UNIVERSITY OF CALIFORNIA, SAN DIEGO

Multi-view, broadband, acoustic classification of marine animals

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Electrical Engineering (Applied Ocean Sciences)

by

Paul L. D. Roberts

Committee in charge:

Jules S. Jaffe, Chair
Mohan M. Trivedi, Co-Chair
William S. Hodgkiss
Kenneth Kreutz-Delgado
William A. Kuperman

2009

The dissertation of Paul L. D. Roberts is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____
Co-Chair

_____
Chair

University of California, San Diego

2009

DEDICATION

To my family: Jessica, Paul, Anne, Jane, Tenaya, and Charlie.

EPIGRAPH

"It would be a great mistake to suppose that vague knowledge must be false. On the contrary, a vague belief has a much better chance of being true than a precise one, because there are more possible facts that would verify it."

—Bertrand Russell, 1923

Climb the mountains and get their good tidings.

—John Muir, 1912

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

TS — Target Strength

SL — Source Level

NL — Noise Level

SNR — Signal to Noise Ratio

SRR — Signal to Reverberation Ratio

SAR — Synthetic Aperture Radar

WGN — White Gaussian Noise

TOF — Time of Flight

DWBA — Distorted Wave Born Approximation

FFT — Fast Fourier Transform

STFT — Short Time Fourier Transform

DCT — Discrete Cosine Transform

PPD — Posterior Probability Density

PDF — Probability Density Function

BDR — Bayes Decision Rule

CCD — Class Conditional Density

GLRT — Generalized Likelihood Ratio Test

GRNN — Generalized Regression Neural Network

GMM — Gaussian Mixture Model

PCA — Principle Component Analysis

LDA — Linear Discriminant Analysis

CCA — Canonical Correlation Analysis

FF — Feature Fusion

DF — Decision Fusion

CF — Collaborative Fusion

NN — Nearest Neighbor

K-NN — K-Nearest Neighbor

MLP — Multi-Layer Perceptron

RBF — Radial Basis Function

SVM — Support Vector Machine

RVM — Relevance Vector Machine

SBL — Sparse Bayesian Learning

MPA — Marine Protected Area

LIST OF SYMBOLS

a — radius

L — length

ρ — density

c — longitudinal sound speed

ω — frequency in radians

f — frequency in Hz

g — ratio of density in the scatterer to that of the surrounding medium

h — ratio of sound speed in the scatterer to that of the surrounding medium

$\gamma_\rho, \gamma_\kappa$ — gamma contrast

$f_b s$ — complex backscattering amplitude

p — continuous time pressure signal

s_0 — model transmit signal

s — scatterer impulse response

S — complex scattering amplitude

$\hat{S}$ — noisy complex scattering amplitude

k — wave number

k — integer index

$\mathbf{k}$ — wave vector

$\mathbf{k}_i$ — incident wave vector

$\mathbf{k}_s$ — scattered wave vector

$\mathbf{e}_i$ — unit vector in incident wave direction

$\mathbf{e}_s$ — unit vector in the scattered wave direction

θ — orientation angle in the horizontal plane

ϕ — orientation angle in the vertical plane

M — the number of views or receivers

$\mathcal{A}$ — system aperture

$\mathcal{V}$ — the number of views in the system

$\mathbf{\Gamma}$ — scatterer parameter matrix

$\mathbf{\Xi}$ — set of scatterer parameter matrices

$\beta_{j,\zeta}$ — beam shape factor for element j

$\mathcal{M}$ — match filtered output

P — magnitude of the Fourier transform of the matched filter output

$\mathbf{y}$ — general transformed feature vector

$\mathbf{y}_j$ — transformed feature vector in the j^{th} receiver

$\mathbf{Y}$ — matrix of feature vectors

E_j^{dct} — discrete cosine transform of data on element j

$\mathbf{C}$ — frequency correlation matrix

p — probability density function

P — probability mass function

$\mathcal{E}_j$ — echo envelope on the j^{th} recevier

a_β — approximation wavelet coefficients at level β

d_β — details wavelet coefficients at level β

c^* — optimal class label

$\mathbf{\Phi}$ — design (or kernel) matrix

$\mathbf{w}$ — kernel machine weight vector

$\mathbf{w}_c$ — kernel machine weight vector for separating class c from the rest

C_{SL} — sidelobe ratio; measures classifier confidence

P_t — projected ellipsoid thickness

ACKNOWLEDGEMENTS

The completion of this dissertation marks the end of nearly seven years spent working and studying at UCSD as a graduate student. During that time, I have had the good fortune to interact with outstanding people, many of whom played key roles in shaping the research presented herein. I owe them a great debt for helping me become the person I am today.

First and foremost, I would like to thank my advisor, Dr. Jules S. Jaffe, whose enthusiasm, optimism, and endless encouragement gave me a wonderful environment in which to pursue my research. Secondly, I would like to thank my co-advisor Professor Mohan M. Trivedi who welcomed me into his team, gave me valuable insight and ideas for my research, and took the time to help me express my own ideas. I would also like to thank the members of my doctoral committee: Professor William Kuperman, Professor William Hodgkiss, and Professor Kenneth Kreutz-Delgado for all of their constructive input that helped guide my research in a useful direction.

I would like to thank the California Sea Grant, Office of Naval Research, and the Seaver Foundation for funding my research.

I would like to thank Professor Peter J. S. Franks for giving me unending support during my time at UCSD.

I would like to thank Prof. Robert Hecht-Nielsen for his ECE 271 class which I had the good fortune to participate in during my fourth year as a graduate student. It provided a wonderful environment to develop professional skills, learn critical analysis, and develop into a responsible scientist.

I would like to thank Michael J. Buckingham, Fernando Simonet, and David C. Barclay for giving me the opportunity to join them on their Makai research trip. It was a great adventure, and rejuvenated my interest in oceanography at a time when it was needed most.

I would like to thank all of the staff at SIO and ECE who provided a solid foundation for a student trying to handle the often challenging job of straddling two departments. Specifically, Evelyn Doudera, Karol Previte, Megan Scott, and Rachael Pope who gave me much help, support, and education.

I would like to thank Eddie Kisfaludy and the SIO machine shop who gave me tremendous help in implementing the experimental side of my research.

I would like to Fernando Simonet, Rob Glatts, Ben Maurer, Erdem Karakoylu, Jonah Steinbuck, Yuriy Platoshyn, Chris Tung, David Martin, Casey Ta who all provided a fantastic team to work with and share experiences.

I would like to thank all of the past and present teams members of CVRR who made me feel welcome on their team, gave me helpful input about my work, and took the time to tell me about their research.

I would like to thank the countless number of developers of Free and Open Source Software who have created the tools I have used for my thesis.

I would like to thank Ben Maurer and Claire Debever for playing a key role in helping me manage two lives in separate cities, and for putting up with my biking obsession[1].

I would like to thank all of my friends in San Diego and Salt Lake City who were always there to help me remember that there is more to life than work and studying.

Finally, I would like to thank my family: Jessica, Paul, Anne, Jane, Tenaya, and Charlie. You have given me unconditional support in everything that I have ever tried. I dedicate this dissertation to you, and thank you from the bottom of my heart.

The text of Chapter 2 is in part and under some rearrangements a reprint of the material as it appears in Paul L. D. Roberts and Jules S. Jaffe, "Multiple angle acoustic classification of zooplankton," *J. Acoust. Soc. Am.*, **121** (4), 2060–2070, April 2007. The dissertation Author was the primary researcher and author, and the co-author listed in this publication directed and supervised the research which forms the basis for this chapter.

The text of Chapter 3 is in part and under some rearrangements a reprint of the material as it appears in Paul L. D. Roberts and Jules S. Jaffe, "Classification of live, untethered zooplankton from observations of multiple-angle acoustic scatter,"

[1] you know you started it!

J. Acoust. Soc. Am., **124** (2), 796–802, Aug. 2008. The dissertation author was the primary researcher and author, and the co-author listed in this publication directed and supervised the research which forms the basis for this chapter.

The text of Chapter 4 is in part and under some rearrangements a reprint of the material as it may appear in Paul L. D. Roberts, Jules S. Jaffe, and Mohan M. Trivedi, "A multi-view, broadband approach to marine fish classification," in prep. *IEEE J. Oceanic Eng.* June 2009. The dissertation author was the primary researcher and author, and the co-authors listed in this publication directed and supervised the research which forms the basis for this chapter.

The text of Chapter 6 is in part and under some rearrangements a reprint of the material as it appears in Paul L. D. Roberts and Jules S. Jaffe, "Multi-view sizing and classification of individual fish," *Proceedings of the 30th Acoustical Imaging Symposium*, Monterey, CA March 2009. The dissertation author was the primary researcher and author, and the co-author listed in this publication directed and supervised the research which forms the basis for this chapter.

VITA

2002	B.S. Computer Engineering *cum laude*, University of California, San Diego
2004	M.S. Electrical Engineering (Applied Ocean Sciences), University of California, San Diego
2009	Ph.D. Electrical Engineering (Applied Ocean Sciences), University of California, San Diego

PUBLICATIONS

P. L. D. Roberts, J. S. Jaffe, and M. M. Trivedi, "A multi-view, broadband approach to marine fish classification," in prep. *IEEE J. Oceanic Eng.* June 2009.

P. L. D. Roberts, J. S. Jaffe, "Multi-view sizing and classification of individual fish," Proceedings of the 30^{th} Acoustical Imaging Symposium, Monterey, CA March 2009.

P. L. D. Roberts, J. V. Steinbuck, F. Simonet, A. Uhlman, C. D. Troy, A. R. Horner-Devine, J. S. Jaffe, P. J. S. Franks, and S. G. Monismith, "Estimation of three-dimensional particle distributions from an in situ stereo particle imaging system," In prep. *IEEE J. Oceanic Eng.* June 2009.

E. M. Karakoylu, P. J. S. Franks, Y. Tanaka, P. L. D. Roberts, and J. S. Jaffe, "Copepod feeding quantified by planar laser imaging of gut fluorescence," *Limnol. Oceanogr.: Methods*, **7**, 33–41, Jan 2009.

J. V. Steinbuck, P. L. D. Roberts, C. D. Troy, A. R. Horner-Devine, F. Simonet, A. Uhlman, J. S. Jaffe, P. J. S. Franks, and S. G. Monismith, "An autonomous open-ocean stereoscopic PIV profiler," submitted *J. Atmos. Ocean. Tech.* 2009.

P. L. D. Roberts and J. S. Jaffe, "Classification of live, untethered zooplankton from observations of multiple-angle acoustic scatter," *J. Acoust. Soc. Am.*, **124** (2), 796–802, Aug. 2008.

P. L. D. Roberts and J. S. Jaffe, "Multiple angle acoustic classification of zooplankton," *J. Acoust. Soc. Am.*, **121 (4)**, 2060–2070, April 2007.

J. S. Jaffe, F. Simonet, P. L. D. Roberts, and A. E. Bowles, "Measurement of the acoustic reflectivity of sirenia (Florida manatees) at 171 kHz," *J. Acoust. Soc. Am.*, **121 (1)**, 158–165, Jan. 2007.

CONFERENCES

P. L. D. Roberts and J. S. Jaffe, 2008: "Application of multiple-angle acoustic scatter to remote fish classification," *Acoustics 08*, Paris, France (**1st place, Student Poster Award**).

P. L. D. Roberts and J. S. Jaffe, 2008: "Remote classification of fish species using multi-vew acoustic scatter, *SEAFACTS*," Bergen Norway.

P. L. D. Roberts and J. S. Jaffe, 2007: "Classifying pelagic animals with multi-view scattered sound," *American Fisheries Society*, San Francisco, California.

P. L. D. Roberts and J. S. Jaffe, 2007: "A new multiple angle scattering apparatus for fish and zooplankton studies, *153rd ASA meeting*," Salt Lake City, Utah.

P. L. D. Roberts and J. S. Jaffe, 2006: "Multiple-angle acoustic scattering and classification of zooplankton," *152nd ASA meeting*, Honolulu, Hawaii (**1st place, Student Paper Award**).

P. L. D. Roberts, J. V. Steinbuck, F. Simonet, A. Uhlman, C. D. Troy, A. R. Horner-Devine, J. S. Jaffe, P. J. S. Franks, and S. G. Monismith, 2006: "Development of a Pelagic Multi-spectral, 3-dimensional, Particle Image Velocimetry System: FIDO PHI PI," *Ocean Optics XVIII*, Montreal Quebec.

P. L. D. Roberts and J. S. Jaffe, 2004: "Identifying stable frequencies for performing animal abundance estimation with a multifrequency system," *148th ASA meeting*, San Diego, California.

ABSTRACT OF THE DISSERTATION

Multi-view, broadband, acoustic classification of marine animals

by

Paul L. D. Roberts
Doctor of Philosophy in Electrical Engineering (Applied Ocean Sciences)
University of California San Diego, 2009
Jules S. Jaffe, Chair
Mohan M. Trivedi, Co-Chair

Acoustical methods provide rapid, non-invasive, and synoptic tools for studying marine ecosystems. Despite the dramatic advances in this technology during the past three decades, there is presently a large disparity between the demand for quantitative information about marine animals and the capability of acoustic systems to deliver this information. A primary reason for this disparity is the strong dependence of acoustic scatter from marine animals on their size, shape, in situ orientation, and taxa. In a typical setting, these parameters are unknown, and are difficult to determine using existing acoustic methods. To mitigate this problem, a multi-view, broadband approach to marine animal classification and size estimation is investigated in this thesis.

Initially, zooplankton classification was investigated for two ecologically important taxa: copepods and euphausiids. Numerical simulations compared physics-based feature transformations, Nearest Neighbor (NN), and Multi-Layer Perceptron (MLP) classifiers. Results indicate that combining frequency-correlation features with a MLP yields an accurate (> 90 % correct) classification algorithm. Based on these promising results, a laboratory system was developed to recorded multi-view, broadband scatter from live, individual copepods and mysids. Results using frequency correlation features indicate that these features yield very good separation between classes with non-overlapping standard deviations computed

from eight individuals per class.

Next, sound scatter data from live, individual fish were used to develop several kernel-machine-based multi-view fusion algorithms. Performance was quantitatively compared as a function of the number of available views, feature spaces, and classification problem type. A collaborative fusion algorithm performs better than the others without requiring any assumption about view geometry, the number of views, or the type of features.

Finally, multi-view fish size and orientation estimation was investigated under three different approaches. Results indicate that classification-based size estimation can be effective with a limited aperture and limited number of views. Model-based and image-reconstruction-based estimation show very good performance with full aperture data.

This thesis demonstrates that the multi-view, broadband approach offers significant advantages for marine animal classification, sizing, and orientation estimation.

Chapter 1

Introduction

Marine ecosystems provide basic and fundamental living resources for humans [1]. Of these resources, perhaps the ones with the most direct influence on human life are fishes and the associated fisheries around the globe. In regard to fish populations in the ocean, there has been increasing evidence showing that many of these resources are being significantly depleted [2, 3], and new strategies of ecosystem-based management and marine protected areas (MPAs) are being recommended [1, 4, 5]. For zooplankton, research has demonstrated a strong correlation between increasing Sea Surface Temperature due to global warming and zooplankton abundance [6], and populations of the California Current have declined by 80 % during the last 60 years [7]. As the complexities of ecosystem management increase with the changes to management paradigms and the need to collect comprehensive data grows, new technologies must be developed, including advanced remote sensing methods for animal abundance estimation [8].

1.1 Marine animal abundance estimation

The estimation of marine animal abundance in the ocean is a fundamental challenge that is to-date still unsolved in many important aspects. In the context of zooplankton, the development of scientific tools for estimating abundance has been underway since the late 1800s [9]. In the context of fish several useful paradigms

for estimating abundance have emerged over the past 50 years [10]. In general these estimation methods can be partitioned into two groups: *direct* and *indirect*. Direct methods physically capture specimens which can be identified and measured either at the time of capture or later in post processing. Direct sampling methods typically use a sampling chamber or net that is deployed in the water column and collects water samples from a specific 3D coordinate in the ocean at a single time or integrated over the tow-path of a sample net. By combining multiple collection devices together (for example a Rosette [9], or a MOCNESS [11]), direct samples can be obtained as a function of both time and space. The primary advantage of these methods is that they offer direct evidence for the type of animals present in a particular region along with some information of their abundance. However, they have the drawback that they are sensitive to the behavior of the animals being studied in the response to the collection system, quantification and enumeration of animals can be laborious, and total volumes sampled can be small compared to other methods.

Indirect sampling methods measure a physical quantity of the animal. The most common of these methods use either sound or light to interrogate the animal and record the scattering of these waves in several dimensions, and in some cases as a function of wave frequency. With the development of faster, smaller, and more powerful computing resources, these methods have grown considerably in the past three decades. These methods offer the distinct advantage over direct sampling methods in that they can be applied to highly mobile animals, do not need to disrupt the echosystem being studied, and are more readily adapted to automated processing to yield quantitative scientific information. The primary drawback of these methods is that their utility is highly dependent on the *specificity* of the measurement system; the degree to which it can distinguish between different types of animals and infer biomass. Specificity usually comes with a tradoff of range and sample volume. The larger the distance from the sensor or the greater the volume sampled, the lower the specificity.

1.2 Modern optical and acoustical sampling methods

Optical methods sample small volumes with very high specificity. In contrast, acoustical methods typically sample large volumes with much lower specificity. Because of the complementary nature of optical and acoustic methods, they are frequently used together [12, 13, 14]. In many cases, direct sampling is also combined to yield ground truth data that can help adjust parameters used to convert estimates from indirect samples into quantitative measures of abundance.

Optical and acoustic image-based methods have been applied to many different types of fish abundance estimation using diver visual surveys [15, 16], LIDAR [17], ROV and HOV based studies [18, 19], and imaging sonar [20]. Optical methods for studying zooplankton have also been well developed over the past two decades [21, 22, 23, 24]. While optical methods offer higher specificity than acoustical methods, they do so at the cost of limited range due to the high attenuation of light waves in seawater. Ranges for optical systems are less than one meter with sample volumes less than one liter. In some cases, water properties or nightime operation preclude the use of optical methods alltogether. Optical methods can also suffer considerably from biofouling [25].

In contrast to optical methods, acoustic methods can perform well at long ranges in seawater, in total darkness, and for long term deployments. Fisheries research has explored the use of acoustic methods for more than half a century [26, 27] and adopted acoustic methods for many areas of study [28, 29, 30]. Recently developed automated processing of acoustic data [31] and new fisheries sonars [32] show promise in obtaining improved abundance estimates as well as permitting the estimation of fish size and behavior. Passive acoustic methods for studying have also been well developed [33].

Since the late 1960s, acoustics has played an important role in estimating zooplankton abundance [34, 35, 36, 28]. Early work focused on multi- or single-frequency abundance estimation using a linear inverse approach based on average models of scattering from known species [35, 36]. These methods were then aug-

mented with development of improved scattering models which cover a wide band of frequencies [37, 37] and broadband classification methods [38, 39]. Recently, acoustic methods for estimating zooplankton abundance estimates have proliferated [40, 41, 42, 43].

Despite the many advantages of acoustic methods as enumerated above, these methods suffer from poor specificity due to ambiguities that arise from the difference in material properties, size, shape, and orientation of the animals being studied. Based on these ambiguities, an acoustic system can easily confuse a small number of animals with high reflectivity with a large number of animals with low reflectivity as it has been shown that the echo energy per unit biomass is widely variable over different taxa of zooplankton [44], and the correlation between volume scattering and zooplankton abundance can be strong for certain taxa, and non-existent for others [45]. Improved estimates of abundance have recently been accomplished by accounting for the taxonomic differences in population structure when applying scattering models, acoustic derived abundance estimates can be improved [46, 47].

While significant research has gone into improving the specificity of acoustic methods for zooplankton abundance estimation by using broadband scattering [38, 37, 39, 48], these methods have only been applied to fish abundance estimation in a few cases [49, 50, 51]. In addition, nearly all applications of acoustical methods to estimation of bio-physical parameters of marine animals have used a co-located transmiter and receiver, collecting monostatic backscatter measurements. Despite the advances made by using broadband scattering, there is a remaining fundamental uncertainty in scattering which is due to the strong dependence of scattering from marine animals on animal orientation. It has been shown [52, 53, 54] that fish orientation plays a key role in the inference problem, to the extent that scattering model errors can be dominated by errors due to uncertainty in fish orientation [54]. This effect has also been documented for zooplankton [55, 56, 57]. These findings are a direct result of the fact that many fish and zooplankton are highly elongate animals, whose length may be several times larger than their other dimensions. As the sound used to insonify these animals is typically in the geometric scattering

regime, this causes a strong dependence on orientation. The animal's scattering cross-section is quite large when it is broadside to the sensor, and quite small when it is off-broadside.

Recently, the idea of using multiple angle scattering to improve the characterization of fish [58, 50], and zooplankton [59, 60] has been explored through simulations and analysis of laboratory data. The basic premise of these methods is that by observing scattering over multiple angles, the estimation of animal size, shape, and taxa can be improved because multiple observations explicitly sample the variability in scattering as a function of animal orientation.

There has also been significant research in fisheries applications devoted to classifying echograms obtained by current scientific echosounders [61], [62]. These methods compute features from an echogram of a fish school and then classify these features with a neural network or kernel machines. In situations where the school is represented by a single species, and the features computed from the echogram are independent of echo intensity, these methods will not suffer from the unknown orientation of the fish in the school. However, these methods are inherently single aspect. There is one recent example [50] of using multiple angle spectra to classify fish in which fish are insonified by a system while they swim around in a tank and a video system tracks their position and orientation relative to the sonar. These data are then used to classify the fish. However, the classification is based on frequency vs. angle spectra when angle is assumed to be known, not on a small set of scattering measurements made when angle is unknown.

1.3 Multi-view target classification

Aspect dependence in target detection and classification has been studied in many sonar and radar applications [63, 64, 65, 66, 67, 68, 69, 70]. In these studies the influence of unknown target orientation on target classification is mitigated by first collecting scattering measurements over multiple angles and then fusing these measurements together. In some cases, the approach allows the data collection to be adapted to the scattering features of the target [71] allowing for improved

classification due to the collection of more discriminant information. In other cases, the data collection model allows for estimation of the unknown target orientation [72].

The most popular formulation employs a Bayesian framework in which the orientation of the target relative to the sensor is assumed to be random and unknown. Because the characteristics of the reflected wave change slowing as a function of view angle, the scattered signal will be highly correlated for small changes in target aspect. Because of this correlation, numerous algorithms which model the unknown orientation as a discrete set of angular "states" have been developed during the past decade [73, 66, 68, 74]. The unknown orientation-state of the target is then modeled as a first-order Markov process, and the probability of a sequence of observations of the target is computed using a hidden Markov model (HMM) framework. Recently, adaptations to the standard HMM structure have been applied which allow for optimization of the number of states [68], [74]. In parallel to the development of HMM based methods have been the development of connectionist methods which train a network to classify features from a single aspect, and then also to fuse classifications from multiple aspects together [63, 70]. These methods typically rely on the fact that targets can be accurately classified using scattering from a single aspect. Finally, there has been recent work in applying collaborative filtering methods to multi-aspect data which allow different aspects and feature transformations to be fused together to yield improved performance [75]. A detailed discusion of the different types of classification algorithms is given in Chapter 4.

1.4 Scope

Based on the perceived potential benefits of combing the use of broadband methods with multi-view techniques, a research program was instituted to examine these benefits for acoustic classification of fish and zooplankton. This dissertation explores the application of multi-view broadband acoustic methods to classification of acoustic scattering from marine animals. The problem is framed in a context

similar to that of the underwater target classification (UTC) literature. Given a set of M observations of sound scattering recorded over a typically *limited* range of angles, the objective is to define a function f which predicts the type of scatterer the observations came from without knowledge of the orientation or length of the scatterer. This dissertation considers the case where the function f is learned from a large set of scattering data collected from the same, or similar scatterers. The primary problem being addressed is that of learning the underlying scattering features of each class which separate it from the other classes. In addition, this dissertation considers the problem of inferring the size, orientation, and shape of fish from observations of multiple angle scatter. This problem in evaluated in the context of both limited- and full-aperture data under three different paradigms: classification-based, model-based, and tomography-based.

The classification and estimation problems are framed in the context of acoustic scattering from individuals (comparable with most UTC methods). It is assumed that a broadband echo from an individual animal has already been detected and segmented from the background, and the remaining problem is to identify which class the animal belongs to, or to estimate a physical parameter of the animal such as its size or shape. A small set of observations of broadband scatter are assumed to be available for this purpose. These observations are generally assumed to collected over a limited range of angles.

Simulation studies of zooplankton classification are performed using realistic numerical scattering models with uniform distributions assumed for animal orientation and length. Using simulated data, the classification problem is posed as a problem of developing feature transformations which combine multi-view data into a single feature set which highlights fundamental differences between classes (Chapter 2). These features are computed from simulated data and evaluated using standard classification algorithms. The performance metric for comparison between feature and classification algorithms is the probability of error as a function of the number of views collected.

Classification of scattering from live animals is evaluated using data from a custom designed multi-view laboratory scattering system. Feature transforma-

tions developed in the simulation study are investigated for broadband multi-view scattering from live, untethered zooplankton (Chapter 3).

Broadband scattering from over 30 different individual fish spanning 9 different species are presented and evaluated in the context of size- and species- based classification using three different multi-view fusion algorithms (Chapter 4), transmit signal bandwidth and array aperture sensitivity (Chapter 5), and fish size and shape estimation (Chapter 6).

1.5 Significance and conclusions

This dissertation presents a comprehensive assessment of both the experimental and computational aspects of the application of multi-view methods to the acoustic classification of marine animals. Results indicate that the multi-view approach offers significant advantages over corresponding single-view methods in terms of lower classification error, or reduced uncertainty in parameter estimates. As described above, there is currently a need for new tools for studying marine animals non-invasively, remotely, and during all hours of the day. This dissertation demonstrates that our laboratory multi-view system, in every case, results in improved accuracy for classifying marine animals. This provides motivation for the implementation of these techniques in a field system, where possible.

As part of the work, a new algorithm has been developed that uses collaborative fusion. The algorithm is general enough to be applicable to a host of multi-class problems without modification. As one advantage of the algorithm, it can be used even when the views are randomly obtained, owing to the fact that the algorithm does not make any assumptions about view geometry in the fusion process. A further advantage is that the algorithm is developed in a way that allows for additional views to be acquired and processed without retraining the algorithm.

During the course of this work, several important challenges have been identified for future research. The primary challenge is related to the effect of limited aperture on multi-view performance. It has been shown both for classification and

size and shape estimation that limited aperture causes significant reductions in performance. A future direction would be to investigate the design of multi-view systems to take advantage of known animal orientation statistics and use this information to maximize the use of limited aperture collection geometries. A second challenge is related to laboratory methods for collecting scattering data from live animals. While the method presented here provides a very controlled and systematic way to collect data, it is time consuming, and suffers from potential biases due to the reaction of the animals under study to the experimental system. An important future direction of research would be to develop a laboratory system that could rapidly collect data from live, unrestricted animals in perhaps an aquarium setting. Finally, the algorithm development and analysis in this dissertation has focused on feature-fusion and decision-fusion algorithms which use kernel methods for classification that do not explicitly model the joint probability of the observed scattering. This is in contrast to hidden Markov models which have shown significant success in underwater target classification [72, 68, 74]. An important future direction is to compare the kernel-based methods developed in this dissertation to HMMs and other probabilistic feature-fusion methods.

Chapter 2

Multiple angle acoustic classification of zooplankton

2.1 Introduction

Zooplankton play a major role in the global ecosystem and the employment of remote sensing techniques for measuring abundance and behavior continues to be a venerable goal. Compared with optically based methods, acoustic ones have an inherent advantage in that sound is attenuated less than light. This leads to both larger detection distances as well as sampling volumes. Unfortunately, problems associated with a lack of specificity have hindered the use of acoustic techniques on a routine basis. Work by McNaught [34] and the subsequent development of multiple frequency methods by Holliday and colleagues have revealed both the great advantages and also challenges that exist when using this technique [35, 76, 36, 77, 78]. Although additional work with broad band sound to discriminate between three groups of zooplankton [38, 56] indicated that it was possible to correctly classify specific examples from each group with reasonable success (80% overall average correct classification), the goal of robustly mapping acoustic volume scattering to biophysical parameters of zooplankton under various oceanographic conditions has remained elusive. One problem has been the confounding influence of both orientation and material properties on backscatter magnitude.

This often prevents investigators from making the necessary link between animal size and backscatter magnitude. Additional complications have been that there is substantial scatter from non-biological sources such as suspended sand, bubbles, and perhaps even microstructure [79]. If a way could be found to discriminate among various taxa acoustically, in spite of these problems, it would be of great value.

In this article the potential increase in classification accuracy that results from observation of reflected sound from multiple angles is considered. It has recently been proposed that sound scattered at multiple angles can be used in order to both size and measure the orientation of fish bladders [58]. The underlying concept being that the spatial structure of the sound field from a single, strong scattering target, has a characteristic pattern related to its size. The success of the method was illustrated with a well known data set [80] and various sampling theorems were proved to obtain unaliased sampling of the scattered sound field. Here, the use of sound scattered at multiple angles in order to discriminate among two zooplankton taxa is explored via forward modeling and subsequent classification.

Many have considered the formulation of acoustic models to predict backscatter as an important component of a program to characterize animals *in situ*. A family of scattering models can be successfully used to predict the acoustic reflectivity of several different types of zooplankton [44, 81, 82, 83]. The situation with respect to crustacean zooplankton is especially good as use of the distorted wave Born approximation (DWBA) has been validated [57, 84]. A web site maintained by Benfield [85] provides public access to several zooplankton models and their morphologies.

Outside of the realm of ocean ecology, recent work in the acoustic classification of stationary targets from multiple views has demonstrated that the multiple views can significantly improve target classification when combined with suitable feature extraction and classification algorithms [63, 68, 71, 70]. Applications such as underwater mine detection [68, 70], airborne target identification [65, 86], and unexploded ordinance detection [87] have been considered. Most algorithms apply a hidden Markov model (HMM) to account for either the unknown sensor-target

aspect [72, 68, 71], or the unknown target type [70]. One approach decomposes the target reflections into a set of discrete angular regions yielding a set of possible states in the HMM [72]. Alternatively, a nonlinear Back Propagated Neural Network (BPNN) has been used to fuse the classification results for multiple views in a wavelet packet based feature space [63]. This formulation demonstrated very good performance in discriminating between mine and non-mine like targets from multiple aspect scattering measurements.

Adaptation of multiple angle scatter techniques to zooplankton classification has promise to confer benefits when used in conjunction with the more traditional backscatter techniques. However, animals are dynamic and therefore require an observation system in which multiple views are obtained almost simultaneously. One solution is to use simultaneous multiple angle scatter measurements. A second issue is related to the feature space used to represent the data. Previous work in target classification considered rigid objects and therefore applied wavelet packets [63, 88, 70] or Matching Pursuit with an elastic scattering based dictionary [72, 68]. However, the resulting feature spaces are not appropriate for the fluid-like weak scatterers considered here. A more appropriate idea for this problem is to exploit the relationship between the shape of the scatterer and the angularly varying scatter amplitude.

This paper explores, through simulation, the use of a one-dimensional array to collect multiple angle scatter and subsequently use these data to discriminate among zooplankton taxa. The case treated is that of differentiating between two taxa of crustacean zooplankton: copepods and euphausiids. The motivation for treating these animals stems from their significance in zooplankton populations of the California Current. As shown here, the large morphological difference between the two groups [89] will allow this discrimination.

2.2 Forward Modeling: Theory and Numerical Implementation

In this section, an acoustic forward model for generating synthetic data is proposed using linear system theory and the distorted wave Born approximation (DWBA). The forward model permits the prediction of the received signal for a known transmit signal using the impulse response of the scatterer. This depends on both the physical properties of the scatterer such as size, shape, and material and also the orientation of the scatterer and the geometry of transmitters and receivers. Under the assumption of linearity, and neglecting effects of spreading and medium attenuation, the received signal $p(t)$ is given by the convolution of the transmitted signal $s_0(t)$ with the impulse response of the scatter $s(t, \mathbf{k}_i, \mathbf{k}_s, \theta, \phi, \mathbf{\Gamma})$

$$p(t) = \int_{-\infty}^{\infty} s_0(\tau) * s(t - \tau, \mathbf{k}_i, \mathbf{k}_s, \theta, \phi, \mathbf{\Gamma}) d\tau, \tag{2.1}$$

where $\mathbf{k}_i$ and $\mathbf{k}_s$ are the incident and scattered wave vectors, θ and ϕ define the orientation of the scatterer, and $\mathbf{\Gamma}$ is a parameter matrix describing the size, shape, and material properties. Assuming values for these parameters permits the prediction of the impulse response of the scatterer using the DWBA. This model does not include propagation effects, however the effect of scatterer position in the beam is included. A description of the multiple angle DWBA is given in Section 2.2.1. Section 2.2.2 defines the size and orientation distributions that are used to generate synthetic data. Section 2.2.3 describes the procedure for generating synthetic data.

2.2.1 Multiple angle DWBA scattering model

The scattering model used to obtain the impulse response of the scatterer is the DWBA[82, 84]. This model relates the size, shape and material properties of the scatterer to the complex scattering amplitude at a particular frequency. The impulse response of the scatterer can be obtained from the complex scattering amplitude by an inverse Fourier transform. The expression for the complex scattering

amplitude $S(\mathbf{k})$ is

$$S(\mathbf{k}) = \frac{\mathrm{k}_1^2}{4\pi} \int\int_R \int \left(\gamma_\kappa(\mathbf{r}_0) - \gamma_\rho(\mathbf{r}_0)\cos\alpha\right) e^{i\mathbf{k}\cdot\mathbf{r}_0} dv_0, \tag{2.2}$$

where

$$\mathbf{k} = \mathrm{k}_2 \left(\mathbf{e}_s - \mathbf{e}_i\right), \tag{2.3}$$

and

$$\cos\alpha = \mathbf{e}_s \cdot \left(-\mathbf{e}_i\right). \tag{2.4}$$

The scalars $\mathrm{k}_1 = 2\pi f/c_1$ and $\mathrm{k}_2 = 2\pi f/c_2$ are the wave numbers in the medium and body of the scatterer respectively, $\mathbf{e}_i$ and $\mathbf{e}_s$ are unit vectors in the direction of the incident and scattered sound waves, and α is the angle between the negative incident wave vector and the scattered wave vector. The term $\gamma_\kappa(\mathbf{r}_0) - \gamma_\rho(\mathbf{r}_0)\cos\alpha$ is the gamma contrast [84, 90].

