University of Alberta

Computer Vision for Computer-Aided Microfossil Identification

by

Adam P. Harrison

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Science

Electrical and Computer Engineering

Spring 2010
Edmonton, Alberta

Examination Committee

Dr. Dileepan Joseph
Supervisor

Dr. H. Vicky Zhao
Examiner

Dr. Martin Jägersand
Examiner

Abstract

Micropalaeontology, a discipline that contributes to climate research and hydrocarbon exploration, is driven by the taxonomic analysis of huge volumes of microfossils. Unfortunately, this repetitive analysis is a serious bottleneck to progress because it depends on the scarce time of experts. These issues propel research into computerized taxonomic analysis, including a promising new approach called computer-aided microfossil identification. However, the existing computer-aided system relies on image-based representations, which severely limits its ability to discriminate specimens. These limitations motivate using computer vision to support richer video and shape-based representations, which is the focus of this thesis. An important contribution is a scheme to localize, capture, and extract video and shape-based representations from large microfossil batches. These representations encapsulate information across multiple lighting conditions. In addition, the thesis describes a method based on photometric stereo to correct misalignments in images of the same object illuminated from different directions. Not only does this correction benefit the application at hand, but it can also benefit a variety of other applications. The thesis also introduces a visual-surface reconstruction method based on maximum likelihood estimation, which constructs usable depth maps even from extraordinarily noisy images. State of the art methods lack this capability. By freeing classification from the bounds imposed by images, these contributions significantly advance computerized microfossil identification toward the ultimate goal of a practical and reliable tool for high-throughput taxonomic analysis.

Acknowledgements

I think I must begin my acknowledgements by quoting Sir Francis Bacon's powerful maxim toward research: "If a man will begin with certainties, he shall end in doubts; but if he will be content to begin with doubts he shall end in certainties." I do this, because although my supervisor, Dr. Dileepan Joseph, taught me countless things, I am most grateful for his influence on my own outlook toward scientific inquiry. Dil rigourously follows a philosophy toward research that may seem simple, but often proves very difficult to follow—an outlook, like Bacon's, that eschews certainty and revels in doubt. This thesis is a direct product of this outlook, and I come away from Dil's supervision with a greater sense of what it is to be a scientist.

I would be remiss if I didn't acknowledge my parents. Without their encouragement and support, I would not have entered into graduate school in the first place. Both of them were always ready to give me much needed perspective when needed, and I was always sure of their support and love. My siblings also helped a great deal in getting me across the finish line. I am especially grateful for the almost daily chats Julie and I shared regarding graduate school and other subjects, which never failed to provide me with much needed shots of encouragement.

The fabulous friendships I fostered in Edmonton made for a particularly good couple of years. Many of these friendship were formed through the University of Alberta Outdoors Club—a club I was heavily involved with that provided me with many weekends in the mountains free from thoughts of computer vision or microfossils.

My fellow students in the Electronic Imaging Lab made for a wonderful environment. Alireza Mahmoodi and I shared many laughs and also many excellent discussions diverting both of us from our work. Cindy Wong made for a fantastic neighbour, and I look forward to keeping up with her own work, which is a fascinating project also related to micropalaeontology. Also, it is no exaggeration to say that I would not have completed this thesis without the benefit of Orit Skorka's continual and reliable restocking of the coffee supply. It was an addiction well-shared. Finally, during the initial year of my studies I worked with Dr. Kamal Ranaweera on a preliminary version of a computer-aided identification system. It was pleasure to work under Kamal's example and benefit from his overwhelming knowledge of

both hardware and software.

I also enjoyed some great relationships with Dr. Martin Jägersand's Vision Group in the Department of Computing Science. Martin and Dana Cobzas were always ready to lend me their vast computer vision expertise with any research problem I had. As well, the Vision Group's reading groups were a great way to learn about fascinating research outside of my own work. Also, I especially enjoyed sharing coffees and illuminating mathematical conversations with Neil Birkbeck.

Finally, I would also like to acknowledge the support I received over the course of this thesis. I was fortunate enough to have received institutional support from the University of Alberta. In addition, the work was also supported by an NSERC CGSM grant and an Alberta Ingenuity iCORE grant.

Table of Contents

1 Introduction 1

1.1 Biostratigraphy: An Important Field 1

1.2 Motivation for a Computerized Solution 2

1.2.1 Biostratigraphy Today 2

1.2.2 Computerized Microfossil Identification 3

1.3 State of the Art of Computerized Solutions 5

1.4 Computer-Aided Identification 8

1.4.1 Method 8

1.4.2 Results 13

1.5 Organization of Thesis 19

2 Computer Vision for Computer-Aided Identification 21

2.1 The Role of Computer Vision 22

2.2 Image Formation and Surface Reconstruction 25

2.3 Maximum Likelihood Estimation 29

2.3.1 The Nonlinear Case 32

2.3.2 The Linear Case 34

2.3.3 Weighted Normals 36

2.4 Conclusion 37

3 Video and Shape-Based Representations 39

3.1 Video Capture Module 41

3.1.1 Equipment 41

3.1.2 Batch Processing 42

3.1.3 Video Capture 44
3.1.4 Dataset Collection 47
3.2 Clustering Module 48
3.2.1 Similarity Measure 49
3.2.2 Clustering Algorithm 50
3.2.3 Template Selection 53
3.3 Shape Extraction Module 54
3.4 Expert Input Module 54
3.4.1 Video-Based Identification 54
3.4.2 Particle-Based Identification 55
3.5 Analysis of Representations 55
3.6 Conclusion 61

4 Image Alignment Using Photometric Stereo 64

4.1 Photometric Alignment 66
4.2 Evaluating the Method 71
4.2.1 Testing with Known Misalignments 71
4.2.2 Quantifying Benefits 75
4.2.3 Experiments with Microfossil Videos 78
4.3 Asymptotic Complexity 81
4.4 Conclusion 82

5 Maximum Likelihood Surface Estimation 85

5.1 Estimating the Gradient Fields 85
5.2 Estimating the Visual Surface 88
5.2.1 Maximum Likelihood Estimation 88
5.2.2 Modified Maximum Likelihood Estimation 92
5.3 Experiments 95
5.3.1 Estimating Known Surfaces 95
5.3.2 Estimating Microfossil Surfaces 103
5.4 Asymptotic Complexity 104

5.5 Conclusion 106

6 Conclusions 109

6.1 Summary 109

6.1.1 Video and Shape-Based Representations 111

6.1.2 Image Alignment Using Photometric Stereo 113

6.1.3 Maximum Likelihood Surface Estimation 114

6.2 Future Work 115

6.2.1 Video and Shape-Based Representations 116

6.2.2 Image Alignment Using Photometric Stereo 117

6.2.3 Maximum Likelihood Surface Estimation 118

Bibliography 121

A Creating Anaglyph Images 130

B First-Order Finite-Difference Operators 133

B.1 1D Finite-Difference Operators 133

B.2 2D Finite-Difference Operators 134

B.3 Surface Estimation Normal Equations 136

List of Figures

1.1 Inside the IODP Bremen Core Repository. 3

1.2 Gene sequencing was immensely accelerated because of the introduction of automation. 4

1.3 The state of the art of computer-aided identification. 9

1.4 The invariant transform. 11

1.5 Accuracy of image-based classification. 16

1.6 Performance of image-based computer-aided identification. . 18

2.1 The significant role of illumination 22

2.2 The effect of illumination direction on similarity scores. . . . 23

3.1 System outline of video and shape-based computer-aided identification of microfossils 40

3.2 Computer-aided system setup 43

3.3 Coordinate system used for microfossil capture 44

3.4 Stage localization errors 46

3.5 Plot of localization error vs. distance of particle to the origin. 47

3.6 Creating a video-based representation. 48

3.7 Obtaining the image corresponding to the alternate resolution of the 180° ambiguity. 51

3.8 Comparison of performance of maximal-clique clustering vs. agglomerative hierarchical clustering. 53

3.9 Archived physical specimens 55

3.10 Comparison of video and shape-based representations. 56

3.11 Histogram of R^2 values for visual surfaces and weighted normals. 58
3.12 Texture-mapped shape-based representations. 60
3.13 Example anaglyph images 62

4.1 Silhouette differences in a centroid-aligned video sequence. . 65
4.2 Example images of the four textures for testing photometric alignment accuracy. 72
4.3 The masking technique for the texture images. 73
4.4 Performance of photometric alignment on the 24 sequences of texture images. 74
4.5 Producing the error-prone centroid-aligned hemisphere images. 76
4.6 Quantitative benefits of using photometric alignment. 77
4.7 Qualitative benefits of using photometric alignment. 78
4.8 Example of two images in a microfossil video using centroid and photometric alignment. 79
4.9 Histogram of R^2 values. 81

5.1 Synthetic surfaces and sample images 97
5.2 Mozart and Shark depth map similarity scores 99
5.3 Mozart surface reconstructions 100
5.4 Shark surface reconstructions 101
5.5 Small-scale Mozart surface reconstructions 102
5.6 Hyperbolic-paraboloid depth map similarity scores 103
5.7 Hyperbolic-paraboloid reconstructions 104
5.8 Surface reconstructions of microfossils 105

List of Tables

3.1 Specifications of the equipment used in the video capture module. 42

3.2 Specimen image variability accounted for by the weighted surface normals and visual surfaces. 59

4.1 Results of applying photometric alignment to microfossil videos. 80

List of Symbols

Z	Visual surface	26
$\mathbf{n}$	Surface normals	26
ρ	Albedo	26
$\boldsymbol{\eta}$	Weighted normals	26
$\boldsymbol{\ell}$	Light direction	26
$\mathbf{L}$	Matrix of light directions	28
p, q	Gradient fields	26
$\boldsymbol{\omega}$	Shadow vector	28
$\boldsymbol{\beta}$	Regression parameters	30
$\mathbf{f}(.)$	Regression model	30
$\mathbf{F}.(.)$	Jacobian of regression model	32
ϵ	Additive noise (typically zero-mean Gaussian)	30
μ	Mean	31
σ	Standard deviation	31
$\boldsymbol{\Gamma}$	Matrix proportional to covariance	31
$\mathcal{N}$	Normal distribution	31
$\mathbf{B}$	Matrix such that $\boldsymbol{\Gamma} = \mathbf{B}^T\mathbf{B}$	31
$\mathbf{g}, \mathbf{h}$	Reparameterization functions	33
$\mathbf{R}$	Constraint matrix	35
$\mathbf{b}$	Constraint values	35
θ	Elevation angle	44
φ	Azimuthal angle	44
R^2	Coefficient of determination	57

List of Abbreviations

AD	Anisotropic Diffusion	91
ADIAC	Automatic Diatomic Identification and Classification	4
AgH	Agglomerative Hierarchical Clustering	52
ANN	Artificial Neural Network	5
API	Application Programming Interface	41
CG	Conjugate Gradients	70
CGR	Correct Genus Rate	14
CSR	Correct Species Rate	14
FC	Frankot-Chellappa	96
GLS	Generalized Least-Squares	31
GQR	Generalized QR Factorization	35
HP	Hyperbolic Paraboloid	96
IID	Identically and Independently Distributed	31
IGR	Incorrect Genus Rate	14
IODP	Integrated Ocean Drilling Program	2
ISR	Incorrect Species Rate	14
ITC	Ideal-Template Classifier	17
LM	Levenberg-Marquart	70
ML	Maximum Likelihood	29
ODP	Ocean Drilling Program	2
OLS	Ordinary Least-Squares	31
PDF	Probability Density Function	31
PTC	Practical-Template Classifier	17

QN	Quasi-Newton	70
RAP	Random A Priori	14
SCG	Scaled Conjugate Gradients	70
SE	Shift Error	71
SEM	Scanning Electron Microscopy	6
SSE	Sum-Squared Error	31
UA	University of Alberta	5
VIDES	Video Identification Expert System	5

Chapter 1

Introduction

1.1 Biostratigraphy: An Important Field

Biostratigraphy, the science of identifying and dating sedimentary rock layers through the study of fossils, is an important area for both climatology and the energy industry. Marine microfossils found within sediment samples are often of species that evolved at a very rapid rate. Consequently, there is a direct correlation between the presence of a certain species and the time period of its sample. Species that are used for time correlation are called index fossils. Ideal index fossils typically possess three characteristics [1]:

- Easily recognized;
- Spread rapidly and widely and then quickly became extinct;
- Easily preserved.

Among other benefits, this correlation between species and time periods allows scientists to determine whether two very distinct sediment samples originate from the same time period or not.

An important field in its own right, biostratigraphy also plays a very important role in academia and industry. For instance, an inherent problem with climate models is that there is no way to run full-scale experiments for testing cause and effect. As a result, climate scientists must use past data to compile a rich dataset. The information held within microfossil samples is key in providing vital information in understanding prehistoric climate [2]. The petroleum industry is another area where microfossil identification plays a significant role. Because biostratigraphy provides supplementary information, industry often employs micropalaeontologists to improve their geological models of an area. Using biostratigraphy to date rock layers, these models provide vital guides in locating hydrocarbon deposits [3]. Thus, biostratigraphy is crucial in scientific and industrial applications that are of great import to our society.

1.2 Motivation for a Computerized Solution

1.2.1 Biostratigraphy Today

In Simmons *et al*'s look ahead to the future of biostratigraphy [4], the authors assert that taxonomy is the driving force behind their field. Taxonomy is key in identifying good index fossils, and in providing reliable relative time scales between biozonation schemes (intervals of rock strata characterized by their fossil species). Detailed taxonomic study is especially required when dealing with biozonation schemes that tend to only possess very localized fossil specimens. Typically, microfossil identification is accomplished through manual study of samples under an optical microscope. Those performing taxonomy usually require three to four months of training [3].

Unfortunately, much taxonomic work still needs to be accomplished. In particular, Simmons *et al* warn that certain biozonation schemes are in serious need of more taxonomic study. Paradoxically, despite the great need for further taxonomy, the authors warn that the field is experiencing decline in interest and funding. The authors point out that in particular biostratigraphy must continually combat the misconception that taxonomic study is “all done” [4]. This problem is so serious that Simmons later devoted an entire article to the subject [5].

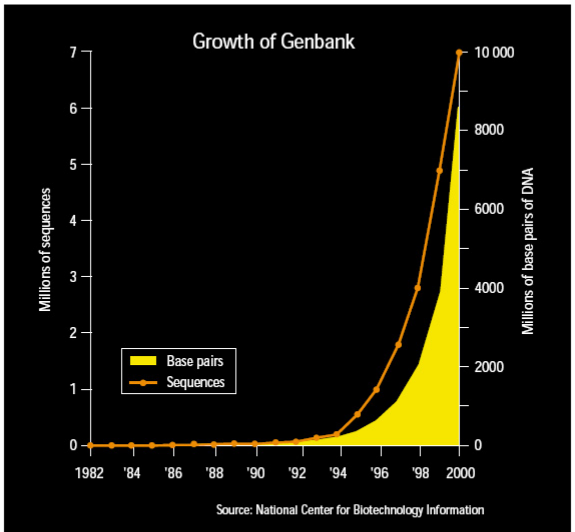
Yet, a major problem with taxonomy is that identifying microfossils in a rock sample is extremely time-consuming. This problem is compounded by the enormous amounts of taxonomic study still required to fill in existing knowledge gaps [4]. Projects such as the Ocean Drilling Project (ODP), and its successor the Integrated Ocean Drilling Project (IODP), have been extraordinarily successful at collecting vast amounts of samples. However, the prospect of classifying species based on this data is daunting to say the least—altogether the ODP collected 130 miles of core samples over the length of its program [6]. As a typical microfossil is the size of a grain of salt, it is apparent that taxonomic classification of these samples is an enormous task. As a result, the problem facing biostratigraphy is two-fold: classification is labour intensive, and a huge amount of taxonomic work still must be performed. Since the industrial and academic applications of biostratigraphy are of such importance to society, solutions should be explored to help overcome the issues with taxonomy that biostratigraphy faces today.

Figure 1.1: Inside the IODP Bremen Core Repository. Each repository holds vast stores of samples. The Bremen repository has a capacity of 1100 m^2, and contains 130 km of deep-sea cores within roughly 100,000 boxes (taken from the Bremen repository website).

1.2.2 Computerized Microfossil Identification

The capacity for engineering skills and techniques to be effectively applied towards important scientific endeavours has been demonstrated in the past. Take, for instance, the decoding of the human genome. Decoding the human genome required sifting through an incredible amount of data. Moreover, traditional gene sequencing is a very lengthy and expensive process requiring the use of expert labour. As Hodgson [7] illustrates (see Fig. 1.2), without automation techniques the human genome project would still be in its infancy and would have had no prospect of finishing within our lifetimes. When automation techniques were brought in, the decoding of the human genome finally became feasible. Hodgson notes that, "a single DNA sequencing machine [in 2000] can produce over 330 000 bases (units of sequence information) per day, more than 100 researchers could manage in a year and a half using manual techniques." When put in that way, the benefits brought about by automation to the human genome project are simply astounding.

In many ways, the problem geneticists faced is analogous to the one biostratigraphers face today. Both must accomplish an analysis of an enormous amount of data by performing repetitive work. However, biostratigraphy's problem is different, as the identification of a microfossil is a purely visual task. Designing an automatic system to identify or discriminate between fossil samples is a challenging problem. Nonetheless, the continuing advances in image processing, pattern recognition, and computer vision indicate that computer engineering is poised to develop breakthroughs for this



Source: National Center for Biotechnology Information

[1] The number of bases of sequenced DNA deposited into the international DNA database system, including GenBank, has skyrocketted since the invention of industrial-scale sequencing systems. The first high-throughput sequencing machine only became available in 1997.

Figure 1.2: Gene sequencing was immensely accelerated because of the introduction of automation (taken from [7]).

important field.

In fact, recent progress in the automatic identification of other biological entities suggest that computerized microfossil identification is an attainable goal. For instance, the Automatic Diatomic Identification and Classification (ADIAC) project has enjoyed promising preliminary success rates in classifying *live* diatoms (unicellular algae) using optical microscopy [8]. While both problems possess their own respective challenges, the success in identifying live diatoms provides much encouragement with respect to microfossils. There is now great reason to believe that a collaborative enterprise between computer engineering and micropalaeontology will significantly accelerate biostratigraphic work. To become a viable contribution to biostratigraphy, such a collaboration must accomplish two tasks:

- Identify microfossil species based on their visual features
- Sort microfossil samples based on their identification.

In addition to an acceleration in classification, such a system would also have the potential to provide consistent accuracy. People, especially those with training, have a considerable capacity for identification and classification. However, even experts are not infallible to fatigue and bias. Culverhouse *et al* studied the accuracy of expert taxonomic classification in the case of phytoplankton specimens and concluded that accuracy was considerably lower than what the authors had expected [9]. The authors concluded that fatigue was a major player in the results. Even though the study focused on phytoplankton, the conclusions of the paper are relevant to microfossil classification, as both deal with taxonomic classification under a microscope. Similar results were reported in the ADIAC project, where the authors concluded that even experience was not a good indicator of classification accuracy [8]. Regarding microfossil identification, the report of the Second Conference on Scientific Ocean Drilling went even so far as to state that inconsistency is a "fundamental weakness" [10]. As a result, if one could develop a computerized microfossil classification system with high accuracy one would avoid having fatigue and bias skew the results.

For these reasons, an automated approach to microfossil identification has been an acknowledged goal within the field for some time [10]. With the advance of trans-disciplinary applications for complex problems, computer engineering offers a unique contribution to particular scientific problems and challenges. Of interest to the field of biostratigraphy is the development of a computerized microfossil identification and sorting system. The topic of this M.Sc. thesis is on the *identification* side of this problem.

1.3 State of the Art of Computerized Solutions

The benefits of developing a computerized microfossil identification system have been recognized by a variety of researchers. Apart from the work conducted at the University of Alberta (UA) Electronic Imaging Lab (formerly the Imaging Science Lab), previous work can be broadly separated into those that use rule sets (the same rules micropalaeontologists use during manual examination) and those that employ a fully-automated artificial neural network (ANN) using supervised learning.

As a way to lesson human error and training requirements, earlier systems often focused on aiding those tasked with manual identification [3, 11, 12]. These systems all rely heavily on human interaction; as such, they do not provide a significant lessening of workload. Of particular note is the most recent system called the Video Identification Expert System (VIDES) [3]. Developed through BP Plc (British Petroleum at the time), VIDES functions by presenting the user with a set of identifying attributes. Based on the sample in question, the user selects values for as many of the attributes as possible. Considerable effort was invested in making attribute selection as easy as possible. As the user selects more attributes, the system infers with greater confidence the identity of the microfossil.

As noted, VIDES does not provide an automated approach to microfossil identification. Even so, using rule sets to infer microfossil species are attractive as they directly correspond to the criteria that micropalaeontologists use for manual identification. Attempting to retain the use of these rule sets, Dr. Thonnat's research group, based at the French National Institute for Research in Computer Science and Control, proposed an identification system that incorporates rule-sets into an automated scheme [13, 14]. The system creates correspondences between mathematical shape descriptors and knowledge-based rules. Although an interesting direction to take, the research group never made clear which low-level features were included and no results were reported. Another serious weakness is that all work was based on holotype images (high quality example of taxa). Yet, practical systems must function "in the wild", where they will encounter specimens of all ranges of quality. Moreover, the system used scanning electron microscopy (SEM) images. Relying on SEM images presents several serious problems. For one, obtaining SEM images is difficult and time-intensive, impacting any time-reductions due to automation. In addition, the system required SEM images from three different views, which is difficult to accomplish under SEM. As well, SEM samples must be treated prior to image acquisition, often making subsequent analysis and sorting impossible.

In contrast to rule-based approaches, other systems use low-level artificial neural networks to classify microfossils. Current examples of ANN approaches all train and classify directly in pixel space. One of the leading examples of an ANN-based classifier is SYRACO 2, developed by Beaufort *et al* [15]. Initially, SYRACO 2 relied on pre-processed images that were normalized with respect to rotation and translation. Unfortunately, the authors found that their normalization technique was unreliable and later developed an improved version [16] that handles translation and rotation by using a somewhat unwieldy combination of 6 parallel neural networks. However, even in its original form, when using very low resolution 64×64 images, SYRACO 2 tuned $800,000$ parameters during training. According to neural network theory, such a network would require a lower bound of $800,000$ images in the training set [17]. As their number of training images was a small fraction of that, the authors were unable to fully explain why their network worked at all. The number of parameters using the bank of parallel neural networks was not reported.

The ANN approach is also used by the Micropaleontology/Geophysiology group at ETH Zurich [18]. The group tested their system, called COGNIS, using reflected-light microscopy, transmitted-light microscopy, and SEM. Out of all of these modalities, only SEM was used in testing whether the system could identify multiple species. Although SEM did produce satisfactory results, the use of SEM poses several problems, as mentioned before. Regarding optical microscopy, the authors performed an experiment testing whether specimens of a certain species (*Florisphaera profunda*) could be identified from a set of images of varying species. While the system had a high correct classification rate, it was also very unreliable, producing a false positive rate of 80%.

As noted, the state of the art can be separated into approaches using traditional rule-sets and approaches using ANNs. Apart from Dr. Thonnat's research group, which did not produce results anyway, the rule-set based approaches are not designed to lessen identification workload. As a result, the state of the art is broadly defined by two extremes in term of automation: those providing little to no automation, represented by rule-set based approaches, and those aiming to provide complete automation, represented by ANN-based approaches. While solutions using rule-sets may lessen training requirements or misclassification rates, they do not address the significant issues regarding the heavy workload and specimen volumes required for detailed microfossil taxonomy. In contrast, the ANN approach does aim to solve these issues, but the practicality of supervised learning is doubtful when applied to the domain of microfossil identification.

For instance, the nature of supervised learning requires that a significant training set be in place that consists of classified data. Developing a large training set is difficult in any situation; however, the time-consuming and expert-labour driven nature of traditional manual microfossil identification makes dataset collection an especially daunting challenge. In fact, it is this difficulty that has motivated research into computerized identification in the first place. Each species one wishes to identify requires a significant number of associated images in the training set. For example, COGNIS [18] reported that on average 70 images were used for each species in its training set. This perhaps explains why SYRACO 2 [16] and COGNIS [18] were only trained to identify 16 and 14 species respectively. When dealing with orders of microfossils possessing very large numbers of species, this problem is especially acute. For instance, estimates of the number of extant benthic foraminifera species (protozoic marine life forms) and diatoms (aquatic algae) are upwards of $40,000$ [19] and $100,000$ respectively [20]. Thus, the prospect of using an ANN-based approach for *general* identification of these microfossils is next to impossible.

If the sheer numbers of microfossil species are not enough of an obstacle for ANN-based approaches, the rigid nature of the training process certainly provides its own significant disadvantages. If one wishes to add or amalgamate species, the system must be completely retrained. As many types of microfossils, including but not limited to benthic foraminifera [21], diatoms [22], and conodont elements (tooth-like microfossils) [23], have an ever-evolving taxonomy, this is not a trivial problem.

For these reasons, the ANN-based approach suffers from serious drawbacks toward becoming a practical and useful implementation. As a result, these problems motivate the alternative direction pursued for this thesis: an unsupervised semi-automated scheme, designated from here on in as *computer-aided* microfossil identification.

1.4 Computer-Aided Identification

Computer-aided identification is a promising alternative to the fully-supervised approaches pursued in the state of the art. Serving as the basis of this thesis, an existing implementation of a computer-aided identification system, developed by the UA Electronic Imaging Lab, has demonstrated feasibility [24]. The author was also a contributor to this work. Sec. 1.4.1 outlines the steps behind this implementation of computer-aided identification. Sec. 1.4.2 discusses its performance.

1.4.1 Method

The challenges associated with the extensive and ever-evolving nature of microfossil taxonomy can be avoided if one implements a system designed to keep experts in the loop. More specifically, one could envision a system that automatically clusters specimens based on their visual similarity (and not pre-existing taxonomic knowledge), chooses a template for each cluster, and then presents the templates to an expert for identification. Experts could either examine templates using a digital representation or the actual particle under a microscope. The template's label would then be applied to all of the particles in its associated cluster. Unlike ANN-based approaches, the goal of such a system would be a significant reduction in identification workload, but *not* an elimination.

Even so, such an approach benefits from an extraordinarily high level of flexibility, as it is not hampered by the uncertain state of microfossil taxonomy and requires no prior training set collection. As a result, a computer-aided identification system could be applied to all species or taxonomic groups, an impractical proposition for identification schemes relying on training sets. Additionally, such a system can be tuned to either aim for fewer or more clusters, thus providing a means to control the tradeoff between work reduction and accuracy. As well, an unsupervised clustering approach also provides more flexibility in the order of steps needed to identify and sort microfossils in a complete system. Unlike a supervised approach, physical specimens can be first sorted into clusters based on their visual similarity and identification of each cluster template can be performed later.

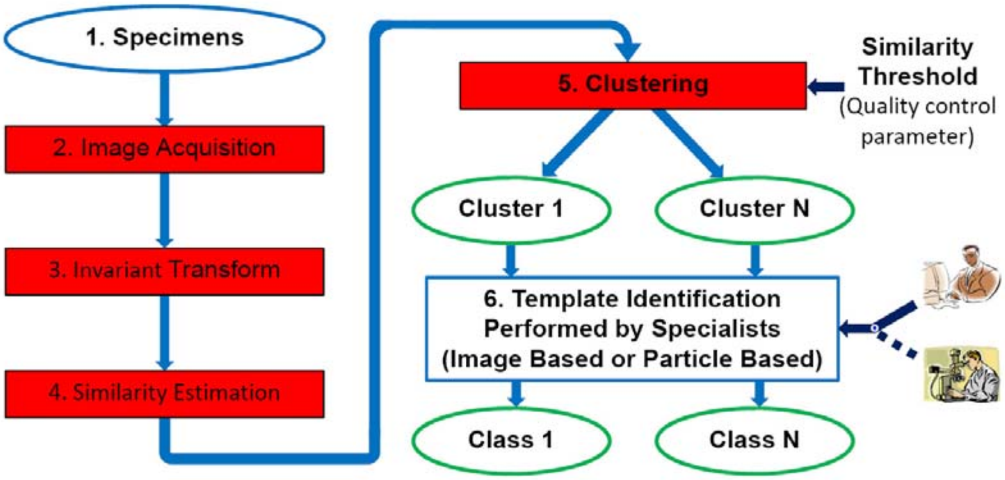


Figure 1.3: The state of the art of computer-aided identification. (1) As a first step, an appropriate specimen set must be collected; (2) Single images of the particles in this set are captured; (3) The objects in the images are canonized so that they are centered and their principal axes are at an orientation of 0°; (4) A similarity matrix for all image pairs in the set is computed; (5) A clustering algorithm uses the similarity matrix as input. A user-defined similarity threshold controls the degree of similarity required between members of the same cluster; (6) For each cluster, a template is automatically chosen. An expert identifies the template using either the specimen image or the actual particle. The template's label is applied to every member of its cluster.

Such an approach, using image-based representations of microfossil particles, has been developed in a preliminary stage at the UA Electronic Imaging Lab [24]. As much of the work in this thesis involves advancing the capabilities of the preliminary system, particularly by incorporating computer vision, it is worthwhile devoting space to the methodology of this implementation. In addition, a first contribution of this thesis is to identify limitations of this preliminary approach.

In its preliminary state, computer-aided microfossil identification can be broadly broken down into several steps. These are illustrated in Fig 1.3. The following subsections explain these individual steps in more detail. For a full explanation of the preliminary system, the reader is encouraged to consult [24].

1.4.1.1 Specimen Set

The specimen set used for the preliminary implementation was comprised of foraminiferal tests (forams). A key oceanic life form for millions of years, countless numbers of foram specimens have been extracted over the years by several different ocean drilling programs, including the ODP. The study of their morphology, as in [25], and their chemical composition, as in [26, 27], has advanced the current understanding of prehistoric climate. In addition, forams have also played important roles in locating hydrocarbon deposits [28].

1.4.1.2 Image Acquisition

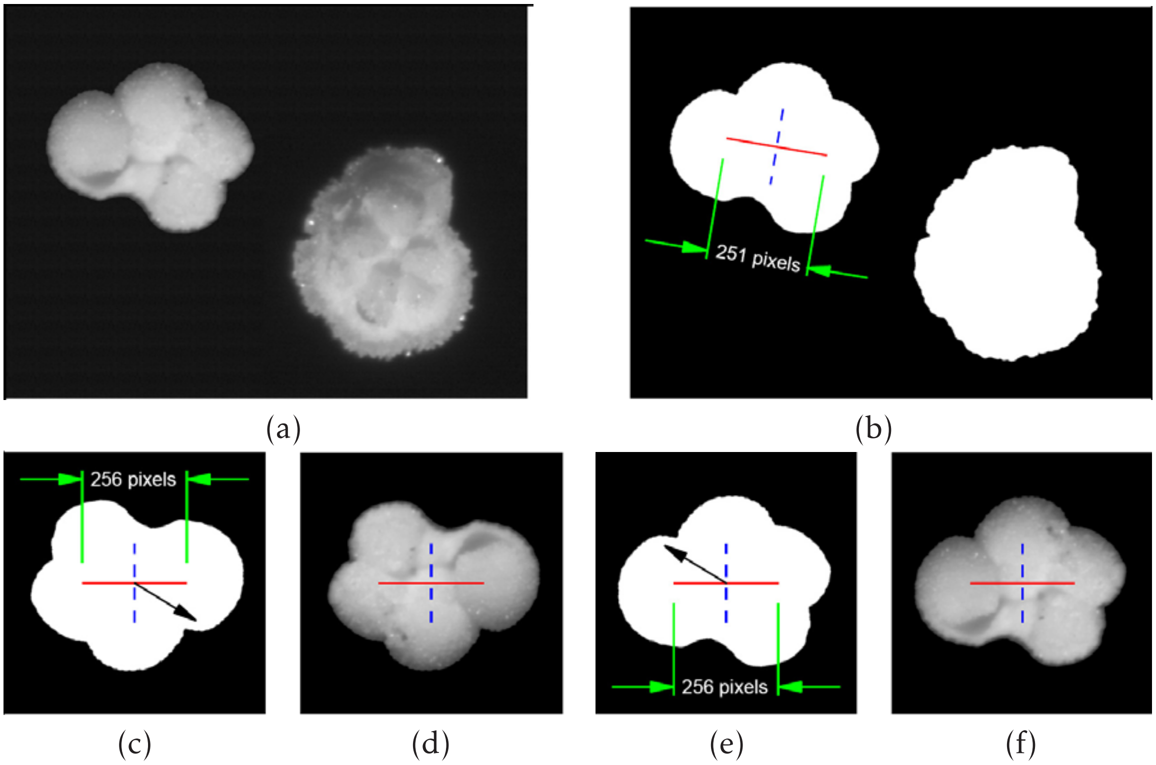
As a first step to image acquisition, a user must sieve the particles and sprinkle them onto an opaque glass slide. The base equipment requirements of the image capture system comprise a microscope, an attached digital camera, and custom-built software to properly capture and store images. As the described equipment is also used for this thesis, further details on the microscope and digital camera can be found in Sec. 3.1.1. Providing an opaque black background, the glass slide allows the system to localize particles by segmenting the current field of view using simple thresholding and searching for any silhouettes within the size ranges allowed by the sieve. Both silhouettes and images of each particle were captured.

1.4.1.3 Invariant Transform

As images of particles sprinkled on a slide exhibit arbitrary rotation, image location, and possibly even scale, normalizing against these factors is an important consideration. For this reason, the system mapped images into a canonical space using an invariant transform. Implementing an invariant transform to guard against these factors requires first computing specimen location, rotation, and scale. As shown in Fig. 1.4(a) and (b), computation of these characteristics used the particle's silhouette. The centroid of the silhouette provides coordinates of the specimen's location in the image while computing the principal components of the silhouette [29] provide the rotation and scale.

With these measurements computed, each image underwent a normalization transform such that:

- The centroid of the object rests in the middle of a 640×640 pixel image.
- The length of the major axis of the object silhouette is 256 pixels.
- The major axis of the silhouette is horizontal.



(a) (b)

(c) (d) (e) (f)

Figure 1.4: The invariant transform (taken from [24]). (a) input photograph containing multiple specimens; (b) silhouette of each specimen after segmentation, with principal components superimposed on one of them; (c) silhouette of one specimen centered in an image of fixed size, with fixed rotation and scale of the first principal component; and (d) canonical image of one specimen. There would be a 180° rotational ambiguity in the canonization, i.e. compare (e) to (c) and (f) to (d), if not for the use of third central moments in the x and y directions, which are indicated by vectors. In both (c) and (e), the third central moment is largest in the x direction. However, the largest third central moment is positive only in (c).

The results of such a transformation applied to an image and its silhouette are shown in Fig. 1.4(c) and (d). However, as demonstrated by Fig. 1.4(c) and (e), orientating an object so that its major axis is horizontal leaves a 180° ambiguity. However, by forcing one of the third-central moments in the x or y direction to be positive, the ambiguity is removed. The best choice is to compute both third-central moments, and force the one with the largest magnitude to be positive [24]. Thus, in the example of Fig. 1.4, the image in (c) represents the correct choice.

1.4.1.4 Similarity Estimation

The next step in the process is to compute a similarity matrix consisting of correlation scores between all pairs of specimen images. A well established measure of similarity in areas of pattern recognition [29], the *correlation coefficient* was used as a similarity measure. Two important reasons motivate the choice of correlation coefficient as a similarity measure [24]. For one, correlation is a well-understood measure, allowing the merits of computer-aided identification to be evaluated without the worry of having a less-established metric confounding results. Secondly, should other similarity measures be evaluated, having results using a proven metric provides a useful basis of comparison.

Removing the effect of background pixels on the computation, the system used a modified form of the correlation coefficient [24]. The modified correlation coefficient, r, between canonical two specimens a and b is defined as:

$$r = \frac{\sum_{k \in I \cup J}(A'_k)(B'_k)}{\sqrt{\sum_{i \in I}(A'_i)^2 \sum_{j \in J}(B'_j)^2}}, \tag{1.1}$$

where A and B represent the image-based representations of a and b respectively. Here, i and j represent pixel coordinates belonging to canonical silhouettes I and J corresponding to a and b, and k represents pixel coordinates belonging to the union $I \cup J$. Differing from the traditional computation of correlation, $A'_{(.)}$ and $B'_{(.)}$ are the pixel intensities of A and B after subtraction of the mean foreground intensity (as opposed to mean intensity of the entire image) from the foreground, where foreground is defined as the region bounded by the canonical silhouette.

1.4.1.5 Clustering

Computing a similarity matrix between specimen pairs is a popular and established method to represent data for clustering [30]. The computer-aided system processed this data by constructing a virtual graph composed of vertices for every specimen and edges only between vertices of visually similar specimens. Two specimens were considered similar if their correlation score was above a user-defined threshold. As the only parameter in the system, the threshold parameter controls whether the system would be inclined to produce fewer or less clusters. Consequently, the threshold parameter determines the degree of importance placed on work reduction vs. classification accuracy. Clusters were extracted from the graph by searching for maximal cliques.

1.4.1.6 Template Identification

With specimens grouped into clusters, a template must be selected from each group. Based on the criteria of *maximum intra-cluster similarity*, the system automatically chose a template from each cluster. Here, an image's intra-cluster similarity is defined as the sum of correlation coefficients that the image has with the other images in its cluster. The specimen with the image holding the highest intra-cluster similarity is chosen as the template.

Up to this point all steps have been automatic. However, for the next step the system requires an expert to identify each template. Two variations exist on the identification process:

- In the *image-based* variation, templates are identified from their images, i.e. a digital representation of physical specimens;
- In the *particle-based* variation, templates are identified by traditional examination under a microscope, i.e. from physical representations.

To enable image-based identification, after images of each specimen were captured, they were automatically uploaded to an online-wiki[1], designed to provide a user-friendly and interactive interface in which experts can search for and identify individual specimens.

In terms of flexibility and ease, image-based classification is the most desirable variation. However, as Sec. 1.4.2.1 will demonstrate, it may not always be possible to identify microfossils from images. Regardless of the identification method used, once an expert applies a class label to a template specimen the same classification is applied to every member of its corresponding cluster.

1.4.2 Results

The UA Electronic Imaging Lab performed two studies pertinent toward testing the capabilities of computer-aided classification. The first study did not test computer-aided identification *per se*, but instead explored the accuracy and reliability of image-based template identification [31]. Since, image-based template identification is preferred due to its flexibility and relative ease, this is an important question to explore. As well, the results of the first study influenced the direction and scope of the second, which examined the performance of the preliminary computer-aided identification system [24]. Both studies reveal important directions to pursue to

[1] http://www.ece.ualberta.ca/ imagesci/microfossil/wiki/wiki.php

further improve computer-aided classification performance and have important implications for this thesis. The results of both studies, which use the metrics described below, are summarized in Secs. 1.4.2.1 and 1.4.2.2.

In the hierarchy of Linnaean taxonomy, the ability to classify the two ranks of *Genus* and *Species* are of primary importance. The correctness or incorrectness of a classification represents a binomial trial with a true or false outcome. For a given microfossil dataset, particle-based classifications of each specimen represents the ground truth. For both studies, four metrics were used to judge the accuracy and reliability of the classifications and are defined as follows [31]:

- The correct genus rate (CGR) is the proportion of specimens, with known genus in the ground truth, that were correctly identified. A similar definition applies for the correct species rate (CSR).

$$CGR = \frac{\text{Number correctly identified}}{\text{Number with known genus}}. \tag{1.2}$$

- The incorrect genus rate (IGR) is the proportion of specimens, with known genus in the ground truth, that were incorrectly identified. A similar definition applies for the incorrect species rate (ISR).

$$IGR = \frac{\text{Number incorrectly identified}}{\text{Number with known genus}}. \tag{1.3}$$

The CGR and IGR, and the CSR and ISR, may not sum to unity as it is possible for a genus or species field to be classified as *unknown*. As any classification scheme will experience varying degrees of success depending on the proportions of each class within a sample set, it is important to construct an appropriate basis of comparison. The random *a priori* (RAP) classifier serves this purpose by performing classifications based on the distribution of classes within the sample. For RAP classifications, the CGR and IGR were computed as follows:

$$CGR = \sum_{i=1}^{N} p_i^2, \tag{1.4}$$

$$IGR = 1 - CGR. \tag{1.5}$$

Here, N represents the number of known genera and p_i represents the ground truth proportion of the i^{th} genus amongst specimens with known genera. The RAP CSR and ISR are similarly defined.

Regardless of the manner of classification, confidence intervals provide a means to extrapolate these results from the sample to the entire population. 95% confidence intervals were computed using the Wilson score method [32].

1.4.2.1 Accuracy of Image-Based Template Identification

As images are the preferred means to identify a template, and are also the representations used to cluster specimens, an important question to explore is whether images contain enough information to reliably identify a specimen. Quantifying how accurately an expert can classify specimens based on their images is one way to determine this. As a precursor to computer-aided microfossil identification, the UA Electronic Imaging Lab conducted a study on exactly this subject, focusing on mainly planktonic foraminifera [31]. This question is not trivial because taxonomists often vary the focal plane and manipulate samples during identification—actions that are impossible when examining a simple digital image. As of yet no one else has reported findings on whether digital images actually contain enough information for identification. The results of this study impacted the preliminary implementation of computer-aided identification and also significantly influenced the work of this thesis.

