

**2D CONTOUR SHAPE ANALYSIS FOR AUTOMATED HERRING ROE
QUALITY GRADING BY COMPUTER VISION**

By

D. Andrew Beatty

BSc Physics, University of Toronto

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES
DEPARTMENT OF COMPUTER SCIENCE

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

December 1993

© D. Andrew Beatty, 1993

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Computer Science
The University of British Columbia
2075 Wesbrook Place
Vancouver, Canada
V6T 1W5

Date:

Dec. 23, 1993

Abstract

A method has been developed to analyze the two-dimensional contour of naturally varying objects and to automatically assess the amount of shape variation that would be perceived by a person. Specifically, a software system was developed to assess the shape quality of herring roe sacs from a single overhead image. The shape difference from first grade to lower grade roe sacs calculated by the system matched well with that perceived by human roe graders, who select shape based on its appeal to the consumer. The principle motivation for this work is its application in the fishing industry, which requires both speed and robustness from the system, as well as grading accuracy. This software forms the basis of an actual industrial application, the prototype of which is currently under development.

Objects in the image are segmented using brightness thresholding and by ensuring they do not overlap. Morphological processing is used to filter out irrelevant shape features such as overlapping parasites and small broken pieces. The major principal axis of area serves as a reference axis for width measurements along the length of the shape. These measurements are the basis for deriving shape features. While this representation restricts the shapes to which the algorithm can be applied, it has worked well for roe sacs. Shape comparison involves taking the magnitude of the vector difference between the smoothed tangent angles at each measurement along the edges of each shape. This provides an effective measure of the human-perceived difference between the shapes, which can be used in combination with an internal database of acceptable shapes (either individuals or large data set clustering) to function as a shape assessment expert. In practice an internal database of 200 roe sacs has proven effective and sufficiently fast for

comparison. A final shape difference measure is arrived at by averaging the 3 smallest shape comparison values obtained when comparing a sample with every database sample.

In tests of sorting first from second grade roe sacs, the system has achieved overall accuracy of 81 ± 2 %. This compares with human grading accuracy, in which the entire three-dimensional shape of the sac is examined, of approximately 92%. The theoretical accuracy limit from two-dimensional contour information is estimated to be within a few per-cent of 85%. To be useful in practice, a grading approach is used in which only roe sacs which are determined to be first grade with high certainty are selected and the remainder are sorted by hand. Using this approach, the system can automatically select out half of a processor's bulk roe with a high (for the industry) accuracy of 97% for the selected sacs.

Table of Contents

ABSTRACT	ii
TABLE OF CONTENTS	iv
LIST OF FIGURES	vii
ACKNOWLEDGEMENTS	viii
1 Introduction	1
1.1 The Industry	1
1.2 The Product	1
1.3 The Task	2
1.4 The Approach	3
2 Background	5
2.1 Applications	5
2.2 Shape Representations	7
3 Method and Approach	10
3.1 Defining the Problem	10
3.1.1 What is Good Shape ?	10
3.1.2 Other Factors	13
3.2 Approaching The Problem	13
3.2.1 The Pilot System	13
3.2.2 System Overview	15
3.2.3 Hardware Subsystem	16

3.2.4	Incremental Development	16
3.3	Shape Representation	18
3.3.1	Choice of Representations	18
3.3.2	Alternative Representations	19
3.3.3	Disadvantages	20
4	Implementation	22
4.1	Image Processing	22
4.1.1	Lighting	22
4.1.2	Smoothing and Thresholding	24
4.2	Binary Image Processing	25
4.2.1	Configuration Labelling	25
4.2.2	Axes and Rotation	25
4.2.3	Morphological Processing	28
4.3	Measurement Extraction	29
4.3.1	Determining The Orientation	29
4.3.2	Taking the Measurements	31
4.4	Classification	32
4.4.1	Performing the Classification	32
4.4.2	Calculating the 'Perfect' Skein Database	33
4.5	Software Design	33
4.5.1	System Platform	33
4.5.2	Software Architecture	34
5	Results and Discussion	35
5.1	Measuring Results	35
5.2	Leave One Out Method	38

5.3	Binary Image Processing	39
5.4	Shape Comparison	39
5.5	Results	42
5.5.1	Two Classifier Accuracy	42
5.5.2	Marginal Grading Results	42
5.5.3	Grading Distributions	44
6	Conclusions	47
6.1	Work to Date	47
6.2	Future Modifications	48
A	Mathematical Background	50
A.1	The Neyman-Pearson Paradigm	50
A.2	Erosion and Dilation	50
B	Experimental Details	52
B.1	Expert Roe Grader Survey	52
B.2	Executable Program Descriptions	52
	Bibliography	57

List of Figures

1.1	A Female Herring with Roe Skein Superimposed	2
1.2	A first grade skein seen from one side	2
1.3	The same skein from the other side	2
1.4	System Overview Diagram	4
3.5	Some First and Second Grade Roe Skeins	11
3.6	Example of Factors Affecting Skein Quality	14
3.7	Computing Subsystem Diagram	17
4.8	The Image Processing Stages	23
4.9	A configuration contained in a bounding box	26
4.10	Stages in Measurement Extraction and Classification	30
4.11	Measurement Axes for the Configuration	31
5.12	Histograms of Difference Measures for Equal Numbers of First and Second Grade Skeins	36
5.13	Marginal Grading Results	45
5.14	Histograms of Difference Measures for Typical Numbers of First and Sec- ond Grade Skeins	46

Acknowledgements

Thanks is due to Ray Gosine for patience and guidance throughout as my supervisor, NSERC for providing funding for this work, BC Packers for providing additional funding and extensive support, David Lowe for his guidance as faculty supervisor, Lon Temereski for ensuring that I never get involved in middle management, Dale Cherchas for allowing his lab to be inundated with smelly fish products, Linda Cao for invaluable aid in collecting data, Murray Chan for being such a wild and craaaazy guy, The Industrial Automation Laboratory denizens for being such gracious losers in the game of minesweeper, Dan Li for ensuring that I wasn't the only one spending Saturday nights in the lab, Ming Wu for ensuring there was always an idle computer available, Jimi Yang for training me to work unhindered in loud construction or percussion drilling sites, Franco Bussani for personally demonstrating the vast benefits of regular Herbalife consumption, Lalith Gamage for his by-example lessons in Sri-Lankan politics, Whistler mountain for being there, and my family.

Chapter 1

Introduction

1.1 The Industry

British Columbia has a substantial herring roe industry which processes roe for a quality-conscious Japanese market. During 1992, approximately 20 companies processed herring roe in B.C., with over C\$100 million in export value. In common with many other agricultural products, export value for roe is determined by a subjective visual assessment of physical properties, such as shape, size and texture. In view of the high volume of herring processed during a short season, increased demand for accurate grading, foreign competition, and the development of high-speed roe extractors, the current labour-intensive approach to roe grading is an expensive bottleneck.

1.2 The Product

While the details of the herring roe preparation process are proprietary, in general the female herring are put through various stages of freezing and salting [Huynh 1984], after which the left and right egg sacs inside the fish (known as skeins) form a rigid mass. Figure 1.1 illustrates the location of a roe skein inside a herring. The skeins are then removed from the fish, further processed, and graded for sale. Figures 1.2 and 1.3 show typical views of a grade one roe skein. Since the colour of the skeins usually doesn't vary, grading depends almost exclusively on size and shape. The product is a delicacy in Japan, where the perfectly shaped skeins, which are classified as first grade, fetch a high

price. Typically, over 80% of the skeins are first grade, but price falls rapidly for second and third grade product. The graded skeins are packed in salty brine for shipping, and after arrival are partially desalinated, bleached, and eaten raw with soy sauce.

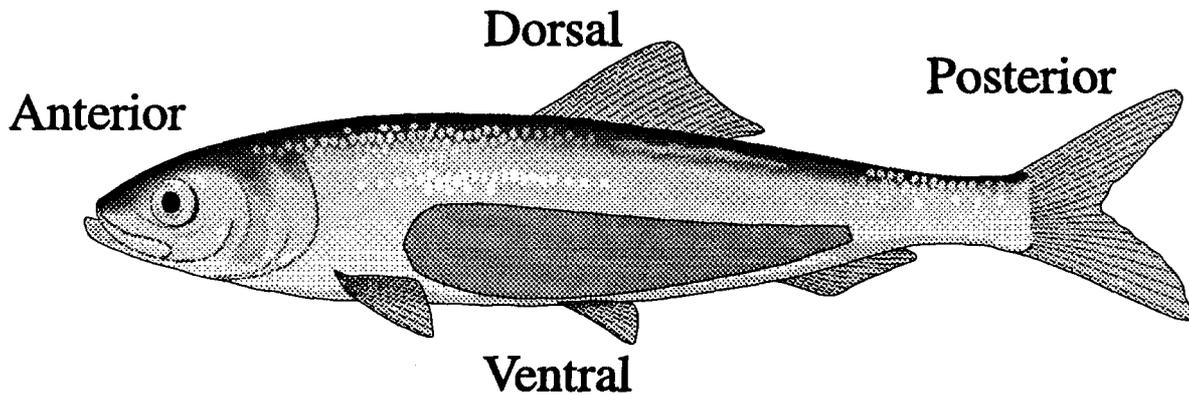


Figure 1.1: A Female Herring with Roe Skein Superimposed

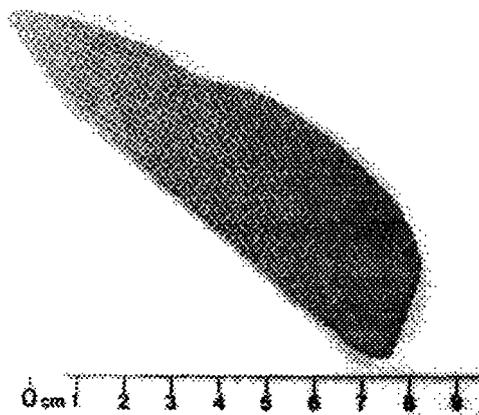


Figure 1.2: A first grade skein seen from one side

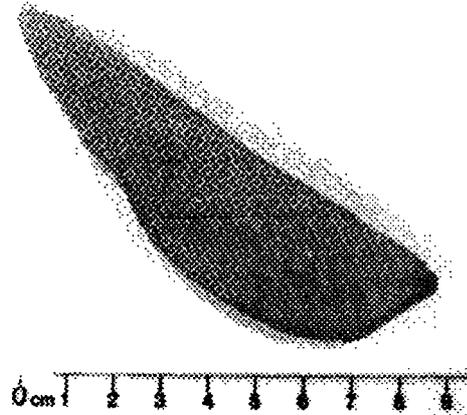


Figure 1.3: The same skein from the other side

1.3 The Task

While the grading process varies between processors, all roe is graded manually by human roe-grading experts. In high-accuracy, low-volume setups, each skein is examined

individually, while in high-volume lower-accuracy setups many skeins are scanned as they pass by the expert graders on a white assembly line. In most setups, the graders divide the skeins into first, second, and third grades.

The goal of this system is not to replace the human graders entirely, but instead to handle the bulk of the grading, leaving the more marginal cases to human expertise. The system is designed to identify only skeins that are first grade within a high degree of certainty, so they can be packaged without human inspection, leaving the remainder to human expertise. The purpose of this is twofold. First is a savings in labour costs. Second, by adjusting the degree of certainty, the accuracy of the automatically packaged samples can significantly exceed that of human grading in high-volume operations, since in high-volume operations it is impossible to find enough expert graders to examine each skein individually.

To remain economical, the speed at which the system can examine skeins must be at least two per second. Below this, the number of installations needed to handle the large bulk of roe becomes unwieldy.

A further point is that the quality and characteristics of the skeins can vary widely between catches made in different areas. It is advantageous for the processing company to apply different grading criteria to the different catches. As a result, the system must provide a way of adjusting itself to the specific qualities of different catches.

1.4 The Approach

The system overview is shown in Figure 1.4. The skeins must be organized into non-overlapping single file prior to passing beneath the visual sensor. The assessment of each skein is then based upon a single greyscale image taken from overhead. Although measuring the length of the skein is trivial, the quality is only dependent upon the (size

independent) shape of the specimen, given that it is above a certain minimum size.