The gamma contrast inside of the volume integral is a function of the density and sound speed of the surrounding medium and the body of the scatterer where (omitting the explicit dependence on position in the body) $\gamma_\kappa = \left(1 - gh^2\right)/gh^2$ and $\gamma_\rho = \left(g-1\right)/g$. The term $g = \rho_2/\rho_1$, is the ratio of the density of the scatterer to the density of the surrounding medium and $h = c_2/c_1$, is the ratio of sound speed in the scatterer to the sound speed in the surrounding medium. equation (2.2) provides the basis for the forward model used in the numerical experiments presented in this paper.

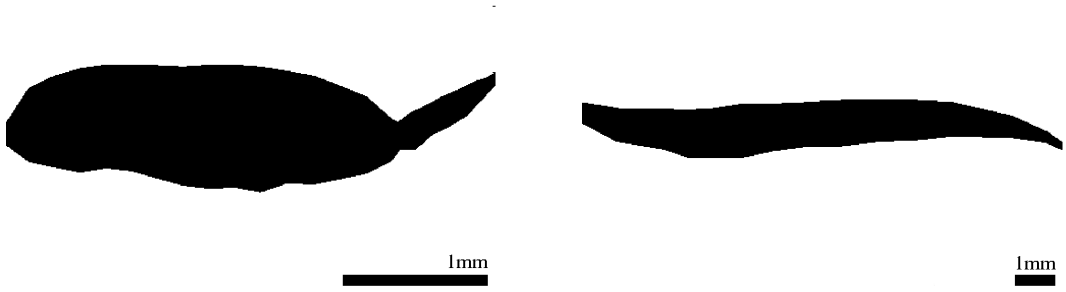
Using equation (2.3), the resultant wave vector $\mathbf{k}$ can be written as

$$\mathbf{k} = \mathrm{k}_2 \frac{sin(\pi - \alpha)}{sin(\alpha/2)} \mathbf{d}, \tag{2.5}$$

where

$$\mathbf{d} = \frac{\mathbf{e}_s - \mathbf{e}_i}{\|\mathbf{e}_s - \mathbf{e}_i\|_2}, \tag{2.6}$$

is the unit vector that points in the direction of the difference between scattered and incident wave vectors. It is apparent that the multiple angle DWBA is closely related to the DWBA for backscatter[84] only now with a scaled and rotated wave vector. This important relationship allows the DWBA for multiple angle scatter to be computed using existing numerical methods for backscatter with only minimal modification.



(a) Copepod (b) Euphausiid

Figure 2.1: Cross-sectional view of the scatterer shapes used to define each class. The copepod shape (a) and euphausiid shape (b) are displayed at different scales. The scale is defined in the lower right corner.

2.2.2 Scatterer size and orientation distributions

An important aspect of the simulations is the choice of size and orientation distributions for the ensemble of scatterers. These data have been generated using a single shape for each class, scaled in volume and rotated in three dimensions. The shapes used for the copepod and euphausiid classes were taken from an online database of zooplankton scattering models [85]. Pictures of the cross-section of the base shape used for each class are shown in Figure 2.1.

The volume scaling is parameterized by a length parameter L, the length of the scatterer from head to tail. In order to simplify the treatment both length classes were drawn from uniform distributions according to $\mathcal{U}[2\text{ mm}, 4\text{ mm}]$ for the copepods and $\mathcal{U}[4\text{ mm}, 15\text{ mm}]$ for the euphausiids. Note that the length distributions overlap slightly and the distribution for the euphausiids is centered around medium length juveniles rather than the larger adults.

Similarly, for the orientation distributions, a simple approach was taken. Representing the orientation of the scatterer by a θ and ϕ angle where θ is the angle relative to the z axis, and ϕ the angle between the x and y axis (Figure 2.2) the orientations were sampled uniformly in three-dimensions according to

$$\theta \sim \arcsin(\mathcal{U}[-1, 1]), \tag{2.7}$$

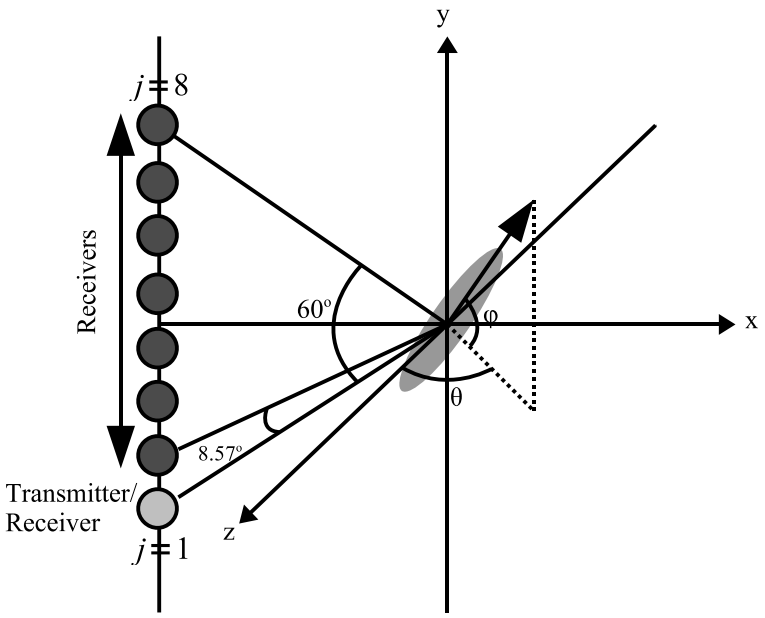


Figure 2.2: View of the array configuration used to generate data. The scatterer is represented by the gray ellipse where the nose of the scatterer is directed along the unit vector defined by the angle ϕ in the x-y plane and θ from the z axis. The angular span of the array is 60 degrees with an angular spacing of 8.57 degrees between elements. The $j = 1$ element functions as both transmitter and receiver, while the $j = 2, ..., 8$ elements receive only.

$$\phi \sim \mathcal{U}[-\pi, \pi]. \tag{2.8}$$

2.2.3 Creation of model realizations

The simulation of a single realization of received scatter on the array is described in this section. The configuration of the array is shown in Figure 2.2. There are eight total elements, $M = 8$. The $j = 1$ element acts as both transmitter and receiver while the $j = 2, ..., M$ elements act only as receivers. The total angular span of the array is 60 degrees with an angular sampling frequency of one sample per 8.57 degrees. The orientation of the scatterer relative to the array is shown in Figure 2.2 and is defined by the angles ϕ and θ as mentioned previously.

The synthetic data is generated by predicting the received pressure signal on each of the eight array elements for a given scatterer orientation. The data

generation process is represented graphically in Figure 2.3. The first step is the selection of a model: copepod or euphausiid. Each three-dimensional scatterer shape is represented as a series of cylindrical segments of thickness 0.016 mm and location x, y, and z corresponding to the center. The segments have radius a, and relative density and sound speed g and h. The i^{th} segment can be represented as the vector $\boldsymbol{\gamma}_i = (x_i, y_i, z_i, a_i, g_i, h_i)^T$ and the entire model of the body by a matrix $\boldsymbol{\Gamma} = (\boldsymbol{\gamma}_1, ..., \boldsymbol{\gamma}_S)$ with S being the total number of segments. Ambient sound speed c, is fixed at 1500 m/s. For all simulations performed in this work, the values of g_i and h_i are held constant throughout the body of the scatterer such that $g_i = 1.035 \ \ \forall \ \ i$ and $h_i = 1.027 \ \ \forall \ \ i$. More information about these parameters, the scattering models, and the algorithm used to compute the DWBA is available from the Acoustic Scattering Models of Zooplankton website[85].

For each realization, a random sample from the distributions for ϕ, θ, L is selected. These parameters are combined with the sound speed c, the incident and scattered wave vectors for each array element: $\mathbf{k}_i^j$ (incident) and $\mathbf{k}_s^j$ (scattered) for $j = 1, ..., M$, and the model for the scatterer $\boldsymbol{\Gamma}$. The DWBA is evaluated for the given model parameters at each frequency yielding the complex scattering amplitude

$$S_j[k] = \text{DWBA}\left(\phi, \theta, L, c, \mathbf{k}_i^j, \mathbf{k}_s^j, \boldsymbol{\Gamma}\right), \tag{2.9}$$

where k represents the index of a particular wave number bin.

To incorporate the effect of a range dependent sample volume in the simulation, for each realization, the scatterer is assigned a uniformly random three-dimensional position relative to the array. The position is defined by the parameter $\boldsymbol{\zeta}$.

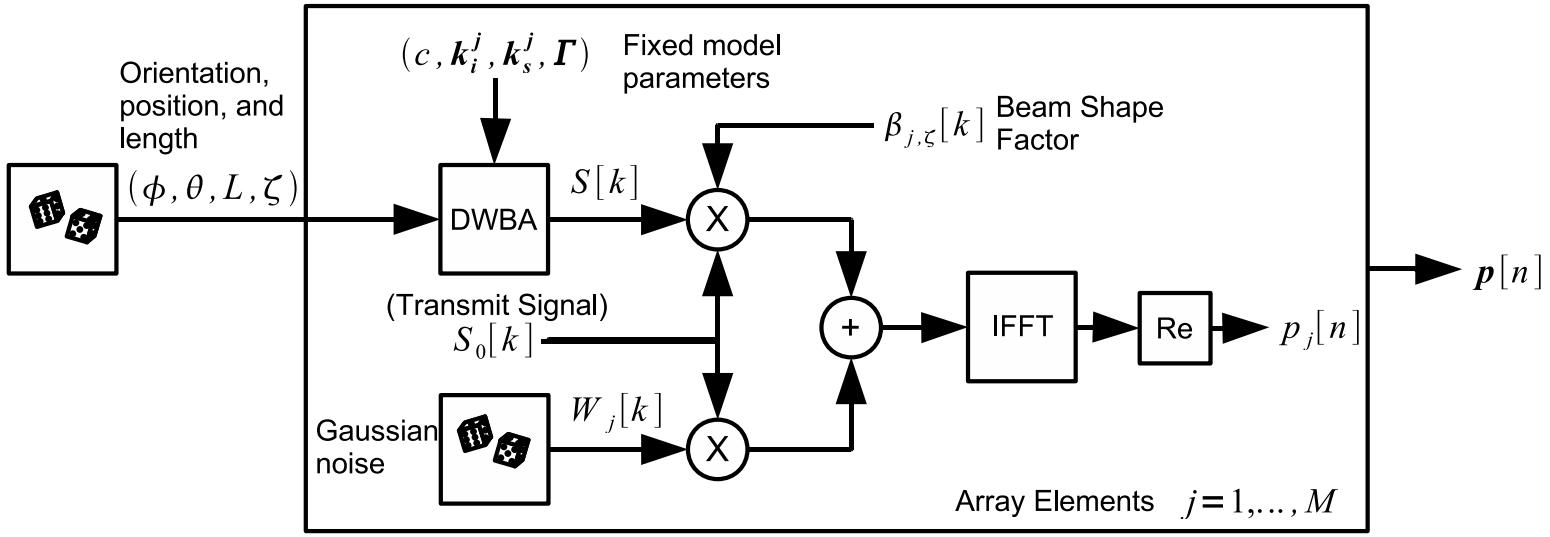


Figure 2.3: Block diagram of the creation of a single realization of scattering on the array. Randomness is included in the model through the parameters ϕ, θ, L, $\boldsymbol{\zeta}$, and the Gaussian noise $W[k]$. The model parameters which are constant for all realizations are the sound speed c, the incident and scattered wave vectors for each array element $\mathbf{k}_i^j$ and $\mathbf{k}_s^j$ for $j = 1, .., M$ and the scatterer model $\mathbf{\Gamma}$. The beam shape factor $\beta_{j,\boldsymbol{\zeta}}[k]$ is computed for each array element and accounts for the position of the scatter in the transmit and receive beam pair. The Gaussian noise $W[k]$ is added to the product of the scattering amplitude $S[k]$ and the beam shape factor $\beta_{j,\boldsymbol{\zeta}}[k]$. The result is then multiplied by the FFT transmit signal $S_0[k]$. The real part of the inverse FFT of the product yields the pressure on the j^{th} array element. The pressure is computed for each of the M array elements. This process is repeated 1000 times for each scatterer class to generate a set of training and test data.

The range dependent sample volume is studied by calculating the position of the scatterer in the transmit and receive beam. As the beam shape changes with frequency, the incident sound intensity, and received sound intensity will vary in a predictable way. For the simulations considered here, the transducers are assumed to be disk shaped in which case the product of the transmit and receive beam shapes is given by

$$\beta_{j,\boldsymbol{\zeta}}[k] = \left| \frac{2J_1(\mathrm{k}r\sin(\eta))}{\mathrm{k}r\sin(\eta)} \frac{2J_1(\mathrm{k}r\sin(\mu))}{\mathrm{k}r\sin(\mu)} \right|, \tag{2.10}$$

where r is the radius of the transducer, $J_1(x)$ is the Bessel function of the first kind of order 1, and η and μ are the angles between the vector from the transducer to the scatterer position, and the incident and scattered wave vectors respectively, for a particular transmitter-receiver pair. For a wave vector $\mathbf{k}$ and scatterer position $\boldsymbol{\zeta}$, the angles are given by

$$\eta = \arccos\left(\frac{\mathbf{k}_i^T\mathbf{k}_i - \mathbf{k}_i^T\boldsymbol{\zeta}}{\|\mathbf{k}_i\|_2\|\mathbf{k}_i - \boldsymbol{\zeta}\|_2} \right), \tag{2.11}$$

and

$$\mu = \arccos\left(\frac{\mathbf{k}_s^T\mathbf{k}_s - \mathbf{k}_s^T\boldsymbol{\zeta}}{\|\mathbf{k}_s\|_2\|\mathbf{k}_s - \boldsymbol{\zeta}\|_2} \right). \tag{2.12}$$

The geometry for the above calculations is shown in Figure 2.4. For all of the simulations, the parameter r is set to 12 mm, and the components of $\boldsymbol{\zeta}$ selected according to $\mathcal{U}[-5mm, 5mm]$. The horizontal distance from $\boldsymbol{\zeta} = \mathbf{0}$ to the array is defined to be 3 m, and thus the majority of scatterer positions are within the -6dB beam width of the array elements at the highest frequency. To compare the effect of the sample volume on the classification performance, the simulations are performed with and without including the beam shape factor. The case without the beam shape factor is equivalent to setting $\beta_{j,\boldsymbol{\zeta}}[k] = 1$ for all realizations.

The received echoes in any practical system will be corrupted by noise due to reverberation, electronics, and other sound sources. Noise due to reverberation will be in the same frequency band as the received echo where as noise from electronics and other sound sources will have energy in other frequency bands as well as the band of the received echo. Since out of band noise can be reduced by

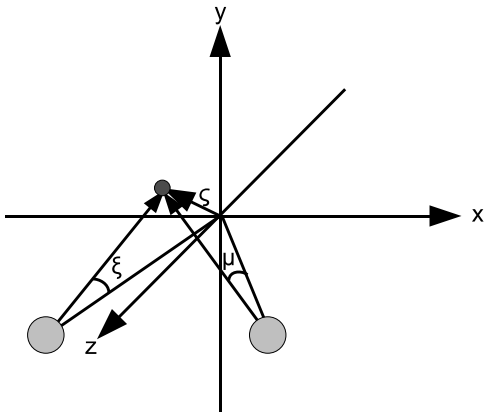


Figure 2.4: View of the method for calculating beam shape for a scatterer positioned at $\boldsymbol{\zeta}$ and a transmitter and receiver pair. The angles between the scatterer position vector, and the vector to each element are defined by η and μ. The angles are used to compute the the change in sound intensity as a function of frequency.

filtering, reverberation noise in the same frequency band as the transmit signal is added to the scattering amplitude. The noise is generated by taking the product of the Fast Fourier Transform (FFT) of the transmit signal $S_0[k]$ with the FFT of a realization of white Gaussian noise $W_j[k]$.

A constant reverberation level is used and thus the signal to noise ratio (SNR) varies as a function of length and orientation of the scatterer. The SNR for the copepod data ranges from -11 to 24 dB whereas the SNR for the euphausiid data ranges from -20 to 35 dB. These ranges for SNR were selected such that the average SNR was close to 7dB for the copepods and 15dB for the euphausiids. These values were deemed to be comparable to what is achieved in practical systems [91, 92]. The noise level can also be defined in terms of an equivalent target strength of -110 dB.

In generating the data, the noise term $W_j[k]$ is first added to the product of the scattering amplitude and the beam shape factor at each wave number bin to yield the noisy scattering amplitude

$$\hat{S}_j[k] = S_j[k]\beta_{j,\boldsymbol{\zeta}}[k] + W_j[k]. \tag{2.13}$$

The convolution defined in equation. (2.1) is accomplished in the frequency domain

by computing the product of the FFT of the transmit signal $S_0[k]$ with the noisy scattering amplitude $\hat{S}_j[k]$. Using equation. (2.13), the pressure signal on the j^{th} element is then obtained by an inverse FFT (IFFT)

$$p_j[n] = \text{Real}\left[\frac{1}{N}\sum_{k=0}^{N-1} S_0[k]\hat{S}_j[k]e^{2\pi ink/N}\right]. \tag{2.14}$$

This process is repeated for each of the M array elements building up the vector

$$\mathbf{p}[n] = (p_1[n], ..., p_M[n])^T. \tag{2.15}$$

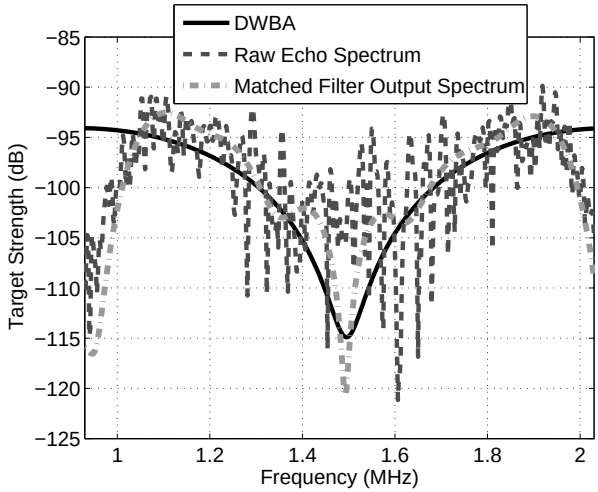
In order to explore the classification success as a function of carrier frequency and bandwidth several different types of signals were used. For the narrow band signals (10% bandwidth) frequencies of 1 and 2 MHz were selected. For the broad band signal, a Linear Frequency Modulated (LFM) chirp was used with a starting frequency of 1 MHz and ending frequency of 2 MHz. The signal duration and energy was kept constant for all signals. The range of frequencies was selected based on past experience with measuring scatter from animals of the size considered here.

To improve the SNR of the data input to the feature extraction algorithms, the raw echo data resulting from the simulation is passed through a matched filter[93]. For the transmit signal $s_0[n]$, and a received echo on array element j defined by $p_j[n]$, the output of the matched filter is

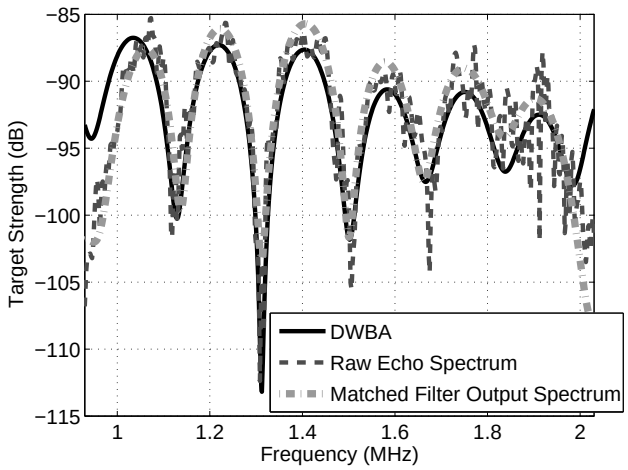
$$\mathcal{M}_j[n] = \sum_{p=-\infty}^{\infty} s_0[p]p_j[p-n]. \tag{2.16}$$

The matched filter output is windowed around the peak in the output with a window size of $W = 50$, corresponding to a time of 5 us or a distance of 7.5 mm. For the broad band signal the time-bandwidth product is 120, yielding a processing gain of roughly 20 dB. For the narrow band signals, the time-bandwidth product is much lower, and thus the processing gain is low as well, around 10 dB.

To visualize the type of noise, and its effect on the estimation of the scattered signal, the model target strength is displayed along with the estimated target strength for the broad band signal type in Figure 2.5.



(a) Copepod



(b) Euphausiid

Figure 2.5: Typical examples of simulated data for both the copepod and euphausiid class. The estimated target strength using the raw echo, and the matched filter output are plotted on top of the target strength as modeled by the DWBA. The high level of noise is clearly visible, as is the improvement obtained from the matched filter.

2.3 Feature extraction

In order to facilitate the classification procedure the data are mapped to a feature space which dramatically reduces the dimensionality of the data while simultaneously highlighting inter-class differences. The three feature spaces used in this work are described in this section.

2.3.1 Single frequency based feature space

To explore the result of using only a single frequency rather than a broad spectrum of frequencies, a single frequency based feature space is defined in which the FFT of the matched filter output on each array element is computed and its magnitude squared is integrated over a small bandwidth. Only the narrow band data is used in this feature space. The resulting sum squared magnitudes on each array element are then combined to form a single feature vector. Specifically, assuming that the j^{th} array element collects N samples, the power in the narrow bandwidth of the signal (k_{min} to k_{max}) is

$$P_j = \sum_{k=k_{min}}^{k_{max}} \left| \sum_{n=0}^{N-1} \mathcal{M}_j[n] e^{-2\pi i n k/N} \right|^2 . \tag{2.17}$$

For the simulations presented here, $N = 1200$. The single frequency feature vector is then defined as

$$\mathbf{y} = (P_1, P_2, ..., P_M)^T . \tag{2.18}$$

The bandwith (k_{min} to k_{max}) is set equal to the transmit signal bandwidth of 10% of the center frequency. The center frequencies used are 1 and 2 MHz.

2.3.2 Discrete cosine transform based feature space

The discrete cosine transform (DCT) has numerous qualities that make it attractive as a feature mapping. For one, the coefficients of the DCT are uncorrelated. It can also be shown, that the DCT can embed most of the energy in the data into a small number of coefficients. While there is no guarantee that such an embedding will yield a discriminant feature space, this is often the case in

practice. The DCT based feature space uses the power spectrum of the matched filter output for the broad band 1-2MHz data. The power spectrum is computed as

$$P_j[k] = \left| \sum_{n=0}^{N-1} \mathcal{M}_j[n] e^{-2\pi i n k/N} \right|^2 . \tag{2.19}$$

For the results presented here $N = 1200$. Having computed $P_j[k]$ for each array element, the DCT of the power spectrum is computed as

$$E_j^{dct}[l] = \sqrt{\frac{2}{N}} \beta[l] \sum_{k=0}^{N-1} P_j[k] \cos\left(\frac{\pi l (2k+1)}{2N} \right), \tag{2.20}$$

where

$$\beta[l] = \left\{ \begin{array}{ll} \frac{1}{\sqrt{2}}, & l = 0 \\ 1, & l = 1, .., N-1 \end{array} \right\}. \tag{2.21}$$

The values of the K largest (ordered by magnitude) coefficients in the transform are retained in the feature vector for the j^{th} array element

$$\mathbf{y}_j = (E_j^{dct}[l^1], ..., E_j^{dct}[l^K])^T, \tag{2.22}$$

where the features are arranged such that $l^1 \leq l^2 \leq ... \leq l^K$. This procedure can be interpreted as an adaptive threshold of the DCT of the power spectrum. Finally, the feature vectors at each array element are combined into a single feature vector

$$\mathbf{y} = \left(\mathbf{y}_1^T, ..., \mathbf{y}_M^T \right)^T . \tag{2.23}$$

The feature vector $\mathbf{y}$ captures the K most energetic wave number bins in the power spectrum of the received signal at each element of the array. A range of values for K were analyzed. It was found that the values $K = 1$, $K = 2$, and $K = 4$ yield the best performance. As K increases, the feature vector is able to capture more of the variability in the frequency response of the scatterer at the cost of a larger feature space dimension. A wavenumber bin width of $\Delta l = 35$ rad/m was used throughout the simulation.

2.3.3 Frequency correlation based feature space

One of the major drawbacks of the previous two feature spaces is that they do not naturally combine the multiple angle data. Instead, the features from each angle of the multiple angle data are lumped together as one big feature vector. This can cause problems in the case of the DCT feature space as the dimensionality of the feature space grows as K times the number of angles M. In this section, a feature mapping which combines the multiple angle data systematically while extracting the features is defined. The features are the eigenvalues of the frequency correlation matrix.

The frequency correlation matrix is obtained by computing the correlation between all pairs of received waveforms in the frequency domain. Specifically, the correlation matrix $\mathbf{C}$ is defined as

$$\mathbf{C} = \mathbf{F}^H \mathbf{F}, \tag{2.24}$$

where

$$\mathbf{F} = (\mathbf{f}_1, ..., \mathbf{f}_M), \tag{2.25}$$

and

$$f_j[k] = \sum_{n=0}^{N-1} \mathcal{M}_j[n] e^{-2\pi i k n/N}, \tag{2.26}$$

is the FFT of the matched filter output for the received data on element j. Thus, the individual elements of the frequency correlation matrix are the cross-correlations between the Fourier transforms of the data on the array elements

$$C_{ij} = \mathbf{f}_i^H \mathbf{f}_j. \tag{2.27}$$

The features are extracted from the frequency correlation matrix using an eigenvalue decomposition[94]

$$\mathbf{\Lambda} = \mathbf{Q}^H \mathbf{C} \mathbf{Q}, \tag{2.28}$$

where $\mathbf{\Lambda} = \text{Diag}\,(\lambda_1, ..., \lambda_M)$ is the diagonal matrix of eigenvalues. The feature vector is then formed by taking the diagonal elements of $\mathbf{\Lambda}$

$$\mathbf{y} = (\lambda_1, ..., \lambda_M)^T. \tag{2.29}$$

The eigenvalues (features) are not simply related to the data at each angle as in the case of the previous two feature spaces. Each eigenvalue is derived from data at all angles. This is the key benefit over the other two feature spaces. The number of non-zero eigenvalues is upper bounded by the number of angles, yet may be lower depending on the degree to which the echoes received at each angle are correlated with one another in the frequency domain. For example, if

$$\mathbf{f}_i^H \mathbf{f}_j \approx \left\{ \begin{array}{l} 0 \ \text{ for } \ i \neq j \\ \kappa \ \text{ for } \ i = j \end{array} \right\}, \tag{2.30}$$

the frequency correlation matrix $\mathbf{C} \approx \kappa \mathbf{I}$ and the eigenvalue spread is nearly flat. In contrast, if

$$\mathbf{f}_i^H \mathbf{f}_j \approx \kappa \ \ \forall \ \ i, j, \tag{2.31}$$

the frequency correlation matrix is approximately rank one, and the eigenvalue value spectrum will be highly peaked at the first eigenvalue. The first example can be thought of as representing a complex shape, where the spectrum of the received signal varies substantially as a function of angle. This second example corresponds to scattering from an angularly symmetric shape. Therefore, in the presence of noise, the variability at each array element is due to only to noise.

2.4 Classification of features

Given a set of features that have been extracted from these data, the next task is to develop a method for assigning a class label to each feature so as to minimize a particular loss function. As is commonly done in pattern classification, the "0-1" loss function is applied which assigns equal penalties to classification errors made for either class [95]. In the zooplankton classification problem considered here, this is a reasonable loss function due to the fact that each class has in effect, equal significance. It can be shown [95] that the classification rule which minimizes the "0-1" loss function is the Bayes decision rule (BDR)

$$i^* = \underset{i}{argmax} \ P_{C|\mathbf{Y}}\left(i \,|\, \mathbf{y}\right), \tag{2.32}$$

where the class i^*, having the maximum *a posteriori* probability given the feature vector $\mathbf{y}$, is chosen. The BDR can be written in terms of the Class Conditional Density (CCD) $p_{\mathbf{Y}|C}(\mathbf{y} \mid i)$ using Bayes Rule, and assuming a prior class distribution $P_C(i)$, as

$$i^* = \underset{i}{argmax}\, p_{\mathbf{Y}|C}(\mathbf{y} \mid i)\, P_C(i)\,. \tag{2.33}$$

In practice, the prior probability may or may not be known. For the procedure considered here, it is assumed that the priors for each class are equal. As a result, that term drops from the maximization. The remaining task is that of maximizing the CCD which is equivalent to computing the Maximum Likelihood estimate of the class label.

Unfortunately, the CCD is almost always unknown. In the best case, only the form of the density is known, but not the parameters that define the actual shape. This is one of the fundamental difficulties encountered in pattern classification and is the point at which *a priori* knowledge or training data must be used to learn about the structure of $p_{\mathbf{Y}|C}(\mathbf{y} \mid i)$.

Here, two popular classifiers are considered: the nearest neighbor (NN) classifier, and the multilayer perceptron (MLP) classifier. The properties of each of these classifiers are briefly reviewed as the implementations used here are standard.

The NN classifier assigns a class label to a new pattern based on the label of the training pattern which is "nearest" to the new pattern according to a particular distance metric. For a given training set $\mathcal{D} = \{(\mathbf{y}_1, i_1), ..., (\mathbf{y}_N, i_N)\}$ where $\mathbf{y}_n$ is a feature extracted from the data according to the methods defined in Section 2.3 and i_n is the associated class label, the NN classifier under the 2-norm assigns the label i_k where k is the index of the nearest neighbor

$$k = \underset{i}{argmin}\, (\mathbf{y} - \mathbf{y}_i)^T (\mathbf{y} - \mathbf{y}_i)\,. \tag{2.34}$$

In contrast to the the NN method, the MLP tries to learn the mapping from feature space to class label space using multiple levels of weighted combinations of the components of the features rather than using the training data explicitly to represent the underlying CCDs. In essence, the MLP learns to approximate $P_{C|\mathbf{Y}}(i \mid \mathbf{y})$ via experience gained from analyzing numerous examples. It has been shown that

this type of classifier can yield very good results in underwater target classification [63] as it is one of the best methods for approximating a high dimensional function.

Given the two class problem, the MLP has two output nodes. The number of input nodes is the same as the number components of the feature vector $\mathbf{y}$. A single hidden layer is used with the number of nodes selected to be twice the number of input nodes. The network is compactly expressed [96] as

$$c_k = U\left(\sum_{j=0}^{2M} \tilde{w}_{kj} V\left(\sum_{i=0}^{M} w_{ji} y_i\right)\right), \tag{2.35}$$

where U and V are non-linear mapping functions, $\tilde{w}_{kj}$ and w_{ji} are network weights, y_i is the i^{th} component of the feature vector, and c_k is the k^{th} component of the classification vector. The weight matrices of the network are initialized randomly at the start of training, and updated at each iteration so as to minimize the error on the training set. Both mapping functions are selected to be the softmax function [95], and the network is trained using the Scaled Conjugate Gradient method. Prior to training and testing, all inputs to the network are z-scaled in the log domain. The training is implemented in MATLAB (The Mathworks; Natick, MA) using the NetLab toolbox [97].

2.5 Results and Discussion

2.5.1 Classifier performance

The classifiers defined in section 2.4 are now evaluated quantitatively on a set of test data mapped into each of the feature spaces defined in section 2.3. The results are displayed as the absolute probability of error as a function of the number of angles (or array elements) that are combined in the classifier. Specifically, the number of angles is equal to the number of array elements included in order starting from element 1. So for example, three angles would correspond to using array elements 1, 2 and 3, and four angles would correspond to using elements 1, 2, 3, and 4. The probability of error is computed according to

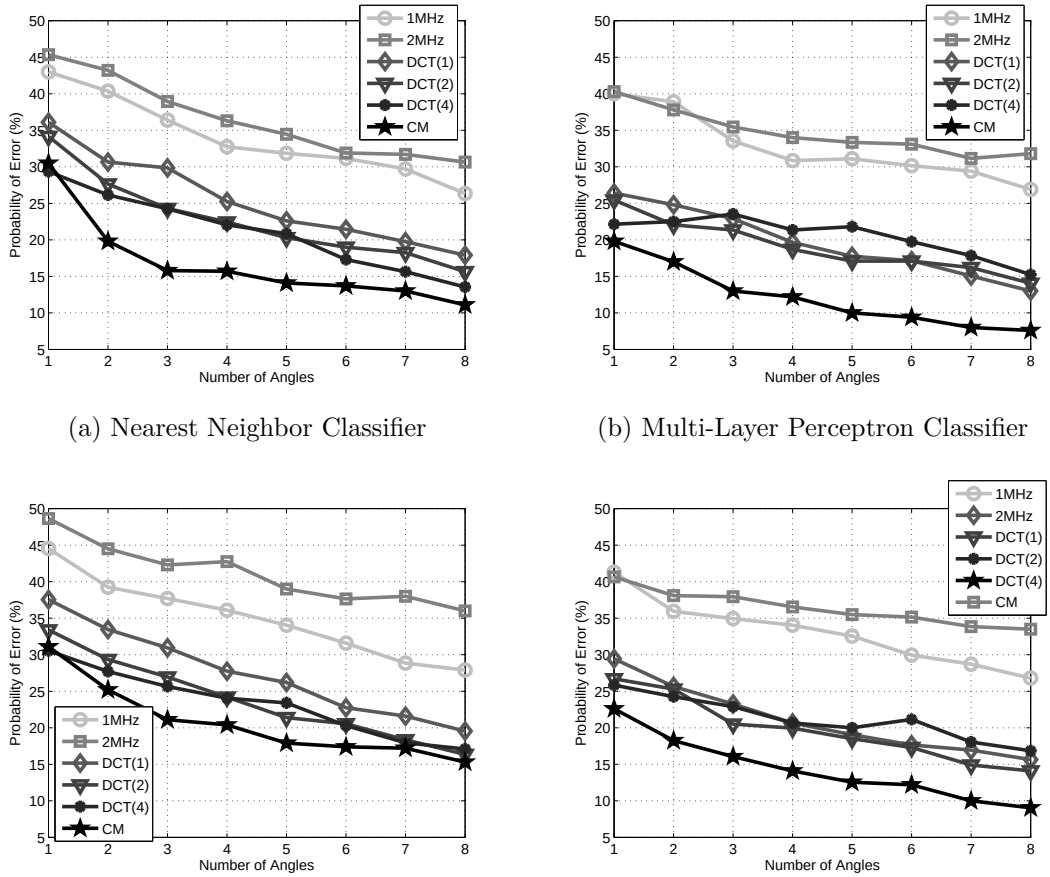
$$p(error) = \frac{p(c|e) + p(e|c)}{2}, \tag{2.36}$$

Table 2.1: The correct and miss-classification probabilities for each classifier when beam shape is excluded from the simulation.