To study this issue, the authors performed experiments quantifying the agreement between image-based classifications of 244 forams and the ground truth, which as mentioned is considered to be the corresponding particle-based classifications of the same forams. As Fig. 1.5 demonstrates, when compared to the RAP classifier, image-based classification performs markedly better at genus identification. Even so, experts were only able to correctly classify specimen genera roughly 80% of the time. At the species level, image-based classifications suffer from a worse correct classification rate. Even so, species level identification still exhibits a low incorrect classification rate. The low IGR and ISR demonstrated in Fig. 1.5 indicate that if an image-based identification of a foram can be performed, it is almost always accurate. However, the expert was not always able to perform image-based classifications, opting to apply an "unknown" label, which reduced the correct classification rates. This problem is especially acute at the species level.

The variability of these results was tested across a variety of factors. Only image quality affected the classification rates with any significance. As a result, the authors concluded that digital images do contain enough information for identification, but only at the *genus* level.

These results underscore the importance of the quality and detail of digital representations presented to experts when performing identification tasks. Should one wish to have experts identify specimens using some digital representation, alternatives other than images will be needed to obtain high correct classification rates.

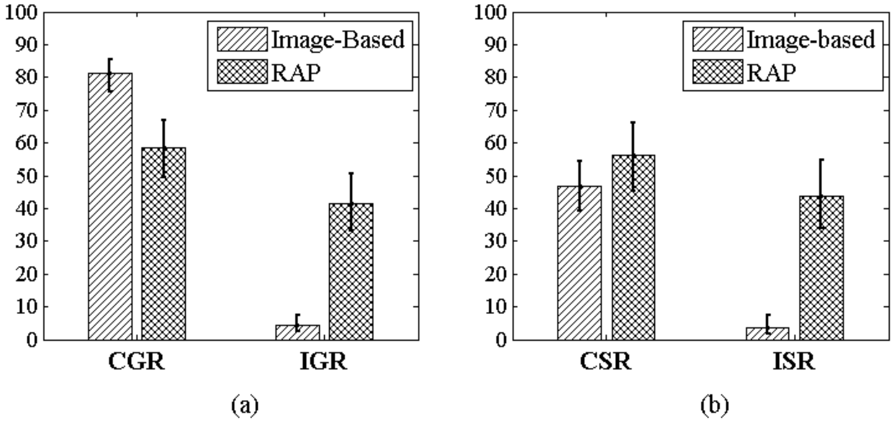


Figure 1.5: Accuracy of image-based classification. (a) correct and incorrect classification rates at the genus level; and (b) correct and incorrect classification rates at the species level. Error bars represent 95% confidence intervals. As can be seen, when identifying specimen genera from images, the expert is correct 81% of the time, and is only incorrect 4.2% of the time. On the other hand, at the species level images can only correctly identify a specimen 47% of the time. Despite the poor correct classification rate at the species level, the incorrect classification rate remained very low.

1.4.2.2 Performance of Computer-Aided Identification

The performance and limitations of image-based identification explored in [31], which the previous section summarized, steered the preliminary implementation of computer-aided identification toward only aiming for genus-level identification. Using the same dataset in [31], the computer-aided classifier in [24] applied the clustering scheme outlined in Sec. 1.4.1 to the samples. This section presents the results in [24], with particular emphasis on elements influencing the course of this thesis.

Ideally, the number of clusters should be as low as possible. While, the correlation coefficient as a similarity measure can be used to cluster specimens together, clusters must also be as homogenous as possible to be useful. In addition, the system must also choose an appropriate template for each cluster and present it to the expert for identification. Finally, an expert must produce accurate classifications of each template. These challenges present the computer-aided classifier with three possible sources of error:

1. Inhomogeneity of clusters;

2. Inability of the template(s) to represent the majority of members in its cluster;

3. Inaccuracy of identifying a template by its image-based representation.

The first error is primarily a judgement on the performance of the clustering algorithm. However, if a cluster is completely homogenous, then the second error, in addition to the first, is completely negated, as all specimens in the cluster are equally appropriate representatives. Even so, for any useful amount of work reduction, heterogenous clusters are unavoidable, which places considerable importance on selecting appropriate templates. This makes it somewhat difficult to separate the contributions of Errors 1 and 2 from each other. Finally, if the expert is performing particle-based template identification, Error 3 does not come into play.

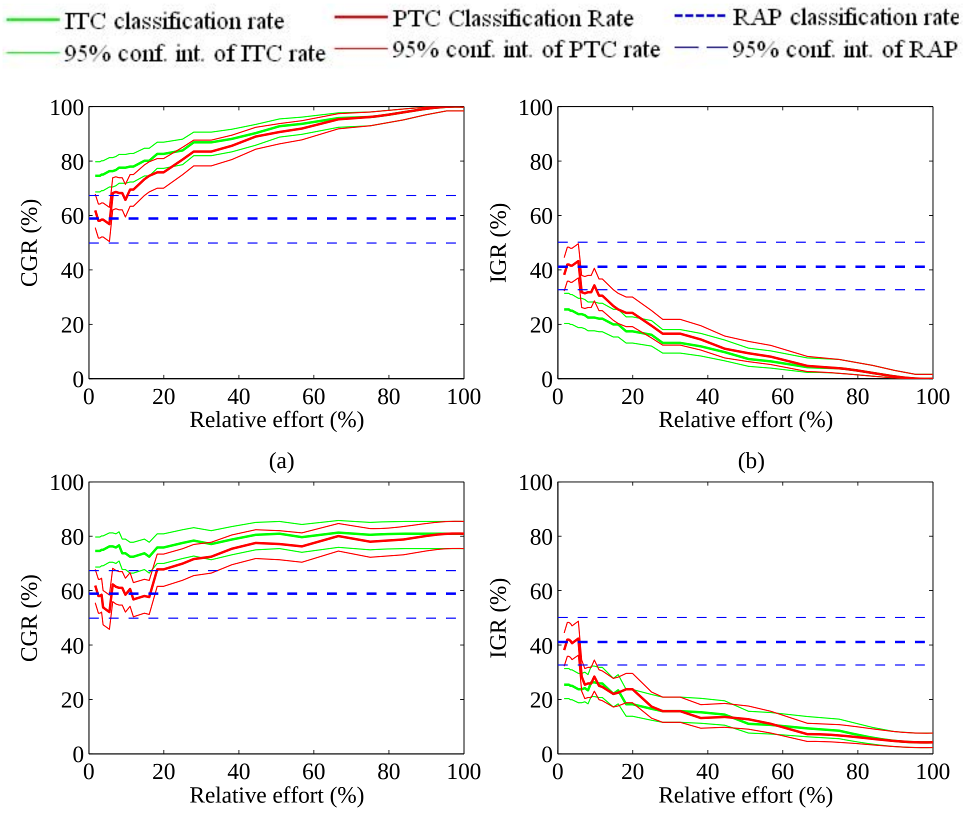
Constructing a classifier that chooses the ideal template every time is an excellent way to measure the contribution of Error 1 without having Error 2 confound the results. This can be accomplished by labelling a cluster with the class belonging to the majority of specimens in the group. Since such a classifier requires that every specimen be classified in the sample set, this is not a practical scheme; however, for the purposes of analysis it does allow the clustering algorithm to be judged on its own merits. Such a classifier will be called an ideal-template classifier (ITC). Alternatively, a practical-template classifier (PTC), the scheme used in an actual implementation of computer-aided identification, chooses templates based on the visual similarity criteria outlined in Sec. 1.4.1.6.

Measuring cluster count vs. similarity threshold revealed that the number of clusters varies nonlinearly with the similarity threshold parameter [24]. The nature of this nonlinear relationship holds sway over the degree of work reduction the system can provide—which is a major criterion by which to judge the performance of the system. Performance of a computer aided classifier is only meaningful when compared against the work it reduces. System accuracy and reliability can be quantified using CGR and IGR values, but quantifying work reductions requires another measure called *relative effort*:

$$\textit{Relative effort} = \frac{\textit{Number of templates examined}}{\textit{Total number of specimens}}. \tag{1.6}$$

Fig. 1.6 demonstrates the contribution of each error to the existing computer-aided classification scheme by graphing the CGRs and IGRs of both the ITC and PTC across different degrees of relative efforts. The top row of Fig. 1.6 depicts classification rates using particle-based template identification, while the bottom row illustrates image-based template identification rates.

The superimposition of the ITC onto Fig. 1.6 offers an excellent way to study the system limitations more closely. For instance, as the particle-



(a)

(b)

(c)

(d)

Figure 1.6: Performance of image-based computer-aided identification. (a) and (b) give the correct and incorrect genus rates of the *particle-based* ITC and PTC; whereas, (c) and (d) do the same for the *image-based* classifiers. In (a) and (b), the CGR and IGR of the ITC demonstrates the performance and error contributions of the clustering algorithm (Error 1). A comparison between the ITC and PTC rates of (a) and (b) illustrates the degree of error resulting from incorrect template selection (Error 2). Finally, comparing the performance between the image-based and particle-based classifiers of the top and bottom rows respectively illustrates errors stemming from identifying templates by their image-based representations (Error 3).

based ITC of Fig. 1.6(a) and (b) chooses ideal templates and identifies them without error, focusing on these results uncovers the contribution of cluster inhomogeneity, or Error 1, to the system limitations. As evidenced by the ITC performance, the clustering algorithm does perform very well at relative efforts of 60% and higher. However, at lower amounts of relative effort the clustering algorithm begins to create clusters with greater heterogeneity. As a result, there is need to improve automatic clustering performance.

Differing from the ITC, the PTC is subject to errors resulting from incorrect template selection. By comparing the ITC and PTC rates of the first row of Fig. 1.6, one can notice that for relative efforts of 40% and above, template selection performs well. However, at lower relative efforts the system begins to increasingly select templates that do not best represent their clusters. While limitations associated with template selection certainly do contribute errors, when compared to issues arising from inhomogeneity of clusters, the magnitude of the errors is relatively small.

In contrast to any issues related to template selection, limitations associated with image-based template identification can clearly become a dominating factor. Although Error 3 is only applicable when templates are identified by their images, the greater ease and flexibility of using a digital representation for template identification establishes the importance of minimizing this error. As previously demonstrated in Sec. 1.4.2.1, when an expert identifies forams by their images, he or she can only correctly identify their genus 81% of the time. It is for this reason that the CGRs of the classifiers using image-based template identification in Fig. 1.6 (c) is also limited to 81%. Even so, the system maintained low IGRs. These results indicate that the limitations of image-based representations for templates can dominate all other factors hampering classification performance. Mitigating this problem requires alternative digital representations for templates.

While the results, especially the consistent low IGRs, are promising preliminary findings, performance must be improved to progress the computer-aided identification system toward practical uses. The results of the preliminary system indicate that cluster inhomogeneity (Error 1) and template identification errors (Error 3) are responsible for the majority of performance limitations. Comparatively, errors in the template selection scheme (Error 2) are negligible. As the following chapter will continue to emphasize, system performance is hampered by the inherent limitations of image-based representations. This affects the system's performance by limiting the ability of experts to identify templates through their digital representations. Alleviating the magnitude of these errors may enable species-level classification of microfossils—a task the preliminary system is simply unable to accomplish. Incorporating alternative digital representations into computer-aided classification represents a crucial area of future work and is a main focus of this thesis.

1.5 Organization of Thesis

As argued in the preceding sections, computer-aided classification offers the most promising direction yet in automating microfossil identification in a practical and usable manner. Although the preliminary computer-aided identification system developed at the UA Electronic Imaging Lab successfully demonstrated feasibility, it suffers from significant limitations. In its broadest sense, the goal of this thesis can be seen as addressing and alleviating some of these limitations by investigating digital representations other than images.

Chapter 2 explores the limitations of the preliminary system further and argues that encapsulating the information across multiple light directions can produce richer digital representations for templates. In doing so, the chapter makes the case for the use of computer vision in computer-aided classification. As part of this exploration, photometric stereo and maximum likelihood concepts are introduced. Chapter 3 details the significant extensions made to the computer-aided system to allow the incorporation of two alternative digital representations into the identification scheme. As well, the chapter summarizes results on data collection using the extended system. Chapter 4 introduces a novel image alignment technique developed to address localization problems in the extended system. The applicability of the alignment technique is broad enough to include any photometric stereo image sequences suffering from relative misalignments from image to image. Finally, Chapter 5 describes a novel visual-surface reconstruction technique using maximum likelihood estimation. This technique is able to be used, amongst other applications, to construct 3D models of microfossils.

Chapter 2

Computer Vision for Computer-Aided Identification

In its current form, the state of the art of computerized identification of microfossils focuses on techniques applied to intensity levels of simple images. In other words, in addition to employing concepts related to machine learning, the current state of computerized microfossil identification uses techniques belonging to the field of image processing. While pixels of single images do contain a great amount of information, image processing techniques are essentially restricted to working with and manipulating 2D data [33]. However, images often contain a great deal of information regarding scene geometry, environment, and motion. Techniques designed to extract this type of information fall under the field of computer vision [33, 34].

While the above description may seem to adequately separate and define the two fields of image processing and computer vision, the reality is there is a considerable degree of uncertainty regarding the scope of computer vision. For instance, Shapiro and Stockman offer the following somewhat ambiguous definition of computer vision as a field whose goal, "is to make useful decisions about real physical objects and scenes based on sensed images" [33]. On the other hand, Forsyth and Ponce define computer vision as a collection of techniques, not necessarily theoretically grounded, designed to extract, "descriptions of the world from pictures or sequences of pictures" [34]. The unclear boundaries of study of computer vision are probably best justified by its relatively young state of development—"an intellectual frontier" [34]. Despite this difficulty in exactly nailing down a definition of computer vision, one commonality holds amongst all aspects of computer vision. Computer vision is concerned with *understanding* images rather than simply working with them. This perhaps explains why Shapiro and Stockman conclude their introduction by equating computer vision with the term *image understanding* [33]—a term general enough to encapsulate any computer vision technique.

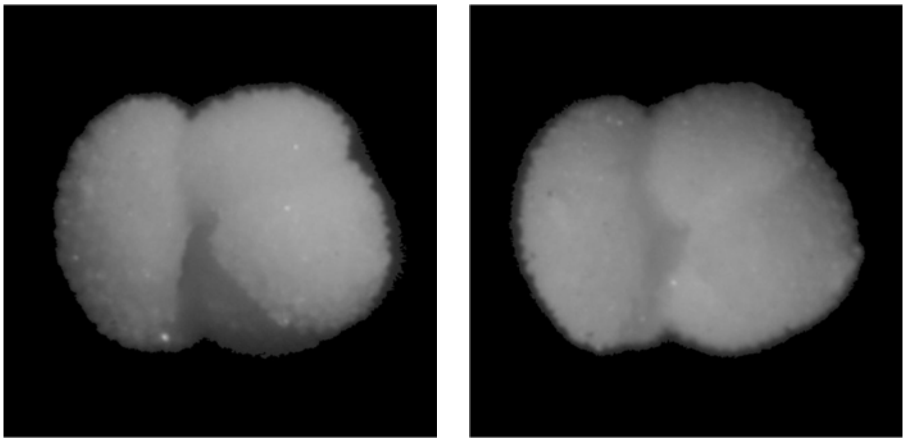
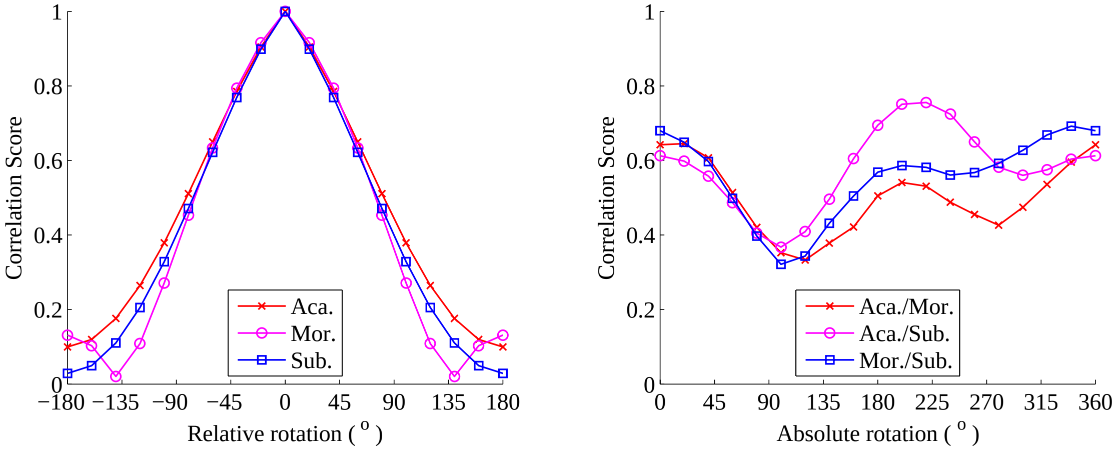


Figure 2.1: The significant role of illumination. Here, an example illustrates the significant role illumination plays in microfossil image formation. This figure depicts the same microfossil specimen (a foraminifera of the genus Acarinina), illuminated from different directions. Notice how different the aperture (opening) in the bottom section of the microfossil looks in both images.

As may be expected, the large scope of computer vision allows a great deal of overlap with image processing. For instance, texts of image-processing [29] and texts on computer vision [33, 34] both deal with topics such as edge detection, linear filters, and texture to name a few. But apart from these areas of overlap, many computer vision techniques can often be thought of as attempting to solve an inverse projection of some sort; given one or more images, they attempt to extract the underlying scene geometry, lighting, motion, or perhaps all three. For the most part, this is a difficult and ill-conditioned task. As a result, much of computer vision is composed of diverse techniques meant to solve problems with very specific tasks and underlying assumptions related to calibration, surface geometry, camera viewpoint, and camera and reflectance models [34].

2.1 The Role of Computer Vision

Although work in [24] demonstrated that effort can be significantly reduced using a semi-automated clustering approach, classification accuracy was still hampered by several limitations. Probably the most serious limitation was that the work in [24] did not control for light direction with respect to a specimen's principal axis. As shown in Fig. 2.1, microfossils can look significantly different depending on illumination direction. Fig. 2.2(a) illustrates the effect that this variability across illumination directions has



(a) (b)

Figure 2.2: The effect of illumination direction on similarity scores (taken from [24]). In (a), the average correlation scores are graphed between images of the *same* microfossil at *different* relative rotations with respect to the light source. In (b), the correlations scores of pairs of *different* specimens at the *same* absolute rotations are graphed. For both (a) and (b) the same three specimens were used, each representing one of the foram genera of Acarinina, Morozovella, and Subbotina. The graph in (b) demonstrates that depending on the absolute illumination direction, clustering thresholds less than 0.75 may group specimens from different genera together. However, as (a) demonstrates, differences in relative illumination direction can produce similarity scores much less than 0.75 between images of the *same* specimen.

on similarity scores. As the figure demonstrates, the correlation scores of the same particle photographed with different illumination directions exhibit extremely low values. But, scores of 0.75 or higher were needed for accurate classification. Consequently, without controlling for illumination there is no guarantee that particles of the same genus will be clustered together at that threshold. As a result, extending the system to control illumination direction is a necessary improvement toward increasing classification performance.

Should illumination direction be controlled, one could continue to use single images, but ensure that all microfossils are photographed having the same angle to the light source with respect to their principal axes. This strategy promises to reduce intra-class variability, allowing the system to cluster a greater number of specimens together. Yet, doing so requires fixing the angle of illumination relative to each specimen's principal axis to an arbitrary value. However, as Fig. 2.2(b) demonstrates, correlation values between images are strongly dependant on the absolute angle of illumination. As a result, the choice of which illumination angle to fix will affect classification performance.

These problems associated with calculating similarity can be avoided by capturing more than one image of each specimen, with each image illuminated from a different direction. Assuming the dataset consists of N images per specimen, one could, for instance, calculate similarity between all N possible illumination directions for each pair and choose the median score. Thus, similarity is guaranteed to be calculated between images illuminated with identical conditions, and there is no need to fix absolute illumination direction to an arbitrary value.

While such a scheme addresses problems with calculating similarity, the fact remains that experts are still unable to identify template specimens using image-based representations. In Sec. 1.4.2.2, this inability to identify templates using image-based representations was denoted Error 3 and was responsible for a large degree of performance limitations. While these issues can be avoided by using particle-based template identification, such a scheme is not nearly as practical as using digital representations. For one, digital representations allow for fast and remote classifications. As well, there are no equipment needs other than a computer. In addition, another expert can easily identify the same digital representation without having to physically inspect the same physical particle. Thus, there is no need to retrieve physical particles from their repository. For these reasons, identifying specimens using their digital representations is much more desirable than using particle-based identification.

One problem with image-based classification is that microfossils are distinguished by the peculiarities and characteristics of their 3D morphology [35]. While an image may adequately represent one or more of these features, it will not always be able to represent all of them at once. For instance, the severity of the aperture at the bottom of the microfossil is apparent in the image of Fig. 2.1(a), while hardly so in Fig. 2.1(b). However, the image in Fig. 2.1(b) succeeds at better depicting the division of lobes on the right-hand side of the fossil compared to Fig. 2.1(a). These inherent limitations in images call for an alternative direction on digital representation—one moving away from image-based techniques.

The example in Fig. 2.1 illustrates that multiple images, each illuminated by their own respective light direction, provides a better insight into the actual 3D morphology of a microfossil than just using a single image. Controlling for illumination direction is key to unlocking this information. This insight motivates using a *video-based* representation for templates rather than an image-based one. By providing a video of a single specimen successively illuminated at different angles, experts would have an increased ability to *understand* the underlying geometry of template fossils.

This could alleviate the genus-level limitation currently holding back template identification. As Sec. 3.1.3 will discuss, capturing video sequences free of relative misalignments requires the application of computer vision techniques.

While videos would certainly provide a richer template representation than images, computer vision techniques can supply an even more powerful option. Similar to what the human visual system unconsciously performs, one promising avenue is to exploit the image variability across differing light directions and extract a *shape-based* representation of the particle in the form of a visual surface. Allowing experts to control viewing angle and lighting conditions, 3D models would serve as excellent representations to use for templates and could provide even greater benefits than their video-based counterparts.

Representing another motivation for the incorporation of computer vision into the system, 3D models of microfossils are in of itself a useful endeavour for micropalaeontologists. Detailed 3D models of macrofossils of extinct species, such as mollusca [36] and crustaceans [37], have contributed knowledge toward the prevailing understanding of these species. In the context of microfossil study, 3D modelling would also have the potential to improve current taxonomic understanding. However, the two cited works employed micro-grinding, which is a destructive and time-consuming process. An ideal modelling method would use computer-vision techniques to develop sufficiently detailed models—the direction approached for this thesis.

Moving away from the previous work's reliance on image-based representations, video and shape-based representations of templates opens up the possibility of a computer-aided classifier able to effectively reduce work not only at the genus level, but potentially at the species level too. Controlling illumination direction is key for both video-based and shape-based representations.

2.2 Image Formation and Surface Reconstruction

The inherent limitations of images in their ability to represent 3D microfossil morphology motivates work using shape-based digital representations. As mentioned, computer-vision techniques are often designed to solve ill-posed and very specific problems. Fortunately, the particulars of microfossil examination satisfy the conditions of a well understood and relatively mature computer vision technique—visual-surface reconstruction using photometric stereo.

In its simplest formulation, reconstructing a visual surface requires a sequence of images with known and controlled illumination all under the same viewpoint. As a result, the only information produced is based on what is visible from the viewing angle. For this reason, the extracted shape, $Z(x,y)$, is often described as 2.5D instead of 3D. The result can also be treated as an image, where each pixel location has an associated depth value. In this work, the extracted shape will often be called a *depth map*.

When viewing an object, the shading of a surface patch is related to its orientation relative to the light source. Under certain assumptions including Lambertian reflectance, orthographic cameras, a principal light source at infinity, and no inter-reflections or shadows, this relationship is linear in nature. Fortunately, microfossils possess surfaces with minimal specularity and highlights, acting as good approximations to Lambertian surfaces. Using these assumptions, the intensity of a noiseless image pixel at coordinates x and y can be expressed in a simple form:

$$I(x,y) = \boldsymbol{\ell}^T \cdot \eta(x,y), \tag{2.1}$$

$$\eta(x,y) = \rho(x,y) \cdot \mathbf{n}(x,y), \tag{2.2}$$

where $\mathbf{n}(x,y) = (n_x, n_y, n_z)^T$ represents surface normals, $\rho(x,y)$ is the surface albedo, and $\boldsymbol{\ell}$ is the light direction expressed as a unit vector. Here, $\eta(x,y)$ represents the surface normals multiplied by the albedo, and will be referred to as *weighted normals*. Typically, coordinates are expressed using a coordinate system based on the supporting surface of the object, with the positive z axis pointing toward the camera. Unless the specification of a particular point is needed, for convenience the rest of this discussion will drop the coordinate points in the notation. As well, without loss of generality, it is assumed that images are of $n \times n$ dimensions.

Since surface normals are defined by the underlying object they describe, one can use the depth map itself as a generator responsible for image formation. First, denote the gradients of the depth map or surface by the following notation:

$$p = \frac{\partial Z}{\partial x}, q = \frac{\partial Z}{\partial y}. \tag{2.3}$$

Denoted this way, the gradients share the following relationship with surface normals [38]:

$$\mathbf{n} = \frac{1}{\sqrt{p^2 + q^2 + 1}} \begin{pmatrix} -p \\ -q \\ 1 \end{pmatrix}. \tag{2.4}$$

From the above formulation, it is clear that the surface normals, like the depth map and its two gradient scalar fields, p and q, are functions of x and

y and not of z. This is a by-product of the 2.5D nature of depth maps and the orthographic camera assumptions. As a result, for any constant c, Z and $Z+c$ will produce identical surface normals.

With surface normals formulated in this manner, the noiseless image formation of (2.1) can be reexpressed as:

$$I = \boldsymbol{\ell}^T \frac{\rho}{\sqrt{p^2+q^2+1}} \begin{pmatrix} -p \\ -q \\ 1 \end{pmatrix}. \tag{2.5}$$

Assuming one possesses a set of images, each under identical viewpoints but differing light directions, one can combine (2.5) for every image into a set of equations. Doing so constructs a nonlinear system of equations to simultaneously solve for Z and ρ. However, in the discrete setting of images, incorporating partial derivative terms in the formulation requires the use of finite differences. Consequently, Z values at specific locations cannot be calculated in isolation of one another. As the size of the image dimensions increase, the formulation also quickly balloons into a very large scale problem, requiring the simultaneous solution of n^2 parameters. Thus, solving for Z directly is a difficult nonlinear problem. For this reason, surface reconstruction is often executed using an intermediate step.

Known as photometric stereo, this intermediate step computes the surface normals and albedo of an object from a set of images. Originally developed by Woodham [39], photometric stereo requires three or more images of an object, given Lambertian reflectance assumptions, under identical viewpoints, but with a known varying principal light source. When executed under these assumptions, the process of calculating normals has not changed significantly since its formulation.

If a sequence of $m \geq 3$ images of the same object, under identical viewpoints, is taken with differing light sources, then each respective image formation equation contributes to the following system of equations:

$$\begin{pmatrix} I_1 \\ \vdots \\ I_m \end{pmatrix} = \begin{pmatrix} \boldsymbol{\ell}_1^T \\ \vdots \\ \boldsymbol{\ell}_m^T \end{pmatrix} \boldsymbol{\eta}. \tag{2.6}$$

When it is over-determined, i.e. $m > 3$, the least-squares solution to the weighted-normal vector, $\boldsymbol{\eta}$, is obtained. Estimated normal and albedo values, $\mathbf{n}$ and ρ, are computed from the weighted-normal estimate using (2.2). Because all images are taken from the same viewpoint, surface patches directly correspond to pixel coordinates x and y. Apart from the linear nature of the photometric stereo step, the appeal of executing this intermediate step lies in the fact that surface normal and albedo solutions can be computed separately for each pixel location. As a result, solving a large nonlinear system is not necessary.

Although most of the assumptions for classic photometric stereo hold for the purposes of microfossil identification, the assumption of no shadows is a frequently violated condition. Shadows can typically be divided into cast and attached shadows. A cast shadow occurs when certain geometries of an object, such as a peak, block illumination from reaching other areas of the object. On the other hand, an attached shadow occurs when a surface normal faces away from the illumination direction. Mathematically, this is expressed as $\boldsymbol{\ell}^T \mathbf{n} < 0$. The most intuitive way to handle shadows is to incorporate a vector $\boldsymbol{\omega}(x,y) = \{\omega_1(x,y)...\omega_n(x,y)\}$ into (2.6) such that:

$$\omega_k(x,y) = \begin{cases} 0 & \text{if } I_k(x,y) \text{ is shadowed} \\ 1 & \text{otherwise} \end{cases} . \tag{2.7}$$

The system of equations to solve for the normal then becomes:

$$\begin{pmatrix} I_1 \\ \vdots \\ I_m \end{pmatrix} = \begin{pmatrix} \omega_1 \boldsymbol{\ell}_1^T \\ \vdots \\ \omega_m \boldsymbol{\ell}_m^T \end{pmatrix} \boldsymbol{\eta}. \tag{2.8}$$

One simple method to determine values for the $\boldsymbol{\omega}(x,y)$ vector is to exclude pixels below a certain value or to compute a histogram of pixel values at each pixel location and simply exclude those in the bottom percentage [40]. Unfortunately, choosing a pixel value threshold or percentage for exclusion is entirely arbitrary. As well, the appropriateness of a choice will not be valid across all images and noise conditions.

Ideally, methods to determine shadows should not rely on arbitrary parameter values. One such method to determine attached shadows, not found in the literature, uses a simple heuristic to determine light directions facing away from the surface normal. First, a normal estimate, $\mathbf{n}$, is produced using all available light directions. Using this estimate, cosines of the angles between the normal and each light direction can be computed using the dot product between them: $\mathbf{Ln}$, where $\mathbf{L} = \{\boldsymbol{\ell}_1...\boldsymbol{\ell}_m\}^T$. Light directions facing away from the normal estimate correspond to negative cosine values. These light directions are then excluded. The shadowed photometric stereo equation of (2.7) is then used to compute a new normal estimate. This process is repeated until no light directions are deemed facing away from the current normal estimate. In the context of this thesis, this method produces better results than the simple threshold or histogram method, making it the technique of choice.

Together, these albedo and surface normals are sufficient to describe the object but do not provide a visual surface. To reconstruct a surface, one must first convert surface normals to surface gradients:

$$\begin{pmatrix} p \\ q \end{pmatrix} = \begin{pmatrix} \eta_x/\eta_z \\ \eta_y/\eta_z \end{pmatrix}. \tag{2.9}$$

While solving for surface normals is a linear problem, computing the gradients requires the nonlinear relationship in (2.9). Fortunately though, (2.9) can be solved separately for each pixel location. Once a suitable estimate of the p and q values at each pixel location is obtained, the two gradient fields can be integrated to reconstruct the object's depth map. Although the integration involves simultaneously solving the depth map at all pixel locations, the problem is now linear in nature.

Unlike Lambertian photometric stereo, visual-surface reconstruction remains an open problem. In particular, state of the art methods struggle to account for image noise when constructing depth maps. Addressing this shortfall, this thesis develops a visual-surface estimation method robust to image noise. Detailed in Chapter 5, this method employs concepts related to maximum-likelihood estimation, which are discussed in the following section.

2.3 Maximum Likelihood Estimation

The image formation equation of (2.1) does not include a noise model. Yet, noise is an unavoidable aspect of images. In the context of visual-surface construction, several authors have also recognized the importance of addressing noise [41–46]. As well, incorporating stochastic noise terms into surface normal estimation provides key insight into the validity of the estimates, including modelling characteristics related to the uncertainty or confidence of the normal estimate values. These are important in their own right, but they also can, and arguably should, play a crucial role when estimating the 3D surface from the normals—something that has only been touched upon in the literature. As noise is typically modelled as some random process following a known distribution, estimating parameter values in the presence of said noise is a task falling under the scope of maximum likelihood (ML) estimation. As a result, ML estimation plays an important role in this thesis, and it is worthwhile devoting space toward a summary of its concepts and notation.

The topic of ML estimation comes into play when an estimate of one or more parameters is desired given a set of observations and an underlying relationship or model. This type of problem also falls under the scope of regression analysis. Restricting the discussion to the univariate case, a relationship can be expressed formally as:

$$y \approx f(\mathbf{x}), \tag{2.10}$$

where y is the observed or dependant variable, f is the underlying model, often some sort of physical process, and $\mathbf{x} = (x_1, x_2, \ldots, x_k)^T$ is a $k \times 1$ vector of *independent* variables (also sometimes called explanatory or regressor variables) whose values in combination with the model explain the behaviour of y. As is often the case, observations are often susceptible to uncertainty based on random errors. For this reason, the relationship in (2.10) is expressed as an approximation. Although not always the case, the uncertainty can often be expressed as additive errors. The restriction to additive errors or noise greatly simplifies estimation, and for the rest of this discussion noise will assumed to be of that form.

When one or more parameters of the model is unknown, (2.10) is expressed as:

$$y = f(\mathbf{x}, \boldsymbol{\beta}) + \epsilon, \tag{2.11}$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)^T$ is a $p \times 1$ vector of unknown variables to be estimated. Note that (2.11) is expressed using the additive error assumption; thus, unlike (2.10), the relationship is no longer an approximation.

In the midst of the uncertainty provided by ϵ, parameter estimation can be based on different criteria. For instance one may estimate parameters to minimize a modelling metric like the sum of the squared errors, to realize the minimum parameter variance, or to maximize the likelihood of observations (ML estimation). As will be seen, under certain assumptions, a parameter estimate satisfying one of these criterions will inherently satisfy in some manner all three. Regardless of the estimation method, the regression problem can be expressed as determining $\hat{\boldsymbol{\beta}}$, the best estimate of $\boldsymbol{\beta}$, which can usually only be done given more than one observation of the type in (2.11). Mathematically, this can be written as:

$$\mathbf{y} = \mathbf{f}(\boldsymbol{\beta}) + \boldsymbol{\epsilon}, \tag{2.12}$$

where $\mathbf{y} = (y_1, y_2, \ldots, y_m)^T$ is a vector of m observations and $\boldsymbol{\epsilon}$ is the set of associated random errors. The $\mathbf{f}(\boldsymbol{\beta})$ notation represents the model output based on the associated independent vector for each observation and a parameter estimate common to all observations. Formally, let $f_i(\boldsymbol{\beta}) = f(\mathbf{x}_i, \boldsymbol{\beta})$ be the prediction based on the parameter vector $\boldsymbol{\beta}$ and one set of independent variables. Then define $\mathbf{f}(\boldsymbol{\beta})$ as $(f_1(\boldsymbol{\beta}), f_2(\boldsymbol{\beta}), \ldots, f_m(\boldsymbol{\beta}))^T$.

Focusing on ML estimation, an estimate of this type maximizes the likelihood of the given observations, based on an assumption of the underlying distribution of the error variables. If $\mathscr{L}$ denotes the likelihood operator, and the probability density function (PDF) of the errors is given as $p(\boldsymbol{\epsilon})$, then an ML estimate of $\boldsymbol{\beta}$ is one that satisfies:

$$\hat{\boldsymbol{\beta}} = \underset{\tilde{\boldsymbol{\beta}}}{\operatorname{argmax}} \mathscr{L}\left\{\mathbf{y} \mid \tilde{\boldsymbol{\beta}}, p\left(\mathbf{y} - \mathbf{f}(\tilde{\boldsymbol{\beta}})\right)\right\}, \tag{2.13}$$

where $\tilde{\boldsymbol{\beta}}$ can be any valid parameter estimate. Should the components of the error vector be identically and independently distributed (IID), (2.13) can be simplified with a product of m univariate PDFs, $p(\epsilon_i)$:

$$\hat{\boldsymbol{\beta}} = \underset{\tilde{\boldsymbol{\beta}}}{\operatorname{argmax}} \prod_{i=1}^{m} p(y_i - f_i(\tilde{\boldsymbol{\beta}})). \tag{2.14}$$

The PDF of the errors greatly affects the manner in which the ML estimate of $\boldsymbol{\beta}$ is calculated. The normal distribution is a useful and common assumption [47], and will be used throughout the rest of this discussion. Here, a normal distribution with mean $\boldsymbol{\mu}$ and an $m \times m$ covariance matrix $\sigma^2\boldsymbol{\Gamma}$ is denoted as $\mathscr{N}(\boldsymbol{\mu}, \sigma^2\boldsymbol{\Gamma})$. When the errors are all homogenously distributed according to $\mathscr{N}(\mathbf{0}, \sigma^2\mathbf{I})$, ML estimation proves equivalent to minimizing the sum-squared error [48]. If the sum-squared error (SSE) is expressed as:

$$SSE(\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{f}(\boldsymbol{\beta}))^T(\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})), \tag{2.15}$$

then the ML estimate of (2.14) is:

$$\hat{\boldsymbol{\beta}} = \underset{\tilde{\boldsymbol{\beta}}}{\operatorname{argmin}}\, SSE(\tilde{\boldsymbol{\beta}}). \tag{2.16}$$

Minimizing the expression in (2.16) is equivalent to an ordinary least-squares (OLS) problem. However, often errors are heterogeneously distributed according to some covariance matrix $\sigma^2\boldsymbol{\Gamma} \neq \sigma^2\mathbf{I}$, and the simplification of (2.14) no longer applies. Situations of this type are equivalent to generalized least squares (GLS) problems. Usually though, the covariance matrix can be decomposed, i.e. into $\sigma^2\boldsymbol{\Gamma} = \sigma^2\mathbf{B}^T\mathbf{B}$. It may be possible to determine $\mathbf{B}$ directly, or alternatively $\mathbf{B}$ can be calculated using Cholesky factorization. Assuming that $\boldsymbol{\Gamma}$ is fully ranked, (2.12) can then be pre-multiplied by $\mathbf{B}^{-T}$, transforming the GLS problem into an OLS formulation [48]:

$$\mathbf{B}^{-T}\mathbf{y} = \mathbf{B}^{-T}\mathbf{f}(\boldsymbol{\beta}) + \mathbf{B}^{-T}\boldsymbol{\epsilon}, \tag{2.17}$$

$$\mathbf{y}' = \mathbf{f}'(\boldsymbol{\beta}) + \boldsymbol{\epsilon}', \tag{2.18}$$

$$\boldsymbol{\epsilon}' \sim \mathscr{N}(\mathbf{0}, \sigma^2\mathbf{I}). \tag{2.19}$$

This is equivalent to minimizing the weighted sum-squared error:

$$SSE(\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{f}(\boldsymbol{\beta}))^T\boldsymbol{\Gamma}^{-1}(\mathbf{y} - \mathbf{f}(\boldsymbol{\beta})). \tag{2.20}$$

The method of determining the parameter estimate $\hat{\boldsymbol{\beta}}$, which minimizes $SSE(\boldsymbol{\beta})$, depends on whether the model function is linear with respect to the parameters or not. A discussion of each case follows below.

2.3.1 The Nonlinear Case

In its most general form, ML estimation incorporates a nonlinear model function, meaning that estimation requires the techniques of nonlinear regression. Due to the inherent difficulties associated with nonlinear regression, many of the field's methods and results are restricted to model formulations involving IID and normally distributed errors whose means are zero. As a result, texts on nonlinear regression either outright restrict their attention to situations where errors are IID [47], or simply recommend to transform all GLS problems to OLS ones whenever possible [48,49]. For this reason, this discussion will also only focus on the IID errors case. For cases where the model formulation consists of correlated and/or non-identical error variances, the discussion will assume that a suitable form of $\mathbf{B}^{-T}$ is available, transforming the problem into an OLS one as expressed in (2.17).

Regardless of the form of the model function, when the errors follow the above assumptions, the ML estimate of the parameters is the estimate minimizing the sum-squared error. Unfortunately, for finite sample sizes, it is often difficult to accurately describe the estimate's statistical behaviour [48]. However, under certain regularity conditions, the estimator, $\hat{\boldsymbol{\beta}}$, follows some key properties as the sample size, m, increases to infinity. Known as asymptotic properties, one of the most important of these properties is consistency, meaning that as the sample size increases, the estimated parameters converge to the true parameters. In other words, the estimates are asymptotically unbiased. Having asymptotically unbiased estimates is a prerequisite for having asymptotic normality, which is another very useful property. In general, verifying consistency and asymptotic normality is a difficult problem. Most texts on nonlinear regression limit their focus to problem formulations where the estimators are assumed to be asymptotically unbiased and normal [48–50]. This discussion will also assume asymptotic unbiasedness and normality. For a good discussion of this topic, including areas outside that of least-squares estimators please see [51].