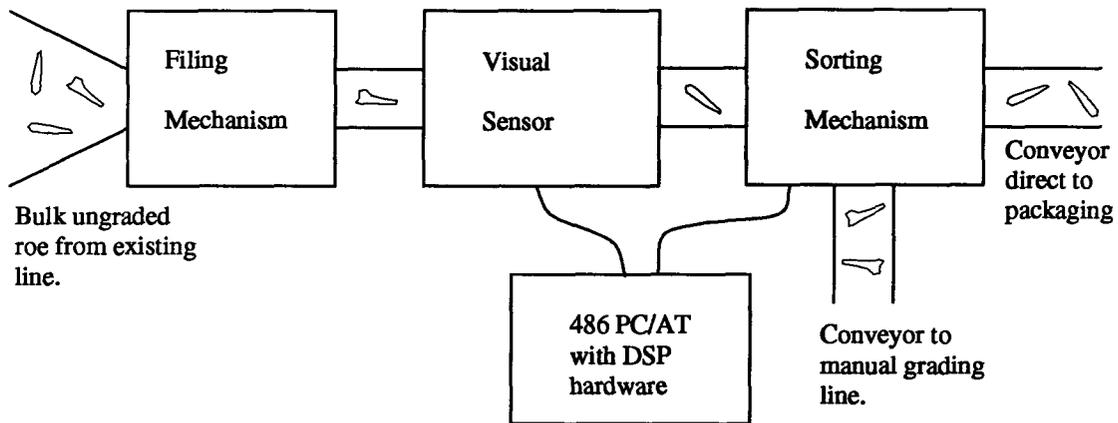


Figure 1.4: System Overview Diagram

The initial goal was to investigate the amount of information available from the contour of a single overhead image of each skein. During this investigation it became clear that an effective system could be based on this information alone. Hence the measurement extraction stage of the system produces a contour representation for the skein, based on a single greyscale image. This representation is derived from 100 width measurements taken along each side of the length of the skein, from the anterior end to the posterior.

A classification stage, which uses an internal database describing possible variations in the shape of a first grade skein, follows the measurement extraction stage. The difference between measured samples and the internal database of acceptable shape variations is calculated using an empirically developed algorithm, and a grade is assigned to the skein based on this difference measure.

Chapter 2

Background

2.1 Applications

Inspection of food products is a large application area for computer vision systems, and many vision-based products have been developed for industry. The technical details of most of these systems are not public. Of the publicly available research which has been done, the majority of actual working applications use only rudimentary computer vision, such as color spectral analysis and binary object moments. A range of agricultural applications of computer vision are described in [ASAE] and [BSRAE].

Work using only spectral analysis has been done to measure the quality of bell peppers [Shearer 1990(1)] and to detect defects in apples and peaches [Miller 1991], [Rehkugler 1989]. In all cases, the basic approach is to examine the brightness in two or three spectra, and use a simple pattern classifier to determine if the color is acceptable. In some cases this involves some segmentation of the image, which is handled by placing the product on an easily differentiable background.

Some applications involve some more spatial awareness, but still rely principally upon spectral analysis. Detection of ripe citrus fruit in trees [Molto 1992] involves some segmentation based on spectral features, and identification of plant species from images of canopies [Shearer 1990(2)] utilizes color texture. A combination of morphological operators and spectral analysis has been used to determine the cutting contour for a shrimp deheader with randomly oriented shrimp [Ling 1991]. Tellingly, however, the spectral

feature was the most reliable determiner of the proper cut position.

Work using binary object moments and morphology has been done to discriminate broken corn kernels [Zayas 1990]. This work is typical of much practical work in image analysis. Binary objects are segmented by thresholding, and area, perimeter, length, width and convex perimeter are features used for pattern classification.

A more complex analysis of corn kernels [Reid 1991] for the detection of cracks makes use of edge detection and Hough transforms. The kernel is a priori segmented by brightness contrast, and its orientation determined from its principle axes. A processing window is established over the central area of the kernel, and the detection of stress cracks proceeds from there. A similar approach used spectral analysis instead of edge detection in this window to analyze maize and soy kernels for quality [Paulsen 1989].

Rudimentary shape recognition is used in a system that automatically inspects potatoes for size and shape [Marchant 1990]. The potatoes are lined up in rollers as they pass beneath the camera, which means that their entire circumference can be viewed (over several images), and that they are touching horizontally. Potato and background are segmented by brightness, but neighbouring potatoes must be separated by the detection of the cusps in the collective contour. A degree of shape analysis is also used in oyster orientation [Tojeiro 1991]. The 2D shell contour is examined from above, the length axis is determined by the major principle axis of area, and the width is measured at a few strategic places to determine the orientation.

The inspiration for the representation described in this thesis came from an application described in [Arnason 1991], in which fish species were differentiated automatically. The fish was segmented a priori using thresholding, and its major principle axis of area determined. The fish was then divided into five sections along its length, and the average width over each section was extracted. Relating these five width measures provided sufficient shape discrimination for distinguishing species. The method presented herein

is significantly more complex, but owes its initial inspiration to this application.

2.2 Shape Representations

As long as they only need to work on test images, schemes for 2D contour representation and comparison abound. According to [Mokhtarian 1992] any shape representation should satisfy the following three criteria:

- **Invariance:** If two curves have the same shape, they should have the same representation.
- **Uniqueness:** If two curves do not have the same shape, they should have different representations.
- **Stability:** If two curves have a small shape difference, their representations should also have a small difference, and vice versa.

The reference to “small shape difference” is poorly defined. It could be defined as what humans perceive as the amount of shape difference, or more aptly it could be defined according to the requirements of each specific problem being solved. For the purpose of this thesis, the specific problem to be solved requires that the amount of shape difference correspond to the perceived shape difference by humans, which is reflected in a skein’s grade.

By this definition, contour parametric shape representations are not stable, and cannot be used as a basis for robust shape analysis. For example, consider the shape difference when a thin hair-like extension projects from the edge of a roe skein (as is the case with parasites). It would vastly change a contour length parametric description, but have almost no effect on the human perceived shape. Worse, the resulting shift of the

other features in the parameterization would destroy the comparison, or necessitate an expensive (in time) searching technique.

A representation such as the curvature scale space presented by [Mokhtarian 1992] which looks at contours in multiple scales would still not react appropriately to such a situation. The length of the hair-like extension would distort the underlying sac shape, regardless of how thin this extension is, because the representation is parameterized by contour length. Even a corrective approach such as smoothing without shrinkage [Lowe 1989] would not have the desired effect. The specific case of the parasite is used only to illustrate a point. In practice, isthmus-like shape features can be removed using morphological operators, but that still leaves open the question of whether a contour-parametric representation is a good choice in light of those shape features that are not quite thin enough or convex enough to be morphologically filtered out. Comparison of contour-length parameterized approaches with the axial-parameterized approach chosen is discussed in Section 3.3.2.

Another problem facing contour-length parameterized descriptions occurs when only part of the contour is available. This usually happens in one of two ways; a part of the object is missing or part of the object is obscured. There are ways to get around this, but they are expensive. Part of the difficulty of this matching is that the scale invariance usually obtained by normalizing the contour-length range of the measurements is no longer valid, since the occluding object, for instance, could have a much longer contour than the part it occludes. The Fourier-Mellin correlation described by [Franz 1991] attempts to address the problem. Although it can find such matches without a search, it is computationally expensive. More problematic, though, the smaller the section of contour which matches, the smaller the matching response, since the match of the entire contour is still being evaluated.

For these reasons, I will not address in more detail any contour length parameterized representations, such as the centroidal profile, cumulative angular, and curvature, which are outlined in [Bebis 1992], the autoregressive model approach [Dubois 1986], the hidden Markov model [He 1991], multiscale contour approximation [Bengtsson 1991], or arch height function [Lin 1992].

Another 'representation' is the Hough transform. Its main use is the recognition of simple shapes with a few degrees of freedom. This thesis does not try to recognize roe skeins, but instead to analyze them. Also, there are many degrees of freedom in the shape of a first grade roe skein, which, furthermore, are not even well-defined.

As mentioned in Section 2.2 the use of moments, contour length to area ratios, morphologically processed images and similar simple extracted features have often been used in practice for machine vision in the agriculture industry. While these methods are fast and effective, they are not informative enough to distinguish the subtle shape differences differentiating first and second grade roe skeins.

Chapter 3

Method and Approach

3.1 Defining the Problem

3.1.1 What is Good Shape ?

Figure 3.5 shows some first grade roe on the left and some second grade roe on the right. Factors affecting the grade are the overall shape, including thickness and twist as well as contour, broken ends or notches in the side, and rough variations on the surface. There is not only the question of how to weight all these factors, much less compute them, but how to objectively determine them.

To get some idea of what kind of sensory input might be necessary to adequately determine skein quality, a survey of two human expert roe graders was taken. They were asked to grade 200 samples of first and second grade skeins. They were then asked to grade both greyscale and thresholded pictures of the same skeins (in shuffled order). The results (See Appendix B.1) revealed that one might be able to get 80%-90% accuracy from the thresholded images alone, assuming a good enough algorithm (in fact, later results indicate that the accuracy limit is within a few per-cent of 85%). The reference to accuracy here is the two-classifier accuracy given equal numbers of first and second grade roe skeins of randomly chosen quality. This accuracy number is used throughout the project as a benchmark, and shall henceforth only be referred to as the two-classifier accuracy. For a detailed treatment, see Section 5.1.

Since a lot of grade information can be determined from the shape of the skein contour

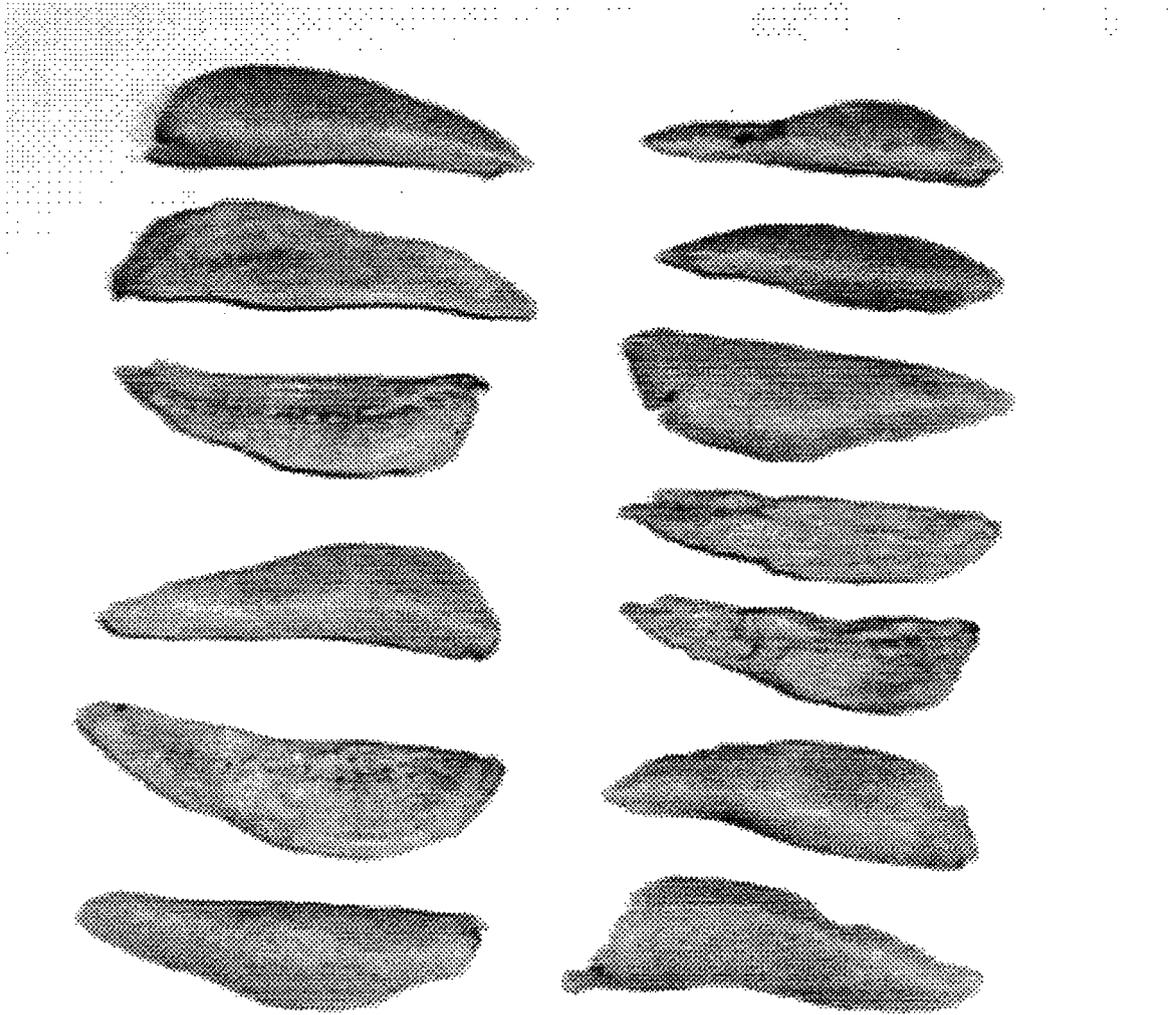


Figure 3.5: Some First and Second Grade Roe Skeins

alone, it was necessary to get some idea of exactly how the contour shape affects grade. Consultation with expert graders over time revealed several important aspects of contour shape. First, there is an ideal shape which all first grade skeins must approximate, and some types of deviations from this shape are far more deleterious to the grade than others. Below are listed some of the shape properties affecting the grade.