	Copepod		Euphausiid		Total
Classifier	Correct	Miss	Correct	Miss	$p(error)$
NN/1MHz	76.1%	23.9%	71.2%	28.8%	26.3%
NN/2MHz	74.1%	25.9%	64.6%	35.4%	30.6%
NN/DCT(4)	87.5%	12.5%	85.4%	14.6%	13.5%
NN/CM	90.8%	9.2%	87.0%	13.0%	11.1%
MLP/1MHz	76.3%	23.7%	70.4%	29.6%	26.6%
MLP/2MHz	70.3%	29.7%	66.1%	33.9%	31.8%
MLP/DCT(4)	85.0%	15.0%	84.5%	15.5%	15.2%
MLP/CM	91.0%	9.0%	93.8%	6.2%	7.6%

where $p(c|e)$ is the probability of classifying a euphausiid as a copepod, and $p(e|c)$ is the probability of classifying a copepod as a euphausiid. Here, the fact that each class is equally likely in this simulation has been used. A consequence of the equal representation for each class is that a system which randomly guesses the class would have a probability of error of 50%. Therefore, 50% probability of error can be achived with no effort, and any classification strategy should have an error below 50%. The classification experiment is performed both for the case where beam shape is neglected from the simulation, and where the beam shape is included. The results of the classification for both cases of data are shown in Figure 2.6.

The 1 and 2 MHz curves result from using the single frequency feature space and the respective narrow band data. The DCT and CM curves result from using the DCT and frequency correlation feature spaces with the broad band 1-2 MHz LFM chirp data. The DCT(1), DCT(2), and DCT(4) curves apply the DCT method outlined in Section 2.3 with $K = 1$, $K = 2$, or $K = 4$ respectively. Figure 2.6 illustrates that there is a general trend of decreasing probability of error as more angles are used in the classification. The amplitude and frequency response of acoustic scatter from crustacean zooplankton is directly related to the scatterer shape, size and orientation. The addition of more angles in the classifier can be interpreted physically as observing the the scatterer from multiple views. The probability of error is reduced as more angles are used in essence because there



(a) Nearest Neighbor Classifier

(b) Multi-Layer Perceptron Classifier

(c) Nearest Neighbor Classifier (beam shape)

(d) Multi-Layer Perceptron Classifier (beam shape)

Figure 2.6: Comparison between the NN classifier and the MLP classifiers for the case of no beam shape (a,b) and beam shape (c,d), in 6 different feature spaces. The 1MHz and 2MHz curves correspond to the single frequency based feature space. The DCT(1), DCT(2), and DCT(4) feature spaces use the DCT method outlined in Section 2.3 with $K = 1$, $K = 2$, or $K = 4$, bin indices included at each angle. The CM curves results from the frequency correlation feature space outlined in Section 2.3

.

Table 2.2: The correct and miss-classification probabilities for each classifier when beam shape is included in the simulation.

	Copepod		Euphausiid		Total
Classifier	Correct	Miss	Correct	Miss	$p(error)$
NN/1MHz	74.1%	25.9%	70.1%	29.9%	27.9%
NN/2MHz	67.8%	32.2%	60.2%	39.8%	36.0%
NN/DCT(4)	83.5%	16.5%	82.3%	17.7%	17.1%
NN/CM	85.8%	14.2%	81.5%	18.5%	16.3%
MLP/1MHz	75.8%	24.5%	74.2%	25.8%	25.0%
MLP/2MHz	69.9%	30.1%	63.1%	36.9%	33.5%
MLP/DCT(4)	85.3%	14.7%	81.0%	19.0%	16.8%
MLP/CM	91.2%	8.8%	90.7%	9.3%	9.0%

is less uncertainty about the shape of the scatterer.

When only one angle is used, observing scatter over a broad range of frequencies (1-2 MHz) reduces the probability of error substantially over the single frequency case. This is a consequence of the fact that single frequency scatter is much more sensitive to scatterer orientation and size then broad band scatter due to the effect of coherent interference at a given frequency. As broad band data excites many frequencies, it is far more robust to changes in scatterer shape and orientation.

The effect of including beam shape on the probability of error can be clearly seen by comparing the top and bottom rows of Figure 2.6. The effect is essentially to shift the curves towards higher probability of error. This is a consequence of the fact that the random position of the scatterer in the beam adds another kind of noise to the data. However, it is possible that this kind of noise can be corrected by exploiting the angular diversity of the array to locate the scatterer in the beam, and correct for the beam shape. While not considered in this paper, this process will be investigated in future work.

The frequency correlation feature space clearly outperforms all of the other feature spaces for both classifiers. This is to be expected since it is the only feature space which naturally combines the echoes at each angle when extracting features. It may be inferred from the poor performance when only a single angle is used, that the variation in the echo as a function of angle is most discriminant between the two

classes. The frequency correlation feature space efficiently extracts this information from these data. The performance (in terms of correct and miss-classification) of each classifier using all eight angles with, and without including beam shape is displayed in Table 2.1, and 2.2. Interestingly, there is a wide variation in the correct and misclassification results for the different feature spaces. The improvement in classification performance as additional angles, and as a result increased array aperture, are used is a direct consequence of having additional independent views of the scatterer. The additional views reduce the uncertainty in the shape of the scatter by way of the intimate link between scatterer shape, and angularly varying scattering amplitude as defined in equation (2.2). Since the two classes of scatterers have distinct shapes, the reduced uncertainty in shape leads to improved classification performance. In general, all of the feature spaces, and classifiers have slightly higher accuracy for the copepod class rather than the euphausiid class except for the frequency correlation feature space which is inconsistent between the two classification algorithms. This is likely a consequence of the large number of orientations for which scattering from the euphausiid is very weak due to the elongated body. There is also a systematic increase in probability of error for the 2 MHz data over the 1 MHz data for both classifiers. This is likely caused by a greater similarity in scattering amplitude at 2 MHz than 1 MHz between the two classes. This could be caused by the fact that at 2 MHz, the scattering is further into the geometric regime, and thus the scattering amplitude is less sensitive to the scatter size. Finally, the best results, are miss-classifications of 9.0% and 6.2% for the copepod and euphausiid respectively. As a result, the total absolute probability of error is 7.6% in the best case. This gives an improvement over random guessing of 84.8%.

2.6 Conclusions and future work

In this paper, the use of multiple angle acoustic scatter to discriminate between two classes of ecologically important zooplankton has been explored using simulations. The research is motivated by the current need for more descriptive

acoustic sensors for studying zooplankton *in situ*. Past work in this area has been limited by the inherent ambiguity in discrimination ability due to the sensitivity of acoustic scattering to material properties and scatterer orientation. These difficulties have been confirmed here where it has also been shown that it is possible to use scatter measured over a multiplicity of angles to achieve a higher rate of correct classification. Using synthetic data, generated via the use of the distorted wave Born approximation, two ecologically important classes of zooplankton: copepods and euphausiids were classified. The classification performance, measured in terms of probability of error, is dramatically improved over single angle observation methods via the use of additional angles. This improvement is even more substantial when broad band scatter is used.

The simulations performed here were geared towards a practical system which could be deployed in the field. Therefore, constraints were placed on the bandwidth of the transmit signal, and the angular distribution of the receivers in this context. The length distributions for both classes were chosen to be typical of those encountered in the Southern California region [98, 99]. In order to understand the ramifications of the proposed method in the presence of noise, a constant level of noise that resulted in an average SNR of 7dB for the copepod and 15 dB for the euphausiid, or an equivalent target strength of -110 dB, was used. This noise level is consistent with practical systems that have been used in the field [91, 92]. In addition, although a strong effort was expended in order to make the work realistic, the performance of a field system may be limited by issues that have not been considered in this work. Specifically, the models used here, while accurate for weak sound scattering, do not include variability due to individual shape, or body pose. Furthermore, uniform orientation distributions were used here, where as in the field, the distributions may be different. Given the promising results observed here, these additional degrees of freedom certainly warrant further investigation through more complex simulations, as well as observation of live animals.

A curious, but potentially very helpful aspect of our result is that a one dimensional array is capable of capturing enough information from a random three-dimensional orientation to yield good classification performance. It may therefore

be that simple array geometries, which can dramatically reduce the development and deployment cost associated with such systems, constitute a pragmatic solution to the *in situ* classification of zooplankton after all.

2.7 Acknowledgements

The authors would like to thank D. E. McGehee, M. Benfield, D. V. Holliday, and C. Greenlaw for development and maintenance of the Advanced Multifrequency Inversion Methods for Classifying Acoustic Scatters website, two anonymous reviews for helpful comments on the manuscript, and California Sea Grant for funding this research.

The text of Chapter 2 is in part and under some rearrangements a reprint of the material as it appears in Paul L. D. Roberts and Jules S. Jaffe, "Multiple angle acoustic classification of zooplankton," *J. Acoust. Soc. Am.*, **121 (4)**, 2060–2070, April 2007. The dissertation author was the primary researcher and author, and the co-author listed in this publication directed and supervised the research which forms the basis for this chapter.

Chapter 3

Classification of live, untethered zooplankton from observations of multiple-angle scatter

3.1 Introduction

Crustacean zooplankton play a major role in the ocean's ecosystem, so it is important to develop non-invasive methods to measure their abundance and behavior. Instruments deployed in the laboratory and field have measured sound scatter from a wide range of animals [34, 36, 100, 37, 84, 47]. In addition, scattering models and classification algorithms have been formulated [38, 81, 84, 48] with the ultimate goals of quantification of animal size and abundance, identification of different taxa [36, 28, 77], and measurements of in situ behavior [101, 92].

The fundamental challenge to achieving these goals arises from the vast diversity of zooplankton in the ocean and the confounding influence of size, shape, orientation, and material properties on acoustic scatter [56, 55]. Variations of these factors lead to substantial ambiguities in using acoustics to both identify and count animals. Reducing these ambiguities, while retaining a system that is practical for fieldwork, would be of great value.

One potential solution is to observe sound that has been reflected at dif-

ferent angles. In the context of diffraction theory, Jaffe (2006) demonstrated that swim-bladder size could be accurately inferred from multiple-angle acoustic reflections observed from a single fish [58]. Roberts and Jaffe (2007) used numerical methods to demonstrate that individual copepods and euphausiids could be classified using multiple-angle, wide-band reflections [59]. The study indicated that the multiple-angle method was more accurate than other techniques using either narrow- or wide-band sound with a single transceiver.

Here the multiple-angle technique was applied to two types of marine zooplankton: copepods and mysids. Copepods are an order of crustacean zooplankton (typically 1-4 mm in length) distributed throughout the world's marine and fresh waters. Mysids are common coastal inhabitants similar in size, shape, and composition to euphausiids (typically 5-10 mm in length). Copepods and euphausiids are dominant taxa of marine ecosystems and there is great interest in quantifying their distributions with remote, non-invasive methods.

A laboratory scattering apparatus was constructed to record simultaneous reflections from zooplankton at multiple observation angles. To be compatible with available hardware, eight receivers were evenly spaced on a line, forming a 2-m-long array. The length of the array was chosen so that it could eventually be deployed on an autonomous underwater vehicle (AUV) or glider. The even spacing of receivers sampled the available aperture uniformly. No optimization of receiver position or array length was attempted, but the final geometry was consistent with simulated multiple-angle experiments [59]. During recording of acoustic reflections, two video cameras simultaneously recorded the size, position, and orientation of animals. Multiple-angle observations were analyzed using a correlation matrix approach designed to highlight changes in scattering across the array. Eigenvalue analyses of these matrices demonstrate that, in our laboratory setup, multiple-angle acoustic data can be used to accurately discriminate between copepods and mysids.

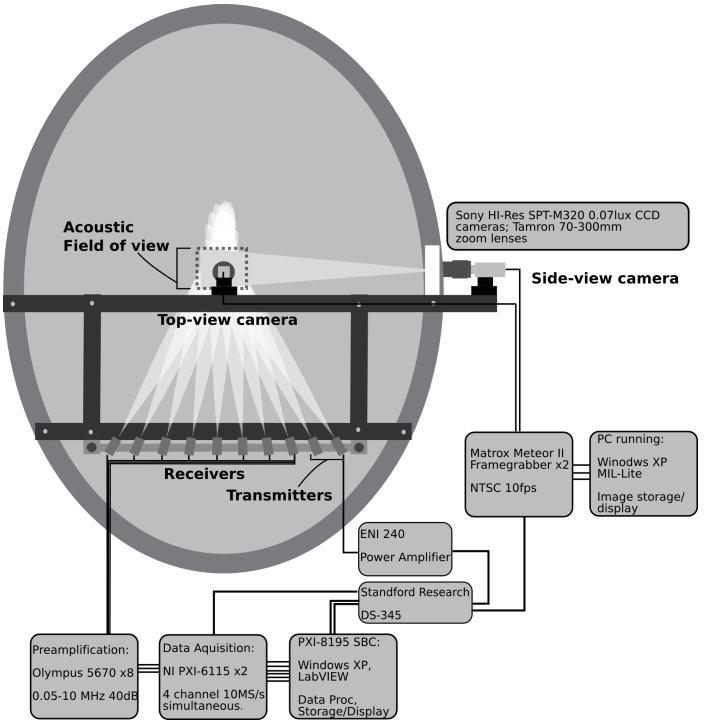
3.2 Experimental methods and data processing

3.2.1 Acoustic scattering apparatus

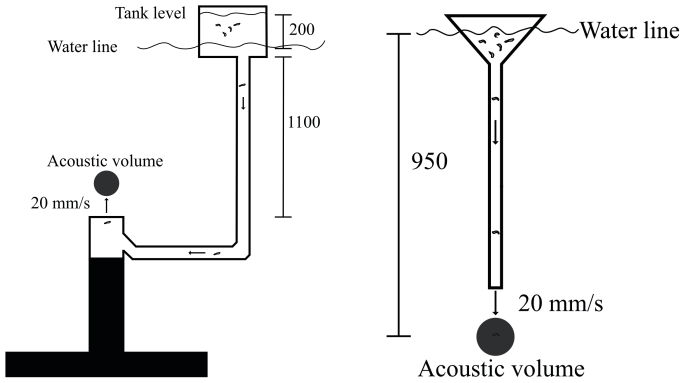
The scattering apparatus consisted of a linear array of 10 disk transducers (Panametrics, Waltham, MA). Eight were used as receivers and two as transmitters. All receivers and one transmitter were 19-mm-diameter, 2.25-MHz broadband transducers (V305-SU). The other transmitter had a diameter of 38 mm (V395-SU). The larger transmitter was used to increase the source level, improving the signal to noise ratio (SNR) of echoes from smaller animals. The field of view (FOV), defined by the overlap among all acoustic beams, was 42 ml with the 38-mm transmitter and 328 ml with the 19-mm transmitter. Optical images within the FOV were acquired using two high-sensitivity VGA cameras (Sony SPT-M320). The acoustic array and cameras were rigidly connected by a Unistrut (Unistrut Corporation, Wayne, MI) system to ensure minimal relative movement during experiments [Figure 3.1(a)]. A custom-built rail system was used to slide the array in and out of the water. This allowed the array to be kept dry when not being used, and then precisely positioned and locked in place for experiments. All experiments were performed in a 3.0-m-wide, 4.2-m-long, 2.4-m-deep elliptical tank with view ports 1.2 m above the bottom. The tank was filled with chilled, filtered seawater maintained at a temperature of 14.9^o C throughout the experiments.

Acoustic data were acquired by a National Instruments (Austin, TX) PXI-8195 controller running Windows XP (Microsoft, Redmond, WA) with two PXI-6115, 10 MHz, 12 bit, 4-channel simultaneous sampling boards with 64 MB of on-board memory. The output from each receiver was fed through a Panametrics 5670 broadband preamplifier prior to digitization. The transmitter was driven by an ENI (Bell Electronics, Kent, WA) AP400B 400-W power amplifier. Waveforms sent to the power amplifier were generated by a Stanford Research (Sunnyvale, CA) DS345 arbitrary waveform generator. Software developed in LabVIEW (National Instruments) controlled the acquisition, recording, and real-time display of data.

A single PC running Windows 2000 controlled the stereo video system. Images were “grabbed” from each camera – at an adjustable rate controlled by



(a)



(b) (c)

Figure 3.1: (a) The multiple-angle scattering apparatus (viewed from above) showing the acoustic and optic elements, Unistrut (Unistrut Corporation) frame, experimental tank, and associated data acquisition components. (b) The bottom-up pump. Animals are drawn out of a tank above the surface and injected below the field of view. (c) The top-down pump. Animals are sedated and allowed to sink through the pipe into the field of view. All distances are in millimeters.

the acoustic transmissions – using synchronized Matrox (Matrox, Dorval, Canada) Meteor II frame grabbers. Software developed in C++ used the Matrox Imaging Library to read images from the boards, bundle them together, display them in real-time, and save them to disk.

To obtain high-contrast images of animals, a 200-mW laser was used for illumination (Aixiz 200 mW, 650-nm module). A wavelength of 650 nm was selected as it is almost invisible to the animals yet suffers limited attenuation through the medium. The laser beam was spread with a diverging lens to yield a cone of light that intersected the FOV.

3.2.2 Experimental setup

Preliminary experiments revealed that scatter from tethers maintaining zooplankton in the FOV dominated observations at these high frequencies. Therefore, a substantial challenge was to position live, untethered animals in the FOV. Copepods were pumped from a small holding tank through a system of hoses and out through a 3-in-diameter pipe positioned directly under the FOV [Figure 3.1(b)]. Copepods typically stayed near the FOV for several seconds after exiting the pipe, whereas mysids quickly swam away from the FOV. To mitigate this problem, mysids were sedated by placing them into a dilute (1% by volume) water bath of clove oil and filtered seawater. They were kept in the bath until they ceased swimming (typically 30-60 s) but retained leg movement. Sedated mysids were then transferred immediately to a funnel system [Figure 3.1(c)] that guided them into the FOV while they sank slowly. These mysids eventually recovered as inferred by their swimming behavior.

3.2.3 System alignment and calibration

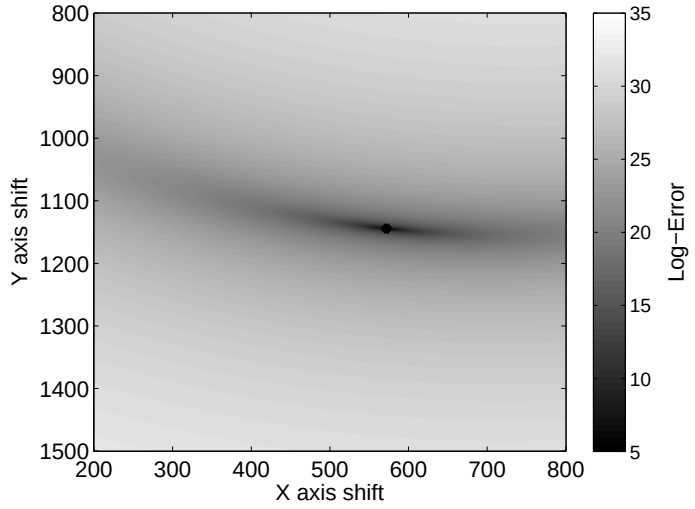
Alignment and calibration of the multiple-angle system were more complicated than for a typical mono-angle system. Transducers were aligned using a long strand of nylon monofilament suspended perpendicular to the array at a point that was designated the system's origin. The monofilaments diameter of 75 μm resulted

in an acoustic scattering pattern that was omni-directional in the horizontal plane, but narrow in the vertical plane. Pointing angles for all transmitters and receivers were iteratively adjusted to maximize reflections.

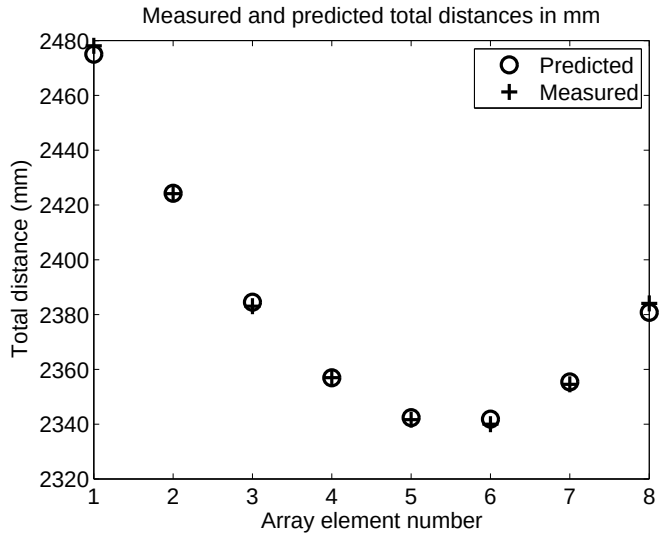
Three-dimensional transducer positions were estimated using the time-of-arrival at each receiver of the echo scattered from the monofilament. With the array elements located on a line, their Euclidean offsets from the origin were determined by computing the mean squared error between the observed and predicted arrival time of echoes at each receiver, for all possible offsets. A clear minimum was found [Figure 3.2(a)] and the model output, compared to measured data, yielded excellent agreement [Figure 2(b)]. The array was off center with a total angular span of roughly 46.7^o (Figure 3.3). The horizontal distance from the center of the array to the origin was 1.14 m and the angular spacing between array elements was on average $5.8^o +/- 1^o$.

The video system was aligned with the acoustic system using a small sphere suspended in the middle of the FOV. One camera was mounted so that it looked through the tank's view port (Side View). The second camera was mounted in a waterproof housing with a small view port and submerged almost directly above the FOV (Top View). Both cameras were mounted to the Unistrut frame using heavy-duty, three-axis telescope mounts (Losmandy DCM2, Los Angeles, CA). These mounts allowed each camera to be rotated until the target was in the center of each frame. They were then locked in place.

Before and after experiments, acoustic calibrations were performed using a 1-mm-diameter tungsten-carbide sphere, following the calibration procedure outlined by Foote [102]. Echoes were collected while the sphere was translated within the FOV. Since the echo spectrum varied only slightly for displacements of the sphere within the FOV, the beam pattern was assumed to be constant in that region. The echo spectrum from the sphere was then used to convert recorded echo spectra to target strength (TS). This process was repeated with the nylon monofilament using the scattering model described in [103]. To check consistency of the calibration, echo spectra –converted to target strength using each calibration method –were compared and found to be within 3 dB. This was adequate for the



(a)



(b)

Figure 3.2: Calibration results for the array. (a) The error surface for the x and y shift parameters using the calibration data, the inferred sound speed, and the measured element spacing. (b) The modeled and measured total acoustic path distances for each receiver. The total acoustic path is the path from the transmitter, to the origin, and then to the receiver.

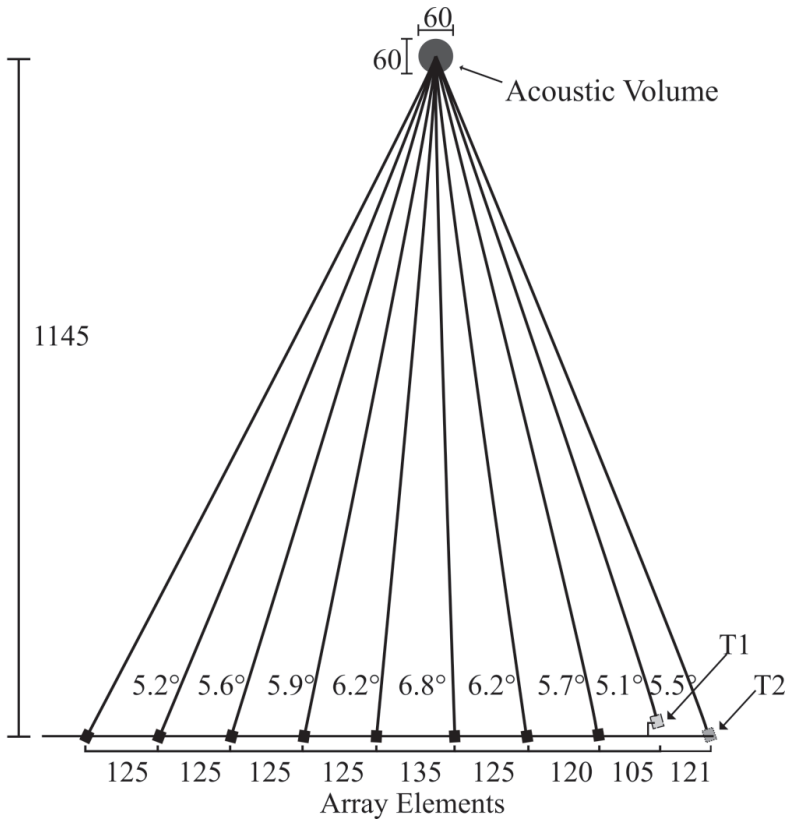


Figure 3.3: The geometry of the acoustic array. All distances are in millimeters. The two transmitters are at the right end of the array. Receiver element 1 is at the left end of the array and element 8 is at the right end of the array.

subsequent analyses as the method only relies on the relative calibration among receivers.

3.2.4 Data acquisition and processing

Multiple-angle data were collected during a series of experiments spanning several months. All zooplankton were collected in the La Jolla Cove area (San Diego, CA) by small boat and immediately brought back to the lab. Mysids were collected by gently dragging a mesh butterfly net across the kelp at the surface of the kelp forest. This procedure typically resulted in 10-100 mysids ranging in length from 8-12 mm. Copepods were collected in a series of net tows using

a 250-μm cod-end, 1-m-diameter net. This yielded several thousand copepods (primarily calanoid) ranging in length from 1 to 4 mm. Animals were allowed to equilibrate with seawater siphoned directly from the FOV for 30 minutes prior to being injected into the tank. During experiments, between one and ten copepods were injected toward the FOV at a time. Only one mysid was injected at a time. Of those animals injected, roughly 10% actually passed through the FOV.

The transmit signal was a linear frequency-modulated (LFM) chirp from 1.5-2.5 MHz with a cosine-squared envelope and a duration of 500 μs. The data-acquisition system recorded ten sequential echoes with 100-ms delay between them. The large tank and relatively small volume of intersection between transmitter and receivers allowed a nearly reverberation-free echo. However, in data from the mysid experiments, there were some small artifacts caused by the injection pipe. These were coherently removed in post processing using data recorded from the pipe alone.

Raw acoustic data were matched filtered with a synthetic model of the transmit signal [104, 93, 55]. The matched-filter output was then windowed to localize the echo from the animal. A window length of 250 time samples (25 μs at 10 MHz sample rate) was selected to capture the longest possible echo for the largest animal insonified. Due to the extended length of the transmit signal, the matched-filter processing gave an SNR improvement of roughly 23 dB over a very short pulse of equivalent power. This processing gain was critical for obtaining good SNR from these weakly scattering animals. The same matched filter was also applied to calibration data. A 4800-point fast Fourier transform (FFT) was used to estimate power spectra of echoes recorded by each receiver. The FFT of each received echo was multiplied by the power spectrum predicted by the calibration model and divided by the power spectrum of the calibration echo. This corrected for the shape of the transmit pulse and the small variation in element sensitivity across the array.

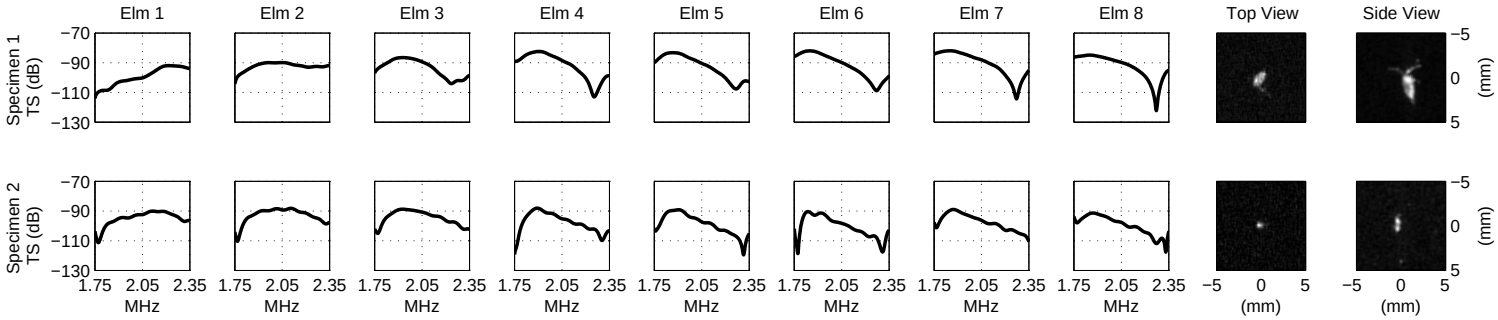


Figure 3.4: Example data for two different copepods. Data are plotted for each receiver element 1 though 8 as calibrated target strength versus frequency. Images on the far right show top and side views of the copepods at the time of insonification. These images are oriented so that the array is at the bottom of the top-view image with element 1 on the left. For the side-view image, element 1 is oriented into the page, and element 8 is out of the page.

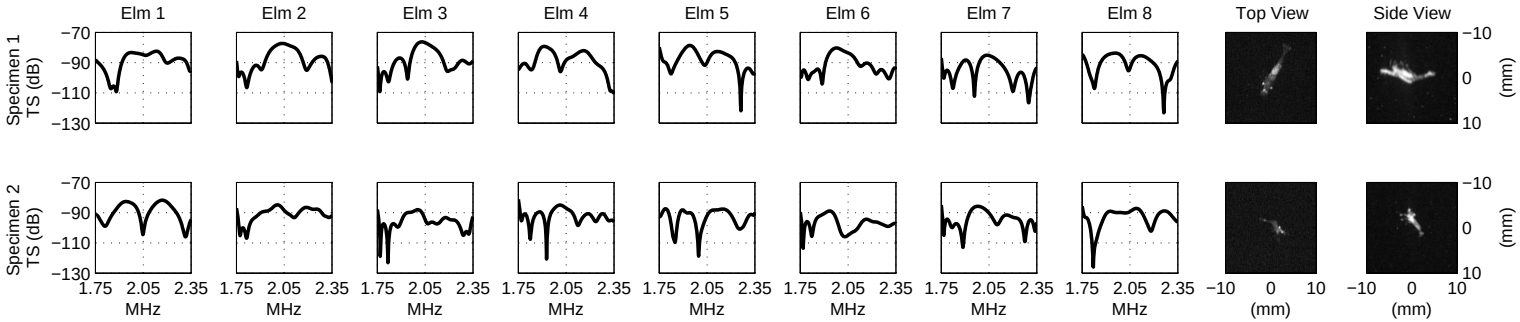


Figure 3.5: Example data for two different mysids. Data are plotted for each receiver element 1 though 8 as calibrated target strength versus frequency. Images on the far right show top and side views of the copepods at the time of insonification. Images are oriented identical to Figure 4. Note that the image scale is changed from Figure 3.4.

To highlight fundamental differences between echoes from each class of scatterers, a frequency correlation algorithm was developed. Let the M-point FFT of the windowed, matched-filter output for the j^{th} element be $F_j[k]$. Then define the cross-correlation between the positive frequency coefficients of the FFTs of two elements a and b as

$$X_{a,b}[m] = \sum_{k=0}^{k=N} F_a^H[((k))_N] F_b[((k-m))_N] \text{ for } 1 \leq a \leq b \text{ and } 1 \leq b \leq b, \quad (3.1)$$

where $((x)))_N$ denotes "x modulo N" and $N = (M+2)/2$. The cross-correlation with the maximum magnitude was then used to form an approximately Hermitian, positive-semidefinite matrix

$$G_{a,b} = X_{a,b}[m^*], \quad (3.2)$$

where

$$m^* = \underset{m}{argmax} \ |X_{a,b}[m]| . \quad (3.3)$$

As explained in [59], the matrix G will be nearly rank 1 if there is an equal correlation between all pairs of receivers. In contrast, if there is a weak correlation between non-identical pairs of receivers G will be similar to a scaled identity matrix.

To quantify the degree of correlation the eigenvalue decomposition [94] was computed as

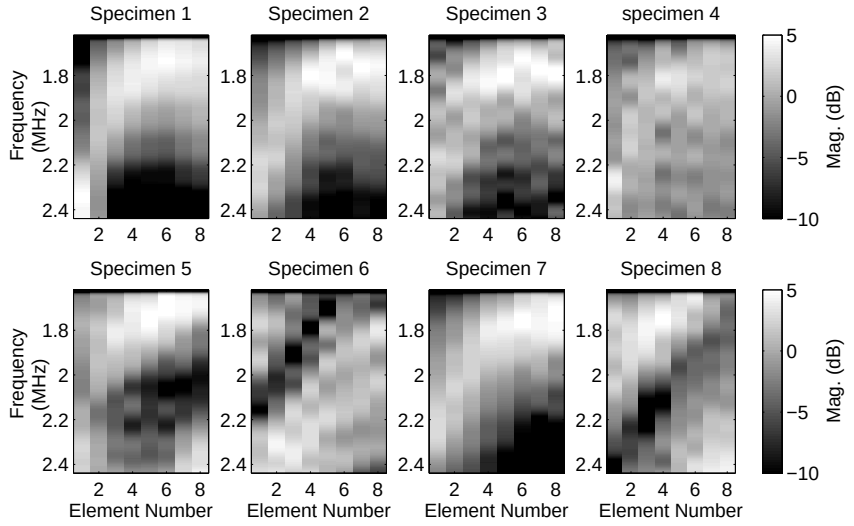
$$\Lambda = Q^H G Q, \quad (3.4)$$

where $\Lambda = \text{Diag}(\lambda_1, ..., \lambda_8)$ is a diagonal matrix of eigenvalues. In practice there was a substantial amount of correlation among echoes across the array, and the eigenvalues deceased logarithmically. Therefore, the log-eigenvalue spectrum was used for analysis.