Provided that consistency and asymptotic normality hold, the distribution of $\hat{\boldsymbol{\beta}}$ obeys certain properties, with a sufficiently large m. Notably, $\hat{\boldsymbol{\beta}}$ approaches a minimum possible variance [48]. Defining $\mathbf{F}_{\cdot}(\boldsymbol{\beta})$ as the Jacobian matrix of the model function:

$$\mathbf{F}_{\cdot}(\boldsymbol{\beta}) = \frac{\partial \mathbf{f}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T}, \tag{2.21}$$

where $\mathbf{F}_.(\boldsymbol{\beta})$ is dependant on both the underlying model function and the values of the parameters, then a minimum bound on the covariance of $\hat{\boldsymbol{\beta}}$ is expressed as $\sigma^2\boldsymbol{\Gamma}$, where $\boldsymbol{\Gamma}^{-1} = \mathbf{F}_.(\boldsymbol{\beta})^T\mathbf{F}_.(\boldsymbol{\beta})$ [48]. The distribution of $\hat{\boldsymbol{\beta}}$ can now be described as asymptotically belonging to $\mathcal{N}(\boldsymbol{\beta}, \sigma^2\boldsymbol{\Gamma})$. Since $\boldsymbol{\beta}$ is not known, $\mathbf{F}_.(\boldsymbol{\beta})$ must be approximated by $\mathbf{F}_.(\hat{\boldsymbol{\beta}})$, which produces an approximation of the asymptotic covariance, $\sigma^2\hat{\boldsymbol{\Gamma}}$. How closely $\hat{\boldsymbol{\beta}}$ follows these asymptotic properties depends on whether m is sufficiently large. In addition, the closer models are to linear models, the closer their associated parameters behave according to the asymptotic limits [47]. While an appropriate value of m cannot be analytically determined, the degree of nonlinearity of a model can. Based on analysing curvature, these measures can be used to predict the degree of deviation from the asymptotic limits [48]. Although certain types of curvature are intrinsic to the model, other types can be mitigated through reparameterization [48].

With this in mind, in both of his books Ratkowsky espouses reparameterization of a model to a form that is as "close" to linear as possible [47,50]. For this reason, Ratkowsky differentiates between intrinsic models, and model functions resulting from different parameterizations. If a model is reparameterized into a "more" linear model using:

$$\mathbf{f}_r(\mathbf{g}(\boldsymbol{\beta})) = \mathbf{f}(\boldsymbol{\beta}), \tag{2.22}$$

$$\mathbf{y} = \mathbf{f}_r(\boldsymbol{\phi}) + \boldsymbol{\epsilon}, \boldsymbol{\phi} \qquad = \mathbf{g}(\boldsymbol{\beta}), \tag{2.23}$$

$$\tag{2.24}$$

where the reparameterized vector $\boldsymbol{\phi}$ is only dependant on values of $\boldsymbol{\beta}$ and not on any elements of the intrinsic model, then the statistical properties of $\boldsymbol{\phi}$ will behave more closely to the asymptotic limits. This is particulary advantageous when the sample size is low. In addition, it will be easier to numerically determine the ML estimate, $\hat{\boldsymbol{\phi}}$, as the linear approximations typically used in iterative techniques will have less error [50]. As well, (2.22) can be reversed to provide a parameterization producing $\boldsymbol{\beta}$ from $\boldsymbol{\phi}$:

$$\boldsymbol{\beta} = \mathbf{h}(\boldsymbol{\phi}), \tag{2.25}$$

Unfortunately, there is no universal technique of determining optimal reparameterizations. In certain situations, suitable reparameterizations to pursue are obvious. For more difficult model functions, suitable parameterizations tends to involve case-by-case analyses by determining the nonlinear contribution each parameter brings to the model function and also by assessing the improvements, if any, of a candidate reparameterization [50].

Relationships also exist between covariances of two different parameterizations. Denoting $\sigma^2\boldsymbol{\Gamma}_{\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}}$ and $\sigma^2\boldsymbol{\Gamma}_{\hat{\boldsymbol{\phi}}\hat{\boldsymbol{\phi}}}$ as the covariances of the nonlinear and more linear parameterizations respectively, the relationship between them evaluates as [50]:

$$\sigma^2 \mathbf{\Gamma}_{\hat{\beta}\hat{\beta}} = \sigma^2 \mathbf{H.}(\phi)^T \mathbf{\Gamma}_{\hat{\phi}\hat{\phi}} \mathbf{H.}(\phi)\ , \tag{2.26}$$

where $\mathbf{H.}(\phi)$ is the Jacobian of $\mathbf{h}(\phi)$ with respect to ϕ evaluated at $\hat{\phi}$. Relying on an accurate first-order Talyor series approximation, the equation in (2.26) is also equivalent to the propagation of error equation [52].

Up to this point, the discussion has only centered on the properties of $\hat{\beta}$, and not on methods to determine the actual parameter estimates that minimize $SSE(\beta)$. Typically, current methods derive in some manner from the Gauss-Newton method of computing least-squares estimates [48]. Relying on a first-order Taylor series approximation to the model function, the Gauss-Newton method essentially searches the solution space using a succession of these first order approximations. However, convergence of the Gauss-Newton method is often poor, and more effective methods implement modifications to the base routine. Some of the most effective and widely used algorithms are based on the Levenberg-Marquart scheme [48]. For a detailed discussion of the methodology behind typical least-squares estimation schemes please see Chapter 14 of [48].

2.3.2 The Linear Case

When the model function $\mathbf{f}(\beta)$ is linear, (2.12) can be expressed as:

$$\mathbf{y} = \mathbf{X}\beta + \epsilon, \tag{2.27}$$

where the model function is replaced by an $m \times p$ matrix, and β represents the parameters to be estimated. The linear nature of the model function allows a direct calculation of $\hat{\beta}$, the parameter estimate minimizing $SSE(\beta)$. As before, ϵ is assumed to be distributed according to $\mathcal{N}(\mathbf{0}, \sigma^2\mathbf{\Gamma})$. Assuming that both $\mathbf{X}$ and $\mathbf{\Gamma}$ are fully ranked, then the least-squares solution, and thus the ML solution, of (2.27) is [53]:

$$\hat{\beta} = (\mathbf{X}^T \mathbf{\Gamma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{\Gamma}^{-1} \mathbf{y}. \tag{2.28}$$

The equations expressed in (2.28) are also known as the solution to the normal equations. In addition to being the ML estimate, $\hat{\beta}$ is also an estimator satisfying several other desirable properties. For one, it is an unbiased estimator; moreover, it possesses the minimum variance out of all other unbiased estimators of (2.27). Additionally, $\hat{\beta}$ is normally distributed according to [53]:

$$\beta \sim \mathcal{N}(\mathbf{0}, \sigma^2 (\mathbf{X}^T \mathbf{\Gamma}^{-1} \mathbf{X})^{-1}). \tag{2.29}$$

If the covariance matrix of the noise is of the convenient form $\sigma^2\mathbf{I}$, then the regression problem reduces to an OLS problem, which can be solved using QR factorization [54]. Compared to the method of normal equations, QR factorization reduces the potential for information loss and does not worsen the conditioning of the system beyond its inherent characteristics. The QR approach to the OLS problem also scales well to large and sparse systems, as fast and efficient sparse OLS solvers have been developed (for instance, when appropriate, the MATLAB operation $\mathbf{x} = \mathbf{A}\backslash\mathbf{b}$ uses sparse calculations based on [55]).

In general, the noise covariance is not in a diagonal and homogenous form. One option, as shown in (2.17), is converting the GLS problem into an OLS one. In cases where $\mathbf{B}$ is well-conditioned, or if $\mathbf{B}^{-T}$ can be readily calculated (for example if $\boldsymbol{\Gamma}$ is in block diagonal form), then conversion to OLS may be a viable choice. In such situations, solving the problem using the normal equations may also prove convenient.

However, when $\mathbf{B}$ is ill-conditioned, an alternative formulation proposed by Paige provides a better conditioned and more numerically stable solution [56]. Although Paige did not coin this term, this alternative method is now referred to as generalized QR factorization (GQR). The method involves a QR factorization followed by an RQ factorization, the details of which are described from a programming perspective in [57]. When preserving sparsity is not an issue, GQR is often the desirable choice over conversion to OLS [54]. Unfortunately, in its typical formulation GQR requires the storage and use of the orthogonal factor from the QR decomposition, which in general is very dense. As a result, current GQR methods are not usually appropriate for large and sparse systems.

Depending on the situation, the assumption of a fully ranked covariance matrix may not hold. In these cases, a singular covariance matrix applies certain implicit constraints on the system [53]. Magnus and Neudecker develop a variety of algebraic solutions based on different assumptions regarding the range space of $\mathbf{X}$ and $\boldsymbol{\Gamma}$ [53]. However, like (2.28), these algebraic forms are not particularly practical. Fortunately, the GQR method is general enough to accept rank deficient covariance matrices, provided QR factorizations with pivoting are used [56]. Alternatively one can extract the implicit constraints, as described in [53], and use them as explicit constraints. However, doing so requires solving an alternate linear system related to the original system.

Apart from implicit constraints arising from a singular covariance matrix, often the system under study also supplies a set of explicit constraints of the form $\mathbf{R}\boldsymbol{\beta} = \mathbf{b}$. If there are r constraints, then $\mathbf{R}$ is of dimension $r \times p$. Solving (2.27) is now a constrained problem. While algebraic formulations of the constrained solution can be derived, as in [53], the GQR approach can easily be extended to this case by augmenting the system in (2.27) directly with the linear constraints:

$$\begin{pmatrix} \mathbf{y} \\ \mathbf{b} \end{pmatrix} = \begin{pmatrix} \mathbf{X} \\ \mathbf{R} \end{pmatrix} \boldsymbol{\beta} + \begin{pmatrix} \boldsymbol{\varepsilon} \\ \mathbf{0} \end{pmatrix}, \tag{2.30}$$

where the rows corresponding to the linear constraints are free of any stochastic elements. Letting $\mathbf{u} = (\boldsymbol{\varepsilon}^T, \mathbf{0}^T)^T$, then the covariance matrix of $\mathbf{u}$, denoted as $\boldsymbol{\Lambda}$ is:

$$\boldsymbol{\Lambda} = \begin{pmatrix} \sigma^2\boldsymbol{\Gamma} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}, \tag{2.31}$$

which is a singular matrix. As a result, even if the original system possessed a non-singular covariance matrix, augmenting the system using (2.30) will always produce rank-deficient covariances. However, as discussed above, ML estimates of systems with singular matrices can be determined using GQR.

Up to now it has been assumed that $\mathbf{X}$ is fully ranked. Yet, in general this will not always be the case. In these situations, an ML estimate of $\boldsymbol{\beta}$ is only possible with the addition of explicit linear constraints as in (2.30) [53]. More specifically, an ML estimate requires that the rank of $(\mathbf{X}^T, \mathbf{R}^T)^T$ is p.

For all of these cases the statistical properties of the parameter estimates, such as covariance, can be computed. For an exhaustive review and theoretical development of the algebraic solutions of linear regression, including the statistical properties of the parameter estimates, the reader is encouraged to consult [53].

2.3.3 Weighted Normals

At this point, it is beneficial to revisit the weighted-normal estimation procedure of Sec. 2.2 as an ML problem. Focusing only on one pixel, under the assumptions of photometric stereo, noisy image formation may be expressed as:

$$I = \boldsymbol{\ell}^T \boldsymbol{\eta} + \epsilon. \tag{2.32}$$

Noise is assumed to be additive, which is a valid assumption as noise is typically uncorrelated with pixel intensity and location [29]. A good model of image noise is a combination of additive zero-mean Gaussian noise coupled with salt-and-pepper noise [29]. Salt-and-pepper noise can be effectively removed using a median filter or one of its variants [29], leaving only the Gaussian noise. As well, image noise is assumed to be linearly independent with the noise terms in all other pixel locations and images. For the rest of this thesis, image noise will be represented by independent zero-mean Gaussian-distributed stochastic terms. As well, since salt-and-pepper noise was not a significant source of noise for the system setup used, median filtering was not performed.

When three or more images are taken at the same viewpoint with differing principal light sources, then for each pixel (2.32) leads to a linear system of equations:

$$\mathbf{i} = \mathbf{L}\boldsymbol{\eta} + \boldsymbol{\epsilon}, \tag{2.33}$$

where $\mathbf{i}$ is a vector of all image pixels, $\mathbf{L} = (\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, \ldots, \boldsymbol{\ell}_m)^T$ acts as the regressor matrix $\mathbf{X}$ of (2.27), and the weighted normals, $\boldsymbol{\eta}$, acts as the parameter vector $\boldsymbol{\beta}$. Since image noise is assumed to be zero-mean IID Gaussian terms, the least-squares solution to (2.32) used in photometric stereo is also an ML estimate. This is a conclusion not always recognized in the literature. Additionally, the estimates are unbiased and exhibit the minimum variance of all possible unbiased estimators. Moreover, by using ML estimation on weighted normals, one can use (2.29) to model the behaviour of the estimates. In this case, the weighted-normal estimates, $\hat{\boldsymbol{\eta}}$, belong to the following distribution: $\mathcal{N}(\boldsymbol{\eta}, \sigma^2(\mathbf{L}^T\mathbf{L})^{-1})$. Note that using the shadowed equation of (2.8) affects the distribution, as appropriate rows of $\mathbf{L}$ must be omitted from the covariance term.

Framing photometric stereo as an ML estimate delivers key implications regarding visual-surface construction. Most notably, by providing a model of weighted normal stochastic behaviour, one can also determine an ML estimate of the object's visual surface. However, unlike weighted-normal estimation, the literature has yet to produce a practical ML estimation scheme for visual surfaces. Addressing this topic, Chapter 5 develops techniques to construct an ML estimate of visual surfaces.

2.4 Conclusion

By examining the limitations of the preliminary implementation of computer-aided classification, this chapter uncovered the need to control for illumination. Apart from simply removing a confounding factor in the classifier, controlling for illumination opens up the possibility to free computer-aided identification from the inherent limitations of image-based template representations. An implicit factor in the relationship between light direction and pixel intensity, the underlying 3D shape and texture of an object governs the image formation process. By presenting an expert with an image sequence of a microfossil illuminated from successively varying directions, video-based representations offer a powerful improvement over image-based representations.

In addition, this chapter motivated using shape-based representations of microfossils. By controlling for illumination, the underlying shape governing the image formation process can be explicitly extracted using computer vision techniques . Using Lambertian reflectance assumptions, the photometric stereo method can extract estimates of the microfossil surface normals and albedo. These parameter estimates can then be used to extract a visual surface. However, unlike Lambertian photometric stereo, the visual surface estimation is an ongoing focus of research. For instance, current visual surface estimation methods struggle under the presence of image noise.

Addressing image noise within the context of visual surface estimation requires employing ML concepts and techniques. As a result, this chapter introduced linear and nonlinear ML theory, both of which will prove crucial in constructing ML estimates of visual surfaces. As asymptotic limits play a key role nonlinear ML theory, they constitute an important part of this chapter's discussion on ML estimation. The chapter concluded by framing photometric stereo in the context of ML estimation, presenting a model of the behaviour of weighted-normal estimates.

Using video and shaped-based representations instead of images could provide the key to improving classification performance beyond its current restrictions. Incorporating computer vision techniques into the computer-aided classification system brings a host of challenges, implications, and also new exciting directions of inquiry. These issues constitute the focus of this dissertation. Representing a crucial component of these challenges, integrating video and shaped-based representations into computer-aided identification requires key extensions to the existing system. Implementing and investigating these new representations in the context of computer-aided identification is the focus of the next chapter.

Chapter 3

Video and Shape-Based Representations

As noted, previous work on computerized microfossil identification relies on image-based representations to discriminate between specimen classes. This characterization is equally true for the preliminary computer-aided system described in [24] and summarized in Sec. 1.4. Sec. 2.1 argued that freeing computer-aided classification from its current limitations requires moving beyond image-based representations for template identification. Controlling for illumination to construct video and shape-based representations offers a powerful alternative to images.

Since the preliminary system can only capture and handle images, supporting video and shaped-based representations requires important extensions to the system. Even so, the system developed for this thesis shares much in common with the preliminary system. As shown in Fig. 3.1, the system can be divided into four main modules.

The *Video Capture Module*, described in Sec. 3.1, is tasked with capturing sequences of images of each specimen at appropriate illumination directions. After collecting videos of each specimen, the system must cluster them appropriately. The *Clustering Module* is responsible for clustering sample sets of specimens and choosing an appropriate template for each cluster. Sec. 3.2 outlines this module, which uses a novel method for image alignment described in Chapter 4. The *Shape Reconstruction Module* computes surface normals of the microfossils from the sequences of images and extracts visual surfaces using a novel ML formulation. The surface estimation method is explained in Chapter 5. Anaglyph video sequences are used to disseminate the shape-based representations online. The final stage of the system, the *Expert Input Module*, requires an expert to identify the templates of each cluster either by physical inspection or, preferably, through the specimen's digital representation. This module is explained in Sec. 3.4.

With the system extensions in place, an analysis of the efficacy of video

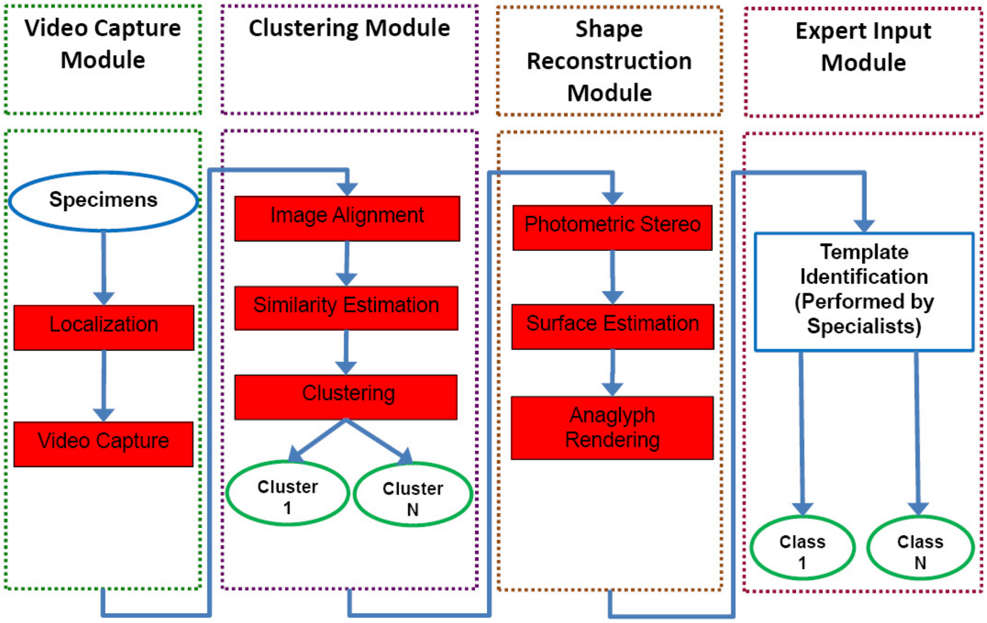


Figure 3.1: System outline of video and shape-based computer-aided identification of microfossils. The system shares much in common with the one outlined in Fig. 1.3; however, in this case it has been reorganized into four main modules. As well, because the system captures image sequences instead of just single images, the two set-ups differ in key areas. Also note that, although not shown in the figure, the clustering step requires a similarity threshold.

and shape-based representations can be made even without expert classifications. This leads to a set of three tasks:

- Collect a sufficiently large dataset of particles and their associated video sequences;
- Apply computer vision techniques to properly extract shape-based representations of microfossils;
- Compare the capabilities of video and shape-based representations of templates.

Sec. 3.5 concludes the chapter with such an analysis, and introduces a digital representation combining the relative strengths of video and shape-based representations.

3.1 Video Capture Module

Developing a computer-aided microfossil identification system using video and shape-based representations requires extensive datasets consisting of image sequences of each individual specimen. This necessitates a system able to automatically localize and capture large batches of specimens. Implementing such a scheme involves several key steps and pieces of equipment.

3.1.1 Equipment

Table 3.1 lists the equipment used for capturing video sequences of microfossil specimens. A picture of the setup is depicted in Fig. 3.2. The following is a description of the criteria used in selecting the equipment.

The linchpin of the system is a suitable microscope with a digital camera attachment. Due to its simplicity and inexpensiveness, optical microscopy is the modality used for this thesis; however, other groups have published results based on SEM [13, 14, 18]. An optical microscope also requires an accompanying light source. Preferable light source features include minimal production of heat (such as the fibre-optic source used in this system) and aperture and colour temperature control. Many optical microscopes can be purchased with an optional digital camera mount. Selection of an appropriate digital camera should at the very least be based on its pixel resolution, data transfer rate, and availability of an application programming interface (API).

Since the light source remains static with respect to the microscope, capturing the specimen under different illumination directions requires actually rotating the particle. Accomplishing this in an automated fashion requires a motorized x-y-phi stage. Appropriate stage dimensions depend on the specific microscope base and stand setup used in the system. The motorized stage must possess enough accuracy and repeatability for specimens on the order of $100\,\mu$m. Like the camera, the stage controller should provide an external API. Another important decision to make is whether to use a stepper or servo motor to drive the stage. A stepper motor was chosen as they and their controllers are typically less expensive than their servo counterparts. As well, high quality stepper motors can provide very good precision without requiring feedback, although occasional calibrations are still required. An aluminum base was designed to securely mount the stage onto the microscope stand. This was manufactured in the UA Electrical and Computer Engineering Machine Shop.

A custom multi-threaded and user-friendly software program designed using C++ in the Visual Studio .NET environment controls the image cap-

Microscope	
Model	Zeiss Stemi 2000-C
Zoom	max ×5
Cost	$5,704
Camera	
Model	PixelLink PL-A774 CMOS
Resolution	1600 × 1200
Frame Rate (fps)	20
Cost	$2,985
Lighting	
Equipment	Schott KL 1500 LCD
Colour Temperature (K)	3000
Wattage (W)	150
Cost	$1,995
Motorized Stage	
Company	Micos USA
Configuration	x-y-phi
Horizontal Range (mm)	150
Vertical Range (mm)	150
Linear Bi-directional Repeatability (um)	±15
Linear Resolution (um)	0.1
Rotational Repeatability (°)	±0.008
Controller	Internal PCI Card with Motor Drivers
Cost	$6,855

Table 3.1: Specifications of the equipment used in the video capture module. The total cost of the equipment was $17,539.

ture system. Much of the required image processing was implemented through the use of the OpenCV software library[1]. The software system also handles uploading the particle images and accompanying information to an online wiki and database.

3.1.2 Batch Processing

As microfossils typically fall within a certain size range, specimens are first sieved to weed out broken or fused particles prior to being photographed. The remaining specimen particles are then sprinkled onto a slide. To ease

[1] http://opencv.willowgarage.com/wiki/Welcome

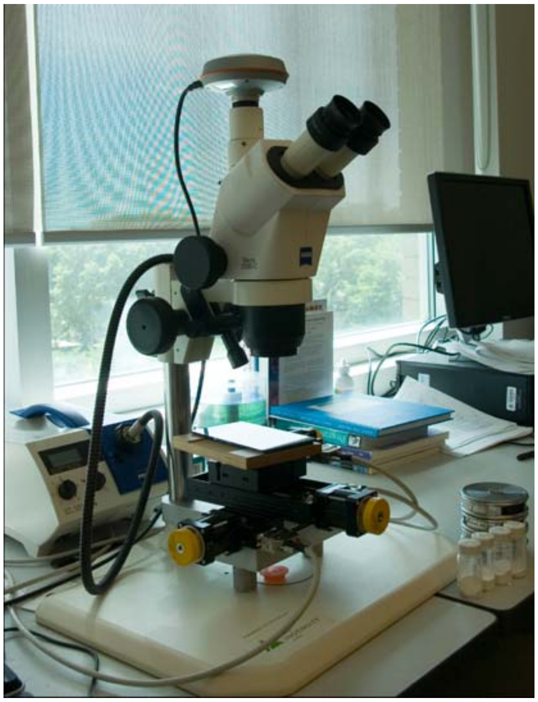


Figure 3.2: Computer-aided identification system setup. This picture illustrates the setup used in the computer-aided classification system of this thesis. Included in this setup is a microscope, motorized stage, custom stand, light source, digital camera, and sieving tools.

the later image segmentation steps, an opaque glass slide was used. After being sprinkled on the slide, the specimens are localized by scanning the entire area of the slide using the motorized stage. To accomplish this, the stage is translated in increments of half the current field of view, ensuring that specimens are not missed while also minimizing overlap. At each scanning increment, an image is captured (without gamma correction) and constrained to have pixel values between 0 and 1. As the system has no *a priori* knowledge of the numbers of specimens on the slide, it must determine on its own whether it has found a specimen or not. But, since particles are sieved beforehand, the particle size range is known. Thus, the presence of a particle can be determined by first segmenting the current field of view, using simple thresholding, and filtering out any objects falling outside the allowed range. If an object in the field of view is found, the location of the particle is calculated and recorded using the current position of the stage combined with the distance of the object's centroid to the center of the image.

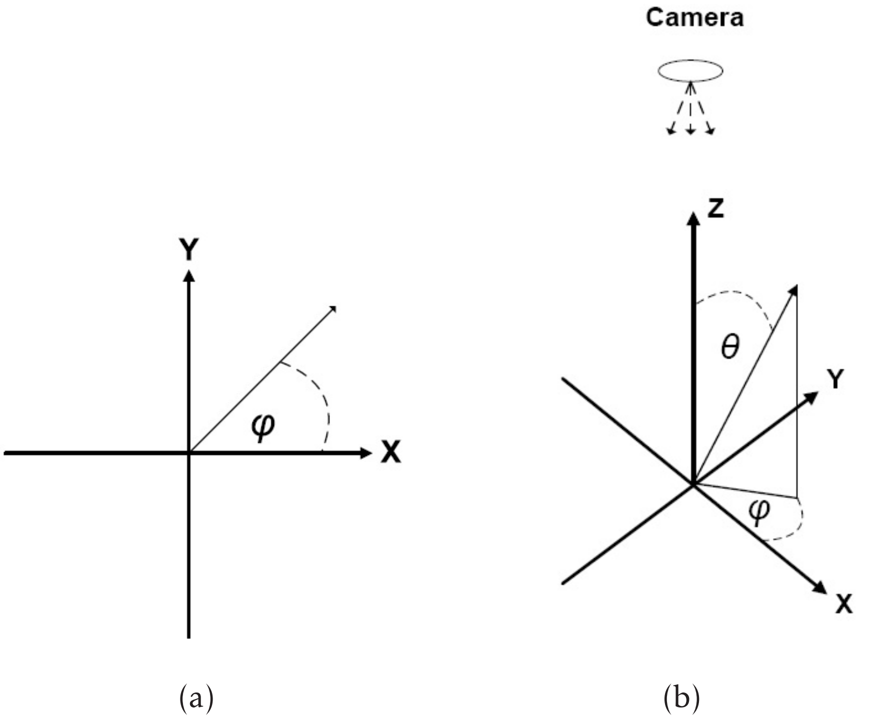


Figure 3.3: Coordinate system used for microfossil capture. The graph in (a) illustrates the coordinate system and the azimuth angle (φ) as seen from the camera's viewpoint, while (b) demonstrates the orientation of the z axis along with the elevation angle (θ).

3.1.3 Video Capture

Recall that the capture of image sequences serves two purposes. The first purpose is to provide video-based representations of microfossils. The second is to provide an image sequence appropriate for application of photometric stereo normal estimation and visual-surface reconstruction techniques. Photometric stereo requires images of an object at identical viewpoints but illuminated at differing and known light directions.

Fig. 3.3 demonstrates the coordinate system used for microfossil capture. Parallel to the image plane, the coordinate system's x and y axes rest on the supporting surface of the object being viewed. The z axis is aligned with the optical axis, but points towards the camera. The figure also illustrates how the elevation and azimuth angles are defined, θ and φ respectively, which are both two crucial angles in the system.

To obtain sequences of images, once all specimens have been localized, the system captures images of each particle at illuminated from different directions with respect to its principal axis. The principal axis is determined using the same method as the invariant transform (please see Sec. 1.4.1.3 and [24] for further details). As the illumination direction is fixed with respect to the microscope, illumination direction is manipulated by physically rotating the particles using the motorized stage. Performed this way, only the azimuth angle of illumination changes, while the elevation angle remains fixed. Increments of 20° are used, producing 18 images per microfossil. Light direction is easily calculated, as the elevation of the current setup is fixed to an angle of 30°, and the azimuth angle of illumination is known through the calculation of the principal axis orientation, the fixed 20° increments, and the system's reference azimuthal angle of 90°. Lighting configurations that incorporate equally spaced azimuth angles and a constant angle of elevation are optimal configurations in reducing the effect of noise for photometric stereo [58].

As is typical with motorized x-y-phi stages, the rotational element of the stage is mounted on top of the translation elements. Consequently, when rotating a particle, its real-world (x, y) location will also change. As a result, keeping an object within the center of the field of view requires both rotating the slide by the desired amount and also translating the stage to take into account the object's new real-world (x, y) coordinates. Ideally, the object's expected new (x, y) coordinates match-up with its actual ones, allowing the object to always remain in the middle of the field of view. However, unavoidable inaccuracies in the motorized stage and calibration of the system cause discrepancies between a particle's actual and expected coordinates. Fig. 3.4 demonstrates an example of this.

Analogous to travelling around the arc of a circle, the degree of change in (x, y) coordinates is proportional to the Euclidean distance of the object from the axis of rotation. Thus, particles located at greater radial distances from the axis of rotation will be more prone to the localization errors illustrated in Fig. 3.4.

To gain a perspective on the degree of localization error the system experiences, 32 microfossils, specifically forams, were localized on the slide. Images of each microfossil were captured after rotating the particles by increments of 20°. For each image, the distance from the particle's centroid to the center of the field of view was calculated. In addition, the radial distance of the particle to the axis of rotation of the stage was recorded. Thus, each foram has 18 different localization errors, one for each rotation in the sequence. Fig. 3.5 graphs the median localization error of each foram versus its radial distance from the stage's axis of rotation. As the figure demonstrates, particles further away from the origin of the stage's coordinate system are more susceptible to localization errors. Considering that images are 640×640 pixels, the scale of the localization errors demonstrated by Fig. 3.5 are quite significant.

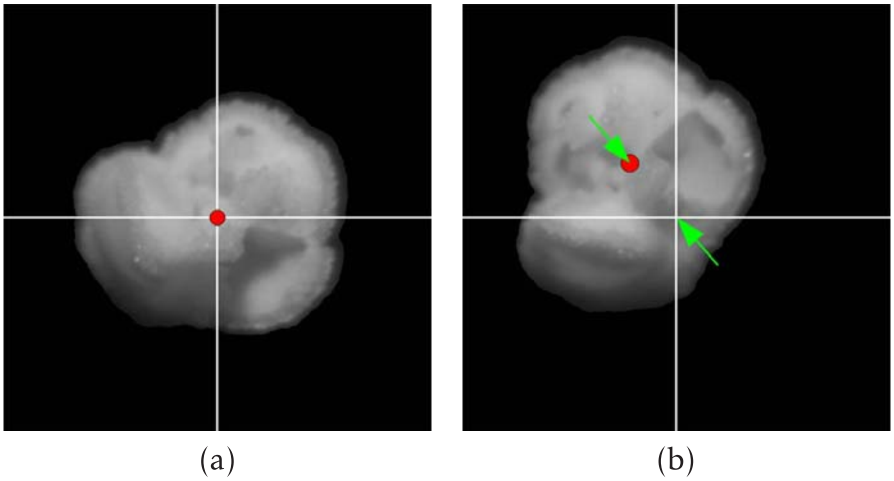


Figure 3.4: Stage localization errors. Errors in particle localization using the motorized stage arise due to inaccuracies in pixel-dimension approximations, stage calibration, and the electro-mechanical system of the apparatus. For instance, in (a) a particle has been localized and centered. In (b) the same particle has been rotated 60°, and while the object has also been translated to its expected (x, y) coordinates, it is no longer centered.

Without these localization errors, each image of the specimen after rotation could have been captured by simply extracting the same region of interest from the middle of the field of view. However, these localization errors displace the specimen from the middle of the image after each rotation. Since the amount of localization error is impossible to predict, image alignment requires determining the center of the object in some manner. As in the localization step, the center of a specimen is treated as the centroid of its silhouette. Thus, after each rotation, an image is captured by extracting a region centered at the specimen's centroid.

Unfortunately, using the specimen's centroid to align images is not without its shortfalls. While the specimen's silhouette is fairly robust against illumination changes, differing relative light directions at each rotation unavoidably cause enough differences in the silhouettes from one image to another to affect centroid position. Thus, each image in the sequence will be centered somewhat differently, resulting in alignment errors. Although smaller than localization errors, they are nonetheless significant. More details on the scale of this problem, and a technique to address it are explored in Chapter 4. However, the rest of the discussion will assume images have been aligned properly.

As it is the speciments that are actually rotated rather than the light source, the principal axes of captured images are at successively increas-

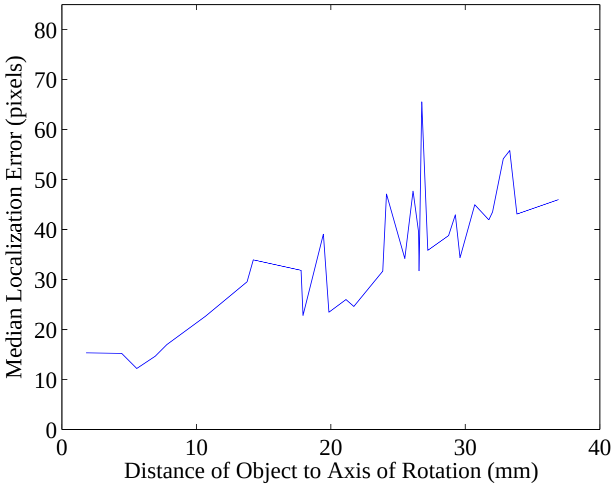


Figure 3.5: Plot of localization error vs. distance of particle to the origin. The median localization errors of 32 forams were computed. Each pixel is roughly equivalent to one μm (pixel size was 1.066 μm). The general trend of increasing localization error with increasing distance to the origin can be seen in the plot.

ing angles with respect to the horizontal. As a result, each image must be rotated back, as in Fig. 3.6, so that each axis is at an angle of 0°. In addition, images are padded so that they are all 640 × 640 pixels. These images are then composed into a video-based representation. Viewing a video gives the impression that the light source is rotating around the specimen. Once every image is captured, the videos are uploaded to the online wiki (Sec. 3.4.1), and the actual particles are archived onto indexed slides. The image sequences are also provided as input to the clustering and shape extraction modules.

3.1.4 Dataset Collection

Using the video-capture module extension to the system, 500 videos of foram specimens were captured. Their associated physical particles were archived as well. As motivated in Sec. 1.4.1.1, forams play a crucial role in academic research and industrial applications, serving as excellent microfossils to study. All specimens were collected from the B-core of the ODP Pacific site 865 at depths ranging from approximately 100 m to 140 m. The ODP Pacific site 865 is a location consisting of species important to paleoclimatological research [25]. The sieving stage prior to video capture constrained specimens to only those with diameters between 250 μm and

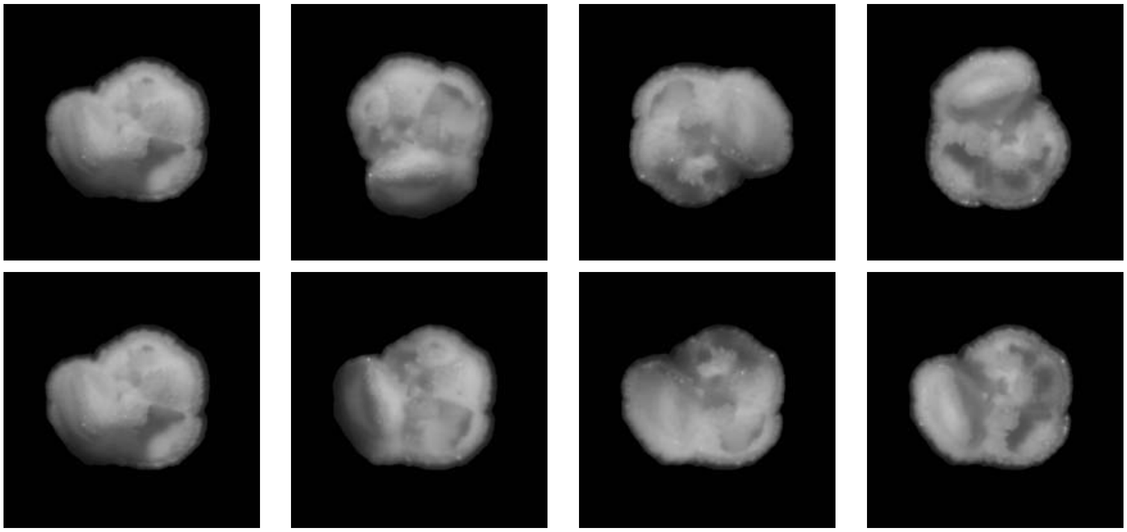


Figure 3.6: Creating a video-based representation. In the first row are images of the same foram rotated so that its principal axis is at angles of 0°, 80°, 180°, and 260°, respectively, from the horizontal. In the second row are the same images, except they have been rotated so that the principal axis is always at an angle of 0°, giving the impression that the foram is stationary while the light source revolves to azimuth angles of 90°, 10°, −90°, and −170° respectively. These frames are combined together to create a video.

300 μm. As mentioned above, videos consist of 18 images illuminated from a constant elevation angle of 30° and from azimuth angles at 20° increments.

3.2 Clustering Module

Once image sequences of each specimen have been captured, the system is ready to automatically cluster the specimens and choose a template for each cluster. Generally speaking, for each microfossil, the clustering module accepts image sequences at a fixed viewpoint and varying light directions as input. The module outputs a set of clusters, each possessing a corresponding template specimen.

Dealing solely with the analysis of data, the clustering module has no equipment requirements other than using appropriate computational software. Ultimately, to progress the system beyond the prototype stage, the module should be implemented using a fast software package, such as C++, that supports a graphical user interface. However, in the clustering module's current prototypical stage, MATLAB, with occasional use of MEX scripts, acts as the computational environment.

The clustering module is comprised of five steps:

1. Align image sequences. Correcting for misalignment required the development of a novel alignment technique explained in Chapter 4;

2. Calculate a similarity matrix between all pairs of specimens. Similarity is discussed in Sec. 3.2.1;

3. Cluster specimens together based on visual similarity. A description of the algorithm used and a reasoning behind the decision is explained in Sec. 3.2.2;

4. Select an appropriate and representative template for each cluster. This step is dealt with in Sec. 3.2.3.

3.2.1 Similarity Measure

As explained in Sec. 2.1, the preliminary system's failure to control for relative differences in illumination direction with respect to the principal axis introduced intra-genus variability. In addition, as shown in Fig. 2.2(b), absolute azimuth angles of illumination affect similarity scores between specimens of different genera. However, the provision of video-based representations consisting of images under 18 different lighting conditions can mitigate these confounding illumination effects.

First, since lighting direction is controlled, similarity may be computed between image pairs with identical illumination conditions. Secondly, each pair of specimens possesses 18 similarity scores—one for every possible absolute azimuthal angle. As a result, choosing the median score of the 18 possibilities essentially avoids any confounding effects caused by absolute angles of illumination. This is the approach taken for this dissertation.

Yet, the utility of such a scheme still depends largely on the strength of the similarity measure used between images. As argued in Sec. 1.4.1.4 and [24], the correlation measure is a powerful and well-understood metric particularly appropriate for prototypical implementations of classification schemes. As a result, for two specimens, a and b, their similarity at azimuthal angle φ can be written as:

$$sim(a,b,\varphi) = r(A_\varphi, B_\varphi), \tag{3.1}$$

where A_φ and B_φ are the images at the azimuth angle φ of a and b, respectively, and r is calculated using the correlation score of (1.1).

However, this similarity scheme suffers from a key problem related to the normalization process of Sec. 1.4.1.3. As mentioned, simply aligning the principal component of an object's silhouette with the horizontal axis leaves a 180° ambiguity. Using the invariant transform developed in [24], which is described in Sec. 1.4.1.3, the video capture module resolves this 180° ambiguity using third-central moments. While the method introduced in [24] was more stable than related methods in the literature, analysis of its performance using the 500 specimens of the foram dataset indicates that it was still susceptible to instability, especially for objects with elliptical shapes. As a result, the invariant transform cannot always be relied upon to resolve the ambiguity in a consistent manner. Rather than attempt to resolve this ambiguity, a better approach is to consider both options of the 180° ambiguity and choose the maximum score of the two. However, as will be shown, this can only be accomplished using images from more than one illumination direction—data available with the extended system.

Using this scheme, similarity between specimen images is formally expressed as:

$$sim(a,b,\varphi) = \max(r(A_{\varphi},B_{\varphi}), r(A_{\varphi},B'_{\varphi})), \tag{3.2}$$

where B'_{φ} represents the image obtained should the 180° ambiguity have been resolved in the opposite manner. Since pixel values of images are dependant on illumination direction, one cannot simply rotate B_{φ} by 180° to produce B'_{φ} as the illumination direction will also rotate by 180°. Fig. 3.7(b) depicts this problem visually. Fortunately, as part of the video capture process, the system captures a sequence of images for multiple illumination directions at 20° increments. Thus, if attempting to compute similarity by using image B_{φ}, the system also has access to $B_{\varphi+180^\circ}$. As Fig. 3.7(c) and (d) demonstrate, rotating $B_{\varphi+180^\circ}$ by 180° will produce the appropriate representation of B'_{φ}.