- General Shape (see Figure 3.5). There seems to be an ideal roe shape which is approximated by all first grade samples. There is some variation in where curves and cusps appear, and how pronounced they can be, but for simplicity I shall refer to a 'perfect' skein contour shape.
- Skein size. The grade is independent of size above a certain minimum size, below which it is automatically downgraded.
- Width. Shapes similar to the ideal but slightly wider or thinner are not penalized, except in the case of very thin pieces.
- Contour smoothness. Jagged contours are heavily penalized, as are gouges or a broken skein. The most common break occurs when the fragile, thin anterior end breaks off.
- Parasites (see Figure 3.6). Parasites or occasionally bits of gut tissue obstruct part of the contour. These are usually sharp and thin projections from the contour, and have no effect on the grade.
- Tail. The posterior tip of the skein, due to its fragility, varies widely in shape, and does not have much affect on the grade.
- Probiscus (see Figure 3.6). The dorsal side of the anterior end of the skein occasionally has an outcropping of variable length, which is common enough that it

does not affect the grade of the skein.

3.1.2 Other Factors

Since the success of the system depends on its ability to grade all samples, the grading algorithm must also be robust. Sensitivity to small overlapping debris such as broken pieces of other skeins or parasites is very undesirable.

In addition, the skein could be lying at any orientation and any position within the field of view of the camera. If it is lying partly off the edge of the image this must be detected, otherwise the position and orientation of the skein must be determined. The system must be able to identify the anterior and posterior ends of the skein, and identify the dorsal and ventral sides.

The final consideration is speed; the system must work fast enough to be useful. In the herring roe industry, this implies a throughput of two or three skeins per second, but even for development purposes a calculation time of greater than a couple of seconds would cause unacceptable delays when testing incremental improvements on large sample sets.

3.2 Approaching The Problem

3.2.1 The Pilot System

From the expert roe grader survey, we might optimistically expect that an 80% two-classifier accuracy could be achieved from the two-dimensional skein contours alone. This is too low an accuracy to be used as a replacement to human grading, which is roughly 90% in high-volume grading installations. However, we can use the information from a computerized system to reduce the amount of human grading that must be done as follows.

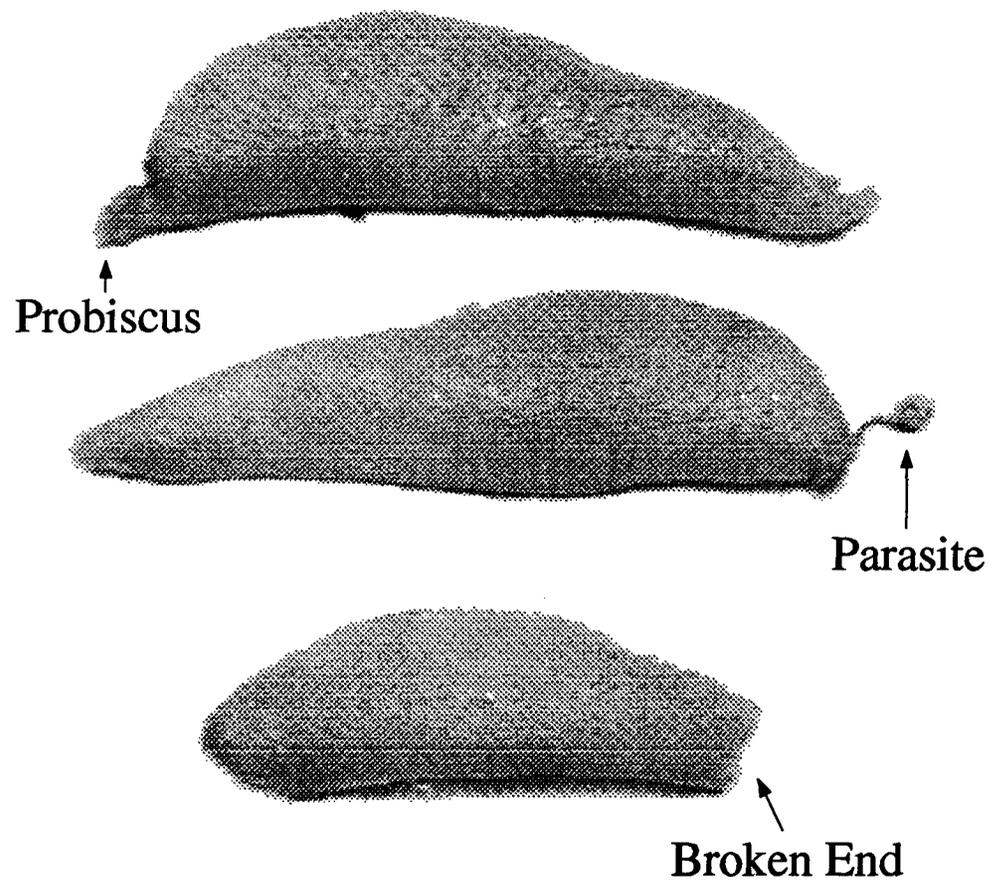


Figure 3.6: Example of Factors Affecting Skein Quality

Of all the incoming roe skeins which reach the current (human) grading installation, 80% are typically first grade, although this percentage may be as high as 90% and as low as 65%. The Neyman-Pearson paradigm for hypothesis testing provides the basis for developing a grading system that is accurate on a well-defined subset of the bulk roe, namely the first grade skeins. Details of the Neyman-Pearson paradigm are given in Appendix A.1. If we increase the power of the Neyman-Pearson test for first grade shape, we reduce the number of false acceptances, and increase the accuracy of first grade acceptances above the two-classifier accuracy. Those skeins not accepted as first grade would be a mix of first and second grade skeins and would still have to be graded by hand, but of course with a more accurate the grading system fewer skeins would be graded by hand. As will be discussed, the roe grading software yields a difference measure, representing the difference between a sample skein's contour shape and that of a first grade skein. By adjusting the acceptance threshold of this difference measure we can change the power of the Neyman-Pearson test.

This approach has been incorporated into the design of the 'pilot system'. The remainder of this chapter outlines the system as a whole, and details its development. The next two chapters offer a detailed description of the software process used to produce a difference measure.

3.2.2 System Overview

Figure 1.4 illustrates the conceptual design of the system as it would exist in a prototype installation. The system consists of a feeder, a conveyor, a diffuse lighting setup, a re-direction mechanism, and the vision system. The feeder must ensure that the skeins are in non-overlapping single file as they move onto the conveyor. The conveyor is white, to facilitate differentiation of the roe from the background, and the lighting is highly diffuse, to minimize shadow. As a skein passes beneath the camera, consecutive video-rate images

are examined to ensure the skein is fully within the video frame. Once such an image has been captured, analysis can proceed, and the result queued until the skein has travelled to the re-direction mechanism, where its path is accordingly directed. Skeins classified as first grade with the required degree of certainty are directed to automatic packaging, and the remaining skeins are processed by hand in the conventional manner.

3.2.3 Hardware Subsystem

The software was developed on an 80486-33 PC with a colour CCD video camera and a frame grabbing and image processing card, as shown in Figure 3.7. The camera was a JVC TK-1070U, used with a 16mm lens, and without any of the image pre-processing options enabled. The blue channel of the RGB output was used as the input to the frame grabber. The frame grabbing and image processing card was the Sharp GPB-1 (General Purpose Board), which consists of a video-rate greyscale frame grabber, 12 image memories (512 by 512 by 8 bits), and three special-purpose DSPs. The DSPs, in combination with accompanying driver software, supply a large number of image processing functions, many of which can be applied at video rates. The functions most relevant to this project were convolution, look-up table conversion, labelling, and erosion/dilation.

3.2.4 Incremental Development

As with any research project, development proceeded up several blind alleys along the way to the final incarnation of the system. These alleys include attempts to integrate depth information from stereo, linear methods of dimensionality reduction (principle components analysis), and many comparison algorithms. These efforts will be discussed along with the methods actually chosen, as they shed light on some of the decisions made.

After the project reached a certain stage evaluation of different options became an automatic part of the process. Once the system was mature enough to be applied to data

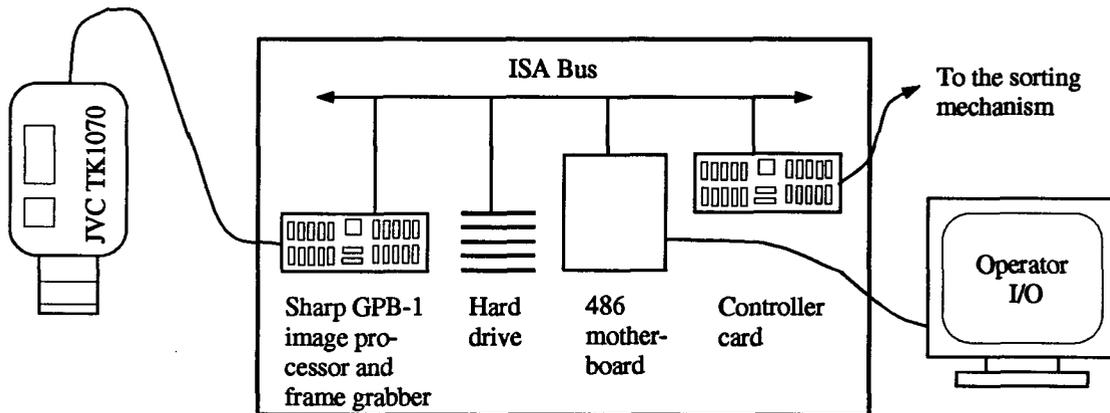


Figure 3.7: Computing Subsystem Diagram

sets without any preselection, a development cycle of error analysis and correction was applied. The system was written as a suite of batch programs which separated the functions of measurement extraction, dimensional reduction, and classification. This made it easy to expand the test set to a couple of thousand skein images, allowing for test results with a precision of two percentage points (in the two-classifier). In any experimental measurement, error is discussed in terms of precision and accuracy. Precision can be defined as the standard deviation of the measurement distribution if it were to be repeated many times, and accuracy can be defined as the systematic error in the measurement. For more discussion, see [Taylor 1982].

After each change to the system, the batch programs are run to automatically generate the test results for the new algorithm. These results include the two-classifier accuracy, a graph of the marginal grading capability (to be discussed), and diagnostic data files of the misgraded samples. If the new results are better than the old by greater than the precision of the measurements, (see Section 5.1) there is valid statistical evidence that the new algorithm is better than the old.

Depending upon the experimental results, new changes can be abandoned or incorporated into the system. Motivation for new changes can be gleaned from the data files of misgraded samples. These are read by the interactive version of the skein comparison program, which displays an image and detailed measurement graphs of each skein being compared. By examining these diagnostics, systematic sources of error can quickly be identified. An algorithmic fix can then be proposed and implemented, and the development cycle begins anew.

The details of this software are described in Section 4.5.

3.3 Shape Representation

3.3.1 Choice of Representations

An internal representation based on width measurements from the major principle axis of area was chosen [deSilva 1992] because it was felt that such a representation would maximize the robustness of the system. In Section 2.2 it was argued that contour length parametric representations were not robust; not because they actually throw away information about the contour, but because of the problems with the comparison methods that they necessitate.

Since all first grade skeins will definitely have distinct principle axes, there is a reliable starting point for all such measurements. By distinct principle axes, it is meant that the shape of the object have an unmistakable skew, so that small variation around the shape of an object will have little affect on the resulting calculation of the principle axes.

When a skein does not have distinct principle axes, it cannot be first grade, and the accurate starting point for measurements is not important, since whatever is extracted will almost certainly be downgraded.

One hundred width measurements are taken from the major principle axis of area

to each of the dorsal and ventral edges of the skein. These measurements are then smoothed and differentiated to yield 99 slope measurements for each side of the skein. The inverse tangent of each slope measurement is then calculated to yield 99 tangent angle measurements for each side of the skein.

The representation and method of shape comparison evolved through several stages, as discussed in Section 5.4, and the final method is to take the magnitude of the vector difference between the tangent angle representations.

The vector difference in tangent angles makes more sense for shape comparison than the vector difference in widths. To understand why, consider that it is not important whether a skein is a little fatter or skinnier than the average first grade skein, but it does matter if the contour varies from thick to thin over a region where first grade skein contours are generally flat.

3.3.2 Alternative Representations

In Section 2.2 it was argued that contour length parameterized representations tended to obscure shape information relevant to the human-perceived shape differences. For this reason, and the computational expense of extraction and comparison for these methods, a contour length parameterized approach was not used.

To understand why an axial parameterized representation might be more robust, consider shape variations which are small in terms of an observer's measure of shape difference. The variation between the two samples could occur at any number of points along the contour. Each such variation will cause some alignment shift in the measurement comparison using a contour-length parametric representation, having the undesirable effect of mismatching the entire comparison to some extent. To correct this with a search is very time expensive since there are many places along the contour in which the samples could differ, and hence many dimensions to a matching search.

The axial-parameterized representation might be considered to have only four degrees of freedom of misalignment. Alignment at the anterior and posterior ends of the skein, and alignment of the angle and offset (transverse to the major principle axis) of the measured principle axis. Misalignment at the ends of the skein is minimized by morphological processing methods discussed in Section 4.2.3. Misalignment of the angle and transverse offset of the major principle axis are certainly a possibility. In the example given in Section 2.2, the effect of a thin filament extending from the edge of the configuration was considered. It is clear that although the alignment of a contour-length parameterized representation is significantly affected in this case, it will have relatively little effect on the axial-parameterized representation alignment, due to the small amount of area involved.

