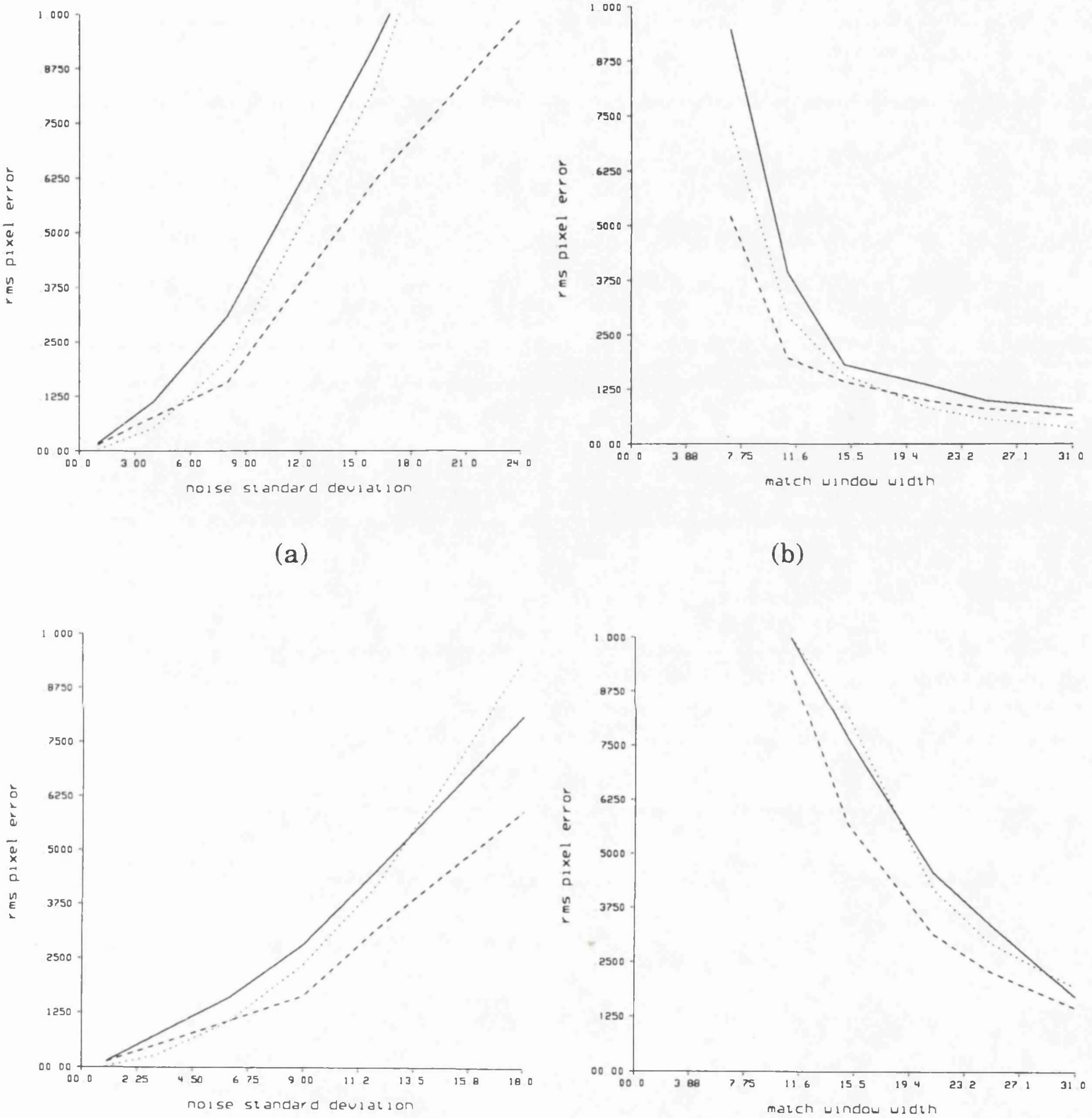


Figure 8.9. The left-view fundus image from Data Set C, indicating three regions selected for experimentation.

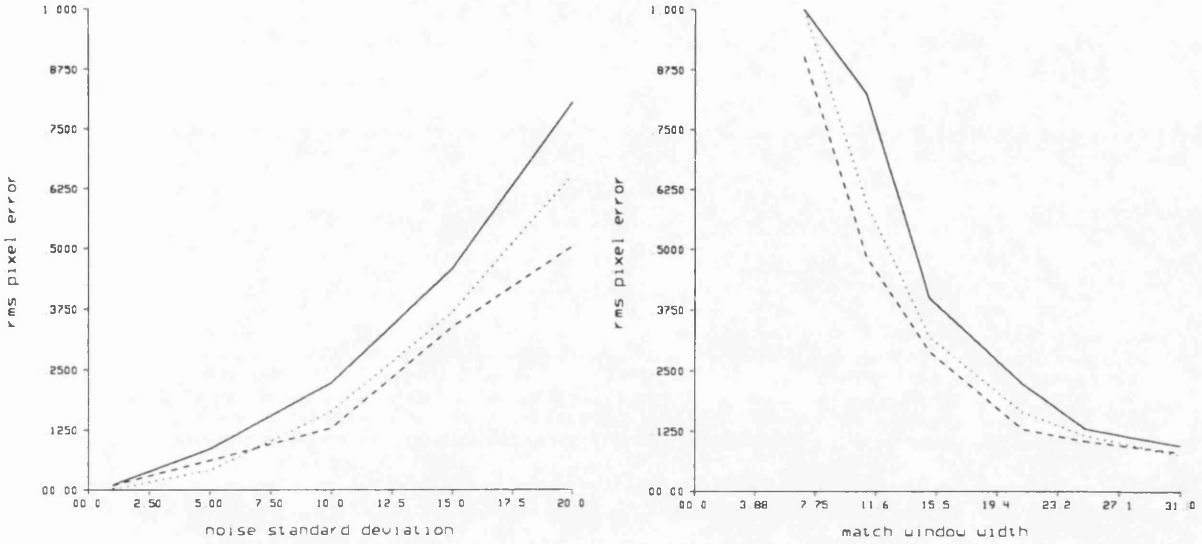
Figures 8.10 shows the plotted results for each of the image regions shown in figure 8.9. The grey level statistics for each image region, the model parameters, and the corresponding experimental methods were similar to those outlined for the image regions from Data Set B and are not detailed here.



(a) (b)

(c) (d)

Figure 8.10. Data Set C image regions - comparison of theoretical and experimental matching error in the presence of noise. For the nerve fibre region (see figure 8.9), (a) plots the rms pixel error versus increasing noise, and (b) plots the rms pixel error versus match window size. For the optic cup region, (c) plots the rms pixel error versus increasing noise, and (b) plots the rms pixel error versus match window size. (continued...)



(e) (f)

Figure 8.10. (cont.) For the retinal vessel region (see figure 8.9), (e) plots the rms pixel error versus increasing noise, and (f) plots the rms pixel error versus match window size. In each plot the solid line represents the experimental error, the dashed line represents the mean LSEE measure, and the dotted line represents the theoretical CRBE error measure based on the image model.

For all the experimental results plotted in figures 8.6 through 8.10 the solid line represents the experimental error, the dashed line represents the mean LSEE measure, and the dotted line represents the theoretical prediction from the CRBE of each model image. The results obtained for all image regions indicate that the predicted error based on the model image agrees quite well with the experimental error measured from the real image. Further, the plotted curve for the mean LSEE is closely aligned with the predicted error, bringing further evidence that the both the *a priori* CRBE and the *a posteriori* LSEE measure accurately predict and reflect the magnitude of the error of the match. Surprisingly, the retinal vessel image regions, which one might expect to be difficult to model, yielded the best results. The nerve fibre image region also yielded good results. These results are discussed further in the following section.

8.6 Discussion

With the experimental evidence that both the error prediction formulae and the least-squares error estimator can provide meaningful measures for assessing matching error in both synthetic and real image data, a discussion of the limitations of such theoretical tools is presented in this section. It is important not only to assess the utility of the tools developed in chapter 7, but also to define the conditions under which the tools are not applicable.

As demonstrated by the experiments conducted in section 8.3, the error prediction formulae, which are based on the auto-correlation function of an image, were shown to be valid when used with synthetic images, where the image model and the underlying auto-correlation function were known. In this sense, the image model, derived auto-correlation, and error prediction formulae were shown to be self consistent. What was encouraging, though, was that the same image model and error prediction formulae applied to real image data gave good agreement between theory and experiment. Specifically, the application of these theoretical tools to the fundus image data gave an indication that such tools might be used in a systematic way to measure, assess, and control matching cross-correlation error.

The set of experiments presented in this chapter have shown that the Cramer-Rao bound ([Whalen71], [Bogler86]) estimate might be used as an *a priori* measure of match error from which a suitable threshold can be set for an acceptable level of match error. This threshold might be used to filter match results using the *a posteriori* LSEE measure of match precision. A match result with a LSEE below the threshold (or some scalar multiple of it) could be accepted, and any wildly inaccurate results could be discarded (see also [Ryan80,81]). This process is consistent with the *region growing* technique [Otto89] used in conjunction with Gruen's algorithm, detailed in chapter 5, where the starting parameters for match points (i.e. the x and y coordinates) are estimated from nearby points for which a match has already been obtained. If the match error could be assessed during computation, then the region growing process could proceed from the best match first, thereby improving the overall results. This type of match result filtering requires further research.

It is important to recall the criteria by which the image model was developed, specified in chapter 7, so as not to be over critical of the limitations of the model. Images which appear to be highly structured macroscopically, are often relatively homogeneous (and isotropic) in texture microscopically. The image model is to be used as a tool for studying image cross-correlation, and match window sizes are usually of modest extent (typically less than 50×50 pixels). Therefore, the image model was used in the experiments to model only small regions in an image. The model is based on the amplitude components of an image Fourier spectrum, since image spectra often exhibit a characteristic functional shape, whose parameters correspond directly to the types of visual image parameters one might wish to vary.

The application of the image model to three fundus image regions in section 8.4 highlights some of the empirical strengths and limitations of the model. Both the nerve fibre image region and the retinal vessel region were relatively easy to model on a local scale, where the image textures were homogeneous and their spectra were easily modelled by a symmetric amplitude function. It was encouraging that the image model could accurately represent two such different image textures as the nerve fibre image region (figure 8.6) and the retinal vessel image region (figure 8.8). The results indicate that for relatively low levels of noise (SNR > 1:1), there was good agreement between theory and experiment. As the match window size increased, the experimental error approached the Cramer-Rao bound estimate as shown, for example, in figure 8.6d.

The results for the optic cup image region (figure 8.7), however, were not as encouraging. The form of the amplitude components of the image spectrum were difficult to model, and the resulting synthetic image was visually not the best representation of the real image. The optic cup image region was an example of an image with little grey level structure, where a small amount of additive noise can cause a large deviation in the grey level statistics of the image. Further development of the image model is required to cope with more structured or complex image textures. It is most important to obtain an accurate measure of matching error in the optic cup region of the fundus image (since the topography of the optic disc and cup is the main interest), and further development and assessment of the image model and the Cramer-Rao bound estimate is needed. When the form of the amplitude of the image spectrum is difficult to model, as with the optic cup image region, the model is less than ideal.

The experimental results for the synthetic data are representative of an ensemble or series of images, each with the same spectral form, where each image is a different realisation of a particular image model. The cumulative results for each image series, in effect, average out the bias introduced by any individual image in the series. Therefore, the experimental results (figures 8.2 and 8.3) are in good agreement with the theory. The experiments involving the real fundus image data were representative of a specific scene structure, where the same original image was used in each trial. Thus, the results may contain some bias, due to the specific structure in these fundus image regions.

It is evident that the image model can provide a good representation of real image data when the scene texture is relatively "busy" and homogeneous in pattern. Further, the scene spectrum should be similar in form in both the x and y axis (isotropic), where the form or shape of the amplitude components is functionally either circularly symmetric or linearly separable (examples of this type of spectrum are presented in chapter 7). When the scene spectrum exhibits a different form in each axis, the incorporation of scale factors in both the x frequency and y frequency or a more complex separable functional form for the amplitude components must be considered. The image model could be developed to cope with this type of scene spectrum. When the scene texture is relatively "flat" or constant in grey value, or if the scene contains a unique (non-homogeneous) texture pattern, the image model and corresponding error prediction formulae are not suitable.

The model was used to represent image regions, with the assumption that sub-regions of the image (as taken within a match window) have a spectrum which is similar in form to the larger (surrounding) image region. As the match window size approaches the full extent of the image region, this assumption is valid. Small match window sub-regions may contain little of the global grey level structure, and may have a spectrum which is very dissimilar in form to the spectrum of the whole image region. The image models which represent the fundus image regions were thus too idealistic, and there was some divergence between experiment and theory. The real image spectra did not have the ideal form of their model counterparts, and the spatial grey level structure was not as uniform in the real image regions. The error predictions based on the model image do not take into account the specific structures of any individual real image, because they are intended to represent an average value for many image regions with the same spectral form.

Apart from the general development of the image model and the associated error prediction formulae to handle a broader range of image domains, there are a number of other areas which could be studied further. For any individual set of matching errors, it would be interesting to study their distribution to see if a characteristic shape for the distribution could be found (for a given image model and amplitude function). This would require a much larger set of data and experimental results than available here in order to yield a statistically confident assessment. By characterising the form of the distribution of errors (e.g. Poisson distributed), it would then be possible not only to predict the magnitude of the matching error, as demonstrated in this chapter, but also to give the associated probability of such an error occurring. With a more rigorous mathematical treatment of the error prediction formulae, it may be possible to derive suitable analytical probability measures.

8.7 Summary

In this chapter the image model, correlation formulae, and error prediction formulae developed in chapter 7 were tested and applied. The application of the stereo matching algorithm, Gruen's algorithm, to geometrically distorted image data was assessed in order to verify the degree of matching accuracy in the absence of additive noise. The use of a mathematical image model (equation (7.5)) allows for the derivation of specific auto-correlation functions given specific amplitude functions. With an analytical expression for the auto-correlation function for a given model image, the matching error in the presence of additive noise can be predicted using the Cramer-Rao bound estimate developed in chapter 7.

The Cramer-Rao bound error prediction formula, which is based on the auto-correlation function of an image, was shown to be valid when used with synthetic images, where the image model and the underlying auto-correlation function are known. The image model developed in chapter 7 was shown to be applicable to two real image data; Data Set A, the plaster model image, and Data Set B, a typical fundus image. In image matching applications where ground truth verification of match results is not possible, the tools developed in chapter 7 can be used to accurately reflect the expected magnitude of cross-correlation matching error. The set of experiments presented in this chapter show that the Cramer-Rao bound estimate can be used as an *a priori* measure of match error, and can be compared to the *a posteriori* least-squares error estimator obtained for each match point. The application of the image model to the fundus image data, and the assessment of both the theoretical error prediction formula and the least-squares error estimator gives an indication that such theoretical tools might be used in a systematic way to measure, assess, and control matching cross-correlation error.

Chapter 9

Conclusions and Suggestions for Further Work

9.1 Summary of the Research

Stereoscopic photography of the human fundus and the subsequent calculation of the topography of the optic disc is of great potential clinical value if accurate measurements of the various parameters that characterise the changing shape of the optic disc topography can be provided. The research presented in this thesis has been an investigation into the problems of obtaining useful clinical measurements from stereo photographs of the optic disc, through automation of the stereometric procedure by digital stereo matching and image analysis techniques.

Following the introduction of the research topic in chapter 1, an overview of the research was presented and the objectives of the research were defined in chapter 2. The general problem of stereoscopy of the human fundus was reviewed together with a discussion of the clinical objectives, data acquisition and data sources, measurement by stereo matching, and methods for modelling the data. At the end of the chapter a set of research objectives was proposed, and these objectives are reviewed in the following section.

A survey of the current clinical methods for stereoscopic measurement of the optic disc was presented in chapter 3. From the literature it was apparent that the problem of topographic analysis of the optic nerve head from stereo photographs was by no means solved, even by commercially available systems. A brief historical survey of stereoscopic fundus photography was given so that some insight could be gained into the difficulties that may arise in the application of digital stereo matching to the fundus data.

Fundus image data acquisition, stereo geometry, limitations of resolution and accuracy, and other relevant physical constraints related to fundus imaging were investigated in chapter 4. In order to understand the difficulties associated with stereo photography of the optic disc, various physical constraints were examined. The parameters which ultimately define the accuracy to which topographic measurements of the optic disc can be made include the size of the optic disc, the stereoscopic baseline, eye movements, motion blur, optical defocus, and illumination variations. The conclusions drawn from chapter 4 indicate that the results obtained from stereoscopic analysis of photographs of the optic disc must be shown to be consistent, accurate to an acceptable level, and reproducible. Not until the stereoscopic effect can be standardised will stereo analysis of the optic disc prove reliable.

In chapter 5, a survey of digital stereo matching algorithms was presented and their strengths and weaknesses were explored, specifically as they related to the suitability of the algorithm for use with the fundus image data. Stereo matching algorithms including feature-based matching, area-based matching, optical flow techniques, and elastic matching were reviewed and evaluated. The potential problems for stereo matching algorithms specific to the fundus image data were also addressed, leading to a set of criteria on which to base the selection of an algorithm. In considering the relevant algorithms in the literature, an area-based algorithm based on the least-squares matching technique was selected. The mathematical foundations of Gruen's algorithm [Gruen85] and its extensions [Otto89] were described in detail.

The algorithm for computer implementation of the Gruen-Otto stereo matching algorithm was presented in chapter 6 together with a brief analysis of the computational requirements. The experimental methods applied to four test data sets were discussed, including data acquisition, image pre-processing, stereo matching, and post-processing and analysis. The data included a physical model for which ground truth could be measured and thus real matching error could be assessed. Three representative sets of fundus image data were also used; non-simultaneous and simultaneous stereo images obtained from standard fundus cameras, and non-simultaneous stereo images obtained from the Scanning Laser Ophthalmoscope.

In review of the results presented in chapter 6, the importance of data standardisation became apparent, and a more consistent data format should be adopted by clinicians if the methods employed here are to be reliable. Similarly, the importance of a large stereoscopic baseline, discussed in chapter 4, was also well demonstrated. An analysis of the results highlighted a number of important issues with respect to determining matching accuracy, and several of the key problems associated with the stereo matching algorithm and the fundus image data were discussed.

A mathematical model of two-dimensional image formation was developed in chapter 7 together with its corresponding auto-correlation function. The model was used as a tool for exploring key problems with respect to the stereo matching of fundus images. Specifically, measures for predicting correlation matching error were developed and applied. Such measures were shown to be of use in applications where the results of image correlation could not be independently verified, and meaningful quantitative error measures were required.

The local structure of the fundus image data was examined using well known image descriptors, and this led to a functional description of image formation. A generic image model was developed which characterised the image structure of discrete image data. Explicit formulae for the auto-correlation of the image model were derived in both one and two dimensions. These formulae were then extended to mean square difference auto-correlation. With an analytical expression for the auto-correlation function for a given model image, the matching error in the presence of additive noise could be predicted using the Cramer-Rao bound estimate ([Ryan80], [Bogler86]). The utility of this measure was demonstrated on both one- and two-dimensional image models.

In chapter 8 the image model, correlation formulae, and error prediction formulae developed in chapter 7 were tested and applied. The application of the stereo matching algorithm to geometrically distorted image data was presented in order to assess the degree of matching accuracy in the absence of additive noise. The Cramer-Rao bound error prediction formula was shown to be valid when used with synthetic image data in the presence of additive noise. The image model developed in chapter 7 was shown to be applicable to three real image data sets; Data Set A, the plaster model image, and Data Sets B and C, two typical fundus images. In image matching applications where ground truth verification of match results is not possible, the tools developed in chapter 7 can be used to accurately reflect the expected magnitude of cross-correlation matching error. The set of experiments presented in chapter 8 show that the Cramer-Rao bound estimate could be used as an *a priori* measure of cross-correlation matching error. The application of the image model to the fundus image data, and the assessment of both the theoretical error prediction formula and the least-squares error estimator gives an indication that such theoretical tools could be used in a systematic way to predict, assess, and control cross-correlation error.

9.2 Review of the Research Objectives

It is of interest to reflect upon the original research objectives presented in chapter 2 and assess how well each objective has been met. The primary objective was to ascertain whether or not reliable stereoscopic measurements could be obtained from photographs of the optic disc, and this objective has been thoroughly addressed in chapters 4, 5, and 6. The issues regarding image acquisition were addressed in chapter 4 for both non- and simultaneous stereo photographs. A suitable stereo matching algorithm was selected in chapter 5, and the application of the algorithm was demonstrated in chapter 6. This provided an automated method from which to generate a depth contour map of the optic disc region from stereoscopic fundus image pairs. It has been shown that depth contour maps of the optic disc which are of clinical relevance can be obtained using the methods described in this thesis. The desire to assess the accuracy of the depth estimates, in the absence of ground truth, promoted the development of the image model and error prediction formulae in chapter 7. With these tools the magnitude of the matching error can be predicted, moving some way toward addressing the issue of reliability in topographic measurement.

The second research objective was to study the physical constraints related to stereoscopy of the human fundus, and to study any potential limitations in the measurement and analysis of the optic disc. Chapter 4 addressed this objective. The constraints imposed by the camera geometry, imaging through the pupil of the eye, as well as the small extent of the optic disc area all impose limitations on the accuracy to which stereo information can be extracted. It was shown that the most problematic parameter in non-simultaneous stereo fundus photography was the camera separation or baseline. There are several key parameters which are beyond the control of the clinician, and these include patient movement and change of fixation. In simultaneous stereo fundus photography, optical defocus, image resolution, and image contrast are of key importance.

Another objective of this research was to select a suitable stereo matching algorithm for use on the fundus image data. In chapter 5, a survey of digital stereo matching algorithms was presented and their strengths and weaknesses were explored, specifically as they relate to the fundus image data. Under a set of criteria on which to base the selection of an algorithm, a least-squares matching algorithm was selected; namely Gruen's least-squares matching algorithm [Gruen85]. The algorithm could be extended, as shown by Otto and Chau [Otto89], to automatically generate a dense set of disparity estimates from a stereo fundus image pair.

The fourth objective was to determine the utility of automated stereo measurement of the optic disc by applying the selected matching algorithm to a representative set of fundus image data. Chapter 6 presented the results of applying the stereo matching algorithm to four test data sets. The results indicated that biologically plausible estimates of the optic disc topography could be obtained, but that further work was required to standardise the procedure and to assess the accuracy of the results.

In the absence of ground truth information for the fundus image data, it was natural that the fifth research objective was to use a physical model (for which ground truth was known) to test the performance of the stereo matching algorithm, and to develop a set of mathematical tools which would permit theoretical predictions on the performance of the stereo matching algorithm. The application of the stereo matching algorithm to the plaster model data and the subsequent analysis of the real matching error in chapter 6 indicated that the algorithm could perform well under favourable conditions. The development of a set of mathematical tools for theoretical predictions of cross-correlation error was undertaken in chapter 7. Although these tools require further development, experimental evidence presented in chapter 8 gave an indication that, at least for the fundus image data, the magnitude of cross-correlation error could be predicted and controlled.

The final research objective was to investigate methods for extracting clinical parameters from the topographic measurements obtained from stereo matching fundus image data. This would provide a means by which changes to these parameters could be monitored over time. This objective was not addressed directly in this thesis. However, some of the key points are considered in the following section.

9.3 Tools for Clinical Measurement

In conventional stereo photography of the fundus, a skilled clinician or trained observer manually processes each stereo pair, estimating parameters such as cup-disc ratio and cup volume. If the topography of the optic disc has been computed using stereo matching techniques, the three-dimensional structure of the disc can be represented as a contour map, as a wire-frame model, or as a series of depth profiles (one-dimensional cross-sections). Although the clinician may find locating the optic disc margin an easier task when viewing a contour map or wire-frame model of the optic nerve head, the visual cues from the photographs themselves are no longer used. Even with additional three-dimensional information, the task of reliably locating the optic disc margin (or the edge of the cup) is not a trivial one.

The disc boundary is difficult to extract reliably from the three-dimensional topographic model of the fundus, and it is equally difficult to determine the cup edge position [Peli89]. A method to automate the task of locating the optic disc boundary from a topographic representation of the fundus is of key importance.