With the ambiguity resolved, the median score over all 18 possibilities forms the core calculation in constructing a similarity matrix between all pairs of specimens. It can be written formally as:

$$sim(a,b) = \operatorname*{med}_{\varphi}\{sim(a,b,\varphi)\}, \tag{3.3}$$

where med denotes the median operator.

3.2.2 Clustering Algorithm

Clustering is a very heavily researched area of study that includes a huge variety of different techniques and approaches. This section uses the terminology and clustering method taxonomy described in [30]. For a very good review of clustering the reader is encouraged to consult that manuscript. Many clustering techniques require the computation of a similarity matrix between all patterns [30], and this work is no exception.

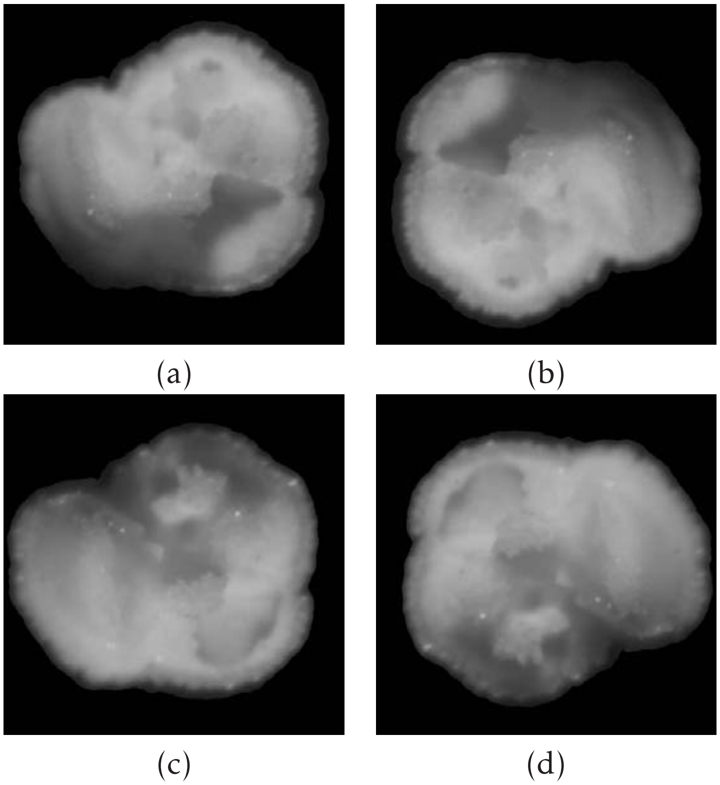


Figure 3.7: Obtaining the image corresponding to the alternate resolution of the 180° ambiguity. For a particular specimen *b*, it is desirable to obtain two images *B* and *B*′ which represent the two possible resolutions of the 180° ambiguity. Representing *B*, the image of a foram specimen in (a) is illuminated from an azimuth angle of 90°. The image in (b) is the same as in (a) but rotated is by 180°. As the image in (b) is illuminated from an azimuth angle of 270°, it is an incorrect representation of *B*′. On the other hand, (c) is an image of the same specimen in (a), but is illuminated from the opposite direction. Rotating (c) by 180° produces the image in (d), which is the correct image to use for *B*′ because it is illuminated from an azimuth angle of 90°.

The previous implementation of computer-aided classification used a clustering algorithm based on maximal cliques of a non-weighted graph [24]. Vertices represented individual patterns, and edges only connected two patterns if their similarity score was above a certain threshold. While the method is valid, the magnitude of similarity between two specimens is not taken into account when computing the maximal cliques. For instance, consider a simple example with only three specimens *a*, *b*, and *c*. Assume *a* and *b* possess a similarity score very close to 1. On the other hand, assume *a* and *c* have a similarity score only slightly above the chosen threshold value.

With all else being equal, the maximal-clique finding algorithm will group a and c together with equal likelihood as it would a and b, when in fact the clustering algorithm should favour the latter.

Agglomerative hierarchical clustering (AgH) is a simple but effective alternative to the maximal clique algorithm that directly incorporates similarity scores. In addition, since it does not require an approximate solution to an NP hard problem (unlike maximal-clique finding), its computation is significantly faster. By using the threshold parameter as the stopping criteria, this work uses a slightly modified version of AgH than the one outlined in [30]. The steps are as follows:

1. Compute a similarity matrix between all pairs of specimens. Set all specimens to be singleton clusters.

2. Find the most similar pair of clusters and merge them into one single cluster. Update the similarity matrix to reflect this merger.

3. If there are no similarity scores greater than the threshold parameter or if all specimens are in one cluster, stop. Otherwise, go to step 2.

The algorithm is hierarchial as for each threshold parameter there exists a corresponding grouping of specimens. The agglomerative characterization describes the grouping progression, which starts from a set of singleton clusters and through mergers constructs larger clusters. The hierarchical nature of the algorithm means that groupings based on a lower threshold directly follow from the higher threshold groupings. As a result, one can obtain a collection of groupings by performing the above steps using the lowest desired threshold as the stopping criterion. If mergers are recorded properly, groupings based on the higher thresholds can be easily retrieved. This feature of AgH offers a distinct advantage over maximal-clique clustering, as the latter requires the use of a computationally expensive maximal-clique finding algorithm at each threshold.

For the updating portion of Step 2, either single-link or complete-link options are typically used [30]. For either case, when two clusters merge the parent cluster inherits the similarity scores obtained from the two child clusters in some manner. In single-link algorithms, for each pair of similarity scores, the parent cluster chooses the maximum score from the two children. In contrast, complete-link clusters opt for the minimum similarity scores. Complete-link clustering tends to produce more compact clusters. Since homogeneity of clusters is extraordinarily important for microfossil grouping, this factor favours complete-link over single-link. As well, the literature indicates that complete-link clustering produces better results for

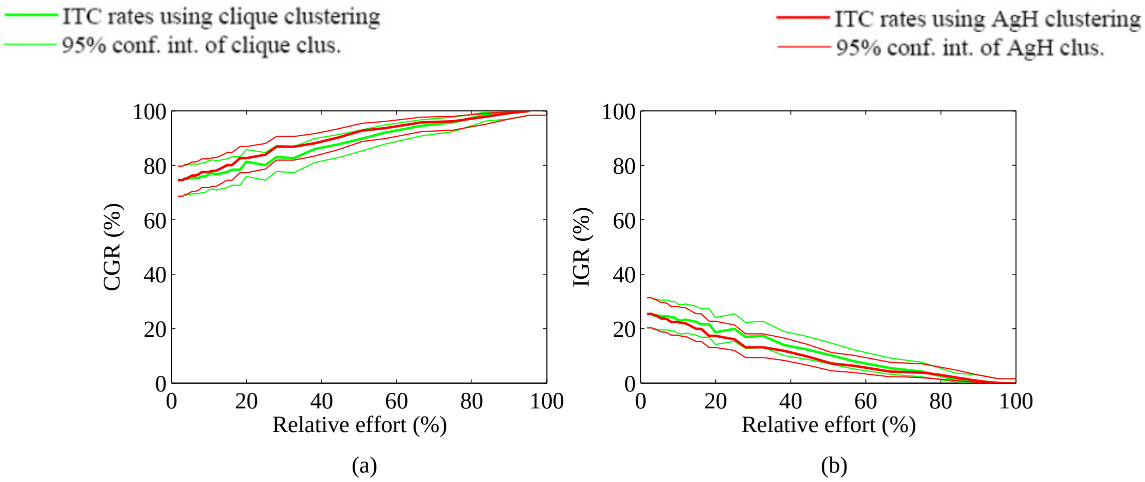


Figure 3.8: Comparison of performance of maximal-clique clustering vs. agglomerative hierarchical clustering. Correct and incorrect genus rates of the particle-based ITC are shown using the dataset from [24]. With regards to performance, agglomerative hierarchical clustering edges out maximal-clique clustering, especially at lower relative efforts.

many practical applications [30] and this is also supported by experiences in the context of this dissertation.

In the context of microfossil identification, the performance of a clustering algorithm must be judged without other sources of error confounding results. As explained in Sec. 2.1, clustering performance can be judged on its own by examining the CGR and IGR of the particle-based ITC. As a result, to perform a comparison between the previous and new clustering algorithms, the classification rates of the particle-based ITC based on AgH clustering were computed using the same dataset and methodology of [24]. These classification rates were then compared with the rates of the particle-based ITC using maximal-clique clustering. Fig. 3.8 graphs the results. As can be seen, AgH clustering performs better than maximal-clique clustering at all relative efforts. However, at no point is the difference in performance significant. However, even if performance between the two algorithms are equivalent, the significantly faster speed of AgH computation over maximal-clique clustering is reason enough to opt for the former algorithm over the latter.

3.2.3 Template Selection

As Sec. 1.4.2.2 demonstrated, template selection was not a major source of error in the previous version of computer-aided classification. For this reason, this system selects templates in the same manner as what was described in Sec. 1.4.1.6 and [24].

3.3 Shape Extraction Module

With templates chosen for each cluster, the system is ready to extract template shapes. First, using the photometric stereo method explained in Sec. 2.2, surface normal and albedo values are estimated for each template. Afterwards, using the ML surface estimation method of Chapter 5, the system extracts shape-based representations from the normals.

3.4 Expert Input Module

Once the sample set is clustered, the system requires human input to identify each cluster template. The most convenient manner to present digital representations are through online tools. In a previous study [31], an online wiki was developed to enable an expert to identify microfossils through their image-based representations. Yet, while shape-based representations are the digital representation of choice, the capabilities of the current wiki do not support depicting 3D models. As a result, at this point identification is performed by either viewing each template's video-based representation, or by actually inspecting each template particle under a microscope. Future work will enhance the capacity of the online wiki to display shape-based representations.

3.4.1 Video-Based Identification

Providing a natural language interface for the user, the wiki allows experts to perform classifications remotely. As of now, forams are the only type of microfossils available in the wiki; however, in principle the wiki can support any type of microfossil. For batches of samples open to the public, any interested person is able to view video-based representations of forams, their video, and particle-based classifications, and even make classifications of their own (provided they have registered). This wiki was also a crucial component of the preliminary computer-aided classifier outlined in Sec. 1.4 [24].

Programmed in *PHP*, the wiki uses a *MySQL* back-end to store data and control access. For the purposes of this thesis, the capabilities of the wiki were extended to allow it also to display videos of microfossils. Users can play, stop, and watch videos frame-by-frame. In addition, any data used for the wiki is freely available for download in convenient formats such as .csv

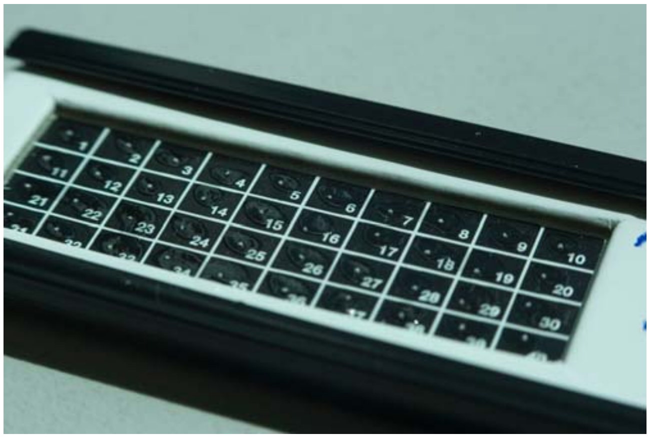


Figure 3.9: Archived physical specimens. Physical particles are glued onto sample slides. Slides are indexed and archived for later use.

or .mat files. In the case of the latter format, a PHP library to create .mat files was developed during the course of this thesis.

3.4.2 Particle-Based Identification

While video-based identification of templates is the more desirable option, accuracy is best when experts identify template specimens through physical examination under a microscope. To enable particle-based identification, each specimen was indexed and archived onto slides after image capture so that a cross-reference exists between the templates chosen by the clustering module and their corresponding physical particles. A typical slide is depicted in Fig. 3.9.

3.5 Analysis of Representations

Arguments in Chapter 2 supporting the use of computer vision relied in a large part on the line of reasoning that shape-based representations can successfully incorporate the rich collection of information encapsulated and revealed by sets of images under varying illumination directions. In addition, by providing experts with the option to examine the object at viewpoints other than straight above, shape-based representations can present information not available with simple videos.

To test the ability of shape-based representations to capture image information, it is instructive to test whether visual surfaces can accurately reconstruct images. Using the photometric stereo procedure explained in Sec. 2.2 and the ML surface estimation method in Chapter 5, experiments

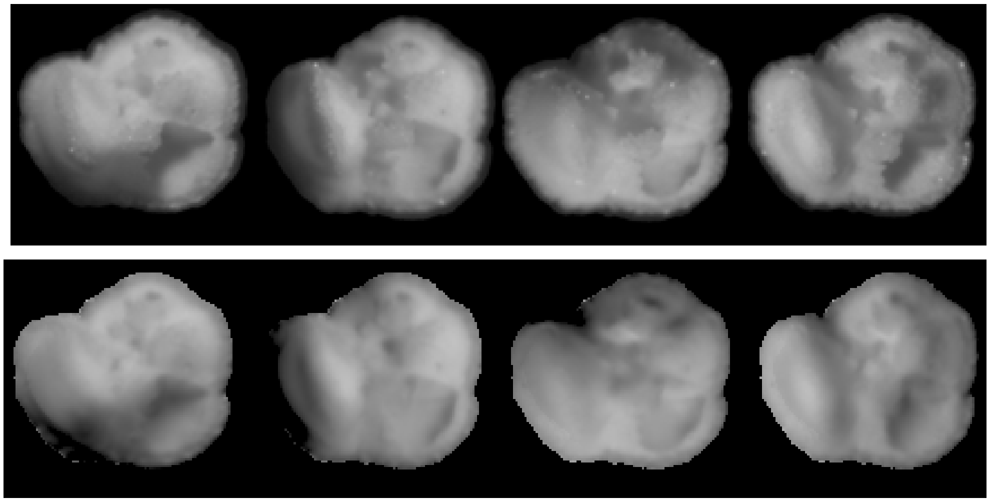


Figure 3.10: Comparison of video and shape-based representations. In the first row are frames taken from the video-based representation. The second row consists of frames reconstructed using the visual surface and the Lambertian albedo.

estimated surfaces using the 18 images of each of the 500 specimens in the foram dataset. Weighted normals and albedo were also calculated as part of this process. Using the image formation equation of (2.5), each of the 18×500 images were reconstructed from the visual surface. As part of this process, surface gradients were approximated using centered finite-differencing.

Unfortunately, as Fig. 3.10 illustrates, reconstructed images suffer from significant detail loss compared to their video-based counterparts. In addition, the reconstructions lack cast shadows. However, cast shadows are important, as they emphasize strong features. By depicting both cast shadows and fine detail and texture, the video-based frames in Fig. 3.10 are much better at emphasizing key geometrical features, such as the spire-like structures rising out of the microfossil. As well, texture loss alone would impact classification accuracy, as in many cases microfossils are distinguished by observing fine details such as texture [35]

This detail loss can be quantified by measuring reconstruction errors of the 500 visual surfaces. However, with regards to the ability of shape-based representations to encapsulate information across all available lighting directions, it is instructive to also measure reconstruction errors of the weighted normals. Visual surfaces combined with albedo can be thought of as a way to reduce the dimensionality of the data from 18 images to 2 basis images. However, weighted normals calculated as an intermediate step also provide a basis—in this case 3 basis images, one for each component of the weighted normals. In fact, assuming that all photometric assumptions hold, weighted normals are a perfect basis for all possible lighting configurations [40, 59]. However, the photometric assumptions can never be satisfied in full, particularly with regards to attached and cast shadows [60]. As a result, in actuality more than 3 basis images are needed to fully account for image variability due to illumination. Nonetheless, empirical and theoretical research has demonstrated that 3 basis images can account for more than 90% of image variability [60]. These results though, are based on eigenvector analysis of a set of images, and do not necessarily apply to weighted-normal basis images.

Taking these considerations into account, experiments should calculate reconstruction errors for both visual surfaces and weighted normals. Testing the reconstruction error of weighted normals serves to measure how well the Lambertian assumptions hold in the context of foram specimens. Doing so requires reconstructing images using the original image formation equation of (2.1) rather than (2.5). On the other hand, differences between reconstruction errors of the visual surfaces and the weighted normals sheds light on how much information is lost when estimating surfaces from normals and albedo. These are both important metrics to consider.

The coefficient of determination or R^2 value provides a measure of the improvement of one hypothesis over another, and can be expressed as:

$$R^2 = \frac{SSE(H_0) - SSE(H_1)}{SSE(H_0)}, \tag{3.4}$$

where H_0 and H_1 represent the null and test hypotheses respectively. In the context of quantifying explained image variability, the mean image across all 18 images serves as H_0, while images generated from the visual surface or weighted-normal estimates correspond to the test hypothesis. R^2 performed this way provides a value for explained variance [61]. The weighted-normal R^2 value for a single image sequence is expressed formally as:

$$R^2 = \frac{SSE(\bar{I}) - SSE(\hat{\eta})}{SSE(\bar{I})}, \tag{3.5}$$

where $\bar{I}$ denotes the mean image across the entire sequence and the sum-squared error incorporates every pixel in every image. Similarly, the R^2 values for a visual surface is calculated using:

$$R^2 = \frac{SSE(\bar{I}) - SSE(\hat{Z})}{SSE(\bar{I})}. \tag{3.6}$$

Fig. 3.11 provides a histogram of the 500 R^2 values for both visual-surface and weighted-normal representations, while Table 4.1 summarizes

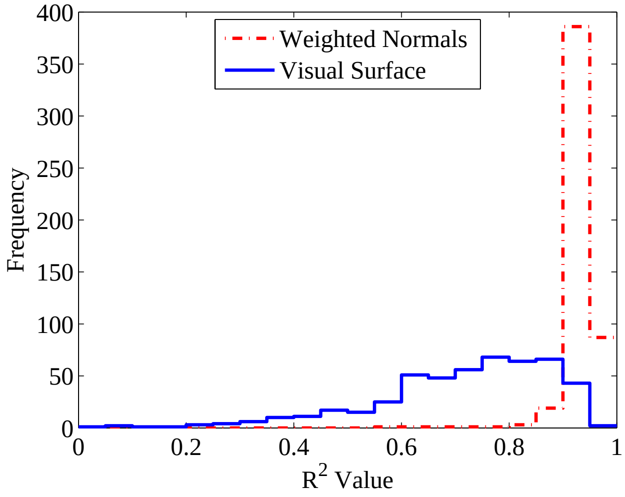


Figure 3.11: Histogram of R^2 values for visual surfaces and weighted normals. As this graph demonstrates, most of the weighted-normal estimates account for close to 90% of image variability. Additionally, most weighted-normal R^2 values are clustered together. On the other hand, visual-surface R^2 values are spread more uniformly between 60% and 90%, meaning they are not as reliable in accounting for image variability.

the values for the 500 foram dataset. The minimum R^2 values depicted by Table 4.1 for the weighted normals and visual surfaces are worrisome; however, it should be noted that outlying cases of poor explained variance corresponded to the presence of confounding factors. These include dust, other secondary small objects in a subset of the images, or microfossils that are lying on an unstable side, which often creates an overhang. In the latter case, using the motorized stage to shake the slide prior to image capture should minimize the occurrence of such situations.

As is evident, weighted normals successfully accounted on median for roughly 94% of image variability. As well, the proximity of the first and third quartile values to the median, and the limited R^2 variability in Fig. 3.11, indicate that 94% is a representative value.

Note that, because a constant angle of elevation is used across the image sequence, it must be stressed that these R^2 values give no indication whether the computed weighted normals serve as a good basis for *all* elevation angles of illumination. Even so, the weighted normals serve as an effective basis for the 18 images captured for each specimen in the foram dataset.

Nonetheless, when reconstructing images using weighted normals, on median 6% of image variability is lost. Responsibility for this shortfall

	Weighted Normals	**Visual Surface**
Maximum	97.05%	96.32%
Third Quartile	94.74%	83.62%
Median	93.79%	74.44%
First Quartile	92.50%	61.99%
Minimum	57.91%	0.029%

Table 3.2: Specimen image variability accounted for by the weighted normals and visual surface. This table presents statistics describing the percentage of image variability explained by the weighted-normal basis and the visual surface basis using shape extracted from the foram video dataset.

lies in non-Lambertian aspects of microfossil surfaces. These results indicate that to better encapsulate information across all lighting directions, weighted-normal estimation may need to incorporate more sophisticated reflectance models.

On the other hand, visual surfaces were less able to account for image variability than weighted normals. With a median R^2 value of 74% and a comparatively high degree of variability in its histogram plot, the results demonstrate that visual surface representations lack a significant degree of information compared to weighted normals. The reduction from 94% to 74% of explained image variability and the increase in the spread of R^2 values from weighted normals to visual surfaces indicate that the estimation process used to create visual surfaces is a significant source of information loss. As a result, further surface refinement is needed to better capture microfossil shapes. This refinement, coupled with non-Lambertian reflectance models, would serve to construct more accurate visual surfaces. Thus, using Lambertian reflectance assumptions to estimate visual shapes falls short as an effective template representation. This motivates further work on microfossil shape extraction, one incorporating non-Lambertian reflectance and fine detail.

However, in the meantime, the inherent strength of video-based representations leads to an interim solution, one that allows the system to retain the use of shape-based representations. In contrast to shape-based representations with Lambertian albedos, videos inherently capture all reflectance and texture characteristics, including cast shadows. By restricting possible light directions to the ones used to create the videos, one can apply the associated image frames onto a surface. In doing so, the same characteristics captured in the video are incorporated into the shape-based representation. When a user wishes to change the light direction, the accompanying texture is changed as well.

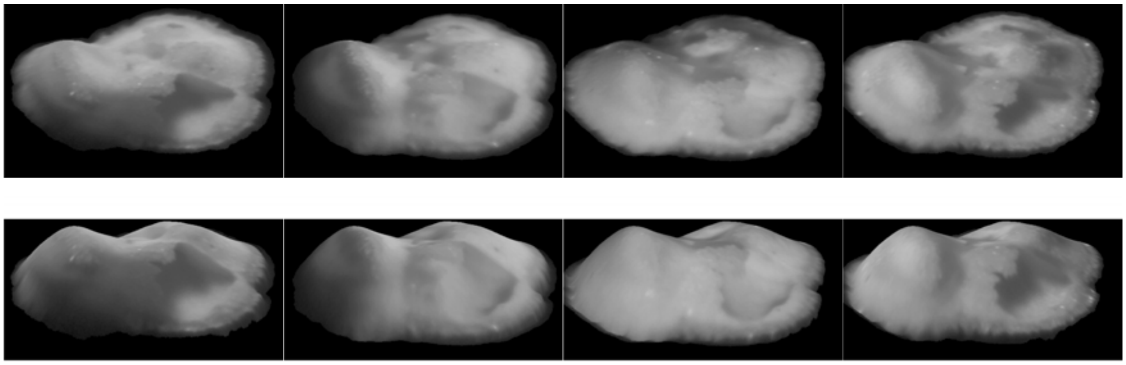


Figure 3.12: Texture-mapped shape-based representations. This figure depicts texture-mapped shape-based representation of the same microfossil in Fig. 3.10. Texture maps are taken from the image frames in the first row of Fig. 3.10. In this figure, viewing elevation angles for the first and second rows are 15° and 30° respectively.

Unfortunately, this means that users are restricted to the 18 light directions used in the video-based representation. As well, the effectiveness of this approach diminishes as the viewing direction deviates from the camera viewpoint used to capture images. Nonetheless, when viewing the texture-mapped shape from above, the visual information remains identical to the video-based representation. As well, as long as the viewpoint does not drastically deviate from an elevation angle of 0°, the textured shape is able to provide more insight into the microfossil geometry than a video.

Fig. 3.12 depicts an example of a texture-mapped visual surface at elevation angles of 15° and 30°. Being at a less severe elevation angle, the viewing direction of the first row of Fig. 3.10 presents a useful viewpoint that retains important details such as the microfossil spires. However, in the second row of the figure, the elevation angle is at 30°, making the viewpoint closer to the ground plane. As a result, the detail lost in the shape-based representations makes more of an impact, as features evident on the texture map, such as the spires, are deemphasized at such viewing angles. This example illustrates the restrictions on viewpoint elevation that the texture-mapped shape-based representations require. Nonetheless, as long as elevation angles are not too severe, texture-mapped shape-based representations can match and exceed the abilities of videos as digital template representations.

However, as the wiki and most browsers do not inherently support rendering shapes, texture-mapped shapes are difficult to deploy online. One powerful and easily deployable representation that incorporates shape information are anaglyph videos (videos consisting of image frames providing stereoscopic 3D effects). By using the shape-based representation and perspective projection to nonlinearly warp the video-based representations, anaglyphs can depict both fine detail and texture, and still provide insight into 3D geometry. Fig. 3.13 illustrates individual frames from an example anaglyph video sequence. Since anaglyphs can to depict 3D information without deviating from the original viewpoint, the suffer from none of the detail loss of texture-mapped shapes. Moreover, they are a striking and easily disseminated means to convey texture and 3D shape that requires no 3D rendering or other significant upgrades on the part of the online wiki. The only equipment requirements are inexpensive anaglyph glasses. Thus, with anaglyph videos, shape-based representations can be used as online template representations with no additional enhancements to the wiki. Appendix A provides more details on constructing analgyph videos.

3.6 Conclusion

Since images are limited representations for computer-aided identification, Chapter 2 argued that further improvement to microfossil classification performance requires alternative template representations such as video and shape. This chapter detailed the significant extensions applied to the existing computer-aided system to provide control of illumination direction and ultimately the ability to support both video and shape-based representations of microfossil specimens. Since light direction is fixed for typical microscopes, major extensions included incorporating an automatic x-y-phi stage into the system. Moreover, to inject as much autonomy as possible, an automatic scheme localizing, capturing, and disseminating very large batches of specimens was developed using a custom multi-threaded and user-friendly C++ software program.

Additional improvements included updating the clustering algorithm to a hierarchical scheme. Being both faster, and slightly more accurate than maximal-clique clustering, AgH clustering represents an important aspect of system improvements. In addition, as the online wiki serves as the desired channel to provide templates to experts, enhancing the capabilities of the wiki to display videos instead of single images constitutes another significant system extension. However, at this point the wiki lacks the ability to display shape-based representations. This represents an important aspect of future work.

Despite improvements provided by centroid alignment, image sequences collected by the system are still susceptible to significant relative misalignments. These misalignments affect both video and shape-based representations. Solving this problem required developing an image alignment rou-

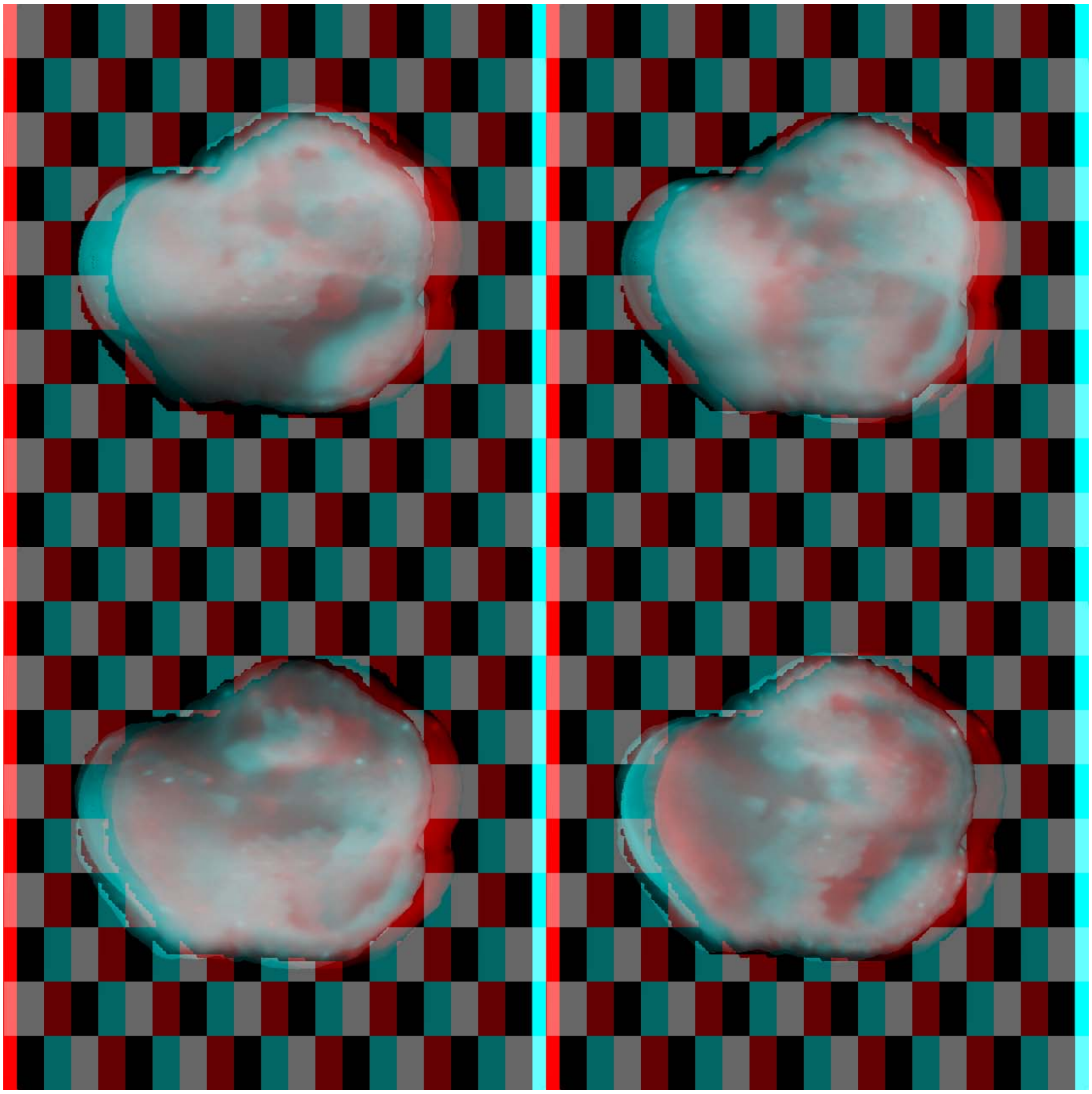


Figure 3.13: Example anaglyph images. This figure illustrates individual frames of an anaglyph video sequence. Red-cyan stereoscopic glasses are required to view these images. Top-left and top-right images are illuminated from 90° and 10° azimuthal angles respectively. Bottom-left and bottom-right images are illuminated from −90° and −170° azimuthal angles respectively. All images are illuminated from an elevation angle of 30°.

tine based on ML estimation and photometric stereo. Constituting a significant topic in its own right, Chapter 4 describes this alignment routine.

With the system's capabilities extended, a dataset incorporating multiple images of specimens under varying lighting conditions was collected. Comprised of 500 foram specimens, the dataset includes 18 images of each specimen, all illuminated from an elevation angle of 30° and from differing azimuth angles at 20° increments. Shape-based representations were extracted from these image sequences using methods from Chapter 5. Analysis of the dataset led to several conclusions regarding the capabilities of shape-based representations. When applied to the foram dataset, the visual surfaces accounted for 74% of image variability. In contrast, weighted normals successfully accounted for 94% of image variability, demonstrating that significant information is lost in the process of estimating surfaces from weighted normals. In addition, the 6% of image variability unaccounted for by the weighted normals indicates that Lambertian reflectance assumptions are another important source of error. However, information loss in the visual surface estimation step remains the greatest source of error.

Future work must focus on increasing the capabilities of methods used to extract shape-based representations. However, as an interim solution, the chapter introduced a texture-mapped visual surface that provides as much detail as the video-based representations when viewing the shape from above. In addition, the texture-mapped shapes offer insight into the 3D geometry of the microfossil by allowing viewpoint other than the one used to capture the videos. Complementary to texture-mapped shapes, the chapter also introduced the concept of anaglyph representations of microfossils. Sharing the same data structure as videos, anaglyph representations offer experts the same detail as video-based representations while simultaneously depicting the 3D information in the shape-based representations. Most importantly, employing anaglyph videos may be readily disseminated online. For this reason, anaglyph videos currently serve as the digital template representation of choice for computer-aided identification.

Chapter 4

Image Alignment Using Photometric Stereo

Using computer vision techniques, shape-based representations can be extracted from video-based representations. When applying photometric stereo to estimate surface normals, the video-based representations are treated as image sequences of the same object under the same viewpoint, but with differing and known illumination conditions. One vital assumption of photometric stereo is that every image in the sequence is aligned properly. Normally, this poses no problems, as photometric stereo is usually applied using an image sequence with a fixed camera and object and different known illumination directions. In contrast, the microfossil video capture system uses a fixed light source accompanied by an x-y-phi stage. As a result, when capturing individual frames of microfossil videos, the system must first rotate the specimen and then rotate the captured images back in the opposite direction. As Sec. 3.1.3 explained, because the rotational element is fixed on top of the translational elements of the stage, a specimen will have different (x, y) locations at each rotation.

While the expected (x, y) coordinates can be calculated, they are sensitive to errors in stage calibration, stage repeatability, and also the distance of the specimen to the origin of the stage's coordinate system. Thus, after rotation, the system attempts to mitigate these errors by centering the specimen in the field of view using its silhouette centroid. This is called *centroid alignment*. Unfortunately, when an object is illuminated from different directions, the extracted silhouettes will be slightly different for each image. Fig. 4.1 demonstrates the degree of difference between silhouettes of a typical centroid-aligned specimen. The variability between silhouettes at the boundary indicates that silhouettes have a degree of sensitivity to illumination direction. As a result, centered images of the same object will not always align properly. These errors can be considered as having introduced relative horizontal and vertical shifts in the images, resulting in a

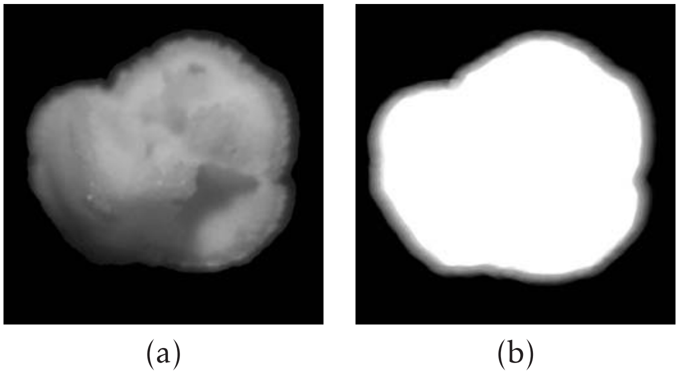


Figure 4.1: Silhouette differences in a centroid-aligned video sequence. One of the 18 image frames of a typical centroid-aligned video sequence of a microfossil is displayed in (a). In (b) the silhouettes of each image in the sequence have been added together, and the result has been rescaled so that areas of complete overlap are represented by pure white. Areas encompassing fewer image silhouettes range from black to grey.

misaligned image sequence. These misalignments affect both the video and shape-based representations collected by the computer-aided system.

Aligning or registering images is typically approached by either dealing with the intensity-based information incorporated in an image's pixels or by only dealing with their 2D shape or silhouettes. Intensity-based registration methods usually involve selecting features between a target and source image, determining correspondences between these features, and finally computing a transform aligning the target image with the source [62]. Typical features can include corners, lines, and image patches. Methods to extract and align based on these features constitute a heavily researched field relevant to subjects such as medical imaging, computer vision, and many industrial and commercial applications. Two very well known examples of image patch-based registration techniques are cross-correlation [29] and optical flow [63]. Regardless of the features used for alignment, they are all based in some way upon the intensity levels of the pixels in the images. Unfortunately, when aligning images of the same object illuminated from different light directions, pixel intensities will intrinsically change along with illumination angle. As a result, traditional image patch-based methods will fail under these conditions, and the changing intensity levels across images make consistent and reliable detection of lower-level features, such as lines or corners, a much more difficult problem. These reasons motivate the use of features outside of traditional intensity-based ones.

Apart from intensity-based alignment, the literature also includes much work on aligning based on silhouettes. Work on this subject often falls under the category of silhouette-matching, as matching two silhouettes often requires determining a transformation that best aligns them or determining similarity measures invariant to translation, rotation, and scale transformations [64, 65]. Probably, the most well known global silhouette descriptors are moments. In fact, since the misalignments introduced by the video capture module are limited to only horizontal and vertical displacements, performing centroid alignment is the appropriate moment-based method to use. However, global silhouette-based methods, such as moments, are sensitive to errors or changes in the silhouette boundary [65]. Since the extracted silhouette boundary of the video sequences is different at every illumination direction, global-based silhouette matching is not the optimal choice. Other more robust alternatives use methods that attempt to model noise and occlusions into the silhouette description. These methods can be based on representations describing the boundary or the entire silhouette [64]. The number of options is quite large, and involve considerations such as noise models, data representation, and a variety of other factors that all considerably affect performance and complexity. Additionally, it is also not clear how to best evaluate the results of silhouette matching [64]; thus, there is no agreed upon way to determine when error-prone silhouettes have been successfully aligned.

Both intensity-based and silhouette-based alignment techniques suffer drawbacks for applications involving image sequences illuminated from differing directions. Silhouette-based techniques are hindered by the inherent limitations of working with silhouettes, and intensity-based techniques rely on consistent pixel intensities across corresponding regions of images. However, if every image in the sequence was illuminated from the same direction, intensity-based alignment would be an attractive choice, as it performs better than its silhouette-based counterparts [65]. Thus, pixel intensities are important pieces of information that should be used when aligning image sequences of microfossil specimens. Incorporating a model of image formation is one powerful way to include intensity-based information in the alignment technique. This chapter describes such an image alignment technique. This novel alignment technique incorporates the Lambertian model of image formation and is designed to align images of the same object illuminated from differing directions.

4.1 Photometric Alignment

As the illumination direction changes from image to image in a sequence, to develop an alignment routine one must incorporate a model of image formation into the process. Using image formation as part of alignment process is called *photometric alignment*. This dissertation uses the Lambertian model of image formation, described in Sec. 2.2, to estimate surface normals. As a result, the Lambertian model is an appropriate choice to also use for photometric alignment.

Given an object with a continuous set of weighted normals η (defined in (2.2), recall that the weighted normals incorporate the albedo) and a one-to-one correspondence between horizontal and vertical coordinates of the real-world and pixel-space, photometric alignment uses the following equation to model the formation of a *misaligned* image (I_k):

$$I_k(x,y) = \boldsymbol{\ell}_k^T \cdot \eta(x + \Delta x_k, y + \Delta y_k) + \epsilon_k(x,y). \tag{4.1}$$

In other words, as defined by (4.1), misaligned images are individual observations of an object with a continuous set of normals under a set of relative shifts. An important assumption implicit in (4.1) is that pixels in background regions of the image are of a constant value. Assuming each sequence has N images, then an image sequence's set of shifts can be expressed as a single shift vector: $\beta = (\Delta x_1, \Delta y_1, \ldots \Delta x_N, \Delta y_N)$. By way of (4.1), determining shift values requires simultaneously determining the continuous surface normals. Since image noise follows IID Gaussian distributions, the ML estimate of β and η is one that minimizes the following sum-squared error of all observations:

$$SSE(\beta, \eta) = \sum_{k=1}^{N} \sum_{x=1}^{n} \sum_{y=1}^{m} r_k(\beta, \eta, x, y)^2, \tag{4.2}$$

where each individual residual term is defined as:

$$r_k(\beta, \eta, x, y) = I_k(x,y) - \boldsymbol{\ell}_k^T \cdot \eta(x + \Delta x_k, y + \Delta y_k), \tag{4.3}$$

and images are assumed to consist of $m \times n$ pixels.

While (4.2) may provide a theoretically valid condition on maximizing likelihood, the equation does not offer an avenue in which to easily determine image shifts. Fortunately, a simple change of variables provides a more practical formulation. Substituting $u = x + \Delta x_k$ and $v = y + \Delta y_k$ with $\Delta u_k = -\Delta x_k$ and $\Delta v_k = -\Delta y_k$, misalignment can be modelled as a set image shifts rather than a set of object shifts. This redefines β as $(\Delta u_1, \Delta v_1, \ldots \Delta u_N, \Delta v_N)$, meaning that the image formation equation is reexpressed as:

$$I_k(u + \Delta u_k, v + \Delta v_k) = \boldsymbol{\ell}_k^T \cdot \eta(u,v) + \epsilon_k(u,v). \tag{4.4}$$

Implied within this change of variables is an important distinction from (4.1); namely, images and their accompanying noise are treated as continuous entities and surface normal values are only considered at a discrete set of locations. Although an image is of course discrete by its nature, it can also be viewed as a continuous function sampled at pixel locations. Interpolating values at locations between pixels approximates the continuous underlying image; thus, providing a means to treat a discrete image as a continuous entity.