A simpler representation based on moments, contour length ratios, curvature variation, and other longstanding features used in machine vision (See Section 2.2) was decided against since it was too simplistic to preserve the shape quality information.

3.3.3 Disadvantages

The main problem with the method is proper alignment of the measurements. There are three cases to consider here. First, the presence of the probiscus, described in Section 3.1.1, might cause misalignment of measurements between two skeins at the anterior end. Second, the variation of tail length, also described in Section 3.1.1, might cause similar misalignment at the posterior end. Finally, the shape features along the length of the skeins may not always be in the same relative position. As will be discussed in Section 5.3, the anterior and posterior end alignment problems are largely addressed by the use of morphological processing, and any remaining alignment problems are circumvented by having a large internal database of allowable skein shape variations.

The other problem is more intrinsic; the inaccuracy in the representation of contour

sections which are close to perpendicular to the major principle axis of area due to the parameterization of the width measurements. Fortunately, first grade roe skein contours only have this problem at the very anterior end, and it has not proven very important.

Chapter 4

Implementation

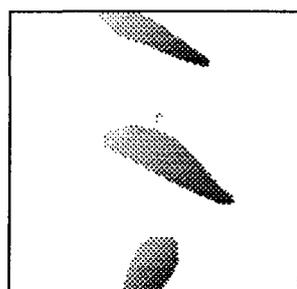
4.1 Image Processing

The image processing stages turned out to be of great importance in this project. Changes made to the image processing stages during development and testing made significant improvements to the performance, whereas the measurement extraction stage remained largely unchanged from its conception. An overview of the image processing stages is given in Figure 4.8.

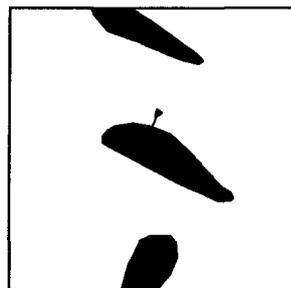
4.1.1 Lighting

According to [Zuech 1988] and many others, one of the basic requirements of a practical machine vision installation is proper lighting. Since the roe skeins are normally carried along a white conveyor belt, segmentation of the skein by thresholding was a logical place to start. Thresholding, however, is very sensitive to shadows and changing light levels across an image. In order to minimize both of these factors, a diffuse light source (actually, many sources) was used. Also, to avoid varying light levels caused by line frequency, only incandescent light was used.

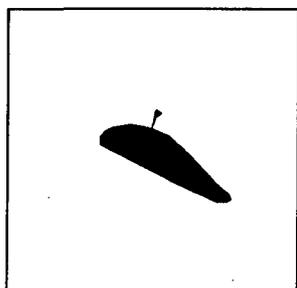
Since roe skeins are yellow-orange in color, their light absorption is highest in the blue spectrum. Hence to maximize the discrimination of roe skeins from background, the greyscale images are taken in the blue light spectrum.



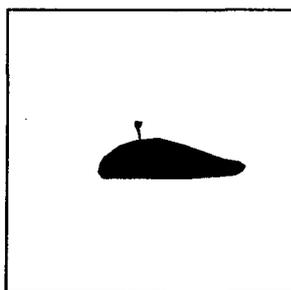
Step 1: Initial Image



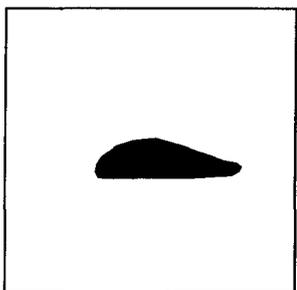
Step 2: Thresholded



Step 3: Object Selected



Step 4: Object Rotated



Step 5: Morphologically Filtered

Figure 4.8: The Image Processing Stages

4.1.2 Smoothing and Thresholding

After thresholding, connected binary areas in the image are labelled. If no smoothing is done, there are occasionally noisy areas in which there are a lot of unconnected small areas (mostly single dots), which run the risk of overflowing the labelling process. Since the labelling process takes place on the image processing hardware and cannot easily be modified, and the best course of action is to avoid overflow altogether. To this end some smoothing of the image must be done either before or after the thresholding.

A visual comparison of smoothing before (using a Gaussian filter approximation) and smoothing after the threshold and then re-thresholding (effectively an erosion operation that only erodes points with few neighbours) revealed that the former method was far more effective both for reducing binary noise and not affecting the finer resolution of the areas of interest.

One consequence of smoothing is the potential systematic effect of fixed scale smoothing on roe of different sizes. As discussed in Section 3.1.1, the grade of a roe skein is size independent above a certain minimum size, hence the risk of more relative smoothing of shape features for smaller skeins. This is not a significant problem, as it turns out, for two reasons. First, the amount of smoothing is small relative to the size of detail examined by the system. Second, first grade skein sizes only vary by at most a factor of three in length.

The threshold level is set during an initial calibration stage. Several automatic threshold algorithms were tried, but none provided a robust method to determine the optimum threshold level. The posterior tips of the skeins are often thin and lightly tinted, and in the extreme cases this necessitates a trade-off between losing the tip of the skein or gaining some shadow in the threshold. This is not significantly detrimental to the system for two reasons. First, with appropriate lighting and carefully set threshold, the extreme

case seldom happens. Second, the anterior tip of the skein is not particularly important, to be discussed later.

4.2 Binary Image Processing

4.2.1 Configuration Labelling

We borrow the terminology from [Sharp Inc. 1991] and refer to a connected region of the binary image (whether four or eight-connected) as a 'configuration'. For this system, four-connectedness was used, but eight-connectedness would likely work just as well.

Once the configurations in the image are labelled, their areas are measured, and significant configurations are identified by a significant (say greater than a square centimeter) area. In a prototype implementation, those configurations with significant area would be tracked and examined in sequence. For the purposes of this pilot system, we assume the largest configuration is the object of interest and examine it. Now that the configuration of interest has been identified, it is easily extracted by throwing away all other configurations.

The bounding box of a configuration is the upper and lower limits of the X and Y components of position over the area of the configuration. The bounding box of the isolated configuration is determined, and for efficiency further processing is restricted to the confines of the bounding box. Figure 4.9 illustrates a configuration of a skein which has been isolated in a bounding box.

4.2.2 Axes and Rotation

Now that the configuration of interest has been identified, its orientation (ie the angle of the principle axis of area) must be calculated, and for future processing it must be rotated to the horizontal. The first moments, $\sum x$ and $\sum y$ of the configuration are calculated,

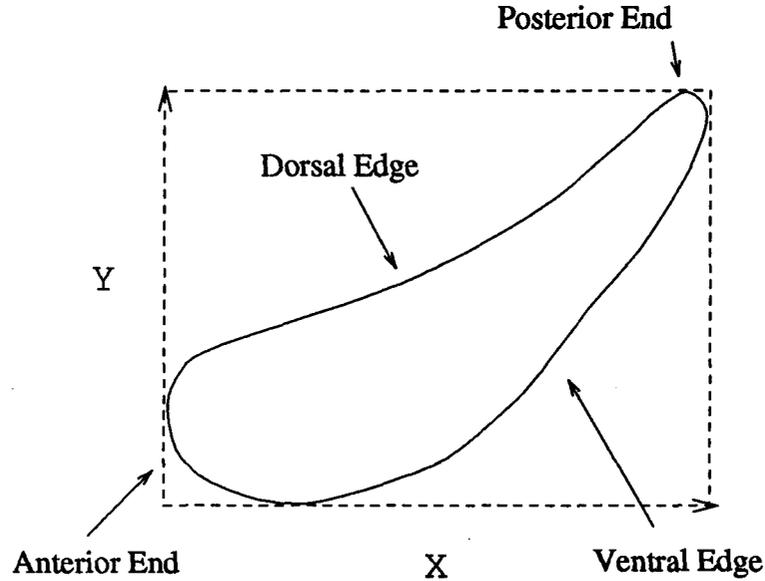


Figure 4.9: A configuration contained in a bounding box

which yield the centre of area (\bar{x}, \bar{y}) , where x and y are the axes of the bounding box as shown in Figure 4.9. Next the second moments are calculated as:

$$M_{x^2} = \sum (x - \bar{x})^2$$

$$M_{y^2} = \sum (y - \bar{y})^2$$

$$M_{xy} = \sum (x - \bar{x})(y - \bar{y})$$

From these, we get the angle of the major principle axis of area, θ , by:

$$\theta = \frac{1}{2} \tan^{-1} \left(\frac{M_{xy}}{M_{x^2} - M_{y^2}} \right) \quad (4.1)$$

A derivation of this equation is given in [Horn 1986], Section 3.2.2. Figure 4.11 shows a configuration with the centre of area and principle axes identified. When $M_{x^2} - M_{y^2}$ approaches zero, θ in Equation 4.1 approaches $\frac{\pi}{2}$. When both $M_{x^2} - M_{y^2}$ and M_{xy} approach zero, no distinct principle axis of area exists, and the skein can be rejected immediately. The moments required to calculate θ are quickly summed by the GPB

hardware, but the board cannot rotate an image, so the rotation must be done in main memory. The binary contents of the bounding box are transferred to main memory where the rotation can take place. A fast algorithm is used, taking advantage of the knowledge that we are operating on a binary image and will re-threshold the image afterward. Since we do not yet concern ourselves with which end is the anterior, the following rotation matrix can always be used to make the principle axis of the configuration horizontal:

$$\begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

The contents of the bounding box are rotated around the centre of the bounding box. Since the furthest pixels in the bounding box from the centre of rotation are at the corners, we make the rotated box a square with sides of length $\sqrt{l_x^2 + l_y^2}$, where l_x and l_y are the width and height of the original bounding box respectively. Using this dimension for the rotated box guarantees that all pixels in the original bounding box will fall within the rotated box.

The complications in rotation arise because of the rasterized discretization of the image. The algorithm used scans through the source image bounding box. For each pixel in the bounding box, the position of its centre, relative to the centre of the bounding box, is rotated using the matrix shown above. The new position is now considered relative to the centre of the rotated box. The distance to the centres of the closest pixels in the rotated box are calculated. If this distance in pixels, d , is less than 1, then a $1 - d$ weight is added to the raster value of that destination pixel in the rotated box.

Once the entire source bounding box has been scanned in this way, the destination rotated box is loaded back to the GPB board and thresholded around the value 0.5 (Actually an integer equivalent to this weighting scheme is used).

4.2.3 Morphological Processing

In concept, this stage should precede the last (rotation) stage. However, due to the hardware limits of the GPB board, the erosion window is only a three by three matrix, and hence not isotropic. A description of the erosion function is given in Appendix A.2.

Now that the configuration is rotated to the horizontal, the effects of erosion and dilation will be similar on similar configurations, regardless of the initial orientation. The pilot system erodes the image, then dilates it an equivalent amount, to remove irrelevant and misleading small-scale shape features. For a discussion of the effects of the erosion/dilation, see Section 5.3.

A problem which arose in practice was that of background pixels appearing in the configuration. Since the roe is moist, even in diffuse light some pixels within the configuration may be bright enough to be thresholded as background. If ignored, these pixels will then act as a seed for erosion, and if the perimeter of that erosion meets the erosion from the real contour of the skein, the shape is irretrievably distorted. To prevent this from happening, it was necessary to label all connected background valued pixel configurations, and map all but the largest such configuration (the actual background) to the foreground.

The other problem that arose was when a fairly large broken piece or parasite was thinly connected to the configuration. In this case, the erosion stage would disconnect the two, but would not fully remove the unwanted parts. In order to prevent the re-expansion of these parts during the subsequent dilation stage, the image is relabelled after erosion and the largest configuration, if there are multiple configurations present, is extracted.

The optimal amount of erosion/dilation was determined by repeated adjustment and retesting using the batch software configuration described in Section 3.2.4.

This ends the image processing and binary image processing stages. The processed image is now loaded into main memory for all the remaining calculations described below.

4.3 Measurement Extraction

An overview of the measurement extraction and classification stages is shown in Figure 4.10. Note that classification (section 4.4 below) actually refers to two steps; calculation of a difference measure, followed by classification based on that difference measure and a classification threshold.

4.3.1 Determining The Orientation

After the measurement extraction stage, an internal shape representation is generated consisting of one hundred width measures of the dorsal side of the skein along the length of the major principle axis from the anterior to the posterior end, and similarly one hundred width measures along the ventral side. In order to do this, the anterior end and the dorsal side of the configuration must be identified as illustrated in Figure 4.9.

Since the anterior end is fatter than the posterior, it is reliably identified by noting which end is closer to the centre of area. This is done after morphological processing since thin filaments such as parasites could otherwise invalidate this test. In practice, this test has been 100% accurate on the approximately 2500 roe skeins tested.

Since the skeins are curved along their length, the dorsal side can reliably be determined by summing the value $\sum x^2y$ over the configuration, where x is the major principle axis of area, and y is the minor, as shown in Figure 4.11. Again, for robustness, this test is applied after morphological processing, and it only fails about 0.2% of the time on first grade skeins.