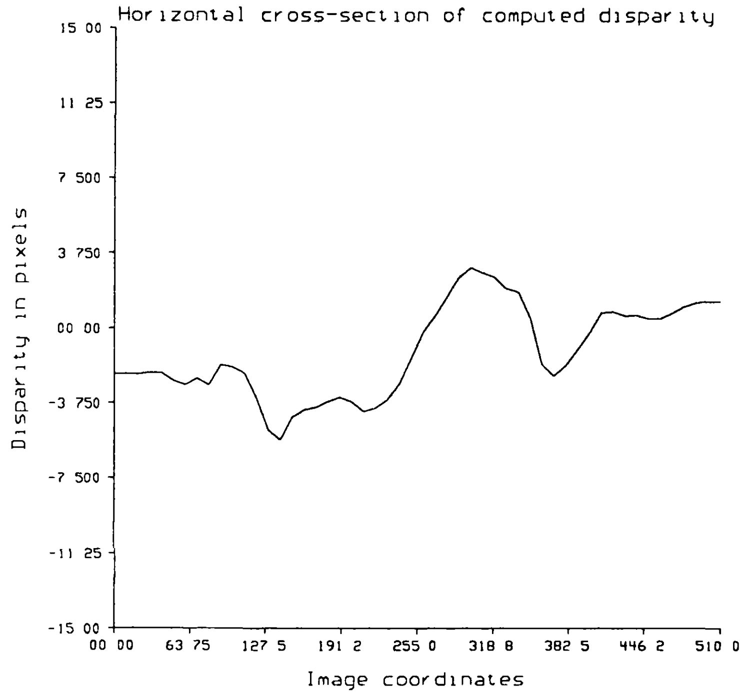


Figure 9.1. A horizontal cross-section of computed depth through the optic disc from Data Set B, highlighting the difficulty in specifying the location of the disc boundary.

It was initially assumed that with a computer generated depth map, locating the optic disc boundary (and the optic cup) would be a relatively easy task. Therefore, most of the effort in this research was directed toward first obtaining an accurate depth map. This assumption, however, was mis-directed. By examining typical contour map results, like those in presented in chapter 6 (figures 6.5 and 6.7), one can begin to see that locating the disc boundary in the general case is itself a major research task. The complexity of the problem was grossly underestimated, and a disproportionate amount of time went toward obtaining accurate topographic measurements, with little time remaining to address the problem of obtaining clinical parameters.

Figure 9.1 shows a one-dimensional cross-section through the contour depth map of Data Set B (as shown in figure 6.4c). Note that for this particular profile, the left-hand boundary of the optic disc is easily sited, but that a procedure for automatic location of the right-hand boundary is not easily specified. Similar representations of disc topography shown in the literature are typically only a simple 1-D profile through a measured optic disc, avoiding the difficulties associated with automatically locating the disc boundary in two dimensions (see, for example, [Johnson79], [Algazi85], and [Lee91a]). More recently, Lee and Brady [Lee91b] have taken a monocular approach to optic disc boundary detection, where the optic disc region is segmented from its surround using morphological image operations. The difficulty with this approach is that the bright central region of the optic nerve head and the scleral ring or bright crescent that often surrounds the optic disc do not necessarily correspond to the location of the disc boundary.

Commercially available systems for optic disc analysis, discussed in chapter 3, require a human operator to mark key boundary points along the optic disc (see [Peli89]). Before an automated procedure to detect the boundary of the optic disc can be developed, a greater understanding of the heuristics by which clinicians now do the task manually is required. With hindsight, the problem of automating what clinicians do is an example of a class of vision problems where machine vision has not yet reached the complexity of human vision.

The ground work of the research in this thesis now provides a suitable stereo matching algorithm for fundus image data, an understanding of the source and magnitude of matching error, and a method for automatic generation of contour maps of the optic disc topography. This is a sound base from which to progress in solving the problems of automated optic disc measurement.

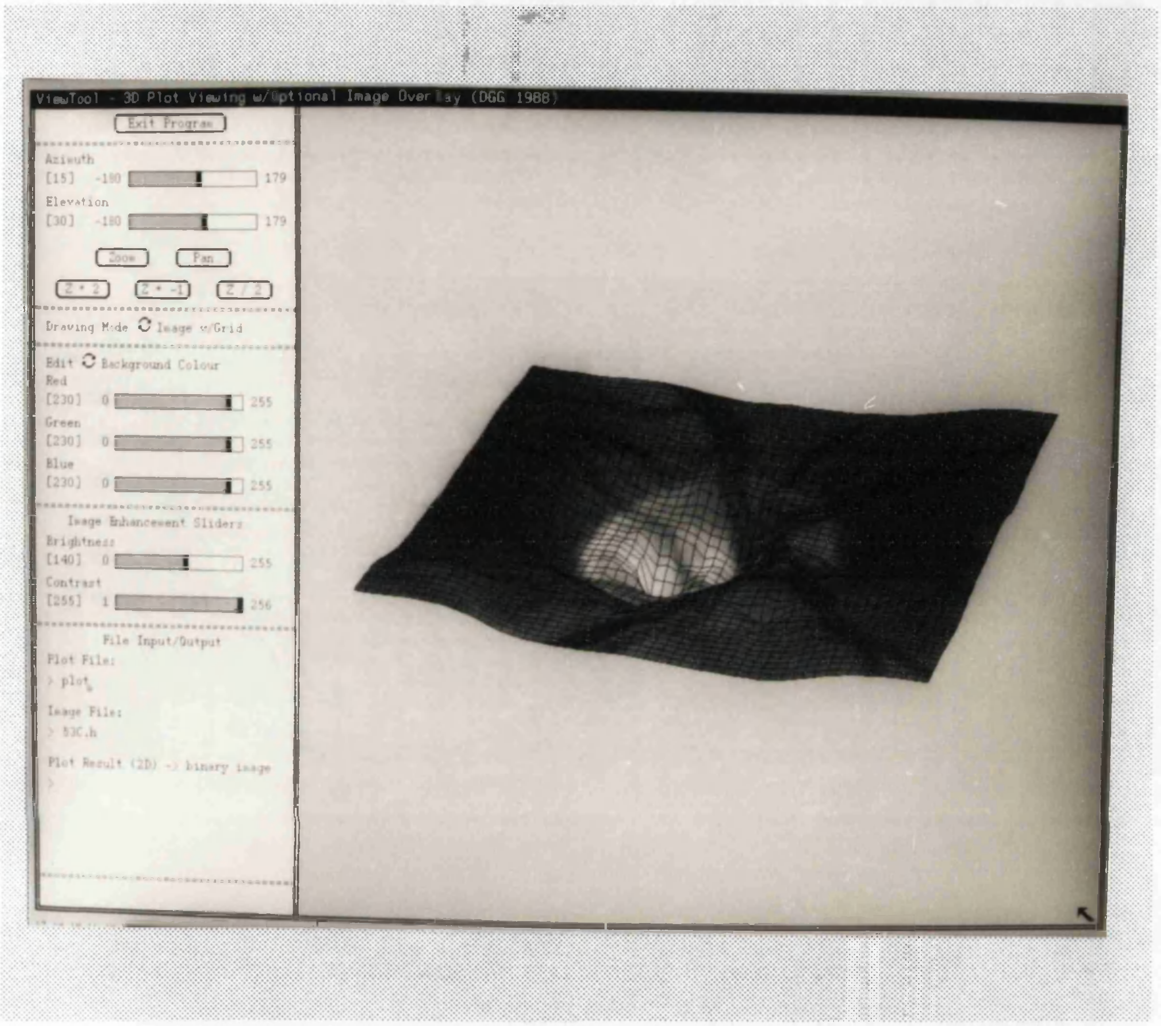
9.4 Suggestions for Further Work

Finally, some suggestions for further work which build on the research presented in this thesis are given. To indicate what may represent the future of stereoscopic analysis of the optic disc, figure 9.2 shows the graphical projection of the fundus image from Data Set C onto the computed surface topography. Appendix I presents some additional results of applying Gruen's algorithm to several simultaneous stereoscopic fundus image pairs obtained using the Donaldson stereo fundus camera (similar to Data Set C). Although these results are not discussed in detail, they give an indication of how well the algorithm can perform on a varied set of fundus image data.

There are a number of suggestions for further work which are a natural extension to the research presented in this thesis. As discussed in the previous section, it is of key importance to investigate methods for automatic disc boundary detection. Demarcating the optic disc boundary is still very much a subjective task, even with the aid of additional three-dimensional information from stereoscopic analysis of optic disc topography. If a robust method can be developed, it will be important to compare clinicians' estimates with those obtained automatically, and to make some assessment of the reliability and reproducibility of the method.

Another area which requires further study is in the improvement of the stereo matching method. For example, it would be useful to investigate the effect that image pre-filtering may have on matching error. Gruen has reported that for certain types of image, a small amount of spatial pre-filtering, such as local averaging, can improve both the convergence and accuracy of the match [Gruen86]. Pre-filtering of the fundus image data prior to stereo matching requires further investigation. Other improvements to the basic Gruen stereo matching algorithm, such as dynamic match window size, parameter weighting, and run-time error analysis, are all interesting areas for further study.

The theoretical and experimental results presented in this thesis indicate that the error prediction formulae developed in chapter 7 might be used to



ViewTool - 3D Plot Viewing w/Optional Image Overlay (DGG 1988)

Azimuth
[15] -180 179
Elevation
[30] -180 179

Drawing Mode ↻ Image w/Grid

Edit ↻ Background Colour
Red
[230] 0 255
Green
[230] 0 255
Blue
[230] 0 255

Image Enhancement Sliders
Brightness
[140] 0 255
Contrast
[255] 1 256

File Input/Output
Plot File:
> plot
Image File:
> 53C.h
Plot Result (2D) -> binary image
>

Figure 9.2. The fundus image from Data Set C projected onto its computed surface topography. The super-imposed grid indicates the points at which disparity estimates have been computed.

predict, assess, and control cross-correlation error. However, these theoretical tools were developed primarily to study the associated error in stereo matching the fundus image data. In order for such tools to become more generally applicable to a wider class of image domains, enhancements to the image model and error prediction formulae would be required. For example, as noted in chapter 8, it would be of great interest to characterise the distribution of matching error, so that it would be possible not only to predict the magnitude (standard deviation) of the matching error, but also to give the associated probability of such an error occurring. With a more rigorous mathematical treatment of the error prediction formulae, it might be possible to derive analytical error probability measures.

One final recommendation for further research is to give more consideration to the latest form of fundus image source, the Scanning Laser Ophthalmoscope (SLO). Fitzke *et al* [Fitzke91] are exploiting the confocal properties of the optics of the SLO in order to obtain optical sections of the optic nerve head. Optical sectioning through small increments of depth could potentially revolutionise the clinical study of the optic nerve head. It would be of great interest to combine or compare the topographic measurements of the optic disc obtained from stereoscopic analysis and optical sectioning.

The work presented in this thesis provides a sound base from which to progress in solving the problems of automated optic disc measurement. Recent advances in the fields of image processing, lasers, optics, computing, and ophthalmology are converging toward a fully automated system for diagnostic evaluation of the human fundus.

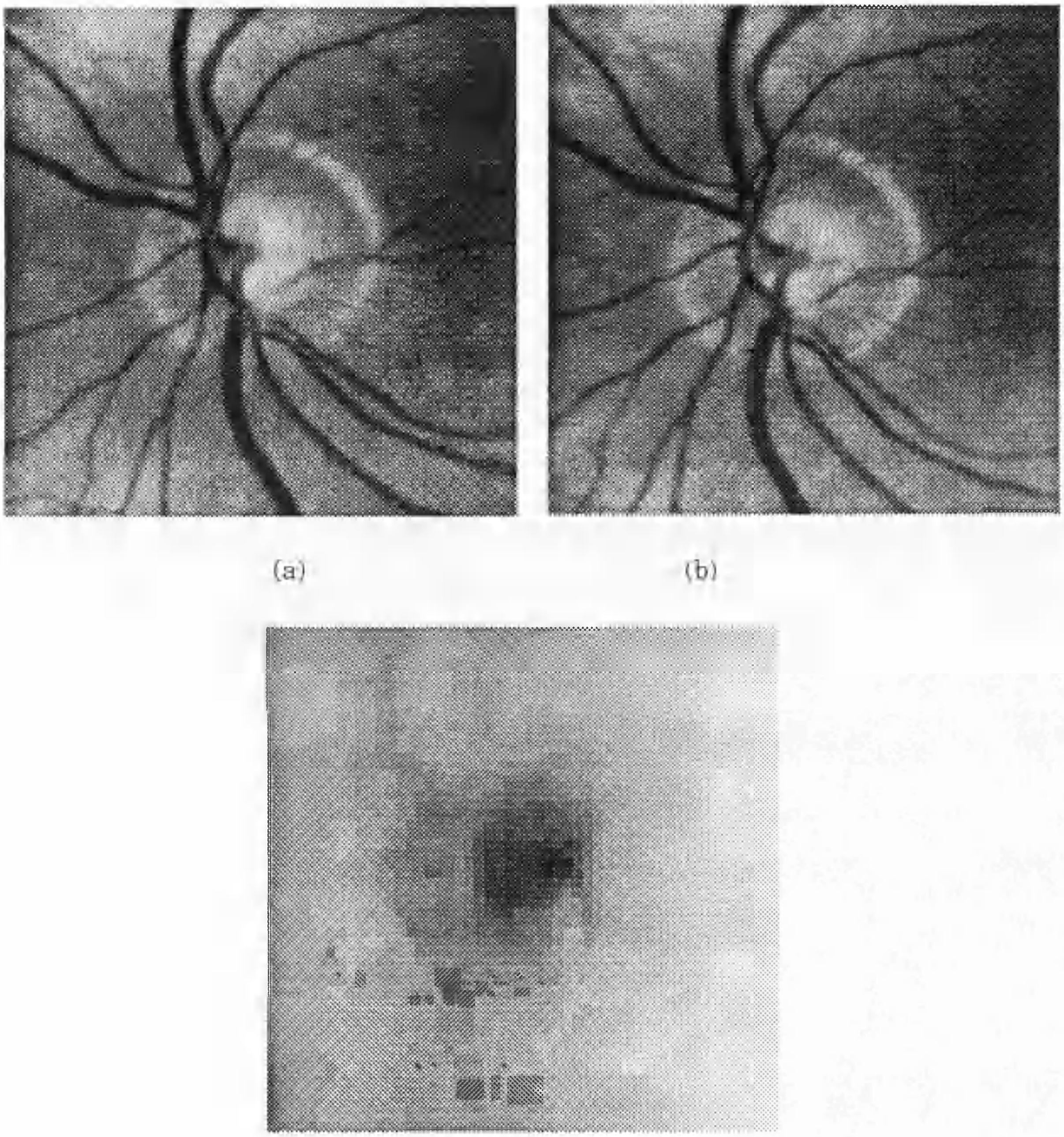
Appendix I
Some Further Results

Presented here are some additional results of applying Gruen's algorithm to several simultaneous stereoscopic fundus image pairs obtained using the Donalson stereo fundus camera (similar to Data Set C). Although these results are not discussed in detail, they provide additional illustrations of the computed topography of the optic disc.

The methods employed to obtain each digitised image were the same as those used for Data Sets B and C, discussed in chapter 6, section 6.3.1. For each stereo image pair, the Gruen-Otto algorithm was applied after manually selecting a small set of seed match points. The selection of the match window size was determined by applying the theoretical techniques presented in chapters 7 and 8. Figures 8.6 through 8.10 indicated that an average match window size of 21×21 pixels would yield better than 0.5 pixel r.m.s error in match location for a representative set of fundus image regions, assuming that the signal-to-noise ratio is better than 1:1. All the results presented here were matched using a 21×21 match window size.

The Gruen-Otto algorithm is a region-growing, stereo matching algorithm which results in a grid-like coverage of the left-hand image (see chapter 5, section 5.6.3). This set of stereo images was matched with a neighbourhood spacing of 5 pixels, and thus the results are similar to those presented for Data Sets B and C. The time to process each image pair was approximately 1 hour running on a Sun Microsystems SparcStation 4/370.

The presentation of the stereo match results for each image pair is similar to that used in chapter 6 in figures 6.4 through 6.9. Additionally, the graphical projection of the left-hand image from each pair onto the corresponding computed surface topography is shown in order to better illustrate the results.

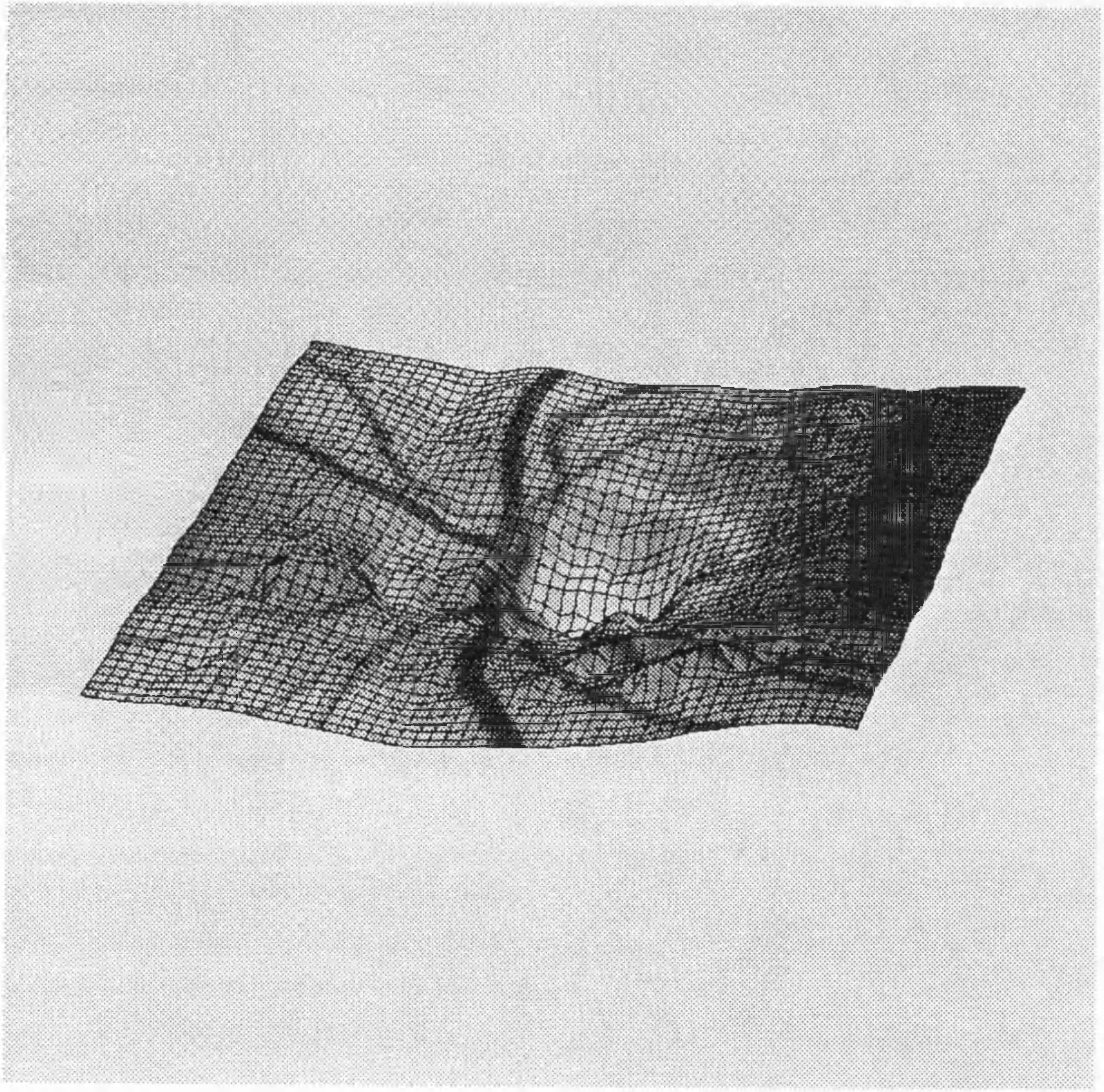


(a)

(b)

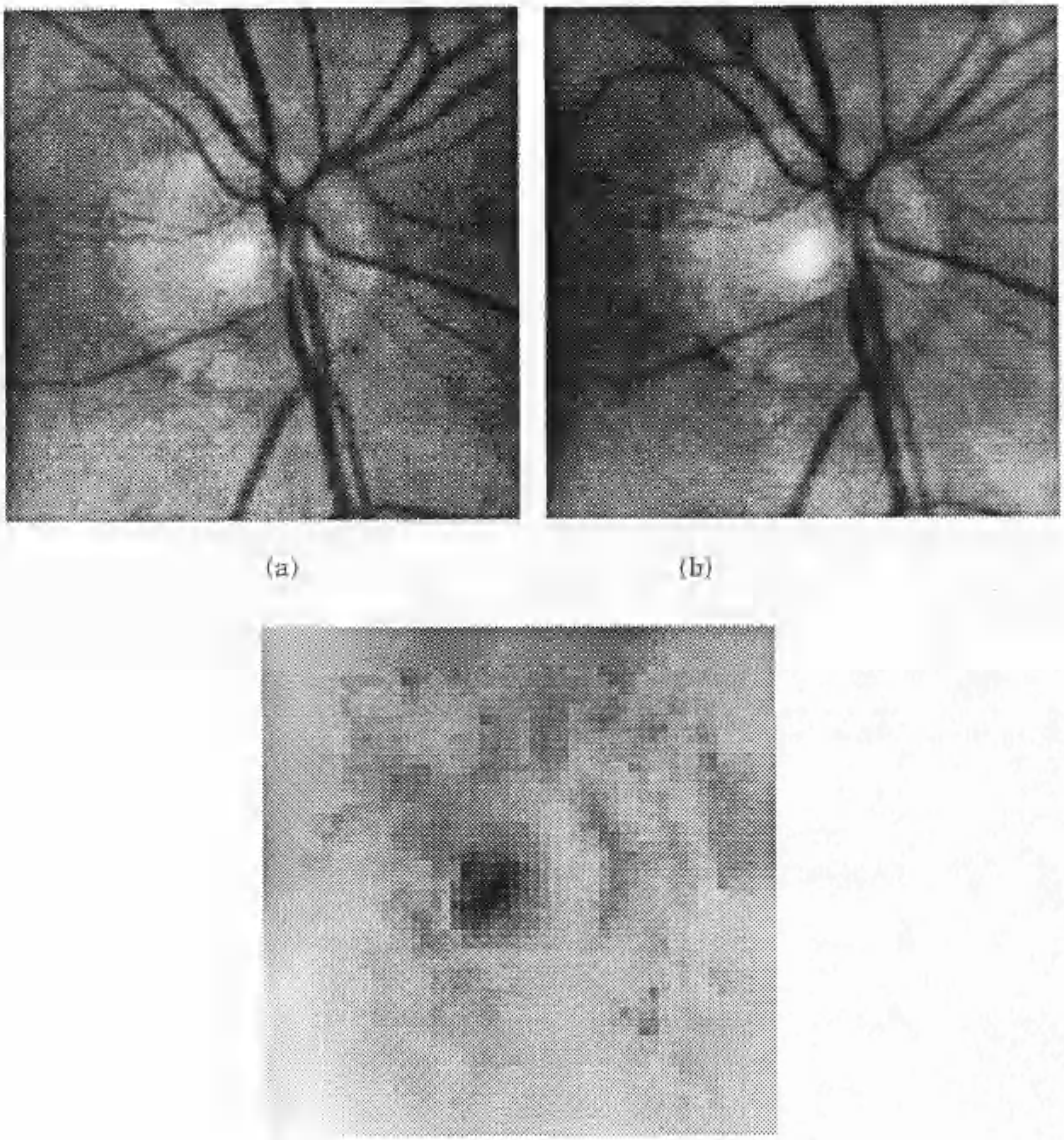
(c)

Figure A.1. Image pair #1: stereo match results of a Donaldson stereo fundus image pair, showing (a) original left-hand image, (b) original right-hand image, and (c) grey-scale (depth-shaded) image of disparity values. (continued...)