On the other hand, being for the most part a discrete phenomenon arising from processes within pixels, there is no continuous analogue for image noise. As a result, there is no appropriate noise distribution at locations in between pixels. Thus, interpolating at location $(u+\Delta u_k, v+\Delta v_k)$ results in a weighted sum of normally distributed error terms. While a sum of normally distributed and weighted variables is also normally distributed, variances from one image to another and across pixels will no longer be equal. This means that the ML estimate requires a GLS formulation. However, as variance values depend on the current set of shifts, properly accounting for the changing noise variance introduces a significant degree of complication. As well, the interpolation couples neighbouring noise terms with each other. Consequently, for the sake of simplicity, image noise shall be assumed to be IID normally distributed for all locations, even those residing in between pixels.

With misalignment modelled as a set of image shifts, the current shift estimate completely determines the current weighted normal estimate, $\hat{\boldsymbol{\eta}}$. Using the photometric stereo equation, the weighted normal estimate given a set of shifts is:

$$\hat{\boldsymbol{\eta}}(\beta, u, v) = (\mathbf{L}^T\mathbf{L})^{-1}\mathbf{L}^T\mathbf{i}(\beta, u, v), \tag{4.5}$$

where $\mathbf{i}(\beta, u, v)$ is defined as:

$$\mathbf{i}(\beta, u, v) = \begin{pmatrix} I_1(u+\Delta u_1, v+\Delta v_1) \\ I_2(u+\Delta u_2, v+\Delta v_2) \\ \vdots \\ I_N(u+\Delta u_N, v+\Delta v_N) \end{pmatrix}. \tag{4.6}$$

As a result, rather than simultaneously solving for image shifts and surface normals, only the correct image shift vector β need be estimated. This simplifies the sum-squared error formulation, as the SSE is now only a function of the shift vector β:

$$SSE(\beta) = \mathbf{r}(\beta)^T\mathbf{r}(\beta). \tag{4.7}$$

Here the residuals have been flattened into a single vector as defined by:

$$\mathbf{r}(\beta) = (r_1(\beta, 1, 1), \ldots r_1(\beta, n, m), r_2(\beta, 1, 1), \ldots, r_N(\beta, n, m))^T, \tag{4.8}$$

with each individual residual term defined as:

$$r_k(\beta, u, v) = I_k(u + \Delta u, v + \Delta v) - \boldsymbol{\ell}_k^T \cdot \hat{\eta}(\beta, u, v). \tag{4.9}$$

Note that (4.9) differs from (4.3) by incorporating the current weighted normal estimate, $\hat{\eta}$, which must be calculated using (4.5). Despite this difference, minimizing (4.7) is equivalent to minimizing (4.2) if one disregards the issues arising from modelling image noise as a continuous phenomena.

Minimizing the SSE determines the ML estimate of the shift vectors. This forms the basis of the photometric alignment technique. This is in effect a nonlinear function minimization scheme, where the only independent variables are the sets of image shifts β. Note that in using such a scheme, the shadowed version of the photometric stereo equation (2.8) cannot be used as that will introduce discontinuities, which many nonlinear least-squares and function minimizers do not handle. A great benefit of constructing residuals in this manner is that minimization also produces the best set of weighted-normal estimates, $\hat{\eta}$, under the photometric stereo assumptions outlined in Sec. 2.2. As a result, in contrast to schemes involving silhouette alignment, the ultimate goal of performing image alignment in the first place is incorporated directly into the residual term.

Many nonlinear minimization schemes require gradient or Jacobian evaluations. This requires partial derivative computations of the *SSE* term with respect to one of the variables in the shift vector β. These variables can represent shifts in either the u or v directions for one particular image. For instance, one can consider the partial derivative of (4.7) with respect to shifts in the u direction of I_ℓ. The partial derivative of (4.7) with respect to Δu_ℓ evaluates as:

$$\frac{\partial SSE(\beta)}{\partial \Delta u_\ell} = 2\mathbf{r}(\beta)^T \frac{\partial \mathbf{r}(\beta)}{\partial \Delta u_\ell}. \tag{4.10}$$

To compute (4.10), the partial derivative of $\mathbf{r}(\beta)$ at each pixel location and image with respect to Δu_l must be evaluated. This can be expressed as the difference between two terms:

$$\frac{\partial r_k(\beta, x, y)}{\partial \Delta u_\ell} = \frac{\partial I_k(u + \Delta u_k, v + \Delta v_k)}{\partial \Delta u_\ell} - \boldsymbol{\ell}^T \cdot \frac{\partial \hat{\eta}(u, v)}{\partial \Delta u_\ell}. \tag{4.11}$$

The first term of (4.11) is simply the partial derivative of the image I_k with respect to shifts of I_ℓ in the u direction. When $k = \ell$, the first term of (4.11) evaluates as the spatial derivative of I_ℓ in the u direction, which is denoted here as U_ℓ. For images other than I_ℓ the partial derivative will evaluate to 0. This is expressed mathematically as:

$$\frac{\partial I_k(u + \Delta u_k, v + \Delta v_k)}{\partial \Delta u_\ell} = \begin{cases} U_\ell(u + \Delta u_\ell, v + \Delta v_\ell), & \text{if } k = \ell \\ 0, & \text{otherwise} \end{cases}. \tag{4.12}$$

For the second term of (4.11), the only variables that change with respect to image shifts are the current weighted-normal estimates. Similar to the photometric stereo equation of (2.6), evaluating the partial derivatives of the normals requires solving an over-determined system; however, in this case the partial derivatives of the images with respect to shifts of I_ℓ in the u direction are used in the righthand side of the equation:

$$\begin{pmatrix} \boldsymbol{\ell}_1^T \\ \vdots \\ \boldsymbol{\ell}_N^T \end{pmatrix} \cdot \frac{\partial \eta}{\partial \Delta u_\ell} = \begin{pmatrix} \partial I_1(u + \Delta u_1, v + \Delta v_1)/\partial \Delta u_\ell \\ \vdots \\ \partial I_N(u + \Delta u_N, v + \Delta v_N)/\partial \Delta u_\ell \end{pmatrix}, \tag{4.13}$$

where the values on the righthand side evaluate as in (4.12). Thus, (4.11) can be evaluated for every location, incorporated into a column vector, and finally combined with (4.10) to result in the gradient calculation with respect to u_ℓ. An identical approach is used for evaluating the gradients with respect to the v direction.

Minimizing the nonlinear SSE function of (4.7) can be accomplished using function minimization schemes. Quasi-Newton (QN) methods are a popular technique often applied to nonlinear functions [48]. Conjugate gradients (CG) is another good and well-researched approach. See [66] for a detailed explanation of CG. One issue with nonlinear CG is the incorporation of a line search requiring significant amounts of function and gradient evaluations. Depending on the size of the images, both (4.7) and (4.10) may be very expensive to compute. An alternative is to use scaled conjugate gradients (SCG) [67], which is designed to minimize gradient evaluations.

While function minimization schemes are a viable approach, the problem of aligning images can also be solved using methods tailored for nonlinear least-squares problems. If there are N images, instead of using one single error term, N error terms can be formulated. As the error term would consist of an $N \times 1$ column vector, a Jacobian would have to be calculated instead of a gradient. Such a scheme would then attempt to determine a set of shifts that best fit the N error terms to zero. For medium scaled problems such as this one, the Levenberg-Marquart (LM) [48] algorithm is typically used. In addition to a medium-scale formulation, the alignment problem could also be treated as a large-scale least-squares problem, with $N \times m \times n$ error terms. However, the scale of the problem is so large, especially when computing gradients, that it quickly becomes intractable. As a result, this option was not considered in testing.

Photometric alignment only corrects for relative misalignments between images in a sequence. As a result, every image in a sequence can be shifted by the same amount, and photometric alignment would not be able to differentiate between the two cases. Thus, the entire sequence may not be centered properly. To fix an absolute location for the sequence, one can anchor one of the images in the sequence, forcing the other images to align with the chosen image. However, the choice of which image to use is arbitrary. Another approach does not require anchoring an image, but simply allows the routine to first determine the set of optimal relative shifts. Afterwards, combining the silhouettes of each image produces a silhouette for the entire sequence. The sequence is then centered using the centroid of this *combined* silhouette. This is what was done for this dissertation.

4.2 Evaluating the Method

The usefulness of photometric alignment hinges on two factors: the accuracy of the alignment method and the benefits, if any, of using the technique. As stated in the previous section, photometric alignment can be performed using methods that include QN, CG, SCG, and LM. Sec. 4.2.1 focuses on determining whether the performance of photometric alignment depends on the minimization method used. As well Sec. 4.2.1 tests how accurately photometric alignment can correct misalignments. Sec. 4.2.2 and Sec. 4.2.3 then explore the benefits of applying photometric alignment to microfossil videos captured by the computer-aided system.

4.2.1 Testing with Known Misalignments

The accuracy of photometric alignment was tested by shifting a sequence of images by known amounts. If the algorithm functioned perfectly, the outputted shifts should be the exact opposite of the shifts applied to the sequence. Thus, the disparity between the known and correcting shifts serves as a good metric in which to judge the algorithm's accuracy.

Image sequences were obtained from the PMTex database[1]. Composed of square 512 × 512 images of rock textures illuminated from known directions, the PMTex database provides excellent image sequences in which to test the alignment routine. As well, the database supplies a large variety of illumination options for each rock texture. A good measure of the accuracy of the algorithm is the shift error (SE), defined as the Euclidean distance of the error in the u and v directions:

$$SE = \sqrt{(\Delta u_{actual} + \Delta u_{correcting})^2 + (\Delta v_{actual} + \Delta v_{correcting})^2}, \tag{4.14}$$

where $\Delta(.)_{actual}$ and $\Delta(.)_{correcting}$ are the actual and correcting shifts respectively. Ideally, SEs should be as low as possible, since if photometric alignment functioned perfectly, the correcting shifts would be the exact opposites of the applied shifts.

[1] http://people.pwf.cam.ac.uk/jw566/research/pmtexdb/index.htm

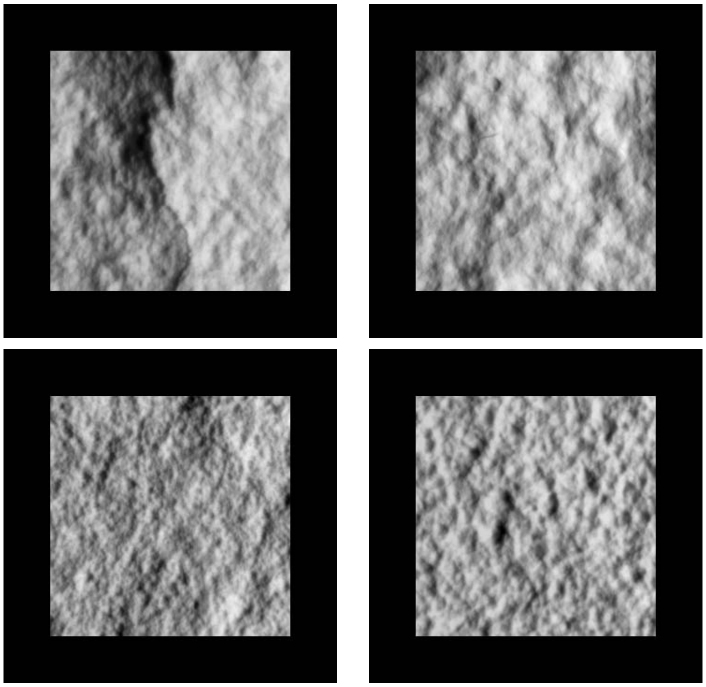


Figure 4.2: Example images of the four textures for testing photometric alignment accuracy. Each of these images are illuminated from an elevation angle of 45° and an azimuth angle of 180°. The depicted images were all padded with zero-intensity pixels.

Four different textures were chosen. Examples of each texture can be seen in Fig. 4.2. In the actual system set-up of the video-capture module, there are two important characteristics. For one, light directions for the image sequences of the specimens exhibit constant elevations but changing azimuth angles. As a result, the light directions for the image sequences of the rock textures were constrained in the same way.

A second important characteristic, which is the cause of misalignments in the centroid-aligned images in the first place, is a changing silhouette from image to image in the same specimen sequence. For this reason, in addition to shifting the images by known amounts, experiments should also simulate the effects of a changing silhouette between images of the same sequence. A mask, designed to vary across different light directions, can approximate the changing silhouette across images. As Fig. 4.3(b) demonstrates, for a particular image, the masking process begins by generating a simulated image of a Lambertian perfect hemisphere with a constant albedo of 1 illuminated from the same direction. As the hemisphere image will have lower pixel values in regions opposite of the light direction, thresholding the sphere image forms a varying circle-based mask that approximates

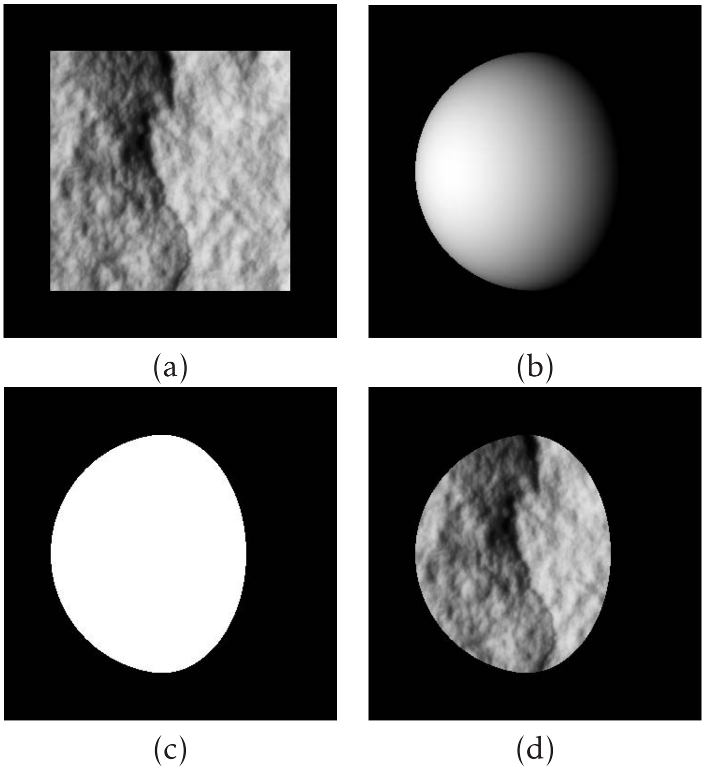
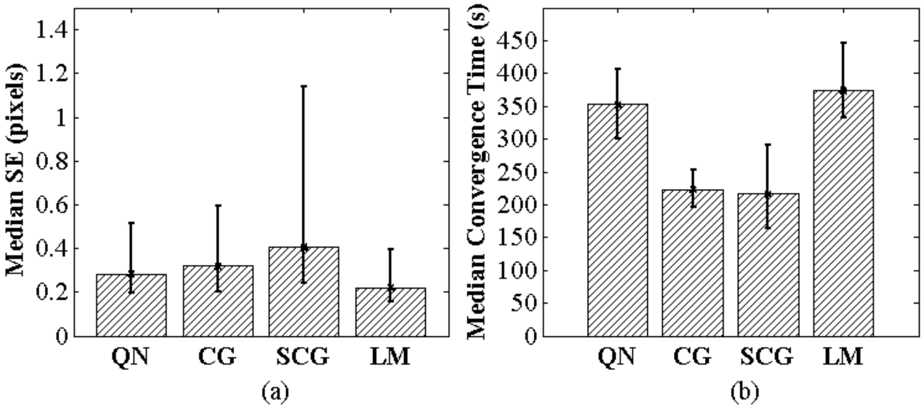


Figure 4.3: The masking technique for the texture images. (a) example texture photograph after padding, illuminated from an azimuthal angle of 180° and an elevation angle of 45°; (b) simulated image of a perfect hemisphere illuminated from the same direction as the example picture in (a); (c) a mask generated by thresholding the hemisphere image by a value of 0.1; (d) The example image in (a) after applying the mask in (c) to it.

how an image sequence's silhouette changes with illumination. For the purposes of this experiment, a threshold value of 0.1 was chosen. Fig. 4.3(c) illustrates the masked version of Fig. 4.3(b). As Fig. 4.3(d) demonstrates, a set of these masks applied to every corresponding image in the sequence, results in silhouette dependant upon illumination direction.

Three image sequences of each rock texture were used for testing, each possessing fixed elevation angles of either 45°, 60°, or 75°, resulting in 12 separate sequences. For every image sequence, the azimuth angle of illumination was set to increase by increments of 30° from 0° to 330°, resulting in 12 images per every sequence. Prior to experimentation, the images were first padded with pixel values of 0 and then resized, for computational speed purposes, to 90×90 pixels. The images were then randomly shifted in the u,v directions with integer values ranging from -5 to 5. As a result, the maximum absolute shift is over 5% of the image dimensions, and the



	QN	CG	SCG	LM
Max SE (pixels)	4.73	2.79	3.99	2.6
Max Time (s)	668.1	328.3	473.3	577.1

(c)

Figure 4.4: Performance of photometric alignment on the 24 sequences of texture images. In both (a) and (b) median values are displayed, with error bars representing the first and third quartile values. In (c) the max SE and convergence times of all four methods are displayed. The error values graphed in (a) indicate that QN, CG, and LM demonstrate comparable accuracy, with LM slightly better than the rest. SCG's third quartile is value is significantly worse than the other three methds. In (b), CG and SCG have the best convergence times.

maximum relative shift between two images is over 10% of the image dimensions. This was performed twice for each image sequence, resulting in 24 separate tests. As each image sequence is composed of 12 images, experimentation results in 288 separate shift error values. Each test recorded the performance of QN, CG, SCG, and LM. Experiments were performed in MATLAB. The NETLAB [2] implementation of CG and SCG were used, and for QN and LM the built-in MATLAB implementation was used. The stopping criteria for each method were termination tolerances of 10^{-4} for both the objective function and the shift values.

Fig. 4.4 graphs the median SE and convergence times of photometric alignment using the four minimization methods. All of the minimization options performed well. Out of the three methods, QN, CG and LM aligned the images with the best accuracy. However, QN's max SE value was the worse of all four methods. As well, LM edged out all other methods in terms

[2] http://www.ncrg.aston.ac.uk/netlab/index.php

of correction accuracy. However, as Fig. 4.4(c) demonstrates, the variability and max convergence times of CG are much smaller than that of QN or LM. Since convergence times are difficult to compare across different MATLAB implementations, these results do not indicate that CG will inherently converge faster than QN or LM methods. Yet, since the accuracy of CG is comparable to that of QN or LM, and the available implementation is almost twice as fast, all future experiments use CG as the minimization routine.

The similarity in results across minimization techniques indicate that photometric alignment is not dependant on a specific minimizer. It is important to note that in an effort to test the algorithm's robustness, conditions used in this experiment were purposely designed to shift images by values larger than typical real-world situations. As well, silhouettes in the tested image sequences underwent more change from image to image than in the real specimen dataset. Consequently, these results indicate that photometric alignment is very robust, even when applied under extraordinarily difficult circumstances.

4.2.2 Quantifying Benefits

Having established that photometric alignment can accurately correct for misalignments, the benefits of including the technique as part of the shape-based representation extraction process must also be quantified. Doing so requires recreating the conditions in which a sequence of images becomes misaligned in the first place. As differences in centroid coordinates across the silhouettes of image sequences is the cause of shortfalls of using centroid alignment, the source of these misalignments must be reproduced.

Representing a key characteristic of the misalignment process, the severity of alignment errors is directly related to the angle of elevation across all images. More specifically, illumination angles closer to the ground plane will result in greater differences in object masks as image variability across different angles of azimuth will be greater. As a result, any potential benefits from aligning images will depend in part on the angle of elevation of the light source.

To test this, perfect hemisphere image sequences with uniform albedo were created. The hemispheres were illuminated at constant elevation and changing azimuth. Masks of each image can be produced using simple thresholding. These masks were then applied to each hemisphere image. Each image was then centroid-aligned. As the area of the mask that is thresholded out changes with azimuth angle, each image possesses a unique set of centering shifts. This results in a sequence of misalignments approximating the centering errors inherent in image capture with the x-y-phi motorized stage and centroid alignment. Fig. 4.5 illustrates the process behind

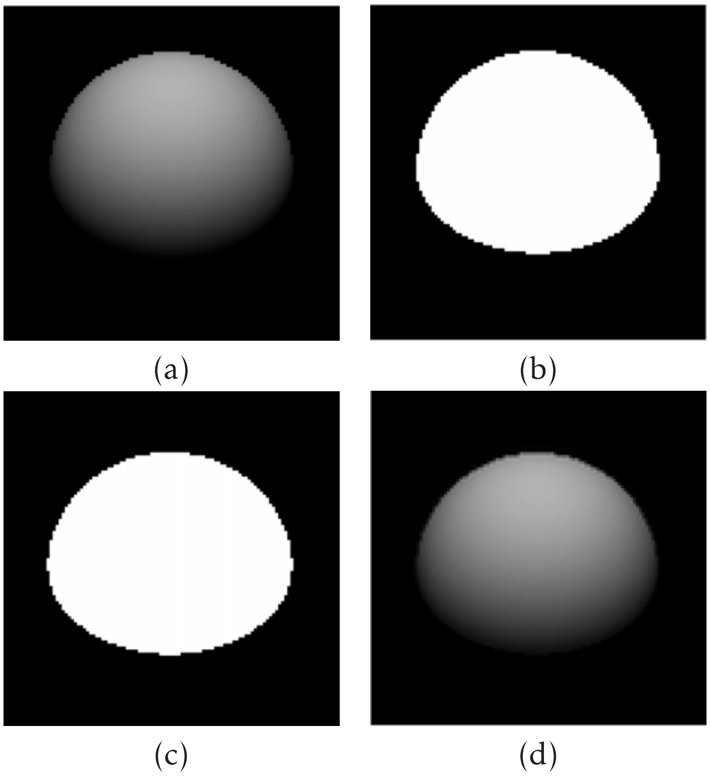


Figure 4.5: Producing the error-prone centroid-aligned hemisphere images. (a) a perfect hemisphere illuminated from an elevation angle of 45° and an azimuth angle of 90°; (b) mask of (a) at a threshold value of 0.04; (c) the mask in (b) shifted so that its centroid is centered in the image; (d) the image in (a) shifted using the same shifts used to center its mask in (c).

generating the error-prone centroid-aligned image sequences.

By applying photometric alignment to the centroid-aligned images, one can obtain a corrected set of images. The similarity score of (3.3), which computes median correlation value between pairs of image sequences, can measure similarity between the photometric-aligned images and the original images. The same can be done with the centroid-aligned images.

Experiments consisted of hemispheres with a radius of 40 pixels, illuminated from elevation angles ranging from 15° to 70°. All sequences consisted of 18 images, with azimuth angles varying from 0° to 340° by increments of 20°. Using the same value as the video capture module of the system, the image threshold was set to 0.04.

Fig. 4.6 graphs similarity scores across different elevation angles between the ground truth and the photometric and centroid-aligned image sequences. As expected, similarity scores of the centroid-aligned images decrease as the elevation angles and degree of misalignment increases. The

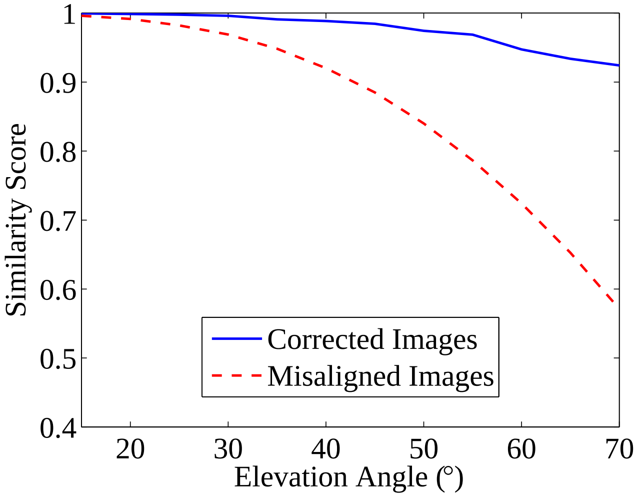


Figure 4.6: Quantitative benefits of using photometric alignment. Similarity scores between photometric and centroid-aligned images and the ground truth were computed using (3.3). Similarity scores of centroid-aligned images decreases significantly as the severity of elevation angle increases. In contrast, photometric alignment successfully mitigated similarity score reduction, retaining relatively constant correlation values even at elevation angles of 70°.

similarity scores of the photometric-aligned images also decline, but at a very gradual rate. In addition, they exhibit higher similarity scores at every elevation angle. The differences in similarity between the centroid and photometric-aligned images are very pronounced at angles greater than 35°. These results indicate that when under conditions involving horizontal and vertical misalignments, performing photometric alignment is a crucial step prior to measuring similarity between image sequences.

The qualitative benefits of using photometric alignment on shape extraction can be demonstrated by using the methods of Chapter 5 to estimate visual surfaces from both the centroid and photometric-aligned images. Depth maps and their cross-sections from an elevation angle of 45° are illustrated in Fig. 4.7. As shown by Fig. 4.7(b) and (e), the depth map of the centroid-aligned image sequence produces a flattened hemisphere. In contrast, the photometric-aligned depth map of Fig. 4.7(c) and (f) is much more hemispherical in nature, and is much more consistent with the ground truth depth map. Consequently, photometric alignment provides both quantitative benefits, when measuring similarity, and also qualitative benefits when constructing 3D models.

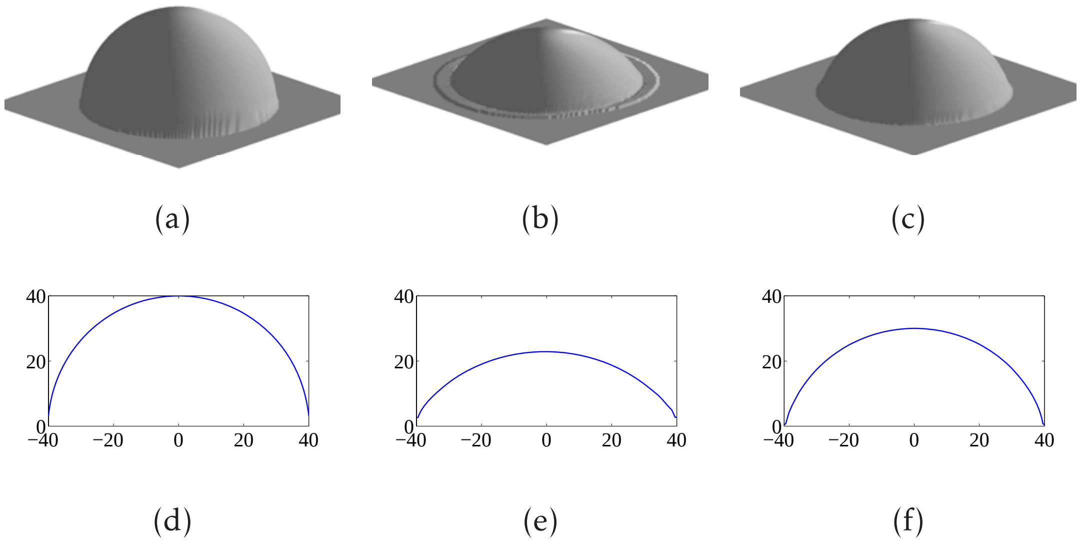
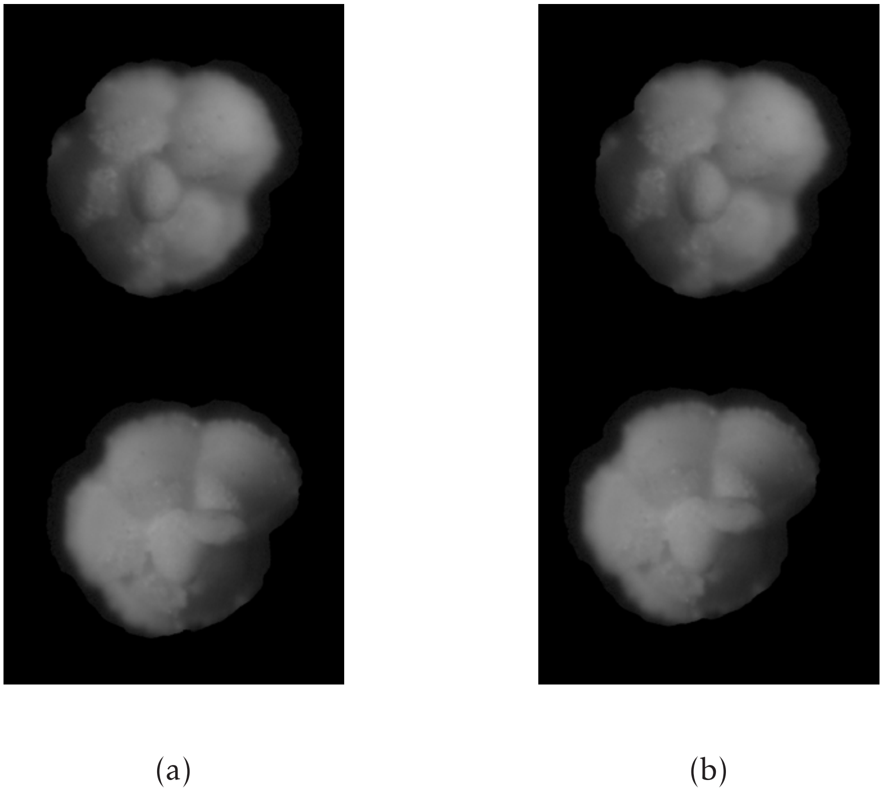


Figure 4.7: Qualitative benefits of using photometric alignment. (a) ground truth depth map of the hemisphere; (b) depth map produced from a centroid-aligned image sequence illuminated by a light source elevated at 45°; (c) depth map of the photometric-aligned sequence; (d-f) Cross sections of the ground truth, centroid-aligned, and photometric-aligned depth maps respectively. Depth maps in (b) and (c) were produced using the Modified-ML Surface Estimation method introduced in Chapter 5.

4.2.3 Experiments with Microfossil Videos

With the accuracy and usefulness of photometric alignment demonstrated, the remaining question to answer is how well it fares on sets of centroid-aligned microfossil video-based representations collected by the computer-aided system. As mentioned in Sec. 3.1.3, individual frames of the video-based representations were centroid-aligned after image capture in an effort to correct for misalignments caused by errors in the motorized stages. Photometric alignment was then applied to the 500 specimen foram dataset. For reasons of computational speed, prior to photometric alignment, the 640×640 pixel images were resized to 160×160 pixels. Afterwards, the computed shifts were applied to the original images, with each shift scaled by a factor of 4.

As Fig. 4.8 qualitatively illustrates, photometric alignment can successfully correct misalignments in the foram image sequences not corrected by centroid alignment. Although the accuracy of the shift values cannot be measured, as in Sec. 4.2.1, considering the R^2 values of the residuals is one



(a) (b)

Figure 4.8: Example of two images in a in a microfossil video using centroid and photometric alignment. The top and bottom images are of a foram illuminated from an azimuth angle of 70° and 260° respectively. On the left are the images after centroid alignment, and on the right are the same images after photometric alignment. Features on the left-hand images do not line-up properly. However, after application of photometric alignment, corresponding features are lined-up much better on the right-hand images.

way to quantitatively evaluate photometric alignment's performance. In this case, the weighted normals produced from the centroid-aligned image sequences and the weighted normals produced from the photometric-aligned image sequences represent the two hypotheses. Denoting the centroid and photometric-aligned weighted normals by $\hat{\eta}_{centroid}$ and $\hat{\eta}_{photometric}$ respectively, the expression for R^2 is written as:

$$R^2 = \frac{\mathrm{SSE}(\hat{\eta}_{centroid}) - \mathrm{SSE}(\hat{\eta}_{photometric})}{\mathrm{SSE}(\hat{\eta}_{centroid})}. \tag{4.15}$$

The R^2 value can be viewed as a description of how much of the difference between the actual images and their generative model counterparts

Shift Values	
Min	4.04×10^{-4} pixels
Median	1.19 pixels
Max	7.68 pixels
R^2 **Values**	
Min	1.87%
First Quartile	37.4%
Median	47.3%
Third Quartile	55.7%
Max	88.2%

Table 4.1: Results of applying photometric alignment to microfossil videos. Applying the algorithm to 500 160×160 pixel centroid-aligned microfossil image sequences resulted in median shifts of 1.19 pixels. Despite these somewhat small shift amounts, median R^2 values were 47.3%, indicating that photometric alignment can successfully mitigate much of the error between the actual images and their generative model counterparts.

can be attributed to misalignments within the centroid-aligned video-based representations that were later corrected by photometric alignment. Measurement errors, noise, any remaining uncorrected misalignments, and the inherent limitations of the chosen generative model are responsible for any remaining residual values. An R^2 value of 1 would indicate that photometric alignment corrected for all errors. Table 4.1 summarizes the results of the alignment technique and Fig. 4.9 plots the distribution of R^2 values.

As the table indicates, photometric alignment reduced reconstruction errors by on median 47%. The histogram demonstrates that this median value is representative of the different R^2 values across the microfossil dataset. The histogram also indicates that R^2 values can also correspond to very high or very low values with reduced frequency. However, for the most part, misalignments not corrected by centroid alignment accounted for much of the difference between pixel values of the actual and reconstructed images. This means that photometric alignment successively decreased a significant amount of error in the reconstructed images. As a result, applying photometric alignment to microfossil videos is an essential step in increasing the accuracy of depth map extraction.

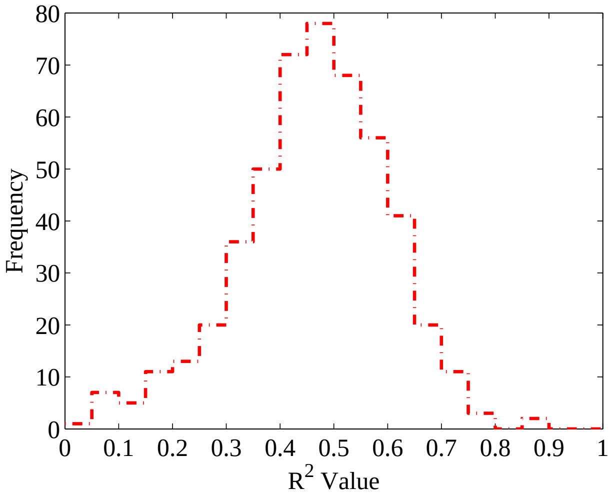


Figure 4.9: Histogram of R^2 values. This figure graphs the distribution of R^2 values and illustrates the reduction in reconstruction error provided by photometric alignment. The majority of R^2 values cluster around 50% reduction in error, with the frequency falling off relatively symmetrically on either side.

4.3 Asymptotic Complexity

As photometric alignment relies on nonlinear minimization, it is difficult to provide guarantees on the asymptotic number of arithmetic operations. However, as weighted normals can be estimated on their own at every image shift value, the scale of the nonlinear problem is not large, as the minimization routines need only determine image shifts. An important aspect that can be determined is the complexity of evaluating the function and its gradient, which are two steps frequently performed by the nonlinear minimization routines. As function evaluations consist of computing residuals of the weighted-normal estimates, the complexity involved in computing these estimates is important. Suppose there are N images, each composed of $n \times n$ pixels. At each pixel, weighted-normal estimation involves executing a least-squares solution involving an $N \times 3$ matrix. With these characteristics, the complexity of each least-squares computation is $O(N)$ [54]. Since this occurs for each pixel, the total complexity of weighted-normal estimation is $O(Nn^2)$. In addition, the residual error of each estimate must also be calculated. But as that also consumes $O(Nn^2)$ arithmetic operations, it does not add to complexity. The process of evaluating the gradient is essentially N separate function evaluations. As a result, gradient evaluations consume $O(N^2n^2)$ arithmetic operations.

In terms of memory use, Newton-based methods such as QN and LM consume $O(N^2)$ memory, where N is the number of parameters [48]. On the other hand, conjugate gradients consume $O(N)$ memory [66]. However, as N is typically not very large (2×18 for the microfossil dataset), memory use is not a significant issue.

4.4 Conclusion

This chapter presented a novel alignment routine called photometric alignment that is designed specifically to correct for relative misalignments in a sequence of photometric stereo images. The routine requires that the object in each image is illuminated by known directions and that misalignments are restricted to horizontal or vertical shifts. As well, pixels are assumed to be constant at all background locations. Enabling it to align images of distinct intensities, photometric alignment incorporates the Lambertian model of image formation directly into its error term. Consequently, a direct result of the routine are normal estimates that best correspond with the given images (under the assumptions described above).

The chapter also described experiments testing the accuracy, robustness, and benefits of photometric alignment. To test accuracy, photometric alignment was applied to sequences of images shifted by known amounts. The accuracy of photometric alignment was only off by a median 0.3 pixels when using CG, QN, or LM minimization. These results demonstrate the accuracy of photometric alignment. Even so, despite the low median SE value, the routine did produce high max SE values, indicating that the algorithm is susceptible to local minima. However, such occurrences are outliers, as demonstrated by the low third quartile values graphed in Fig. 4.4. As well, it should be noted that the conditions of this experiment were particularly strenuous, as the silhouettes from image to image varied by considerably more than what is typical in the microfossil dataset. The median absolute shift value was 2.5 pixels, or roughly 2.8% of the image dimensions (90×90 pixel images). In contrast, when photometric alignment was applied to the microfossil images dataset, the algorithm corrected the sequences with a median absolute shift value of 1.2 pixels, which is roughly 0.8% of the dimensions of the 160×160 pixel foram images. Although it may be confounded by local minima in very difficult conditions, photometric alignment still manages to align image sequences with very good accuracy.

In addition to presenting results on the photometric alignment's accuracy, this chapter also demonstrated the worth of using such an alignment scheme through two separate experiments. The first such experiment explored the detrimental effect of misalignments not corrected by centroid alignment, as seen in the microfossil dataset. As well, the experiment explored the beneficial effects of correcting for these latent misalignments through the use of photometric alignment. To accomplish this, a perfect hemisphere was artificially illuminated under similar conditions as that of the microfossils. The silhouette of each image was artificially truncated at several different angles of elevation. Centroid alignment was then used to center each image, recreating the cause of the errors seen in the microfossil dataset. With higher elevations come greater truncation, and thus greater degree of error in the centroid-aligned image sequences. As a result, the correlation score between centroid-aligned images and their original counterparts quickly fell off as the degree of misalignment increased. On the other hand, photometric-aligned image sequences produced consistent and high similarity scores. Qualitatively, the centroid-aligned depth maps were much flatter as misalignment increased, while photometric-aligned depth maps exhibited a form much closer to the desired hemisphere.

In addition, the benefits of applying photometric alignment to the centroid-aligned microfossil dataset were quantified. When applying photometric alignment to the microfossil videos, the representations are treated as sequences of individual image frames. Since no ground truth of the shift values or depth map is known beforehand, the only available metric is measuring R^2 values of residual terms before and after photometric alignment. These residual terms measure the error between the actual and reconstructed images making up the microfossil video sequences. With a median R^2 value of 47%, photometric alignment successively mitigated significant amounts of the error, meaning that misalignments not corrected by centroid alignment accounted for over 45% of the error in the microfossil dataset. As a result, estimated normals of the specimen surfaces were significantly more consistent with the given images. These results demonstrate the great worth in applying photometric alignment to the microfossil dataset.

Photometric alignment is applicable to any situation involving horizontal and vertical misalignment in a sequence of images. In many situations, to automatically obtain images of an object illuminated from differing directions, it is more feasible to rotate the object rather than the light source. In the case of the system setup used for this work, the microscope used has a fixed light source. This is also the standard for most other microscopes. As well, since a motorized stage is already required for (x, y) localization of the microfossil particles, it is much more practical to attach an additional rotational element, than to design a non-static light source. This would also be the case for many other objects small enough to require microscopes. However, the applicability of photometric alignment is not limited to microscopic objects, as the alignment algorithm would also prove beneficial to any situation where pixel correspondence between images can no longer be guaranteed.

Chapter 5

Maximum Likelihood Surface Estimation

Reconstructing shape from weighted normals requires methods to estimate surface. Research into visual shape extraction is ongoing. Although significant progress has resulted in several useful shape integration techniques, most current methods do not incorporate an *image* noise model into their formulations. The methods that do so only work under limited conditions [46] or use complicated nonlinear minimization routines [41,42]. However, incorporating an image noise model into surface reconstruction is important, as it allows the generation of ML estimates of the surface. This chapter develops an ML surface estimation technique that avoids imposing additional constraints and does not suffer from practicality issues.

5.1 Estimating the Gradient Fields

Developing a surface reconstruction routine that handles noise within its formulation requires a model of noise and its propagation through all steps of the reconstruction process. Yet, as Noakes and Kozera note, many of the classic surface extraction techniques that integrate gradients, such as [68–70], only produce an ML estimate should the gradient fields and not the image observations be corrupted by uniform Gaussian noise [41,42]. Although Gaussian noise is a reasonable assumption for image observations, the nonlinear transformation producing gradient estimates will not result in simple IID additive Gaussian noise [41]. Variations of these classic approaches have often attempted to handle the noise present in the gradients. However, these more recent techniques either again assume uniform Gaussian noise in the gradients [43], or base their techniques on heuristic weighting schemes [45].

On the other hand, more recent work on surface extraction correctly models gradient noise as a function of the uniform Gaussian noise in image observations [46]. Unfortunately, the assumptions made to formulate the derivations of gradient noise break down if surface normals have an angle greater than 6% with respect to the optical axis (meaning the surface must be almost horizontal [46]).