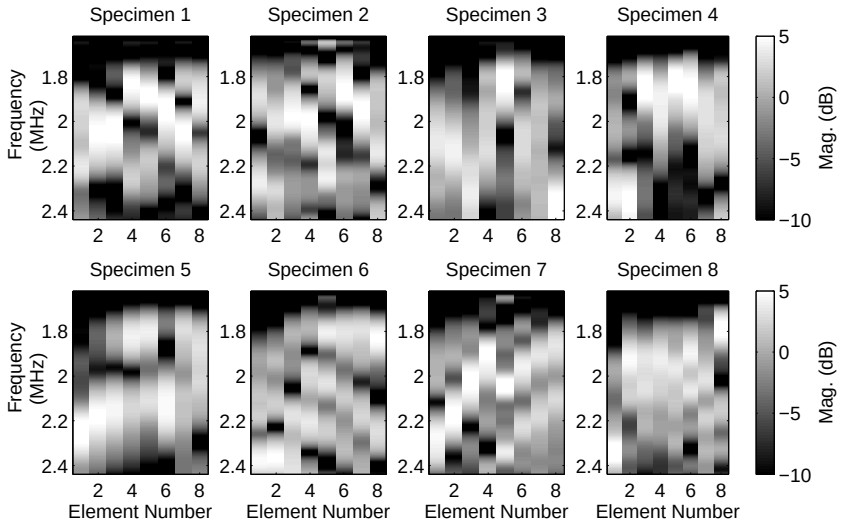
3.3 Results

3.3.1 Scattering apparatus and experiment analysis

The transducer array, video system, and data-acquisition hardware worked well for collecting repeatable, multiple-angle scattering measurements when animals were positioned in the FOV. The combination of a rigid frame, with robust



(a)



(b)

Figure 3.6: (a) Normalized magnitude of the scattering spectrum as a function of frequency and array element number for all eight copepod specimens. These data highlight the slowly varying nature of target strength as a function of frequency and look angle. (b) Normalized magnitude of the scattering spectrum as a function of frequency and element number for all eight mysid specimens. These data highlight the relative increase in complexity of target strength as a function of frequency and look angle when compared with (a).

rotation mounts for all sensors allowed simple and straightforward alignment that remained unchanged throughout the experiments. The rail system for translating the array in and out of the water proved invaluable and was constructed using standard, off-the-shelf parts at small cost. The camera system's low resolution, coupled with the interlaced video signal and poor laser-beam characteristics, yielded images of only moderate quality, though they were adequate for this study.

Positioning live specimens in the FOV without causing artifacts in the data was the most challenging aspect of the experiments. Numerous methods were evaluated during the development of the system (including a wide variety of tethers) but the injection method provided the only artifact-free data. Unfortunately, the injection frequently added bubbles to the acoustic field, or the animal would move out of the FOV before reflections were recorded. Therefore, few artifact-free echoes were obtained during experiments. Data are presented here from eight individual copepods and eight individual mysids. The small size of the data set is solely a consequence of the lack of an efficient means for repeatedly placing untethered, live animals in the FOV.

3.3.2 Multiple-angle data analysis

Multiple-angle data require additional processing to characterize the target animals. In the first step, the spectra of echoes at each angle were computed (Figures 3.4, 3.5). Despite the small sample size, patterns in target strength data were clear. These highlighted the influence of animal orientation on echo spectra, motivating the frequency correlation processing. For copepods, target strength curves (Figure 3.4) were very similar among receivers. In addition, target strength was slowly varying across frequencies. Video observations indicated that both specimens shown in Figure 3.4 were nearly broadside to the array. However, a small tilt can be inferred from a null in the spectra moving from elements 1 to 8. Multiple-angle data can therefore offer enhanced insight into the animal's orientation. For mysids, target strength is less similar among receivers than with copepods, and varies more as a function of frequency (Figure 3.5). Video data showed that, for the two specimens shown in Figure 3.5, orientation in the horizontal plane differed

by roughly 90o (Figure 3.5, top-view images). Specimen 1 was nearly end-on to element 1 and nearly broadside to element 8. Specimen 2 was nearly broadside to element 1 and nearly end-on to element 8. Orientations can also be inferred from acoustic data: there is a decrease in spectral complexity as the animal orientation becomes closer to broadside. Near broadside, spectra are smoother with a well-defined null. This null is likely a result of interference between sound reflected from the sides of the mysids body closest to and farthest from the array element. To highlight differences among data sets, multiple-angle spectra for each specimen were combined together to form an image. The spectrum for each angle was first normalized by its standard deviation to remove relative differences in the average reflected energy. When plotted as a function of frequency and angle (Figure 6) this normalized spectral magnitude shows substantial similarity among frequency and angle for copepods [Figure 3.6(a)] with more variability for mysids [Figure 3.6(b)]. Frequency correlations among spectra of angular data were computed using equations (1) and (2). Log-eigenvalue spectra (Figure 3.7) indicate that echoes at each angle are more correlated for copepods than for mysids: the slope of the average spectrum for copepods is roughly twice that for mysids. Furthermore, the average spectrum for copepods decreases steadily, whereas for mysids it is rather flat up to the sixth eigenvalue at which point it steadily decreases.

3.4 Discussion and conclusions

Experiments using a multiple-angle acoustic receiver array and live copepods and mysids have shown that it is possible to use the scattered acoustic signal to distinguish between these zooplanktonic taxa. Preliminary experiments showed that signals from animals tethered in the FOV were dominated by scatter from the tether. Techniques were developed to introduce live, untethered animals into the FOV; however, data quantity was limited by the low success rate in positioning animals in the FOV. In a real pelagic environment these factors would not be an issue.

Multiple-angle data (Figures 3.4, 3.5 and 3.6) exemplify a fundamental

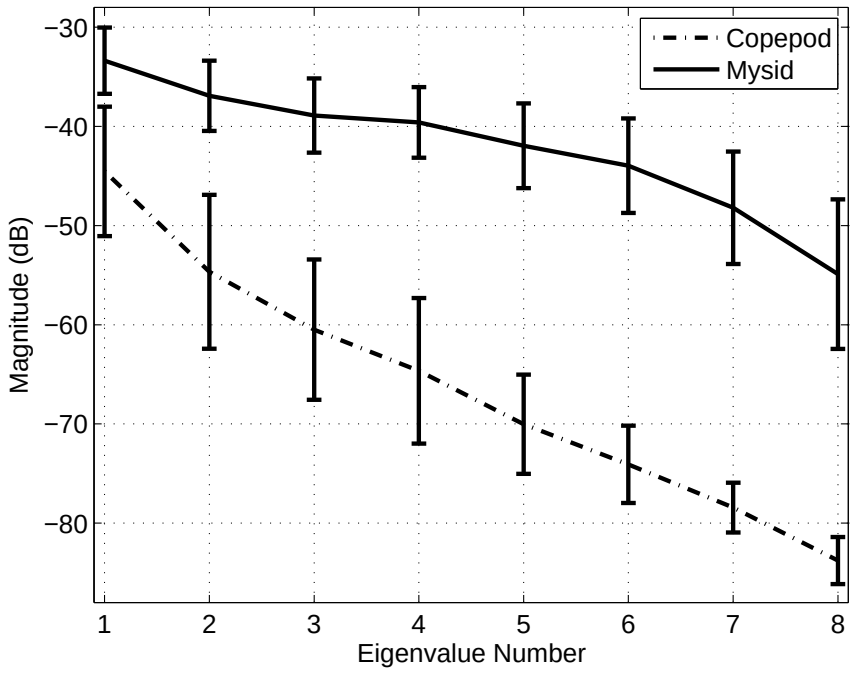


Figure 3.7: (a) Normalized magnitude of the scattering spectrum as a function of frequency and array element number for all eight copepod specimens. These data highlight the slowly varying nature of target strength as a function of frequency and look angle. (b) Normalized magnitude of the scattering spectrum as a function of frequency and element number for all eight mysid specimens. These data highlight the relative increase in complexity of target strength as a function of frequency and look angle when compared with (a).

principle of sound scattering from weak scatterers: The scattered sound field in the immediate vicinity of a target, and its radiated pattern in the far field are Fourier transform pairs [90]. Therefore, variability in target strength is controlled by the size and shape of the scatterer. A thicker scatterer at a given orientation permits more variability over frequency for a fixed bandwidth. Likewise, a more elongate scatterer permits more variability over angle. This can complicate the interpretation of single-angle, wide-band scatter when animal orientation is unknown [105]. A comparison of data from single angles (Figure 3.6) reveals that single broadband echoes from copepods and mysids can be quite similar depending on the orientation of the animal relative to the system. However, when multiple angles are considered this similarity is dramatically reduced.

A further advantage of the method presented here is that the transducers only need to be inter-calibrated and not absolutely calibrated. This method is therefore relatively immune to biofouling, so long as the biofouling occurs equally for all array elements. A potential disadvantage is that multiple elements require more sophisticated hardware and computer processing than a single-element system. However, several existing systems use multiple transducer configurations for measuring Doppler shifts to infer currents, so this is not a problem – even in a battery-powered instrument. Furthermore, the computational burden is modest.

While this study has demonstrated the utility of a multiple-angle array for zooplankton identification, there remain several research issues requiring further consideration. One concerns optimizing the array geometry: the number of elements, the distance between them, and the overall length of the array. Another concerns extension of the processing methods, once the data are in hand. The eigenvalue method described here works well in analyzing the presented lab data; however, alternate analyses should be explored.

Naturally, the real challenge in zooplankton sensing is to implement the method in the ocean. Beyond the basic issue of putting electronic instruments in a corrosive, high-pressure environment, there is the added problem of the large diversity of animals: the many compositional types and body morphologies make enumeration and identification difficult. One approach might be to combine the

use of broadband, multiple-angle scattered sound with target-strength observations of individuals. Such a system would generate information about both the size and reflectivity of the animals, with reduced sensitivity to their absolute orientation. This will greatly reduce the number of candidate animals corresponding to a measured set of reflections, enhancing our ability to discriminate among them. Incorporating these ideas into a sea-going system will increase our knowledge about the marine planktonic ecosystem and the role of zooplankton in the regulating the dynamics and fluxes through that ecosystem.

3.5 Acknowledgements

The authors would like to thank the SIO machine shop for assistance with construction of the scattering apparatus, Eddie Kisfaludy, Erdem Karakoylu, Fernando Simonet, Ben Maurer, and Robert Glatts for technical consulting on the scattering apparatus design and help with experiments, two anonymous reviewers and Peter Franks for helpful comments on the manuscript, and the California Sea Grant for funding this research.

The text of Chapter 3 is in part and under some rearrangements a reprint of the material as it appears in Paul L. D. Roberts and Jules S. Jaffe, "Classification of live, untethered zooplankton from observations of multiple-angle acoustic scatter," *J. Acoust. Soc. Am.*, **124** (2), 796–802, Aug. 2008. The dissertation author was the primary researcher and author, and the co-author listed in this publication directed and supervised the research which forms the basis for this chapter.

Chapter 4

A multi-view, broadband approach to acoustic classification of marine fish

4.1 Introduction

Acoustic methods for studying fish populations offer the potential for a rapid, non-invasive assessment that can be cost effective and operate over considerable ranges, day or night. These methods have become routine in most fisheries applications [106, 107, 108]. They provide a fundamental datum that can define far reaching fisheries management decisions and increase our understanding of the number and distribution of fishes in the ocean. One of the major challenges associated with extracting information about fish from acoustic scatter is that the scattering from fish is strongly influenced by fish orientation [52, 109]. In some cases, this sensitivity has been shown to dominate over errors in scattering models [53] and lead to biases in estimates of abundance [54]. This difficulty arises primarily from the conventional way acoustic methods are typically employed in fish estimation problems. A single-frequency echo recorded from either a fish school or individual fish is converted into an estimate using a model which has many unknown parameters. These parameters are typically assigned values using historical

data, net-tow data, or avoided all together by using averaged models. However, this measurement paradigm may not yield enough information to constrain these parameters and therefore they ultimately must be estimated from data.

In addition to estimating fish size, shape, and abundance using acoustic scatter, there has also been considerable interest in classification of fish. Species classification has proved to be a challenging problem [107], however recent studies have shown some success in this area using echogram analysis [110, 62, 61], multi-frequency methods [111], and broadband methods [51]. In general, methods that use behavioral information encoded in echograms (see [62] for example) show very good performance and methods that use energy-based properties of the scattered sound have also shown some success [51]. Due to the typical lack of specificity (echo energy is not inherently discriminant at a given frequency due to coherent interference) associated with acoustic scatter, approaches which take into consideration behavioral information (for example school shape, position tracking, or swimming speed) should be more successful than those that use echo data alone. However, observation of fish behavior can be a time consuming and costly process, and modifications to existing echo-energy methods that improve accuracy would be highly valuable.

The subject of classifying targets from observations of scattering has received significant attention during the last three decades (Table 4.1). In general, the classification methods can be grouped into three major categories:

1. • Hidden Markov model methods [72, 64, 112, 71, 70].
2. • Multi-layer Perceptron (MLP) methods [113, 114, 63, 115, 116].
3. • Information-theoretic and Bayesian probability methods [117, 118, 119, 67].

Most of these methods have been developed in radar and sonar applications. More recently, these methods have been applied in fisheries acoustics [117, 50, 108]. In addition to HMM and MLP based methods, collaborative agent methods for improving the performance of decision-fusion classifiers have been recently been developed for underwater object detection in sonar imagery [75, 120]. These methods

improve upon the very good performance obtained with decision-fusion methods such as [63], but do not require a constant or known number of observations.

Classification of fish in situ using a single view is challenging because of the strong influence of fish orientation on acoustic scatter, the great degree of varability in fish size and shape, and the sensitivity of acoustic scatter to internal structures in the fish body such as the swim bladder. During the past four decades many methods have been developed to improve on length and orientation estimation [121], individual fish classification [122, 113, 114], and classification of fish schools [61, 62]. A primary advantage of these methods is that they typically operate within the existing acoustic paradigm employed in current practice. This allows a direct application of these methods to current data. However, because most of them are single-view by design, they remain sensitive to unknown fish orientation. Because of the strong dependence of scatter on fish orientation, it is possible that by directly measuring scattering over multiple views, additional information about the length, orientation, and possibly taxa of the fish could be collected and used to improve estimates of these unknown parameters.

During the last decade, several methods have been developed for processing multi-view scattering (Table 4.1). Although the application areas are different, and the scatterers themselves (typically man-made strong scattering objects) are quite different, many of these methods can be applied to fish classification with good success.

Table 4.1: List of echo classification methods developed in fisheries acoustics and sonar target classification during the past three decades. Methods are grouped by application area and classification approach. Feature transformations are listed with references to representative works.

Application / Classification Approach	Representative works (Feature transforms)
Fisheries acoustics	
• Gaussian mixture models	Volume backscatter [117]
• Echo statistics	Echo energy [123, 122, 124], Frequency vs. angle spectra [50]
• Length estimation, Multi-beam	Echo duration [108]
• Wide-band tank reverberation	Total target strength [125]
• MLP	echo envelope [113, 114, 51] cepstral features [126]
• GLRT	Echo intensity [118]
Sonar target classification	
• Decision fusion, MLP and linear	Wavelet packets [63]
• MLP	Wavelets [127, 115], STFT [116], CCA [128, 129]
• RVM, Gibbs sampling	Time reversal imaging [69]
• Adaptive KNN, MLP	LDA [130]
• Adaptive multi-aspect HMM	Matching pursuit [71, 74]
• Multi-aspect HMM	Matching pursuit [72] [68], Scattering centers [112]
• Dual-HMM	Wavelets [64]
• GRNN	PCA [131]
• Aspect-independent classifier	Eigenvalues of wavepackets [132]
• Combined MLP-HMM	Wavelets [133], Wavelet packets [70]
• Full-spectrum classify-before-detect	STFT, GLRT [134]
• Collaborative multi-aspect agents	Wavelet packets [75, 120]

A useful means for categorizing multi-view classification strategies is the method used to combine information for each view together. Feature-fusion methods [72, 70] typically estimate the joint probability of the set of views using a model for how the statistics of the views change. Other feature-fusion methods include the cosine-transform and correlation-based based algorithms in [59] and the frequency vs. angle spectra processing in [50]. One of the main advantages of feature fusion is that it can take advantage structural changes in the echoes from each view to increase accuracy. This is especially advantageous in cases where the scattering from each class is strongly dependent on angle, and the scatterers have significant inter-class variability in size and shape.

In contrast to feature-fusion methods, decision-fusion methods combine the outputs of classification algorithms which can be either real valued or binary valued. Typically, these methods compute posterior class probabilities for each view independently and then combine the views together using a fusion algorithm that is optimized using training data [63]. The primary advantages of these methods is that in each step they make discriminant decisions and can potentially model more complex features. In addition, these methods can combine views together in a way that is discriminant. The draw back is that they generally can not be adapted to handle additional views without retraining the system so they have limited application in problems where the number of sonar reflections collected for a given target is variable.

A third type of fusion can be defined as collaborative fusion [120]. In collaborative fusion, each view defines an independent agent who makes decisions based on their observed feature vector and also on the decisions of the other agents. By selecting probabilistic methods for combining agent decisions, the algorithm can achieve better performance than decision fusion without the constraint of a fixed number of views. This approach has all of the benefits of decision fusion, but does not require a fixed or known geometry allowing it to be applied to problems were views are collected randomly, and the total number of views is not known a priori. In addition, collaborative fusion allows decisions made from single views to be compared with decisions made from multiple views. Because of this last feature,

collaborative fusion can outperform decision-fusion [120].

Adapting multi-view fusion methods to fish classification could offer significant benefits for improving classification accuracy with only modest changes (adding additional receivers) to current measurement systems. As fish are quite different scatterers than the man-made elastic scattering objects studied in most multi-view classification studies, different feature spaces and classification algorithms must be developed to process scattering from these animals in an appropriate manner. This paper explores the development of three different classification algorithms and three different feature spaces for classifying broadband, multi-view scatter recorded from individual fish in a laboratory. The problem is framed in the ecologically relevant context of classifying fish as a function of their size, or species. It is assumed that a set of training data is available for each fish which includes observations of broadband scattering collected by eight receivers spanning 42^o of aperture for several hundred uniformly randomly orientations of the fish. Algorithms are then tested on data unseen during training. The primary objective is to develop a classification algorithm that: achieves low error, yields the most dramatic reductions in error as more views become available, and makes the least number of assumptions about collection geometry.

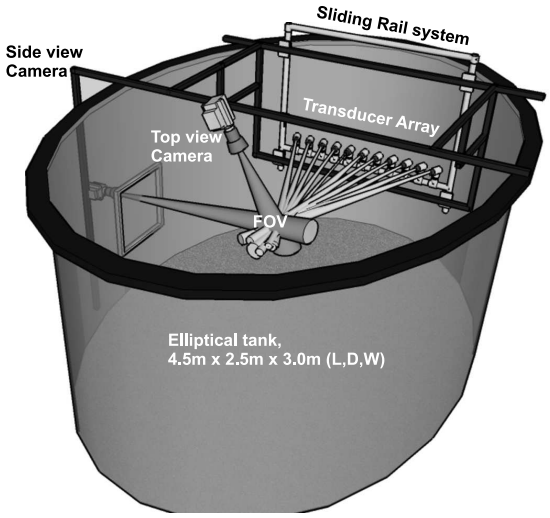
4.2 Multi-view acoustic scattering data

Broadband acoustic reflections from several species of fish were recorded using a recently developed laboratory scattering apparatus (Figure 4.1a). Fish were sedated using an MS-222 protocol (UCSD IACUC, protocol # S07191) and harnessed in the acoustic field by a 75 μm diameter, monofilament line. Each fish would typically come to rest so that its head and tail were in the acoustic plane of the receivers. Once stabilized, fish were rotated through 360^o several times about the vertical axis while echoes were recorded simultaneously on all eight receivers. Synchronous with acoustic reflections, stereo video data were recorded and used to estimate animal size, shape, and orientation. The system alignment and calibration are discussed in detail in Chapter 3.

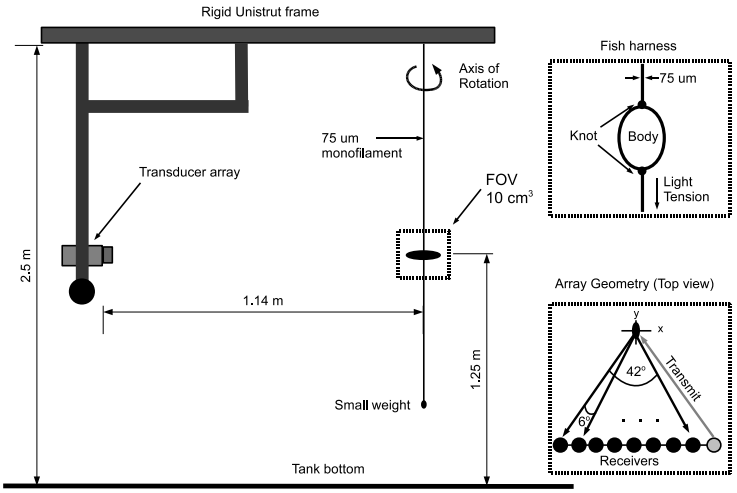
The transducer array was positioned 1.25 m off the bottom of the tank. The fish was positioned 1.14 m away from the array, approximately in the center of the tank. The acoustic field of view was 5 cm^3 for the highest frequencies used and grew larger for lower frequencies. A very small harness was made from 75 μm nylon monofilament (Figure 4.1b) which held the fish body roughly horizontal in the tank and permitted rotation about the vertical axis. The use of an acoustically minimal harness was critical to ensure that the signal from the harness did not dominate the scattered signal and bias measurements of scattering. A small weight was used to apply enough tension on the harness so that the fish could not easily escape, and to ensure that the fish would stay roughly in the same horizontal position while rotating.

Data were collected in two groups in February 2007 and September 2007 using different transmit signals. The first group spanned eight different species ranging in length from 1 to 14 cm (Table 4.2). Transmit signals were broadband linear frequency-modulated (LFM) chirps with center frequencies of 525 and 775 kHz, 250 kHz of bandwidth, and a cosine-squared envelope.

Acoustic and video data collected during the experiments were initially analyzed by hand to check for artifacts, and remove records where the fish had moved out of the field of view, or the field was contaminated by bubbles. Once the records for each fish had been checked, an automated algorithm was used to estimate the orientation of the fish from the top view camera. This algorithm was supervised to ensure that the orientation estimates were meaningful. Using the estimated orientation, the data sets were indexed to extract samples distributed approximately uniformly in θ.



(a)



(b)

Figure 4.1: (a) Drawing of the scattering apparatus showing the transducer array, sliding rail system for moving the array in and out of the water, unistrut frame, top- and side-view cameras, and elliptical tank. (b) The geometry of the problem. The transducer array is located at 1.25 m off the bottom of the tank. The transmitter and receivers are aligned to intersect at a common point 1.14 m away from the array. The total angular span of the receivers (bottom right) is approximately 42^o with 6^o angular sampling. The fish are held in the field of view (FOV) of the array by a 75 μm nylon monofilament harness (top right) which maintains the fish in a horizontal aspect throughout the experiments. Very light tension is applied to the harness using a small weight to ensure that the fish rotates about the vertical axis and does not drift out of the FOV.

Table 4.2: The set of all specimens used in laboratory experiments. Data were collected at two non-overlapping bandwidths of B_1 = 400-650 kHz and B_2 = 650-900 kHz. Fish species, length range, number of individuals studies, transmit signal bandwidth, and total number of echoes collected are listed.

Species	Lengths	Specimens	Bandwidth	# Echoes
Northern Anchovy (Engraulis mordax): S_1	10-12 cm	2	B_1	2000
Pacific Sardine (Sardinops sagax): S_2	12-14 cm	2	B_1	2000
Blacksmith (Chromis punctipinnis): S_3	5.75-6.25 cm	2	B_1	2000
White Sea Bass (Atractoscion nobilis): S_4	6.8-7.1 cm	2	B_1	1000
Garibaldi (Hypsypops rubicundus): S_5	4-5 cm	2	B_1	1000
Blue Damselfish (Chrysiptera cyanea): S_6	3-5 cm	6	B_1, B_2	3000
Four-Striped Damselfish (Dascyllus melanurus): S_7	1.5-3 cm	6	B_1, B_2	3000
Green Damselfish (Chromis viridis): S_8	2-5 cm	6	B_1, B_2	3000
Blue Yellowtail Damselfish (Chrysiptera parasema): S_9	3-4 cm	6	B_1, B_2	3000

4.3 Acoustic data processing

Several approaches are typically used to extract acoustic features from echoes. Wavelets [63, 133], matching pursuit [72, 68], and echo energy or shape [118, 113, 114] have demonstrated good performance. For scatterers that have significantly different scatter complexity as a function of view angle, cross-correlation in frequency between multiple views lead to a marked decrease in error as opposed to other methods [59]. In this paper, several different feature spaces are explored.

4.3.1 Data preprocessing

The fish scattering experiments performed in February 2007 and Sept 2007 yielded 500 GB of scattering data which included thousands of echoes from the background, miscellaneous scatterers in the tank, fish (the intended subjects) and a host of other data which was not usable for a variety of reasons. To transform data to a usable form, several prepossessing steps were applied.

Raw acoustic data were filtered with a matched filter derived from a synthetic model of the transmit signal. This allowed echoes from the fish to be localized with high resolution and separated from other scatterers in the tank. The matched filter output is computed as [93]

$$\mathcal{M}_j[n] = \sum_{p=-\infty}^{\infty} s_0[p] p_j[p-n], \tag{4.1}$$

where $p_j[n]$ is the raw pressure signal on the j^{th} receiver and $s_0[n]$ is the transmit signal model. Due to the pulse-compression gain associated with this broadband signal, this processing step yields a SNR increase of roughly 25 dB. The output signals $\mathcal{M}_j[n]$ are windowed around the peak with a window size of 1000 time samples, or 100 μs.

The most dominant features evident in the acoustic data were the shape and duration of the echo which varied systematically with fish orientation, size, and shape. To extract echo shape, a Hilbert transform [135] was used to estimate the envelope of the echo.

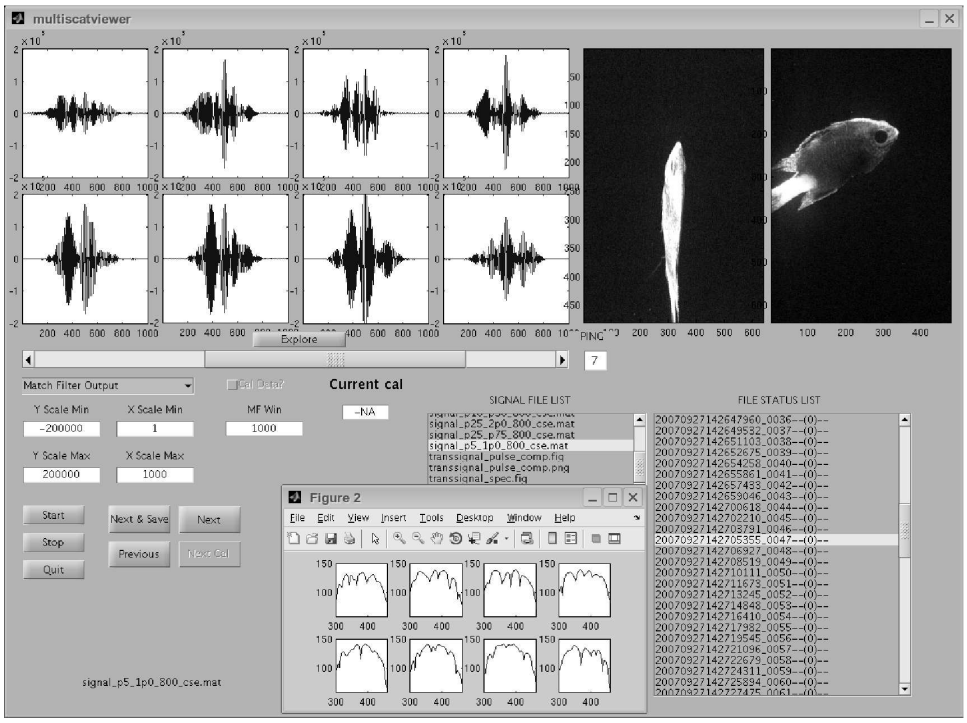


Figure 4.2: The Multiscat software package showing the analysis of a single snapshot of scattering from a damselfish. The viewer shows the matched filtered output on each receiver (top left) along with the estimated echo spectra (bottom middle). Available data files are displayed in a list which can be easily browsed and selected for viewing (bottom left). Different signal models and calibration data can be loaded and applied to data. The top- and side-view images of the fish at the moment of insonification are displayed in the top right of the figure. Data can be browsed by ping number, set number, or file index.

A custom software package (Figure 4.2) written in MATLAB (Mathworks, Natick, MA) Guide was developed for manually investigating data. This was critical in the early stages of preprocessing to identify usable sections of data, check for artifacts both in video and acoustic data, and evaluate possible adjustments to the data collection method that should be applied in subsequent experiments.

The manual data parsing yielded a very large set of data (roughly 3000 pings per fish) which sampled the available orientations of the fish in a non-uniform manner. In order to avoid biasing the estimates of classification performance by comparing records with different average fish orientations, data were processed according to the diagram given in (Figure 4.3).

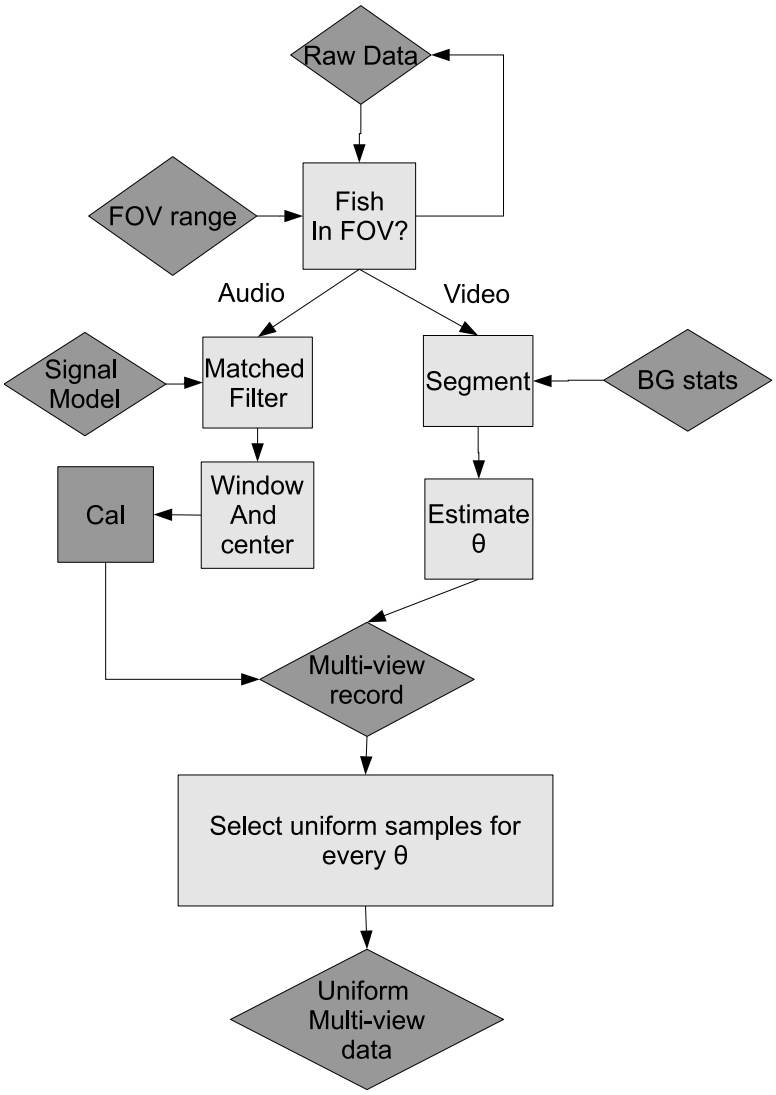


Figure 4.3: The data preprocessing used to parse through the raw experimental data and generate a set of uniformly sampled (in θ) data records when fishes were in the field of view of the system. The system automatically assigns an orientation to each fish as defined by its appearance in the top-view camera image. Assigned orientations are used to ensure data for each fish can be sampled uniformly.

The automated orientation detection was accomplished by applying background subtraction to video images with a unimodal Gaussian distribution for each pixel [136] and a threshold that was set to three standard deviations above the mean. This segmentation typically yielded many foreground objects which could then be removed from consideration by computing the area of each object (counting the 8-neighbor connected foreground pixels, [137]) and selecting the object with the largest area. The orientation of the object was then estimated by computing the covariance matrix between the x_1 and x_2 coordinates of the connected pixels of the object and defining the orientation as the direction of the principle eigenvector of the matrix. For most fish this yielded an accurate estimate for their orientation. These orientation data were then sampled to ensure a nearly uniform distribution to build up data sets of scattering for each fish.

The raw signal to noise ratio was typically between 10 and 20 dB. Matched filtering increased this to 35 to 45 dB which is comparable to achievable SNR for broadband field systems [51]. However, to reduce the possibility of artifacts from the experimental system biasing results, and to yield a more conservative estimate of performance, white Gaussian noise was added to the matched filter outputs to yield an effective SNR of 30 dB.

4.3.2 Cosine-based features

The discrete cosine transform (DCT) can yield features useful for discrimination between scatters of different shape and size [59]. Here, the first 44 coefficients of the DCT of the echo envelope were used as features

$$y_j[l] = \sum_{k=0}^{N-1} \mathcal{E}_j[k] \cos\left[\frac{\pi l(2k+1)}{2N}\right], \quad \text{for } l = 0, 1, ...43, \tag{4.2}$$

where $\mathcal{E}_j$ is the echo envelope estimated with the Hilbert transform. These 44 coefficients typically captured over 98% of the energy in the echo envelope.