(d)

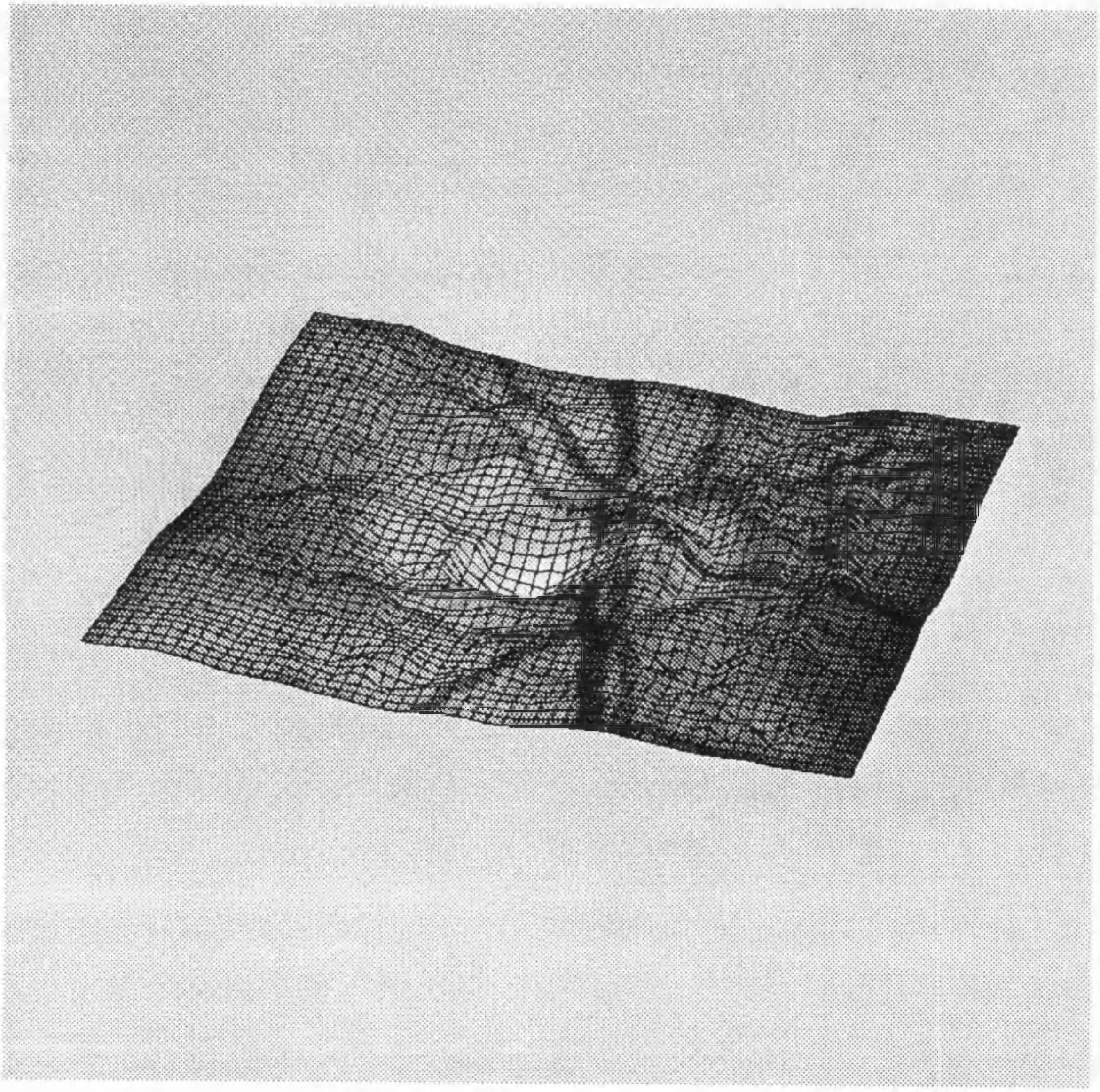
Figure A.1 (cont.). (d) the left-hand image in (a) projected onto its computed surface topography.


(a)

(b)

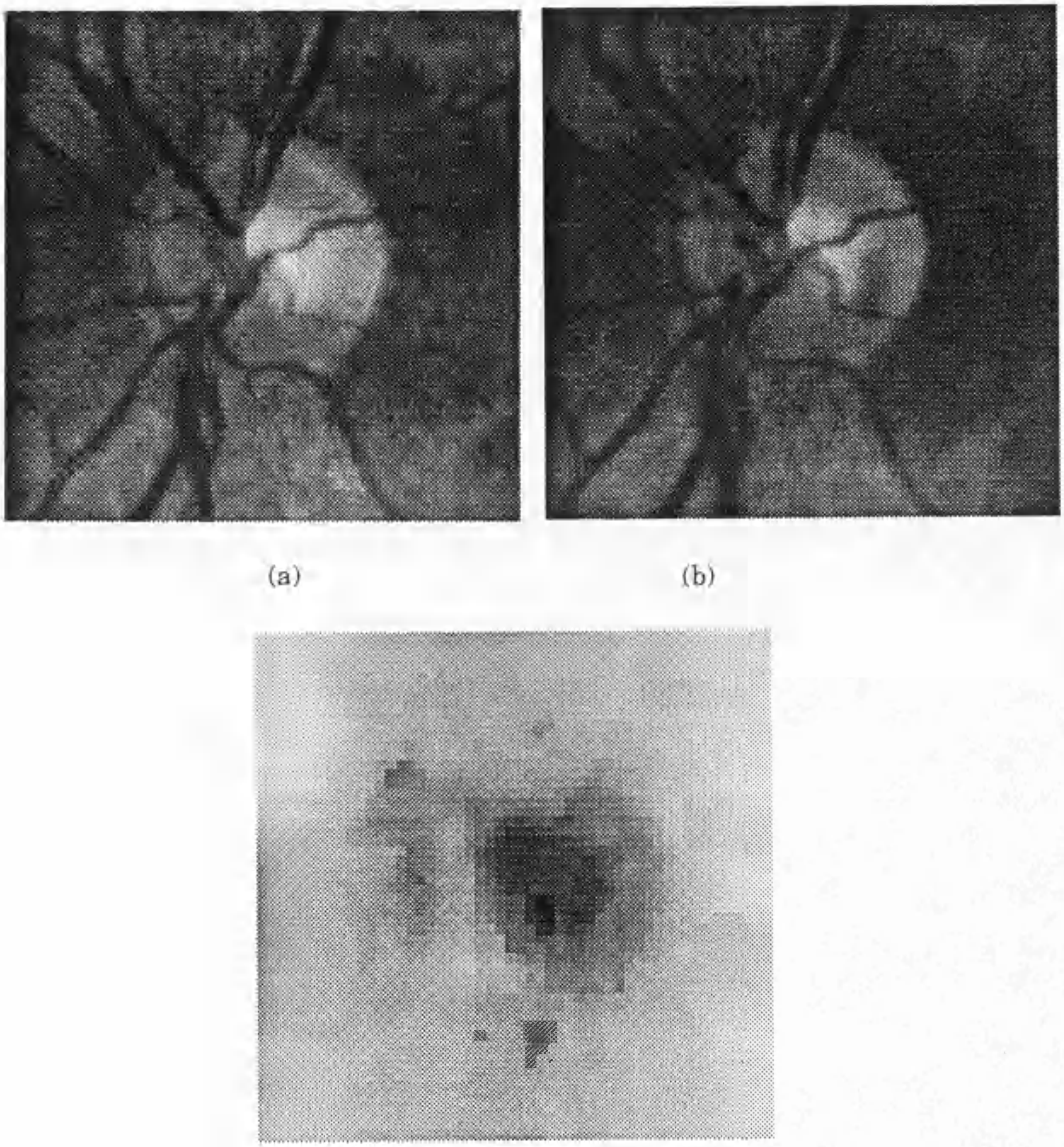
(c)

Figure A.2. Image pair #2: stereo match results of a Donaldson stereo fundus image pair, showing (a) original left-hand image, (b) original right-hand image, and (c) grey-scale (depth-shaded) image of disparity values. (continued...)



(d)

Figure A.2 (cont.). (d) the left-hand image in (a) projected onto its computed surface topography.

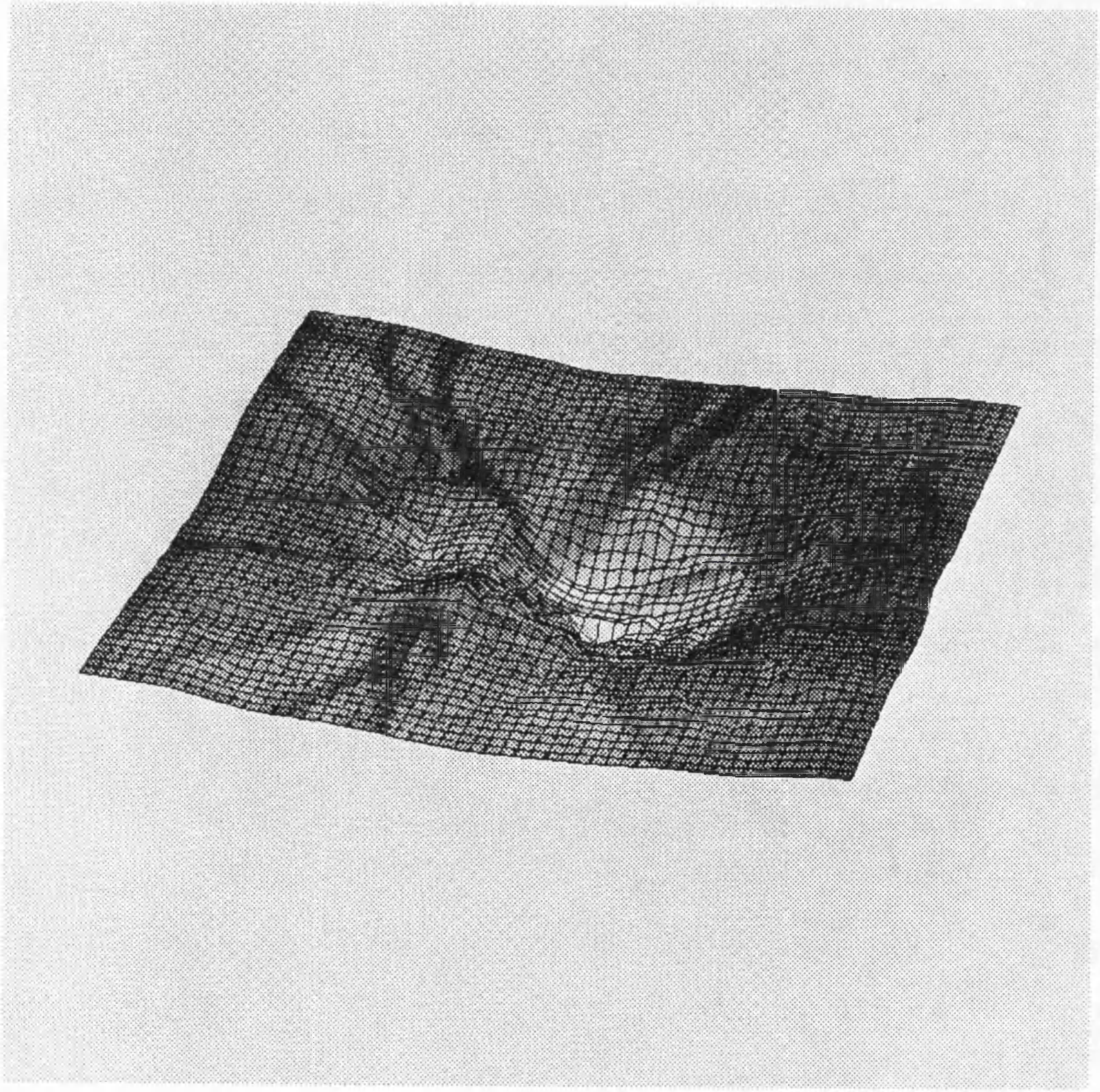


(a)

(b)

(c)

Figure A.3. Image pair #3: stereo match results of a Donaldson stereo fundus image pair, showing (a) original left-hand image, (b) original right-hand image, and (c) grey-scale (depth-shaded) image of disparity values. (continued...)



(d)

Figure A.3 (cont.). (d) the left-hand image in (a) projected onto its computed surface topography.

Appendix II
Gruen Program Source Code

This appendix contains the 'C' source code listing of the Gruen program, together with a brief description of how the program can be used (in the form of a UNIX manual page). This implementation of Gruen's algorithm [Gruen85,86] and the region-growing extensions of Otto and Chau [Otto89] follows the pseudo-code description of the algorithm presented in chapter 6.

NAME

Gruen - stereo matching using Gruen's method with sheet growing extension

SYNOPSIS

Gruen [**-p** *patchradius*] [**-t** *xy_threshold*] [**-s** *shaping_threshold*] [**-i** *maxiters*] [**-g** *startx nx incx starty ny incy*] [**-e** *eigen_threshold*] [**-c** *conrad*] [**-x** *radiometric_opt*] [**-d** *distort_threshold*] [**-m** *maxrs*] [**-n** *minrs*] [**-u** *maxrg*] [**-v** *minrg*] [**-b**] *image1 image2*

DESCRIPTION

Gruen uses Gruen's adaptive least squares correlation algorithm and a sheet growing algorithm to produce an array of disparity points for *image1* and *image2*. A number of seed points are necessary as the start off points for the algorithm. The seed points are read from standard input. The algorithm for sheet growing is based on a "best" first growing algorithm. The match with current best "score" is used to predict an initial guess for the match of its neighbours. The "score" used is determined by the x and y terms in the covariance matrix of the parameters. (See the section on "precision" in Gruen 1985.) The largest eigenvalue of this submatrix defines the score (thus, the nearer the eigenvalue to 0, the better the match). A priority queue based on this score is used to implement this strategy. Hence, the matching "grows" out from the best matched points first.

The output format is the same as the input format and is as follows:-

field1 field2 ..... field13

where,

field1,field2,field3,field4
field1 and *field2* are the sample,line of the left hand image. Similarly *field3* and *field4* are the sample,line of the corresponding point in the right hand image.

*field5*This is the disparity measured for the match, determined from *fields1-4.*

field6,field7,field8,field9
These are the 4 shaping parameters corresponding to the affine matrix elements s11, s12, s21 and s22 respectively.

field10
This is the additive radiometric parameter where 0.0 means that there is no additive shift.

field11
This is the multiplicative radiometric parameter where 1.0 means that there is no multiplicative gain.

field12
This is an integer field which is used to denote how many iterations the algorithm required before reaching the termination thresholds.

field13
This field is the largest eigenvalue for the covariance matrix mentioned above. The larger the number, the less confident the stereo match.

OPTIONS

-p *patch_radius*

Modify the default patch radius. The patch will be square, with each side of length 2×*patch_radius*+1. The default radius is 12.

-t *xy_threshold*

Modify the default xy threshold, which specifies the termination threshold for the x and y parameters — when the changes are less than this threshold (in magnitude) the iteration terminates. The default value 0.05.

-s *shaping_threshold*

Modify the default shaping paramter threshold, which specifies the termination threshold for the affine shaping parameters — when the changes are less than this threshold (in magnitude) the iteration terminates. The default value 0.05.

-i *maxiters*

Modify the default maximum no. of iterations. The default is 16. There is a limit on the maximum number of iterations because Gruen's algorithm may not converge.

-g *startx nx incx starty ny incy*

These parameters will specify the area of the image in which the sheet growing algorithm should be working. The default for startx and starty are (0,0) i.e. top left and incx and incy are 5 i.e. sample every 5 pixels. The nx and ny are set to size of input image / 5.

-e *eigen_threshold*

Modify the default eigenvalue threshold, which specifies the threshold used on the largest eigenvalue of the x-y covariance matrix. If the threshold is 0.0 (default), then this constraint is not used. Otherwise, any point with the eigenvalue less than this threshold will be considered as valid.

-c *convergence radius*

Modify the convergence radius threshold, which sets the threshold limit of the difference in the calculated disparity (after gruen) and the predicted disparity (from the sheet growing). The default is 3.0 pixels, i.e. it will allow a shift of 3.0 pixel and the point is still valid.

-x *radiometric option*

Modify the default radiomteric option used as match parameters. The options are 0) no radiometric parameters, 1) multiplicative radiometric gain only, 2) additive radiometric shift only, 3) both additive and multiplicative radiometric parameters. The default is option 3.

-d *distort_threshold*

Modify the default distortion threshold, which specifies the threshold used on the standard deviation of the 2 rotational parameters in the distortion matrix. The default is 0.1. Any point with the standard deviation less than this threshold will be considered as valid.

-m *maxrs*

Modify the default threshold used on the additive radiometric parameter of the final match. This sets a limit on the maximum radiometric shift allowed for each point. The default is 128.0.

-n *minrs*
Modify the default threshold used on the additive radiometric parameter of the final match. This sets a limit on the minimum radiometric shift allowed for each point. The default is -128.0.

-u *maxrg*
Modify the default threshold used on the multiplicative radiometric parameter of the final match. This sets a limit on the maximum radiometric gain allowed for each point. The default is 4.0.

-v *minrg*
Modify the default threshold used on the multiplicative radiometric parameter of the final match. This sets a limit on the minimum radiometric gain allowed for each point. The default is 0.25.

-b Various diagnostic messages are output for debugging and testing purposes. The default is no debugging.

DIAGNOSTICS

Assorted difficulties and errors will produce messages on standard error. These are intended to be self-explanatory.