The nonlinear nature of gradient field noise is a potent motivation for skipping gradient estimation altogether and reconstructing surfaces directly from image observations. For instance, Noakes and Kozera developed a technique to solve for Z directly using an iterative technique entitled 2D Leap Frog [41, 42]. Since their optimization problem works directly from images to the surface, noise is kept IID, allowing the solution to correspond to an ML estimate. While their technique demonstrates promise, its practicality is questionable. First, their technique assumes a unit albedo throughout the object surface, a restriction certainly not reflected in most real-world objects. Additionally, like all iterative procedures, the 2D Leap Frog technique relies on the convergence properties of the chosen algorithm and the initial guess. Initial guesses in published experiments use the *true* surface corrupted by adding uniform Gaussian noise. Such experiments are not very instructive of the technique's practical capabilities. As crucial as these two issues are, convergence performance remains the biggest challenge with the 2D Leap Frog technique. This challenge prevents the algorithm from being applied to realistic dimension sizes [71]. These performance issues have led the authors to investigate implementations using parallel architectures [71].

While estimating the surface directly from image observations without any intervening steps is certainly theoretically convenient in terms of ML estimation, practical issues related to the nonlinear relationship between image intensities and surface values remain a serious obstacle. Nonlinear regression theory provides an alternative that retains the desirable two step process of first estimating gradients, and then linearly estimating the surface from these gradients.

To accomplish this, the nonlinear aspect of gradient and albedo estimation must be considered. Assuming a weighted-normal estimate using (2.33) has been obtained and allowing $\boldsymbol{\beta}$ to denote $(p, q, \rho)^T$, then for each pixel location:

$$\begin{pmatrix} \hat{p} \\ \hat{q} \\ \hat{\rho} \end{pmatrix} = \begin{pmatrix} \hat{\eta}_x / \hat{\eta}_z \\ \hat{\eta}_y / \hat{\eta}_z \\ \sqrt{\hat{\eta}_x^2 + \hat{\eta}_y^2 + \hat{\eta}_z^2} \end{pmatrix} \simeq \begin{pmatrix} p \\ q \\ \rho \end{pmatrix}, \tag{5.1}$$

$$\hat{\boldsymbol{\beta}} = \mathbf{h}(\hat{\boldsymbol{\eta}}), \tag{5.2}$$

where $\mathbf{h}(\boldsymbol{\eta})$ represents the relationship mapping the weighted normals to the gradients and albedo. Note that $\mathbf{h}(\boldsymbol{\eta})$ is valid for all values of $\boldsymbol{\eta}$ except for when $\eta_z = 0$. In addition, apart from this degenerative case, there is a one-to-one correspondence between the values of $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$.

Using the nonlinear parameterization of (5.1), one can incorporate the gradient fields and albedo into a system of image formation equations:

$$\mathbf{i} = \mathbf{L}\frac{\rho}{\sqrt{p^2+q^2+1}}\begin{pmatrix} -p \\ -q \\ 1 \end{pmatrix} + \boldsymbol{\epsilon}, \tag{5.3}$$

$$\mathbf{i} = \mathbf{f}(\boldsymbol{\beta}) + \boldsymbol{\epsilon}, \tag{5.4}$$

$$\mathbf{i} = \mathbf{f}(\mathbf{h}(\boldsymbol{\eta})) + \boldsymbol{\epsilon}, \tag{5.5}$$

where $\mathbf{i}$ is a vector incorporating pixels from each light direction. While (5.3) provides a parameterization directly incorporating the desired terms, in practice it is more convenient to use the linear parameterization using the weighted normals, as the solution can be solved directly using linear least squares. As well, the linear regression solution and transformation is equivalent to the nonlinear regression solution. After solving for the weighted normals, computing $\hat{\boldsymbol{\beta}} = \mathbf{h}(\hat{\boldsymbol{\eta}})$ determines the gradient and albedo estimates.

As discussed in Sec. 2.3.1, under certain regularity conditions, nonlinear ML parameter estimates can follow certain asymptotic properties, most notably asymptotic unbiasedness and normality. Additionally, the discussion introduced the following expression for the asymptotic limit of the parameter estimate covariance matrix:

$$\sigma^2\boldsymbol{\Gamma}_{\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}} = \sigma^2(\mathbf{F.}(\boldsymbol{\beta})^T\mathbf{F.}(\boldsymbol{\beta}))^{-1}, \tag{5.6}$$

where $\mathbf{F.}(\boldsymbol{\beta})$ is the Jacobian of the model function with respect to $\boldsymbol{\beta}$ evaluated at the true parameter value. In practice, the Jacobian is evaluated at $\hat{\boldsymbol{\beta}}$, producing an estimate, $\hat{\boldsymbol{\Gamma}}_{\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}}$, of the covariance. While (5.6) is perfectly valid and can also provide a convenient form of the covariance inverse, the inverse in the expression prevents it from offering an analytical expression for the actual covariance. Obtaining such an analytical expression requires using an alternative formulation. Presenting a relationship between the covariance of the linear parameter estimates, $\hat{\boldsymbol{\eta}}$, and that of $\hat{\boldsymbol{\beta}}$, the expression in (2.26) provides such an alternative formulation. As explained in Sec. 2.3.3, assuming the photometric assumptions are valid, the covariance matrix of the weighted normal estimates is $\sigma^2(\mathbf{L}^T\mathbf{L})^{-1}$. Should the shadowed photometric equation of (2.8) be used, the appropriate rows of $\mathbf{L}$ must be excluded from the distribution. As a result, using (2.26), the covariance of the nonlinear parameters can be written as:

$$\sigma^2\boldsymbol{\Gamma}_{\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}} = \sigma^2\mathbf{H.}(\boldsymbol{\eta})\boldsymbol{\Gamma}_{\hat{\boldsymbol{\eta}}\hat{\boldsymbol{\eta}}}\mathbf{H.}^T(\boldsymbol{\eta}), \tag{5.7}$$

$$\sigma^2\boldsymbol{\Gamma}_{\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}} = \sigma^2\mathbf{H.}(\boldsymbol{\eta})(\mathbf{L}^T\mathbf{L})^{-1}\mathbf{H.}^T(\boldsymbol{\eta}), \tag{5.8}$$

where $\mathbf{H}.(\boldsymbol{\eta})$ is the Jacobian of $\mathbf{h}(\boldsymbol{\eta})$ evaluated at the true parameter values. As with $\mathbf{F}.(\boldsymbol{\beta})$, the Jacobian is evaluated at $\hat{\boldsymbol{\eta}}$, resulting in a covariance estimate $\hat{\boldsymbol{\Gamma}}_{\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}}$. In this case, the Jacobian of $\mathbf{h}$ offers an easily expressed form:

$$\mathbf{H}.(\boldsymbol{\eta}) = \begin{pmatrix} -\frac{1}{\eta_z} & 0 & \frac{\eta_x}{\eta_z^2} \\ 0 & -\frac{1}{\eta_z} & \frac{\eta_y}{\eta_z^2} \\ \frac{\eta_x}{\sqrt{\eta_x^2+\eta_y^2+\eta_z^2}} & \frac{\eta_y}{\sqrt{\eta_x^2+\eta_y^2+\eta_z^2}} & \frac{\eta_z}{\sqrt{\eta_x^2+\eta_y^2+\eta_z^2}} \end{pmatrix}. \tag{5.9}$$

Thus, using asymptotic theory, the error distributions of the nonlinear parameter estimates approximately follow:

$$\boldsymbol{\epsilon}_{\hat{\boldsymbol{\beta}}} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \hat{\boldsymbol{\Gamma}}_{\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}}). \tag{5.10}$$

The prevalence of η_z in the denominator of elements of $\mathbf{H}.(\boldsymbol{\eta})$ indicates that small values of η_z produce large values in the covariance matrix of the gradients and albedos. Translated to its physical meaning with regard to object normals, this means that gradients corresponding to near vertical areas of the object surface are more susceptible to image noise. Several authors have also recognized this fact [43, 46]. This observation also corresponds with Agrawal *et al*'s heuristic treatment of noise, where the authors regard p and q values with large magnitudes as more vulnerable to noise [45]. Yet, while having η_z in the denominator will often produce gradient terms of higher value, this correspondence is not guaranteed to be true every time. As well, since η_z incorporates albedo into its value, the expression in (5.9) reveals that regions of the object with low albedo are also prone to producing more severe gradient errors. This agrees with expected behaviour, as the effect of noise on low intensity pixels is greater than on higher intensity ones. Agrawal *et al*'s scheme fails to include this effect of low albedo.

The expression in (5.8) characterizes an additive and normally distributed model of gradient error. Since the relationship between the gradients and surface is linear, integrating the gradients into an ML surface estimate remains a linear least squares problem.

5.2 Estimating the Visual Surface

5.2.1 Maximum Likelihood Estimation

Estimating the gradient values and their covariance at every pixel location comprises the first step in depth map estimation. The gradients are then related to the surface through a simple relationship. First, since images are discrete by nature, matrix notation should be used to represent the sampled scalar fields of the previous section. Using matrix notation, the gradient estimates are then related to the actual surface values through the following expression:

$$\mathbf{P} = \mathbf{Z}_x + \mathbf{E}_p, \tag{5.11}$$

$$\mathbf{Q} = \mathbf{Z}_y + \mathbf{E}_q, \tag{5.12}$$

where $\mathbf{Z}_x$ and $\mathbf{Z}_y$ represent partial derivatives of the surface in the x and y directions respectively, and the distributions of the errors $\mathbf{E}_p$ and $\mathbf{E}_q$ follow the properties outlined in (5.10) (note that (5.10) includes albedo variance-covariance terms in its expression, but these are not needed for surface estimation).

The matrix notation for the gradient fields also reflects that in the discrete case, finite differences represent partial derivatives, meaning that surface estimation at specific locations are related to their neighbours' values. Consequently, the estimation of $\mathbf{Z}$ is not characterized by individual systems at each location, but is a large-scale problem simultaneously incorporating every location. But, to continue working with matrix algebra requires reordering the 2D coordinates of both the surface and gradient fields into vector form (for example using column-major ordering). Written formally, the following expression frames surface estimation into a linear regression problem:

$$\begin{pmatrix} \mathbf{p} \\ \mathbf{q} \end{pmatrix} = \begin{pmatrix} \mathbf{D}_x \\ \mathbf{D}_y \end{pmatrix} \mathbf{z} + \boldsymbol{\epsilon}, \tag{5.13}$$

where $\mathbf{D}_x$ and $\mathbf{D}_y$ are finite difference operators in the x and y directions respectively. As well, $\boldsymbol{\epsilon}$ represents the vector of all noise terms $(\boldsymbol{\epsilon}_{\mathbf{p}}^T, \boldsymbol{\epsilon}_{\mathbf{q}}^T)^T$. Here, the parameter $\mathbf{z}$, the gradient estimates, and the error terms have been expressed as vectors. The derivative operators, $\mathbf{D}_x$ and $\mathbf{D}_y$, can take on forms corresponding to the desired order of derivative accuracy. The distribution of the errors incorporates every individual noise term, which all behave according to (5.10). The covariance of the noise is large and sparse, as the noise terms in $\mathbf{p}$ and $\mathbf{q}$ estimates are only cross-correlated if they share identical locations in the depth map. Written mathematically, $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}_{\boldsymbol{\epsilon\epsilon}})$, with $\boldsymbol{\Gamma}_{\boldsymbol{\epsilon\epsilon}}$ denoting the large and sparse covariance matrix.

As noted earlier, the addition of any constant to $\mathbf{z}$ will still produce the same gradient values. As a result, there is an inherent ambiguity in (5.13). Not mentioned however was a more insidious ambiguity, as different types of finite differencing schemes will introduce their own respective ambiguities [41]. Algebraically, these ambiguities can be treated as rank deficiencies. For this reason, (5.13) cannot be solved on its own. Nonetheless, including additional explicit constraints can usually resolve any ambiguities.

These constraints take on the form:

$$\mathbf{b} = \mathbf{R}\mathbf{z}. \tag{5.14}$$

As the ambiguities arising from integrating the two gradient fields are essentially related to determining the proper offset, constraints must restrict a sufficient number of surface locations to specific values. An easy to implement scheme simply involves restricting locations considered to be part of the background to zero. In those cases, the rows of **R** consist entirely of zeros except for a single entry corresponding to the background location, which is set to 1. Background locations can be determined by masking the original images. Failing that, another approach is to simply set all boundary locations to zero. Regardless of the manner in which the constraints are formed, they serve to augment (5.13) into a solvable system:

$$\begin{pmatrix} \mathbf{p} \\ \mathbf{q} \\ \mathbf{b} \end{pmatrix} = \begin{pmatrix} \mathbf{D}_x \\ \mathbf{D}_y \\ \mathbf{R} \end{pmatrix} \mathbf{z} + \boldsymbol{\epsilon}', \tag{5.15}$$

where $\boldsymbol{\epsilon}' = (\boldsymbol{\epsilon}^T, \mathbf{0}^T)^T$ denotes the augmented error terms. The covariance of the augmented noise, $\sigma^2 \boldsymbol{\Gamma}_{\boldsymbol{\epsilon}'\boldsymbol{\epsilon}'}$, is identical to the original noise covariance matrix, $\sigma^2 \boldsymbol{\Gamma}_{\boldsymbol{\epsilon}\boldsymbol{\epsilon}}$, except for the addition of zero terms corresponding to variance and covariances belonging to the constraints:

$$\boldsymbol{\Gamma}_{\boldsymbol{\epsilon}'\boldsymbol{\epsilon}'} = \begin{bmatrix} \boldsymbol{\Gamma}_{\boldsymbol{\epsilon}\boldsymbol{\epsilon}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}. \tag{5.16}$$

Constraining background or boundary pixels to zero height implicity assumes that the border regions of the object shape are also at zero height. However, the object may be raised from the supporting surface by shape features not visible from the camera viewpoint. The height of the object border relative to the supporting surface may not be uniform. As a result, border regions may posses differing relative depths from one another. As a result, it is important to note that constraining background pixels to zero may introduce errors in the estimated depth values of the border regions of the object.

Since (5.15) is a generalized least-squares problem, for small-scale depth maps, the dense GQR technique offers an excellent method in which to solve the overdetermined system. However, as mentioned in Sec. 2.3, there is no appropriate sparse and large-scale version of GQR. Until such an scheme is developed, alternative solutions must be used. Using the normal equations to solve the overdetermined system provides such an alternative. Returning to the original least-squares system in (5.13), which is not yet augmented by the constraints, the normal equations for surface integration are:

$$\begin{pmatrix} \mathbf{D}_x^T & \mathbf{D}_y^T \end{pmatrix} \boldsymbol{\Gamma}_{\boldsymbol{\epsilon\epsilon}}^{-1} \begin{pmatrix} \mathbf{p} \\ \mathbf{q} \end{pmatrix} = \begin{pmatrix} \mathbf{D}_x^T & \mathbf{D}_y^T \end{pmatrix} \boldsymbol{\Gamma}_{\boldsymbol{\epsilon\epsilon}}^{-1} \begin{pmatrix} \mathbf{D}_x \\ \mathbf{D}_y \end{pmatrix} \hat{\mathbf{z}}. \tag{5.17}$$

Note that (5.17) formulates a system of equations solving for the ML parameter *estimate*, $\hat{\mathbf{z}}$. As such, it does not describe the generative model; thus, there are no additive noise terms in the formulation.

As noise cross-correlation is confined to **p** and **q** terms sharing the same pixel locations, the noise covariance matrix is tridiagonal, meaning that the product of its inverse with another sparse matrix is easily computable [54] and will remain sparse. While (5.17) offers a convenient formulation of the normal equations, it can be simplified further. As Appendix B demonstrates, by assuming periodicity the transposes in (5.17) can be dropped, simplifying the expression to:

$$\begin{pmatrix} \mathbf{D}_x & \mathbf{D}_y \end{pmatrix} \boldsymbol{\Gamma}_{\boldsymbol{\epsilon\epsilon}}^{-1} \begin{pmatrix} \mathbf{p} \\ \mathbf{q} \end{pmatrix} = \begin{pmatrix} \mathbf{D}_x & \mathbf{D}_y \end{pmatrix} \boldsymbol{\Gamma}_{\boldsymbol{\epsilon\epsilon}}^{-1} \begin{pmatrix} \mathbf{D}_x \\ \mathbf{D}_y \end{pmatrix} \hat{\mathbf{z}}. \tag{5.18}$$

In an algebraic sense, $\boldsymbol{\Gamma}_{\boldsymbol{\epsilon\epsilon}}^{-1}$ corresponds to a weighting scheme, conforming with Agrawal *et al*'s treatment of noise reduction based on heuristic weighting schemes [45]. In their work, Agrawal *et al* favoured a weighting scheme stemming from the anisotropic diffusion (AD) approach to image restoration [45]. This approach lacks theoretical justification. Moreover, while anisotropic diffusion has demonstrable benefits in image restoration, it is unclear whether it is an appropriate tactic for use in surface integration as the nonlinear effects of image noise on gradients are not considered. The other heuristic approaches developed in [45] suffer from similar issues.

Since the system in (5.18) is not fully-ranked, the constraints in (5.14) must be incorporated. But, unlike the regression model of (5.13), (5.18) represents a fully-determined system of equations, each corresponding to one particular element or location in the parameter estimate $\hat{\mathbf{z}}$. In addition to these equations, the constraint matrix **R** also supplies its own equations, restricting certain **z** locations to certain values in **b**. To incorporate these constraints, equations in (5.18) corresponding to constrained $\hat{\mathbf{z}}$ locations should simply be replaced by their respective constraint equations in (5.14), e.g. at the boundary.

With the constraints incorporated, (5.18) presents an ML scheme for estimating the surface. Entitled ML Surface Estimation, the accuracy of the formulation depends in part on the finite-differencing scheme used. However, more accurate finite-differencing reduces the system sparsity, thereby increasing computational complexity and memory requirements. Thus, the cost and benefits of increasing accuracy must be considered when deciding on the differencing scheme. For the purposes of this dissertation, the ML Surface Estimation method employs centered-differencing with second-order accuracy:

$$\mathbf{Z}_x(x,y) = \frac{\mathbf{Z}(x+1,y) - \mathbf{Z}(x-1,y)}{2}, \tag{5.19}$$

$$\mathbf{Z}_y(x,y) = \frac{\mathbf{Z}(x,y+1) - \mathbf{Z}(x,y-1)}{2}. \tag{5.20}$$

Centered-differencing schemes are denoted $\mathbf{D}_x^c$ and $\mathbf{D}_y^c$ for the x and y directions respectively.

5.2.2 Modified Maximum Likelihood Estimation

While ML Surface Estimation provides a powerful scheme for surface integration, it is an estimation based on finite-differencing approximations to partial derivatives. The effects of finite-differencing are sometimes subtle. For instance, consider the normal equations of (5.18) under the presence of IID *gradient* noise, the implicit assumption of earlier works [68–70]. As IID noise reduces $\mathbf{\Gamma}_{\boldsymbol{\epsilon\epsilon}}^{-1}$ to an identity matrix, (5.18) simplifies to:

$$\mathbf{D}_x^c\mathbf{p} + \mathbf{D}_y^c\mathbf{q} = ({\mathbf{D}_x^c}^2 + {\mathbf{D}_y^c}^2)\hat{\mathbf{z}}. \tag{5.21}$$

Since $({\mathbf{D}_x^c}^2 + {\mathbf{D}_y^c}^2)$ is a discrete approximation to the Laplacian operator, (5.21) is simply a discrete approximation to the following continuous Poisson equation:

$$\text{div}(\mathbf{p},\mathbf{q}) = \nabla^2\mathbf{z}, \tag{5.22}$$

where ∇^2 and div(.,.) denote the divergence and Laplacian operator respectively:

$$\text{div}(\mathbf{p},\mathbf{q}) = \frac{\partial\mathbf{p}}{\partial x} + \frac{\partial\mathbf{q}}{\partial y}, \tag{5.23}$$

$$\nabla^2\mathbf{z} = \frac{\partial\mathbf{z}^2}{\partial^2 x} + \frac{\partial\mathbf{z}^2}{\partial^2 y}. \tag{5.24}$$

Several other authors have arrived at the same formula as (5.22), but they approach the problem using variational calculus to minimize a continuous energy functional analogue of the least-squares cost [44, 45, 68, 70]. In fact, when gradient noise is IID, or when noise is simply not accounted for, (5.22) corresponds to the Euler-Lagrange equation describing the conditions by which the continuous least-squares functional is minimized (please see Agrawal *et al* for an excellent derivation [45]). As well, Agrawal *et al* [45] demonstrated the basic equivalence between the variational approach of shape integration, and that of approaches using some type of orthogonal basis, such as [43, 69, 72]. However, the correspondence between direct least-squares minimization and the variational approach has never been sufficiently demonstrated.

Unfortunately, by employing $(\mathbf{D}_x^{c\,2} + \mathbf{D}_y^{c\,2})$ to represent the Laplacian, the normal equations using direct least-squares minimization results in a second derivative approximation spanning twice the step size. This produces the following filter:

$$\begin{pmatrix} 0 & 0 & 0.25 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0.25 & 0 & -1 & 0 & 0.25 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.25 & 0 & 0 \end{pmatrix}. \tag{5.25}$$

Consequently, replacing $(\mathbf{D}_x^{c\,2} + \mathbf{D}_y^{c\,2})$ with a superior finite-differencing scheme will produce a better approximation of the continuous Poisson equation of (5.22). For example, an alternative is to construct a finite differencing scheme off of the following popular filter mask:

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{pmatrix}. \tag{5.26}$$

Written formally, this alternative formulation can be expressed as:

$$\mathbf{D}_x^c \mathbf{p} + \mathbf{D}_y^c \mathbf{q} = (\mathbf{D}_{2x}^c + \mathbf{D}_{2y}^c)\hat{\mathbf{z}}, \tag{5.27}$$

where $(\mathbf{D}_{2x}^c + \mathbf{D}_{2y}^c)$ represents the sparse Laplacian operator using the finite differencing scheme in (5.26). Since replacing $(\mathbf{D}_x^{c\,2} + \mathbf{D}_y^{c\,2})$ with $(\mathbf{D}_{2x}^c + \mathbf{D}_{2y}^c)$ only modifies the right-hand side of the normal equations, (5.27) must be formulated explicitly, meaning that the system is no longer strictly equivalent to the original linear regression system in (5.13).

Yet, when noise is not IID, the normal equations present a more complicated expression:

$$\begin{pmatrix} \mathbf{D}_x^c & \mathbf{D}_y^c \end{pmatrix} \boldsymbol{\Gamma}_{\boldsymbol{\epsilon\epsilon}}^{-1} \begin{pmatrix} \mathbf{p} \\ \mathbf{q} \end{pmatrix} = \begin{pmatrix} \mathbf{D}_x^c & \mathbf{D}_y^c \end{pmatrix} \boldsymbol{\Gamma}_{\boldsymbol{\epsilon\epsilon}}^{-1} \begin{pmatrix} \mathbf{D}_x^c \\ \mathbf{D}_y^c \end{pmatrix} \hat{\mathbf{z}}. \tag{5.28}$$

The inclusion of $\boldsymbol{\Gamma}_{\boldsymbol{\epsilon\epsilon}}^{-1}$ prevents trivial substitutions of more accurate second derivative finite-differencing schemes into the right-hand side of (5.28).

One approach abandons centered-differencing completely, and simply uses forward-differencing in the original linear regression model of (5.13):

$$\left(\begin{array}{c}\mathbf{p}\\ \mathbf{q}\end{array}\right)=\left(\begin{array}{c}\mathbf{D}_x^f\\ \mathbf{D}_y^f\end{array}\right)\mathbf{z}+\boldsymbol{\epsilon}. \tag{5.29}$$

where $\mathbf{D}_{(.)}^f$ denotes forward finite-differencing. As Appendix B demonstrates, if ones assumes periodicity in the surface, $(\mathbf{D}_{(.)}^f)^T=-\mathbf{D}_{(.)}^b$, where $\mathbf{D}_{(.)}^b$ denotes backward finite-differencing. This reduces the normal equations derived from (5.29) to a mixture of forward and backward-differencing schemes:

$$\left(\begin{array}{cc}\mathbf{D}_x^b & \mathbf{D}_y^b\end{array}\right)\boldsymbol{\Gamma}_{\boldsymbol{\epsilon\epsilon}}^{-1}\left(\begin{array}{c}\mathbf{p}\\ \mathbf{q}\end{array}\right)=\left(\begin{array}{cc}\mathbf{D}_x^b & \mathbf{D}_y^b\end{array}\right)\boldsymbol{\Gamma}_{\boldsymbol{\epsilon\epsilon}}^{-1}\left(\begin{array}{c}\mathbf{D}_x^f\\ \mathbf{D}_y^f\end{array}\right)\hat{\mathbf{z}}. \tag{5.30}$$

Although not justified this way, this result is equivalent to the approach used by Agrawal *et al* in their heuristic-based weighting [45]. The appeal of (5.30) lies in that should gradient noise be IID, the formulation reduces to:

$$\mathbf{D}_x^b\mathbf{p}+\mathbf{D}_y^b\mathbf{q}=(\mathbf{D}_{2x}^c+\mathbf{D}_{2y}^c)\hat{\mathbf{z}}, \tag{5.31}$$

which incorporates the desired finite-differencing form of the second derivative operator. This result follows from the well-known fact that efficient centered-difference approximations to second derivatives are equal to the successive application of forward and backward approximations to first derivatives [73].

Similarly, substituting backward-differencing instead of forward-differencing into (5.29) results in an equally valid set of normal equations:

$$\left(\begin{array}{cc}\mathbf{D}_x^f & \mathbf{D}_y^f\end{array}\right)\boldsymbol{\Gamma}_{\boldsymbol{\epsilon\epsilon}}^{-1}\left(\begin{array}{c}\mathbf{p}\\ \mathbf{q}\end{array}\right)=\left(\begin{array}{cc}\mathbf{D}_x^f & \mathbf{D}_y^f\end{array}\right)\boldsymbol{\Gamma}_{\boldsymbol{\epsilon\epsilon}}^{-1}\left(\begin{array}{c}\mathbf{D}_x^b\\ \mathbf{D}_y^b\end{array}\right)\hat{\mathbf{z}}. \tag{5.32}$$

Unfortunately, weighting schemes (5.30) and (5.32) use an undesirable differencing scheme in their left-hand sides, namely backward-differencing or forward-differencing, which leads to the discrepancy between (5.27) and (5.31). To realize centered-differencing on the left-hand side of (5.31) one cannot simply replace $\mathbf{D}_{(.)}^f$ or $\mathbf{D}_{(.)}^b$ by $\mathbf{D}_{(.)}^c$ on the left-hand side of (5.30) or (5.32) as that will break the symmetry in the manner in which weights are treated on both sides of the equation. Fortunately, a solution exists to this problem.

A useful way to view centered-differencing is as an average of forward and backward differencing. Written formally, this is expressed as:

$$\mathbf{D}_x^c = \frac{\mathbf{D}_x^f + \mathbf{D}_x^b}{2}, \tag{5.33}$$

$$\mathbf{D}_y^c = \frac{\mathbf{D}_y^f + \mathbf{D}_y^b}{2}. \tag{5.34}$$

Using this concept, instead of averaging at the outset, one can execute averaging after the two respective normal equations produced by forward and backward-differencing have been formulated. The benefit of averaging the two normal equations of (5.30) and (5.32) together is that it constructs a centered-difference scheme on the left-hand side of the equation. This results in the following set of equations:

$$\begin{pmatrix} \mathbf{D}_x^c & \mathbf{D}_y^c \end{pmatrix} \boldsymbol{\Gamma}_{\epsilon\epsilon}^{-1} \begin{pmatrix} \mathbf{p} \\ \mathbf{q} \end{pmatrix} = \frac{1}{2}\left[\begin{pmatrix} \mathbf{D}_x^f & \mathbf{D}_y^f \end{pmatrix} \boldsymbol{\Gamma}_{\epsilon\epsilon}^{-1} \begin{pmatrix} \mathbf{D}_x^b \\ \mathbf{D}_y^b \end{pmatrix} + \begin{pmatrix} \mathbf{D}_x^b & \mathbf{D}_y^b \end{pmatrix} \boldsymbol{\Gamma}_{\epsilon\epsilon}^{-1} \begin{pmatrix} \mathbf{D}_x^f \\ \mathbf{D}_y^f \end{pmatrix} \right] \hat{\mathbf{z}}. \tag{5.35}$$

When gradient noise is IID, (5.35) reduces to the desirable formulation of (5.27), which incorporates a Poisson equation using centered-differencing on the left-hand side of the system.

The system of equations in (5.35) forms the core of an alternative integration scheme entitled Modified-ML Surface Estimation. As with ML Surface Estimation, in order to produce a solution, constraints must be incorporated into (5.35). The benefits of using Modified-ML Estimation over the original ML Estimation scheme depend in a large part on the severity of errors arising from using $(\mathbf{D}_x^{c\,2} + \mathbf{D}_y^{c\,2})$ to approximate the Laplacian operator. As a general rule, smaller-scale depth maps will suffer more from approximation errors than large-scale depth maps.

5.3 Experiments

Several experiments were performed to observe the performance of the ML-based methods compared to competing methods from the literature. All experiments and tests were performed in the MATLAB environment.

5.3.1 Estimating Known Surfaces

Experiments tested the merits of ML Surface Estimation vs. Modified-ML Surface Estimation, and also against the following leading surface integration methods:

1. FC - The Frankot-Chellappa method [69], a classic method based on using Fourier basis functions, often used as a base of comparison for other methods, e.g. in [45, 72]. These experiments use the implementation provided by Agrawal *et al* in the code accompanying [45].

2. Poisson - The direct solution to the unweighted Poisson equation of (5.27), first implemented in [70] using discrete-cosine transforms; however, these experiments use an implementation based on sparse matrix algebra.

3. AD - The weighted Poisson formulation based on Anisotropic Diffusion, which Agrawal *et al* showed is superior to FC, Poisson, and on par if not better than other weighting schemes in reconstructing test surfaces using images corrupted by noise [45].

Together, these alternative integration schemes are grouped together under the term literature methods. All methods, except for FC, constrain boundary pixels to the true surface values. As FC uses Fourier basis functions, there is no clear way to incorporate spatial boundary constraints. However, FC does require a constraint on the average surface height (the DC term in Fourier space) [69]. The implementation used in these experiments constrains average surface values to 0.

Throughout this discussion three synthetic surfaces will serve to test depth map extraction routines. Figs 5.1 illustrates the three surfaces along with noiseless and noisy sample images. The Mozart depth map is a surface commonly used in the literature, and was also used as a test surface in [45]. The Shark depth map presents a more challenging surface, particularly at the fin regions whose thin and steep slopes are very susceptible to image noise. Generating images for these two surfaces requires calculating surface normals at each pixel value. Surface normal values were calculated by first computing gradient values using centered-differencing approximations. The expression in (2.4) can then convert gradient values to surface normals.

Although the Mozart and Shark depth maps are both interesting test surfaces, a problem with relying on finite-differencing to produce “true” surface normal values is that these true values are approximations themselves. As well, estimation schemes that incorporate the same finite-differencing schemes in their formulation, such as ML Estimation, will have an unfair advantage. For this reason, it is important to test depth maps with analytical gradients. To fulfill this need, experiments also tested a third depth map—a hyperbolic paraboloid (HP). The following simple function con-

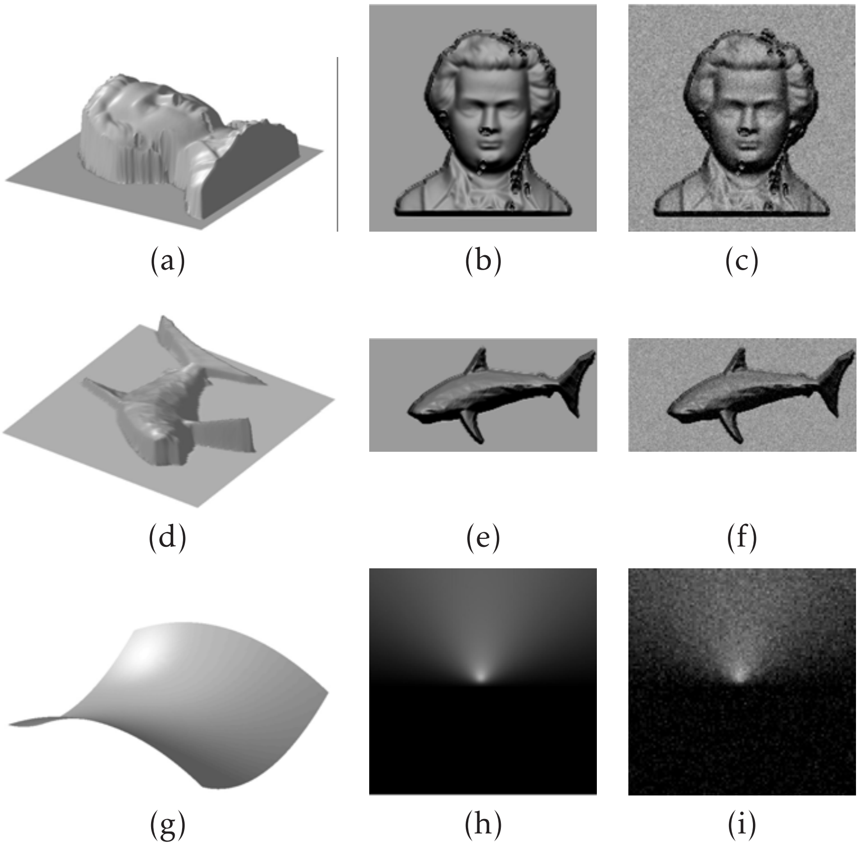


Figure 5.1: Synthetic surfaces and sample images. (a)-(c) the Mozart depth map; (d)-(f) the Shark depth map; (g)-(i) the Hyperbolic-Paraboloid surface. The first column displays the synthetic surfaces. The second and third columns display sample images illuminated from an elevation of 30° and from an azimuth angle of 90°; however, the third column's images are corrupted by zero-mean Gaussian noise at 5% standard deviation.

structs an HP:

$$Z(x,y) = \frac{x^2}{a^2} - \frac{y^2}{b^2}. \tag{5.36}$$

Analytical derivatives, and hence normals, are easily computed at each pixel location. These experiments used a value of 2 for both a and b, with the origin resting in the middle of the image.

Images were generated directly from the depth maps using the image formation equation of (2.1) with universal albedos of 0.7. Sequences were illuminated using the same conditions as the computer-aided identification system setup, meaning at angles of elevation of 30° and varying azimuth angles ranging from 0° to 340° at increments of 20°. As a result, each sequence consisted of 18 separate images. Pixels corresponding to object normals facing away from the light source were set to 0.

To test the capabilities of each integration method under noisy conditions, normally distributed error terms were added to the images. Standard deviation values of the stochastic noise were either 0 (no noise), 0.1%, 0.3%, 0.5%, 1%, 2.5%, 5%, 7.5%, and 10% of the full scale, which is always 1 for this dataset. The noisy images were then used to estimate the gradient fields, **P** and **Q**, and also their covariance using (5.8). These estimates then served as input into the five surface integration methods. To gain a clear picture of the stochastic performance of each method, this process was repeated 50 times for each noise level, where each iteration possessed its own realization of the stochastic noise.

Similarity between the output of each integration method and the true surface provides a measure of each scheme's performance. Despite representing different types of information, the data in depth maps and images are structured the same way. As a result, measures of image similarity are equally appropriate to apply to depth maps. Removing the effects of both scale and offset from influencing the result, correlation, as formulated in Sec. 1.4.1.4, serves as an excellent similarity measure. Similarity scores of each method under each noise level for each of the 50 iterations were recorded.

Fig. 5.2 illustrates the median similarity scores, along with the first and third quartiles, for each method when applied to the Mozart and Shark depth maps. As the figure demonstrates, when no noise is present, the ML method perfectly reconstructs both surfaces. Additionally, the methods from the literature all reconstruct a surface very close to the original one. As well, while the Modified-ML reconstructs a surface very similar to the true surfaces, its performance reconstructing the Mozart surface at low noise levels is slightly worse than all other methods.

The addition of noise causes the similarity scores of the literature methods to rapidly fall off. Specifically, the Poisson and FC methods are particularly susceptible to image noise, producing very low similarity scores relative to the other methods. The AD method suffered least of all the literature methods. As both Poisson and FC integration are unweighted formulations, the difference in results between the two unweighted formulations and the AD method support Agrawal *et al*'s conclusion that using weights mitigates noise effects [45]. Even so, AD exhibited considerable variability and struggled to handle high noise levels. In contrast, similarity scores of the ML and Modified-ML methods are stable across all noise levels and much higher, indicating that the ML weighting scheme is more effective than Agrawal *et al*'s heuristic weights. In fact, although there is variability, the ML-based methods are so stable that their quartile similarity scores are indistinguishable from the median values in the graph.

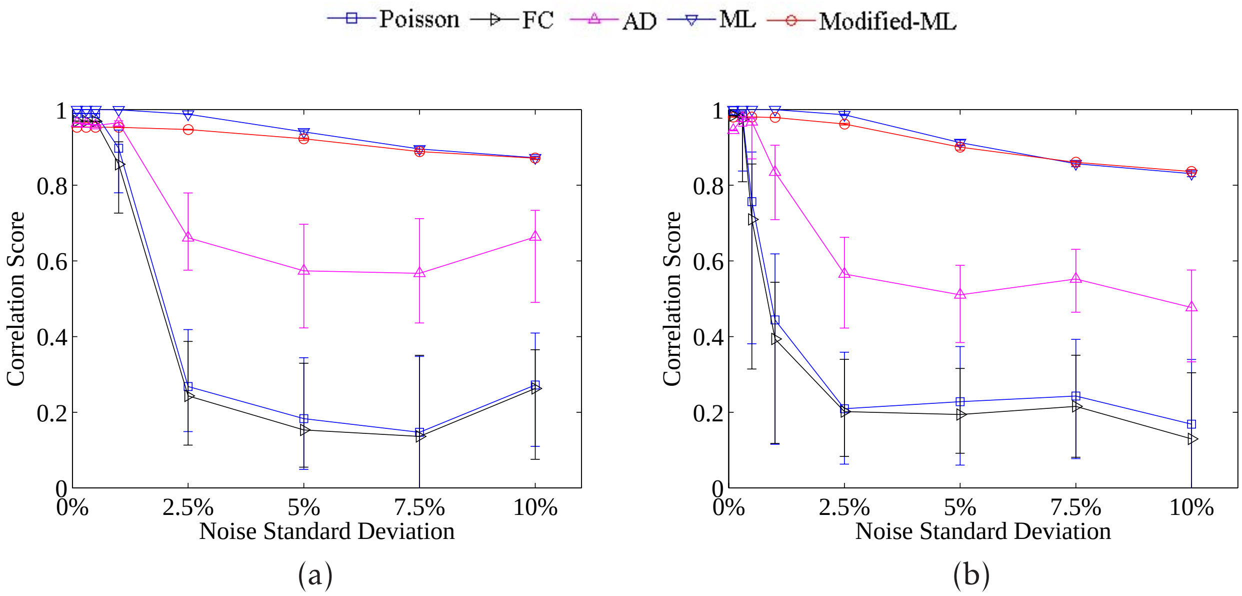


Figure 5.2: Mozart and Shark depth map similarity scores. (a) similarity results for the Mozart depth map; (b) similarity results for the Shark depth map. Both graphs display median scores with error bars representing the first and third quartiles. For both depth maps, the ML Method provides perfect reconstructions at 0% noise and almost perfect similarity results at 1% noise. As noise increases, the similarity scores of the ML and Modified-ML begin to converge and remain high. In contrast, similarity scores of the other methods fall off rapidly.

The effectiveness of the ML and Modified-ML methods in handling noise is also reflected in the visual quality of the reconstructions. As Figs. 5.3 and 5.4 illustrate, reconstructions using the literature methods are so corrupted as to become essentially unusable at noise levels of 1% or higher. On the other hand, both ML and Modified-ML are extraordinarily robust to noise, producing excellent reconstruction results even in the presence of 10% noise. It should be noted that in Agrawal *et al*'s tests on the Mozart surface in [45], the authors depicted superior visual results than Fig. 5.3 for all three literature methods, even under the presence of considerable noise. However, the experiments of this thesis were unable to replicate Agrawal *et al*'s results.