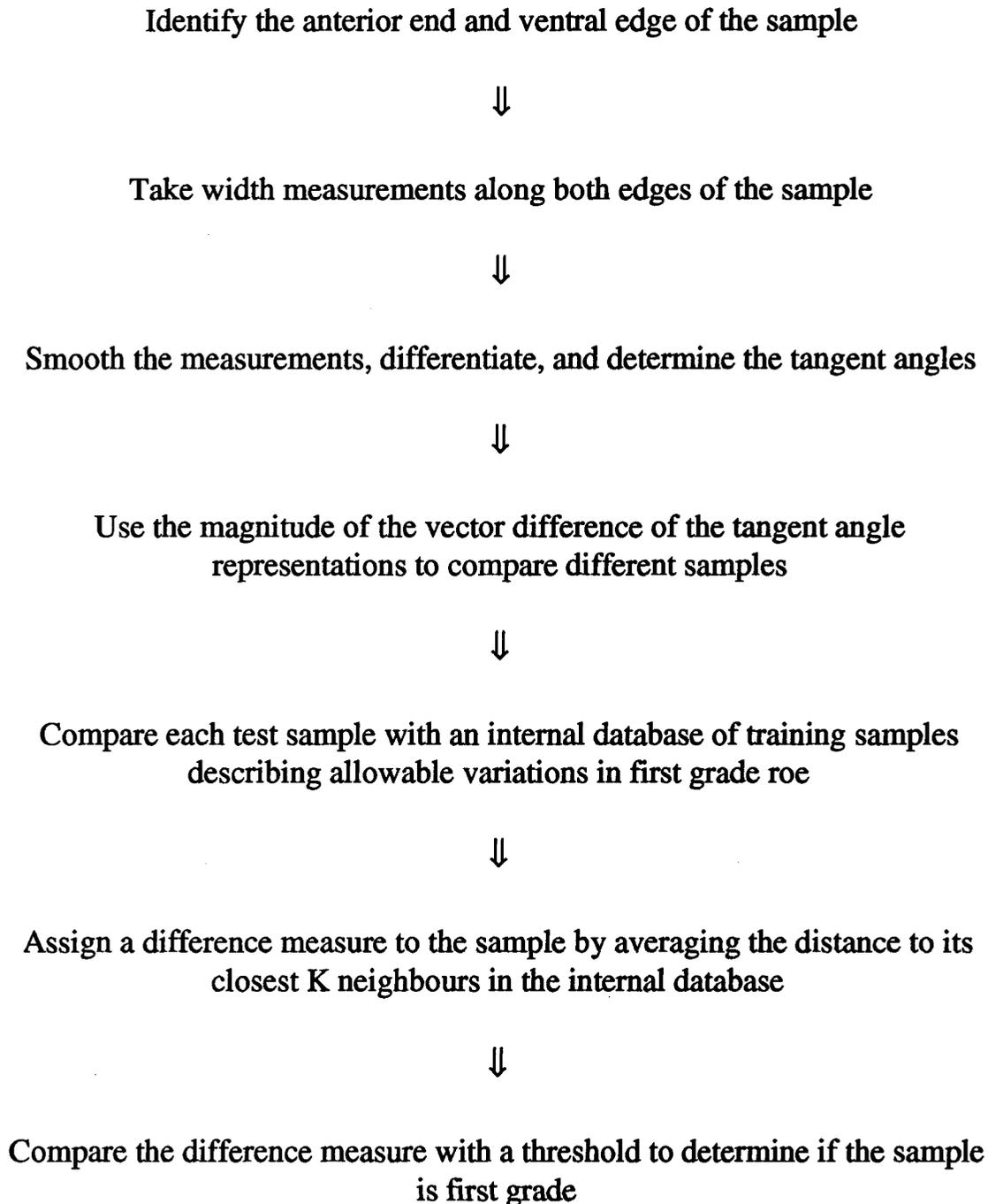


Figure 4.10: Stages in Measurement Extraction and Classification

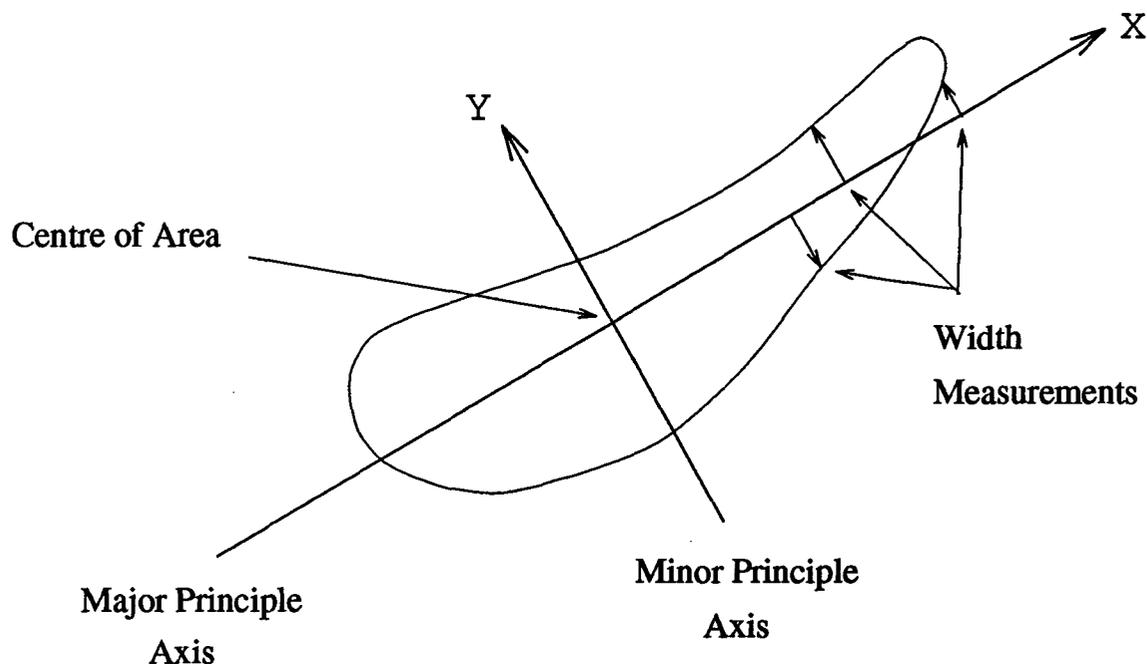


Figure 4.11: Measurement Axes for the Configuration

4.3.2 Taking the Measurements

Since the length of the configuration in pixels is not likely to be an even multiple of 100 (the number of width measures in the extracted representation), the averaging of the widths at each pixel value for x must be interpolated. The algorithm counts through the pixels along the x axis. For each pixel that lies fully within one of the 100 measures, the width of the configuration at that particular pixel column is added to the width count for that measure. For pixels that lie over a boundary of the 100 measures, the pixel column width is weighted according to the area of overlap before being added to the appropriate width count. The width counts are all weighted by $\frac{100}{l}$, where l is the configuration length in pixels. This yields a set of 100 interpolated average width measures.

The length of the skein in pixels is also recorded, and the width measurements are re-expressed proportionately to length. The length is then re-expressed in inches, using

a calibration constant determined in an off-line calibration phase. Skeins which are too small to be first grade are automatically rejected at this point.

4.4 Classification

4.4.1 Performing the Classification

Each skein is represented by 200 width measures, a dimensionality far too high to successfully approach with general pattern recognition techniques. The standard approach would be to try and reduce the dimensionality while preserving the quality information, but this is not likely to work well in this case since shape quality is not well-defined, and it is very unlikely that shape quality will just happen to map well onto simple features such as average and variance of curvature. Instead, a comparison technique is used that does not require dimensionality reduction.

First, the width measures are smoothed by convolution with a Gaussian approximation with $\sigma = 2.0$ (width measures), another empirically determined number. Note that smoothing at this stage is independent of the size of the skein, so no bias is introduced.

After smoothing the measurements are differentiated, however the order of these operations is irrelevant since the operators commute. The resulting 99 smoothed slope measures of each side of the configuration are converted into tangent angles by taking the inverse tangent. This tangent angle representation shall be referred to as the 'final representation'.

In the initial pilot system, there was a single internal representation of a 'perfect' skein, which was the average set of extracted measurements of many first grade skeins. Each sample to be graded was compared with this 'perfect' skein, and a difference measure was evaluated. The difference measure was calculated as the magnitude of the vector difference between the final representations of the sample and the 'perfect' skein.

In its final form, the pilot system uses not one, but many internal 'perfect' skeins. This constitutes a database of allowable variations in the skein contour shape. A difference measure is calculated as above, for each 'perfect' skein in the database, and the lowest k difference measures are averaged to yield the final difference measure. In practice, the most (empirically) effective number was $k = 3$, with an internal database of two hundred samples (a number chosen because it did not adversely affect the speed of the algorithm).

4.4.2 Calculating the 'Perfect' Skein Database

In the final version of the pilot system, the database of allowable first grade skeins is picked either at random, or by hand (it makes little difference) from first grade samples. Since each database sample represents an individual skein, this makes it convenient to apply a leave-one-out test to a large data set, which is what is currently done.

In the interest of maintaining an acceptable grading rate, there is only time to compare the final representation of a skein with a couple of hundred others. In order to try and distill the information of a large training set, say a couple of thousand first grade skeins, it is necessary to find a subset representation of vectors which best characterize this training set. To do this, automatic clustering of the full database is the best option, and [Duda 1973] offers several well known automatic clustering algorithms. The issue of clustering was not pursued in this thesis due to the lack of samples available for study.

4.5 Software Design

4.5.1 System Platform

The software was designed to run from a DOS platform. This choice was mandated by the driver software for the GPB-1 board, but is nonetheless a good choice since it avoids the overhead of a modern operating system, which is not necessary for a dedicated system

such as this one. In addition, PC-AT computers are a very cost-effective computing platform. Since the memory and speed requirements of the system were large, it was necessary to use a 32 bit memory model. The choice of compiler and DOS extender was again mandated by availability of driver software for the Sharp card. The Metaware High C/C++ compiler version 3.0 was used to compile the source code to object files (optimized for the 486 CPU) and the Pharlap linker version 4.0 was used to link the object files and libraries to form the executables.

For small ancillary programs, for which 32 bit executables were not necessary, the Microsoft C compiler version 6.0 was used.

4.5.2 Software Architecture

The project was implemented as a set of executable tools, communicating via files. This allowed for easy experimentation by altering only the functions under consideration, and re-processing the same intermediate data, either interactively for error examination, or in batches for performance evaluation. Not until the end of the entire project was an executable constructed which could perform all the grading functions at once, interactively.

Descriptions of the executables are given in Appendix B.2, the organization of which revolves around the files which they operate on. The first of these file types is the greyscale image. Using background thresholding and Stacker ©, we were able to store about three thousand sample images on a 200MB drive. The next file type contains the raw extracted width measurements, length in inches, and greyscale filename (if appropriate) of arbitrary numbers of skeins. The final file type contains lists of floating point numbers corresponding to final difference measures calculated for a group of skeins, using some comparison method. The difference measure files are then used to automatically generate the two-classifier accuracy and the marginal grading results chart.

Chapter 5

Results and Discussion

5.1 Measuring Results

Results were measured in terms of the two-classifier accuracy, which we define here as the maximum accuracy achieved over different settings of the power of the test (see Appendix A.1), when classifying equal numbers of first and second grade skeins taken randomly from ungraded bulk roe. Since the properties of the roe vary with catch area (called 'lots' in industry), a lot with the lowest quality skeins was chosen. The lowest quality skeins have the property of being the most difficult to grade properly, and hence our two-classifier accuracy should represent a lower bound over all the lots.

The power of the test is indirectly controlled by adjusting the threshold level below which a difference measure is deemed to represent a first grade skein of roe. The two-classifier accuracy is achieved when the threshold level is set such that the likelihood of rejecting a first grade skein is equal to that of accepting a second grade skein. When dealing with equal numbers of first and second grade skeins, this will occur where the histograms of their difference measures intersect, as shown in Figure 5.12.