BUGS

In this implementation, the image patches must be square, with each point weighted equally.

```
+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++

/*
**      GRUEN SOURCE CODE LISTING
**
**      This is the entire source code listing for the Gruen stereo matching
**      program. This code was designed and written solely by the author.
**
**      Author: David G. Greenwood
**      Date:   June 1991
*/

+++++++++++++++++++++++++++++++ File: gruen.h ++++++++++++++++++++++++++++++++++

/*
** gruen.h - generic include file
**
**      Header file with globally declared variable types
*/

#ifndef GRUEN_H
#define GRUEN_H

#include <stdio.h>
#include <math.h>

#define REAL double

typedef struct array {
        REAL **start;
        int numrow;
        int numcol;
} ARRAY, *ARRAYPTR;

#define ACCESS(_arrayptr, _row, _col) (_arrayptr)->start[_col][_row]

typedef short BOOLEAN;

typedef struct point {
        REAL xpos;
        REAL ypos;
} POINT, *POINTPTR;

typedef struct im {
        unsigned char *data;
        int width;
        int height;
} IMAGE, *IMAGEPTR;

typedef struct list {
                POINT left;
                POINT right;
                REAL s1;
                REAL s2;
                REAL s3;
                REAL s4;
                REAL rs_plus;
                REAL rs_mult;
                BOOLEAN seed;
                REAL score;
                struct list *ptrleft;
                struct list *ptrright;
} LIST, *LISTPTR;

#define PATCH_ONLY !0
#define SEED    !0
#define NON_SEED 0

#define UNMATCHED 0
#define MATCHED 1
#define WORKING 2

#define NUM_PARAMS 8    /* # of patch parameters */

/* thresholds */
REAL T_SP;      /* shaping parameter threshold */
REAL T_XY;      /* translation parameter threshold */
REAL T_RS;      /* radiometric parameter threshold */
REAL CONRAD;    /* convergence radius */
REAL MAX_RS, MIN_RS;
REAL MAX_RG, MIN_RG;
int  RS_ON;

int PATCHRADIUS;
int MAXITERS;
int SEED_MAXITERS;
REAL STARTX, STARTY;
int NUMX, INCX;
int NUMY, INCY;
REAL VARI_THRESHOLD;
REAL EIGN_THRESHOLD;
REAL DIST_THRESHOLD;
BOOLEAN Debug;

#endif GRUEN_H

++++++++++++++++++++++++++++++ File: main.c ++++++++++++++++++++++++++++++++++

/*
** main.c - main program
**
**      This module contains the main flow control to run
**      Gruen's algorithm with region-growing extensions.
*/
```

```
#include "gruen.h"

#define sqr(_t) ((_t)*(_t))

/* extern ARRAYS declared from memory.c */
extern ARRAY identity_matrix;
extern ARRAY shape_matrix;
extern ARRAY t_shape_matrix;
extern ARRAY threshold_values;

main(argc, argv)
  int argc;
  char **argv;
{
  FILE    *fd1, *fd2, *fderr, *file_must_open(), *fopen();
  IMAGE   left_image, right_image;
  POINT   left_origin, right_origin;
  POINT   l_origin, r_origin, right_pred;
  int     dir, arg, iters;
  REAL    rs_plus, rs_mult, vari, eig;
  REAL    t_rs_plus, t_rs_mult;
  REAL    eigen();
  BOOLEAN seed;


        if((arg=read_commandline(argc, argv))==0)
                print_format(argv[0]);

        /* read image headers  & images */
        fd1 = file_must_open(argv[arg]);
        fd2 = file_must_open(argv[arg+1]);

        read_image(fd1, &left_image);
        read_image(fd2, &right_image);
        close(fd1); close(fd2);

        /* default grid/neighbour spacing */
        if(STARTX == -1.0) STARTX = 0.0;
        if(STARTY == -1.0) STARTY = 0.0;
        if(INCX == -1) INCX = 5;
        if(INCY == -1) INCY = 5;
        if(NUMX == -1) NUMX = left_image.width/INCX;
        if(NUMY == -1) NUMY = left_image.height/INCY;

        /* initialize & memory allocs */
        init_lists(STARTX, NUMX, INCX, STARTY, NUMY, INCY);
        init_matrices(PATCHRADIUS);
        print_header(argv[arg], argv[arg+1]);

        while (read_points(&left_origin, &right_origin, &shape_matrix,
                &identity_matrix, &rs_plus, &rs_mult) != EOF)
```

```
{
    iters = gruens(&left_image, &right_image, &shape_matrix, PATCHRADIUS,
        &left_origin, &right_origin, MAXITERS, &threshold_values,
        &rs_plus, &rs_mult);
    eig = eigen(PATCHRADIUS);

    if(check_constraints(&left_origin,&right_origin,&right_origin,
        &shape_matrix,eig,rs_plus,rs_mult,iters)) {
        store_list(&left_origin, &right_origin, &shape_matrix, eig,
                rs_plus, rs_mult, SEED);
        matched(&left_origin);
        print_results(stdout, &shape_matrix, &left_origin, &right_origin,
                iters, rs_plus, rs_mult, PATCHRADIUS, vari, eig);
    }
}


while(get_list(&l_origin, &r_origin, &t_shape_matrix, &t_rs_plus,
        &t_rs_mult, &seed))
{
    if(INCX == 0 && INCY == 0) break;

    /*
    ** temps are needed for shape_matrix, rs_plus and rs_mult because
    ** each attempt at a neighbour will corrupt them.
    */

    for(dir=0; dir < 4; dir++) {

        /* restore origin, shape_mtrix, rs_plus and rs_mult */
        left_origin = l_origin;
        right_origin = r_origin;
        ACCESS(&shape_matrix,0,0) = ACCESS(&t_shape_matrix,0,0);
        ACCESS(&shape_matrix,0,1) = ACCESS(&t_shape_matrix,0,1);
        ACCESS(&shape_matrix,1,0) = ACCESS(&t_shape_matrix,1,0);
        ACCESS(&shape_matrix,1,1) = ACCESS(&t_shape_matrix,1,1);
        rs_plus = t_rs_plus; /* 0.0 */
        rs_mult = t_rs_mult; /* 1.0 */

        /* try and predict a new match */
        if(predict_match(&left_origin, &right_origin, &shape_matrix,
                &right_pred, &left_image, &right_image, seed, dir))
            continue;

        /* run this match */
        iters = gruens(&left_image,&right_image,&shape_matrix,PATCHRADIUS,
                &left_origin,&right_origin,MAXITERS,&threshold_values,
                &rs_plus, &rs_mult);
        eig = eigen(PATCHRADIUS);

        /* check results */
        if(check_constraints(&left_origin,&right_origin,&right_pred,
```

```
                &shape_matrix,eig,rs_plus,rs_mult,iters)) {
                store_list(&left_origin, &right_origin, &shape_matrix,
                    eig, rs_plus, rs_mult, NON_SEED);
                matched(&left_origin);
                print_results(stdout,&shape_matrix,&left_origin,&right_origin,
                    iters, rs_plus, rs_mult, PATCHRADIUS, vari, eig);
            }
        }
    }

    free_matrices();
    exit(0);
}

+++++++++++++++++++++++++++++ File: gruen.c ++++++++++++++++++++++++++++++++

/*
** gruen.c - gruen algorithm control routines
**
**      This module contains the main gruen routine plus support routines.
**
**      The algorithm is as follows:
**
**      resample left hand patch
**      iterate
**              resample right hand patch
**              construct design matrix A and vector l
**              solve A alpha = l for alpha
**              update transformation parameters
**      until (change < termination or maxiters)
*/
#include "gruen.h"

/* extern global variables */
extern ARRAY identity_matrix;
extern ARRAY left_patch,right_patch;
extern ARRAY right_xd_patch;
extern ARRAY right_yd_patch;
extern ARRAY design;
extern ARRAY rhs;
extern ARRAY larger_patch;
extern ARRAY X;


/*
** gruens:
**
**      main gruen routine - as specified above.
*/

gruens(left_image, right_image, shaping_matrix, patchradius, left_origin,
        right_origin, maxiters, terminate_conds, rs_plus, rs_mult)
```

```
  IMAGEPTR left_image, right_image;
  ARRAYPTR shaping_matrix;
  int patchradius;
  POINTPTR left_origin, right_origin;
  int maxiters;
  ARRAYPTR terminate_conds;
  REAL *rs_plus, *rs_mult;
{
  BOOLEAN done, terminate();
  int no_iters;

        no_iters = 0;
        if (sample_patch(left_image, left_origin, &identity_matrix, &left_patch,
                NULL, NULL, patchradius, PATCH_ONLY) == -1)
                return(0);

        do {
        if (sample_patch(right_image, right_origin, shaping_matrix, &right_patch,
                &right_xd_patch,&right_yd_patch,patchradius,!PATCH_ONLY) == -1) {
                done = 1;
                no_iters = -no_iters;
        } else {
                construct_design_and_rhs(patchradius, rs_plus, rs_mult);
                solve(&design,&rhs, patchradius);
                update_params(shaping_matrix, right_origin, rs_plus, rs_mult);
                done = terminate(terminate_conds);
                no_iters++;
               }
        } while (!done && no_iters < maxiters);
        if (no_iters == maxiters) no_iters = -no_iters;
        return(no_iters);
}


/*
** sample_patch:
**
**      sample a patch of size patchradius from image, and compute derivatives
*/
sample_patch(image,origin,transformation,result,xd,yd,patchradius,patch_only)
  IMAGEPTR image;
  POINTPTR origin;
  ARRAYPTR transformation;
  ARRAYPTR result;
  ARRAYPTR xd, yd;
  int patchradius;
  BOOLEAN patch_only;
{
  int xp, yp, size;
  int width, height;
  REAL s1, s2, s3, s4;
  REAL ox, oy;
```

```
  REAL interpolate(), x_derivative(), y_derivative();

        width = image->width;
        height = image->height;
        size = patchradius;


        s1 = ACCESS(transformation,0,0);
        s2 = ACCESS(transformation,0,1);
        s3 = ACCESS(transformation,1,0);
        s4 = ACCESS(transformation,1,1);
        ox = origin->xpos;
        oy = origin->ypos;

        for (yp = (-size); yp <= size; yp++)
            for (xp = (-size); xp <= size; xp++) {
                REAL xi,yi;

                xi = s1 * (REAL)(xp) + s2 * (REAL)(yp) + ox;
                yi = s3 * (REAL)(xp) + s4 * (REAL)(yp) + oy;

                if (!(xi >= 0 && yi >= 0 && xi < (width-1) && yi < (height-1))) {
                        /* printf("patch out of bounds: xi=%f  yi=%f\n",xi,yi);*/
                        return(-1);
                }

                /* note: remember access is by (row, col) , ie. (y,x) */

                ACCESS(result,(yp+patchradius),(xp+patchradius)) =
                                interpolate(image,xi,yi);

                if (!patch_only) {

                        ACCESS(xd, (yp+patchradius),(xp+patchradius)) =
                                        x_derivative(image, xi, yi);

                        ACCESS(yd, (yp+patchradius),(xp+patchradius)) =
                                        y_derivative(image, xi, yi);
                }
            }

        return(0);
}

/*
** x_derivative:
**
**      x derivatives image at (REAL) (x,y) using forward difference.
*/
REAL x_derivative(image, x, y)
  IMAGEPTR image;
  REAL x, y;
```

```
{
  int width;
  REAL temp, fracx, fracy;
  unsigned char *ptr;

        width = image->width;
        ptr = (unsigned char *)image->data + ((int)y)*width + ((int)x);
        temp = (REAL)(*(ptr+1)) - (REAL)(*ptr);
        return(temp);
}


/*
** y_derivative:
**
**      y derivatives of image at (REAL) (x,y) using forward difference.
*/
REAL y_derivative(image, x, y)
  IMAGEPTR image;
  REAL x, y;
{
  int width;
  REAL temp, fracx, fracy;
  unsigned char *ptr;

        width = image->width;
        ptr = (unsigned char *)image->data + ((int)y)*width + ((int)x);
        temp = (REAL)(*(ptr+width)) - (REAL)(*ptr);
        return(temp);
}


/*
** interpolate:
**
**      bilinear interpolation of the image at (x,y)
*/
REAL interpolate(image, x, y)
  IMAGEPTR image;
  REAL x, y;
{
  int ix, iy;
  REAL fx;
  REAL t1, t2;
  REAL   ret;
  unsigned char *ptr;

        ix = (int)x;
        iy = (int)y;
        fx = x - ix;

        ptr = (image->data) + iy*(image->width) + ix;
```

```
        t1 = (REAL)(*ptr) + fx*((REAL)(*(ptr+1)) - (REAL)(*ptr));
        ptr += (image->width);
        t2 = (REAL)(*ptr) + fx*((REAL)(*(ptr+1)) - (REAL)(*ptr));
        ret = t1 + (y - (REAL)(iy))*(t2 - t1);

        return(ret);
}


/*
** construct_design_and_rhs:
**
**      set up design matrix and the rhs vector
*/
construct_design_and_rhs(patchradius, rs_plus, rs_mult)
  int patchradius;
  REAL *rs_plus, *rs_mult;
{
  int i, j, offset, size;
  REAL x_deriv, y_deriv;
  REAL xpos, ypos;
  REAL trm;

        size = 2 * patchradius + 1;
        offset = 0;
        trm = *rs_mult;

        for (i = 0; i < size; i++)
                for (j = 0; j < size; j++) {

                        x_deriv = ACCESS(&right_xd_patch,j,i);
                        y_deriv = ACCESS(&right_yd_patch,j,i);

                        xpos = (REAL)(i-patchradius);
                        ypos = (REAL)(j-patchradius);

                        ACCESS(&design,offset,0) = trm * x_deriv;
                        ACCESS(&design,offset,1) = trm * xpos * x_deriv;
                        ACCESS(&design,offset,2) = trm * ypos * x_deriv;
                        ACCESS(&design,offset,3) = trm * y_deriv;
                        ACCESS(&design,offset,4) = trm * xpos * y_deriv;
                        ACCESS(&design,offset,5) = trm * ypos * y_deriv;

                        if(RS_ON == 0) /* none */
                        ACCESS(&rhs,offset,0) = ACCESS(&left_patch,j,i)
                                        - ACCESS(&right_patch,j,i);

                        if(RS_ON == 1) /* mult only */
                        {
                        ACCESS(&design,offset,6) = ACCESS(&right_patch,j,i);
                        ACCESS(&rhs,offset,0) = ACCESS(&left_patch,j,i);
                        }
```

```
                        if(RS_ON == 2) /* plus only */
                        {
                        ACCESS(&design,offset,6) = 1.0;
                        ACCESS(&rhs,offset,0) = ACCESS(&left_patch,j,i)
                                        - ACCESS(&right_patch,j,i);
                        }

                        if(RS_ON == 3) /* both */
                        {
                        ACCESS(&design,offset,6) = 1.0;
                        ACCESS(&design,offset,7) = ACCESS(&right_patch,j,i);
                        ACCESS(&rhs,offset,0) = ACCESS(&left_patch,j,i);
                        }
                        offset++;
                }
}


/*
** terminate:
**
**      Compare the conditions to see if terminate
*/
BOOLEAN terminate(conditions)
  ARRAYPTR conditions;
{
  BOOLEAN stat;
  REAL fabs();

        stat = !0;
        stat &= ( ACCESS(conditions,0,0) > fabs(ACCESS(&X,0,0)) );
        stat &= ( ACCESS(conditions,1,0) > fabs(ACCESS(&X,1,0)) );
        stat &= ( ACCESS(conditions,2,0) > fabs(ACCESS(&X,2,0)) );
        stat &= ( ACCESS(conditions,3,0) > fabs(ACCESS(&X,3,0)) );
        stat &= ( ACCESS(conditions,4,0) > fabs(ACCESS(&X,4,0)) );
        stat &= ( ACCESS(conditions,5,0) > fabs(ACCESS(&X,5,0)) );
        return(stat);
}


/*
** update_params:
**
**      update shaping parameters, point origin, and radiometric shift
*/
update_params(shaping, origin, rs_plus, rs_mult)
  ARRAYPTR shaping;
  POINTPTR origin;
  REAL *rs_plus, *rs_mult;
{
        ACCESS(shaping,0,0) += ACCESS(&X,1,0);
```

```
        ACCESS(shaping,0,1) += ACCESS(&X,2,0);
        ACCESS(shaping,1,0) += ACCESS(&X,4,0);
        ACCESS(shaping,1,1) += ACCESS(&X,5,0);

        origin->xpos += ACCESS(&X,0,0);
        origin->ypos += ACCESS(&X,3,0);

        if(RS_ON == 1)
                *rs_mult = ACCESS(&X,6,0);

        if(RS_ON == 2)
                *rs_plus = ACCESS(&X,6,0);
        if(RS_ON == 3)
        {
                *rs_plus = ACCESS(&X,6,0);
                *rs_mult = ACCESS(&X,7,0);
        }
        return(0);
}

++++++++++++++++++++++++++++++ File: solve.c ++++++++++++++++++++++++++++++++

/*
** solve.c - numerical solution to Ax = b.
**
**      This module contains the routines to compute the solution
**      to Ax = b. First compute inv(AtA) and (At)b, and then
**      solve by x = inv(AtA)*(At)b.
*/

#include "gruen.h"

#define sqr(_t)  ((_t)*(_t))

extern ARRAY C; /* workspace for AtA */
extern ARRAY D; /* workspace for At*b */
extern ARRAY Y; /* workspace */
extern ARRAY X; /* solution vector */
extern ARRAY GJ;/* Gauss-Jordan elimination workspace */

/*
** solve:
*/
solve(A, B, patchradius)
  ARRAYPTR A;
  ARRAYPTR B;
  int patchradius;
{
  int i,j,k;
  int rows, cols;
  REAL sum;
  int d;

        d = 2 * patchradius + 1;
        rows = d * d;           /* number of equations */
        cols = A->numcol;       /* number of unknowns */

        /* compute C = AtA */
        for(j=0; j < cols;j++) {
                for(i=0; i < cols; i++) {
                        sum = 0.0;
                        for(k=0; k < rows;k++)
                                sum += ( ACCESS(A,k,i) * ACCESS(A,k,j) );

                        ACCESS(&C,i,j) = sum;
                }
        }

        /* compute inverse of C */
        GJ_elimination(&C, &GJ);

        /* compute d = Atb */
        for(j=0; j < cols; j++) {
                ACCESS(&D,j,0) = 0.0;
                for(i=0; i < rows; i++)
                        ACCESS(&D,j,0) += ( ACCESS(A,i,j) * ACCESS(B,i,0) );
        }

        /* compute x = (AtA)inv * (Atb) */
        for(i=0; i < cols; i++) {
                sum = 0.0;
                for(k=0; k < cols;k++)
                        sum += ( ACCESS(&GJ,i,k+cols) * ACCESS(&D,k,0) );

                ACCESS(&X,i,0) = sum;
        }
}

++++++++++++++++++++++++++++++ File: pred_constr.c ++++++++++++++++++++++++++

/*
** pred_constr.c
**
**      This module contains the routines for predicting
**      a neighbouring match, and verifying the result
**      against the known constraints:
**          predict_match() and check_constraints().
*/

#include "gruen.h"

/*
** predict_match() - predict a neigbouring pixel match
*/
```

```
predict_match(left,right,shape_matrix,pred_right,leftim,rightim,seed,direction)
  POINTPTR left,right,pred_right;
  ARRAYPTR shape_matrix;
  IMAGEPTR leftim, rightim;
  BOOLEAN seed;
  int direction;
{
  POINT p_inc,p_temp;

        /* See if the point is one of the grid point */
        if (seed == NON_SEED) {
                p_inc.xpos = (left->xpos - STARTX) / (REAL)INCX;
                p_inc.ypos = (left->ypos - STARTY) / (REAL)INCY;
                switch (direction) {
                        case 0: p_inc.ypos -= 1; break;
                        case 1: p_inc.xpos += 1; break;
                        case 2: p_inc.ypos += 1; break;
                        case 3: p_inc.xpos -= 1; break;
                }
        } else { /* SEED point */
                p_inc.xpos = floor((left->xpos - STARTX) / (REAL)INCX);
                p_inc.ypos = floor((left->ypos - STARTY) / (REAL)INCY);
                switch (direction) {
                        case 0: p_inc.xpos += 1; p_inc.ypos += 1; break;
                        case 1: p_inc.xpos += 1;break;
                        case 2: break;
                        case 3: p_inc.ypos += 1;break;
                }
        }

        /* check if point out of grid bounds */
        if (p_inc.ypos >= NUMY || p_inc.xpos >= NUMX || p_inc.xpos < 0
                || p_inc.ypos < 0) return(!0);

        /* the new left hand point */
        p_temp.xpos = p_inc.xpos * (REAL)INCX + STARTX - left->xpos;
        p_temp.ypos = p_inc.ypos * (REAL)INCY + STARTY - left->ypos;
        left->xpos += p_temp.xpos;
        left->ypos += p_temp.ypos;

        /* check if already matched */
        if(already_matched(left)==MATCHED)
                return(!0);

        /* predict the new right hand point */
        right->xpos += p_temp.xpos * ACCESS(shape_matrix,0,0) +
                       p_temp.ypos * ACCESS(shape_matrix,0,1);
        right->ypos += p_temp.xpos * ACCESS(shape_matrix,1,0) +
                       p_temp.ypos * ACCESS(shape_matrix,1,1);

        pred_right->xpos = right->xpos;
        pred_right->ypos = right->ypos;
```

```
        /* check if patch will walk off images */
        if( (left->xpos - PATCHRADIUS) < 0
         || (left->xpos + PATCHRADIUS) >= leftim->width
         || (left->ypos - PATCHRADIUS) < 0
         || (left->ypos + PATCHRADIUS) >= leftim->height
         || (right->xpos - PATCHRADIUS) < 0
         || (right->xpos + PATCHRADIUS) >= rightim->width
         || (right->ypos - PATCHRADIUS) < 0
         || (right->ypos + PATCHRADIUS) >= rightim->height)
                return(!0);

        return(0);
}


/*
** check_constraints() - verify resulting match to meet various constraints
*/
check_constraints(leftpt,rightpt,pred,shape_matrix,eigen,rs,rg,iters)
  POINTPTR leftpt, rightpt, pred;
  ARRAYPTR shape_matrix;
  REAL eigen, rs, rg;
  int iters;
{
  static REAL sum_s10=0.0, sum_s11=0.0, nmr_pt=0.0;
  REAL s10, s11;

        /* check iterations */
        if(iters <= 0) return(0);

        /* check convergence */
        if( fabs(rightpt->xpos - pred->xpos)>CONRAD
         || fabs(rightpt->ypos - pred->ypos)>CONRAD)
                return(0);

        /* check precision */
        if(EIGN_THRESHOLD != 0.0 && eigen > EIGN_THRESHOLD)
                return(0);

        /* check distortion */
        if(nmr_pt > 1.0) {
                s10 = fabs(ACCESS(shape_matrix,1,0) - (sum_s10/nmr_pt));
                s11 = fabs(ACCESS(shape_matrix,1,1) - (sum_s11/nmr_pt));
                if(s10 > DIST_THRESHOLD || s11 > DIST_THRESHOLD)
                        return(0);
        }

        /* check radiometric ranges */
        if(rs < MIN_RS || rs > MAX_RS || rg < MIN_RG || rg > MAX_RG)
                return(0);
```

```
        sum_s10 += ACCESS(shape_matrix,1,0);
        sum_s11 += ACCESS(shape_matrix,1,1);
        nmr_pt += 1.0;
        return(!0);
}

++++++++++++++++++++++++++++++ File: error.c +++++++++++++++++++++++++++++++++

/*
**  error.c - least-squares error estimation routine
**
**      This module contains the routines to compute the error
**      associated with a given match.  This includes forming
**      the (co)variance matrix K, and computing the standard
**      error.  The largest eigenvalue of the 2x2 sub-matrix of
**      K, formed from the entries corresponding to the disparity
**      measures x and y, is to be computed.  The eigenvalue will
**      be an indicator of the variance of the disparity for the
**      match.  Better still, is the standard deviation, which
**      will be an indicator of the magnitude of the error in pixels.
*/

#include "gruen.h"

extern ARRAY design;    /* design matrix or A */
extern ARRAY rhs;       /* Right-hand side or b */
extern ARRAY X;         /* current solution vecor x */
extern ARRAY V;         /* error vector v = Ax - b */
extern ARRAY GJ;        /* workspace for GJ-elimination, holds inv(AtA) */

#define sqr(_t) ((_t)*(_t))

/*
** eigen():
**
**      Compute the largest eigenvalue of the 2x2 matrix
**      formed from the (co)variance matrix K.
**      See Gruen 1985 for details.
*/
REAL
eigen(patchradius)
  int patchradius;
{
  int i,j,k, rows,cols;
  REAL sum, vari;
  REAL a11, a12, a21, a22;
  REAL b, c, eig1, eig2, redund;
  REAL compute_sigma();

        cols = design.numcol;
        rows = design.numrow;
```

```
        /* calculate variance: sigma^2 */
        vari = compute_sigma(patchradius);

        /*
        ** here we would calculate AtA and then calculate inv(AtA)
        ** but matrix GJ is undisturbed from solve(), so use it.
        */

        /*
        ** form 2x2 submatrix of K - this bit is hardcoded!!
        ** assume x and y are in position 0 and 3
        */
        a11 = vari * ACCESS(&GJ,0,cols);
        a12 = vari * ACCESS(&GJ,0,cols+3);
        a21 = vari * ACCESS(&GJ,3,cols);
        a22 = vari * ACCESS(&GJ,3,cols+3);

        /*
        ** calculate largest eigenvalue using quadratic formulae
        */
        /* a = 1 */
        b = (a11 + a22);
        c = (a11*a22 - a12*a21);

        eig1 = (b + sqrt(sqr(b) - 4.0*c))/2.0;
        eig1 = sqrt(eig1);
        return(eig1);
}


/*
** compute_sigma:
**
**      compute the standard error of unit weight for
**      this match.
**              v = Ax - l
**      Return the variance (err^2).
*/
REAL
compute_sigma(patchradius)
  int patchradius;
{
  int i,j, rows, cols, size;
  REAL sum, vari, redund;

        size = 2 * patchradius + 1;
        cols = design.numcol;
        rows = design.numrow;

        /* compute Ax - l */
        for(j=0; j < rows; j++) {
                sum = 0.0;
```

```
                for(i=0; i < cols; i++) {
                        sum += ACCESS(&design,j,i) * ACCESS(&X,i,0);
                }
                ACCESS(&V,j,0) = sum;
        }
        for(j=0; j < rows; j++) {
                ACCESS(&V,j,0) -= ACCESS(&rhs,j,0);
        }

        vari = 0.0;
        for(j=0; j < rows; j++) {
                vari += sqr(ACCESS(&V,j,0));
        }

        /* redundancy */
        if((RS_ON == 1) || (RS_ON == 2))
                redund = (REAL)(size*size - 7);
        else if (RS_ON == 3)
                redund = (REAL)(size*size - 8);
        else
                redund = (REAL)(size*size - 6);

        vari = vari/redund;
        return(vari);
}


/*
** GJ_elimination():
**
**      Used to calculate the inverse of a matrix Q.
**      The array GJ will contain the result.
*/
GJ_elimination(Q, GJ)
  ARRAYPTR Q, GJ;
{
  int rows, cols;
  int p, q, r;
  REAL t;

        rows = GJ->numrow;
        cols = GJ->numcol;
        q = Q->numcol;

        /* copy Q to GJ, and append identity to GJ */
        for(r=0; r < rows; r++) {
                for(p=0; p < cols; p++) {
                        if(p < q)
                                ACCESS(GJ,r,p) = ACCESS(Q,r,p);
                        else if((p-q) == r)
                                ACCESS(GJ,r,p) = 1.0;
                        else
```

```
                                ACCESS(GJ,r,p) = 0.0;
                }
        }


        for (r=0; (r< ((rows==cols)?(rows-1):rows) )&&(r<cols); r++) {

                if (ACCESS(GJ,r,r) == 0.0) {

                        for (p=r+1; (p<rows)&&(ACCESS(GJ,p,r)==0.0); p++)
                                ;

                        /* swap 2 rows */
                        if (p<rows)
                                for (q=r; q<cols; q++) {
                                        t = ACCESS(GJ,r,q);
                                        ACCESS(GJ,r,q) = ACCESS(GJ,p,q);
                                        ACCESS(GJ,p,q) = t;
                                }
                }

                /* row <- row * constant */
                if (ACCESS(GJ,r,r) != 0.0) {
                        for (q=r+1; q<cols; q++)
                                ACCESS(GJ,r,q) /= ACCESS(GJ,r,r);
                        ACCESS(GJ,r,r) = 1.0;
                }

                if (ACCESS(GJ,r,r) != 0.0) {

                    for (p=0; p<rows; p++)
                        if (p!=r) {

                                /* row <- row + constant * row */
                                for (q=r+1; q<cols; q++)
                                        ACCESS(GJ,p,q) -= ACCESS(GJ,p,r)
                                                * ACCESS(GJ,r,q);

                                ACCESS(GJ,p,r) = 0.0;
                        }
                }
        }
}

++++++++++++++++++++++++++++++ File: list.c ++++++++++++++++++++++++++++++++++

/*
** list.c - List management routines
**
**      This module contains the routines for the list_to_be_grown_from
**      and already_matched control.
*/
```

```
#include "gruen.h"

char *done_list=NULL;
LISTPTR grow_from=NULL;

int seed_counter;

/*
** init_lists():
**
**      Create the "already matched" list, and init the grow_from list.
*/
init_lists()
{
  int i, nelem;

        nelem = (NUMX * INCX) * (NUMY * INCY);
        done_list = (char *)calloc(nelem, sizeof(char));
        if (done_list == NULL) {
                fprintf(stderr,"calloc failed in init_lists() - exit\n");
                exit(1);
        }
        for(i=0; i < nelem; i++)
                done_list[i] = UNMATCHED;

        grow_from = NULL;
        seed_counter = 0;
}


/*
** already_matched():
**
**      Return MATCHED if this point is already matched,
**      WORKING if being worked on, otherwise UNMATCHED.
*/
int
already_matched(origin)
  POINTPTR origin;
{
  int pos;
        if(NUMX==0 || NUMY==0) return(0);
        pos = NUMX * (int)((origin->ypos - STARTY)/INCY)
                + (int)((origin->xpos - STARTX)/INCX);
        return((int)(done_list[pos]));
}


/*
** matched():
**
```

```
**      Mark the point as successfully matched.
*/
matched(origin)
  POINTPTR origin;
{
  int pos;
        if(NUMX==0 || NUMY==0) return(0);
        pos = NUMX * (int)((origin->ypos - STARTY)/INCY)
                + (int)((origin->xpos - STARTX)/INCX);
        /* not on grid */
        if((origin->xpos/(REAL)INCX - floor(origin->xpos/INCX))!=0.0
                || (origin->ypos/(REAL)INCY - floor(origin->ypos/INCY))!=0.0)
                done_list[pos]=UNMATCHED;
        else /* on grid */
                done_list[pos] = MATCHED;
}

/*
** get_list():
**
**      Pop the matched point with the minimum "score" off the
**      list to be grown from.
*/
get_list(l_origin, r_origin, shape, rs_plus, rs_mult, seed)
  POINTPTR l_origin, r_origin;
  ARRAYPTR shape;
  REAL *rs_plus, *rs_mult;
  BOOLEAN *seed;
{
  LISTPTR ptr, deletemin();

        ptr = deletemin(&grow_from);

        if(ptr == NULL) return(0);

        l_origin->xpos = ptr->left.xpos;
        l_origin->ypos = ptr->left.ypos;

        r_origin->xpos = ptr->right.xpos;
        r_origin->ypos = ptr->right.ypos;

        ACCESS(shape,0,0) = /* 1.0; */ ptr->s1;
        ACCESS(shape,0,1) = /* 0.0; */ ptr->s2;
        ACCESS(shape,1,0) = /* 0.0; */ ptr->s3;
        ACCESS(shape,1,1) = /* 1.0; */ ptr->s4;

        *rs_plus = ptr->rs_plus;
        *rs_mult = ptr->rs_mult;

        *seed = ptr->seed;

        /* now delete the node */
```

```
        cfree(ptr);
        return(!0);
}


/*
** store_list():
**
**      store this match data in the list of seeds to be
**      grown from.
*/
store_list(l_origin, r_origin, shape, score, rs_plus, rs_mult, seed)
  POINTPTR l_origin, r_origin;
  ARRAYPTR shape;
  REAL score;
  REAL rs_plus, rs_mult;
  BOOLEAN seed;
{
  LISTPTR new;

        /* create a new node and insert into tree */

        new = (LISTPTR)(calloc((unsigned)1, sizeof(LIST)));
        if(new == NULL) {
                fprintf(stderr,"calloc failed in store_list() - exit\n");
                exit(1);
        }

        new->left.xpos = l_origin->xpos;
        new->left.ypos = l_origin->ypos;
        new->right.xpos = r_origin->xpos;
        new->right.ypos = r_origin->ypos;
        new->s1 = ACCESS(shape,0,0);
        new->s2 = ACCESS(shape,0,1);
        new->s3 = ACCESS(shape,1,0);
        new->s4 = ACCESS(shape,1,1);
        new->rs_plus = rs_plus;
        new->rs_mult = rs_mult;
        new->seed = seed;
        new->score = score;
        new->ptrleft = NULL;
        new->ptrright = NULL;

        insert_list(new, &grow_from);
}


/*
** insert_list():
**
**      Insert the new node into the binary tree.
*/
```

```
insert_list(new, start)
  LISTPTR new, *start;
{
        /* binary tree insert */

        if(*start == NULL)
                *start = new;
        else if (new->score < (*start)->score)
                insert_list(new, &((*start)->ptrleft));
        else
                insert_list(new, &((*start)->ptrright));
}

/*
** deletemin():
**
**      Delete the minimum node from a binary tree,
**      and return a pointer to that node.
*/
LISTPTR
deletemin(start)
  LISTPTR *start;
{
  LISTPTR temp;

        if(*start == NULL) return(NULL);

        if ((*start)->ptrleft == NULL) {
                temp = *start;
                *start = (*start)->ptrright;
                return(temp);
        }

        return(deletemin(&((*start)->ptrleft)));
}

/*
** print_tree():
**
**      Print in-order tree traversal for debug.
*/
print_tree(start)
  LISTPTR start;
{
        if(start == NULL) return;

        if(start->ptrleft == NULL && start->ptrright == NULL) {
            printf("(%4.0f,%4.0f) : %f\n", start->left.xpos, start->left.ypos,
                start->score);
        }
        else {
            print_tree(start->ptrleft);
```

```
            printf("(%4.0f,%4.0f) : %f\n", start->left.xpos, start->left.ypos,
                 start->score);

            print_tree(start->ptrright);
        }
}

++++++++++++++++++++++++++++++ File: io.c ++++++++++++++++++++++++++++++++++++

/*
** io.c - input/output routines
**
**      This module contains routines for handling all I/O.
**      These include opening files, image input, results output,
**      and any debug routines.
*/

#include "gruen.h"
#include <pixrect/pixrect_hs.h>

/*
** file_must_open:
**
**      open file routine
*/
FILE *
file_must_open(filename)
  char *filename;
{
  FILE *fd, *fopen();

        fd = fopen(filename,"r");
        if (fd == NULL){
                fprintf(stderr,"unable to open %s\n",filename);
                exit(1);
        }
        return(fd);
}


/*
** parse:
**
**       parse the command line.
*/
parse(count, List, number)
  int count, number;
  char **List;
{
  float t;
  int d;
```

```
if(count < 2)
        return(0);

if(number >= count)
        return(0);

if (List[number][0] != '-')
        return(1);

switch(List[number][1]) {
        case 's' : /* shaping thresholds follow */
                   if((count-number-1) < 1) return(0);
                   if(sscanf(List[number+1],"%f", &t) != 1) return(0);
                        T_SP = (REAL)t;
                   number += 2;
                   return(number);
                   break;
        case 't' : /* translation threshold follow */
                   if((count-number-1) < 1) return(0);
                   if(sscanf(List[number+1],"%f", &t) != 1) return(0);
                        T_XY = (REAL)t;
                   number += 2;
                   return(number);
                   break;
        case 'i' : /* max iterations follows */
                   if((count-number-1) < 1) return(0);
                   if(sscanf(List[number+1],"%d", &d) != 1) return(0);
                        MAXITERS = (int)d;
                   number += 2;
                   return(number);
                   break;
        case 'p' : /* patchradius */
                   if((count-number-1) < 1) return(0);
                   if(sscanf(List[number+1],"%d", &d) != 1) return(0);
                        PATCHRADIUS = (int)d;
                   number += 2;
                   return(number);
                   break;
        case 'g' : /* grid/neighbour spacing */
                   if((count-number-1) < 6) return(0);
                   if(sscanf(List[number+1],"%f", &t) != 1) return(0);
                        STARTX = (REAL)t;
                   if(sscanf(List[number+2],"%d", &d) != 1) return(0);
                        NUMX = (int)d;
                   if(sscanf(List[number+3],"%d", &d) != 1) return(0);
                        INCX = (int)d;
                   if(sscanf(List[number+4],"%f", &t) != 1) return(0);
                        STARTY = (REAL)t;
                   if(sscanf(List[number+5],"%d", &d) != 1) return(0);
                        NUMY = (int)d;
                   if(sscanf(List[number+6],"%d", &d) != 1) return(0);
                        INCY = (int)d;
```

```
                           number += 7;
                           return(number);
                           break;
                case 'e' : /* eigenvalue threshold */
                           if((count-number-1) < 1) return(0);
                           if(sscanf(List[number+1],"%f", &t) != 1) return(0);
                                EIGN_THRESHOLD = (REAL)t;
                           number += 2;
                           return(number);
                           break;
                case 'r' : /* radiometric threshold */
                           if((count-number-1) < 1) return(0);
                           if(sscanf(List[number+1],"%f", &t) != 1) return(0);
                                T_RS = (REAL)t;
                           number += 2;
                           return(number);
                           break;
                case 'd' : /* distortion threshold */
                           if((count-number-1) < 1) return(0);
                           if(sscanf(List[number+1],"%f", &t) != 1) return(0);
                                DIST_THRESHOLD = (REAL)t;
                           number += 2;
                           return(number);
                           break;
                case 'c' : /* convergence radius */
                           if((count-number-1) < 1) return(0);
                           if(sscanf(List[number+1],"%f", &t) != 1) return(0);
                                CONRAD = (REAL)t;
                           number += 2;
                           return(number);
                           break;
                case 'm' : /* max radiometric shift */
                           if((count-number-1) < 1) return(0);
                           if(sscanf(List[number+1],"%f", &t) != 1) return(0);
                                MAX_RS = (REAL)t;
                           number += 2;
                           return(number);
                           break;
                case 'n' : /* min radiometric shift */
                           if((count-number-1) < 1) return(0);
                           if(sscanf(List[number+1],"%f", &t) != 1) return(0);
                                MIN_RS = (REAL)t;
                           number += 2;
                           return(number);
                           break;
                case 'u' : /* max radiometric gain */
                           if((count-number-1) < 1) return(0);
                           if(sscanf(List[number+1],"%f", &t) != 1) return(0);
                                MAX_RG = (REAL)t;
                           number += 2;
                           return(number);
                           break;
                case 'v' : /* min radiometric gain */
                           if((count-number-1) < 1) return(0);
                           if(sscanf(List[number+1],"%f", &t) != 1) return(0);
                                MIN_RG = (REAL)t;
                           number += 2;
                           return(number);
                           break;
                case 'x' : /* radiometric option */
                           if((count-number-1) < 1) return(0);
                           if(sscanf(List[number+1],"%d", &d) != 1) return(0);
                                RS_ON = (int)d;
                           if(RS_ON < 0 || RS_ON > 3) return(0);
                           number += 2;
                           return(number);
                           break;
                case 'b' : /* debug ON */
                           if((count-number-1) < 1) return(0);
                           Debug = !0;
                           number += 1;
                           return(number);
                           break;
                default: return(0);
        }
}


/*
** read_commandline:
**
*/
read_commandline(argc, argv)
{
  int arg, savearg;

        /* set defaults */
        T_SP =  0.05;
        T_XY =  0.05;
        T_RS =  0.05;   /* not used */
        CONRAD = 3.0;
        MAX_RS = 128.0; MIN_RS = -128.0;
        MAX_RG = 4.0;   MIN_RG = 0.25;
        MAXITERS = 16;
        SEED_MAXITERS = 30;
        PATCHRADIUS = 12;
        STARTX = -1.0; NUMX = INCX = -1;
        STARTY = -1.0; NUMY = INCY = -1;
        EIGN_THRESHOLD = 0.0;
        DIST_THRESHOLD = 0.1;
        RS_ON = 3; /* both */
        Debug = 0;

        arg = 1;
```

```
        do {
                savearg = arg;
                arg = parse(argc,argv, arg);
                if(arg == 0) return(0);
        } while(arg != 1);

        if((argc-savearg) >= 2)
                return(savearg);
        else
                return(0);
}


/*
** read_image:
**
**      read in Sun Pixrect format image
*/
read_image(fd,image)
  FILE *fd;
  IMAGEPTR image;
{
  int i,j;
  unsigned char *ptr;
  colormap_t *colourmap=NULL;
  struct pixrect *mpr, *pr_load();

        mpr = pr_load(fd, colourmap);

        if (mpr == NULL) {
                fprintf(stderr,"error loading image\n");
                exit(1);
        }

        /* make data contiguous */
        if((ptr=(unsigned char *)
            malloc(sizeof(unsigned char)*mpr->pr_width*mpr->pr_height))==NULL) {
                fprintf(stderr,"memory allocation error\n");
                exit(1);
        }

        image->data = ptr;
        for(j=0; j < mpr->pr_height; j++)
                for(i=0; i < mpr->pr_width; i++)
                        *ptr++ = (unsigned char)pr_get(mpr, i,j);

        image->width = mpr->pr_width;
        image->height = mpr->pr_height;

        pr_destroy(mpr);
}
```

```
/*
** read_points:
**
**      seed points are read from stdin, with optional parameters
*/
read_points(left_origin, right_origin, new_shaping_matrix, def_shaping_matrix,
                rs_plus, rs_mult)
  POINTPTR left_origin, right_origin;
  ARRAYPTR new_shaping_matrix;
  ARRAYPTR def_shaping_matrix;
  REAL *rs_plus, *rs_mult;
{
  char line[256];
  float x1, x2, y1, y2;
  float t1, t2, t3, t4, ts1, ts2, dummy;
  REAL s1, s2, s3, s4, rs1, rs2;
  int no_items;

        if (gets(line) == NULL)
                return(EOF);

        /*
        ** read changed to match output format, hence the "dummy"
        ** to ignore the disparity column.
        */
        no_items=sscanf(line,"%f %f %f %f %f %f %f %f %f %f %f",
                &x1,&y1,&x2,&y2,&dummy,&t1,&t2,&t3,&t4, &ts1, &ts2);

        if (no_items == EOF) return(EOF);

        if (no_items < 4)
                perr("Error in standard input file. No. of items < 4");

        left_origin->xpos = (REAL)x1;
        left_origin->ypos = (REAL)y1;
        right_origin->xpos = (REAL)x2;
        right_origin->ypos = (REAL)y2;

        if (no_items == 4) {
                /* use default shaping parameters */
                s1 = ACCESS(def_shaping_matrix,0,0);
                s2 = ACCESS(def_shaping_matrix,0,1);
                s3 = ACCESS(def_shaping_matrix,1,0);
                s4 = ACCESS(def_shaping_matrix,1,1);
                rs1 = 0.0;
                rs2 = 1.0;
        }

        if (no_items > 4 && no_items != 11)
                perr("Error in standard input file. No. of item > 4 < 10");
        else if (no_items == 11) {
```

```
                /* use user-supplied shaping parameters */
                s1 = (REAL)t1;
                s2 = (REAL)t2;
                s3 = (REAL)t3;
                s4 = (REAL)t4;
                rs1 = (REAL)ts1;
                rs2 = (REAL)ts2;
        }

        ACCESS(new_shaping_matrix,0,0) = s1;
        ACCESS(new_shaping_matrix,0,1) = s2;
        ACCESS(new_shaping_matrix,1,0) = s3;
        ACCESS(new_shaping_matrix,1,1) = s4;

        *rs_plus = rs1;
        *rs_mult = rs2;

        return(!EOF);
}

/*
** print_patch:
**
**      debug patch dump
*/
print_patch(patch, patchradius)
  ARRAYPTR patch;
{
  int i,j,d;

        d = 2 * patchradius + 1;

        printf("Patch of size %d * %d\n", d, d);
        for (i = 0; i < d; i++) {
                for (j = 0; j < d; j++)
                        printf("%10.8f ",ACCESS(patch,i,j));
                printf("\n");
        }
        fflush(stdout);
}

/*
** print_matrix:
**
**      debug matrix dump
*/
print_matrix(matrix, r , c)
  ARRAYPTR matrix;
  int r, c;     /* pass in r,c to avoid printing oversize matices */
{
  int i,j;
```

```
        printf("Matrix of size %d * %d\n", r, c);
        for (i = 0; i < r; i++) {
                for (j = 0; j < c;j++)
                        printf("%-10.4f ",ACCESS(matrix,i,j));
                printf("\n");
        }
        fflush(stdout);
}


print_header(file1,file2)
  char *file1, *file2;
{
    printf("    '%s <-> %s' p:%2d t:%5.3f s:%5.3f d:%5.3f c:%5.3f e:%5.3f",
        file1,file2,PATCHRADIUS,T_XY,T_SP,DIST_THRESHOLD,CONRAD,EIGN_THRESHOLD);
    printf(" i:%3d g:%2d %2d rs:%1d\n", MAXITERS, INCX,INCY, RS_ON);
    printf("    lx      ly      rx      ry      disp        shaping parameters");
    printf("            rs+     rs*   iter  eign\n");
    fflush(stdout);
}


/*
** print_results:
**
**      print out final results for this match
*/
print_results(fd, shaping, left_origin, right_origin, iters, rs_plus, rs_mult, ps,
                vari, eig)
  FILE *fd;
  ARRAYPTR shaping;
  POINTPTR left_origin;
  POINTPTR right_origin;
  int iters;
  REAL rs_plus, rs_mult;
  int ps;
  REAL vari;
  REAL eig;
{
  REAL disparity();

    fprintf(fd,"%8.2f%8.2f%8.3f%8.3f%8.3f%8.4f%8.4f%8.4f%8.4f%8.2f%8.2f%5d%8.5f\n",
        left_origin->xpos, left_origin->ypos, right_origin->xpos,
        right_origin->ypos, left_origin->xpos - right_origin->xpos,
        ACCESS(shaping,0,0), ACCESS(shaping,0,1), ACCESS(shaping,1,0),
        ACCESS(shaping,1,1), rs_plus, rs_mult, iters, eig);

    fflush(fd);
}

/*
** perr:
```

```
**
**      print error string and exit.
*/
perr(s)
char *s;
{
        perror(s);
        exit(2);
}

/*
** print_format:
**
**      print command line format and exit
**
*/
print_format(progname)
  char *progname;
{
  FILE *fp;
  fp=stderr;
  fprintf(fp,"format: %s [options] left_image right_image", progname);
  fprintf(fp," (<seed_points)\n");
  fprintf(fp,"\toptions:\n");
  fprintf(fp,"\t-p: patchradius (1 int), def: %d\n", PATCHRADIUS);
  fprintf(fp,"\t-t: translation parameter threshold (1 float), def: %6.4f\n", T_XY);
  fprintf(fp,"\t-s: shaping parameter threshold (1 float), def: %6.4f\n", T_SP);
  fprintf(fp,"\t-i: maximum iterations (1 int), def: %d\n", MAXITERS);
  fprintf(fp,"\t-g: startx numx incx starty numy incy (6 int), def: all space 5\n");
  fprintf(fp,"\t-e: eigenvalue threshold (1 float), def: %6.4f\n", EIGN_THRESHOLD);
  fprintf(fp,"\t-c: convergence radius (1 float), def: %6.4f\n", CONRAD);
  fprintf(fp,"\t-x: radiometric option 0=none,1=mult,2=plus,3=both (1 int), def: %d\n",
RS_ON);
  fprintf(fp,"\t-d: distortion threshold (1 float), def: %6.4f\n", DIST_THRESHOLD);
  fprintf(fp,"\t-m: max radiometric shift (1 float), def: %6.4f\n", MAX_RS);
  fprintf(fp,"\t-n: min radiometric shift (1 float), def: %6.4f\n", MIN_RS);
  fprintf(fp,"\t-u: max radiometric gain (1 float), def: %6.4f\n", MAX_RG);
  fprintf(fp,"\t-v: min radiometric gain (1 float), def: %6.4f\n", MIN_RG);
  fprintf(fp,"\t-b: debuggin on\n");
  exit(1);
}

++++++++++++++++++++++++++++++ File: memory.c ++++++++++++++++++++++++++++++++

/*
** memory.c - memory allocation and de-allocation routines
**
**      This module contains the routines to handle all memory
**      allocation and initialization.
**
**      All global arrays, etc. are declared here.
*/
```

```
#include "gruen.h"

ARRAY identity_matrix;          /* identity matrix as default    */
ARRAY left_patch,right_patch;   /* image patches to be matched   */
ARRAY right_xd_patch;           /* x derivatives of right patch  */
ARRAY right_yd_patch;           /* y derivatives of right patch  */
ARRAY design;                   /* matrix of equations (A)       */
ARRAY rhs;                      /* right-hand-side or (l)        */
ARRAY shape_matrix;             /* matrix of shaping parameters  */
ARRAY t_shape_matrix;           /* matrix of shaping parameters  */
ARRAY threshold_values;         /* termination conditions        */
ARRAY larger_patch;             /* patch+1 for computing derivs. */

ARRAY C;                        /* temp matrix (AtA)             */
ARRAY D;                        /* new rhs (Atb)                 */
ARRAY Y;                        /* intermediate sol.             */
ARRAY X;                        /* solution vector (updates)     */
ARRAY V;                        /* error terms v = Ax -b         */
ARRAY GJ;                       /* Gauss-Jordian elim. workspace */

/*
** REAL_array_alloc:
**
**      Routine to allocate 2D Floating point (REAL) array
**      set up array as row major (as in C)
*/
REAL_array_alloc(new_array,rows,cols)
  ARRAYPTR new_array;
  int rows, cols;
{
  unsigned int size;
  int i;
  REAL *vector;
  REAL **index;
  char *calloc();

        size = (unsigned)(rows * cols);

        if ((vector = (REAL *)calloc(size,sizeof(REAL))) == NULL)
                perr("gruen: calloc failed in REAL_array_alloc()");

        if ((index = (REAL **)(calloc((unsigned)cols,sizeof(REAL *)))) == NULL)
                perr("gruen: calloc failed in REAL_array_alloc()");

        /* create column index */
        new_array->start = index;
        new_array->numrow = rows;
        new_array->numcol = cols;
        for (i = 0; i < cols; i++){
                *index++ = vector;
                vector += rows;
```

```
        }

        /* init */
        vector = new_array->start[0];
        for(i=0; i < size; i++)
                *vector++ = 0.0;

}

/*
** init_matrices:
**
**      Init (calloc) all matrices
*/
init_matrices(patchradius)
  int patchradius;
{
  int d,n,t;

        d = (2 * patchradius + 1);
        n = d * d;
        t = NUM_PARAMS;

        if(RS_ON == 1 || RS_ON == 2) t += 1;
        if(RS_ON == 3) t += 2;

        REAL_array_alloc(&design,n,t);
        REAL_array_alloc(&rhs,n,1);
        REAL_array_alloc(&larger_patch, d+2, d+2);
        REAL_array_alloc(&left_patch,d,d);
        REAL_array_alloc(&right_patch,d,d);
        REAL_array_alloc(&right_xd_patch,d,d);
        REAL_array_alloc(&right_yd_patch,d,d);

        REAL_array_alloc(&identity_matrix,2,2);
        REAL_array_alloc(&shape_matrix,2,2);
        REAL_array_alloc(&t_shape_matrix,2,2);
        ACCESS(&identity_matrix,0,0) = 1;
        ACCESS(&identity_matrix,0,1) = 0;
        ACCESS(&identity_matrix,1,0) = 0;
        ACCESS(&identity_matrix,1,1) = 1;

        REAL_array_alloc(&threshold_values,8,1);
        ACCESS(&threshold_values,0,0) = T_XY;
        ACCESS(&threshold_values,1,0) = T_SP;
        ACCESS(&threshold_values,2,0) = T_SP;
        ACCESS(&threshold_values,3,0) = T_XY;
        ACCESS(&threshold_values,4,0) = T_SP;
        ACCESS(&threshold_values,5,0) = T_SP;
        ACCESS(&threshold_values,6,0) = T_RS;
        ACCESS(&threshold_values,7,0) = T_RS;
```

```
        /* solution vars */
        REAL_array_alloc(&C,t,t);
        REAL_array_alloc(&D,t,1);
        REAL_array_alloc(&Y,t,1);
        REAL_array_alloc(&X,t,1);
        REAL_array_alloc(&V,n,1);
        REAL_array_alloc(&GJ,t,t*2);
}

/*
** free_matrices:
**
**      Be nice and free up memory.
*/
free_matrices()
{
        REAL_array_dealloc(&design);
        REAL_array_dealloc(&rhs);
        REAL_array_dealloc(&larger_patch);
        REAL_array_dealloc(&left_patch);
        REAL_array_dealloc(&right_patch);
        REAL_array_dealloc(&right_xd_patch);
        REAL_array_dealloc(&right_yd_patch);

        REAL_array_dealloc(&identity_matrix);
        REAL_array_dealloc(&shape_matrix);
        REAL_array_dealloc(&t_shape_matrix);

        REAL_array_dealloc(&threshold_values);

        REAL_array_dealloc(&C);
        REAL_array_dealloc(&D);
        REAL_array_dealloc(&Y);
        REAL_array_dealloc(&X);
        REAL_array_dealloc(&V);
        REAL_array_dealloc(&GJ);
}

/*
** REAL_array_dealloc:
**
**      Free memory for this array.
*/
REAL_array_dealloc(old_array)
  ARRAYPTR old_array;
{
        cfree(old_array->start[0]);
        cfree(old_array->start);
}
```