4.3.3 Wavelet features

The wavelet transform, as well as several variations (wavelet packets, for example) have been applied to underwater target classification problems [63, 127, 115, 133]. The hallmark of the wavelet transform is its general ability to compress real-world signals into a very small number of coefficients. This property is due in part to the fact that the transform can be represented as a linear filter bank with constant Q filters, where Q is the ratio of the center frequency of the filter to its bandwidth. The wavelet transform approximation $a_\beta[n]$ and details $d_\beta[n]$ coefficients satisfy the recursive relations

$$a_\beta[n] = \sum_i h[i-2n]a_{\beta-1}[i], \tag{4.3}$$

and

$$d_\beta[n] = \sum_i g[i-2n]a_{\beta-1}[i], \tag{4.4}$$

where β is the level of the transform, h is the low-pass FIR filter, g is the high-pass FIR filter, and $a_0[n] = s[n]$ where $s[n]$ is the original signal. Equations (4.3) and (4.4) show that the approximation coefficients are formed by recursively low-pass filtering the input signal and downsampling. In contrast, the detail coefficients at level β are formed by high-pass filtering $a_{\beta-1}$. For a given level β, the combination of the approximation and detail coefficients simply give the low- and high-pass filter outputs of $\beta-1$-level approximation to the original signal. Here, the wavelet coefficients of the echo envelope $\mathcal{E}_j[k]$ for $\beta = 6$ are used as features

$$\mathbf{y}_j = (\mathbf{a}_6^T, \mathbf{d}_6^T)^T, \tag{4.5}$$

where $\mathbf{a}_6$ and $\mathbf{d}_6$ are the sixth-level coefficients computed from the j^{th} echo envelope.

4.4 Classification algorithms

The problem of interest is to classify a sequence (either in time, or space, or both) of observations. Application of the Bayes Decision Rule (BDR) under the "zero-one" loss function yields an optimal classifier [95]

$$c^* = \underset{c}{\operatorname{argmax}}\ P(c|\mathbf{y}_1, .., \mathbf{y}_M). \tag{4.6}$$

While Equation. (4.6) offers an optimal solution in general, for a specific application the posterior density is typically unknown and difficult to estimate directly from limited data. Therefore, approximate methods must be used. When the observations are separated in time and space, application of a hidden Markov model has shown considerable success [72, 66, 64, 68, 70]. However, Multi-layer perceptrons have also been applied to fuse multiple observations [63]. A comparison study showed that out of several different methods, Support Vector Machines yielded the lowest classification error [88] however this study did not consider multi-view fusion.

The enhancement in classification accuracy as a result of combining several views depends on two key factors: (1) the change in scatter between views is class-dependent, and (2) the single view scatter is sufficiently unique between classes. Both (1) and (2) are typically satisfied when the classes consist of different size or shape scatterers. Scatterers of equivalent size and shape may also be discriminated if their internal structure is sufficiently unique [72]. However, (2) is typically violated when scatterers from a given class show considerable intra-class variability in size as this typically dominates the scattering signal.

Three classification strategies are explored in this work. The algorithms investigate feature- , decision-, and collaborative-fusion based on kernel machine classification and estimation of posterior class probabilities.

4.4.1 A_1: Multi-view feature-fusion algorithm

A commonly used algorithm is feature fusion (Figure 4.4) in which features computed from all of the views are combined together to form a single feature vector. A special case of feature fusion is where features are computed from individual views and then stacked together to form one large feature vector. In this case, the feature vector used for classification is given by

$$\mathbf{Y} = (\mathbf{y}_1, ..., \mathbf{y}_M)^T, \tag{4.7}$$

where $\mathbf{y}_j$ is the feature vector computed from the j^{th} view. As this approach was evaluated for zooplankton classification [59] and yielded good performance, it is adopted here as well. One of the potential drawbacks is that the dimension of the feature vector grows linearly with the number of views. To compensate for this increase in dimension, Principle Components Analysis (PCA) or Linear Discriminant Analysis (LDA) can be applied to reduce the dimension of the feature space while retaining features with significant variance (in the case of PCA) or the best linear discrimination (LDA).

Given the set of combined feature vectors $\mathbf{Y}$, a support vector machine (SVM) classifier is trained using the one-vs-rest method [138] where C binary SVM classifiers are trained to separate one of the C classes from the rest. The decisions made by each of the binary classifiers are then combined together to yield the final classification. The width of the kernel function σ and the soft-margin parameter of the SVM were selected using 5-fold cross-validation for each feature space. The SVM used a radial basis function (RBF) kernel of the form

$$K(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{||\mathbf{x} - \mathbf{y}||^2}{2\sigma^2}\right). \tag{4.8}$$

It was found that the best value for σ increased as additional views were added to and therefore σ was set according to

$$\sigma(M) = 4 + 1.5 * M, \tag{4.9}$$

where M is the number of views to be fused. This result held independently of the feature transform used as the features were z-scaled prior to training and testing. Optimization of the SVM learning problem was carried out using the SVM^{light} [139] code and interfaced into MATLAB using the Spider toolbox [140].

Although the SVM classifier has been shown to outperform several other popular methods for underwater target classification [88], two other classification algorithms based on the K-Nearest Neighbor (K-NN) [95] and Multi-class

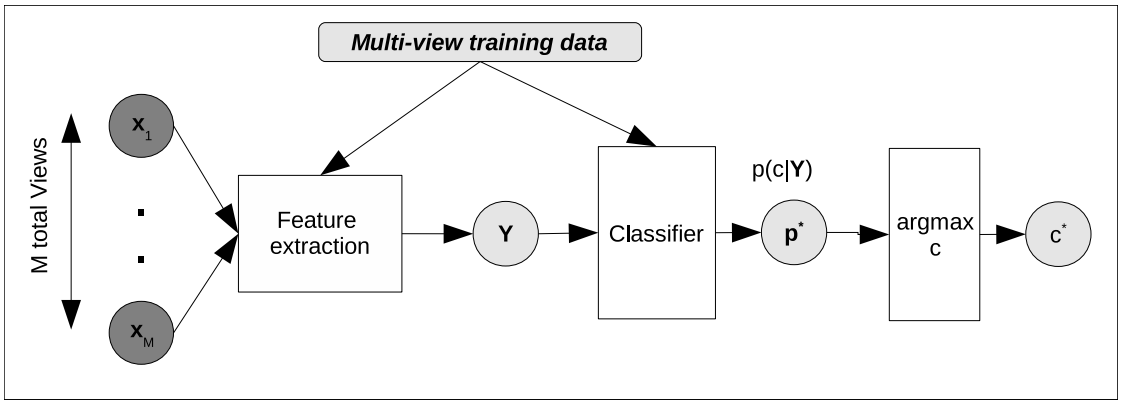


Figure 4.4: The feature-fusion algorithm. Data vectors from each view are combined together and a feature transformation is computed to yield a combined feature feature vector $\mathbf{Y}$ incorporating information from all M views. A single classifier is trained using a multi-view training set.

AdaBoost [141, 142] classifiers were also tested. For the K-NN classifier, cross-validation was used to select the number of neighbors. For all of the feature transforms, a value of $K = 3$ was selected. For the Multi-class AdaBoost algorithm, decision trees were used as the weak-learners, and the improvement in performance was found to level off after ten iterations.

4.4.2 A_2: Multi-view decision-fusion algorithm

The decision-fusion algorithm used in this work is described in the diagram given in Figure 4.5. This algorithm extends the approach used in [63] to the case of SVM classifiers with posterior class probability outputs and multi-class classification. In the first level of the algorithm, M SVMs are trained to classify features from each view independently. The resulting classifications are converted to probabilities, and passed as a whole to a final SVM which is trained using the same data set used to train each individual classifier. In the following description the term posterior class probability is used interchangeably with confidence and does not imply a strict probability.

The classifiers used to compute the single-view posterior class probabilities

are of the general form

$$f_\theta(\mathbf{x}) = \mathbf{w} \cdot \mathbf{\Phi}(\mathbf{x}) + b. \tag{4.10}$$

Many of the popular nonlinear classification algorithms (eg. SVM, RVM, MLP) take this form [143, 144]. Here, SVMs are used for both the single-view feature classification and the multi-view fusion. The SVM output has the form [138]

$$f_\theta(\mathbf{x}) = \text{sign}(\mathbf{w} \cdot \mathbf{\Phi}(\mathbf{x}) + b). \tag{4.11}$$

By removing the sign operation in equation 4.11, the SVM returns real-valued numbers which indicate the distance of the example to the separating hyperplane in feature space. Therefore the SVM does not inherently provide a way to obtain posterior class probabilities. Several methods have been developed to estimate posterior probabilities from the SVM output. However, it is known that converting outputs from sparse learning algorithms into posterior probabilities can be error prone. A simple explanation for this is that only the examples that lie on or in the margin will impact the definition of the SVM classification boundary and therefore the distance of points which lie near training examples that are not support vectors is not directly related to the probability that the example belongs to the given class. When using the SVM with one-vs-rest classification, C SVMs are trained and each one outputs the distance of the example to the separating plane which divides the training data into one class vs. the other $C-1$ classes. A simple way to convert these distances into probability estimates is to use the softmax function

$$p(c|\mathbf{x}) = \frac{exp(\mathbf{w}_c \cdot \mathbf{\Phi}(\mathbf{x}) + b)}{\sum_{c=1}^{C} exp(\mathbf{w}_c \cdot \mathbf{\Phi}(\mathbf{x}) + b)}. \tag{4.12}$$

Equation (4.12) treats the distance on the positive side of the margin for each of the C SVMs as an indicator for relative degree to which the example may belong each class. The output is bounded to be on the interval $[0, 1]$, and for examples which lie on or within the margin, the softmax function gives a good estimate for the posterior probability. For examples which lie far away from the margin, the probability estimate is poor, however in these cases, the examples are typically well separated and a precise estimate of the probability is not as important.

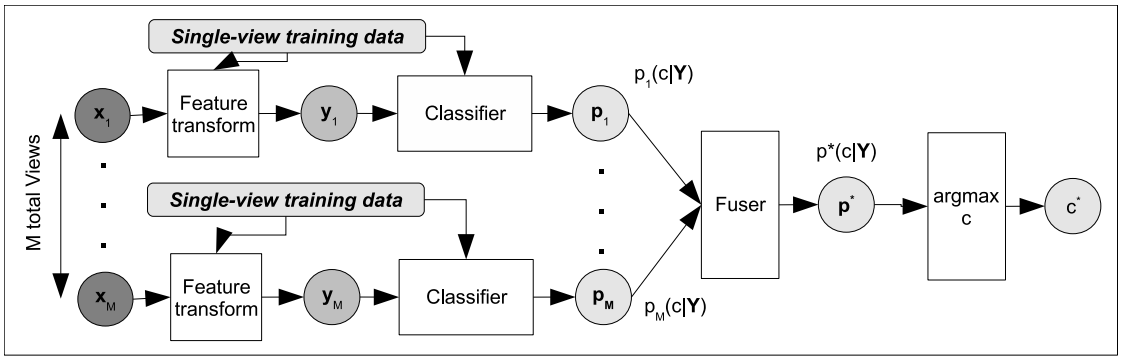


Figure 4.5: The decision-fusion algorithm. An SVM is used to classify features from each view. Each SVM outputs an estimate for the posterior probability using the softmax function. These outputs are combined into a probability vector which is classified by a second SVM.

To fuse the posterior probability estimates from each view together, a probability vector is defined for each view

$$\mathbf{p}_j = (p(c = 1|\mathbf{y}_j), ..., p(c = C|\mathbf{y}_j))^T . \tag{4.13}$$

The posterior probability vectors for every view are then grouped together to yield one large probability vector containing the probabilities for every class based on each view

$$\mathbf{P} = (\mathbf{p}_1, ..., \mathbf{p}_M)^T . \tag{4.14}$$

This vector is then used for classification by training a second SVM to map $\mathbf{P}$ to a class label. The same RBF kernel and widths were used as in equations (4.8) and (4.9). Several other approaches were tested using for example MLPs or RVMs, but the results were not found to be significantly different. The SVM offers much faster training than the RVM and many fewer parameters to adjust than the MLP and was used throughout this work.

4.4.3 A_3: Multi-view collaborative-fusion algorithm

The primary advantage of the decision-fusion algorithm, over feature-fusion algorithms such as HMMs, is that the decisions are made in a discriminant manner in each step of the process. Preliminary testing of HMM-based algorithms on

fish scattering data revealed that common probabilistic models such as Gaussian mixtures were typically not able to yield good separation between classes. Furthermore, the implicit sequential nature of the HMM was not technically satisfied by the data collection process as the views were collected in a bistatic, simultaneous fashion. Recently, collaborative fusion strategies have been developed to address some of the limitations of feature- and decision-fusion while maintaining the very desireable quality of the HMM that additional views can be incorporated without retraining [120]. They key component of these methods is decision feedback, where decisions made using a subset of views are sent back to be used in updating the decisions made based on each view. In addition, emphasis is placed on combing decisions from multiple views without training a classification algorithm which permits the trained algorithm to incorporate new views without retraining. The fusion algorithm used here is an extension of the one developed in [120] which incorporates SVMs for single-view classification, applies to multi-class problems, and uses confidence measures to obtain weights for combining decision feedback. The general form of the algorithm is given in Figure 4.6.

The algorithm defines a set of M agents, one for each view. These agents collaborate with each other to reach a final prediction about the class of the unknown target. At the start of the algorithm, the j^{th} agent estimates the posterior class probability given the j^{th} view, $p(c|\mathbf{y}_j)$ using the probabilistic SVM method defined in Section 4.4.2. The process yields a set of posterior probability estimates for each class. In the collaborative agent context, these probabilities are treated as measures of confidence in the unknown target belonging to one of the C classes. The algorithm given in [120] was applied to binary classification where the confidence was defined to be a scalar for each view, μ_j on the range $[0, 1]$. For the multi-class problem considered here, the confidence is defined by the vector of posterior probabilities from the j^{th} view

$$\boldsymbol{p}_j = (p(c = 1|\mathbf{y}_j), ..., p(c = C|\mathbf{y}_j))^T. \tag{4.15}$$

Each agent transmits its probability vector $\boldsymbol{p}_j$ to all of the other agents. Therefore, after initially yielding a single-view prediction about the class, each agent receives

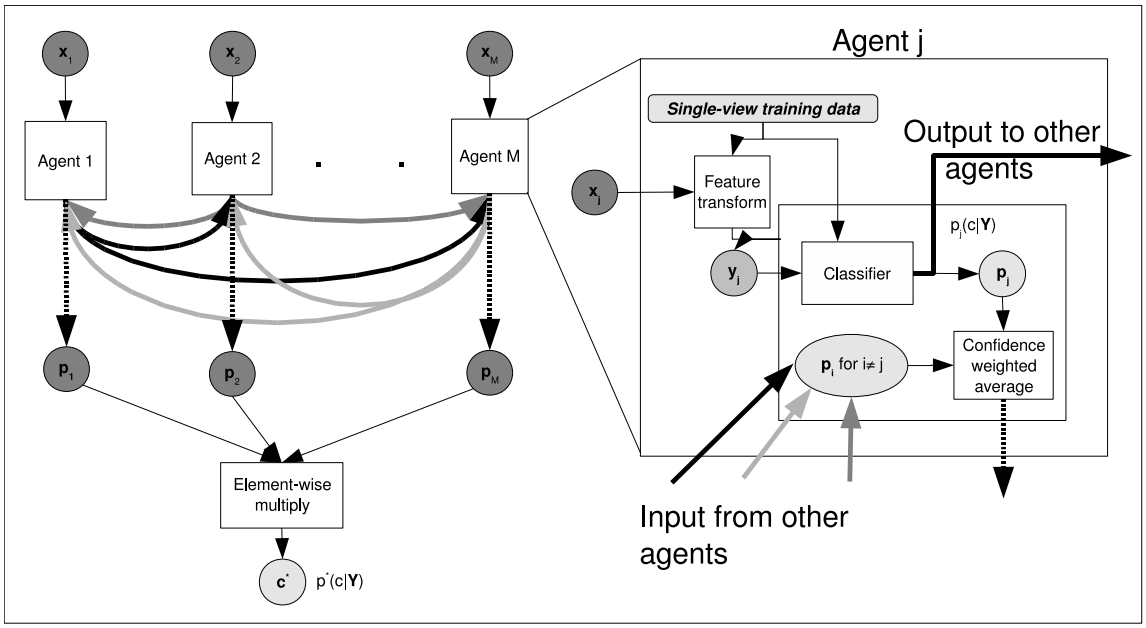


Figure 4.6: Collaborative fusion algorithm. Each view defines an "agent" which computes an initial posterior probability estimate using an SVM similar to algorithm A_2. These probability estimates are then sent from each agent to all the other agents. Upon receiving estimates from all other agents, each agent computes a final posterior estimate by combining their initial estimate and the estimates from all other agents using confidence weighting. The final outputs from each agent are multiplied together to yield a joint posterior probability for each class.

a set of probability vectors $(\boldsymbol{p}_1, ..., \boldsymbol{p}_{j-1}, \boldsymbol{p}_{j+1}, ..., \boldsymbol{p}_M)$ that can be used to adjust the initial prediction. Under the assumption of independence between the probability vectors, the joint probability for a given class is computed as

$$\mathbf{p}_{joint}(c = k|\mathbf{Y}_j) = \prod_{i \neq j}^{M} \mathbf{p}_i[k]. \tag{4.16}$$

Given the initial posterior probability $p_j(c|\mathbf{y}_j)$ and the joint probability $p_{joint}(c|\mathbf{Y}_j)$, the next step is to combine the two decisions together. There are many ways to do this, including using a second SVM, although that would destroy the ability of the algorithm to incorporate new views without retraining. The method used here is based on the observation that the classifiers typically give accurate results when the posterior probability of the class label is close to 1 for one of the classes and close to zero for the others. When the probability is nearly even between the classes, the class with slightly greater probability is frequently the wrong class

label. This is a consequence of the fact that examples which lie within the margin for most of the trained SVMs are given evenly distributed probabilities between classes by equation (4.12).

To quantify the degree to which the probability is spread between classes, the sidelobe ratio [145] is used. The ratio is defined as

$$C_{SL} = \frac{p(c_1|\mathbf{Y}) - p(c_2|\mathbf{Y})}{p(c_1|\mathbf{Y})}, \tag{4.17}$$

where

$$p(c_1|\mathbf{Y}) \geq p(c_2|\mathbf{Y}) \geq ... \geq p(c_C|\mathbf{Y}). \tag{4.18}$$

When most of the probability is given to c_1, C_{SL} is close to 1, when the probabilities are roughly equation between classes, C_{SL} is close to 0. To combine the individual predictions with the joint probabilities, the sidelobe ratio is computed for each and scaled to be unity. This gives the weights

$$\mathcal{W}_{init} = \frac{C_{SL}^{init}}{C_{SL}^{init} + C_{SL}^{joint}}, \tag{4.19}$$

and

$$\mathcal{W}_{joint} = \frac{C_{SL}^{joint}}{C_{SL}^{init} + C_{SL}^{joint}}. \tag{4.20}$$

The final prediction output from the j^{th} agent is then given as a vector of posterior probabilities for each class label

$$p_j^f(c|\mathbf{Y}) = p_j(c|\mathbf{y}_j)\mathcal{W}_{init} + p_{joint}(c|\mathbf{Y}_j)\mathcal{W}_{joint}. \tag{4.21}$$

Using these final predictions from each agent, the algorithm again computes the joint posterior as

$$p^*(c|\mathbf{Y}) = \prod_{j=1}^{M} p_j^f(c|\mathbf{Y}). \tag{4.22}$$

Several remarks are in order to explain the probabilistic computations in this algorithm. First, there is an implicit assumption in computing the joint distributions that the posterior estimates from each view are independent. This is not technically satisfied in our application, but is weakly satisfied for the SVM classifier, especially for examples that are difficult to classify correctly from a single view.

Second, because the predictions from the individual and the group are combined using weights derived from the sidelobe ratio there is no explicit bias given to the group prediction over the individual prediction.

4.5 Results and algorithm evaluation

Data sets listed in table 4.2 are divided up and regrouped together to test the application of the classification algorithms to several classification problems over varying degrees of difficulty. The definition of these data sets are given in table 4.3. The difficulty of the classification problem is qualitatively described in the table. Each data set is given a set of two ratings describing the amount of inter-class variability in fish length and the amount of intra-class variability in fish length. Data sets which consist of a single fish per class (D_2, D_3, D_4) have no intra-class variability and typically result in lower errors. Data sets D_1 and D_6 are particularly difficult because there are several individuals per class (for the same amount of training data) and there is significant intra-class variability in length.

Table 4.3: Data sets formed from combinations of the available scattering data. These data sets represent several different types of classification problems. Methods for sampling from available data within a given species are listed after the number of individuals per class. They are organized as follows: *long* selects the longest individuals from each class; *short* selects the shortest individuals from each class; *inter* selects individuals that have the maximum inter-class variability in length; *intra* selects the individuals that have the maximum intra-class variability in length. The final two columns outline the relative intra- and inter-class variability in length over the data set. As length is a dominant feature of the echoes, greater intra-class variability and lower inter-class variability are expected to yield higher error rates.

Set name	Species	Classes	# Per class	Samples	Intra-	Inter-
D_1	S_6, S_7, S_8, S_9	4	4, *all*	(240 / 60)	High	Med
D_2	S_6, S_7, S_8, S_9	4	1, *long*	(240 / 60)	None	Med
D_3	S_6, S_7, S_8, S_9	4	1, *short*	(240 / 60)	None	Med
D_4	S_6, S_7, S_8, S_9	4	1, *inter*	(240 / 60)	None	Med
D_5	S_6, S_7, S_8, S_9	4	2, *inter*	(240 / 60)	Med.	High
D_6	S_6, S_7, S_8, S_9	4	2, *intra*	(240 / 60)	High	High
D_7	S_1, S_2, S_3, S_4, S_5, S_6, S_7, S_8	4	4, *all*	(240 / 60)	Med.	High
D_8	S_1, S_2, S_3, S_4, S_5, S_6, S_7, S_8	8	2, *all*	(240 / 60)	Low	High
D_9	S_1, S_4, S_5, S_8	4	2, *inter*	(240 / 60)	Low	High
D_{10}	S_2, S_3, S_6, S_7	4	2, *inter*	(240 / 60)	Low	High

4.5.1 Classifier performance

The performance of the classification algorithms is evaluated as the average probability of error computed using 5-fold cross-validation. Standard errors of the cross-validation estimate are computed using the adjusted variance estimate [146, 147] of the form

$$\hat{\sigma}^2_{CV} = \left(\frac{1}{F} + \frac{p}{1-p}\right) \mathrm{Var}[x], \tag{4.23}$$

where $p = \frac{1}{F+1}$, F is the number of folds in the cross-validation, and $Var[x]$ is the sample variance. The variable p attempts to capture an overestimate for the correlation between training sets and therefore an estimate of the variance which is slightly higher, yielding a conservative measure of classifier performance.

Before evaluating the performance of the three algorithms defined in Section 4.4, the feature-fusion algorithm (Section 4.4.1) is evaluated on each data set using all eight available views, feature spaces defined by the DCT, Db4, and Haar wavelet coefficients, and the SVM, K-NN, and AdaBoost classification algorithms (Table 4.4). The results show a wide range of performance depending on the data set, classifier, and feature space. The SVM classifier performs as well or better than the K-NN or AdaBoost algorithms for all data sets and feature spaces. The AdaBoost algorithm generally performs better than the K-NN algorithm for most data sets except for the most challenging ones (D_1, D_5, and D_6). This is likely a consequence of the fact that the decision tree used as a weak learner for AdaBoost is not able to synthesize structural information in the features any better than the K-NN distance function, and therefore does not perform better when there is strong overlap between the classes.

The wavelet-based feature spaces generally outperform the DCT features for each data set. Of the wavelet feature spaces, the Haar wavelets typically outperform the Db4 wavelets which is somewhat unexpected as the Db4 wavelet typically gives better compression than the Haar wavelet for these kinds of echoes and has shown better classification performance for scattering from elastic shelled targets [64]. However, the performance improvement may be a result of the Haar

transform retaining additional discriminant information which is left out of the Db4 transform due to retaining only the coarsest approximation and detail coefficients.

The relative error between data sets is correlated with the number of individuals per class, the intra-class variability in length, and also the inter-class variability in length. These results are independent of classifier and feature space used. Given that the SVM shows better performance than either the K-NN or AdaBoost classifiers, it is used for the remainder of the results for the feature-fusion algorithm.

Table 4.4: Average probability of error using all eight views and the feature-fusion algorithm with different classifiers. Error is estimated using five-fold cross-validation with adjusted standard errors. Ten data sets (rows), three feature transformations, and three classification algorithms are compared under the feature-fusion (Section 4.4.1) algorithm.

Set	SVM			K-NN			ADABoost		
	DCT	Db4	Haar	DCT	Db4	Haar	DCT	Db4	Haar
D_1	50 ± 1.7	47 ± 1.5	43 ± 1.9	49 ± 0.6	53 ± 0.6	46 ± 0.7	51 ± 0.3	53 ± 0.6	53 ± 0.7
D_2	24 ± 0.5	13 ± 1.0	13 ± 1.2	33 ± 1.3	45 ± 1.4	34 ± 0.6	34 ± 0.8	34 ± 0.9	34 ± 0.5
D_3	16 ± 1.5	12 ± 1.0	10 ± 1.0	28 ± 1.2	37 ± 2.2	26 ± 1.7	28 ± 1.6	33 ± 0.6	31 ± 1.1
D_4	21 ± 1.0	12 ± 1.0	10 ± 1.0	32 ± 1.2	38 ± 1.8	29 ± 1.3	33 ± 1.1	35 ± 1.6	32 ± 1.0
D_5	35 ± 0.2	30 ± 2.7	26 ± 1.1	35 ± 0.9	41 ± 1.1	35 ± 1.1	44 ± 0.9	48 ± 0.5	46 ± 1.0
D_6	43 ± 1.4	36 ± 0.7	33 ± 3.2	46 ± 1.0	51 ± 1.1	44 ± 1.0	42 ± 0.5	43 ± 0.9	44 ± 1.0
D_7	27 ± 1.9	22 ± 1.9	20 ± 1.4	37 ± 1.0	40 ± 2.7	39 ± 2.2	28 ± 1.0	26 ± 0.8	28 ± 0.6
D_8	25 ± 0.8	22 ± 1.2	18 ± 0.9	33 ± 0.7	31 ± 1.0	30 ± 1.3	31 ± 0.8	34 ± 0.9	33 ± 0.5
D_9	16 ± 0.9	10 ± 1.0	9 ± 1.3	21 ± 2.5	22 ± 2.0	21 ± 2.3	15 ± 0.6	11 ± 1.0	11 ± 1.0
D_{10}	13 ± 1.7	15 ± 1.8	14 ± 1.0	24 ± 0.9	23 ± 1.2	22 ± 1.0	18 ± 1.0	15 ± 0.9	18 ± 0.7

4.5.2 Comparison of multi-view fusion algorithms

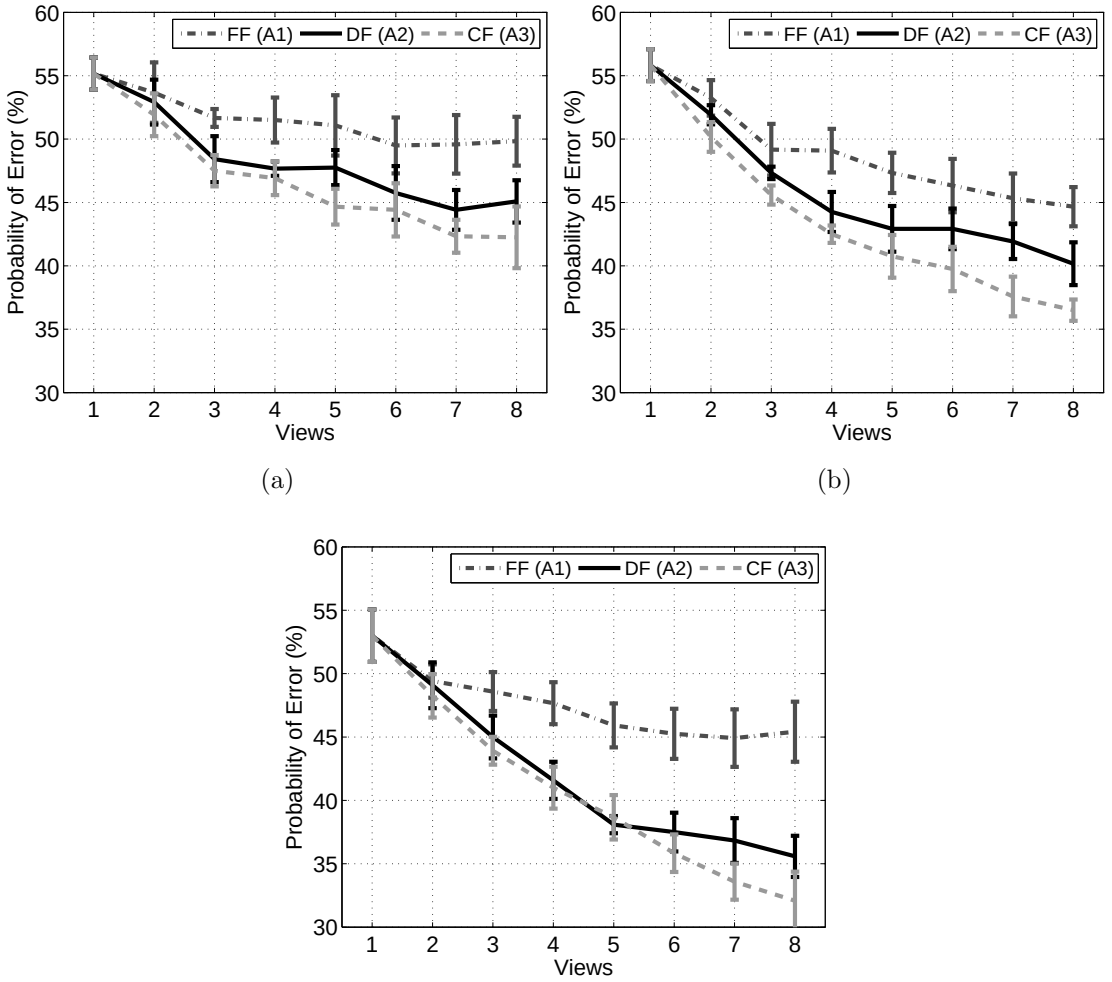
Using the SVM classifier for feature fusion, the three classification algorithms (Section 4.4) are compared as a function of feature space and the number of views used for classification on the D_1 (Figure 4.7) and D_8 (Figure 4.8) data sets. The collaborative-fusion algorithm (A_3) shows the best performance of the three algorithms with the decision-fusion algorithm giving comparable performance, and the feature-fusion algorithm performing significantly worse. Again, the Haar wavelet feature space results in lower error than either the DCT or the Db4 feature space.

A key feature of the decision- and collaborative-fusion algorithms is that the performance improvement over the feature fusion algorithm increases as more views are available. This result is likely a consequence of the fact that the feature space dimension grows as Md, where d is the dimension of the feature space, and M is the number of views. In contrast, the decision-fusion algorithm feature space grows as Cd where C is the number of classes, and the collaborative fusion algorithm feature space dimension is d independent of the number of views. For all algorithms and feature spaces, the average probability of error is typically monotonically decreasing as additional views are added. However, the collaborative fusion algorithm error curve generally decreases faster and more smoothly than the others. The roughness in the error curve is related to the effect of adding more views on increasing the feature space dimension and also increasing discrimination capabilities. When the feature space dimension is increased without adding discriminant information one might expect the error rate to stay the same, or possibly get worse. Because the collaborative-fusion algorithm computes weighted averages of initial probabilities from each agent, it results in a generally smoother curve with a reduction in error that is more rapid as additional views are added.

The error curves for the other data sets (D_2, D_3, D_4, D_5, D_6, D_7, D_9, D_{10}) show similar trends to D_1 and D_8 (Appendix C). These two data sets represent problems with high intra-class variability with four classes (D_1) and low intra-class variability with eight classes (D_8). The errors are much higher for D_1 than for D_8. The collaborative fusion algorithm almost always performs as well or bet-

ter than the decision-fusion algorithm, and the feature-fusion algorithm performs significantly worse in all cases. However, relative errors between the algorithms vary significantly as a function of the classification problem. Error curves for the decision- and collaborative-fusion algorithms are very similar for D_1. In contrast, they are significantly different for D_8 with the collaborative fusion algorithm showing significantly lower error. The shape of the error curve highlights differences in difficulty between data sets and the different behavior of the algorithms in response to these different data sets. For example, D_1 has four individuals per class, and contains the maximum amount of intra-class variability where as D_8 has much lower intra-class variability, but twice as many classes. The error curves for D_1 decrease much more gradually as views are added than for D_8 indicating that adding additional views is more beneficial in cases where there is less intra-class variability in length.

In order to compare the performance of the algorithms across all data sets and feature spaces, the average probability of error using all eight views is computed and presented in Table 4.5. The general performance trends are similar. In addition, it can be seen that the absolute error increases as more individuals are added to each class. This is likely a result of having fewer training examples per individual, but also a result of the addition of more individuals broadening the feature distributions and causing more overlap between classes. This effect is highlighted for D_1, D_5, and D_6 which show markedly higher errors than the other data sets.



(a) (b)

(c)

Figure 4.7: Probability of error on the D_1 vs. the number of views used for classification, all three classification algorithms, and the DCT (a), Db4 (b) and Haar (c) feature spaces.