To calculate the two-classifier accuracy, a number of first and second grade skeins (picked randomly) are graded. Denote the fraction of first grade skeins rejected (ie. considered not first grade) by the system as α , the significance of the test. Clearly the fraction accepted is $1 - \alpha$. Denote the fraction of second grade skeins accepted (as first grade) by β , then the power of the test is defined as $1 - \beta$. Now the two-classifier accuracy

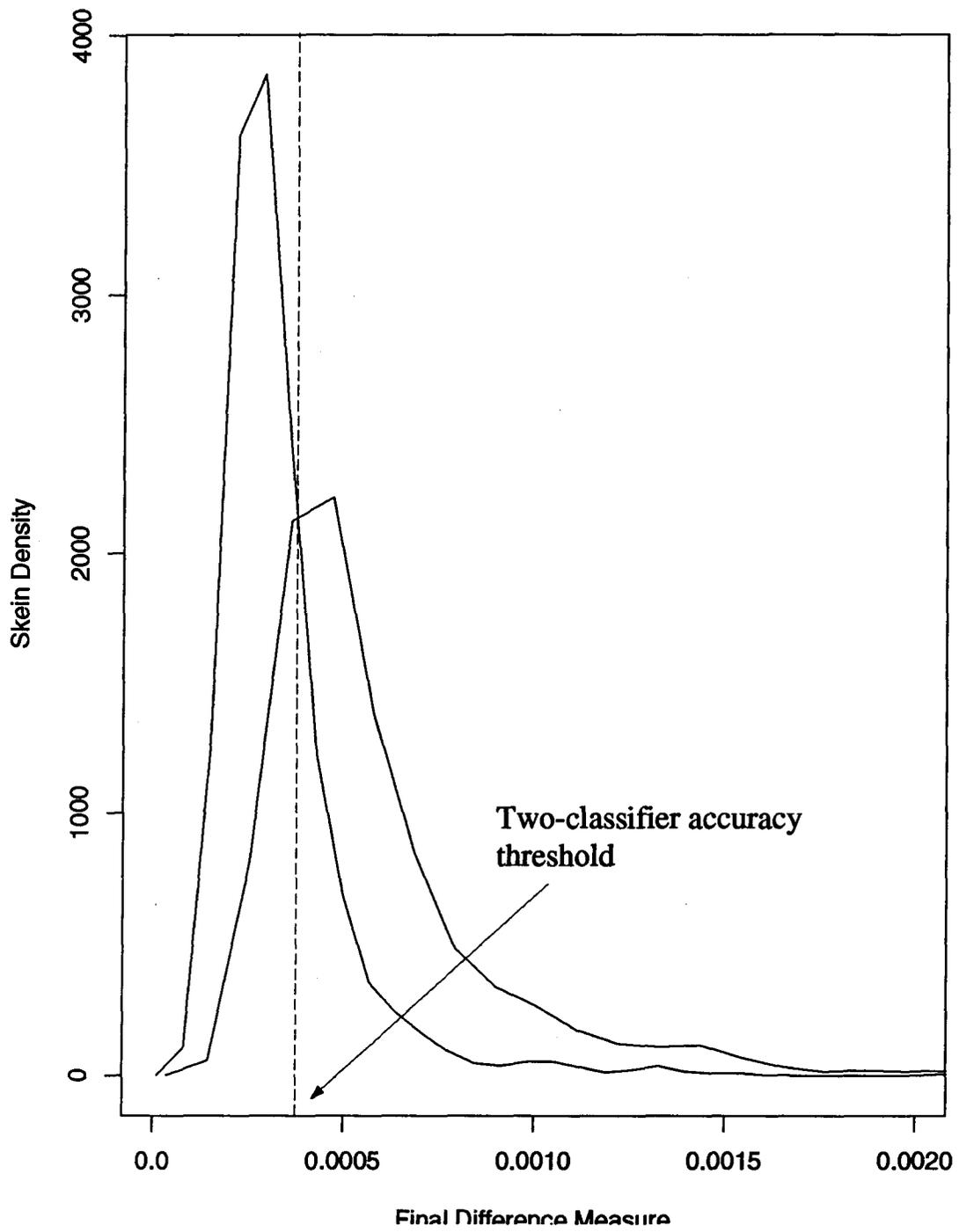


Figure 5.12: Histograms of Difference Measures for Equal Numbers of First and Second Grade Skeins

is:

$$A = 1 - \frac{\alpha + \beta}{2} \quad (5.2)$$

Given a number of roe of the same actual grade, n , which are to be assigned a grade by the system, the probability distributions of the number of skeins classified as first and as second grade are binomial distributions, given by:

$$P(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Where k is the number of correctly graded skeins (of the n total skeins), p is the probability that each skein be correctly graded, and $P(k)$ is the probability that k skeins will be correctly graded. It is well known [Rice 1988] that the variance of $P(k)$ is given by:

$$\sigma^2 = np(1-p) \quad (5.3)$$

The performance tests were done with 760 skeins of first grade and 760 of second grade. Since approximately 80% of skeins are correctly graded by the system p is about 0.8. This is true for both first and second grade skeins, so both grading distributions will have the same variance. To calculate this variance, we substitute $p = 0.8$ into Equation 5.3:

$$\sigma^2 = 760 \times 0.8 \times 0.2 \simeq 122$$

$$\sigma \simeq 11$$

So the relative precision in the measured number of correctly graded skeins is:

$$\frac{\sigma}{n} \simeq 0.014$$

Since the two-classifier accuracy is a function of the grading accuracy of the first and second grade skeins (at a given power of the Neyman-Pearson test for first grade), we can apply some basic experimental error analysis to derive the precision of our measurement for the two-classifier accuracy. An introduction to experimental error analysis is [Taylor 1982]. From the above we know that the relative precision in both α and β is 0.014. From Equation 5.2 we can calculate the precision in the two-classifier accuracy as:

$$\sigma_A = 0.014 \times \sqrt{2} \simeq 0.02$$

Now if two two-classifier accuracy measures are compared, in order to analyze the relative performance of some algorithmic adjustment, the difference between the two will have a precision of:

$$\sigma_{A_1-A_2} = 0.014 \times 2 \simeq 0.03$$

So there is a full 3% standard deviation in any comparison of two-classifier accuracies, with our test data.

5.2 Leave One Out Method

In order to get the aforementioned precision, the system must be tested on all samples. Since the system works by using an internal database of perfect roe, care must be taken to avoid testing on any skeins which have been used in the training process. To do this, the leave-one-out method (sometimes known as cross-validation) was used. For the test results given in this thesis, the first grade training set was created by randomly selecting 200 first grade skein measurements. Then, for each skein to be analyzed, a check is made to see if that skein is in the training set. If it is, then its training instance is ignored, and the identical grading process is applied with a training set of 199 skeins.

5.3 Binary Image Processing

One effective change made during the development of the algorithm was the inclusion of the erosion/dilation stage to remove extraneous contour information. This followed an error analysis that identified several related problems, and had three desired effects. First, it cut off most of the probiscus, if present, largely removing the problem of measurement alignment at the anterior end, as discussed in Section 3.3.3. Second, it cut off the tail at a certain thickness, alleviating the same measurement alignment problem for the posterior end. Finally, it helped to remove configuration pixels made by parasites or small broken fragments which fall across the contour of the skein. If the broken fragment is fat enough, however, it still leaves a residual interruption of the contour, so there is a trade-off, in the amount of erosion/dilation, between removing larger extraneous shape features and removing relevant shape features.

Conceptually, the erosion/dilation can be thought of as a squelch on the binary object scale. Features with high contour concavity are thrown away without affecting larger-scale features of the shape.

5.4 Shape Comparison

When the dimensionality of a feature space is as large as 200, traditional pattern classification methods of partitioning the space are not feasible. Hence the internal database and shape comparison method described earlier. The key to success for this approach is a comparison method which yields a reasonable estimate of the similarity in shape. Over the course of development of the system, many variations on this comparison method were tried, a few of which are described below.

- Initially, comparisons were made between the current method and the direct method of taking the magnitude of the vector difference between the raw width measurements. For reasons postulated in Section 3.3.1, this method offered distinctly poorer performance.
- An attempt was made to reduce the dimensionality of the measurements using principle components analysis, described in [Fukunaga 1990]. The basic idea is to find a few axes within the 200 dimensional measurement representation which best encapsulate the variation in some sample set. The axes are picked subject to the constraint of minimizing the loss of information when the sample set is projected onto these axes. The hope is that the dimensions encapsulating the greatest variance will also encapsulate most of the shape quality information.

This method was used to reduce the number of dimensions to 3, 5 and 10. In all cases grading using k nearest-neighbour techniques was very poor. This is not surprising in retrospect, since the blind application of a linear dimensionality reduction technique is not likely to work well for a complex notion like similarity in shape.

- A later comparison method used was to take the magnitude of the vector difference between the tangent slopes. This was used for a while until the error sourcing process revealed that contour sections with high slope were being unfairly weighted, by virtue of the fact that slopes became very large as their corresponding tangents approach the vertical. The remedy was to re-express these values as tangent angles by taking the arc-tangent of the slope values. This improved the two-classifier accuracy by a significant 5%.

- Before the inclusion of an internal database of shape variations, a single 'perfect' skein shape was used to compare all samples with [Beatty 1993]. This 'perfect' skein was actually the average of select first grade skeins of the highest quality. With this internal model, a comparison method using variable alignment of the anterior end of the sample was applied. This was achieved by removing successive measurements from the anterior end of the sample measurement (with the assumption that we may be removing a probiscus). Each time a measurement is dropped from the anterior end, a new vector is (conceptually) created by interpolating the remaining measures into a final representation of 99 measurements. This vector is then used to create a difference measure against the 'perfect' skein as described previously. A gradient descent of these difference measures is followed, until a local minimum is reached. At this point it is assumed that any probiscus has been optimally removed, and the local minimum becomes the final difference measure. This functionality was aimed mainly at eliminating the probiscus problem, and the results were a measurable improvement of about 3% in the two-classifier accuracy. Even after the expansion of the internal model to encompass many first grade skein samples, this dynamic alignment improved results, but after the inclusion of the morphological processing, which takes care of most of the alignment problems (for reasons discussed in Section 5.3), this method ceased to improve performance, and was not incorporated into later versions of the pilot system.
- Another adjustment to the comparison method was to consider each skein as two separate halves. Then the (squared) magnitude of the vector difference for each half could be compared with the internal database separately, effectively squaring the size of the internal database. This was done by simply dividing the width measurements into two edge sets of fifty measures per side. The difference measure

of each half is calculated against the internal database separately, then they are added to form a final difference measure. The success of this method, however, relies the two halves of the roe contributing to the entire shape quality independently, which was clearly not the case, since the results were not measurably improved.

- Rather than weight each vector component identically by taking the magnitude of the vector difference between the measurement values (tangent angles) for two skeins, the Mahalanobis distance [Duda 1973], [Fukunaga 1990] was used. In this method, the difference between each pair of vector components is weighted inversely to the variance of that component over some training set. In other words, each component is weighted in terms of its statistical significance. This method yielded no improvement in performance, indicating that the discriminatory significance of a given measure is not sufficiently related to its Mahalanobis distance.

5.5 Results

5.5.1 Two Classifier Accuracy

The final two-classifier accuracy of the system was $81 \pm 2\%$. This is a good result, since we originally estimated the theoretical maximum accuracy to be between 80% and 90%, due to the limited amount of information contained in the contour shape of a skein. It is not likely that the theoretical maximum has actually been achieved, however, so in retrospect the theoretical limit for two-classifier accuracy has been refined to the range 85% to 90%.

5.5.2 Marginal Grading Results

The two-classifier accuracy was a convenient way of measuring overall accuracy of the system, but what is really important is the result the system will achieve in actual

operation. As mentioned in Section 3.2.1, a marginal grading scheme is proposed as a means to grade only a certain percentage of the bulk roe with high accuracy.

A file containing a list of difference measures made from the test set of first grade samples and a similar list made from the test set of second grade samples is produced by the software, as discussed in Section 4.5.2. A threshold value is picked, and all difference measures below that threshold are deemed to represent first grade samples, and all difference measures above deemed to represent second grade samples. The number of misgraded samples is then the number of first grade difference measures which fall above this threshold plus the number of second grade samples which fall below it. By decreasing the threshold, the number of second grade skeins which are accepted (as first grade) can be reduced to a small fraction of the total, however a significant number of first grade skeins will be rejected, requiring human grading of the rejected skeins in order to retrieve the valuable first grade ones.

If a number of different thresholds are picked, different values for the power and significance of the test for first grade shape (see Section A.1) result, denoted by $1 - \beta_i$ and α_i respectively. Recall that the power is the fraction of second grade skeins which are rejected and the significance is the fraction of first grade skeins which are rejected.

Denote the percentage of bulk ungraded skeins which are first grade by f_1 . The fraction of the skeins classified as first grade is given by:

$$F_i(f_1) = f_1(1 - \alpha_i) + (1 - f_1)\beta_i \quad (5.4)$$

And the accuracy of all skeins graded as first grade is:

$$A_i(f_1) = \frac{1 - \alpha_i}{F_i(f_1)} \quad (5.5)$$

In Figure 5.13, a number of different difference measure thresholds were used, denoted

by different values of i in Equations 5.4 and 5.5. For each threshold value, $F_i(f_1)$ and $A_i(f_1)$ were calculated for $f_1 = \{65, 70, 75, 80\}$. So for example, if the desired grading accuracy is 97% for first grade skein detection, and the bulk ungraded skeins were 80% first grade, then we find where the $f_1 = 80\%$ curve crosses the Y-axis value of 97%, and measure the X-axis position of this point to indicate that the system would select 50% of the bulk skeins (as first grade) at this accuracy.

5.5.3 Grading Distributions

In order to understand the shape of the curves in Figure 5.13, it is useful to examine the difference measure distributions of the first and second grade test samples. Figure 5.14 shows a graph of this distribution for $f_1 = 80$. The high accuracy achieved with a low difference measure threshold is a result of the rapid rise in the histogram of first grade skein difference measures, while that of the second grade skeins rises much more slowly.