References

[Abramowitz65]
Abramowitz, M. and I. A. Stegun, *Handbook of Mathematical Functions*, Dover Publications, Inc., New York, 1965.

[Akita82]
Akita K. and H. Kuga, "A Computer Method of Understanding Ocular Fundus Images", *Pattern Recognition*, vol. 15, No. 6, pp. 431-443, 1982.

[Algazi85]
Algazi, V. R., J. L. Keltner, and C. A. Johnson, "Computer Analysis of the Optic Cup in Glaucoma", *Invest. Ophthal. Vis. Sci.*, vol. 26, no. 12, pp. 1759-1770, 1985.

[Allen90]
Allen, R. E. (Editor), *The Concise Oxford Dictionary of Current English* 8th Edition, Clarendon Press, Oxford, 1990.

[Arnold80]
Arnold, R. D. and T. O. Binford, "Geometric Constraints in Stereo Vision", *Proc. SPIE*, vol. 238, pp. 281-292, 1980.

[Badique88]
Badique, E., Y. Komiya, N. Ohyama, T. Honda, and J. Tsujiuchi, "Use of Color Image Correlation in the Retrieval of Gastric Surface Topography by Endoscopic Stereopair Matching", *Applied Optics*, vol. 27, no. 5, 1988.

[Bajcsy82]
Bajcsy, R. and C. Broit, "Matching Deformed Images", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 351-353, 1982.

[Bajcsy89]
Bajcsy, R. and S. Kovacic, "Multiresolution Elastic Matching", *Compter Vision, Graphics, and Image Processing*, vol. 46, pp. 1-21, 1989.

[Ballard82]
Ballard, D. H. and C. M. Brown, *Computer Vision*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1982.

[Barnard80]
Barnard, S. T. and W. B. Thompson, "Disparity Analysis of Images", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-2, no. 4, pp. 333-340, 1980.

[Barnea72]
Barnea, D. I. and H. F. Silverman, "A Class of Algorithms for Fast Digital Image Registration", *IEEE Transactions on Computers*, vol. C-21, no. 2, pp. 179-186, 1972.

[Bedell27]
Bedell, A. J., "Photographs of the Fundus Oculi", *New York Journal of Medicine*, vol. 27, pp. 951-971, 1927.

[Bengtsson77]
Bengtsson, B. and C. E. T. Krakau, "Some Essential Optical Features of the Zeiss Fundus Camera", *Acta Ophthalmologica*, vol. 55, pp. 123-131, 1977.

[Bishop87]
Bishop, K. and T. Krupin, "Variability of Optic Disc Topography Measurements on the Rodenstock Optic Disc Analyzer", *Invest. Ophthal. Vis. Sci. (ARVO suppl.)*, 1987.

[Bogler86]
Bogler, P. I., "An Analytical Expression for Correlation Accuracy", IEEE Transactions on Aerospace and Electronic Systems, vol. AES-22, no. 5, pp. 648-657, 1986.

[Bougrenet88]
de Bougrenet de la Tocnaye, J. L. and F. Ghorbes, "Scale-rotation invariant pattern recognition applied to image data compression", *Pattern Recognition Letters*, vol. 8, pp. 55-58, 1988.

[Bracewell65]
Bracewell, R., *The Fourier Transform and Its Applications* McGraw-Hill, Inc., New York, 1965.

[Brzakovic86]
Brzakovic, D. and B. White, "Surface Orientation Using Texture Gradient Derived from Two Views of a Scene", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1045-1047, 1986.

[Caelli88]
Caelli, T. M. and Z. Q. Liu, "On the minimum number of templates required for shift, rotation and size invariant pattern recognition", *Pattern Recognition Letters*, vol. 21, pp. 205-216, 1988.

[Caprioli87]
Caprioli, J. and J. M. Miller, "Topographic Measurements of the Optic Nerve Head in Glaucoma", *Invest. Ophthal. Vis. Sci. (ARVO suppl.)*, 1987.