Comparing the two ML-based methods for both the Mozart and Shark surfaces, ML outperforms the Modified-ML at low noise levels. However, at higher noise levels the two methods converge, with similarity scores actually favouring Modified-ML over ML. As well, while perhaps not visually apparent in the printouts of Figs. 5.3 and 5.4, the large-scale Mozart and Shark ML surface estimates exhibit visually noticeable roughness at

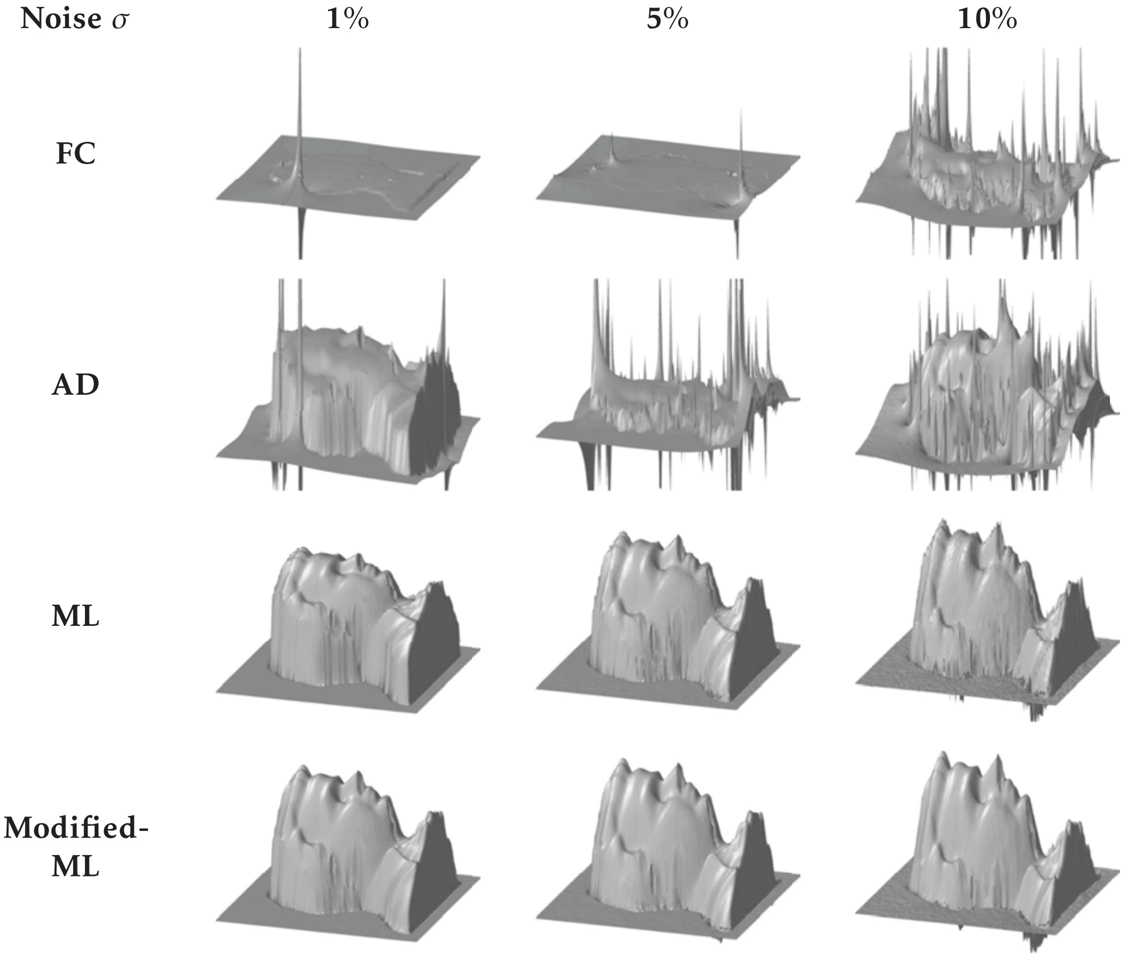


Figure 5.3: Mozart surface reconstructions. These results present typical samples reconstructions of the Mozart test surface using images corrupted by 1%, 5%, and 10% zero-mean Gaussian noise. Results of the FC, AD, ML, and Modified-ML methods are displayed. Results for the Poisson method are very similar to those of the FC method.

high noise-levels. This decline in the ML method's performance can be attributed to its relative weakness in approximating the Laplacian operator vs. the Modified-ML method.

In fact, for smaller scale depth maps, the severity of these approximation errors are much greater. For instance, consider Fig. 5.5, which demonstrates reconstruction results using a smaller-scale version of the Mozart depth map. As the figure illustrates, at 5% and 10% noise levels, the ML surface estimates are very rough and choppy, indicating that the lower accuracy approximations inherent in the ML method have a more significant impact for smaller scales.

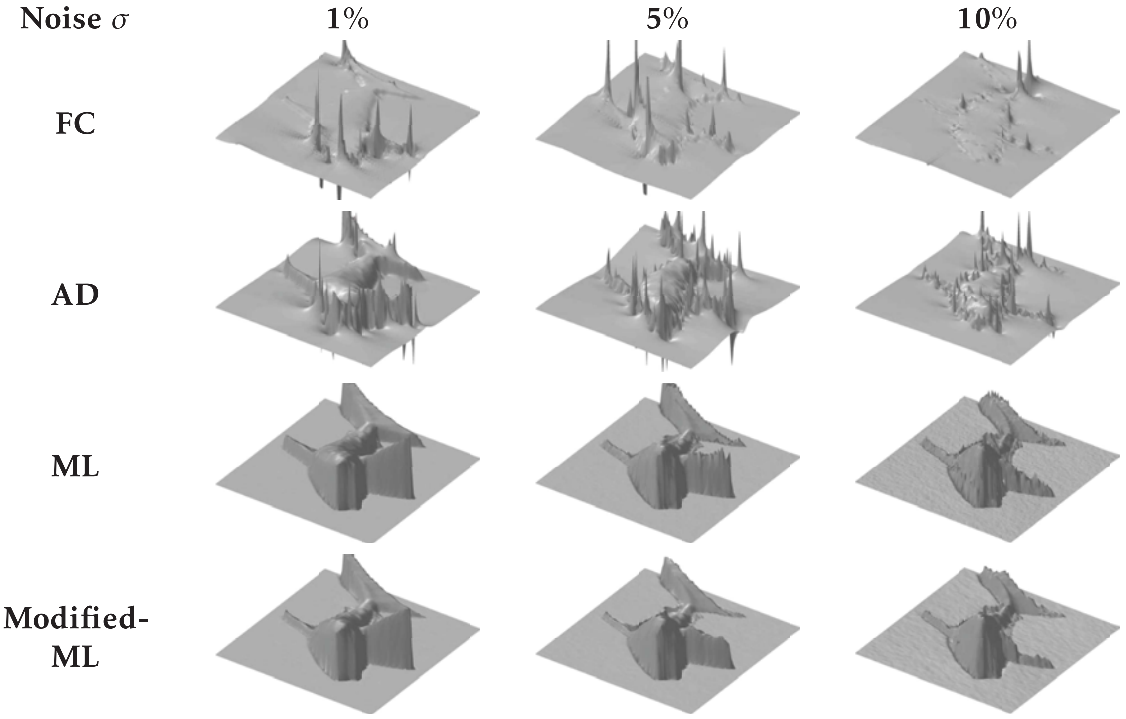


Figure 5.4: Shark surface reconstructions. These results present typical samples reconstructions of the Shark test surface using images corrupted by 1%, 5%, and 10% zero-mean Gaussian noise. Results of the ML, Modified-ML, Frankot-Chellappa (FC), and AD methods are displayed. Results for the Poisson method are very similar to those of the FC method.

While the Modified-ML method clearly outperforms the ML method at high noise levels, the correlation results suggest that the ML method is superior to the Modified-ML method at zero to low noise. However, the true surface normals for both the Mozart and Shark depth maps were estimated using the same centered-differencing scheme incorporated into the ML method. As a result, when testing performance using the Mozart and Shark depth maps, the ML method enjoys an unfair advantage over the Modified-ML method. For this reason, it is instructive to consider the performance of both methods on the HP surface, which permits analytical formulations of its surface normals. Unlike the first two surfaces, the HP depth map and its accompanying gradients provides appropriate conditions in which to judge the merits of both ML-based methods with equal footing. Fig. 5.6 graphs similarity scores of all of the tested integration schemes on the HP surface.

As the figure demonstrates, at zero noise both the ML and Modified-ML methods perfectly reconstruct the surface. This differs from the previ-

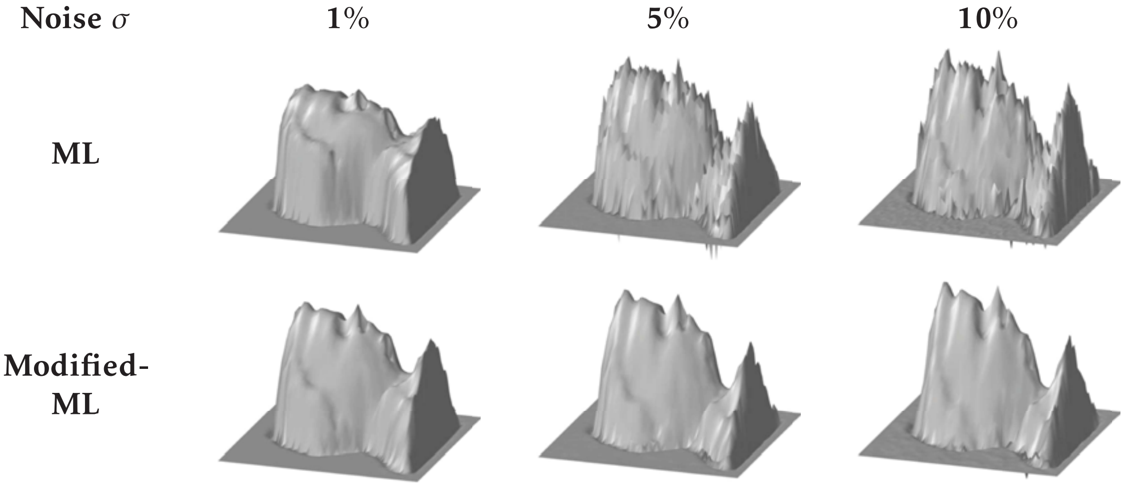


Figure 5.5: Small-scale Mozart surface reconstructions. These results present typical small-scale sample reconstructions of the Mozart test surface using images corrupted by 1%, 5%, and 10% zero-mean Gaussian noise. Results of the ML and Modified-ML methods are displayed.

ous surface results, where Modified-ML reconstruction could not match ML Surface Estimation's performance at zero to low noise. As the HP surface is a second-order surface, and the finite-differencing schemes for both methods are accurate to the same order, this is not surprising. When greater levels of noise are introduced, both ML-based methods performed extremely well, scoring very close to 1 at all noise levels.

Additionally, similar to the previous results, the Poisson and FC methods perform well at zero to low noise, but suffer greatly under the presence of higher image noise values. In contrast, the AD method scores well across all noise levels. These similarity scores suggest that AD, ML, and Modified-ML surface estimation share very similar performance. Yet as Fig. 5.7 demonstrates, despite this apparent similarity, the quality of visual results of these methods are quite different.

The poor visual quality of AD surface reconstruction is perhaps the most immediate observation pulled from Fig. 5.7. As a result, despite its good similarity scores and a good underlying idea, AD can be discounted because of problems with heuristic weighting. Comparing the two remaining sets of example surfaces, the ML method reconstructs a much rougher surface under the presence of noise. At higher noise levels, the roughness of the ML surfaces is even more pronounced. Thus, as with the first two test surfaces, Modified-ML provides superior reconstructions to ML estimation at high noise levels. But, differing from the previous tests, Modified-ML

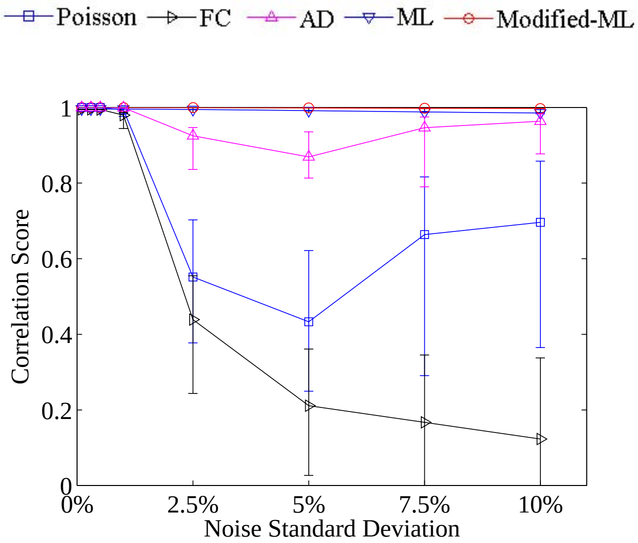


Figure 5.6: Hyperbolic-paraboloid depth map similarity scores. ML and Modified-ML methods provide perfect reconstructions at 0% noise and almost perfect similarity results at 1% noise. As noise increases, similarity scores of the ML and Modified-ML remain close to 1, with the AD method also exhibiting good correlation performance. In contrast, similarity scores of the other methods fall off rapidly.

is as proficient as ML estimation at reconstructing the HP surface at zero noise. In addition to demonstrating the importance of employing analytical surfaces in experimentation, these results justify choosing Modified-ML over ML Surface Estimation.

5.3.2 Estimating Microfossil Surfaces

Although the Modified-ML technique proved very capable of reconstructing the synthetic surfaces, its abilities in reconstructing microfossil shapes is also an important consideration. Ultimately, with a powerful enough reconstruction scheme, estimated microfossil surfaces can serve as template representations and even taxonomic tools. Fig. 5.8 displays typical reconstructions using Modified-ML Surface Estimation on the foram dataset.

Reconstruction successfully captured the broad and large-scale features of the microfossils. However, finer details were lost. Additionally, the morphology of the reconstructed shapes do not exhibit the sharp peaks and other near-discontinuities of actual microfossil shapes. Although, the integration process is certainly responsible for part of this detail loss, it is not the only factor preventing better reconstructions. For instance, latent

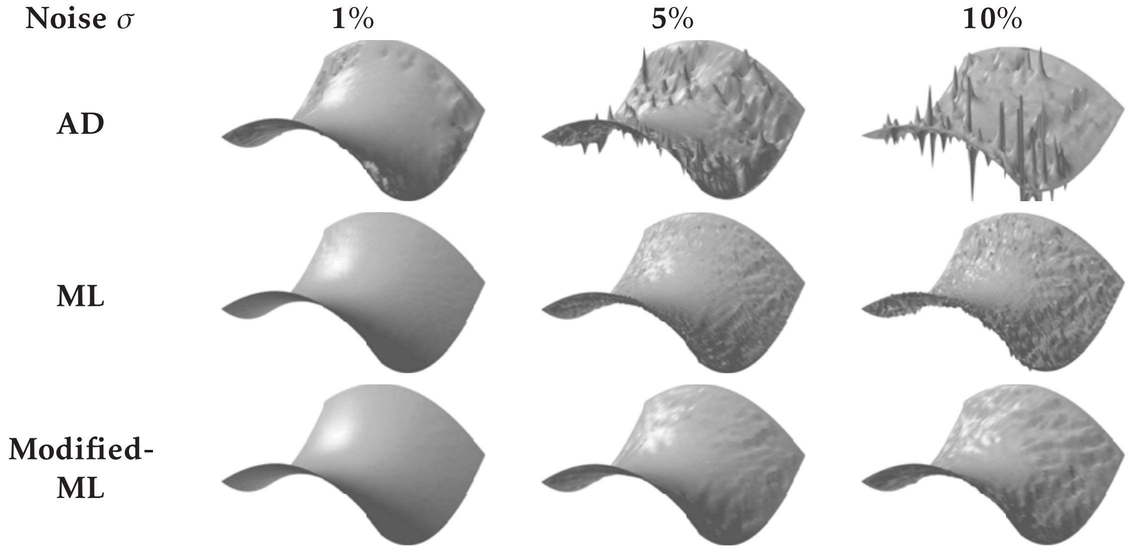


Figure 5.7: Hyperbolic-paraboloid reconstructions. These results present typical reconstructions of the HP test surface using images corrupted by 1%, 5%, and 10% zero-mean Gaussian noise. Results of the ML, Modified ML, and Diffusion methods are displayed.

misalignments not completely corrected by the image alignment routine in Chapter 4 may explain some of this detail loss. Additionally, blurriness due to limited depth of field resulting from microscope magnification also hinder fine detail reconstruction. As well, violations of the photometric stereo assumptions, including Lambertian reflectance, light source at infinity, orthographic cameras, and no shadows would also be responsible for reconstruction errors.

Nonetheless, the microfossil surface estimation results are encouraging as they demonstrate that shape features can be extracted. Additionally these results highlight areas of focus of future work, such as reconstructing micro-features and near-discontinuities. With further work, it should be feasible to provide highly detailed shape-based template representations. Modified-ML Surface Estimation provides an important contribution toward this goal.

5.4 Asymptotic Complexity

Executing either ML-based method requires constructing an expression for the inverse of the large and sparse matrix $\boldsymbol{\Gamma}_{\epsilon\epsilon}$. Yet, as correlation is confined to only gradient values sharing the same pixel, $\boldsymbol{\Gamma}_{\epsilon\epsilon}^{-1}$ can also be viewed as a weighting scheme for each pixel, where sets of weights are equiva-

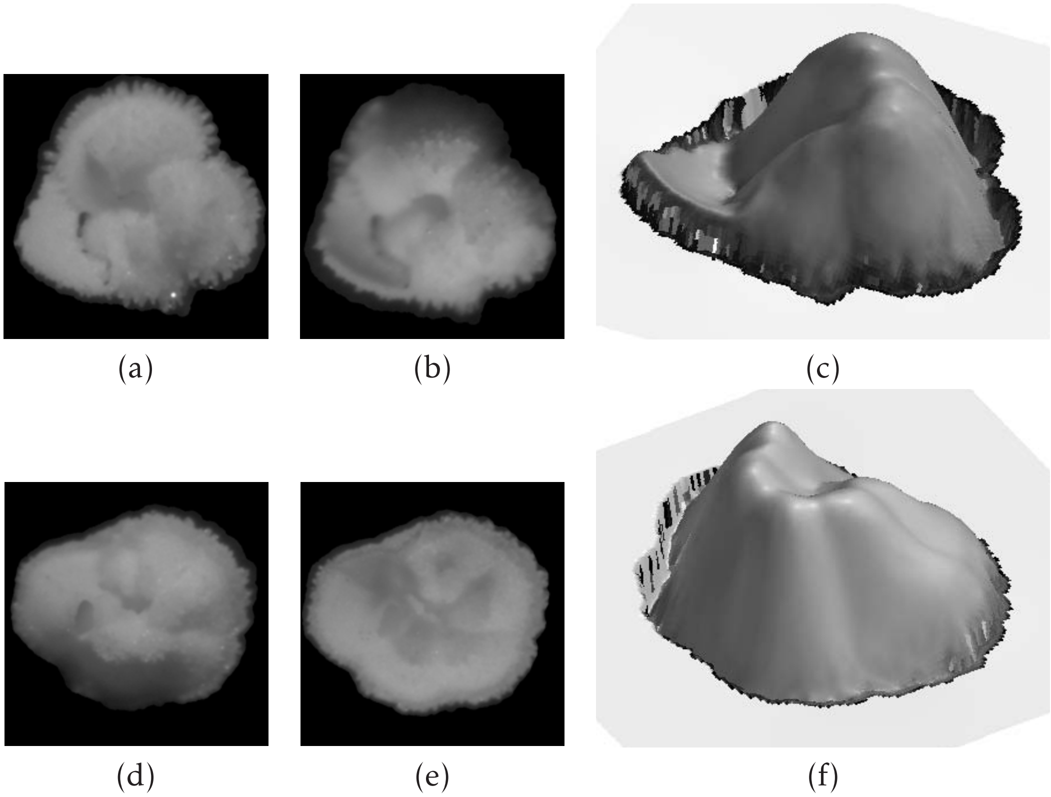


Figure 5.8: Surface reconstructions of microfossils. Illustrated in the first two columns are images from typical samples in the foram dataset. Images in both columns are illuminated from an elevation angle of 30°, with angles of azimuth of 90° and 270° respectively. The third column depicts surfaces estimated using the Modified-ML method.

lent to elements from the 3×3 matrix $\mathbf{\Gamma}^{-1}_{\hat{\beta}\hat{\beta}}$, whose values can be easily obtained using (5.6). Such a computation executes in constant time for one pixel. As a result, assuming images are $n \times n$ dimensions, the complexity of computing weights is $O(n^2)$ for all pixels. Computation requirements of the AD-based weights of Agrawal *et al* [45] possess equivalent complexity. Assuming sparse matrix libraries are used, memory requirements are also $O(n^2)$.

Whether weights are used or not, all methods except for FC, estimate $\mathbf{z}$ using a large and sparse system of the form $\mathbf{Ax} = \mathbf{b}$, e.g. (5.18) or (5.35). In addition to the fill factor of $\mathbf{A}$, the memory and complexity bounds of solving such a system depend on the sparsity pattern of the matrix and the capabilities of the sparse library used. As both $\mathbf{\Gamma}^{-1}_{\epsilon\epsilon}$ and finite differencing operators possess favourable sparsity patterns, $\mathbf{A}$ will also possess a useful sparsity pattern. For instance, $\mathbf{A}$ will have 7 and 9 diagonals respectively for Modified-ML and ML Surface Estimation. For Agrawal *et al*'s AD surface estimation method, its sparsity pattern is identical to that of Modified-ML estimation. On the other hand, solving the unweighted Poisson equation of (5.27) produces a pentadiagonal system. For all cases, the memory use of **A** is proportional to $O(n^2)$.

Efficiency of solving these cases depends on the sparse solver's ability to handle wide-banded symmetric systems. Assuming the complexity of each case is asymptotically proportional to the number of nonzeros in the system, solving each of the systems should take $O(n^2)$ time [54]. In reality, actual complexity will depend on the narrowness of the matrix bandwidths, the efficiency of the base sparse matrix operations, and the effectiveness of the fill-reducing ordering used [74]. As many of the ordering methods are heuristic in nature [74], complexity guarantees are difficult to provide. However, whether or not the sparse system takes $O(n^2)$ time or longer to solve, computing weights will not add to the asymptotic complexity of the algorithms. In the case of the FC method, its complexity is equivalent to that of the 2D Fast-Fourier Transform, which is $O(n^2 log(n^2))$ [29].

In terms of execution times of the implementations used for experimentation, all five methods scaled linearly with the total number of pixels. Unsurprisingly, the three weighted estimation techniques, AD, ML, and Modified-ML executed slower than the unweighted methods. Out of the three weighted techniques, AD executed noticeably faster than the ML methods. However, the ML implementations computed weights by looping through every pixel, whereas AD weights were computed using MATLAB vectorization. Thus, the differences in computation times can be attributed to the way in which the ML methods were programmed in MATLAB. As this is a characteristic specific to MATLAB, these slower times do not reflect differences in algorithmic complexity.

Comparing the two unweighted schemes, FC executed considerably faster than the Poisson method. Limitations in current sparse methods used to solve wide-banded systems are responsible for this difference. As direct solvers of sparse symmetric systems continue to improve, execution times of the Poisson, AD, and ML methods will continue to decrease.

5.5 Conclusion

While there are many existing surface reconstruction methods based on photometric stereo, current formulations either fail to properly model the effect of image noise or employ highly demanding nonlinear optimization techniques. To fill this gap, this chapter describe methods using a combination of nonlinear and linear regression to provide ML surface estimates.

Like many visual-surface reconstruction techniques, ML estimation consists of two main steps. The first step computes gradient estimates for every pixel location. However, differing from the state-of-the art, the ML techniques of this chapter model disturbances in gradient estimates as resulting from noise in image observations. Using nonlinear regression theory and asymptotic approximation, gradient errors are modelled as following a zero-mean Gaussian distribution that is correlated only amongst parameters sharing the same pixel location.

With the stochastic nature of gradient estimates modelled, the second step incorporated the estimates into a linear GLS formulation, producing an ML estimate of the surface. Together, both steps comprise ML Surface Estimation. However, due to inherent characteristics of finite-differencing approximations to derivatives, the ML formulation incoporates an ineffective a discrete approximation to the Laplacian operator. As a result of these approximation issues, ML Surface Estimation tends to produce rough and choppy depth maps under the presence of noise. Mitigating this effect, adjustments to the ML formulation result in superior approximations to the Laplacian. Entitled Modified-ML Surface Estimation, this formulation can be implemented using normal equations.

When tested against noisy images, Modified-ML Surface Estimation demonstrated excellent robustness to image noise. In fact, the technique produced usable surfaces even under the presence of very high levels of image noise. Contrasting these results, leading methods from the literature were unable to handle image noise robustly and reconstructed unusable surfaces under the presence of noise with 1% standard deviation or higher.

The Modified-ML method detailed in this chapter is the first visual-surface reconstruction technique able to reliably handle image noise and its propagation through all estimation steps in a readily computable way. While other techniques, such as [41, 46], do treat image noise in a rigorous ML manner, practical implementations are lacking or their applicability is limited.

When executing Modified-ML estimation, modelling noise does not introduce additional asymptomatic complexity. Thus, Modified-ML Surface Estimation offers both an effective and efficient reconstruction scheme. As noise is an unavoidable phenomena present in images from real cameras, Modified-ML estimation represents a significant advancement in the state of visual-surface reconstruction.

Although applicable in low-noise conditions, the benefits of Modified-ML become particulary clear in situations involving high noise intensities. Objects studied under a microscope, such as microfossils, provide a case in point. Due to the optics of magnification, depth of field is often very limited in microscopy. Decreasing the aperture is an approach that could be used to offset this effect. However, this approach also decreases the amount of light transmitted to the image sensor, increasing the relative magnitude of noise. As a result, the Modified-ML technique is particularly relevant to visual-surface estimation of microfossil and other microscopic objects.

Finally, while the uniform zero-mean Gaussian assumption for image noise is a well-accepted and frequently-used model, situations exist where different noise distribution assumptions are more appropriate. In these cases, ML estimates no longer correspond to least-squares computations. Even so, although the covariance computations detailed in this chapter would no longer be valid, the ML framework introduced is equally applicable to alternative noise models. As work progresses on ML estimation techniques for noise distributions outside that of the Gaussian model, e.g. the work on least-absolute deviations applicable to Laplacian-distributed noise [75], the practical scope of the framework provided here will continue to expand.

Chapter 6

Conclusions

6.1 Summary

The continual advancement of the capabilities of computer engineering and computer science do not only benefit a single or even small set of fields. In actuality, computer engineering consistently offers new opportunities for collaboration with a diverse range of research fields. Automation represents a common trait of these collaborations and the field of micropalaeontology provides a perfect case in point. Micropalaeontological work, in particular biostratigraphy, is crucial to many extraordinarily important enterprises—notably hydrocarbon exploration [3] and ongoing attempts to further understanding of current and prehistoric climates [2]. Detailed taxonomic requirements characterize much of micropalaeontological work, meaning that specimens must be classified and sorted. While microfossil samples can be readily collected in immensely large volumes, as evidenced by the work of the IODP, performing detailed identification and sorting of specimens remains a significant challenge in the field. Traditionally these tasks are performed manually by trained specialists using microscopes. The labour required is repetitive, tiring, and can also safely be described as tedious.

These difficulties hinder effective analysis of large scale microfossil datasets. Moreover, the correlation between inconsistency in specialist classifications and fatigue [9] introduces additional problems when identifying large volumes of samples. Thus, as it stands now, microfossil identification and sorting acts as a sort of bottleneck for many aspects of micropalaeontological research. As it has done with other fields, computer engineering has the potential to automate much of this micropalaeontological work. Focusing on the task of computerized identification of microfossils, this thesis explored how the field of computer vision can aid in solving this problem.

The approach to computerized identification of microfossils pursued in this thesis differs from the state-of-the-art. Coined as computer-aided classification, the approach clusters specimens together based solely on visual similarity, and not based on pre-existing taxonomic knowledge or training sets. By presenting representative samples or templates from each cluster to an expert for identification, the system effectively reduces the amount of required classifications specialists need to make. As discrimination is based only upon visual similarity, the choice of digital representation used in template identification and automatic clustering plays an extremely important role.

To evaluate the feasibility of this approach, the UA Electronic Imaging Lab developed a preliminary computer-aided classifier, of which the author was a contributor [24]. Image-based representations were used for both templates and automatic clustering. Although the study demonstrated the feasibility of computer-aided identification, higher performance is needed for a practical classification aid. An important published contribution of this thesis is the exploration of the effect of uncontrolled illumination on classification performance. In the context of automatic clustering, these limitations can be overcome by capturing multiple images of microfossils under different light directions.

However, as experts are unable to identify microfossils through their image-based representations, simply controlling for illumination on its own does not address issues related to template representation. As experts are unable to identify microfossils at the species level using images and have trouble doing so at the genus level [31], template representations must move beyond images. To increase template classification performance, alternative digital representations must incorporate more information and detail than images. Since the variation of shading and shadows across different light directions provide important indicators to an object's 3D shape, representations using and incorporating this information are a viable and powerful alternative to images. Controlling for and manipulating illumination direction is key to unlocking this information. Since the light source is fixed in typical microscope setups, including the one used for the thesis, controlling illumination necessitates the incorporation of a motorized x-y-phi stage.

Amalgamating image sequences of the microfossil under differing light directions into a video is an effective representation to present to experts. By allowing experts to observe the effects of changing light direction, these video-based representations reveal aspects of the microfossil's 3D morphology that are hidden using simple images.

While videos offer an improved representation than images, they are not the only option. Where humans can intuitively gain understanding of 3D morphology using light and reflectance, computer vision techniques can explicitly extract shaped-based representations. Specifically, photometric stereo can compute values of the surface normals and albedos for every pixel location. Using these surface normals, estimation techniques can construct a visual surface. Allowing the manipulation of viewpoint and illumination, 3D models offer potential to increase an expert's confidence and ability to reliably identify a specimen. Additionally, 3D models have aided in the understanding of organisms other than microfossils [36, 37], indicating that reliable 3D modelling could bring its own unique benefits to the field of micropalaeontology.

The unpublished contributions of this thesis revolve around efforts to employ these alternative template representations to computer-aided classification of microfossils. Computer vision plays a crucial role towards this goal. First, incorporating video-based and shape-based representations into a computer-aided classification scheme required key modifications to the existing system setup. These advances to the equipment and methodology of the existing computer-aided system comprise an important area of this thesis' contributions. Secondly, due to unavoidable inaccuracies in the motorized stage, image sequences of specimens exhibit significant misalignments. Corrupting both video and shape-based representations, these misalignments were addressed through an ML estimation formulation entitled photometric alignment. The third and final main contribution of this work focuses on extracting shape-based representations for templates. Still constituting an open problem, visual surface construction is a very active topic of research. This thesis addresses the topic of surface construction from normals under the presence of image noise using ML estimation. A discussion of the results and significance of all three unpublished contributions follow in the next three subsections.

6.1.1 Video and Shape-Based Representations

The preliminary computer-aided system described in [24] lacked the capabilities to produce and employ video and shape-based template representations. As such, incorporating these alternative representations requires significant modifications to the preliminary system. Controlling for illumination directions represents one of the most important extensions.

Satisfying the equipment requirements of such a goal, an x-y-phi motorized stage was incorporated into the system. Enhancements to the existing C++ image capture program, developed during the course of the preliminary system, enabled the control and manipulation of stage position. These advancements allowed the development of an automatic scheme in which to localize and image large batches of specimens at once.

Datasets captured with the extended system consist of sequences of images captured at successive azimuthal angles of illumination. As the light source remains fixed with respect to specimens' principal axes, manipulating light direction required rotating the physical particle under view and canonizing the resulting images. As the rotational element of the stage mounts above the linear elements, rotating the specimen also involves translation. Unfortunately, as the stage is not 100% accurate, the expected and actual locations of particles do not coincide. To mitigate these errors, particles were centered in the field of view using silhouette centroid calculations. However, despite this additional fine-tuning, discrepancies in the silhouettes introduced latent misalignments within image sequences. This problem impacts both video and shaped-based representations. Solving this problem required developing an image alignment routine based on ML estimation and photometric stereo. Described in Chapter 4, the alignment routine performs its corrections prior to any data analysis.

Apart from enhancing the system to support video and shape-based representations, extensions also included employing a faster and slightly more accurate clustering algorithm than the preliminary system's. In addition, as the online wiki serves as the desired channel to provide templates to experts, improvements made to the wiki to display videos instead of images constitutes another significant system extension. Further work is required to support online dissemination of shape-based representations.

With the system's capabilities extended, a dataset comprised of 500 foram specimens was collected. The dataset includes 18 images of each specimen, all illuminated at an angle of elevation of 30° and with azimuth angles increasing by increments of 20°. Using this dataset, methods from Chapter 5 extracted shape-based representations. When applied to the foram dataset, visual surfaces accounted for 74% of image variability. The 26% shortfall indicates that non-Lambertian illumination effects and information loss in estimating surfaces from weighted normals significantly affect the capabilities of shape-based representations to represent microfossil morphology.

To overcome this limitation, the chapter introduced a modification to the shape-based representations that applies texture maps based on video frames. Providing an identical representation to that of videos when the visual surface is viewed from above, the texture-mapped shape-based representations nonetheless allow experts to view microfossils at modest deviations from the camera viewpoint. A benefit of these texture-mapped surfaces is that they lead to the promising of producing anaglyph videos. Representations of this sort are useful, as anaglyph videos represent a shape-based representation that can provide a similar set of information as texture-mapped surfaces without any additional rendering requirements on the part of the online wiki. In addition, by providing insight into 3D geometry at a viewpoint straight above the microfossil, there is no need to view anaglyphs from different elevations. Thus, anaglyphs avoid the detail loss associated with texture-mapped shapes.

Supporting video and shape-based representations constitutes a significant departure from the state of the art of microfossil identification. Freeing computerized-classification from the inherent limitations of images, the ramifications of such an approach are potentially far-reaching. Representing another long-lasting contribution, the system extensions effected during the course of this thesis enables the collection of extraordinarily useful and rich datasets. As work on shape-based microfossil representations advances, the data collection techniques developed during this dissertation will continue to provide the requisite datasets.

6.1.2 Image Alignment Using Photometric Stereo

A key assumption of photometric stereo is a perfect correspondence between all images in the sequence. Normally, this is a trivial condition to satisfy, as typical image capture scenarios for photometric stereo fix the camera and object and manipulate the illumination direction. Nonetheless, situations do arise where perfect correspondence can no longer be guaranteed. In certain cases, system setups, like the one used in this thesis, are characterized by a fixed light direction; as a result, objects must be rotated to produce changing illumination conditions. These differences are enough to cause significant misalignments in the microfossil dataset collected during the course of this thesis.

Current alignment techniques are generally categorized into intensity or silhouette-based methods. Unfortunately, neither of these approaches are appropriate for image sets illuminated by different lighting conditions between images. Nonetheless, intensity-based alignment methods do possess key benefits; namely, they can be evaluated against a meaningful error metric. Additionally, such an error metric can often be considered as some form of function, providing a surface (albeit nonlinear) in which to search for image shifts. Taking these considerations in stride, a Lambertian generative model of image formation enables the use of image intensities when aligning photometric images. This is called photometric alignment.

By incorporating a model of image formation, photometric alignment simultaneously estimating the best weighted-normals and corrective shifts of the object images. However, weighted normals depend on the current shift estimates, meaning that in practice photometric alignment only need minimize a nonlinear function based on image shifts. As image noise is modelled as following a zero-mean IID Gaussian distribution, minimizing the sum-squared error of the nonlinear function is equivalent to determining the ML estimate. Representing a key added benefit, an implicit goal of photometric alignment is determining the corrective shifts producing the best set of weighted normals.

Several experiments were performed to measure photometric alignment's accuracy and reliability. Using artificially generated shifts, experiments determined that the performance of photometric alignment does not depend on a specific minimization method. Additionally, experiments also demonstrated that photometric alignment can successfully correct for much of the latent error within centroid-aligned images. Photometric-aligned image sequences also produce superior visual surface results, qualitatively illustrating the routine's benefits. These benefits were further demonstrated by applying photometric alignment to centroid-aligned image sequences in the microfossil dataset. After aligning all 500 image sequences, photometric alignment significantly improved agreement between the image sets and their reconstructed counterparts, providing an over 40% reduction in sum-squared error. These results demonstrate the worth of photometric alignment as a crucial component of the computer-aided microfossil classification system.

While image misalignment is an identified problem with the microfossil image capture system, the scope of the problem is certainly not constrained to this sole system or even application. During the course of researching potential motorized stages to purchase, it was found that most motorized stages place the rotational element on top of the translational elements, forcing a change to a particle's horizontal and vertical placement during a rotation. As a result, the equipment used for the computer-aided system are typical of most microscope setups. Moreover, photometric image misalignment can crop up in any situation outside of microscopy where a particle must be rotated instead of the light source.

6.1.3 Maximum Likelihood Surface Estimation

Constructing a surface or depth map from videos has the potential to provide extraordinarily beneficial visual representations. An accurate 3D model could provide a powerful digital template representation to present to users. Additionally, 3D models could also act as useful teaching and analysis tools. Although many popular visual-surface reconstruction methods date back from over 20 years ago [68, 69], efforts are ongoing to continue to improve integration methods. Dealing with image noise is one issue that many integration methods in recent years have focused on [41–46].

However, the state-of-art either fails to incorporate a correct model of image noise and its propagation through all steps of the integration process [44, 45], is only valid in limited circumstances [46], or employs nonlinear optimization techniques that have great difficulty in converging [41,42]. The work of this thesis differs from previous efforts by both employing ML estimation in all steps of the process and retaining the linear step of integrating gradients into visual surfaces.

First, the method applies asymptotic nonlinear regression principals to estimate the behaviour of the gradient fields, obtaining an ML estimate of the gradients and their errors. With a model of the stochastic behaviour of gradients in hand, estimating the surface from the gradients is kept to a generalized *linear* least-squares regression problem.

Using finite-difference versions of the derivative operators, estimating the surface from the gradients involves a large and sparse linear system with constraints. Entitled ML Surface Estimation, the method proved extremely robust to image noise, reconstructing good results even in the presence of significant amounts of noise. However, the results suffer from roughness or choppiness under the presence of noise, affecting visual quality. The source of this problem lies with the discrete version of differentiation used in the system. If one assumes periodicity in the surface and IID noise in gradient fields, the normal equations incorporate a Laplacian operator. Unfortunately, the finite-differencing approximation of the Laplacian used by ML Surface Estimation fails to offer sufficient accuracy.

This problem can be solved through a reasoned combination of forward and backward finite-differencing schemes. Denoted Modified-ML Surface Estimation, this method uses a more accurate approximation to the Laplacian operator. As a result, surfaces produced by Modified-ML estimation suffer from none of the roughness issues of ML estimation and exhibit an equal amount of robustness to image noise.

Serving as a means to incorporate noise modelling into surface integration, the surface reconstruction techniques developed during this thesis demonstrate the power of modelling noise and applying ML estimation to a problem. As noise is an unavoidable feature of images, ML serves a very important role for any application using images as observations. Visual-surface reconstruction is certainly one of these applications, and the Modified-ML estimation technique developed during this thesis serves as a powerful and effective means to handle image noise.

6.2 Future Work

To be an effective means of microfossil identification, computer-aided classification must be freed from the limitations inherent in images. Applying computer vision to computer-aided classification is a potent means to accomplish this goal. The work performed during this thesis explored several facets regarding how to best employ computer vision as a key aspect in classification tasks. The incorporation of video and shape-based template representations, the development of an alignment technique using photometric stereo, and the development of an ML surface estimation technique all constitute significant advancements to the field of computer-aided classification.

An important future goal is to measure the impact on classification performance that these enhancements provide. Measuring performance requires a fully classified dataset—a task whose difficulties motivate work on computerized identification in the first place. Consequently, obtaining results relating to the accuracy and reliability of the extended computer-aided system is challenging. For this reason, the UA Electronic Imaging Lab is currently increasing its capabilities to perform in-house manual classifications. It is expected that the microfossil dataset described in this thesis will be fully classified by late summer 2010. Additionally, the work done so far has garnered interest from the Natural History Museum in London, England, currently possessing one of the most impressive repositories of flora and fauna spanning the globe, including microfossils. As well, an ongoing project at the UA Electronic Imaging Lab is currently exploring means of dataset classification that do not completely depend on the full participation of micropalaeontologists or trained experts.

As a result, the most immediate feature of future work to pursue is to fully classify the foram dataset collected during this thesis. While it is expected that the extensions to the computer-aided system will significantly increase performance, improvements to the system can easily be envisioned, especially for a prototypical implementation such as this one. The sections below focus on these areas of improvement.

6.2.1 Video and Shape-Based Representations

Since errors in template identification were such a major source of error in [24], measuring the improvements in template identification is an important first task once a fully classified microfossil dataset is obtained. In addition, the performance benefits of controlling for illumination for automatic clustering must be evaluated. It is anticipated that the combination of controlled illumination and alternative template representations will increase classification performance, possibly even at the species level. Analysing performance will also help uncover priorities for further improvements.

Providing control of elevation angle represents one possible improvement to the system. As the current system is restricted to an illumination elevation angle of 30°, observations of object reflectance are limited to one elevation angle. Should more sophisticated reflectance models be incorporated into shape extraction methods, a varied set of elevation angles will prove crucial in performing visual-surface estimation. Additionally, a set of elevation angles would also increase the possible light directions available to the texture-mapped visual surfaces.

In addition to increasing available illumination angles, the online wiki is another area that would benefit from further work. Although the wiki can currently support video-based representations, presenting shape-based representations of microfossils would provide even greater advantages. When implemented in an open manner and providing data access to all interested parties, tools like the online wiki have the potential to act as useful means of education, analysis, and collaboration. The incorporation of 3D models into the wiki, with the option to manipulate viewpoint and lighting, would only magnify such benefits.