This diagram also makes it clear that marginal grading to remove skeins which are clearly second grade is not possible. Because of the relative abundance of the first over second grade skeins, the $f_1 = 80$ histograms shown have similar values at larger difference measures. This means that even for large threshold values to reject 'clearly second grade' skeins, the rejected skeins would still contain a large percentage of first grade skeins.

Improvement of the grading system amounts to reducing the overlap between these two distributions.

Roe Grading Yield for Lot 125 (2 class. acc. 81%)

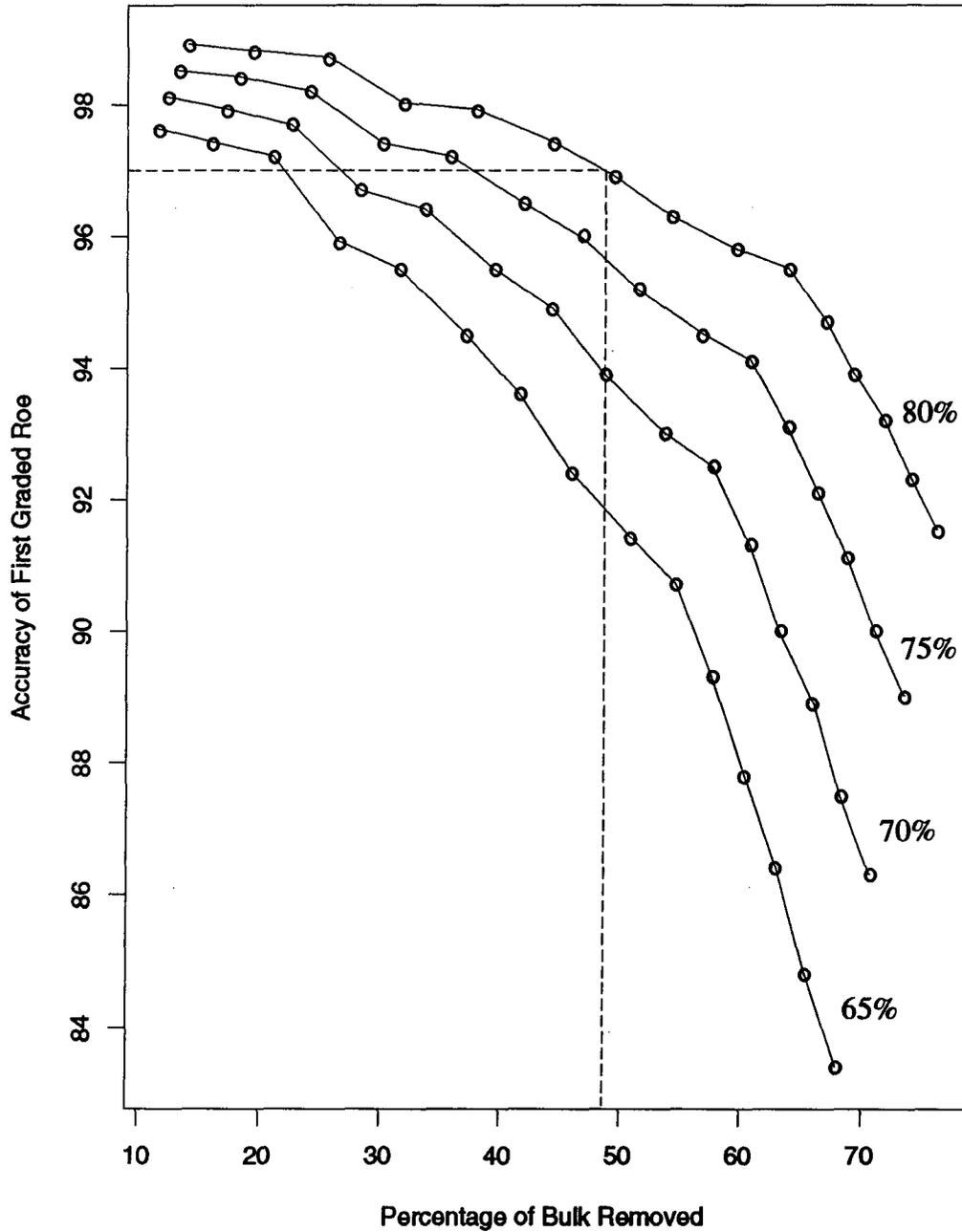


Figure 5.13: Marginal Grading Results

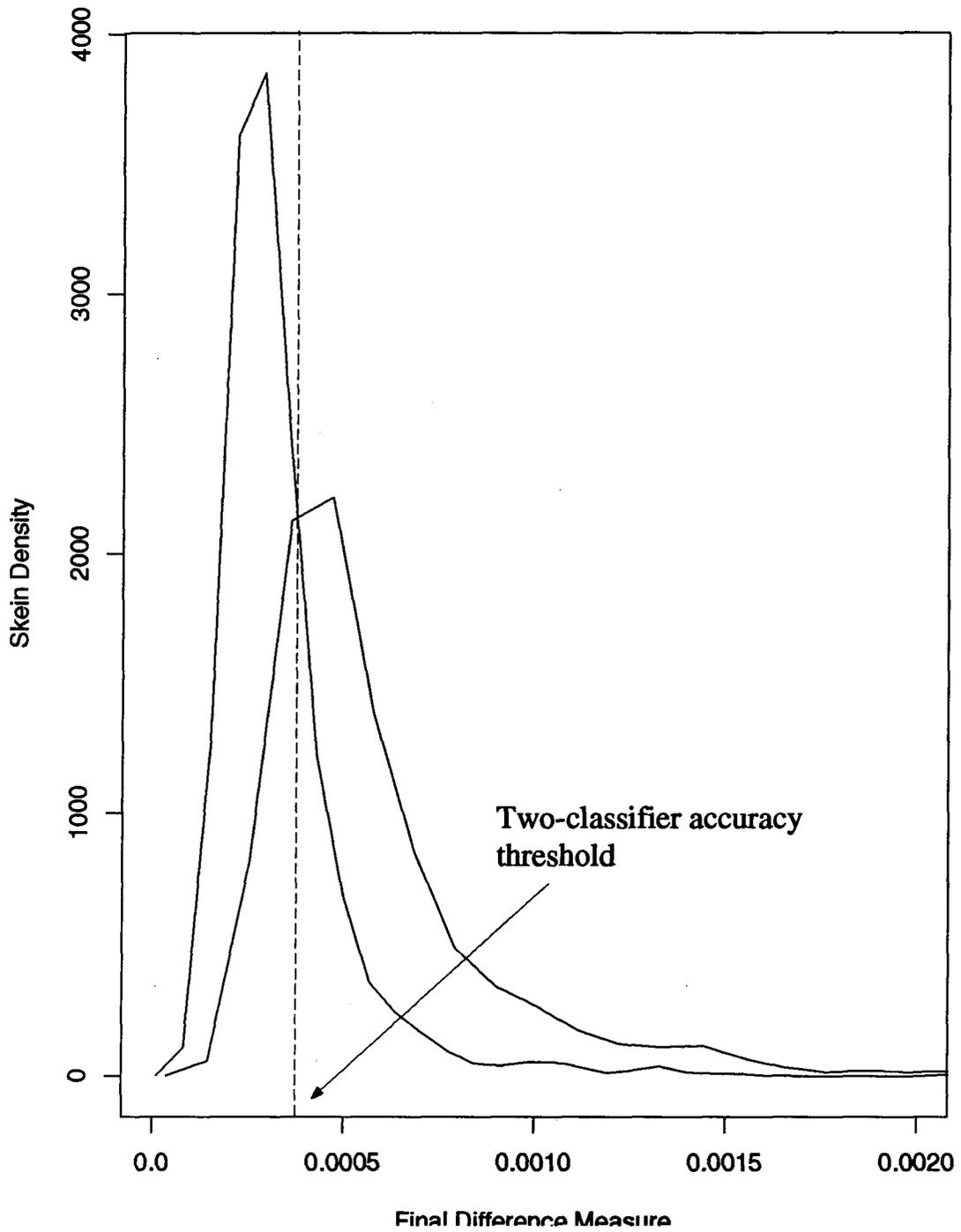


Figure 5.14: Histograms of Difference Measures for Typical Numbers of First and Second Grade Skeins

Chapter 6

Conclusions

6.1 Work to Date

A method of measuring the similarity between two-dimensional contour shapes which approximates the similarity perceived by people has been developed. There are restrictions on the types of contours which can be compared. First, the shape must have a distinct major principle axis of area, in order that an a priori reference axis can be established. Second, the tangent normal of the shape contour cannot be close to parallel to the major principle axis except at the two ends of the object, as defined by that axis. Exactly how close the normal can get to this angle depends on the desired accuracy of the representation, since the closer it gets, the more rough the internal representation of the contour, as compared to a length-parametric contour representation. Specifically, the ratio of the width-measure sampling frequency used in this system to a length parametric one is given by:

$$R(\theta) = (\sin \theta)^{-1}$$

Where θ is the angle between the tangent normal and the major principle axis of area.

Nor does the system described address the problem of identifying the roe skein in a less constrained environment. By using thresholding and binary connectivity methods, the software assumes that most of the work of segmenting the roe skein from the background has already been accomplished by lighting and placement of the skeins. None of these limits should be surprising, however, given the complexity of visual perception and

recognition, and the requirements that this system run in real-time and be robust.

The success of this system can be attributed mainly to two factors. First is the final representation and comparison algorithm, and second is the binary morphology operations used. The latter techniques have been widely used in computer vision, but I am not aware of any previous mention of the former.

The key contribution of the binary morphology operations was, as discussed, to act as a squelch on the scale of detail of the shape. This made the system robust even in the face of the many irrelevant protrusions that could occur in the contour.

The principle benefit of the final representation and comparison method used was a comparison of shape which agreed well with human observations. Human observations are, after all, how the quality of shape is defined in this application.

6.2 Future Modifications

The most apposite follow-up to this work is the extension of the contour representation and comparison method to more general shapes. In two dimensions this presents two difficulties. First, and most difficult, is when a shape does not have distinct principle axes of area. Since the method relies upon an a priori axis for representation, determining an appropriate axis beforehand may well require an entirely different representation. In simpler cases there may be some case-specific attribute allowing for a priori determination of axes, or the shape may lend itself to the determination of axes by searching through a series of angles starting with a guess, using some gradient descent variant.

The second problem in generalizing on two-dimensional shapes is the case when the tangent normal of the contour is parallel to the major principle axis at points other than the two endpoints. The most successful approach in this case might be to segment the object into parts, based on the representation from a single axis, and create an axis for

each part.

A more direct extension of the algorithm is that to three dimensions, or to range images. This would likely be most effective with a cylindrical raw width representation around the major principle axis of area. Final representation would be surface normal vectors for each cylindrically tessellated width measure. In this case, the restrictions on shape would be similar; there must be distinct principle axes of area, and the surface normals cannot be parallel to the major principle axis at any points but the endpoints.

To increase the accuracy of the system in its current application, investigation of additional measurements should be considered. Two which may prove useful in adding shape discrimination information are a measure of high-frequency angular variance in the contour, taken from the binary image which has not been morphologically processed, and a small number of average width measures, taken at strategic intervals along the length. The high-frequency angular variance measure would add a finer-scale quality measure to the lower-scale shape analysis currently used. The width measures may provide a check for cases where small differences in angle are cumulative, and relative widths vary significantly even though tangent angles are everywhere similar.

Appendix A

Mathematical Background

A.1 The Neyman-Pearson Paradigm

Under the Neyman-Pearson paradigm, a good treatment of which appears in [Rice 1988] we evaluate the probability of measuring the current situation for the case in which an hypothesis, H_0 , is true, and for the case in which it is false. In this application, we let H_0 mean that the skein being examined is of first grade. The vision system must either reject or accept H_0 for each skein which is examined. The system will be correct when it accepts H_0 , given a first grade skein, or rejects H_0 given a second grade skein. The probability that H_0 is rejected when it is false is called the power of the test. Two distinct errors can occur. A type I error occurs when H_0 is rejected for a first grade skein. The probability that H_0 will be rejected given that it is true, denoted by α , is the significance level of the test. A type II error occurs when H_0 is accepted for a non-first grade skein. The probability that H_0 will be accepted given that it is false is denoted by β . Clearly $1 - \beta$ equals the power of the test.

A.2 Erosion and Dilation

Erosion and dilation are converse, although not inverse, operations used to process binary images. Although their definition can be extended to greyscale images [Dougherty 1993], that is not a concern of this paper. In theory, an erosion operation maps all configuration pixels in the source which are within a certain distance from a background pixel

to the background. Configuration pixels not within this distance from a background pixel remain configuration pixels. Conversely, dilation maps background pixels close to configuration pixels to configuration pixels. In practice, rasterization allows only an approximation of the concept of distance, and in the case of the GPB-1 image processing hardware used, we are limited to the use of a three by three matrix window. The operation can clearly be repeated to the desired amount of erosion/dilation, but the operation is not isotropic. This, as mentioned in section 4.2.3, is the reason that the configuration representing the skein is rotated to the horizontal before application of the erosion/dilation.