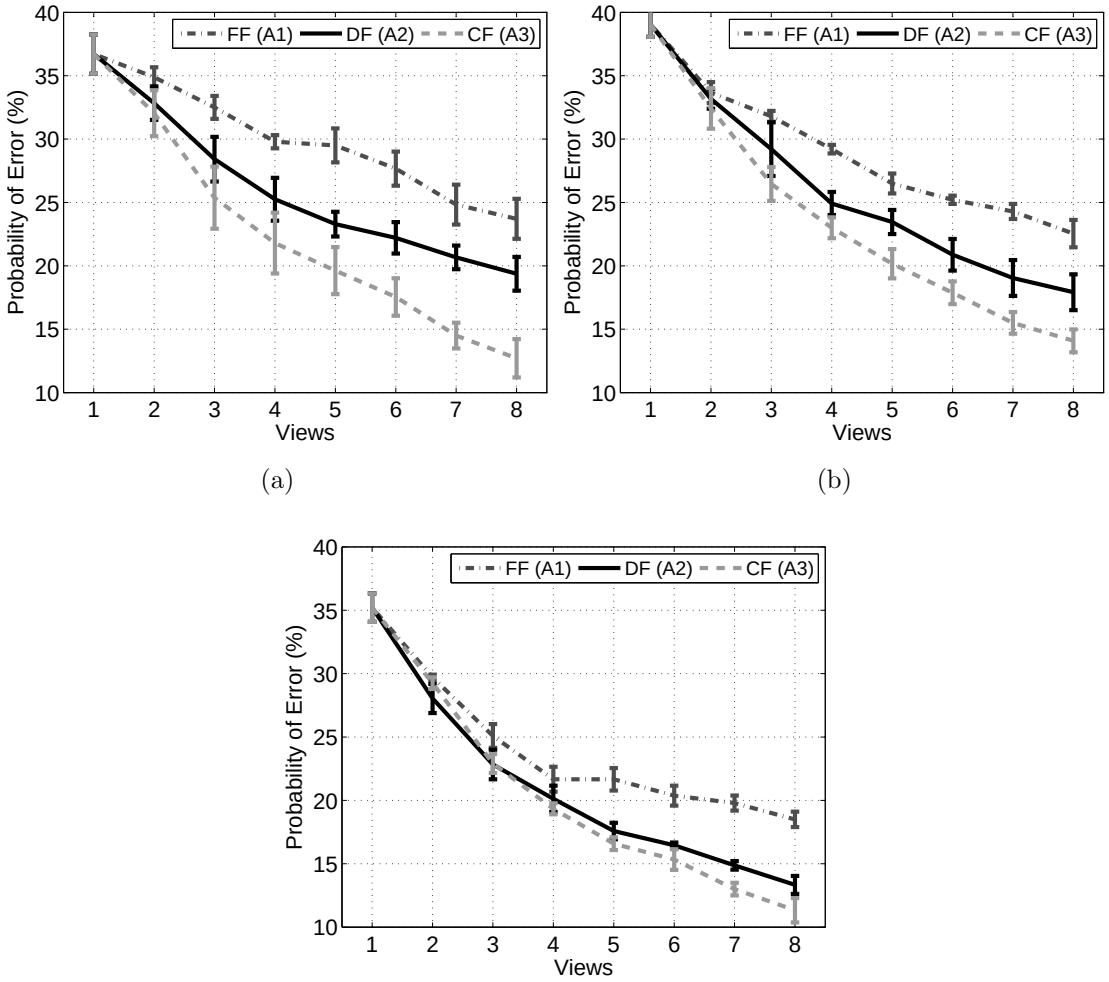


Figure 4.8: Probability of error on the D_8 vs. the number of views used for classification, all three classification algorithms, and the DCT (a), Db4 (b) and Haar (c) feature spaces.

Table 4.5: Average probability of error using all eight views and all three classification algorithms with the SVM classifier used for feature-fusion. Error is estimated using five-fold cross-validation with adjusted standard errors. Ten data sets (rows) and three feature transformations are compared.

Set	A_1, Feature fusion			A_2, Decision fusion			A_3, Collaborative fusion		
	DCT	Db4	Haar	DCT	Db4	Haar	DCT	Db4	Haar
D_1	50 ± 1.7	47 ± 1.5	43 ± 1.9	46 ± 2.0	42 ± 3.0	36 ± 3.0	44 ± 2.0	36 ± 2.0	33 ± 1.8
D_2	24 ± 0.5	13 ± 1.0	13 ± 1.2	18 ± 2.0	10 ± 1.0	9 ± 1.0	17 ± 2.0	10 ± 1.0	9 ± 1.5
D_3	16 ± 1.5	12 ± 1.0	10 ± 1.0	13 ± 1.0	9 ± 1.0	8 ± 2.0	12 ± 1.0	9 ± 1.0	7 ± 1.41
D_4	21 ± 1.0	12 ± 1.0	10 ± 1.0	15 ± 1.3	7 ± 0.5	6 ± 1.3	14 ± 1.3	7 ± 1.0	5 ± 1.5
D_5	35 ± 0.2	30 ± 2.7	26 ± 1.1	31 ± 3.0	26 ± 2.4	22 ± 1.0	28 ± 3.0	19 ± 1.3	17 ± 2.0
D_6	43 ± 1.4	36 ± 0.7	33 ± 3.2	35 ± 1.4	27 ± 2.3	24 ± 2.6	33 ± 1.4	26 ± 1.8	22 ± 3.65
D_7	27 ± 1.9	22 ± 1.9	20 ± 1.4	19 ± 1.7	19 ± 1.4	18 ± 1.0	17 ± 1.7	19 ± 1.4	18 ± 1.0
D_8	25 ± 0.8	22 ± 1.2	18 ± 0.9	20 ± 0.7	18 ± 1.2	13 ± 1.1	14 ± 0.6	15 ± 0.5	13 ± 0.9
D_9	16 ± 0.9	10 ± 1.0	9 ± 1.3	9 ± 1.2	9 ± 1.5	5 ± 1.7	9 ± 1.2	7 ± 0.8	4 ± 0.5
D_{10}	13 ± 1.7	15 ± 1.8	14 ± 1.0	12 ± 1.4	15 ± 1.8	11 ± 1.2	8 ± 1.5	10 ± 1.4	9 ± 1.2

4.5.3 Classification performance using random views

Thus far, the classification error has been evaluated under the assumption that the view geometry is fixed, and the angular spacing between views is constant and highly limited. This model fits the general paradigm of existing underwater target classification methods, and also of the work previously performed on zooplankton where it was assumed that the views were collected simultaneously from a receiver array with limited aperture [59, 60].

In some applications (echo counting sonar for example), many observations of individual fish may be collected where each individual has an unknown orientation relative to the sonar. In this application, the multi-view approach could be applied by treating the echo from each individual fish as a realization of multi-view scattering from a single fish class. Under this model, each echo is assumed to provide a uniformly random sample from the orientation distribution of the fish from a given class. That is, eight views are still collected, but the aperture is effectively no longer limited, and the geometry is now random and changes for every sample. Specifically, let k denote the k^{th} sample in the data set, and let $\mathbf{y}_j^k$ denote the feature vector computed from echo data on the j^{th} receiver for k^{Th} sample. The random set of views is then defined by

$$\mathbf{Y} = (\mathbf{y}_1^{k_1}, \mathbf{y}_2^{k_2}, ..., \mathbf{y}_M^{k_M}), \tag{4.24}$$

where k_i for $i = 1, ..., M$ are sampled uniformly randomly and independently from the set of integers $1, ..., S$ with replacement and S is the total number of examples in the data set.

Under this model, one would expect an improvement in performance due to the increase in effective aperture of the samples. However, this comes at the cost of uncertainty in the geometry. This uncertainty is likely to cause the improvement obtained from feature fusion to be significantly less than that obtained from either decision or collaborative fusion.

The probability of error vs. the number of views is again plotted for the D_1, and D_8 (Figures 4.9 and 4.10). The error curves for the other data sets (D_2, D_3, D_4, D_5, D_6, D_7, D_9, D_{10}), computed using random views, show similar trends to D_1

and D_8 (Appendix D). There is generally a significant reduction in absolute error from the fixed and limited aperture case. However, the feature-fusion algorithm sometimes shows an increase in error over the fixed view case for the DCT features. For the Db4 and Haar features, the feature-fusion algorithm shows a reduction in absolute error over the fixed view case, but the reduction is relatively small compared to both the decision- and collaborative-fusion algorithms which show very dramatic reductions for all feature spaces and data sets. This shows a key advantage of the decision- and collaborative-fusion algorithms. Because neither algorithm explicitly depends on the view geometry during training, they can fully utilize the increased effective aperture even when the views are collected randomly.

The performance of the algorithms and feature spaces when all eight views are available (Table 4.6) shows that using random views significantly reduces the classification error for all data sets and algorithms. The reduction is dramatic for the decision- and collaborative-fusion algorithms, and for the data sets which only include one individual per class, the error is very small, sometimes less than 1%. Comparing the error curves (Figures 4.9 and 4.10) for the D_8 and D_1 data sets, it can be seen that the error curves are close to converging to a very low error in the case of the D_8. We conjecture that the error curves for the D_1 data set would likely decrease much further by adding more views.

These results support the idea that systems which collect uniformly random views would likely perform much better than systems which collect a limited set of views. The algorithms not based on feature-fusion are significantly better at taking advantage of the random views as they don't make any explicit assumptions about the view geometry during training.

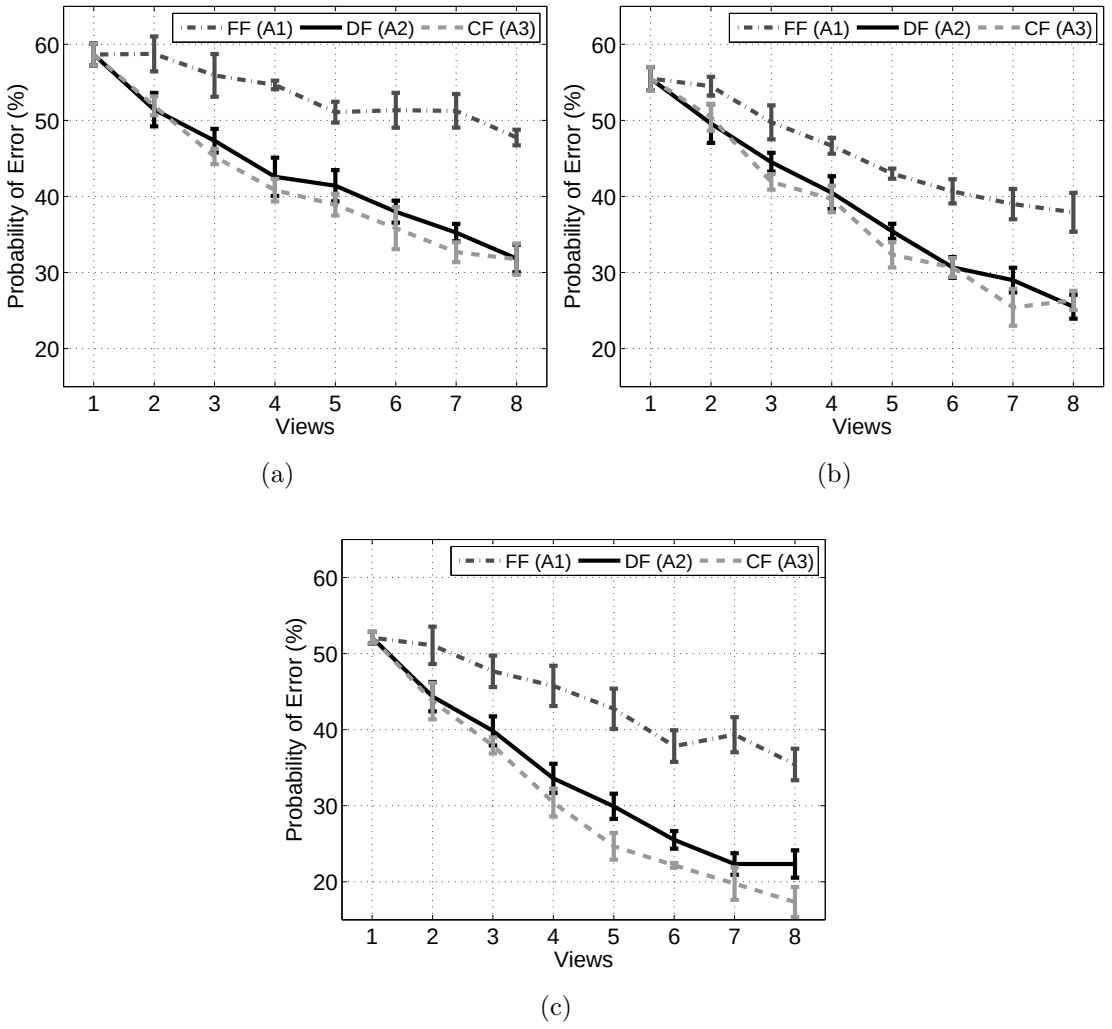
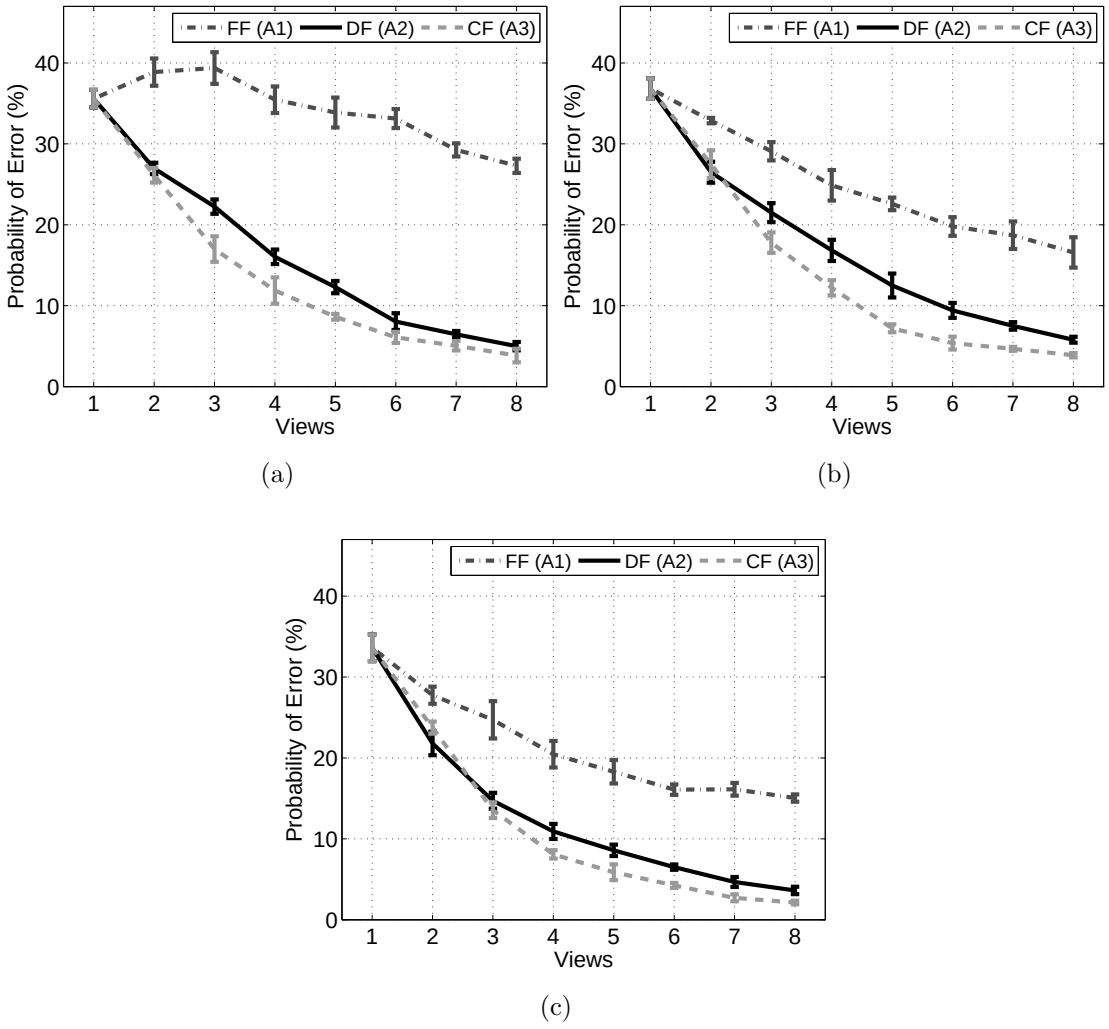


Figure 4.9: Probability of error on the D_1 vs. the number of views used for classification, all three classification algorithms, and the DCT (a), Db4 (b) and Haar (c) feature spaces.



(a) (b)

(c)

Figure 4.10: Probability of error on the D_8 vs. the number of views used for classification, all three classification algorithms, and the DCT (a), Db4 (b) and Haar (c) feature spaces.

Table 4.6: Average probability of error using all eight views and all three classification algorithms with the SVM classifier used as the classifier. The views are not constrained by a fixed geometry and are randomly sampled from available data. Error is estimated using five-fold cross-validation with adjusted standard errors. Ten data sets (rows) and three feature transformations are compared.

Set	A_1, Feature fusion			A_2, Decision fusion			A_3, Collaborative fusion		
D_1	49 ± 1.7	39 ± 1.6	36 ± 0.7	32 ± 1.6	26 ± 4.1	21 ± 2.1	32 ± 1.6	26 ± 1.4	18 ± 1.3
D_2	24 ± 0.5	10 ± 1.0	9 ± 1.2	5 ± 2.0	2 ± 1.0	1 ± 1.0	6 ± 2.0	2 ± 1.0	1 ± 1.5
D_3	24 ± 4.2	12 ± 2.2	11 ± 1.8	2 ± 0.5	1 ± 0.2	1 ± 0.2	3 ± 1.0	1 ± 0.2	1 ± 0.4
D_4	25 ± 3.0	8 ± 0.6	6 ± 0.8	4 ± 0.5	1 ± 0.7	1 ± 0.2	4 ± 0.9	1 ± 0.4	1 ± 0.3
D_5	39 ± 1.2	22 ± 2.6	18 ± 2.7	16 ± 1.2	9 ± 0.6	7 ± 1.4	16 ± 1.8	7 ± 0.6	5 ± 1.5
D_6	46 ± 1.9	30 ± 3.2	30 ± 2.7	22 ± 1.2	13 ± 2.1	10 ± 1.7	23 ± 2.2	12 ± 1.2	10 ± 1.8
D_7	26 ± 2.0	11 ± 2.7	9 ± 1.5	7 ± 1.4	7 ± 1.2	3 ± 1.2	5 ± 0.7	4 ± 1.0	3 ± 0.9
D_8	28 ± 1.0	17 ± 1.3	16 ± 0.7	4 ± 0.9	6 ± 1.0	4 ± 0.9	3 ± 0.7	3 ± 0.6	3 ± 0.4
D_9	11 ± 1.0	4 ± 1.0	4 ± 1.2	1 ± 0.5	1 ± 0.3	1 ± 0.3	1 ± 0.4	1 ± 0.2	1 ± 0.2
D_{10}	18 ± 2.7	12 ± 1.5	12 ± 1.0	2 ± 0.5	3 ± 0.7	2 ± 0.8	1 ± 0.5	2 ± 0.3	1 ± 0.7

4.5.4 Classification performance as a function of problem size

Although the collaborative fusion algorithm appears to yield better performance for many of the data sets and feature transformations, an additional important comparison between algorithms is the rate at which error increases as more classes are added to the problem.

To investigate this measure of performance, the D_8 data set is used, and seven sub-problems are defined by including more classes from the D_8 data set starting with a binary problem and increasing up to the full eight-class problem. The algorithms and feature transforms are applied in the same manner as above and probability of error is estimated using five-fold cross-validation.

It can be seen that the collaborative-fusion algorithm out performs both the feature-fusion and decision-fusion algorithms dramatically in all three feature spaces (Figure 4.11). Interestingly, in the DCT feature space, the collaborative fusion curve begins to flatten out after 4 classes are added and the error remains just over 10 %. This behavior is not seen in any of the other algorithms or feature spaces. There is a sharp cusp in the performance curve at 4 classes. This is due to the order in which the classes are added, where there first three classes are easily separated by all methods. The fourth class is quite similar to the third class, and as a result the error increases significantly after including that class in the problem.

4.6 Conclusion

In this work, a multi-view approach to marine fish classification has been developed using laboratory data from live animals. The approach is based on fusing features or classification results from individual views together to dramatically improve the accuracy of the classifiers. This improvement was typically on the order of 50 % or higher, and in some cases could be as high as 90 %.

To give a comprehensive evaluation of the fish classification problem, three different feature spaces, three different classifiers, and three algorithms were de-

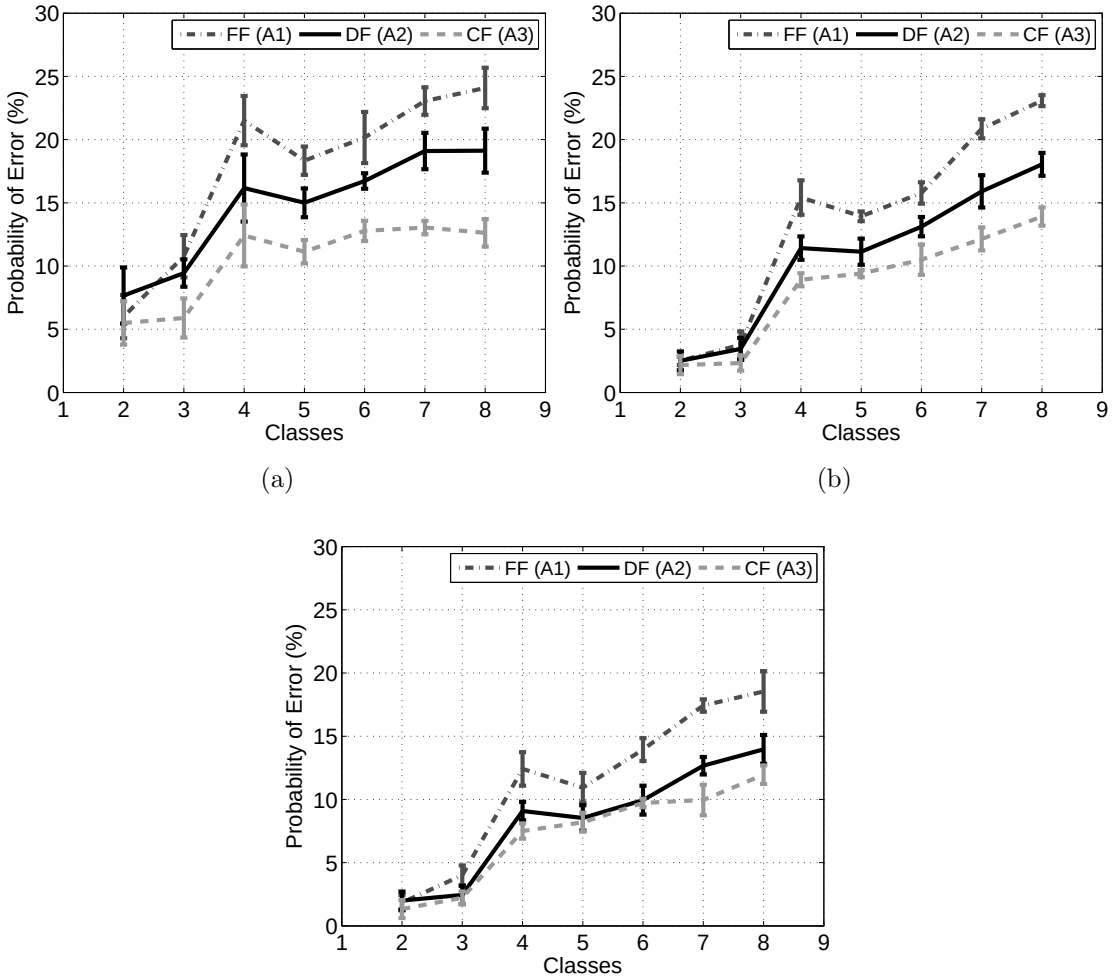


Figure 4.11: Classification error as a function of the the number of classes in the problem for the D_8 data set using DCT (a), Db4 (b), and Haar (c) wavelets. A general trend of increasing error as more classes are added to the problem is seen for all algorithms and feature spaces.

veloped and compared with one another. The features were all based on transformation of the echo envelope recorded from each view of the fish. The classification algorithms represented extensions of modern methods that have been used in underwater target classification to multi-class problems using support vector machines. A collaborative-fusion algorithm based on multi-aspect object classification [120] was developed that yielded the best performance with the least number of assumptions about view geometry and number of available views. This algorithm has the significant advantage that it can be applied to problems with unknown and variable view geometries, and can incorporate additional views without requiring retraining.

While the results of the application are promising, several important areas for future research exist. First, this work dealt only with fusion algorithms which do not explicit model the joint probability of the observation sequence (for example hidden Markov models). Given the success of the collaborative fusion algorithm, it is likely that an algorithm which combines the decision feedback with probabilistic modeling of the observation sequence would yield good results. In this regard, a comparison between the collaborative-fusion and HMMs would be useful. In addition, a potential improvement to the collaborative-fusion algorithm would be to assign an additional weight which incorporates the degree of confidence given to the group rather than the individual. However, the advantage of the unbiased weighting is that it allows strong predictions from one agent to overpower weak predictions from the group. A second issue is one of generalization fish populations with significant intra-species variability in length. Given the sensitivity of the algorithms to scatterer length variability, training the algorithm with only a few individual fish may not be sufficient to yield adequate performance when applied to diverse populations. A natural question that arises when performing laboratory experiments is the similarity to what is observed in the lab to what would be observed in the field. In these experiments, the animals movements were constrained as well as them being anesthetized. An important future direction is to deploy a similar system in a less-artificial setting such as an aquarium where measurements could be recorded from many more individual fish under conditions

which are potentially less stressful for the fish.

Currently, there is a need in fisheries ecology for more specificity in acoustic surveys. In this study, it has been shown that multi-view methods may offer considerable advantages in improving specificity of acoustic systems which could translate to increased classification accuracy over currently used single-view methods.

4.7 Acknowledgments

The authors would like to thank the SIO machine shop for assistance with construction of the scattering apparatus, Eddie Kisfaludy, Erdem Karakoylu, Fernando Simonet, Ben Maurer, and Robert Glatts for technical consulting on the scattering apparatus design and help with experiments, and the California Sea Grant for funding this research.

The text of Chapter 4 is in part and under some rearrangements a reprint of the material as it may appear in Paul L. D. Roberts, Jules S. Jaffe, and Mohan M. Trivedi, "A multi-view, broadband approach to marine fish classification," in prep. *IEEE J. Oceanic Eng.* June 2009. The dissertation author was the primary researcher and author, and the co-authors listed in this publication directed and supervised the research which forms the basis for this chapter.

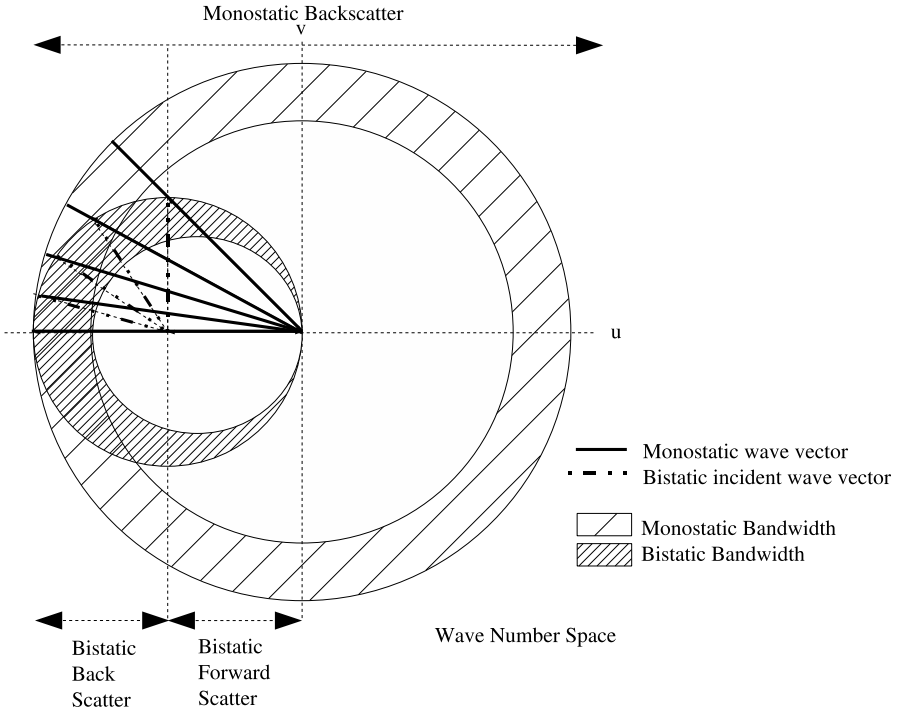
Chapter 5

Study of the effect of bandwidth and aperture on classification performance

5.1 Introduction

The performance of the multi-view classification approach has been studied assuming a realistic, yet fixed set of system parameters as the bandwidth of the transmit signal, and the aperture and number of views of the receivers has been fixed or assumed to be uniformly random. In this chapter, the performance trade-offs in changing these three parameters are explored through simulation and analysis of laboratory data.

The relationship between bandwidth, aperture, and system performance is an important topic which has been studied in many areas related to weak scattering [148, 149, 150, 151]. Due to the Fourier transform relationship between the scattered field and the scatterer [90], the bandwidth and aperture required to recover specific attributes can be defined. Figure 5.1 gives an example of the relationship between scattering measurement geometry and spatial frequencies of the scatterer that are sampled under the Born approximation. It can be seen from the figure that increasing transmitter bandwidth yields spatial frequencies along a sin-



(a)

Figure 5.1: Example of the different spatial frequencies sampled under monostatic and bistatic geometries as a function of bandwidth and aperture for weak scattering under the Born approximation. It can be seen that the coverage of wavenumber space is directly related to bandwidth in the case of monostatic scattering, but not for bistatic scattering. In the case of monostatic scattering, low spatial frequencies are only sampled when low transmit signal frequencies are used. In contrast, forward bistatic scattering samples the DC spatial frequency for any transmit signal frequency.

gle direction in the wavenumber space where as increasing aperture (moving along the perimeter of a circle in wavenumber space) yields directional spatial frequencies. Clearly, combining full aperture and bandwidth yield complete (when both magnitude and phase are recorded) data that uniquely determine the scatterer.

The influence of bandwidth on underwater target classification was evaluated by Yao et. al 2002 [152]. In this study, classification performance using data with 40 kHz and 80 kHz of bandwidth was evaluated and it was shown that the 80 kHz data yielded significant improvements in accuracy over the 40 kHz data. This study however did not investigate the role of aperture as it was limited to

single-view classification.

To explore how bandwidth *and* aperture influence classification accuracy, a pragmatic approach was taken. Using simulated and laboratory data, classification algorithms were trained and tested for different transmit signal bandwidths and receiver array apertures. The influence of these parameters was quantified by computing the average error of the classification algorithm on test data for each combination of bandwidth and aperture.

Bandwidth $\mathcal{B}$ defines the difference between the maximum and minimum frequencies in the transmit signal for which the intensity $S(f)$ is within 3 dB of the peak intensity,

$$\mathcal{B} = f_{max} - f_{min}, \tag{5.1}$$

where $S(fmin) = 0.5S(f_c)$, $S(fmax) = 0.5S(f_c)$, and f_c is the center frequency of the signal. The system is assumed to collect a set of $\mathcal{V}$ views over a given aperture. The aperture $\mathcal{A}$ is then defined by the range of angles over which scattering is measured,

$$\mathcal{A} = \theta_v - \theta_1, \tag{5.2}$$

where it is assumed that the angles are limited to the range [0^o,90^o] and increase from θ_1 to $\theta_{\mathcal{V}}$. Finally, the angular sample frequency F_a is defined to be

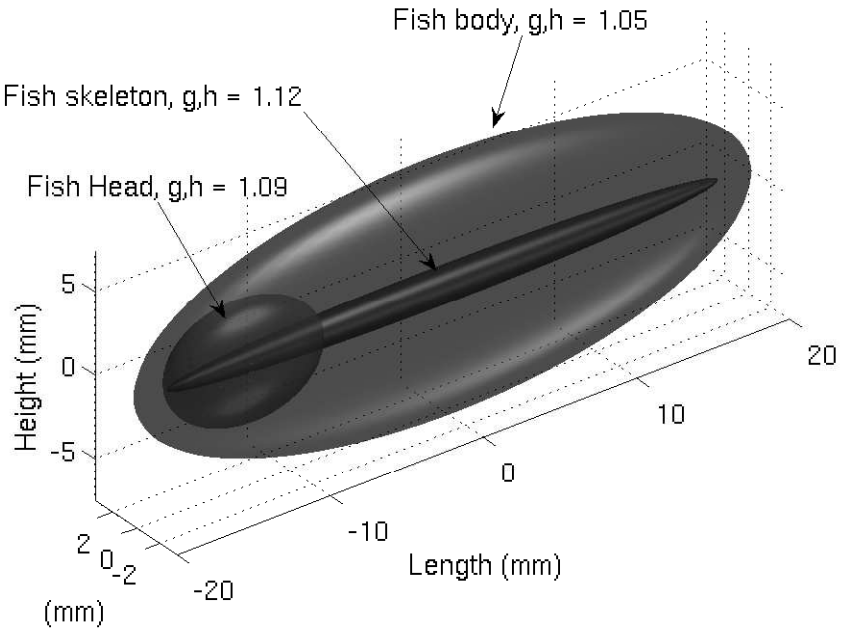
$$F_a = \frac{\mathcal{V}}{\mathcal{A}}. \tag{5.3}$$

5.2 Scattering model and echo simulations

The model used for simulating echoes is a hybrid version of the ellipsoid-DWBA (Appendix A), which allows multiple ellipsoids to be modeled. The model also takes into account the energy loss due to reflection at each ellipsoid interface. In general, the model can be defined as

$$fbs_H(\mathrm{k}, \theta, \mathbf{\Xi}) = \sum_{n=1}^{N} fbs(\mathrm{k}, \theta, \mathbf{\Gamma}_n) e^{i2\mathrm{k}\mathbf{e}^T \mathbf{x}_n}, \tag{5.4}$$

where $\mathbf{\Xi}$ is a set of model parameters $(\mathbf{\Gamma}_1, ..., \mathbf{\Gamma}_N)$, θ is the angle of the incident wave vector relative to the head of the scatterer, and $\mathbf{e}$ is the direction of the



(a)

Figure 5.2: Visualization of the hybrid DWBA model using three ellipsoids. Each ellipsoid is given a separate g and h value. The values are chosen to model a weaker scattering body and stronger scattering head and spine.

incident wave vector, and $\mathbf{x}_n$ is the position of the n^{th} ellipsoid. The function $fbs(x)$ can be any backscattering model.