[Canny83a]
Canny, J. F., "A Variational approach to Edge Detection", *Proc 3rd NAAI*, pp. 54-58, 1983

[Canny83b]
Canny, J. F., "Finding Edges and Lines in Images", *M.I.T. AI Laboratory Technical Report* vol. 720, Cambridge, Ma. 1983.

[Castan85]
Castan, S. and J. Shen, "A Stereo Vision Algorithm Taking into Account the Perspective Distortions", *Proc. 7th Int. Conf. Pattern Recognition*, pp. 444-446, 1985.

[Castleman79]
Castleman, K. R., *Digital Image Processing*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1979.

[Chan88]
Chan, M. H. and H. T. Tsui, "Recognition of partially occluded 3-D objects by depth map modelling", *Pattern Recognition Letters*, vol. 7, pp. 319-327, 1988.

[Conners80]
Conners, R. W. and C. A. Harlow, "A Theoretical Comparison of Texture Algorithms", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-2, no. 3, pp. 204-222, 1980.

[DeSoete84]
De Soete, G., "A least squares algorithm for fitting an ultrametric tree to a dissimilarity matrix", *Pattern Recognition Letters*, vol. 2, pp. 133-137, 1984.

[Donaldson64]
Donaldson, D. D., "A New Camera for Stereoscopic Fundus Photography", *Trans. American Ophth. Soc.*, vol. 62, pp. 439-458, 1964.

[Dvornyechenko83]
Dvornyechenko, V. N., "Bounds on (Deterministic) Correlation Functions with Applications to Registration", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-5, pp. 206-213, 1983.

[Eastman87]
Eastman, R. D. and A. M. Waxman, "Using Disparity Functionals for Stereo Correspondence and Surface Reconstruction", *Computer Vision, Graphics, and Image Processing*, vol. 39, pp. 73-101, 1987.

[ElHage80]

El Hage, S. G. and F. Berny, "Contribution of Crystalline Lens to the Sperical Aberation of the Eye", *Journal of the Optical Society of America*, vol. 63, p. 205, 1973.

[Ffytche73]

Ffytche, T. J., A. R. Elkington, and I. J. Dowman, "Photogrammetry of the Optic Disc", *Transactions of Ophthalmological Society, UK*, vol. 93, p. 251, 1973.

[Fitzke91]

Fitzke, F. W., H. Woon, G. Timberlake, L. Robinson, J. Marshall and A. C. Bird, "Optical Modifications to a Scanning Laser Ophthalmoscope for High Magnification, Noarrow Optical Section Imaging", *Laser and Light in Ophthalmology*, vol. 4, no. 1, pp. 7-14, 1991.

[Forshaw90]

Forshaw, M. R. B. and D. G. Greenwood, "The Structure of Auto- and Cross-Correlation Functions", Report No. 90/1, Image Processing Group, University College London, 1990.

[Gagalowicz81]

Gagalowicz, A., "A New Method for Texture Fields Synthesis: Some Applications to the Study of Human Vision", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-3, no. 5, pp. 520-533, 1981.

[Gemmar82]

Gemmar, P., "Image correlation: processing requirements and implementation structures on a flexible image processing system (FLIP)", in *Multicomputers and Image Processing*, Academic Press, New York, pp. 87-103, 1982.

[Goldmann77]

Goldmann, H. and W. Lotmar, "Rapid Detection of Changes in the Optic Disc: Stereo Chronoscopy", *Graefes Archiv Ophthalmologie*, vol. 202, pp. 87-99, 1977.

[Golub83]

Golub, G. H. and C. F. VanLoan, *Matrix Computations*, North Oxford Academic Publishing Co., Oxford, England, 1983.

[Gonzalez87]

Gonzalez, R. C. and P. Wintz, *Digital Image Processing*, Second Edition, Addison-Wesley Publishing Co., Reading, Mass., 1987.

[Goshtasby85]

Goshtasby, A., "Template Matching in Rotated Images", *IEEE Transactions* *on Pattern Analysis and Machine Intelligence*, vol. PAMI-7, no. 3, pp. 338-344, 1985.

[Goshtasby87]
Goshtasby, A., "Piecewise Cubic Mapping Functions for Image Registration", *Pattern Recognition*, vol. 20, no. 5, pp. 525-533, 1987.

[Gray74]
Gray, H., *Anatomy, Descriptive and Surgical*, T. P. Pick and R. Howden Eds., Running Press, Philadelphia, Pennsylvania, 1974.

[Griffin90]
Griffin, P. M. and B. L. Deuermeyer, "A Methodology for Pattern Matching of Complex Objects", *Pattern Recognition*, vol. 23, pp. 245-254, 1990.

[Grimson81]
Grimson, W. E. L., *From Images to Surfaces: A Computational Study of the Human Early Visual System*, M.I.T. Press, Cambridge, Ma., 1981.

[Grimson82]
Grimson, W. E. L., "A Computational Theory of Visual Surface Interpolation", *Phil. Trans. Roy. Soc. London*, vol. B 298, pp. 395-427, 1982.

[Grimson85]
Grimson, W. E. L., "Computational Experiments with a Feature Based Stereo Algorithm", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-7, no. 1, pp. 17-34, 1985.

[Gruen85]
Gruen, A. W., "Adaptive Least Squares Correlation - A powerful Image Matching Technique", *S. Afr. Journal of Photogrammetry, Remote Sensing and Cartography*, vol. 14, no. 3, 1985.

[Gruen86]
Gruen, A. W. and E. P. Baltsavias, "High Precision Image Matching for Digital Terrain Model Generation", *International Archives of Photogrammetry and Remote Sensing*, vol. 26, no. 3, pp. 284-296, 1986.

[Gullstrand11]
Gullstrand, A., "Einfuhrung in die Methoden der Dioptrik des Auges des Menschen", Leipzig, 1911.

[Haykin86]
Haykin, S., *Adaptive Filter Theory*, Prentice-Hall, Englewood Cliffs, New Jersey, 1986.

[Hensher81]
Hensher, D. A. and L. W. Johnson, *Applied Discrete Choice Modelling*, Halsted Press, John Wiley & Sons, New York, New York, 1981.

[Hitchings76]
Hitchings, R. and G. Spaeth, "The Optic Disc in Glaucoma", *British Journal of Ophthalmology*, vol. 60, p. 778, 1976.

[Hixson87]
Hixson, J. R., "Revolutionary Visualization of the Eye", *Medical Tribune, International Medical News Weekly*, Wed. July 15, 1987.

[Hoff86]
Hoff, W. and N. Ahuja, "Surfaces from Stereo", *Proc. 8th Int. Conf. Pattern Recognition*, pp. 516-518, 1986.

[Horn81]
Horn, B. K. P. and B. G. Schunck, "Determining Optical Flow", *Artifiical Intelligence*, vol. 17, pp. 185-203, 1981.

[Jenkins51]
Jenkins, F. A. and H. E. White, *Fundamentals of Optics*, McGraw-Hill Publishing Co., London, England, 1951.

[Johnson79]
Johnson, C. A., J. L. Keltner, M. A. Krohn, and G. L. Portney, "Photogrammetry of the Optic Disc in Glaucoma and Ocular Hypertension with Simultaneous Stereo Photography", *Invest. Ophthal. Vis. Sci.*, vol. 18, no. 12, pp. 1252-1263, 1979.

[Julesz60]
Julesz, B., "Binocular Depth Perception of Computer Generated Patterns", *Bell Systems Tech. J.*, vol. 39, pp. 1125-1162, 1960.

[Kaneko82]
Kaneko, H. and E. Yodogawa, "A Markov Random Field Application to Texture Classification", *Proc. IEEE* vol. 6, pp. 221-225, 1982.

[Keating85]
Keating, J. P., "Some Practical Aspects of Covariance Estimation", *Pattern Recognition Letters*, vol. 3, pp. 295-298, 1985.

[Kolli87]
Kolli, K. and J. McDermott, "Minimizing Error when Following Patients with Optic Nerve Head Analyses", *Invest. Ophthal. Vis. Sci. (ARVO suppl.)*, 1987.

[Kottler74]

Kottler, M. S., A. R. Rosenthal, and D. G. Falconer, "Digital Photogrammetry of the Optic Nervehead", *Invest. Ophthal. Vis. Sci.*, vol. 13, no. 2, pp. 116-120, 1974.

[Lee84]

Lee, C. and A. Rosenfeld, "An Approximation Technique for Photometric Stereo", *Pattern Recognition Letters*, vol. 2, pp. 339-343, 1984.

[Lee88]

Lee, S. and J. H. Kim, "A Fast Computational method for Minimum Square Error Transforms", *Pattern Recognition Letters*, vol. 8, pp. 143-146, 1988.

[Lee90]

Lee, H. J. and W. L. Lei, "Region matching and depth finding for 3D objects in stereo aerial photographs", *Pattern Recognition*, vol. 23, pp. 81-94, 1990.

[Lee91a]

Lee, S. and M. Brady, "Integrating Stereo and Photometric Stereo to Monitor the Development of Glaucoma, *Image and Vision Computing*, vol. 9, no. 1, pp. 39-44, 1991.

[Lee91b]

Lee, S. and M. Brady, "Optic Disk Boundary Detection", *Proc. British Mach. Vision Conf.*, P. Mowforth, Ed., Springer Verlag, pp. 359-362, 1991.

[LeGrand80]

Le Grand, Y. and S. G. El Hage, *Physiological Optics*, Springer Verlag, Berlin, 1980.

[Levine73]

Levine, M. D., D. A. O'Handley, and G. M. Yagi, "Computer Determination of Depth Maps", *Computer Graphics and Image Processing*, vol. 2, pp. 131-150, 1973.

[Lin86]

Lin, Z., H. Lee, and T. S. Huang, "Finding 3-D Point Correspondences in Motion Estimation", *IEEE*, pp. 303-305, 1986.

[Ling86]

Ling, A., T. Krile, S. Mitra, and Z. Shihab, "Early Detection of Glaucoma using Digital Image Processing", *Invest. Ophthal. Vis. Sci. (ARVO suppl.)*, vol. 27, 1986.

[Littmann82]

Littmann, H., "Determination of the Real Size of an Object on the Fundus of the Living Eye" (in German), Klin. Mbl. Augenheilk., vol. 180, pp. 286-289, 1982.

[Lloyd85]

Lloyd, S. A., "Binocular Stereo Algorithm Based on the Disparity-Gradient Limit and using Optimisation Theory", *Image and Vision Computing*, vol. 3, no. 4, pp. 177-181, 1985.

[Lloyd86]

Lloyd, S. A., "Stereo Matching Using Intra- and Inter-row Dynamic Programming", *Pattern Recognition Letters*, vol. 4, pp. 273-277, 1986.

[Lloyd87]

Lloyd, S. A., E. R. Haddow, and J. F. Boyce, "A Parallel Binocular Stereo Algorithm Utilizing Dynamic Programming and Relaxation Labeling", *Computer Vision, Graphics, and Image Processing*, vol. 39, pp. 202-225, 1987.

[March88]

March, R., "Computation of Stereo Disparity using Regularization", *Pattern Recognition Letters*, vol. 8, pp. 181-187, 1988.

[Marr76]

Marr, D. and T. Poggio, "Cooperative Computation of Stereo Disparity", *Science*, vol. 194, pp. 283-287, 1976.

[Marr79]

Marr, D. and T. Poggio, "A Computational Theory of Human Stereo Vision", *Proc. Roy. Soc. London*, vol. B 204, pp. 301-328, 1979.

[McGillem76]

McGillem, C. D. and M. Svedlow, "Image Registration Error Variance as a Measure of Overlay Quality", *IEEE Transactions on Geoscience Electranics", vol. GE-14, no. 1, pp. 45-49, 1976.*

[Medioni85]

Medioni, G. and R. Nevatia, "Segment-Based Stereo Matching", *Computer Vision, Graphics, and Image Processing*, vol. 31, pp. 2-18, 1985.

[Metzger27]

Metzger, E., "Die Stereophotographie des Augenhintergrundes", *Klin. Monatsbl. Augenh.*, vol. 78, pp. 338-348, 1927.

[Mikelberg84]

Mikelberg, F. S, G. R. Douglas, M. S. Schultzer, T. N. Cornsweet, and K. Wijsman, "Reliability of Optic Disk Topographic Measurements Recorded with a Video-Ophthalmograph", *American Journal of Ophthalmology*, vol. 95, no. 1, pp. 98-102, 1984.

[Mikelberg86]

Mikelberg, F. S, G. R. Douglas, and M. S. Schultzer, "The Correlation Between Cup-Disc Ratio, Neuroretinal Rim Area, and Optic Disc Area Measured by the Video-Ophthalmograph and Clinical Measurement", *American Journal of Ophthalmology*, vol. 101, no. 1, pp. 7-12, 1986.

[Miller87]

Miller, J. M. and J. Caprioli, "Videographic Measurements of Optic Disc Pallor", *Invest. Ophthal. Vis. Sci. (ARVO suppl.)*, 1987.

[Mitra87]

Mitra, S., S. Whiteside, B. Nutter, T. Krile, and R. Brown, "Fundus Image Analysis in Glaucoma with a PC-based System", *Invest. Ophthal. Vis. Sci. (ARVO suppl.)*, 1987.

[Moravec77]

Moravec, H. P., "Towards Automatic Visual Obstacle Avoidance", *Proc. 5th Int. Joint Conf. Artificial Intell.*, Cambridge, Ma., p. 584, 1977.

[Morton84]

Morton, R. A., *Photography for the Scientist*, Academic Press, London, England, 1984.

[Mostafavi78]

Mostafavi, H. and F. W. Smith, "Image Correlation with Geometric Distortion Part I: Acquisition Performance", *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-14, no. 3, pp. 487-493, 1978.

[Mostafavi78]

Mostafavi, H. and F. W. Smith, "Image Correlation with Geometric Distortion Part II: Effect on Local Accuracy", *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-14, no. 3, pp. 494-500, 1978.

[Nagel83]

Nagel, H., "Displacement Vectors derived from Second-Order Intensity Variations in Image Sequences", *Computer Vision, Graphics, and Image Processing*, vol. 21, pp. 85-117, 1983.

[Nagin85a]

Nagin, P., B. Schwartz, and K. Nanba, "The Reproducibility of Computerized

Boundary Analysis for Measuring Optic Disc Pallor in the Normal Optic Disc", *Ophthalmology*, vol. 92, no. 2, pp. 243-251, 1985.

[Nagin85b]
Nagin, P. and B. Schwartz, "Detection of Increased Pallor over Time", *Ophthalmology*, vol. 92, no. 2, pp. 252-261, 1985.

[Nevatia76]
Nevatia, R., "Depth Measurement by Motion Stereo", *Computer Graphics and Image Processing*, vol. 5, pp. 203-214, 1976.

[Nishihara84]
Nishihara, H. K., "Practical Real-Time Imaging Stereo Matcher", *Optical Engineering*, vol. 23, no. 5, pp. 536-545, 1984.

[Nordenson15]
Nordenson, J. W., "Om Centrisk Fotografering au Ogonbotten", *Hygea*, vol. 77, pp. 1538-52, 1915.

[Nordenson27]
Nordenson, J. W., "Stereoskopiche Ophthalmographie durch einfache Aufnahmen", *Upsala lak. forh.*, vol. 32, pp. 338-48, 1927.

[Nordenson30]
Nordenson, J. W., "Ueber stereoskopiche Photographie des Augenhintergrundes", *Upsala lak. forh.*, vol. 35, pp. 216-22, 1930.

[Norton53]
Norton, H. J. Jr., "Absolute Three Dimensional Colored Retinal Photographs", *Trans. Am. Acad. Ophth.*, vol. 57, pp. 612-13, 1953.

[Ohta85]
Ohta, Y. and T. Kanade, "Stereo by Intra- and Inter-Scanline Search Using Dynamic Programming", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-7, no. 2, pp. 139-154, 1985.

[Olsen86]
Olsen, S. I., "Concurrent Solution of the Stereo Correspondence Problem and the Surface Reconstruction Problem", *Proceedings, 8th Int. Conf. Pattern Recognition*, pp. 1038-1040, 1986.

[Oppenheim75]
Oppenheim, A. V. and R. W. Schafer, *Digital Signal Processing*, Prentice-Hall, Englewood Cliffs, New Jersey, 1975.

[Otto89]

Otto, G. P. and T. K. W. Chau, "A "Region-Growing" Algorithm for Matching of Terrain Images", *Image and Vision Computing*, vol. 7, pp. 83-94, 1989.

[Parthasarathy86]

Parthasarathy, G. and B. N. Chatterji, "The use of data windows in feature extraction for high-dimensional PR problems", *Pattern Recognition Letters*, vol. 4, pp. 25-30, 1986.

[Peli87]

Peli, E., R. A. Augliere, and G. T. Timberlake, "Feature-Based Registration of Retinal Images", *IEEE Trans. Medical Imaging*, vol. MI-6, no. 3, pp. 272-278, 1987.

[Peli89]

Peli, E., "Electro-optic Fundus Imaging", *Survey of Ophthalmology*, vol. 34, no. 2, pp. 113-122, 1989.

[Pirenne67]

Pirenne, M. H., *Vision and the Eye*, Science Paperbacks, Chapman and Hall Ltd., London, England, 1967.

[Plesch87]

Plesch, A., U. Klingbeil, and J. Bille, "Digital Laser Scanning Fundus Camera", *Applied Optics*, vol. 26, no. 8, pp. 1480-1486, 1987.

[Pollard85]

Pollard, S. B., J. E. W. Mayhew, and J. P. Frisby, "PMF: A Stereo Correspondence Algorithm Using a Disparity Gradient Limit", *Perception*, vol. 14, pp. 449-470, 1985.

[Pollard87]

Pollard, S. B., J. Porrill, J. E. W. Mayhew, and J. P. Frisby, "Matching Geometrical Descriptions in Three-Space", *Image and Vision Computing*, vol. 5, no. 2, pp. 73-78, 1987.

[Pratt78]

Pratt, W. K., *Digital Image Processing*, Wiley-Interscience, New York, 1978.

[Prince87]

Prince, A. M. and R. Ritch, "Reproducibility of Rodenstock Optic Nerve Analysis in Eyes with Different Cup-Disc Ratios", *Invest. Ophthal. Vis. Sci. (ARVO suppl.)*, 1987.

[Ralston85]
Ralston, A. and P. Rabinowitz, *A First Course in Numerical Analysis*, McGraw Hill Book Co., London, England, 1985.

[Rosenfeld82]
Rosenfeld, A. and A. C. Kak, *Digital Image Processing*, 2nd Edition, vol. 2, Academic Press, New York, 1982.

[Rosenholm87a]
Rosenholm, D., "Multi-Point Matching Using the Least-Squares Technique for Evaluation of Three Dimensioanl Models", *Photogrammetric Engineering and Remote Sensing*, vol. 53, no. 6, pp. 621-626, 1987.

[Rosenholm87b]
Rosenholm, D., "Least-squares matching method: some experimental results", *Photogrammetric Record*, vol. 12, pp. 493-512, 1987.

[Rosenthal77]
Rosenthal, A. R., M. S. Kottler, D. D. Donaldson, and D. G. Falconer, "Comparative Reproducibility of the Digital Photogrammetric Procedure Utilyzing Three Methods of Stereophotography", *Invest. Ophthal. Vis. Sci.*, vol. 16, no. 1, pp. 54-60, 1977.

[Rosenthal80]
Rosenthal, A. R., D. G. Falconer, and I. Pieper, "Photogrammetry Experiments with a Model Eye", *British Journal of Ophthalmology*, vol. 61, pp. 881-887, 1980.

[Ryan80]
Ryan, T. W., R. T. Gray, and B. R. Hunt, "Prediction of Correlation Errors in Stereo-Pair Images", *Optical Engineering*, vol. 19, no. 3, pp. 312-322, 1980.

[Ryan81]
Ryan, T. W. and B. R. Hunt, "Recognition of Stereo-Image Cross-Correlation Errors", in *Progress in Pattern Recognition*, vol. 1, by L. N. Kanal and A. Rosenfeld, North-Holland Publishing Co., Amsterdam, 1981.

[Scheuing86]
Scheuing, A. and H. Niemann, "Computing Depth From Stereo Images using Optical Flow", *Pattern Recognition Letters*, vol. 4, pp. 205-212, 1986.

[Sclove81]
Sclove, S. L., "Pattern Recognition in Image Processing Using Interpixel Correlation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-3, pp. 206-208, 1981.

[Secilla88]

Secilla, J. P., N. Garcia, and J. L. Carrascosa, "Template Location in Noisy Pictures", *Signal Processing*, vol. 14, pp. 347-361, 1988.

[Shapiro87]

Shapiro, J. M. and K. S. Bush, "New Developments in the Analysis of the Optic Nerve Head Topography: Calibration, Automation, Ray Tracing", *Invest. Ophthal. Vis. Sci. (ARVO suppl.)*, 1987.

[Shields87]

Shields, M. B., J. F. Martone, A. R. Shelton, A. R. Ollie, and J. MacMillan, "Reproducibility of Topographic Measurements with the Optic Nerve Head Analyzer", *American Journal of Ophthalmology*, vol. 104, no. 6, pp. 581-586, 1987.

[Shirai87]

Shirai, Y., *Three-Dimensional Computer Vision*, Springer-Verlag, Berlin, 1987.

[Spaeth87]

Spaeth, G. L. and R. Varma, "Optic Disc Vessel Shift in Glaucoma: Image Analysis Versus Clinical Evaluation", *Invest. Ophthal. Vis. Sci. (ARVO suppl.)*, 1987.

[Sutton83]

Sutton, M. A., "Determination of Displacements using an Improved Digital Correlation Method", *Image and Vision Computing*, vol. 1, no. 3, pp. 133-139, 1983.

[Takamoto79]

Takamoto, T., B. Schwartz, and G. T. Marzan, "Stereo Measurement of the Optic Disc", *Photogrammetric Engineering and Remote Sensing*, vol. 45, no. 1, pp. 79-85, 1979.

[Therien84]

Therien, C. W. and K. Fukunaga, "Properties of Separable Covariance Matrices and Their Associated Gaussian Random Processes", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-6, pp. 652-656, 1984.

[Thomason87]

Thomason, M. G. and E. Granum, "Sequential inference of Markov networks by dynamic programming for structural pattern recognition", *Pattern Recognition Letters*, vol. 5, pp. 31-39, 1987.

[Thorner09]

Thorner, W., "Die stereoskopiche Photographie des Augenhintergrundes", *Klin. Monatbl. Augenh.*, vol. 47, pp. 481-490, 1909.

[Toriwaki78]

Toriwaki, J. and T. Fukumura, "Extraction of Structural Information from Grey Pictures", *Computer Graphics and Image Processing*, vol. 7, pp. 30-51, 1978.

[Trevedi85]

Trevedi, H. P. and S. A. Lloyd, "The Role of Disparity Gradient in Stereo Vision", *Perception*, vol. 14, pp. 685-690, 1985.

[Ullman79]

Ullman, S., *The Interpretation of Visual Motion*, M.I.T. Press, Cambridge, Ma., 1979

[VanTrees68]

Van Trees, H. L., *Detection, Estimation, and Modulation Theory, Part I*, John Wiley & Sons, Inc. New York, 1968.

[Varma87]

Varma, R., G. R. Douglas, and G. L. Spaeth, "A Comparative Study of Three Methods of Analysis of Optic Disc Topography", *Invest. Ophthal. Vis. Sci. (ARVO suppl.)*, 1987.

[Webb87]

Webb, R. H., G. W. Hughes, and F. C. Delori, "Confocal Scanning Laser Ophthalmoscope", *Applied Optics*, vol. 26, no. 8, pp. 1492-1499, 1987.

[Whalen71]

Whalen, A. D., *Detection of Signals in Noise*, Academic Press, London, England, 1971.

[Whiteside86]

Whiteside, S., S. Mitra, T. Krile, and Z. Shihab, "Microcomputer Based Fundus Image Processing in RNFL Assessment", *Invest. Ophthal. Vis. Sci. (ARVO suppl.)*, vol. 27, 1986.

[Wolf83]

Wolf, P. R., *Elements of Photogrammetry*, McGraw-Hill Book Company, London, England, 1983.

[Wong77]

Wong, A. K. C. and M. A. Vogel, "Resolution-Dependent Information Measures for Image Analysis" *IEEE Trans. on Systems, Man, and Cybernetics*, vol. SMC-7, no. 1, pp. 49-61, 1977.

[Woodham78]
Woodham, R. J., "Photometric Stereo: A Reflectance Map Technique for Determining Surface Orientation for Image Intensity", *Proc. SPIE 22nd Technical Symposium*, vol. 155, pp. 136-143, 1978.

[Woon90]
Woon, H. W., F. W. Fitzke, G. H. Chester, D. G. Greenwood, and J. Marshall, "The Scanning Laser Ophthalmoscope: Basic Principles and Applications", *Journal of Ophthalmic Photography*, vol. 12 no. 1, pp. 17-23, 1990.