The effect of erosion followed by an equivalent amount of dilation, as used by the pilot system, is to remove small-scale contour variations, especially filaments or any thin extension from the relatively large low-curvature shape of the skein.

Appendix B

Experimental Details

B.1 Expert Roe Grader Survey

Two human expert roe graders were asked to grade 240 samples of first and second grade skeins, in shuffled order. They were then asked to do the same from greyscale images, and then from binary images. The disparity between the two graders for the actual skeins was high (11%), but this is partly due to some confusion about the grading scheme being used. In industry, the number is generally about 8%. The disparity between the two graders in the binary images was 15%, and the discrepancy in grading assignments between the actual skeins and their binarized images was 22%. Since the graders were not familiar with grading skeins by their binary silhouettes alone, the added disparity between their evaluations (over that achieved with actual skeins) indicates that they could have more accurately graded from the binary images with some training. We can extrapolate these results by evaluating the additional uncertainty introduced in the binary evaluation, and then assuming that this portion can be graded as accurately as the remainder. Performing these calculations, and estimating the uncertainty gives us a potential accuracy range of 80%-90%. The results for the greyscale images were similar.

B.2 Executable Program Descriptions

Only those executables relevant to the final implementation are described here. Others, especially the principle components analysis executables which were abandoned early on,

are not described.

- **anal 'file1' 'file2'**

The arguments are files containing the difference measures of the first and second grade skeins, respectively, of some test set. The analyze program then accepts a threshold and gives both the alpha and beta errors around that threshold, optionally printing out a list of the image filenames which were alpha misclassified, and a list of those that were beta misclassified. These lists can then be used with another program (rerror) to produce two diagnostics measurements files for examination of error sources.

- **analchar 'file1' 'file2'**

The arguments are as above, but this executable accepts a range of thresholds and a step size, and loops through this range. For each threshold in this loop, the program loops through several likely possibilities for the percentage of incoming roe which is first grade (65 and prints out pairs of numbers corresponding to the percentage accuracy of the first graded skeins, and the percentage of bulk roe that would be graded out. These numbers can be directed to an ASCII file and exported to a statistics program (such as S+) to print out charts such as Figure fig:results.

- **analhist 'file'**

This program just translates the binary difference data in 'file' into ASCII for use with a statistics program (such as S+), to create histograms of the difference values, such as Figure 5.14.

- **blive**

Allows live camera view of the blue light spectrum.

- **demo1**

This is an interactive demonstration of the whole grading process, which only works with on-line images and requires size and brightness calibration before beginning. A file 'demo' containing any number of training samples in extracted measurement format must be present in the current directory. This file composes the internal database.

- **grab**

Uses the GPB-1 to frame grab and store multiple images for training set data collection. Images are grabbed with every stroke of the keyboard and stored in sequentially numbered files starting with a user-defined root name and base number. The grabbed images can optionally have the background thresholded to allow for maximum compression, whether storing on a compressed partition, or using an archiving program.

- **grabb**

As above except the images are grabbed from the blue light spectrum.

- **grabbit**

As above except the images are thresholded before being saved, for maximum compression. The test images so captured are useful for generating test results with very large numbers of samples, but make subsequent error analysis less effective.

- **hist**

Calculates the histogram of pixel intensities in an image taken in in real-time from the camera, full spectrum, using the GPB-1 board, and draws a histogram to the screen (256 brightness levels).

- **histb**

Same as above except that only the blue light spectrum is captured.

- **ravg 'file1' 'file2' ...**

Takes the average of any number of skein measurements in any number of extracted measurement files and stores the result as a single skein measurement in the file 'average.sys'.

- **qscand -1 'file1' ...-2 'file2' ...-t 'file3' ...**

The first file argument(s) contain extracted measurements of the first grade skeins of the test set, the second file argument(s) contain extracted measurements of the second grade skeins of the test set, and the third file argument(s) contain the measurements composing the training database. The classification process is performed on the test set, and the resultant difference measures are stored in the files 'one.dif' and 'two.dif' respectively.

- **pscand, hscand**

These commands accept the same arguments as 'qscand' above, but they classify using the alternate schemes of dynamic alignment and separating the skeins into halves, respectively, as described in Section sec:comparison.

- **qgscan,pgscan,hgscan**

These commands accept the same arguments as 'qscand' above, but instead of batch processing the classification process using their respective methods, they interactively perform the classification, providing numerical and graphic output to help diagnose sources of error.

- **rerror 'file1' 'file2' 'file3'**

The first filename is an ASCII file containing the filenames of images, generated automatically by '[qph]scand', which were misclassified in a batch test. The second file contains the measurements of these images, along with the rest of the test set, and upon completion, 'file3' will contain the measurements of only those misclassified skeins, for examination using '[qph]gscan'.

- **rselect 'file1' 'file2'**

The image corresponding to each skein in the extracted measurement file 'file1' is displayed, and if selected appended to 'file2'. This allows for hand-picking a training set, which was the preferred method before larger training sets and clustering were used.

- **train**

This program performs the measurement extraction, and creates measurement files from greyscale images. The images can be on-line or stored as files. If stored as files, they can be interactively queried, or batch processed using a file containing the image names. In all cases calibration size and lighting level must be interactively set at the beginning, along with the filename to contain the extracted measurements.

Bibliography

- [ASAE] The American Society of Agricultural Engineers, *Transactions of the ASAE*.
- [Arnarson 1991] H. Arnarson, *Fish and fish product sorting*, in L.F. Pau and R. Olafsson (Eds.), *Fish Quality Control by Computer Vision*, Marcel Dekker Inc : New York (1991). pp. 245-261.
- [Beatty 1993] A. Beatty, R.G. Gosine, C.W. de Silva. *Recent Developments in the Application of Computer Vision for Automated Herring Roe Assessment*. Proc. IEEE Pacific Rim Conference on Communications, Computers, and Signal Processing, pp.698-701 (1993).
- [Bebis 1992] G.N. Bebis, G.M. Papdourakis, *Object Recognition Using Invariant object Boundary Representations and Neural Network Models*. Pattern Recognition, Vol. 25(1):25-44 (1992).
- [Bengtsson 1991] A. Bengtsson, J.O. Eklundh, *Shape Representation by Multiscale Contour Approximation*. IEEE Trans. PAMI 13(1):85-93 (January 1991).
- [BSRAE] The British Society for Research in Agricultural Engineering, *Journal of Agricultural Engineering Research*. Academic Press, London.
- [deSilva 1992] C. deSilva, R. Gosine, Q. Wu, N. Wickramarachi, A. Beatty, *Flexible automation of fish processing*. Eng. Applic. Artif. Intell., (accepted December 1992)
- [Dougherty 1993] Edward R. Dougherty (1993). *Mathematical Morphology in Image Processing*. New York: Marcel Dekker, Inc.
- [Dubois 1986] S.R. Dubois, F.H.Glanz, *An Autoregressive Model Approach to Two-Dimensional Shape Classification*. IEEE Trans. PAMI, 8(1):55-66 (January 1986).
- [Duda 1973] R.O. Duda, P.E. Hart, *Pattern classification and Scene Analysis*. John Wiley and sons Inc. (1973).
- [Franz 1991] E. Franz, M.R. Gebhardt, K.B. Unklesbay, *Shape Description of Completely Visible and Partially Occluded Leaves for Identifying Plants in digital Images*. Trans. of the ASAE, 34(2):673-681 (1991).

- [Fukunaga 1990] K. Fukunaga, *Introduction to Statistical Pattern Recognition, Volume 2*. Academic Press, Inc. (1990).
- [He 1991] Y. Ge, A. Kundu, *2-D Shape Classification Using Hidden Markov Model*. IEEE Trans. PAMI, 13(11):1172-1184 (November 1991).
- [Horn 1986] B.K.P. Horn (1986). *Robot Vision*. Cambridge, Mass.: The MIT Press.
- [Huynh 1984] M.D. Huynh, L. Hildebrand, J. Mackey, *Preprocessing factors affecting quality of herring roe*, Technical Report 11, Fisheries Technology Division, B.C. Research Corporation (1984).
- [Lin 1992] Y. Lin, J. Dou, H. Wang, *Contour Shape Description Based on an Arch Height Function*. Pattern Recognition, Vol. 25(1):17-23 (1992).
- [Ling 1991] P. Ling, S.Starcy, *Feature extraction for a machine-vision based shrimp deheader*. Transactions of the ASAE 34(6):2631-2636 (1991).
- [Lowe 1987] David G. Lowe. *Three-Dimensional Object Recognition from single Two-Dimensional Images*, Artificial Intelligence 31:355-395 (1987).
- [Lowe 1985] David.G. Lowe. *Perceptual Organization and Visual Recognition*, Kluwer: Boston, MA, c1985.
- [Lowe 1989] D.G. Lowe, *Organization of Smooth Image Curves at Multiple Scales*. Int. J. Computer Vision 3. (1989).
- [Marchant 1990] J.A. Marchant, C.M. Onyango, M.J. Street, *Computer vision for potato inspection without singulation*, Comput. Electron. Agric. 4:235-244 (1990).
- [McFarlane 1991] N.J.B. McFarlane, *A computer-vision algorithm for automatic guidance of microplant harvesting*, Comput. Electron. Agric. 6:95-106 (1991).
- [McDonald 1990] T. McDonald, Y.R. Chen, *Applications of Morphological Image Processing in Agriculture*. Transactions of the ASAE 33(4):1345-1352 (1990).
- [Milios 1989] E.E. Milios, *Shape Matching Using Curvature Processes*. Computer Vision, Graphics, and Image Processing 47:203-226 (1989).
- [Miller 1991] B.K. Miller, M.J. Delwiche, *Peach Defect Detection with Machine Vision*. Transactions of the ASAE 34(6):2589-2587 (November-December 1991).

- [Mokhtarian 1992] R. Mokhtarian, A.K. Mackworth, *A Theory of Multiscale, Curvature-Based Shape Representation for Planar Curves*. IEEE Trans. PAMI 14(8):789-805 (August 1992).
- [Molto 1992] E. Molto, F. Pla, F. Joste, *Vision Systems for the Location of Citrus Fruit in a Tree Canopy*. J. agric. Engng. Res. Vol. 52:101-110 (1992).
- [Pau 1991] L.F. Pau, R. Olafsson. *Fish quality control by computer vision*, New York : M.Dekker, c1991.
- [Paulsen 1989] M.R. Paulsen, W.D. Wigger, J.B. Litchfield, J.B. Sinclair, *Computer Image Analyses for Detection of Maize and Soybean Kernel Quality Factors*. J. Agric. Engng. Res. 43:93-101 (1989).
- [Rehkugler 1989] G.E. Rehkugler, J.A. Throop, *Image Processing Algorithm for Apple Defect Detection*. Transactions of the ASAE, Vol. 32(1):267-272 (January-February 1989).
- [Reid 1991] J.F. Reid, C. Kim, M.R. Paulsen, *Computer Vision Sensing of Stress Cracks in Corn Kernels*. Transactions of the ASAE, Vol. 34(5):2236-2244 (September-October 1991).
- [Rice 1988] John A. Rice (1988). *Mathematical Statistics and Data Analysis*. Pacific Grove, California: Wadsworth & Brooks.
- [Sharp Inc. 1991] Sharp Digital Information Products Inc. *Sharp GPB-1 Image Processing Board Manual*, ©1991.
- [Shearer 1990(1)] S.A. Shearer, F.A. Payne, *Color and Defect Sorting of Bell Peppers Using Machine Vision*. Transactions of the ASAE, Vol. 33(6):2045-2050 (November-December 1990).
- [Shearer 1990(2)] S.A. Shearer, R.G. Holmes, *Plant Identification Using Color Co-occurrence Matrices*. Transactions of the ASAE, Vol. 33(6):2037-2044 (November-December 1990).
- [Strachan 1991] N.J.C. Strachan, C.K. Murray, *Image analysis in the fish and food industry*, in L.F. Pau and R. Olafsson (Eds.), *Fish Quality Control by Computer Vision*, Marcel Dekker Inc : New York (1991). pp. 209-223.
- [Taylor 1982] J.R. Taylor, *An introduction to error analysis : the study of uncertainties in physical measurements*, Mill Valley, Calif. : University Science Books, c1982.

- [Throop 1989] J.A. Throop, G.E. Rehkugler, B.L Upchurch, *Application of computer vision for detecting watercore in apples*, Trans. Am. Soc. Agric. Engin. 32(6):2087-2092 (1989).
- [Tillet 1991] R.D. Tillett. *Image Analysis for Agricultural Processes: A Review of Potential Opportunities*, J. agric. Engng Res. 50:247-257 (1991).
- [Tojeiro 1991] P. Tojeiro, F. Wheaton, *Oyster Orientation Using Computer Vision*. Transactions of the ASAE, Vol. 34(2):689-693 (1991).
- [Zayas 1990] I. Zayas, H. Converse, J. Steele, *Discrimination of Whole from Broken Corn Kernels with Image Analysis*. Transactions of the ASAE 33(5):1643-1646 (September-October 1990).
- [Zuech 1988] Nello Zuech. *Applying machine vision*, New York : Wiley, c1988