For the sake of simplicity, three ellipsoids that represent the head, body, and spine of the fish were used for simulations. An example of the model is given in Figure 5.2. By incorporating additional ellipsoids (over the traditional DWBA) the more complex model is able to capture some of the complex features observed in measured scattering patterns (Figure 5.3). This figure shows a comparison between measured scattering data (Figure 5.3 (a)) and a matching model output (Figure 5.3 (b)). The model is able to capture the complex banding structure seen in the measured data quite well throughout a wide range of θ values. However, the exact position of bands in the data and the model is inconsistent. This is indicative of a mismatch in sound speed between the actual scatterer and the model. Despite the strong qualitative similarity between measured data and model output, it should be noted that assessing quantitative agreement between the model and data remains

Table 5.1: Model parameters used for each type of scatterer in the simulation. Units of a, b, c, and x are millimeters.

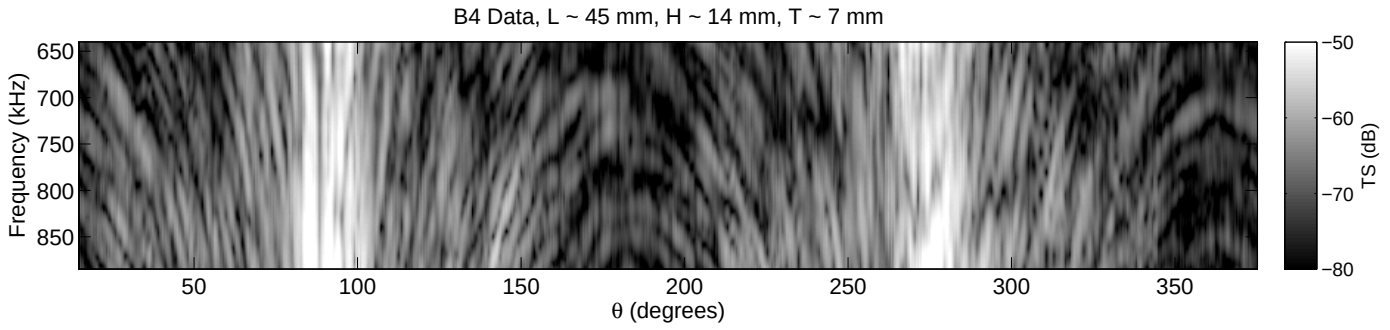
Scatterer	a	b	c	x	g,h
S_1-Body	2	0.75	0.35	0	1.05
S_1-Head	0.5	0.35	0.22	-1.3	1.09
S_1-Spine	1.8	0.10	0.10	0	1.12
S_2-Body	1.5	0.56	0.26	0	1.05
S_2-Head	0.37	0.26	0.16	-0.97	1.09
S_2-Spine	1.35	0.07	0.07	0	1.12
S_3-Body	2	0.1	0.55	0	1.05
S_3-Head	1	0.45	0.32	-1.3	1.09
S_3-Spine	1.8	0.20	0.20	0	1.12

an important area of future research.

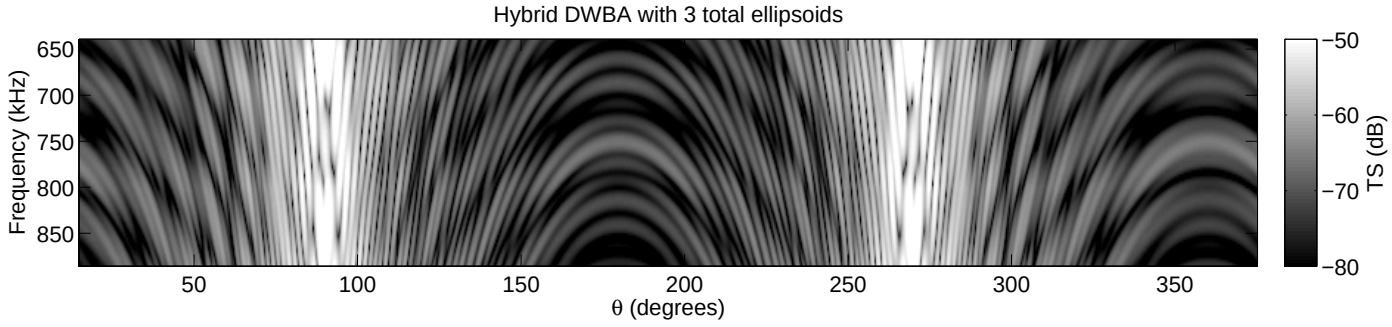
For the simulations presented here, a three-ellipsoid version of the model is used, with only the size and shape being varied. The values for all parameters and each scatterer are given in Table 5.1. Scatterer S_1 is the base model. Scatterer S_2 has the same shape as S_1, but is 25% smaller. Scatterer S_3 has the same body length ($a = 2$ mm) as S_1, but a different shape (wider and shorter). These three scatterer types are used to compare the effect of bandwidth and aperture when discriminating between scatterers of different size (S_1 and S_2) and different shape (S_1 and S_3).

5.3 Scattering simulations

The classification algorithms use the echo envelope to compute features. As such, the differences in echo structure provide the basis for discrimination. Before comparing the effect of bandwidth and aperture on classification performance, the effect of these parameters on the structure of the echo is explored.



(a)



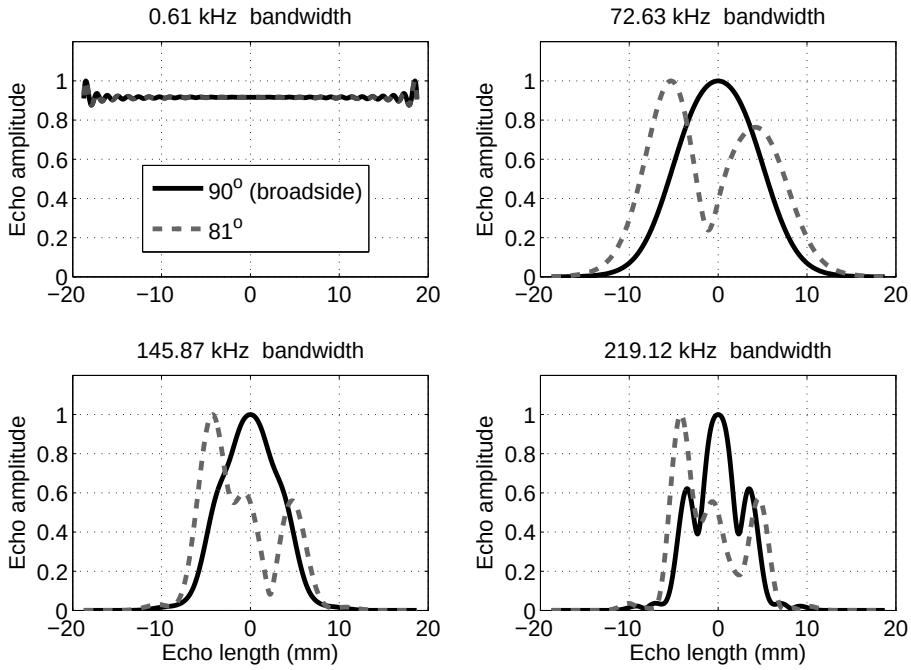
(b)

Figure 5.3: Comparison between measured scattering data from a small Damselfish (Chrysiptera cyanea) (a), and a matching model output from the hybrid DWBA (b). The model output is able to capture many of the complex banding structures in the measured data. Deviations between the measured data and the model are likely due to a mismatch in assumed sound speed in the body of the fish.

To investigate the effect of bandwidth and aperture on echo structure, two separate comparisons are made. First, two views are compared as a function of four different bandwidths (Figure 5.4). Second, two bandwidths are compared as a function of four different views (Figure 5.5)

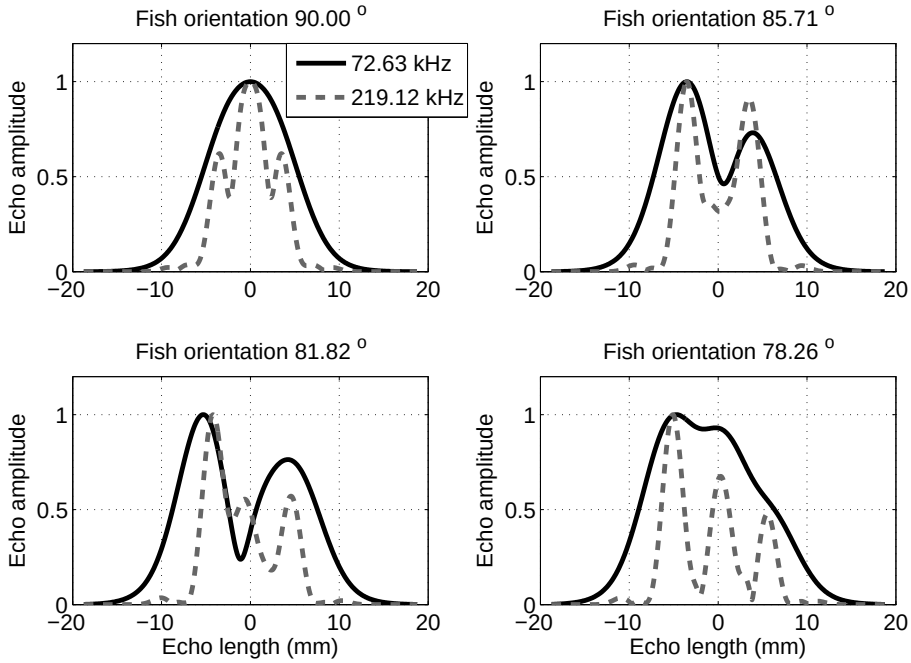
Figure 5.4 shows a direct relationship between bandwidth and resolution of peaks in the echo. For each view, increasing bandwidth yields more detail in the structure of the echo. The primary consequence of the this is that resolving differences between the two views critically depends on bandwidth. When there is very little bandwidth (Figure 5.4 (a)) the normalized echo output is identical between the two views. As the bandwidth increases (Figure 5.4 (b-d)) the two echoes show significant differences. The local maxima in the echoes are due to interferences in the body of the scatterer. Viewed near-broadside with 250 kHz of bandwidth, three peaks can be resolved in the echo, two peaks are due to the front and back wall of the scatterer body, and the third is due to the combination of the head and skeleton ellipsoids. In Figure 5.4 (b-d) a broadening in the the two outer peaks is noticeable in the 81^o view. This is due to the projection of the ellipsoid of the body becoming longer as the view moves away from broadside. Detection of the difference in spacing between the two peaks in Figure 5.4 (b) for each view would be sufficient (in the absence of noise) to estimate the length of the body. However, note that bandwidths greater than 100 kHz (Figure 5.4 (c-d)) are required to resolve the internal structure. These results indicate that, in a scenario where the body length alone was insufficient to distinguish between the classes, increasing the bandwidth of the system would be a possible strategy to improve performance. The performance gain would be dependent of the degree of similarity between classes at finer scales.

Figure 5.5 shows a complex relationship between scatterer orientation, bandwidth, and the resolution of structures in the body of the scatterer. For both 72 and 220 kHz bandwidth, the resolution of the peaks in the echo is *orientation dependent*. The number of resolvable peaks changes between orientations, and in certain cases (Figure 5.5 (d), 72 kHz) the peaks are masked together and are not resolvable. However, while the internal structures may or may not be resolved



(a)

Figure 5.4: Comparison between two views as a function of four different bandwidths.



(a)

Figure 5.5: Comparison between two bandwidths as a function of four different views.

for a given orientation, the duration of the echo increases monotonically and the scatterer orientation moves away from broadside. This result is the key feature which allows fish length and orientation estimation methods ([58],[153]) to work well even in the absence of knowledge about the internal structure of the scatterer.

5.4 Classification error for different parameter combinations

5.4.1 Error analysis on simulated data

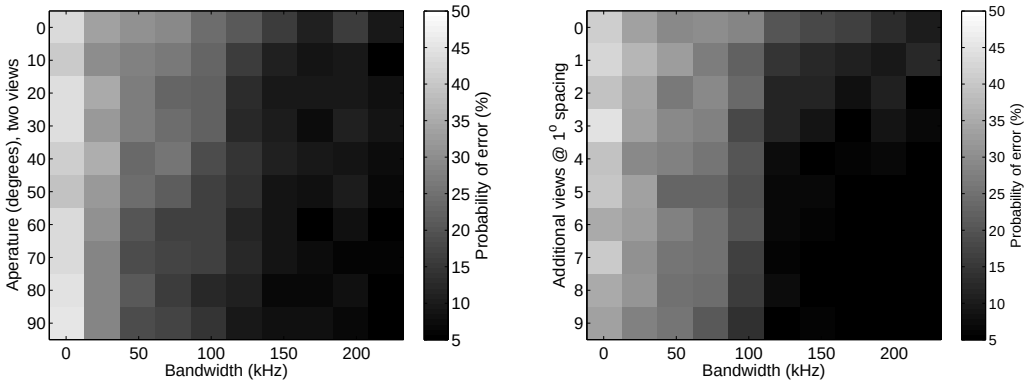
Using the decision-fusion algorithm defined in Section 4.4.2, the hybrid model is used to generate several realizations of scattering from scatterers S_1, S_2, and S_3. White Gaussian noise was added to the model output at a level of -76 dB. A Monte Carlo simulation is performed while varying the bandwith of the

transmit signal from 0 kHz to 250 kHz and also the aperture of the views. As an initial evaluation of the relationship between aperture and bandwidth, the average probability of error for discriminating between S_1 and S_2 was computed while varying the aperture from 0 to 90^o and the bandwidth from 0 to 250 kHz (Figure 5.6 (a)). A systematic relationship between the probability of error and both aperture and bandwidth can be seen. For a given aperture, increasing bandwidth (moving along a row in the figure) significantly reduces error from a high 45% down to a low 5%. This trend holds true for the majority of the apertures. However, it can be seen that for larger aperture, less bandwidth is required to yield a given error reduction, and for the case of very little aperture (top row), the error remains very higher up to full bandwidth. Interestingly, for a given bandwidth, increasing aperture only offers a small reduction in error that can be obtained with even a modest increase in aperture.

Given that bandwidth appears to be the most important factor when only two views of variable aperture are considered, the impact of increasing the number of views (from 1 to 9) is evaluated (Figure 5.6 (b)). Interestingly, it can be clearly seen that increasing the number of views (even though the aperture remains relatively small) yields a significant reduction in error, but only when the bandwidth is greater than 100 kHz. This is likely a result of the fact that for a very limited aperture, difference between views only become apparent when there is enough bandwidth to resolve the differences in length and thickness between S_1 and S_2. This result shows an important connection between aperture and bandwidth.

5.4.2 Error analysis on laboratory data

Using the laboratory scattering data from three different fish, the effect of the number of views and bandwidth are evaluated in a manner similar to Figure 5.6 (b), but with one fewer views owing to the limitations of the experimental system. In this case, the problem is broken down into a size-based classification problem (Figure 5.7 (a)) and a species-based classification problem (Figure 5.7 (b)). In the case of size-based classification, similar results to those of the simulation study (which also considered size-based classification) are obtained. However, certain



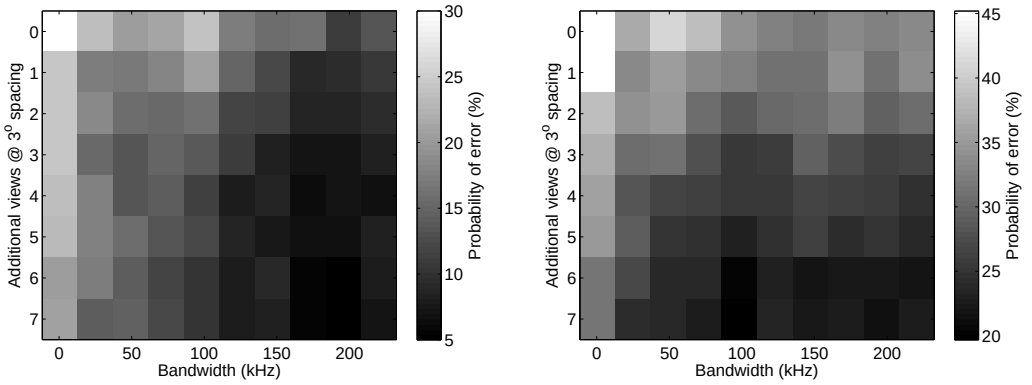
(a) (b)

Figure 5.6: Average probability of error as a function of aperture and bandwidth for a two-class size-based classification problem with simulated data. (a) Two views with increasing aperture (rows) and bandwidth (columns). (b) Increasing number of views and aperture (rows) and bandwidth (columns). Each pixel is the average probability of error over 5 random trails of training and testing with 250 training examples per class and 250 testing examples per class.

bandwidths (175 and 200 kHz) yield much lower error than others when all eight views are used. In addition, the reduction in error obtained by adding more views is higher than it was in simulation. This is likely a consequence of the fact that the experimental setup had a 3^o angular spacing where as the simulation used only a 1^o angular spacing. Despite these differences, the size-based classification yield very similar conclusions in both simulation an experiment.

The species-based classification (Figure 5.7 (b)) shows some very interesting features. In this case, increasing bandwidth does not help as much as increasing the number of views. This is likely due to the two classes being similar in size and shape. Therefore, increasing the resolution of the system by adding bandwidth does not yield any new discriminant features. In contrast, adding more views helps because each view provides some additional discriminant information related to angular changes in scattering due to differences in shape.

The previous results demonstrate that increasing the number of views makes a significant improvement in classification performance which can dominate improvements obtained by increasing bandwidth in the case of species-based classifi-



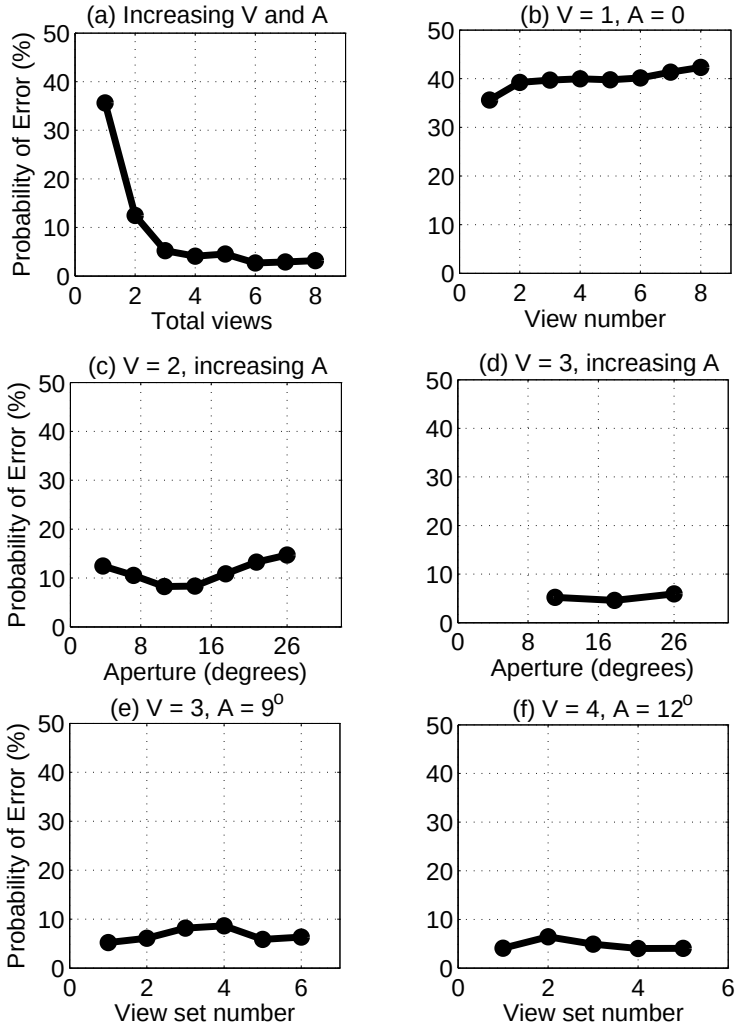
(a) (b)

Figure 5.7: Average probability of error as a function of the number of views and bandwidth for a two-class classification problem with simulated data. (a) Size-based classification problem. (b) Species-based classification problem. Each pixel is the average probability of error over 5 random trails of training and testing with 250 training examples per class and 250 testing examples per class.

cation. In order to evaluate the impact of choosing different views, and changing the aperture and number of views independently, the size-based classification problem with laboratory data is evaluated again using different combinations of views and aperture (Figure 5.8).

5.5 Conclusions

In this chapter, the effects of available bandwidth and aperture on classification performance have been investigated. Results indicate that sufficient bandwidth is critical for yielding good classification accuracy when the classes are of different sizes. The bandwidth required to achieve good accuracy is found to be greater than 100 kHz for the 2-5 cm length fish considered here. When classes are of different shape and similar size, the number of views plays a more significant role with performance increasing as more views are added. In practice, the number of available views will likely be dictated by experimental limits. However, for systems geared towards species classification, an emphasis on larger aperture and high view count would likely be more beneficial than larger bandwidth.



(a)

Figure 5.8: Probability of error as a function of different numbers of views and different amounts of aperture. For each plot, A denotes the total aperture of the views and V denotes the number of views used. It can be seen that in this case, the number of views makes the most dramatic difference in performance. In addition, increasing aperture can reduce error further.

Chapter 6

Multi-view acoustic sizing and classification of individual fish

6.1 Introduction

The advantages of acoustic methods for studying fish populations are rapid and non-invasive assessment that can be cost effective, operate over long ranges, and operate in complete darkness. In pelagic fisheries applications, acoustics has become the standard means for abundance assessment [107]. However, despite these advantages, acoustic methods generally lack specificity, and are highly sensitive to fish orientation [54, 53]. Since fish orientation is typically unknown during the time of insonification, this presents a potential problem for translating measured energy into abundance. In the case of individual fish, acoustic scattering is highly sensitive to fish size, shape, and orientation. This limits the inferences that can be made with conventional single-frequency and single-view methods.

In the context of zooplankton classification, broadband scattering measurements recorded at several different angles can offer dramatic improvements in the discrimination capabilities over both narrowband and single-view methods [59, 60]. Here, the key aspect of improvement is a result of exploiting characteristic shape differences across multiple views for different zooplankton classes. Although fish typically share more shape similarity between classes than zooplankton, these

methods can be naturally extended to fish classification with appropriate modifications to the learning algorithms.

In addition to classification, it is useful to estimate physical parameters of the fish such as size, shape, orientation, and perhaps even internal structure. Such an approach has been presented [58] for estimating swim-bladder length from multiple views using narrow or wideband scattering. This method was based on analysis of variability in echo magnitude as a function of view angle and frequency. In this document, an alternative approach is investigated based on echo duration. Here, the echo duration from multiple views is converted into a "thickness." These thickness estimates are then compared to an ellipsoid model to yield an estimate for ellipsoidal shape.

In addition, it is shown that both echo energy and echo duration can be combined to create a two-dimensional reconstruction that is a projected image of the fish. The method is investigated using full aperture scattering measurements and incorporates a procedure for aligning and windowing these data.

6.2 Laboratory Scattering Measurements

In order to observe multi-view broadband scattering measurements a laboratory system for recording these data was constructed.

6.2.1 Multi-view scattering apparatus

The details of the scattering apparatus and its calibration are described in Roberts et. al 2008 [60]. The system is used in a similar configuration, except that the transmit frequencies are lower, and the fish are harnessed in the field of view under a sedation protocol approved by the University of California, San Diego (IACUC #S07191). Broadband, bistatic scattering is recorded simultaneously on eight receivers for a single transmit pulse. The receiver array is positioned at mid-depth in a large elliptical tank, 1.25 m off the bottom (Figure 6.1). The transmitter and receivers are aligned so that their beams intersect at a common location that is 1.14 m away from the array. The field of view varies with frequency, but is no

smaller than 5 cm3. Two cameras record the position and orientation of the fish while it is being insonified. During insonification, the fish is held roughly horizontal in the plane of the array and rotates about the vertical axis.

6.2.2 Fish scattering experiments

Two sets of experiments were performed using broadband linear frequency-modulated (LFM) chirps, one in February 2007 with 250 kHz of bandwidth with a center frequency of 500 kHz, and the second in September 2007 with 250 kHz of bandwidth and a center frequency of 750 kHz. In total, 50 specimens were insonified by the system. Of those, 36 resulted in usable data where the orientation of the fish was roughly horizontal during insonification, the fish stayed alive during the entire experiment, and data were free from artifacts due to bubbles or other contaminants. Scattering data were preprocessed by matched filtering raw echoes with a transmit signal model [93]. The matched filter output was windowed around the peak with a window duration of 100 μs. Windowing was applied to each receiver independently to account for differences in travel time between the transmitter and receiver. Video data provided an estimate for the orientation of the fish during each ping.

6.3 Multi-view fish classification

There has been substantial research in the area of multi-view target classification during the past decade [63, 68, 70]. This has lead to a progression towards methods that combine both discriminant and probabilistic algorithms [70]. The classification algorithm used in this work is a network based approach applying a cascade of multi-layer perceptions (MLP) and a single support vector machine (SVM) to combine probabilistic predictions from the individual views. The features used were discrete cosine transform (DCT) coefficients computed from received echoes. The algorithm was evaluated on two, three-class problems where the algorithm was trained using a subset of available data and tested on the remaining data. Average error rates were obtained by repeating this process five

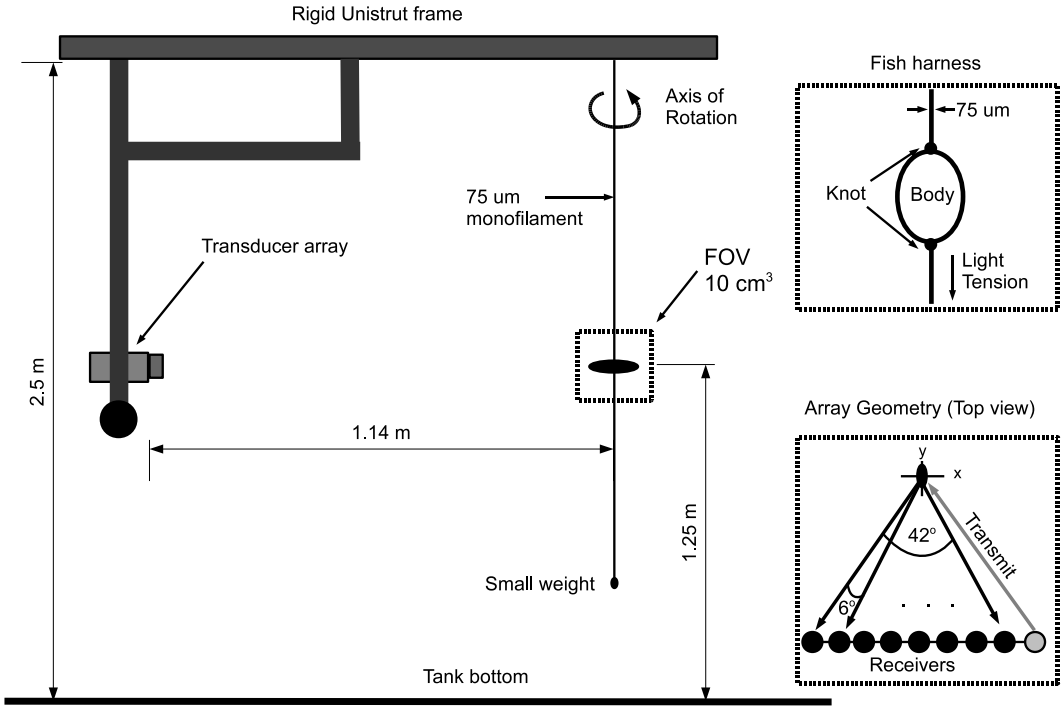


Figure 6.1: Two-dimensional drawing of the experimental setup showing the geometry of the problem, the harness used to hold fish in the field of view, and the array geometry with 6^o spacing and a full aperture of 42^o.

times using different sets for training and testing. The classification results for the two problems are shown in Figure 6.2. Size-based classification is accurate (10 % error), and significant (40 %) reductions in error are obtained using a small number of views. In the case of species-based classification, the absolute error is significantly higher, yet the fractional error reduction remains at roughly 40%. In addition, the error curve does not level out after eight views indicating that further reduction in error could be achieved by acquiring additional views.

6.4 Multi-view size and orientation estimation

6.4.1 Ellipsoid model

Multi-view echo data were combined with an ellipsoid scattering model to estimate fish size and orientation. The scattering model is a special case of the distorted wave Born approximation (DWBA) [84] which assumes the scatterer is an ellipsoid. For a single ellipsoid, the DWBA has been shown to be a spherical Bessel function of the first kind (Appendix A)

$$S(\mathrm{k}) = C(g,h,\alpha)\mathrm{k}_1^2 abc\frac{j_1(\mathrm{k}P_t)}{\mathrm{k}P_t}. \tag{6.1}$$

where

$$P_t = \sqrt{a^2cos^2(\phi)sin^2(\theta) + b^2sin^2(\phi)sin^2(\theta) + c^2cos^2(\theta)}, \tag{6.2}$$

and a, b, and c are the principle radii of the ellipsoid and ϕ, and θ define the direction of the wave vector with wavenumber k with respect to the principle axes of the ellipsoid. The scaling term $C(g,h,\hat{\mathrm{I}}\acute{\mathrm{s}}\alpha)$ is a function of the gamma contrasts of the scatterer [84]. The parameter ϕ was assumed to be zero since the fish rotates in the plane of the array, and the parameter b was fixed to a constant value. The P_t term is a âĂIJthicknessâĂİ parameter, and can be related to an âĂIJequivalent spherical radiusâĂİ due to the fact that the DWBA for a sphere of radius r can be obtained from equation (6.1) by substituting $r = P_t$ and $r^3 = abc$ (Appendix A).

6.4.2 Posterior parameter estimation

The unknown ellipsoid parameters were assumed to be random and unknown. A factorized prior was assumed over the model of the form

$$p(a,c,\theta) = p(a,c)p(\theta), \tag{6.3}$$

where $\mathbf{\Theta} = (a,c)^T$ and

$$p(\Theta) = \frac{1}{sqrt(2\pi)^2|\mathbf{\Sigma}|}\exp\left((\Theta - \boldsymbol{\mu}_\Theta)^T\,\mathbf{\Sigma}^{-1}\,(\Theta - \boldsymbol{\mu}_\Theta)\right), \tag{6.4}$$

and $p(\theta)$ is uniform.

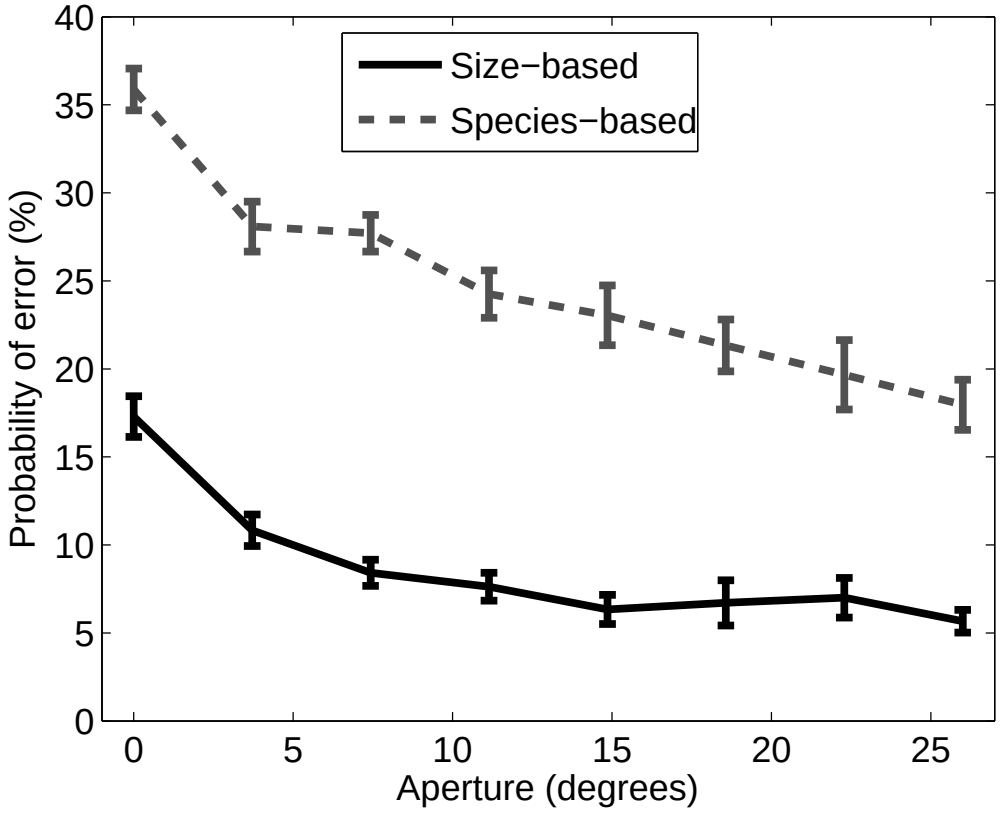


Figure 6.2: Probability of error for the size- and species-based classification problems with three classes. The error is plotted as a function of the effective aperture in degrees resulting from increasing the number of views. Error bars denote one standard deviation. Fractional reductions in error for each problem are above 40% using all eight views for both classification problems.

The likelihood function was assumed to be Gaussian with unknown variance which was estimated from data using the sample variance. The error for the i^{th} view was computed as

$$e_i = \hat{P}_i - P_i(a, c, \theta), \tag{6.5}$$

where $\hat{P}_i$ is the thickness for the i^{th} view estimated from echo data using frequency domain analysis.

Examples of marginal posterior probability density functions (pdfs) for a, c, and θ are shown as a function of the number of views of the fish, when the views cover 360^o (Figure 6.3), and also when the views are increasingly limited (Figure

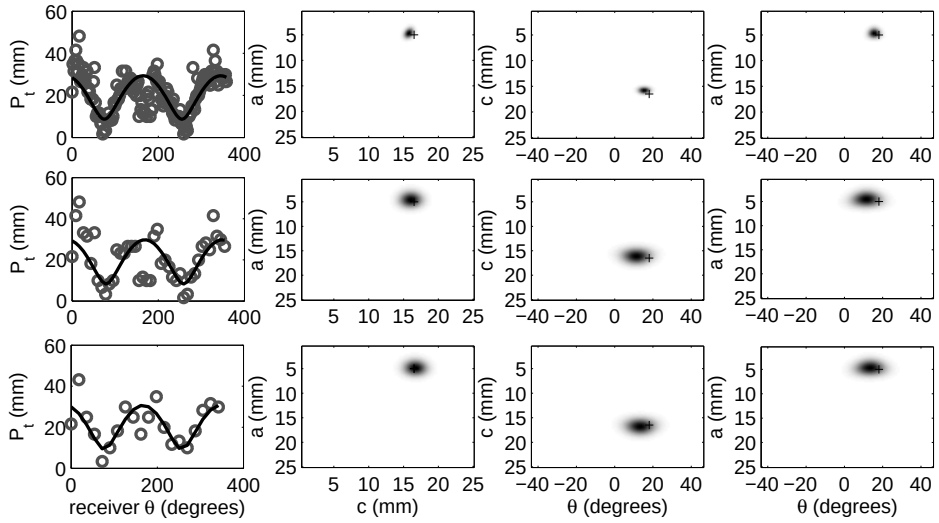


Figure 6.3: Marginal posterior pdfs for different angular sampling frequencies (rows). The left column shows the modeled (black line) and estimated thickness (gray circles) and the other columns show the marginal pdfs. Going from the top to the bottom row, the angular sample frequency is reduced by 50% at each step. The measured parameters are shown by the black "+" signs.