[Xu87]
Xu, G., S. Tsuji, and M. Asada, "A Motion Stereo Method Based on a Coarse-to-Fine Control Strategy", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-9, 1987.

[Yakimovsky78]
Yakimovsky, Y. and R. Cunningham, "A System for Extracting Three-Dimensional Measurements from a Stereo Pair of TV Cameras", *Computer Graphics and Image Processing*, vol. 7, pp. 195-210, 1978.

[Yokoyama78]
Yokoyama, R. and R. M. Haralick, "Texture Synthesis Using a Growth Model", *Computer Graphics, Vision, and Image Processing*, vol. 8, pp. 369-380, 1978.

[Zimmermann88]
Zimmermann, J., "Computer-Eyed", *Physicians Video Guide*, pp. 4ff, Jan./Feb. 1988.

THE STRUCTURE OF AUTO- AND CROSS-CORRELATION FUNCTIONS

Report No. 90/1

M.R.B. Forshaw & D.G. Greenwood

Department of Physics and Astronomy
University College London
Gower Street, London WC1E 6BT

ABSTRACT

Formulae are derived for the shape of the correlation surface when a template image is matched against a scene. Two correlation algorithms are considered, and formulae are given for three types of scene, namely those where the spatial frequency spectra have rectangular, linear fall-off and 1/f shapes. Formulae are also given for one-dimensional data.

13 June 1990

1. INTRODUCTION

Consider a continous, two-dimensional function F(x,y), which might for example be the apparent brightness of a scene. In image processing, and indeed in other areas of science and technology, it frequently happens that a sampled version of F(x,y) is obtained, and that this sampled version, which we shall denote by f(x,y), has to be examined to see if a portion of the sampled data corresponds to a previously stored reference pattern or template, which we shall denote by g(x,y). The size of the template is usually much smaller than the size of f(x,y): for example f(x,y) might be non-zero over 512 by 512 pixels, while g(x,y) might be of size 32 by 32 pixels. The process of searching f(x,y) to see if a portion of it corresponds to g(x,y) is often called *template matching*. A number of algorithms exist for carrying out the matching: most of them are based on two or three variations on the theme of *correlation*.

In correlation the template is moved across the scene and some measure of the match is derived. This correlation measure can be thought of as a two-dimensional surface: the hope is that there exists a well-defined peak, or perhaps a minimum, which represents the location in f(x,y) where a near-copy of g(x,y) may be found. If g(x,y) is identical to a part of f(x,y) then the process can be considered to be *auto-correlation*, though not in the strict mathematical sense, which requires that the template and scene should be identical. If g(x,y) is known to differ in some way from all or part of f(x,y) then the process should strictly be called *cross-correlation*.

The hope of finding a perfect match is often not fulfilled in practice. A copy or near-copy of g(x,y) may not be present in f(x,y); the correlation algorithm and the structure of the functions f(x,y) and g(x,y) may be such that false correlation maxima (or minima) may be larger than the 'correct' match; noisy data may distort the shape of the peak; or the scene may have so little structure that the correlation maximum is excessively broad.

Although it is obviously important to be able to predict the performance of a correlation algorithm for a specific type of scene and template, relatively few theoretical predictions have been published. The correlation functions of elementary types of scene, such as isolated square objects on a uniform field, can obviously be calculated with little difficulty. However, theoretical analyses have usually been confined to more general discussions of unspecified scene structures (see e.g. [De Soete], [Dvornychenko], [Parthasarathy], [Sclove], [Therrien & Fukunaga]). At the other extreme, much of the experimental work which we have located has been either heuristic or empirical in nature (e.g. [de Bougrenet], [Gruen a], [Caelli], [Chan], [Lee & Lei], [Rosenholm a, b]). This is unfortunate: on the one hand, it is sometimes difficult to predict the performance of a particular algorithm from its theoretical description, and on the other hand it is often equally difficult to extrapolate from a limited set of experimental results. It is the purpose of this report to derive explicit formulae which permit relatively simple predictions of the performance of two standard cross-correlation algorithms in terms of a few easily-definable scene parameters. Formulae are also derived for the one-dimensional case. The contents of this report apply only to noise-free scenes: an analysis of the effects of noise will appear in Chapter 7 of D.G.'s PhD thesis.

The report contains a brief review of certain correlation algorithms (section 2). After a description of how the scene structure is defined (section 3), the one-dimensional auto-correlation case is treated first (section 4), followed by an analysis of two-dimensional auto-correlation (section 5). The extensions to cross-correlation are then considered (section 6).

2. CROSS-CORRELATION ALGORITHMS

We assume that f(x,y) has non-zero values only over the range $-N/2 \leq x \leq N/2$, $-N/2 \leq y \leq N/2$, that g(x,y) is non-zero over the range $-M/2 \leq x \leq M/2$, $-M/2 \leq y \leq M/2$, and that $M < N$. The two correlation algorithms which we shall be concerned with are defined as follows:

$$C(d,e) = \sum_{y=-M/2}^{+M/2} \sum_{x=-M/2}^{+M/2} g(x,y) f(x+d,y+e) \tag{1}$$

where x and y are initially assumed to take only integer values and d and e are the displacements of the template origin relative to that of the scene, and

$$C_{ssd}(d,e) = \sum_{y=-M/2}^{+M/2} \sum_{x=-M/2}^{+M/2} (g(x,y) - f(x+d,y+e))^2 \tag{2}$$

The subscript 'ssd' stands for 'sum of squares of differences' and the displacements d and e are assumed to be sufficiently small that the template never falls outside the non-zero range of the scene. Expression (1) is the most-frequently used algorithm, or rather the one which is most frequently described. Under ideal circumstances it produces a correlation peak for the particular values of d and e where a copy of the template can be found in the scene, whereas expression (2) produces a correlation minimum. Expression (2) can be called the minimum-error, minimum-square-error or least-squares algorithm (see e.g. [Lee & Kim]), though it is sometimes misnamed the mean-square-difference or mean-square-error algorithm. There may, of course, be several copies or near-copies of the template in the scene, in which case there will be several maxima or minima. Scaled or normalised versions exist for both of these algorithms, but we will not deal with these versions here, or with other correlation algorithms (see e.g. [Rosenfeld & Kak], section 9.4.1., [Pratt], section 19.1, [Griffin]).

If we expand (2) we obtain

$$C_{ssd}(d,e) = \sum_{y=-M/2}^{+M/2} \sum_{x=-M/2}^{+M/2} g^2(x,y) - 2f(x,y)g(x,y) + f^2(x,y) \tag{3}$$

The first and last terms in (3) are related to the local average values of the template and scene. In many cases we may approximate them by constants, and this is what we shall do in this report. Thus the variable term is the second, which can be seen to be identical to $C(d,e)$, apart from a factor of -2. For the particular case when $f(x,y)$ and $g(x,y)$ are equal then $C_{ssd}(d,e)$ has the value of zero when c and d are both zero. In this instance, and with the assumption about the constancy of the $f^2(x,y)$ and $g^2(x,y)$ terms, $C_{ssd}(d,e)$ reduces to

$$C_{ssd}(d,e) \approx 2C(0,0) - 2C(d,e) \tag{4}$$

We shall use this approximation throughout this report in order to derive expressions for $C_{ssd}(d,e)$ from $C(d)$. The practical advantage of $C_{ssd}(d,e)$ over $C(d,e)$ is that it is less affected by large DC scene components (the *covariance* correlation measure also avoids the effect of DC biases, by calculating the mean, then removing it - see e.g.

[Keating]). Thus, provided that descriptions which are easy to sum or to integrate can be found for g and f, so that we can derive an expression for $C(d,e)$, then we can immediately write down the (approximate) expression for $C_{ssd}(d,e)$.

It is the aim of this work, to derive analytical expressions for the shape of $C(d,e)$ and $C_{ssd}(d,e)$ in the vicinity of the "ideal" correlation maximum or minimum, as a function of the following parameters:

the size of the scene f(x,y), assumed to be N by N pixels square,

the size of the template g(x,y), assumed to be M by M pixels square ($M < N$),

a(r,s), the spatial frequency amplitude spectrum of the scene, where r and s are the spatial frequencies in the x and y directions,

The effects of noise are not discussed in this report: they will be considered in D.G.'s PhD thesis.

It is important to note that our particular aim is to derive formulae which are suitable for describing the performance of a more complicated scene-matching algorithm for use in stereophotogrammetry. The algorithm is a least-squares cross-correlation algorithm, and has been developed by a number of workers - see e.g. [Gruen], [Otto]. It is not our purpose to describe this algorithm in detail, but it is relevant to note that one important feature of this algorithm rests on the assumption that it is possible, under favourable conditions, to obtain *sub-pixel registration accuracy* using correlation methods. In the present context, this implies that the correlation functions $C(d,e)$ and $C_{ssd}(d,e)$, which strictly speaking exist only for integer values of the displacements d and e, have a valid meaning for non-integer values of d and e, particularly in the vicinity of the correlation maximum or minimum: this point is discussed further in Section 3 which follows. In order to obtain formulae which are manageable and not over-burdened by superfluous symbols, we shall make various simplifying assumptions or approximations, which will be discussed as the derivation proceeds. It may comfort the reader to know that these approximations appear to be justified by the good-to-excellent agreement which has been obtained between the theoretical predictions and experimental results.

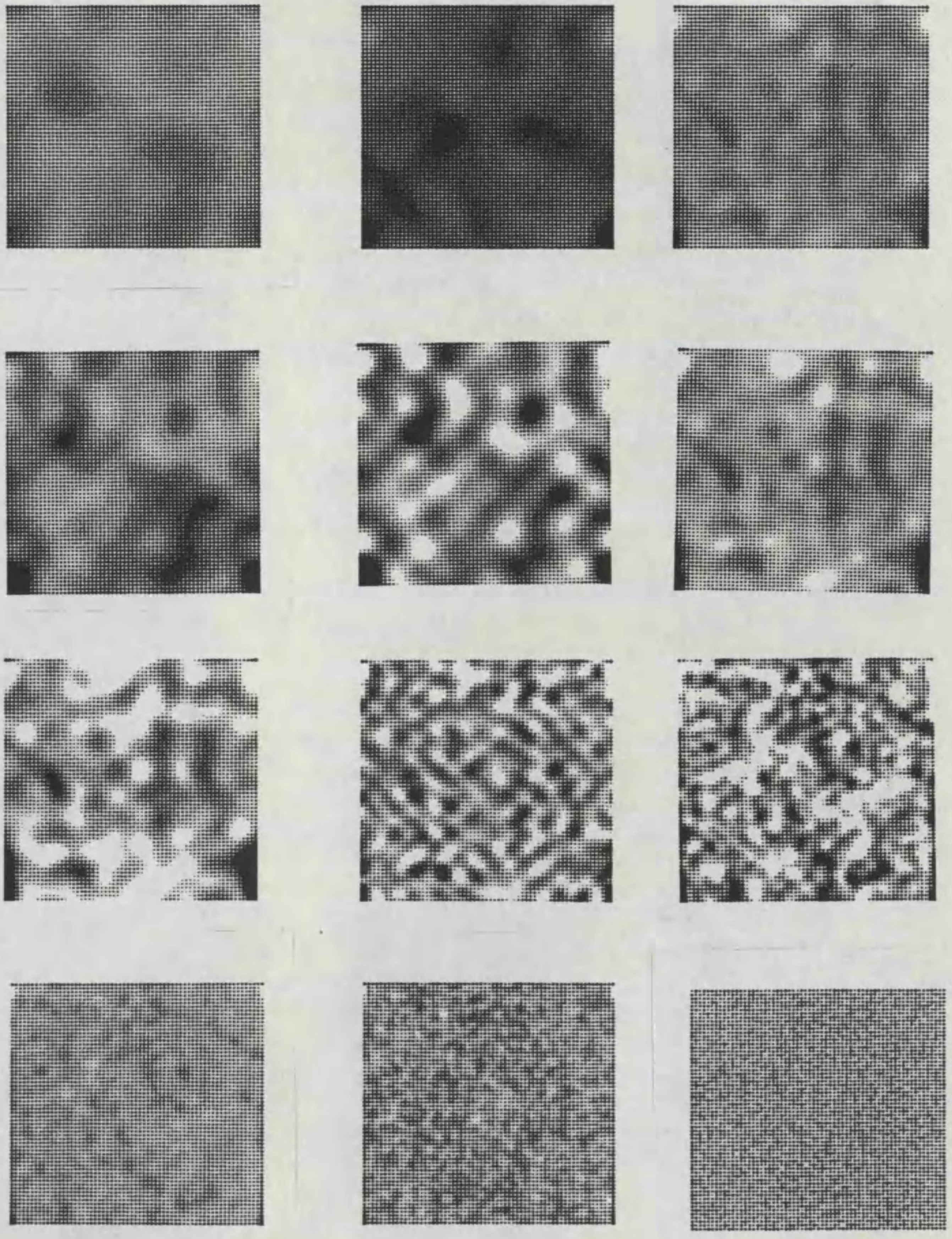


FIGURE 1 Some examples of scene texture.

3. THE TEMPLATE AND THE SCENE

In this report we are particularly interested in scenes which have random texture, in the sense that any part of the scene may be expected *a priori* to have the same statistics as any other part of the scene. We shall assume that the scene, F(x,y), is a continuous function of x and y, and that it can be described by an expression with the following form:

$$F(x,y) = \sum_{s=0}^{s=n_c} \sum_{r=-n_c}^{r=n_c} a(r,s) cos(\frac{2\pi}{N}(rx+sy)+\phi(r,s)) \quad (5)$$

where r and s, the spatial frequencies in the x and y directions, take on only discrete integer values over the scene width (one cycle per image width, two cycles per image width, etc.). The phases $\phi(r,s)$ are assumed to be randomly distributed over the range 0 to 2π. The detailed form of the spatial frequency amplitude spectrum, a(r,s), can be left unspecified for the time being, except that we assume that it is real, and that a(r,s) = a(-r,-s), although $\phi(r,s)$ will not in general equal $\phi(-r,-s)$. Later, we shall choose forms for a(r,s) which are representative of some scene spectra. Note that the ranges of r and s do not cover all four quadrants, because of the assumed symmetry of a(r,s). Specification of $a(r,s)$ over all four quadrants is superfluous and merely results in expressions for the correlation coefficients which are twice as big as those given in this report.

One of the first assumptions which we shall make is that a(r,s) is non-zero over a relatively large range of values of r and s, that is, we implicitly assume that the scene has a fairly complex structure and that the template has the same sort of local texture as the scene. This is roughly equivalent to the assumption that the scene is stationary in a statistical sense. The reason for this assumption is that, if we take as template a small but arbitrary portion of this scene, M by M pixels in extent, then this template will have less structure than the whole scene (because it is smaller and therefore has more low-frequency components relative to its size). The correlation surface $C(d,e)$ or $C_{ccs}(d,e)$ which results from the particular choice of template may, in individual cases, have a very peculiar, atypical shape, depending on the details of the template. The formulae which we shall derive can be looked on in two ways, *either* as the limiting case when M and N are large and the scene is very complicated, *or* as the expectation values which will be obtained with smaller values of N and M, and with less busy scenes, but taken over a very large number of repeated correlation trials with random choices of template and scene. In the latter case we could expect to find that individual realisations of scene and template may produce a correlation surface which departs significantly from the theoretical predictions. A more detailed discussion of the properties of certain scenes will be given elsewhere [DG thesis].

The scene function $F(x,y)$ is continuous in x and y, whereas its sampled counterpart $f(x,y)$ is strictly non-zero only at discrete values of x and y, namely the sampling points. However, in the remainder of this report we shall treat the functions $f(x,y)$ and $g(x,y)$ (and their one-dimensional counterparts) as if they were continuous functions. Our justification for this is based on two assumptions.

First, it is frequently the case that scene structures are relatively highly-correlated from point to point - see for example [Pratt, pp132,565]. This is not always true: for example, a scene may contain small structures against a background with a different texture. There will then be discontinuities in the correlation between neighbouring points across the boundaries of the small structures. However it is sometimes the case, even here, that the template and scene will contain such a large number of the small structures that they may be considered merely as 'texture' which is uniformly distributed over the whole scene. Furthermore, it is often the case that significant correlation exists over a few pixels, due perhaps to image blurring or to signal degradation in the signal-processing chain. In the present instance, we are interested in the case where the template represents a slightly-distorted version of part of the scene. The distortion may be due to rotation, translation, shear, expansion or some more general affine or projective transformation. The underlying problem is to search through the range of possible distortions and to find the best possible match, preferably to within a small fraction of a pixel. This sub-pixel search can only be carried out by resampling the data, using for example bilinear interpolation. Such interpolation is only meaningful if some sort of correlation exists between pixels. References [Sclove] and [Dvornychenko] contain theoretical discussions of some of the effects associated with inter-pixel correlation. We shall assume in this report that a meaningful interpolation can be made between the discrete samples of $f(x,y)$ and $g(x,y)$.

Our second assumption is that real data are always affected to some extent by noise. When the noise is relatively large, the correlation surface will be so corrupted that sub-pixel location of the correlation maximum or minimum is almost impossible, unless the template is very large. There is therefore a regime, where it is almost irrelevant that the data are sampled: the best that one can hope to do is to obtain an estimate of the probability that the correlation maximum lies within a specified set of limits. Under these circumstances it is permissable (with only a moderate amount of hand-waving) to treat the sampled functions $f(x,y)$ and $g(x,y)$ as if they were continuous variables. Although this report does not consider the effects of noise, the inclusion of noise effects is conveniently carried out by using the assumption of continuity.

The reader may have noticed that we have chosen to define the scene structure in terms of its spatial frequency spectrum and not as a Markov random field (MRF). It is our opinion that in many practical cases (though not always: see e.g. [Thomason]) a better understanding and analysis of the correlation function can be obtained using the methods outlined in this report. On the other hand, if one is interested in texture analysis, then MRF methods are perhaps to be preferred over the present approach.

One consequence of our assumption that the discrete sampled functions $f(x,y)$ and $g(x,y)$ can be treated as continuous, is that we can frequently replace summations by integrals. Swapping between summations and integrals is carried out at several points in this report where it appears convenient to do so. The reader is assured that this cavalier treatment could be justified at excessive length, if justification were needed.

4. THE 1-D AUTOCORRELATION FUNCTION

In this section we derive expressions for the one-dimensional autocorrelation functions $C(d)$ and $C_{ssd}(d)$. This is done for two reasons: first, because the 1-D case is interesting in its own right (see e.g. [Oppenheim & Schafer], sections 11.5 & 11.6), and second, because we shall use the 1-D case to develop expressions for the 2-D case. Here we consider only the *autocorrelation* case, where the template is assumed to be identical to some portion of the scene.

Given a one-dimensional scene F(x):

$$F(x) = \sum_{n=0}^{n=n_c} a(n)\cos(px + \phi_n) \tag{6}$$

where $p = 2\pi/N$, we may approximate the one-dimensional autocorrelation function $C(d)$ by

$$C(d) \approx \int_{-M/2}^{M/2} \sum_{m=0}^{m=n_c} a(m)\cos(mpx+\phi_m) \sum_{n=0}^{n=n_c} a(n)\cos(np(x+d)+\phi_n)dx \tag{7}$$

Re-arranging the sums gives

$$C(d) = \int_{-M/2}^{M/2} \sum_{m=0}^{m=n_c} a(m) \sum_{n=0}^{n=n_c} a(n)\cos(mpx+\phi_m)\cos(np(x+d)+\phi_n)dx \tag{8}$$

If we decompose the double summation we obtain

$$\begin{aligned} C(d) = & \int_{-M/2}^{M/2} \sum_{n=0}^{n=n_c} a^2(n)\cos(npx+\phi_n)\cos(np(x+d)+\phi_n)dx \\ & + \int_{-M/2}^{M/2} \sum_{\substack{m=0 \\ m\neq n}}^{m=n_c} a(m) \sum_{n=0}^{n=n_c} a(n)\cos(mpx+\phi_m)\cos(np(x+d)+\phi_n)dx \end{aligned} \tag{9}$$

For n_c sufficiently large, and for d sufficiently small ($d \ll N/n_c$), *and* for ϕ uniformly distributed between 0 and 2π, we may neglect the second set of terms in comparison with the first: its magnitude will be of order $O(n_c < |a| >)$ in comparison with the first term's $O(Mn_c < a^2 >)$. To show this we note that, for small values of npd, the $cos.cos$ term in the first integral will integrate to an approximate value of M/2. The summation over n yields $n_c < a^2 >$. Now consider the second integral. Because the phase terms are assumed to be randomly distributed, the $cos.cos$ term here will in general integrate to the much smaller value of $O(1)$. The double summation of $a(m)a(n)$ over n and m yields a quantity whose r.m.s. value will be of order $O(2^{1/2} n_c < |a| >)$, and it follows that the second set of terms may be neglected in comparison with the first. We therefore obtain the approximate expression

$$C(d) \approx \int_{-M/2}^{M/2} \sum_{n=0}^{n=n_c} a^2(n)\cos(npx+\phi_n)\cos(np(x+d)+\phi_n)dx \tag{10}$$

Using the relation $\cos(a+b) = \cos(a)\cos(b) - \sin(a)\sin(b)$ we obtain

$$\begin{aligned} C(d) \approx & \sum_{n=0}^{n=n_c} a^2(n) \int_{-M/2}^{M/2} cos(npx+\phi_n) cos(npx+\phi_n) cos(npd) dx \\ & - \sum_{n=0}^{n=n_c} a^2(n) \int_{-M/2}^{M/2} cos(npx+\phi_n) sin(npx+\phi_n) sin(npd) dx \end{aligned} \tag{11}$$

$$\begin{aligned} = & \sum_{n=0}^{n=n_c} a^2(n) cos(npd) \int_{-M/2}^{M/2} \cos^2(npx+\phi_n) dx \\ & - \sum_{n=0}^{n=n_c} a^2(n) sin(npd) \int_{-M/2}^{M/2} cos(npx+\phi_n) sin(npx+\phi_n) dx \end{aligned} \tag{12}$$

We make the approximation that the expectation value of the $\cos^2$ term is 1/2 over the range of integration, except for the DC term n = 0, for which we must use a value of unity. This is a good approximation for large values of M and n_c, but it becomes progressively poorer as M and/or n_c tend to unity. In such circumstances the comments in the introduction, namely that the resulting formulae represent average values over many trials, apply. Similarly, we approximate the $sin.cos$ term by its expectation value of zero. This yields

$$C(d) \approx Ma^2(0) + M/2 \sum_{n=1}^{n=n_c} a^2(n) cos(npd) \tag{13}$$

which may be further approximated by

$$C(d) \approx Ma^2(0) + M/2 \int_{n=1}^{n=n_c} a^2(n) cos(npd) dn \tag{14}$$

4.1. Specific Scene Structures

In this section we shall derive expressions for C(d) when the scene is described by specific formulae which represent cases which are often found experimentally. In particular, we consider scene spectra given by the following formulae:

(1) $a(n) = b \ , \ 1 \le n \le n_c;$

$$a(0) = B \tag{15a}$$

(2) $a(n) = b(1 - |n|/n_c), \ 1 \le n \le n_c;$

$$a(0) = B \tag{15b}$$

(3) $a(n) = b/n \ , \ 1 \le n < n_c;$

$$a(0) = B \tag{15c}$$

The first describes a spectrum which is flat out to a cut-off frequency n_c, except for the DC component, which has an amplitude B (the inclusion of a distinct value for the DC term permits us to consider the all-too-frequent cases where the scene structure is relatively low-contrast and the background level is high). The second describes a spectrum whose amplitudes decrease linearly to the cutoff frequency: for this spectrum also, a DC component of amplitude B is included. The third spectrum follows a $1/f$ law between

$f = 1$, where its amplitude is b, and $n = n_c$, the cutoff frequency (note that n_c is *not* the Nyquist frequency). The DC component in this example is once again chosen to have an amplitude B. In all of these cases we assume that the phases of the individual components are randomly distributed between 0 and 2π.

Case 1: Flat Spectrum

The formulae for $C(d)$ is readily derivable for this case:

$$C(d) = MB^2 + M/2 \int_{n=1}^{n=n_c} b^2 \cos(npd)dn \tag{16}$$

$$= MB^2 + \frac{Mb^2}{2pd}\left[\sin(n_c pd) - \sin(pd)\right] \tag{17}$$

Although this expression is $= MB^2 + \frac{Mb^2}{2pd}\left[\sin(n_c pd) - \sin(pd)\right]$ hich are obtained with other scene spectra are not so understandable. It is also the case that the expressions for $C(d)$ become progressively more inaccurate away from the peak. Since we are mainly interested in the shape of the curve near the peak, we expand the expression as a power series in d:

$$C(d) \approx MB^2 + \frac{Mb^2}{2pd}\left[pdn_c - p^3d^3n_c^3/6 + .. - pd + p^3d^3/6 + ..\right] \tag{18}$$

provided that $1 \ll n_c < N$, and $n_c < N/4d$. If we replace p by $2\pi/N$ and neglect the lower limit of the integral in comparison with n_c, we obtain

$$C(d) \approx M\left[B^2 - b^2 n_c/2\right] - \pi^2 d^2 M b^2 n_c^3/(3N^2) \tag{19}$$

Hence, from expression (4),

$$C_{ssd}(d) \approx 2\pi^2 d^2 M b^2 n_c^3/(3N^2) \tag{20}$$

As one might expect from considerations of symmetry, the approximate expressions have a quadratic variation with d.