Apart from further work on templates, employing shape-based representations for automatic clustering constitute another significant aspect of future work. Currently, *automatic clustering* uses similarity matrices based on median image correlation scores between pairs of specimens. However, images only depict information from one lighting direction. On the other hand, shape-based representations can serve in principle as a basis for all available illumination conditions. In particular, as weighted normals accounted for 89% of image variability, they may provide a superior representation in which to measure similarity. However, doing so involves measuring similarity between vector fields, a task requiring further work.

6.2.2 Image Alignment Using Photometric Stereo

The alignment technique developed as part of the system extension proved successful in correcting for horizontal and vertical discrepancies between photometric stereo images. However, as the problem involved the minimization of a nonlinear function, performance depends in part on the minimization algorithm used. This thesis tested four relatively standard routines, QN, CG, SCG, and Levenberg-Marquart, with CG prevailing over the other two options. Alternative minimization schemes besides those three may produce better results.

Exploring the resolution of misalignments outside of horizontal and vertical shifts is another promising and challenging aspect of future work. These may include rotational, affine, and even perspective shifts. The latter two may even be used to approximate any real-world object misalignments, provided their magnitude does not significantly alter the viewpoints. For example, they could model minor variations in the real-world positions of the object and/or camera across all images in the sequence. In these cases the Jacobians needed for alignment could be very similar to the affine and perspective warps derived in [76]. However, unlike translational misalignments, an update to light direction in the image formation equation must accompany rotational, affine, and perspective warps. As well, simple change of variables would no longer suffice to alternate between object and image shifts, making it unclear on how to best retain an ML estimation formulation.

The alternation between object and image shifts also leads to additional aspects of future work. In an effort to retain simplicity, when changing variables to formulate the problem using image shifts, noise was assumed to be zero-mean IID Gaussian terms at all *continuous* locations—an assumption not supported by the primarily discrete nature of image noise. Modelling noise more accurately would involve incorporating the interpolation method used to estimate image values at continuous locations. Most likely, this would mean modelling noise as a weighted sum of IID Gaussian error terms. This would couple noise values at all locations together through some sparse structure. As a result, ML estimates of weighted normals would require solving a large and sparse GLS system simultaneously for all pixel locations.

Yet, this coupling means that solving the GLS problem using current sparse solver techniques would quickly become intractable for typical computers. This also affects the nonlinear step of image shift estimation, as the problem would have to be transformed from a nonlinear GLS formulation to an OLS one using the same covariance matrix. Increasing the capabilities of sparse linear algebra routines so that they can handle GLS problems with coupled covariance matrices would alleviate these problems. Developing a sparse GQR technique may be the best direction to take for this goal. Even more importantly, a fast and reliable sparse GQR routine would benefit innumerable applications outside that of the image misalignment problem.

6.2.3 Maximum Likelihood Surface Estimation

Even though weighted normals accounted for close to 90% of image variability, there is significant room for improvement. For instance, the model of image information used in this dissertation can easily break down in the presence of non-Lambertian reflectance. More advanced methods of normal estimation should be explored, such as using more sophisticated [77] or even non-parametric [78] models of reflectance. However, doing so means abandoning the convenient linear model of image formation and the accompanying convenient means to model statistical behaviour of weighted-normal estimates.

Assuming no shadows in the image represents another key limitation of classic photometric stereo. Although this thesis used a simple heuristic to identify shadows, this approach is certainly not without flaws. For instance, it does not take into account whether the residual error is reduced. As well, the probability of a pixel being shadowed should increase if its neighbours are also shadowed. However, the complexity of considering these issues drastically increases with the number of different viewpoints. Despite these practical issues, the application of shadowed pixel identification methods taking these issues into account, such as [79], should be investigated.

Yet, irrespective of the generative model, using least squares to estimate normals will always produce an ML estimate provided image noise is comprised of zero-mean IID Gaussian additive terms. While normally distributed noise is certainly a mathematically convenient and well-accepted assumption in image processing, arguments exist for alternative parametric models of noise [43]. Additionally, additive noise terms fail to take into account other possible distortions, such as camera blur.

While these considerations motivate incorporating different or even more sophisticated noise models into the image formation equation, doing so introduces difficulties in performing ML estimation, as it may no longer be equivalent to least squares. One possible avenue to explore is applying robust statistics [80] to surface normal estimation. By assuming an underlying parametric model to the data, but with the tacit acknowledgement that the model is not an exact abstraction, robust statistics may offer a means to continue using a Gaussian distribution to model the errors along with all the accompanying mathematical niceties.

Alternatively, situations where image noise is modelled as following longer-tailed distributions are better handled through solving least-absolute deviations [81]. As well, least-absolute deviations produces the ML estimate for Laplacian distributed noise terms [75]. These cases may provide an interesting avenue for future work.

In addition to properly modelling image noise and weighted-normal behaviour, the stochastic characteristics of the *gradient* estimates are another crucial factor in obtaining an accurate 3D model. Since the p and q gradients are both ratios of surface normal components, their estimation is a nonlinear formulation. As a result, for small samples sizes, stochastic behaviour of the estimates is heavily influenced by the nonlinear characteristics of the model function [48]. To avoid this problem, this thesis modelled estimated gradients using asymptotic limits.

In general, a sample size of 18 light directions will not be large enough to allow estimates to reach these asymptotic limits. Consequently, asymptotic variance will generally be smaller than the actual variance and the estimator will also be biased [48, 50]. However, estimates of the curvature of the model at specific values of $\hat{\beta}$ can lead to estimates of the bias and true variance [48]. Despite the time consuming nature of this task, curvature analysis should be considered for future work.

Alternatively, as gradients consist of a ratio of two normally distributed terms (assuming Gaussian IID image noise), methods explicitly modelling this ratio may also prove useful. For instance, Kuparinen describes specific physical conditions where gradient behaviour can be approximated as normally distributed [46]. As well, general conclusions, such as those of Hayya *et al* [82], related to ratios of this type can provide additional insight into gradient parameter estimation and behaviour.

Although the surface integration routine developed during the course of this thesis successfully integrates image noise into its formulation, it still suffers from information loss relative to the weighted normals. While the least-squares technique used to estimate surfaces in photometric stereo is globally accurate, many local features are lost during the integration [83]. While potentially affecting any application, this poses a particular problem in modelling microfossils, as local features and texture often constitute strong identifying features [35]. This detail loss is evidenced by the 74% reconstruction rate of visual surfaces vs. the 89% rate of weighted normals in the context of the foram dataset.

One potential way to solve this problem is to estimate surface values directly from images intensities, mitigating the detail loss from performing two separate least-squares estimation steps. As discussed in Chapter 5, techniques developed by Noakes *et al* already attempt this [41, 42]. However, their work has yet to develop a practical implementation. Another option is to initially estimate a surface using integration methods, such as the one developed in this thesis, and then later refine the surface to include finer details. For instance, one can first perform surface estimation from photometric stereo, and then apply shape-from-shading techniques onto the depth map [83]. Alternatively, one can perform bundle adjustment, a method general enough to encompass many topics in computer vision [84]. Regardless of the manner in which the surface is refined, these and other methods should be a priority in any future work hoping to use depth maps as visual tools for microfossil identification and learning. As a final note, the issues addressed here regarding loss of fine detail are applicable to almost any textured object. Resolving these issues then are of significant importance on their own merits, and should be a priority for the field of visual-surface reconstruction in general.

Bibliography

[1] Patricia Vickers Rich, Pat Vickers Rich, Mildred Adams Fenton, Thomas Hewitt Rich, and Carroll Lane Fenton, *The Fossil Book: A Record of Prehistoric Life*, Courier Dover Publications, 1997.

[2] Kurt R. Geitzenauer, "Coccoliths as late quaternary palaeoclimatic indicators in the subantarctic pacific ocean," *Nature*, vol. 223, no. 5202, pp. 170–172, 1969.

[3] P.A. Swaby, "Vides: an expert system for visually identifying microfossils," *IEEE Expert [see also IEEE Intelligent Systems and Their Applications]*, vol. 7, no. 2, pp. 36–42, Apr 1992.

[4] R.O. Koshkarly B.J. O'Neill R.W. Scott M.D. Simmons, W.A. Berggren and W. Ziegler, "Biostratigraphy and geochronology in the 21st century," in *International Senckenberg Conference: Paleontology in the 21st Century Workshop*, H.R. Lane, J. Lipps, F.F. Steininger, and W. Ziegler, Eds., Frankfurt Germany, 1997, Senckenberg Museum.

[5] M. Simmons, "Biostratigraphy: Surviving extinction," *Palaios*, vol. 13, no. 3, pp. 215–216, 1998.

[6] "Ocean drilling program final technical report," Tech. Rep., Consortium for Ocean Leadership Inc., Lamont-Doherty Earth Observatory of Columbia University, Texas A&M University, National Science Foundation Contracts ODP83-17349 and OCE93-08410, 2007.

[7] J. Hodgson, "Gene sequencing's industrial revolution," *IEEE Spectrum*, vol. 37, no. 11, pp. 36–42, 2000.

[8] Hans du Buf and Micha M. Bayer, Eds., *Automatic Diatom Identification*, vol. 51 of *Machine Perception and Artificial Intelligence*, World

Scientific, New Jersey, 2002.

[9] P. F. Culverhouse, R. Williams, B. Reguera, V. Herry1, and S. González-Gil, “Do experts make mistakes? a comparison of human and machine identification of dinoflagellates,” *Marine Ecology Progress Series*, vol. 247, pp. 17–25, February 2003.

[10] G. Bernard Munsch, Ed., *Second Conference on Scientific Ocean Drilling (COSOD II)*, Strasbourg, France, 1987. Joint Oceanographic Institutions for Deep Earth Sampling, European Science Foundation.

[11] D.R. Brough and I.F. Alexander, “The fossil expert,” *Expert Systems*, vol. 3, no. 2, pp. 76–83, 1986.

[12] W. R. Riedel, “Identify: a prolog program to help identify fossils,” *Comput. Geosci.*, vol. 15, no. 5, pp. 809–823, 1989.

[13] S. Liu, M. Thonnat, and M. Berthod, “Automatic classification of planktonic foraminifera by a knowledge-based system,” *Artificial Intelligence for Applications, 1994., Proceedings of the Tenth Conference on*, pp. 358–364, 1-4 Mar 1994.

[14] S. Yu, P. Saint-Marc, M. Thonnat, and M. Berthod, “Feasibility study of automatic identification of planktic foraminifera by computer vision,” *The Journal of Foraminiferal Research*, vol. 26, no. 2, pp. 113–123, 1996.

[15] D. Dollfus and L. Beaufort, “Fat neural network for recognition of position-normalised objects,” *Neural Netw.*, vol. 12, no. 3, pp. 553–560, 1999.

[16] L. Beaufort and D. Dollfus, “Automatic recognition of coccoliths by dynamical neural networks,” *Marine Micropaleontology*, vol. 51, no. 1-2, pp. 57–73, 2004.

[17] Lionel Tarassenko, *Guide to Neural Computing Applications*, John Wiley and Sons, Inc., New York, NY, USA, 1998.

[18] Jrg Bollmann, Patrick S. Quinn, Miguel Vela, Bernhard Brabec, Siegfried Brechner, Mara Y. Corts, Heinz Hilbrecht, Daniela N. Schmidt, Ralf Schiebel, and Hans R. Thierstein, *Image Analysis, Sed-*

iments and Paleoenvironments, vol. 7 of *Developments in Paleoenvironmental Research*, chapter Automated Particle Analysis: Calcareous Microfossils, pp. 229–252, Springer Netherlands, 2005.

[19] Martin A. Buzas and Stephen J. Culver, “Species diversity and dispersal of benthic foraminifera,” *BioScience*, vol. 41, no. 7, pp. 483–489, 1991.

[20] Frank Eric Round, R. M. Crawford, and D. G. Mann, *Diatoms: biology and morphology of the genera*, Cambridge University Press, 1990.

[21] John W. Murray, “Biodiversity of living benthic foraminifera: How many species are there?,” *Marine Micropaleontology*, vol. 64, no. 3-4, pp. 163–176, 2007.

[22] Eugene F. Stoermer, “Diatom taxonomy for paleolimnologists,” *Journal of Paleolimnology*, vol. 25, no. 3, pp. 393–398, 2001.

[23] Walter C. Sweet and Philip C. J. Donoghue, “Conodonts: Past, present, future,” *Journal of Paleontology*, vol. 75, no. 6, pp. 1174–1184, 2001.

[24] Kamal Ranaweera, Adam P. Harrison, Santo Bains, and Dileepan Joseph, “Feasibility of computer-aided identification of foraminiferal tests,” *Marine Micropaleontology*, vol. 72, no. 1-2, pp. 66 – 75, 2009.

[25] D.C. Kelly, T.J. Bralower, J.C. Zachos, I.P. Silva, and E. Thomas, “Rapid diversification of planktonic foraminifera in the tropical pacific (odp site 865b) during the late paleocene thermal maximum,” *Geology*, vol. 24, pp. 423–429, 1996.

[26] A.E. Rathburn, J.J. Pichon, M.A. Ayress, and DeDeckker, “Microfossil and stable-isotope evidence for changes in late holocene palaeoproductivity and palaeoceanographic conditions in the prydz bay region of antarctica,” *Palaeogeography Palaeoclimatology Palaeoecology*, vol. 131, pp. 485–510, 1997.

[27] J.P. Kennett and L.D. Stott, “Abrupt deep-sea warming, palaeoceanographic changes and benthic extinctions at the end of the palaeocene,” *Nature*, vol. 353, pp. 225–229, Sep. 1991.

[28] S.Q. Breard, A.D. Callender, and M.J. Nault, "Paleoecologic and biostratigraphic models for pleistocene through miocene foraminiferal assemblages of the gulf coast basin," *Gulf Coast Association of Geological Societies, Transactions*, vol. 43, pp. 493–502, 1993.

[29] Rafael C. Gonzalez and Richard E. Woods, *Digital Image Processing (3rd Edition)*, Prentice Hall, January 2006.

[30] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: a review," *ACM Comput. Surv.*, vol. 31, no. 3, pp. 264–323, September 1999.

[31] Kamal Ranaweera, Santo Bains, and Dileepan Joseph, "Analysis of image-based classification of foraminiferal tests," *Marine Micropaleontology*, vol. 72, no. 1-2, pp. 60 – 65, 2009.

[32] Alan Agresti and Brent A. Coull, "Approximate is better than exact for interval estimation of binomial proportions," *The American Statistician*, vol. 52, no. 2, pp. 119–126, 1998.

[33] Linda G. Shapiro and George C. Stockman, *Computer Vision*, Prentice Hall, January 2001.

[34] David A. Forsyth and Jean Ponce, *Computer Vision: A Modern Approach*, Prentice Hall, August 2002.

[35] Howard Armstrong and Martin Brasier, *Microfossils*, Wiley-Blackwell, 2 edition, 2005.

[36] Mark D. Sutton, Derek E. G. Briggs, David J. Siveter, and Derek J. Siveter, "An exceptionally preserved vermiform mollusc from the silurian of england," *Nature*, vol. 410, no. 6827, pp. 461–463, 2001.

[37] David J. Siveter, Mark D. Sutton, Derek E. G. Briggs, , and Derek J. Siveter, "An ostracode crustacean with soft parts from the lower silurian," *Science*, vol. 302, no. 5651, pp. 1749–1751, 2003.

[38] Berthold K. Horn, *Robot Vision*, McGraw-Hill Higher Education, 1986.

[39] Robert J. Woodham, "Photometric method for determining surface orientation from multiple images," *Optical Engineering*, vol. 19, no. 1,

pp. 139–144, 1980.

[40] R. Epstein, Alan L. Yuille, and Peter N. Belhumeur, "Learning Object Representation from Lighting Variations," in *ECCV '96: Proceedings of the International Workshop on Object Representation in Computer Vision II*, London, UK, 1996, pp. 179–199, Springer-Verlag.

[41] Lyle Noakes and Ryszard Kozera, "Nonlinearities and Noise Reduction in 3-Source Photometric Stereo," *Journal of Mathematical Imaging and Vision*, vol. 18, no. 2, pp. 119–127, 2003.

[42] Lyle Noakes and Ryszard Kozera, "Denoising Images: Non-linear Leap-Frog for Shape and Light-Source Recovery," in *Geometry, Morphology, and Computational Imaging*, vol. 2616 of *Lecture Notes in Computer Science*, pp. 143–162. Springer Berlin / Heidelberg, 2003.

[43] B. Karaçali and W. Snyder, "Noise Reduction in Surface Reconstruction from a Given Gradient Field," *International Journal of Computer Vision*, vol. 60, no. 1, pp. 25–44, 2004.

[44] Amit Agrawal, Rama Chellappa, and Ramesh Raskar, "An Algebraic Approach to Surface Reconstruction from Gradient Fields," in *ICCV '05: Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, Washington, DC, USA, 2005, pp. 174–181, IEEE Computer Society.

[45] Amit K. Agrawal, Ramesh Raskar, and Rama Chellappa, "What Is the Range of Surface Reconstructions from a Gradient Field?," in *Proceedings of the European Conference on Computer Vision*, 2006, pp. 578–591.

[46] Toni Kuparinen and Ville Kyrki, "Optimal Reconstruction of Approximate Planar Surfaces Using Photometric Stereo," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 12, pp. 2282–2289, 2009.

[47] David A. Ratkowsky, *Handbook of Nonlinear Regression Models*, Marcel Dekker Inc., 1990.

[48] G. A. Seber and C. J. Wild, *Nonlinear Regression*, John Wiley & Sons,

Inc., 1989.

[49] A Ronald Gallant, *Nonlinear statistical models*, John Wiley & Sons, Inc., New York, NY, USA, 1986.

[50] David A. Ratkowsky, *Nonlinear Regression Modeling*, Marcel Dekker Inc., 1983.

[51] Takeshi Amemiya, *Advanced Econometrics*, chapter 4, Basil Blackwell Ltd., 1985.

[52] Stuart L. Meyer, *Data Analysis for Scientists and Engineers*, John Wiley & Sons Inc., 1975.

[53] Jan R. Magnus and Heinz Neudecker, *Matrix differential calculus with applications in statistics and econometrics*, chapter 13, John Wiley & Sons, 1999.

[54] Gene H. Golub and Charles F. Van Loan, *Matrix computations (3rd ed.)*, Johns Hopkins University Press, Baltimore, MD, USA, 1996.

[55] John R. Gilbert, Cleve Moler, and Robert Schreiber, "Sparse matrices in matlab: Design and implementation," *SIAM J. Matrix Anal. Appl*, vol. 13, pp. 333–356, 1992.

[56] C.C. Paige, "Computer Solution and Perturbation Analysis of Generalized Linear Least Squares Problems," *Mathematics of Computation*, vol. 33, no. 145, pp. 171–183, 1979.

[57] E. Anderson, Z. Bai, and J. Dongarra, "Lapack working note 31: Generalized qr factorization and its applications," Tech. Rep., Knoxville, TN, USA, 1991.

[58] Ondrej Drbohlav and Mike Chantler, "On optimal light configurations in photometric stereo," *Computer Vision, IEEE International Conference on*, vol. 2, pp. 1707–1712, 2005.

[59] Amnon Shashua, "On photometric issues in 3d visual recognition from a single 2d image," *International Journal of Computer Vision*, vol. 21, pp. 99–122, 1997.

[60] Ravi Ramamoorthi, "Analytic PCA Construction for Theoretical Analysis of Lighting Variability in Images of a Lambertian Object," *IEEE Transactions in Pattern Analysis and Machince Intelligence*, vol. 24, no. 10, pp. 1322–1333, 2002.

[61] N. J. D. Nagelkerke, "A note on a general definition of the coefficient of determination," *Biometrika*, vol. 78, no. 3, pp. 691–692, September 1991.

[62] A. Ardeshir Goshtasby, *2-D and 3-D Image Registration: for Medical, Remote Sensing, and Industrial Applications*, Wiley-Interscience, 2005.

[63] B. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision.," in *Proceedings of the International Joint Conference on Artificial Intelligence*, 1981, pp. 674–679.

[64] Sven Loncaric, "A survey of shape analysis techniques," *Pattern Recognition*, vol. 31, pp. 983–1001, 1998.

[65] Remco C. Veltkamp and Michiel Hagedoorn, "State of the art in shape matching," pp. 87–119, 2001.

[66] Jonathan R. Shewchuk, "An introduction to the conjugate gradient method without the agonizing pain," Tech. Rep., Pittsburgh, PA, USA, 1994.

[67] Martin Fodslette Moller, "A scaled conjugate gradient algorithm for fast supervised learning," *Neural Netw.*, vol. 6, no. 4, pp. 525–533, 1993.

[68] Berthold K. P. Horn and Michael J. Brooks, "The Variational Approach to Shape from Shading," *Computer Vision, Graphics, and Image Processing*, vol. 33, no. 2, pp. 174–208, February 1986.

[69] Robert T. Frankot and Rama Chellappa, "A Method for Enforcing Integrability in Shape from Shading Algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 10, no. 4, pp. 439–451, 1988.

[70] T. Simchony, R. Chellappa, and M. Shao, "Direct Analytical Methods

for Solving Poisson Equations in Computer Vision Problems," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 5, pp. 435–446, 1990.

[71] Tristan Cameron, Ryszard Kozera, and Amitava Datta, "A Parallel Leap-Frog Algorithm for 3-Aource Photometric Stereo," in *Computer Vision and Graphics: International Conference, ICCVG 2004*. 2006, Computational Imaging and Vision, pp. 95–102, Springer.

[72] P. Kovesi, "Shapelets Correlated with Surface Normals Produce Surfaces," in *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, Oct. 2005, vol. 2, pp. 994–1001 Vol. 2.

[73] Steven C. Chapra and Raymond Canale, *Numerical Methods for Engineers*, McGraw-Hill, Inc., New York, NY, USA, 2006.

[74] Timothy A. Davis, *Direct Methods for Sparse Linear Systems*, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2006.

[75] Yinbo Li and Gonzalo R. Arce, "A Maximum Likelihood Approach to Least Absolute Deviation Regression," *EURASIP Journal on Applied Signal Processing*, vol. 2004, pp. 1762–1769, 2004.

[76] Simon Baker and Iain Matthews, "Lucas-kanade 20 years on: A unifying framework," *International Journal of Computer Vision*, vol. 56, no. 3, pp. 221–255, February 2004.

[77] Dan B. Goldman, Brian Curless, Aaron Hertzmann, and Steven M. Seitz, "Shape and spatially-varying brdfs from photometric stereo," in *ICCV '05: Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, Washington, DC, USA, 2005, pp. 341–348, IEEE Computer Society.

[78] Michael Holroyd, Jason Lawrence, Greg Humphreys, and Todd Zickler, "A photometric approach for estimating normals and tangents," *ACM Trans. Graph.*, vol. 27, no. 5, pp. 1–9, 2008.

[79] M. Chandraker, S. Agarwal, and D. Kriegman, "ShadowCuts: Photometric Stereo with Shadows," in *Proceedings of the IEEE Conference on*

Computer Vision and Pattern Recognition CVPR '07, 17–22 June 2007, pp. 1–8.

[80] Frank R. Hampel, Elvezio M. Ronchetti, Peter J. Rousseeuw, and Werner A. Stahel, *Robust Statistics: The Approach Based on Influence Functions (Wiley Series in Probability and Statistics)*, Wiley-Interscience, New York, revised edition, April 2005.

[81] Subhash C. Narula and John F. Wellington, "The minimum sum of absolute errors regression: A state of the art survey," *International Statistical Review / Revue Internationale de Statistique*, vol. 50, no. 3, pp. 317–326, 1982.

[82] Jack Hayya, Donald Armstrong, and Nicolas Gressis, "A note on the ratio of two normally distributed variables," *Management Science*, vol. 21, no. 11, pp. 1338–1341, 1975.

[83] U. Sakarya and I. Erkmen, "An improved method of photometric stereo using local shape from shading," *Image and Vision Computing*, vol. 21, no. 11, pp. 941–954, 2003.

[84] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, ISBN: 0521540518, second edition, 2004.

Appendix A

Creating Anaglyph Images

Anaglyphs are stereoscopic images that simulate the binocular effect of our eyes by creating artificial disparity. One way to create an anaglyph is to produce two black and white images, where each image is warped from a perspective belonging to either the left or right eye. These warped intensity-based images are then assigned to different colour channels. Thus, by viewing the image using filtered glasses, the anaglyph produces an artificial disparity, which human brains interpret as 3D depth. Usually, the red channel holds one image, and the blue and green channels holds the other image.

In the context of creating microfossil anaglyphs, the perspective warp employs the shape-based representations to perform two separate nonlinear mappings of the video-based representations. Working directly in pixel space, the depth values in shape-based representations provide a point in 3D space for every pixel location of the video. By specifying two different viewpoints, perspective projection can then map each 3D point into two separate sets of 2D points. Doing so involves working in projective space, which requires using homogenous coordinates for both 3D and 2D points [84]. Denoted $\mathbf{x}$ and $\mathbf{X}$ respectively, 2D and 3D homogenous points are expressed as:

$$\mathbf{x} = \begin{pmatrix} x, & y, & z \end{pmatrix}^T, \tag{A.1}$$

$$\mathbf{X} = \begin{pmatrix} x, & y, & z, & w \end{pmatrix}^T. \tag{A.2}$$

$$\tag{A.3}$$

The Euclidean expressions of $\mathbf{x}$ and $\mathbf{X}$ are $\begin{pmatrix} x/z, & y/z \end{pmatrix}^T$ and $\begin{pmatrix} x/w, & y/w, & z/w \end{pmatrix}^T$ respectively. To convert from Euclidean to projective space requires simply setting either z or w to 1 for $\mathbf{x}$ and $\mathbf{X}$ respectively. The following discussion will assume that the world coordinate frame follows the specifications in Fig. 3.3

Creating anaglyphs requires creating two different projective warps from viewpoints looking directly down to the microfossil video and at equal distances to the left and right of the x-axis origin. These two viewpoints can be parameterized by two values:

- D, the height of the two viewpoints from the image plane;
- d, the disparity between the two viewpoints.

In general, a projective warp **P** mapping **X** to **x** is expressed as:

$$\mathbf{P} = \mathbf{KR}\begin{bmatrix} \mathbf{I} & -\mathbf{C} \end{bmatrix}, \tag{A.4}$$

where **K** is the camera calibration matrix, **R** is a rotation matrix aligning the world coordinate frame with the camera coordinate frame, and **K** is the camera location in world coordinates [84]. Note that since creating anaglyphs using microfossil video and shape representations only requires working in pixel space, world coordinate frame is somewhat of a misnomer.

Working in pixel space simplifies the formulation of the camera calibration matrix to the following expression:

$$\mathbf{K} = \begin{pmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{A.5}$$

where f is the camera focal length. To ensure that different D values will not globally enlarge or shrink the resulting image dimensions, f should be set to D.

As camera coordinates are expressed with the z direction pointing toward the image plane, the z axes of the camera and world coordinate frames point in opposite directions. As a result, **R** is written as:

$$\mathbf{R} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}. \tag{A.6}$$

Finally, given D and d, the camera location in the world coordinate frame is expressed as:

$$\mathbf{C} = \begin{pmatrix} \pm d/2 & 0 & D \end{pmatrix}^T, \tag{A.7}$$

where the x location's sign depends on whether the left or right eye is being simulated.

Thus, **P** can be written formally as:

$$\mathbf{P} = \begin{bmatrix} D & 0 & 0 \\ 0 & D & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 & \mp d/2 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -D \end{bmatrix}. \tag{A.8}$$

Using homogenous coordinate representations of the world 3D points corresponding to the pixel locations and depth map values of the microfossil shapes, the anaglyph image coordinates are computed through:

$$\mathbf{x} = \mathbf{P}\mathbf{X}. \tag{A.9}$$

However, since anaglyphs need only warp in the x direction to create disparity, the nonlinear mappings in the y direction of (A.9) can be ignored. The pixel values of the video-based representations corresponding to $\mathbf{X}$ are then mapped to the image location specified by $\mathbf{x}$. Since (A.9) results in a non-uniform set of image locations, pixel intensities must be interpolated at gridded locations to create an image.

Appendix B

First-Order Finite-Difference Operators

Recall the system of equations framing surface estimation as a linear regression problem:

$$\begin{pmatrix} \mathbf{p} \\ \mathbf{q} \end{pmatrix} = \begin{pmatrix} \mathbf{D}_x \\ \mathbf{D}_y \end{pmatrix} \mathbf{z} + \boldsymbol{\epsilon}, \tag{B.1}$$

Using the normal equations, solving (B.1) is equivalent to solving the following system of equations:

$$\begin{pmatrix} \mathbf{D}_x^T & \mathbf{D}_y^T \end{pmatrix} \boldsymbol{\Gamma}_{\boldsymbol{\epsilon\epsilon}}^{-1} \begin{pmatrix} \mathbf{p} \\ \mathbf{q} \end{pmatrix} = \begin{pmatrix} \mathbf{D}_x^T & \mathbf{D}_y^T \end{pmatrix} \boldsymbol{\Gamma}_{\boldsymbol{\epsilon\epsilon}}^{-1} \begin{pmatrix} \mathbf{D}_x \\ \mathbf{D}_y \end{pmatrix} \hat{\mathbf{z}}, \tag{B.2}$$

where $\boldsymbol{\Gamma}_{\boldsymbol{\epsilon\epsilon}}$ is proportional to the covariance matrix of the errors. However, under certain conditions and depending on the finite-differencing scheme used, the normal equations of (B.2) follow some useful properties. For the rest of this discussion, $\mathbf{D}^f$, $\mathbf{D}^b$, and $\mathbf{D}^c$ will denote forward, backward, and centered-differencing respectively.

B.1 1D Finite-Difference Operators

Imagine a 1D continuous function $f(x)$ and its derivative $f'(x)$ with respect to x. Approximating the function and the derivative operation requires sampling and restricting the domain x of f. Assuming periodicity is one way to restrict the domain. More formally, construct a periodic function $f_*(x)$ where $f_*(x) = f(x)$, $0 \le x < L$ and $f_*(x + kL) = f_*(x)$, $k \in \mathbb{Z}$, and $\forall x$. Under these assumptions, $f'(x) = f_*'(x)$, $0 \le x < L$.

Now denote the vector $\mathbf{f}$ as a sampled version of $f_*(x)$. Let L be an integer and assume the sampling period is 1. By constructing a finite-differencing operator, $\mathbf{D}$, one can calculate $\mathbf{f}' = \mathbf{D}\mathbf{f}$, which is a sampled version of $f_*'(x)$ up to some order of accuracy depending on $\mathbf{D}$. The periodic nature of $\mathbf{f}$ affects the structure of $\mathbf{D}$ at the borders.

Consider the case of centered-differencing. Should $L = 5$, then a centered-differencing version of $\mathbf{D}$ would be:

$$\mathbf{D}^c = \begin{pmatrix} 0 & 1 & 0 & 0 & -1 \\ -1 & 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & 1 \\ 1 & 0 & 0 & -1 & 0 \end{pmatrix}. \tag{B.3}$$

Notice that rows of $\mathbf{D}^c$ corresponding to the borders of $\mathbf{f}$ use the periodicity of the function in assigning elements values. Also notice, that a key feature of the structure of $\mathbf{D}^c$ is that the operator is antisymmetric, meaning $\mathbf{D}^{c^T} = -\mathbf{D}^c$. Assuming periodicity is not the only way to restrict domains to ensure antisymmetry. For example, assuming values of $f(x)$ outside of $[0, L)$ are 0 also produces an antisymmetric derivative operator.

Forward-differencing, on the other hand, follows a different property. Again, allowing L to equal 5, a forward-differencing version of $\mathbf{D}$ would be:

$$\mathbf{D}^f = \begin{pmatrix} -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 \\ 1 & 0 & 0 & 0 & -1 \end{pmatrix}. \tag{B.4}$$

A key property of $\mathbf{D}^f$ is that its transpose is equal to the negative of its backward-differencing counterpart, meaning $\mathbf{D}^{f^T} = -\mathbf{D}^b$. This property also holds should one assume values of $f(x)$ outside of $[0, L)$ are 0.

B.2 2D Finite-Difference Operators

The properties of first-order 1D derivative operators demonstrated in Sec. B.1 extend to the 2D case. Let $f(x,y)$ be some continuous function over 2D space. Consider a 2D periodic equivalent $f_*(x,y)$, where $f_*(x,y) = f(x,y)$, $0 \leq x < L$ and $0 \leq y < L$. Also, $f_*(x+kL, y) = f_*(x, y+mL) = f_*(x,y)$, $k,m \in \mathbb{Z}$, and $\forall x,y$. Without loss of generality, let L be an integer, and the sampling period in both directions be 1. Assuming the function periods are equivalent in both directions in not necessary for this discussion, but it does simplify explanation.

Analogous to the 1D case, formulations for the sampled versions of $f(x,y)$ and its partial derivatives with respect to x and y, denoted respectively as

$f_x(x,y)$ and $f_y(x,y)$, exist. Let the $L \times L$ matrix $\mathbf{F}$ be the sampled version of $f(x,y)$ over $0 \leq x < L$ and $0 \leq y < L$. Calculating discrete approximations of its partial derivatives using linear algebra requires a "flattening" of $\mathbf{F}$ into column vector form. Two orderings are possible to flatten matrices: column or row-major ordering. Keeping in line with MATLAB, column-major ordering is the preferred ordering for this discussion. Denoting $\mathbf{f}^{col}$ as the column-major ordering vector of $\mathbf{F}$ and $\mathbf{f}^{row}$ as the row-major equivalent, the two orderings are related to each other through a permutation, $\mathbf{f}^{row} = \mathbf{\Pi}\mathbf{f}^{col}$. Similarly discrete approximations to $f_x(x,y)$ and $f_y(x,y)$ are denoted respectively as $\mathbf{f}_x^{(.)}$ and $\mathbf{f}_y^{(.)}$, where the superscript indicates the type of ordering used.

Column-major ordering simplifies producing derivatives in the y direction. To illustrate this, one may express the y derivative as: $\mathbf{f}_y^{col} = \mathbf{D}_y\mathbf{f}^{col}$, where

$$\mathbf{D}_y = \begin{pmatrix} \mathbf{D} & & & \\ & \mathbf{D} & & \\ & & \ddots & \\ & & & \mathbf{D} \end{pmatrix}, \tag{B.5}$$

with $\mathbf{D}$ being one of the $L \times L$ matrices defined as in Sec. B.1. The transpose of $\mathbf{D}_y$ offers a convenient form:

$$\mathbf{D}_y^T = \begin{pmatrix} \mathbf{D}^T & & & \\ & \mathbf{D}^T & & \\ & & \ddots & \\ & & & \mathbf{D}^T \end{pmatrix}. \tag{B.6}$$

When employing centered-differencing, (B.6) is expressed as:

$$\mathbf{D}_y^{c\,T} = \begin{pmatrix} \mathbf{D}^{c\,T} & & & \\ & \mathbf{D}^{c\,T} & & \\ & & \ddots & \\ & & & \mathbf{D}^{c\,T} \end{pmatrix} = \begin{pmatrix} -\mathbf{D}^{c} & & & \\ & -\mathbf{D}^{c} & & \\ & & \ddots & \\ & & & -\mathbf{D}^{c} \end{pmatrix} = -\mathbf{D}_y^c. \tag{B.7}$$

Thus, centered-differencing in the y direction is antisymmetric. However, should one use forward-differencing, (B.6) is expressed as:

$$\mathbf{D}_y^{f\,T} = \begin{pmatrix} \mathbf{D}^{f\,T} & & & \\ & \mathbf{D}^{f\,T} & & \\ & & \ddots & \\ & & & \mathbf{D}^{f\,T} \end{pmatrix} = \begin{pmatrix} -\mathbf{D}^{b} & & & \\ & -\mathbf{D}^{b} & & \\ & & \ddots & \\ & & & -\mathbf{D}^{b} \end{pmatrix} = -\mathbf{D}_y^b, \tag{B.8}$$

demonstrating that in the y direction, the forward and backward difference operators follow the same properties as the 1D case.

Somewhat non-intuitively, $\mathbf{D}_y$ can also compute derivatives for the x direction; yet, ordering must be *row-major*. This results in a simple expression of the partial derivative with respect to x: $\mathbf{f}_x^{row} = \mathbf{D}_y \mathbf{f}^{row}$. Notice that both the function and its derivative are expressed in row-major ordering. However, to keep the result consistent with the y derivative, this formulation should be converted to column-major ordering. The conversion is written as:

$$\mathbf{f}_x^{col} = \mathbf{\Pi}^T \mathbf{f}_x^{row} = \mathbf{\Pi}^T \mathbf{D}_y \mathbf{f}^{row} = \mathbf{\Pi}^T \mathbf{D}_y \mathbf{\Pi} \mathbf{f}^{col}. \tag{B.9}$$

As a result, $\mathbf{D}_x = \mathbf{\Pi}^T \mathbf{D}_y \mathbf{\Pi}$. Antisymmetry of centered-differencing operators in the x direction follows from:

$$\mathbf{D}_x^{c\,T} = \mathbf{\Pi}^T \mathbf{D}_y^{c\,T} \mathbf{\Pi} = -\mathbf{\Pi}^T \mathbf{D}_y^c \mathbf{\Pi} = -\mathbf{D}_x^c. \tag{B.10}$$

Thus, both $\mathbf{D}_x^c$ and $\mathbf{D}_y^c$ are antisymmetric matrices. In the forward-differencing case, $\mathbf{D}_x^{f\,T}$ can be expressed as:

$$\mathbf{D}_x^{f\,T} = \mathbf{\Pi}^T \mathbf{D}_y^{f\,T} \mathbf{\Pi} = -\mathbf{\Pi}^T \mathbf{D}_y^b \mathbf{\Pi} = -\mathbf{D}_x^b. \tag{B.11}$$

All of the preceding results can also be demonstrated using row-major ordering.

B.3 Surface Estimation Normal Equations

Although least-squares problems are typically solved using QR based methods, the normal equation formulation of the problem can serve as a useful algebraic illustration of the system being solved. As well, in the case of visual-surface estimation, limitations in sparse GQR techniques obligate the use of the normal equations. If one assumes periodicity in the surface, then $\mathbf{D}_x$ and $\mathbf{D}_y$ take on the forms and the accompanying properties of the matrices in Sec. B.2. Although surfaces are not typically periodic, this is a common assumption when dealing with images and depth maps. For instance, the discrete Fourier transform includes an implicit assumption of periodicity. As a result, methods of surface reconstruction based off of Fourier bases, such as the well-known Frankot-Chellappa method [69] assume a periodic surface.

Should the expression for the surface estimation normal equations in (B.2) employ centered-differencing, the antisymmetric properties of the operator produces the following formulation:

$$\begin{pmatrix} \mathbf{D}_x^c & \mathbf{D}_y^c \end{pmatrix} \mathbf{\Gamma}_{\epsilon\epsilon}^{-1} \begin{pmatrix} \mathbf{p}^* \\ \mathbf{q}^* \end{pmatrix} = \begin{pmatrix} \mathbf{D}_x^c & \mathbf{D}_y^c \end{pmatrix} \mathbf{\Gamma}_{\epsilon\epsilon}^{-1} \begin{pmatrix} \mathbf{D}_x^c \\ \mathbf{D}_y^c \end{pmatrix} \hat{\mathbf{z}}^*, \tag{B.12}$$

where $\hat{\mathbf{z}}$, $\mathbf{p}$, and $\mathbf{q}$ are starred to emphasize their assumed periodicity. Alternatively, the forward-differencing version of the normal equations reduce to:

$$\begin{pmatrix} \mathbf{D}_x^b & \mathbf{D}_y^b \end{pmatrix} \boldsymbol{\Gamma}_{\boldsymbol{\epsilon\epsilon}}^{-1} \begin{pmatrix} \mathbf{p}^* \\ \mathbf{q}^* \end{pmatrix} = \begin{pmatrix} \mathbf{D}_x^b & \mathbf{D}_y^b \end{pmatrix} \boldsymbol{\Gamma}_{\boldsymbol{\epsilon\epsilon}}^{-1} \begin{pmatrix} \mathbf{D}_x^f \\ \mathbf{D}_y^f \end{pmatrix} \hat{\mathbf{z}}^*, \tag{B.13}$$

which takes into account the relationship ${\mathbf{D}^f}^T = -\mathbf{D}^b$. Equivalently, backward-differencing exhibits the following normal equations:

$$\begin{pmatrix} \mathbf{D}_x^f & \mathbf{D}_y^f \end{pmatrix} \boldsymbol{\Gamma}_{\boldsymbol{\epsilon\epsilon}}^{-1} \begin{pmatrix} \mathbf{p}^* \\ \mathbf{q}^* \end{pmatrix} = \begin{pmatrix} \mathbf{D}_x^f & \mathbf{D}_y^f \end{pmatrix} \boldsymbol{\Gamma}_{\boldsymbol{\epsilon\epsilon}}^{-1} \begin{pmatrix} \mathbf{D}_x^b \\ \mathbf{D}_y^b \end{pmatrix} \hat{\mathbf{z}}^*. \tag{B.14}$$