Case 2: Linearly-Decreasing Spectrum

It is possible to obtain an exact solution to (14) when the spectrum is given by (15b), but, given the variety of approximations which are involved, it is not particularly meaningful. We therefore consider only the shape of $C(d)$ in the region of the maximum. For this particular scene spectrum, if we approximate $\cos(x)$ by $1 - x^2/2$,

$$C(d) \approx MB^2 + \frac{Mb^2}{2}\int_{n=1}^{n=n_c} 1 - \frac{2n}{n_c} + \frac{n^2}{n_c^2} dn$$

$$- \frac{M}{2}\frac{b^2p^2d^2}{2}\int_{n=1}^{n=n_c} n^2 - \frac{2n^3}{n_c} + \frac{n^4}{n_c^2} dn \tag{21}$$

$$= MB^2 + \frac{Mb^2 n_c}{6} - \frac{Mb^2p^2d^2n_c^3}{120} \tag{22}$$

if we neglect the lower limit of the integral in comparison with n_c. Replacing p by $2\pi/N$ gives

$$C(d) \approx M\left[B^2 + \frac{b^2 n_c}{6}\right] - \frac{Mb^2\pi^2 d^2 n_c{}^3}{30N^2} \tag{23}$$

whence, using (4),

$$C_{ssd}(d) \approx \frac{Mb^2\pi^2 d^2 n_c{}^3}{15N^2} \tag{24}$$

Case 3: 1/F Spectrum

In this case the correlation coefficient $C(d)$ is given by

$$C(d) \approx MB^2 + \frac{Mb^2}{2}\int_{n=1}^{n_c} \frac{cos(npd)}{n^2} dn \tag{25}$$

Changing variables in the integral yields

$$\int dn = pd\int_{pd}^{n_c pd} \frac{cos(x)}{x^2} dx \tag{26}$$

$$= pd\left[\frac{-cos(x)}{x} - \int \frac{sin(x)}{x} dx\right]_{pd}^{n_c pd} \tag{27}$$

The integral in (27) is the sine integral $Si()$ (see e.g. Abramowitz & Stegun section 5.2). We shall approximate the integrand, for argument values less than $\pi/2$, by $(1 - x^2/6 + ...)$. We thus obtain

$$\int dn \approx pd\left[-cos(x)/x - x + x^3/18 + \cdots\right]_{pd}^{n_c pd} \tag{28}$$

$$= -pd\left[\frac{cos(n_c pd)}{n_c pd} - \frac{cos(pd)}{pd} + n_c pd - pd - \frac{(n_c pd)^3}{18} + \frac{(pd)^3}{18}\right] \tag{29}$$

$$\approx 1 - (pd)^2 n_c/2 \tag{30}$$

Hence

$$C(d) \approx M(B^2 + b^2/2) - \pi^2 M n_c b^2 d^2/N^2 \tag{31}$$

and

$$C_{ssd}(d) \approx 2\pi^2 M n_c b^2 d^2/N^2 \tag{32}$$

Comparison of Correlation Formulae

It is of interest to note that expressions (20), (24) and (32) have almost the same form, but in going from the flat-spectrum case to the 1/f spectrum case the sensitivity of the correlation to the cut-off frequency n_c varies in the ratio

$$\frac{2n_c^3}{3} \; : \; \frac{n_c^3}{15} \; : \; 2n_c$$

In the flat-spectrum case the dependence on the cut-off frequency is quite strong, as one might expect, while the dependence on n_c is much weaker for the $1/f$ spectrum.

5. THE TWO-DIMENSIONAL AUTOCORRELATION FUNCTION

Consider the two-dimensional image function

$$f(x,y) = \sum_{s=0}^{s=n_c} \sum_{r=-n_c}^{r=n_c} a(r,s)\cos\left(\frac{2\pi}{N}(rx+sy)+\phi_{r,s}\right) \tag{33}$$

where we assume that $a(r,s)$ is real and symmetrical in r and s. Note that the summation over r extends over both positive and negative frequencies, in order to avoid asymmetries in the scene structure. It would be possible to extend the range of the s summation as well, but this would merely result in an overall doubling of the scene amplitude. We shall ignore the fact that, strictly speaking, we should not include frequency components for which $r < 0$ and $s = 0$.

The 2-D autocorrelation function $C(d,e)$ of $f(x,y)$ is:

$$C(d,e) = \int_{-M/2}^{M/2} \int_{-M/2}^{M/2} \sum_{r=-n_c}^{n_c} \sum_{s=0}^{n_c} a(r,s)\cos(rpx+spy+\phi_{r,s})$$

$$\sum_{u=-n_c}^{n_c} \sum_{v=0}^{n_c} a(u,v)\cos(up(x+d)+vp(y+e)+\phi_{u,v})dxdy \tag{34}$$

where r, s, u, and v are spatial frequencies, and $p = 2\pi/N$ as before. As for the 1-D case, we factor out the terms in the summation for which $r = u$ and $s = v$. Only this group of terms is retained:

$$C(d,e) \approx \int_{-M/2}^{M/2} \int_{-M/2}^{M/2} \sum_{r=-n_c}^{n_c} \sum_{s=0}^{n_c} a^2(r,s)\cos(rpx+spy+\phi_{r,s})$$

$$\cos(rp(x+d)+sp(y+e)+\phi_{r,s})dxdy \tag{35}$$

since the remaining terms are of order $n_c^2 < |a| >$ in comparison with (35), which is of order $M^2 n_c^2 < a^2 >$ (see the argument for the 1-D case). Re-arranging the summations and integrations in (35) gives:

$$C(d,e) \approx \sum_{r=-n_c}^{n_c} \sum_{s=0}^{n_c} a^2(r,s) \int_{-M/2}^{M/2} \int_{-M/2}^{M/2} \cos(rpx+spy+\phi_{r,s})$$

$$\cos(rp(x+d)+sp(y+e)+\phi_{r,s})dxdy \tag{36}$$

We now expand the cosine terms in the double integral, first by the rpd components:

$$\iint dxdy = \int_{-M/2}^{M/2} \int_{-M/2}^{M/2} \cos(rpx+spy+\phi_{r,s})$$

$$\cos(rp(x)+sp(y+e)+\phi_{r,s})\cos(rpd)dxdy$$

$$-\int_{-M/2}^{M/2} \int_{-M/2}^{M/2} \cos(rpx+spy+\phi_{r,s})$$

$$\sin(rp(x)+sp(y+e)+\phi_{r,s})\sin(rpd)dxdy \tag{37}$$

The equivalent *spe* expansion yields

$$
\begin{aligned}
\iint dxdy \;=\; & \\
& cos(rpd)\,cos(spe)\int_{-M/2}^{M/2}\int_{-M/2}^{M/2}\cos(rpx+spy+\phi_{r,s})\cos(rpx+spy+\phi_{r,s})dxdy \\
& -\;cos(rpd)\,sin(spe)\int_{-M/2}^{M/2}\int_{-M/2}^{M/2}\cos(rpx+spy+\phi_{r,s})\sin(rpx+spy+\phi_{r,s})dxdy \\
& -\;sin(rpd)\,cos(spe)\int_{-M/2}^{M/2}\int_{-M/2}^{M/2}\cos(rpx+spy+\phi_{r,s})\sin(rpx+spy+\phi_{r,s})dxdy \\
& -\;sin(rpd)\,cos(spe)\int_{-M/2}^{M/2}\int_{-M/2}^{M/2}\cos(rpx+spy+\phi_{r,s})\cos(rpx+spy+\phi_{r,s})dxdy
\end{aligned}
\tag{38}
$$

which reduces to

$$
\begin{aligned}
\iint dxdy \;=\; & cos(rpd+spe)\int_{-M/2}^{M/2}\int_{-M/2}^{M/2}\cos^2(rpx+spy+\phi_{r,s})dxdy \\
& -\;sin(rpd+spe)\int_{-M/2}^{M/2}\int_{-M/2}^{M/2}\cos(rpx+spy+\phi_{r,s})\sin(rpx+spy+\phi_{r,s})dxdy
\end{aligned}
\tag{39}
$$

We make the approximation that the first double integral in (39) can be replaced by $M^2/2$, except for the DC term r,s = 0, for which we must use a value of M^2. This is a good approximation for large values of M and r and s, but it becomes progressively poorer as M and/or r , s tend to unity. Similarly, we approximate the *sin.cos* term by its average value of zero. On replacing the sums over r and s, this yields

$$
C(d,e) \approx M^2a(0,0) + \frac{M^2}{2}\sum_{r=-n_c}^{n_c}\sum_{s=1}^{n_c}a^2(r,s)cos(rpd+spe) \tag{40}
$$

As in the 1-D case, we approximate the summation by an integral:

$$
C(d,e) \approx M^2a(0,0) + \frac{M^2}{2}\int_{r=-n_c}^{n_c}\int_{s=1}^{n_c}a^2(r,s)cos(rpd+spe)dx \tag{41}
$$

We now make the assumption that $a(r,s)$ is a separable function of x and y, and further, that the x and y functions are identical (differences in scale along the two axes could easily be incorporated if needed). The assumption of separability is often used (see e.g. [Pratt] for some experimental results and [Therrien & Fukunaga] for a discussion of some of the properties of separable covariance matrices). In the present case we write:

$$
a(r,s) = a(r)a(s) \tag{42}
$$

where the reader will be aware that the a functions on the LHS and RHS are not identical. We expand the cosine term:

$$C(d,e) \approx$$

$$M^2a(0,0) + \frac{M^2}{2}\int_{s=1}^{n_c} a^2(s)\cos(spe)\left[\int_{r=-n_c}^{n_c} a^2(r)\cos(rpd)dr\right]ds$$

$$-\frac{M^2}{2}\int_{s=1}^{n_c} a^2(s)\sin(spe)\left[\int_{r=-n_c}^{n_c} a^2(r)\sin(rpd)dr\right]ds \qquad (43)$$

Consider just the first part-integral in (43) :

$$I_1 = \int_{r=-n_c}^{n_c} a^2(r)\cos(rpd)dr \qquad (44)$$

which equals

$$2\int_{r=0}^{n_c} a^2(r)\cos(rpd)dr \qquad (45)$$

using considerations of symmetry. A similar consideration of the integral

$$I_2 = \int_{r=-n_c}^{n_c} a^2(r)\sin(rpd)dr \qquad (46)$$

shows that its value is zero. Hence

$$C(d,e) \approx M^2 + M^2\int_{s=1}^{n_c} a^2(s)\cos(spe)\left[\int_{r=0}^{n_c} a^2(r)\cos(rpd)dr\right]ds \qquad (47)$$

At this point, hoping that the reader will bear with us, we shall make yet another approximation. Instead of having a lower limit of $s = 1$ in the integral (47) we shall henceforth replace it by the value $s = 0$. We saw in the 1-D case that we could make this approximation when n_c was large in comparison with unity. It is possible to show that a similar approximation holds for the 2-D case. Leaving the limit $s = 1$ merely leads to extra polynomial terms which will only be neglected in the end.

5.1 SPECIFIC SCENE STRUCTURES

The scene structures which we consider are the two-dimensional counterparts of those for the one-dimensional case:

(1) $a(r,s) = b \;,\; 1 \le r,s \le n_c;$

$a(0,0) = B$ (48a)

(2) $a(r,s) = b(1 - |r|/n_c)(1 - |s|/n_c) \,,\; 1 \le r,s \le n_c;$

$a(0,0) = B$ (48b)

(3) $a(r,s) = b/(rs) \,,\; 1 \le r,s \le n_c;$

$a(0,0) = B$ (48c)

As in the 1-D case, these represent flat, linearly-decreasing and 1/f spectra respectively, at least along the r=0 and s=0 axes. These spectra are all linearly separable, whereas the 'real' spectra shown in Figure 1 are circularly symmetrical.

Case 1. Flat spectrum

Since the spectrum is flat out to the cut-off frequency n_c, it is easy to evaluate $C(d,e)$ exactly (subject to the various assumptions which we have made):

$$C(d,e) \simeq M^2B^2 + M^2b^2 \int_{s=0}^{n_c} cos(spe) \int_{r=0}^{n_c} cos(rpd)\, dr\, ds \qquad (49)$$

$$= M^2B^2 + M^2b^2 \frac{\sin(pdn_c)}{pd} \frac{\sin(pen_c)}{pd} \qquad (50)$$

In the present instance we are chiefly interested in the shape of the function near the correlation peak. Expanding the sin functions yields:

$$C(d,e) \simeq M^2(B^2+b^2n_c^2) - M^2b^2p^2n_c^4\frac{(d^2+e^2)}{6} + O(d^2e^2)/36 \qquad (51)$$

If we replace p by $2\pi/N$ then we obtain

$$C(d,e) \simeq M^2(B^2+b^2n_c^2) - 2\pi^2M^2b^2n_c^4(d^2+e^2)/(3N^2) \qquad (52)$$

Hence

$$C_{ssd}(d,e) \simeq 4\pi^2M^2b^2n_c^4(d^2+e^2)/(3N^2) \qquad (53)$$

Case 2. Linearly-decreasing spectrum

Here we consider the particular case where

$$a(r,s) = b(1-|r|/n_c)(1-|s|/n_c), \qquad r,s \neq 0$$

$$a(0,0) = B \qquad (54)$$

This is equivalent to setting the expression for $a(r)$ equal to

$$a(r) = b^{1/2}(1-|r|/n_c), \qquad r \neq 0$$

$$a(0) = (B\ )^{1/2} \qquad (55)$$

with equivalent expressions for $a(s)$. The reason for using separable expressions like [], rather than a radial function of the form $(1-(r^2+s^2)^{1/2}/n_c)$, is quite simple: it produces much simpler expressions for the correlation functions, and these expressions still agree quite well with experiment near the correlation peak, even for radially symmetric frequency distributions.

For a scene spectrum as specified by (55), and approximating $cos(x)$ by $1-x^2/2$ as for the 1-D case, the correlation function takes the form

$$C(d,e) \approx M^2B^2 + M^2b^2\int_{r=0}^{n_c} a^2(s)\cos(spe)\left[I_3\right] ds \tag{56}$$

where

$$I_3 \approx \int_{r=0}^{n_c} 1 - \frac{2r}{n_c} + \frac{r^2}{n_c^2} - \frac{r^2(pd)^2}{2} + \frac{r^3(pd)^2}{n_c} - \frac{r^4(pd)^2}{2n_c^2}\, dr \tag{57}$$

$$= \frac{n_c}{3} - \frac{n_c^3(pd)^2}{60} \tag{58}$$

The second integral in (56) has the same form as I_3. Thus

$$C(d,e) \approx M^2\left[\left[B^2 + \frac{b^2n_c^2}{9}\right] - \frac{n_c^4b^2p^2(d^2+e^2)}{180} + O(d^2e^2)\right] \tag{59}$$

We ignore the term in d^2e^2, as this simply makes the function tend more nearly to a 'square' function of d and e (in line with the separable-axes approximation), while we are more interested in a circularly-symmetric approximation. Thus $C(d,e)$ becomes

$$C(d,e) \approx M^2\left[\left[B^2 + \frac{b^2n_c^2}{9}\right] - \frac{\pi^2n_c^4b^2(d^2+e^2)}{45N^2}\right] \tag{60}$$

As for the 1-D case, we may deduce the correlation function $C_{ssd}(d,e)$ by inspection:

$$C_{ssd}(d,e) = -2\,(\textit{second term of } (60)) \tag{61}$$

or

$$C_{ssd}(d,e) \approx \frac{2M^2b^2n_c^4\pi^2(d^2+e^2)}{45\,N^2} \tag{62}$$

Case 3. 1/F spectrum

Here we have

$$C(d,e) \approx M^2B^2 + M^2b^2\int_{s=0}^{n_c} \frac{\cos(spe)}{n^2}\left[I_4\right] de \tag{63}$$

where

$$I_4 = \int_{r=0}^{n_c} \frac{\cos(rpd)}{n^2}\, dn \tag{64}$$

From Section 4.1., Case 3, we know the value of I_4, at least in the vicinity of the correlation peak:

$$I_4 \approx 1 - \frac{(pd)^2n_c}{2} \tag{65}$$

Therefore

$$C(d,e) \approx M^2B^2 + M^2b^2\left[1 - \frac{(pd)^2 n_c}{2}\right]\left[1 - \frac{(pe)^2 n_c}{2}\right] \tag{66}$$

$$\approx M^2B^2 + M^2b^2\left[1 - p^2 n_c(d^2 + e^2)/2 + O(d^2e^2)\right] \tag{67}$$

Substituting for p yields

$$C(d,e) \approx M(B^2 + b^2) - 2\pi^2 M_2 b^2 n_c (d^2 + e^2)/N^2 \tag{68}$$

and

$$C_{ssd}(d,e) \approx 4\pi^2 M_2 b^2 n_c (d^2 + e^2)/N^2 \tag{69}$$

6. ONE- AND TWO-DIMENSIONAL CROSS-CORRELATION FORMULAE

When the template is no longer identical to part of the scene then matters become much more complicated. There are many reasons why the two may differ, even if we ignore the effects of sampling and quantisation. For example:

1. The template may be a ‘perfect’ representation of some prototype object, but the scene may be corrupted by noise.

2. Both template and scene may be corrupted by noise. This is often the case in stereophotogrammetry.

3. There may be differences in the DC components between scene and template, there may be some ‘gain factor’ which multiplies $g(x,y)$ but not $f(x,y)$, or there may be some slow secular variations in $g(x,y)$ or $f(x,y)$ which must be taken into consideration.

4. The template may contain a distorted version of part of the scene. This distortion may be due to an affine transformation of the template with respect to the scene, or there may be some more complicated non-linear spatial relation between the template and its best match in the scene.

5. Parts of the template may simply be missing from the scene, for example due to obscuration or to temporal changes between the recording of the template and the collection of scene data.

There is an extensive body of literature which relates to the filtering of sampled, one-dimensional, electrical signal data (see e.g. [Haykin]). When the properties of the scene and the noise (or other perturbations) are known then Wiener filtering may be used. If the properties of the template and/or scene vary slowly with time (or its equivalent) then it is sometimes possible to employ adaptive Kalman filtering techniques.

Such methods are less successful for two-dimensional image data matching, though some attempts have been made. A more practical method for handling the many perturbing factors is to treat the problem of finding the best match between the template and the scene as a minimum-square-error (m.s.e.) problem using matrix methods. Here, the template data are deliberately distorted to cover the multi-dimensional parameter space (gain factor, x and y translation, rotation, shear etc.) until that choice of parameters is found which minimises the sum of squares of differences between the template and some part of the scene. Such methods have been explored in [Gruen a,b], [Otto] and elsewhere. D.G.'s thesis (in preparation) applies the technique to stereophotogrammetry of the human retinal fundus. Such error-minimising algorithms are based in effect on the use of the C_{ssd} algorithm: their computational complexity does not at present permit even greater refinements, such as the inclusion of spatial-frequency weighting based on Wiener or Kalman filter theory. In effect, the matrix methods assume that the data are noise-free, and that they need only to test the template against the scene, over some volume of parameter space, to find the perfect match.

If we could ignore the corrupting effects of noise, sampling effects and secular or other slow variations, then such methods would be almost completely successful. Any imperfect match would be equivalent to cross-correlation between a template $g(x,y)$ and a scene $f(x,y)$, and not to auto-correlation. The cross-correlation coefficient C_{ssd} would have some minimum value at the perfect match and a greater-than-minimum value for all other parameter values. We might therefore be tempted to think that it would be useful to derive expressions for C_{ssd} when the scene and template differ along one or other axis in the multi-dimensional parameter space.

In an ideal world this would indeed be the case, because one would then be able to assess the sensitivity of the algorithm to changes in one parameter or another. In practise, it does not seem to be so desirable. Apart from the enormous range of variations in possible scene type, the main limitation would appear to be the fact that the theory outlined in this report is based on the assumption that the scene is more-or-less uniform in 'texture'. At the present stage of development of this class of algorithm, there seem to be more pressing problems, the most important of which are probably

1. the need to assess the effects of noise and

2. the need to introduce an element of adaptability into multi-parameter matching algorithms, in order to cope with changes in scene texture and structure from region to region.

We shall not therefore in this report derive any expressions for cross-correlation which depend upon any particular distortion parameter.

7. SUMMARY

In this report we have derived formulae for the shape of the correlation surface when a template image is matched against a scene. Two correlation algorithms were considered, and formulae were given for three types of scene structure. Formulae were also given for one-dimensional data. Comparisons between theory and experiment will be given in a separate publication.

REFERENCES

J.L. de Bougrenet de la Tocnaye & F. Ghorbes, "Scale-rotation invariant pattern recognition applied to image data compression", Patt. Rec. Lett. **8**, 55-58, 1988.

T.M. Caelli & Z.-Q. Liu, "On the minimum number of templates required for shift, rotation and size invariant pattern recognition", Patt. Rec. **21**, 205-216, 1988.

M.-H. Chan & H.-T. Tsui, "Recognition of partially occluded 3-D objects by depth map modelling", Patt. Rec. Lett. **7**, 319-327, 1988.

G. De Soete,"A least squares algorithm for fitting an ultrametric tree to a dissimilarity matrix", Patt. Rec. Lett. **2**, 133-137,1984.

V.N. Dvornyechenko, "Bounds on (Deterministic) Correlation Functions with Applications to Registration", IEEE Trans. Patt. An. & Mach. Intell. **PAMI-5**, 206-213, 1983.

P. Gemmar, "Image correlation: processing requirements and implementation structures on a flexible image processing system (FLIP)", pp 87-103 of *Multicomputers and Image Processing*, Academic Press Inc., N.Y. 1982.

P.M. Griffin & B.L. Deuermeyer, "A Methodology for Pattern Matching of Complex Objects", Patt. Rec. **23**, 245-254,1990.

A.W. Gruen, "Adaptive Least Squares Correlation - A Powerful Image Matching Technique", presented paper to the ACSM-SP Convention, Washington, DC., 1985.

A.W. Gruen & E.P. Baltsavias, "High Precision Image Matching for Digital Terrain Model Generation", *International Archives of Photogrammetry and Remote Sensing*, **26**, No. 3, 284-296, 1986.

S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, Englewood Cliffs, N.J., 1986.

J.P. Keating, "Some practical aspects of covariance estimation", Patt. Rec. Lett. **3**, 295-298,1985.

H.-J. Lee & W.-L. Lei, "Region matching and depth finding for 3D objects in stereo aerial photographs", Patt. Rec. **23**, 81-94, 1990.

S. Lee & J.H. Kim, "A fast computational method for minimum square error transforms", Patt. Rec. Lett. **8**, 143-146, 1988.

A.V. Oppenheim & R.W. Schafer, *Digital Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1975.

G.P. Otto & T.K.W. Chau, "A 'Region-Growing' Algorithm for Matching of Terrain Images", Image and Vision Computing **7**, 83-94, 1989.

G. Parthasarathy & B.N. Chatterji,"The use of data windows in feature extraction for high-dimensional PR problems", Patt. Rec. Lett. **4**, 25-30,1986.

W.K. Pratt, *Digital Image Processing*, Wiley-Interscience, New York, 1978.

A. Rosenfeld & A.C. Kak, *Digital Image Processing*, 2nd. edition, Vol. 2, Academic Press, New York, 1982.

D. Rosenholm, "Multi-Point Matching Using the Least-Squares Technique for Evaluation of Three-Dimensional Models", Photogramm. Eng. & Remote Sens. **53**, 621-626, 1987.

D. Rosenholm, "Least-squares matching method: some experimental results", Photogrammetric Record **12**, 493-512, 1987.

S.L. Sclove, "Pattern Recognition in Image Processing Using Interpixel Correlation", IEEE Trans. Patt. An. & Mach. Intell. **PAMI-3**, 206-208, 1981

C.W. Therien & K. Fukunaga, "Properties of Separable Covariance Matrices and Their Associated Gaussian Random Processes", IEEE Trans. Patt. An. & Mach. Intell. **PAMI-6**, 652-656, 1984.

M.G. Thomason & E. Granum, "Sequential inference of Markov networks by dynamic programming for structural pattern recognition", Patt. Rec. Lett. **5**, 31-39, 1987.