6.4). In the case of the limited views, a size proportional prior (Equation 6.4) is used to constrain the posterior pdfs. As can be seen, the estimates are tightly peaked near the true parameter values when a complete set of views are available (Figure 6.3). In contrast, when the aperture is decreased, the distributions stay peaked until the aperture decreases below 60^o. When the aperture is highly limited (Figure 6.4, bottom row), the distributions are quite broad demonstrating a fundamental uncertainty between length and thickness of the fish at limited aperture.

6.5 Fish shape reconstruction

In this section, the reconstruction of the 2D image of the fish (as viewed from above) is investigated using the full 360^o of backscattering over 250 kHz of bandwidth centered at 750 kHz. Due to the limited bandwidth, and the use of backscatter, the range of spatial frequencies sampled by the system is limited

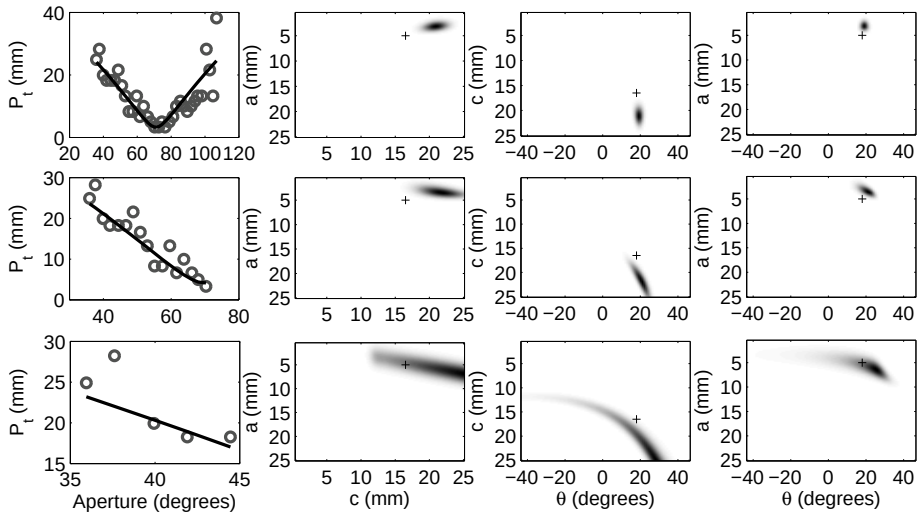


Figure 6.4: Marginal posterior pdfs for different apertures (rows). The left column shows the modeled (black line) and estimated (gray circles) thickness and the other columns show the marginal pdfs. Going from the top to the bottom row, the aperture is 60, 30, and 10 degrees. The measured parameters are shown by the black "+" signs

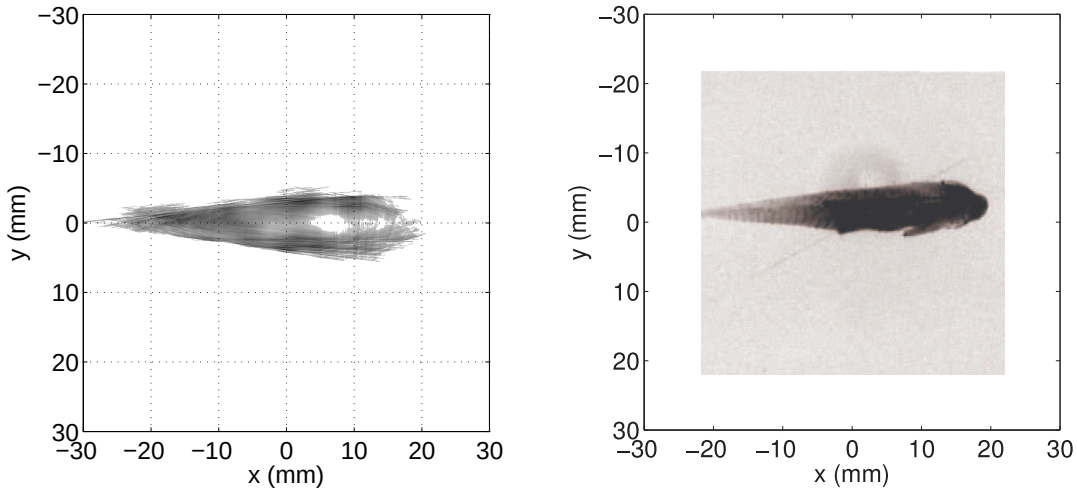
to an annulus centered at the origin with width radial frequencies varying from 650 to 900 kHz. This prevents direction reconstruction from echo data as no low frequency data is available. To mitigate this problem, the echo envelope is used to approximate the impulse response of the fish at a given observation angle. This is similar to creating a time of flight image of strong scatterers on the fish body. The laboratory experiments were not designed with this type of reconstruction in mind, and the echoes must first be registered for different fish orientations. To align and window echoes, sparse Bayesian learning ([154], [144]) is used to estimate an impulsive impulse response $\mathbf{w}$ which satisfies

$$p(\mathbf{y}|\mathbf{w},\sigma^2) = \frac{1}{\sqrt{(2\pi\sigma^2)}}\exp\left(-\frac{1}{2\sigma^2}||\mathbf{y}-\mathbf{\Phi}\mathbf{w}||^2\right), \tag{6.6}$$

such that

$$p(\mathbf{w};\boldsymbol{\gamma}) = \prod_{i=1}^{N}\frac{1}{sqrt2\pi\gamma_i}\exp\left(-\frac{w_i^2}{2\gamma_i}\right). \tag{6.7}$$

These constraints on $\mathbf{w}$ enforce a sparse solution [154]. Given the registered, windowed echo envelopes for all 360 views, the image of the fish is reconstructed using



(a) (b)

Figure 6.5: Reconstructed 2D image of a damselfish (a) showing internal structure in the reconstruction and very good agreement with the video image (b).

an inverse Radon transform [155] (Figure 6.5).

6.6 Conclusions

In this paper, the application of multi-view acoustic methods to sizing and classifying individual fish has been presented. It has been shown that observing multiple views of scattering offers significant advantages for classifying fish both in terms of size classes and species classes. The absolute error in species-based classification is much higher due to the lack of discriminant information in the echoes. Results of this study show that size-based classification is mostly likely to be a viable and accurate strategy for inferring information from limited view data. When full-aperture data is available, the inverse Radon transform approach performed better than the parametric approach due to mismatch between the ellipsoid model and the true shape of the fish. However, the ellipsoid model was able to yield accurate estimates when full-aperture data was used. These results highlight some promising areas of research in applying multi-view methods problems in fisheries and oceanography.

6.7 Acknowledgements

This work was supported by the California Sea Grant. The authors would like to thank Eddie Kisfaludy, Robert Glatts, Fernando Simonet, Ben Maurer, Erdem Karakoylu, and the SIO machine shop for help with the design and implementation of these laboratory experiments.

The text of Chapter 6 is in part and under some rearrangements a reprint of the material as it appears in Paul L. D. Roberts and Jules S. Jaffe, "Multi-view sizing and classification of individual fish," *Proceedings of the 30th Acoustical Imaging Symposium*, Monterey, CA March 2009. The dissertation author was the primary researcher and author, and the co-author listed in this publication directed and supervised the research which forms the basis for this chapter.

Appendix A

Multi-view, broadband scattering from weak scattering ellipsoids

A.1 Introduction

It is well known that plane wave scattering in the far field under the Born approximation can be represented as a Fourier transform of the scatterer [90]. In the context of inffering biophsyical information about the scatterer from measurments of the far field scattering pattern, perhaps the most important consequence of the Fourier transform relationship is that the finite support of the scatterer implies that the far field pattern will be band-limited. If scattering is observed at a fixed angle over a wide band of frequencies, the oscilations in the pattern directly code the support of the scatterer along the direction of wave. Estimation of this oscilatory frequency (or its time domain, echo-duration equivalent) is at the core of most of the estimation methods presented in this dissertation. Therefore, the analytic expresion for this frequency is derived for a general ellipsoid as this shape yields a good model for scattering from many zooplankton, and a reasonable model for scattering from the body of some fishes. The expresion is derived under assumption of homogeneous scattering from a fluid-like ellipsoid with density and sound speed compariable to that of the surrounding medium. In this case, a modification of the Born approximation is applicable.

The distorted wave Born approximation (DWBA) [82, 84] for a spheriod [156], and ellipsoid (Dezhang Chu, Woods Hole Oceanographic Institute, USA, pers. comm., 2007) have been derived previously. Here, a derivation of the DWBA for an ellipsoid is offered which directly applies the projection-slice theorem, naturally incorporates the notion of an "equivalent radius," and results in the proof of an equivalent sphere theoreom.

A.2 DWBA for a homogeneous ellipsoid

The DWBA for a general bistatic configuration can be written as

$$S(\mathbf{k}) = \frac{\mathrm{k}_1^2}{4\pi} \int\int\int_V \left(\gamma_\kappa(\mathbf{x}) - \gamma_\rho(\mathbf{x}) \cos\alpha\right) e^{i\mathbf{k}\cdot\mathbf{x}} d\mathbf{x}, \tag{A.1}$$

where the wavevector $\mathbf{k}$ is derived from the wavevectors for the bistatic incident and scattered wavevectors $\mathbf{k}_i$ and $\mathbf{k}_s$ according to [59]. If the scatterer is homogeneous, the DWBA simplifies to

$$S(\mathbf{k}) = C(g,h,\alpha) \frac{\mathrm{k}_1^2}{4\pi} \int\int\int_V e^{i\mathbf{k}\cdot\mathbf{x}} d\mathbf{x}, \tag{A.2}$$

where

$$C(g,h,\alpha) = \left(\gamma_\kappa - \gamma_\rho \cos\alpha\right), \tag{A.3}$$

and

$$\gamma_\kappa = \left(1 - gh^2\right)/gh^2, \tag{A.4}$$

and

$$\gamma_\rho = (g-1)/g. \tag{A.5}$$

The direction of the wavevector $\mathbf{k} = (k_1, k_2, k_3)^T$ will be defined in terms of the angles ϕ and θ according to

$$k_1 = \mathrm{k}\cos(\phi)\sin(\theta), \tag{A.6}$$

$$k_2 = \mathrm{k}\sin(\phi)\sin(\theta), \tag{A.7}$$

$$k_3 = \mathrm{k}\cos(\theta), \tag{A.8}$$

where k is the wavenumber in the body of the scatterer.

The DWBA for a homogenious ellipsoid can be derived using a generalization of the projection-slice theorem, and the intersections of planes and ellipsoids at arbitrary orienations. The volume of integration is given by the set

$$V = \{\mathbf{x} | \mathbf{x}^T R \mathbf{x} \ \leq 1 \ \forall \ \mathbf{x} \ \in \ \Re^3\}, \tag{A.9}$$

where $R = \text{diag}(a^{-2}, b^{-2}, c^{-2})$, and the quantities a, b, and c make up the lengths of the principle axes of the ellipsoid. Given the definition of the volume in equation (A.9), equation (A.2) can be recognized as the three-dimensional Fourier transform of the function

$$f(\mathbf{x}) = \left\{ \begin{array}{c} C(g,h,\alpha)\frac{\mathrm{k}_1^2}{4\pi}, \quad \text{for} \ \ \mathbf{x} \ \ \in \ \ V \\ 0, \ \ \text{otherwise} \end{array} \right\}. \tag{A.10}$$

Using the projection-slice theorem (Appendix B.2), equation (A.2) can be written as

$$S(\mathrm{k}) = C(g,h,\alpha)\frac{\mathrm{k}_1^2}{4\pi}\int_{-P_t}^{P_t} f_V(P_t, x')e^{i\mathrm{k}x'}dx', \tag{A.11}$$

where x' is a point along the direction $\mathbf{d} = (cos(\phi)sin(\theta), sin(\phi)sin(\theta), cos(\theta))^T$, $f_V(P_t, x')$ is the area of the intersection ellipse between the ellipsoid and the plane normal to $\mathbf{d}$ centered at $x'\mathbf{d}$, and P_t is the length of the vector from the ellipsoid center to the plane that is normal to $\mathbf{d}$ and tangent to the ellipsoid surface. From the derivations given in [157], the expresions for P_t, and $f_V(P_t, x')$ are

$$P_t = \sqrt{a^2cos^2(\phi)sin^2(\theta) + b^2sin^2(\phi)sin^2(\theta) + c^2cos^2(\theta)}, \tag{A.12}$$

and

$$f_V(P_t, x') = \frac{abc\pi}{P_t}\left(1 - \frac{x'^2}{P_t^2}\right). \tag{A.13}$$

Subsituting equation (A.12) and equation (A.13) into equation (A.11) yields

$$S(\mathrm{k}) = C(g,h,\alpha)\frac{\mathrm{k}_1^2 abc}{4P_t}\int_{-P_t}^{P_t}\left(1 - \frac{x'^2}{P_t^2}\right)e^{i\mathrm{k}x'}dx', \tag{A.14}$$

which can be seen as the Fourier transform of a parabola on the interval $[-P_t, P_t]$. This has the solution (Appendix B.1)

$$\int_{-P_t}^{P_t}\left(1 - \frac{x'^2}{P_t^2}\right)e^{i\mathrm{k}x'}dx' = \frac{4}{\mathrm{k}}j_1(\mathrm{k}P_t), \tag{A.15}$$

where $j_1(x)$ is the spherical bessel function of the first kind. Substituting equation (A.15) into equation (A.14) yields

$$S(\mathrm{k}) = C(g,h,\alpha)\mathrm{k}_1^2 abc\frac{j_1(\mathrm{k}P_t)}{\mathrm{k}P_t}. \tag{A.16}$$

A.3 Equivalent sphere theorem

With the DWBA for the general ellipsoid, the DWBA for a sphere of radius r follows from letting $a = b = c = r$. Substituting into equation (A.16) and simplifying yields

$$S(\mathrm{k}) = C(g,h,\alpha)\mathrm{k}_1^2 r^3\frac{j_1(kr)}{kr}. \tag{A.17}$$

It can been seen that equation (A.16) and equation (A.17) are equivalent up to a scale factor under the definition of an equivalent radius

$$r_{eq} = P_t = \sqrt{a^2cos^2(\phi)sin^2(\theta) + b^2sin^2(\phi)sin^2(\theta) + c^2cos^2(\theta)}, \tag{A.18}$$

where the ratio of the DWBA for the ellpsoid to the DWBA for the sphere is the constant

$$K = \frac{abc}{r_{eq}^3}. \tag{A.19}$$

The constant K is the ratio of the volume of the ellipse with semi-axes a, b, and c, to that of the sphere of radius r_{eq}. Note that r_{eq} is exactly equal to the distance between two parallel planes normal to the direction of the wavevector $\mathbf{k}$, the plane through the origin of the ellipsoid, and the plane tangent to the surface of the ellipsoid [157].

Appendix B

Fourier transforms for ellipsoids

B.1 Fourier transform of a finite parabola

We wish to find the Fourier transform of a parabola of the form $(1-x^2/r^2)$ on the interval $[-r, r]$. The Fourier transform is given by

$$F(\mathrm{k}) = \int_{-r}^{r} \left(1 - \frac{x^2}{r^2}\right) e^{i\mathrm{k}x} dx \tag{B.1}$$

$$= \int_{-r}^{r} e^{i\mathrm{k}x} dx - \frac{1}{r^2} \int_{-r}^{r} x^2 e^{i\mathrm{k}x} dx \tag{B.2}$$

$$= 2\frac{\sin(\mathrm{k}r)}{\mathrm{k}} - \frac{1}{r^2}\left(2r^2\frac{\sin(\mathrm{k}r)}{\mathrm{k}} - \left(4\frac{\sin(\mathrm{k}r)}{\mathrm{k}^3} - 4r\frac{\cos(\mathrm{k}r)}{\mathrm{k}^2}\right)\right) \tag{B.3}$$

$$= \frac{4}{\mathrm{k}}\left(\frac{\sin(\mathrm{k}r)}{\mathrm{k}^2 r^2} - \frac{\cos(\mathrm{k}r)}{\mathrm{k}r}\right) \tag{B.4}$$

$$= \frac{4}{\mathrm{k}} j_1(\mathrm{k}r), \tag{B.5}$$

where the final step in the solution made use of the identity

$$j_1(x) = \frac{\sin(x)}{x^2} - \frac{\cos(x)}{x}, \tag{B.6}$$

where $j_1(x)$ is a spherical bessel function of the first kind [158].

B.2 n-1 fold projection-slice theorem

Following the development of the projection-slice theorem in [159], we wish to show that the n-dimensional Fourier transform of an n-dimensional function

evaluated along a line in Fourier space is equivalent to the one-dimensional Fourier transform of the $n-1$ fold projection of the function onto the line. First, define the n-1 fold projection of the function $f(\mathbf{x})$ onto the dimension x_i as the integral over all other dimensions

$$p(x_i) = \int \cdots \int f(\mathbf{x}) dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_n. \tag{B.7}$$

The n-dimensional Fourier tranform of the function $f(\mathbf{x})$ is

$$F(\boldsymbol{\mu}) = \int \cdots \int f(\mathbf{x}) e^{-i\boldsymbol{\mu}^T \mathbf{x}} dx_1 \cdots dx_n. \tag{B.8}$$

Now, define the line along which the transform is evaluated to be $\mu\mathbf{d}$ for $\mu \in [-\infty, \infty]$ and $\|\mathbf{d}\|_2 = 1$. Without loss of generality, we can define the line to lie on the x_1 axis such that $\mathbf{d} = (1, 0, ..., 0)^T$. The transfom, evaluated as a function of μ, then becomes

$$\begin{aligned} F_{\mathbf{d}}(\mu) &= \int \cdots \int f(\mathbf{x}) e^{-i\mu\mathbf{d}^T \mathbf{x}} dx_1 \cdots dx_n & \text{(B.9)} \\ &= \int \cdots \int f(\mathbf{x}) e^{-i\mu x_1} dx_1 \cdots dx_n & \text{(B.10)} \\ &= \int_{-\infty}^{\infty} \left[\int \cdots \int f(\mathbf{x}) dx_2 \cdots dx_n \right] e^{-i\mu x_1} dx_1 & \text{(B.11)} \\ &= \int_{-\infty}^{\infty} p(x_1) e^{-i\mu x_1} dx_1, & \text{(B.12)} \end{aligned}$$

where $p(x_1)$ is the $n-1$ fold projection of the function $f(\mathbf{x})$ onto the x_1 dimension.

Appendix C

Additional error curves for the fixed geometry

Error curves for additional data sets defined in Table 4.3 are shown in the following figures for all three feature spaces and classification algorithms defined in Chapter 4. Error curves in this appendix are computed by treating the view geometry as being fixed and therefore the angular sampling frequency and aperture do not vary between training and testing. The aperture is limited by the experimental setup (Section 4.2).

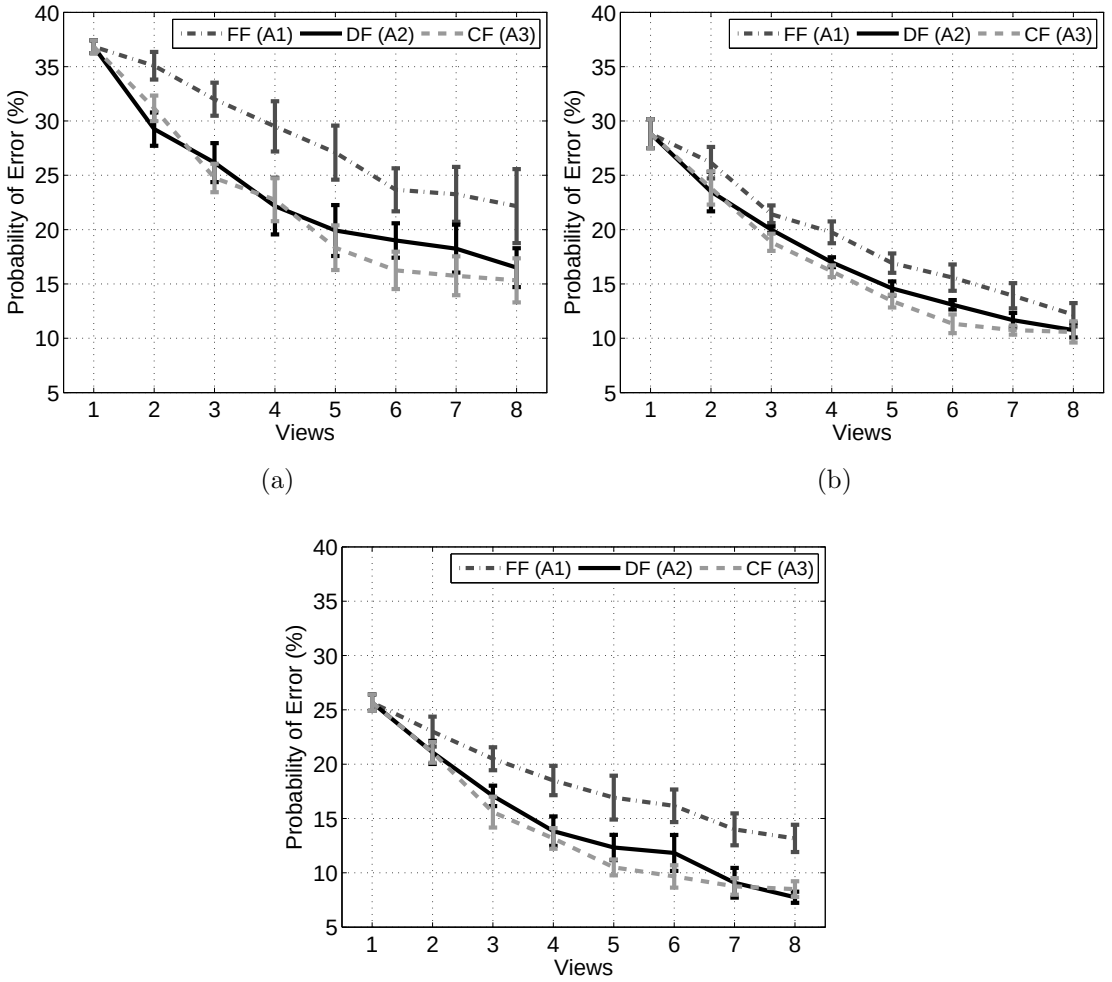


Figure C.1: Probability of error on D_2 vs. the number of views used for classification, all three classification algorithms, and the DCT (a), db4 (b) and Haar (c) feature spaces.

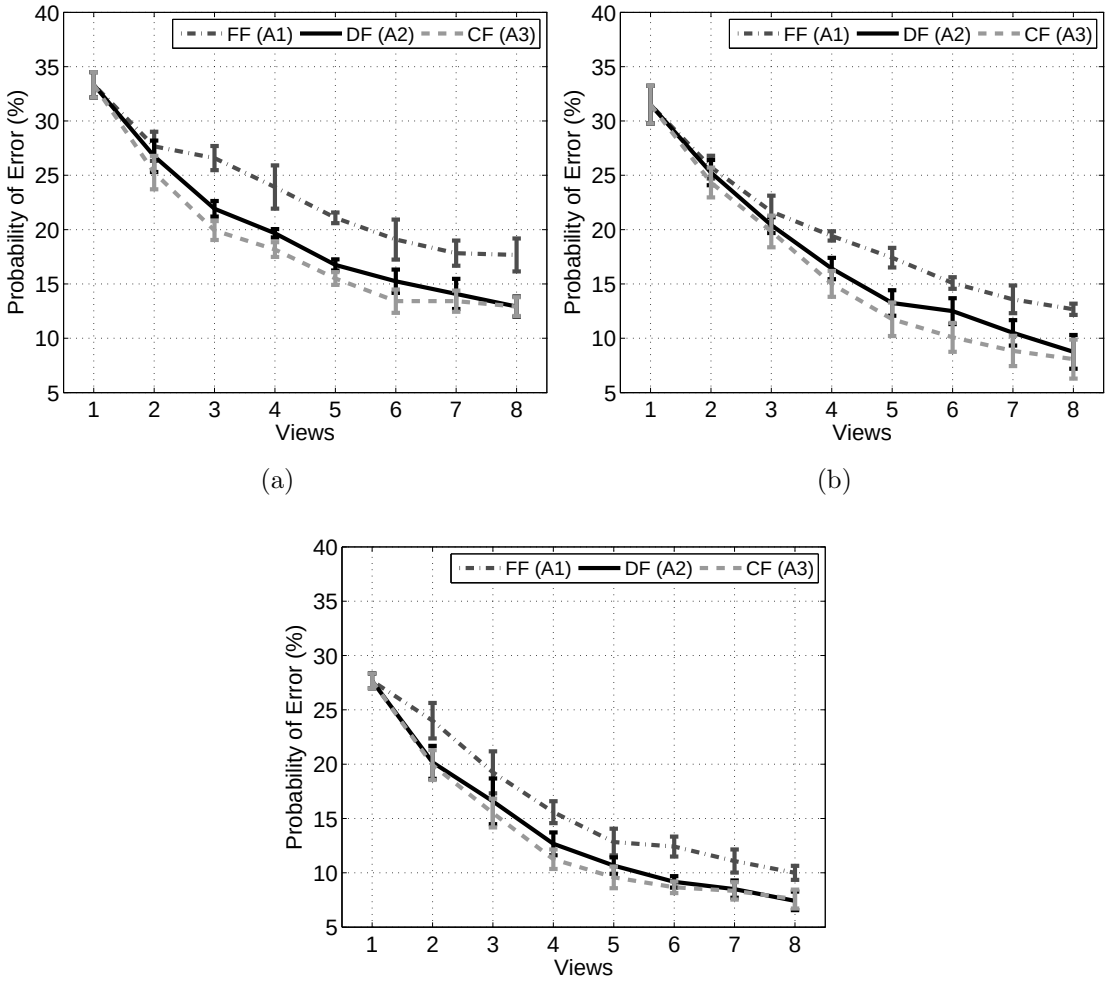


Figure C.2: Probability of error on D_3 vs. the number of views used for classification, all three classification algorithms, and the DCT (a), db4 (b) and Haar (c) feature spaces.

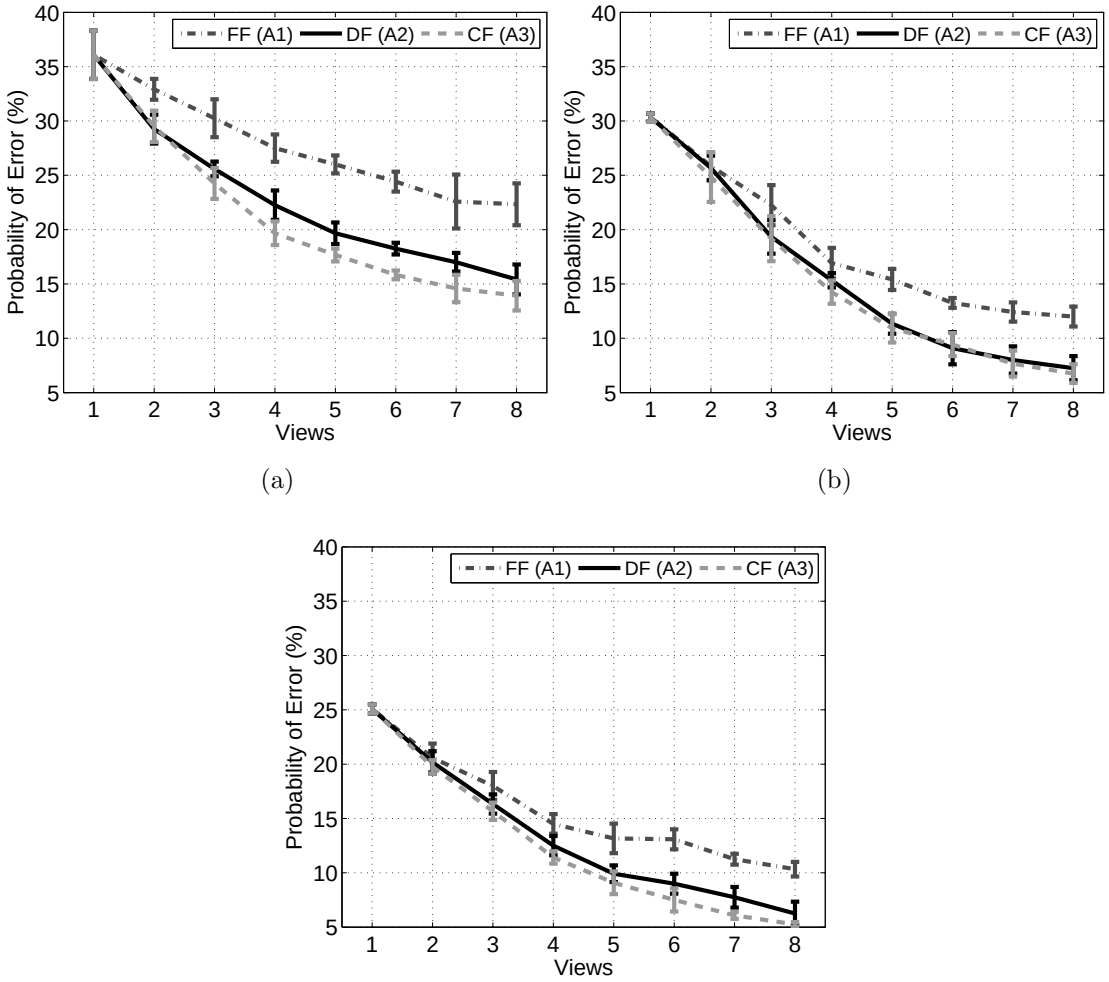
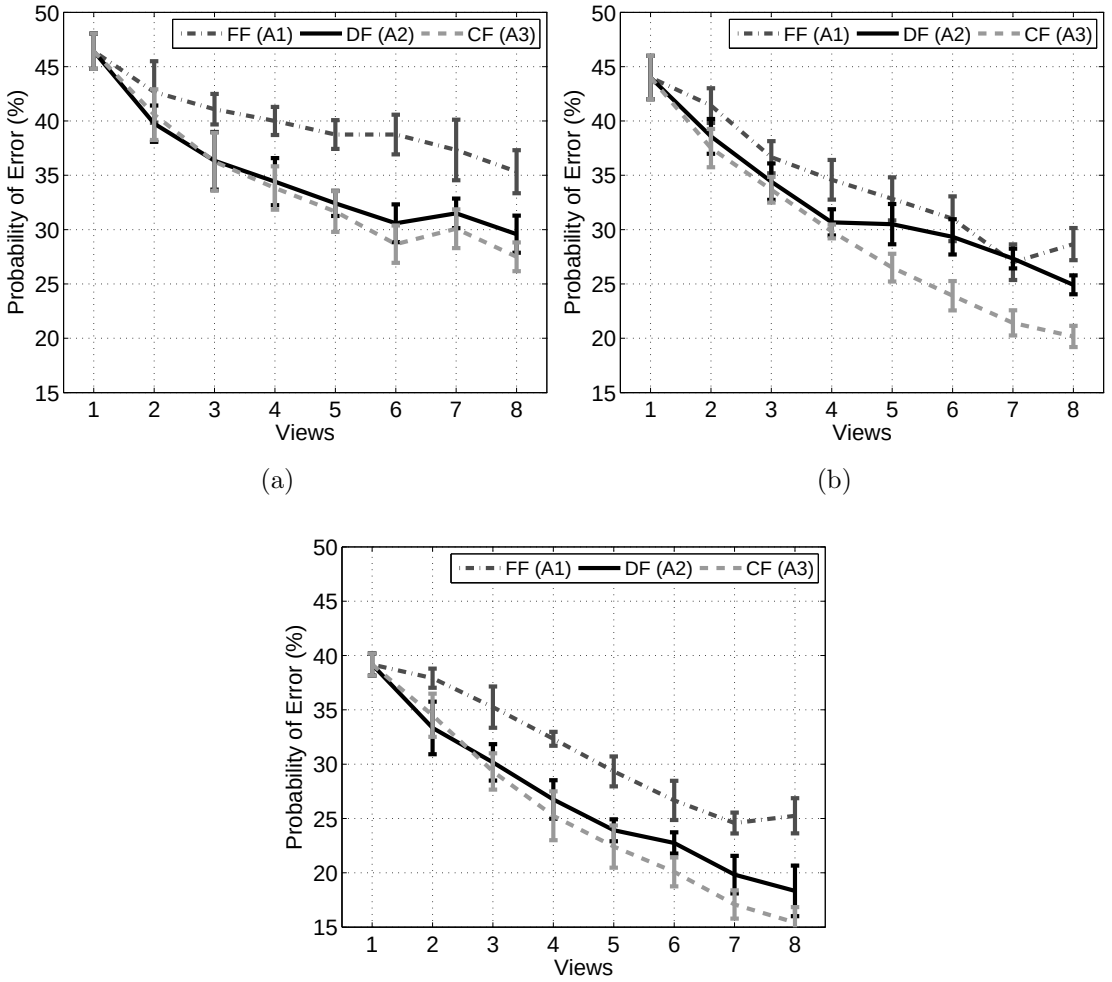


Figure C.3: Probability of error on D_4 vs. the number of views used for classification, all three classification algorithms, and the DCT (a), db4 (b) and Haar (c) feature spaces.



(a)

(b)

(c)

Figure C.4: Probability of error on D_5 vs. the number of views used for classification, all three classification algorithms, and the DCT (a), db4 (b) and Haar (c) feature spaces.

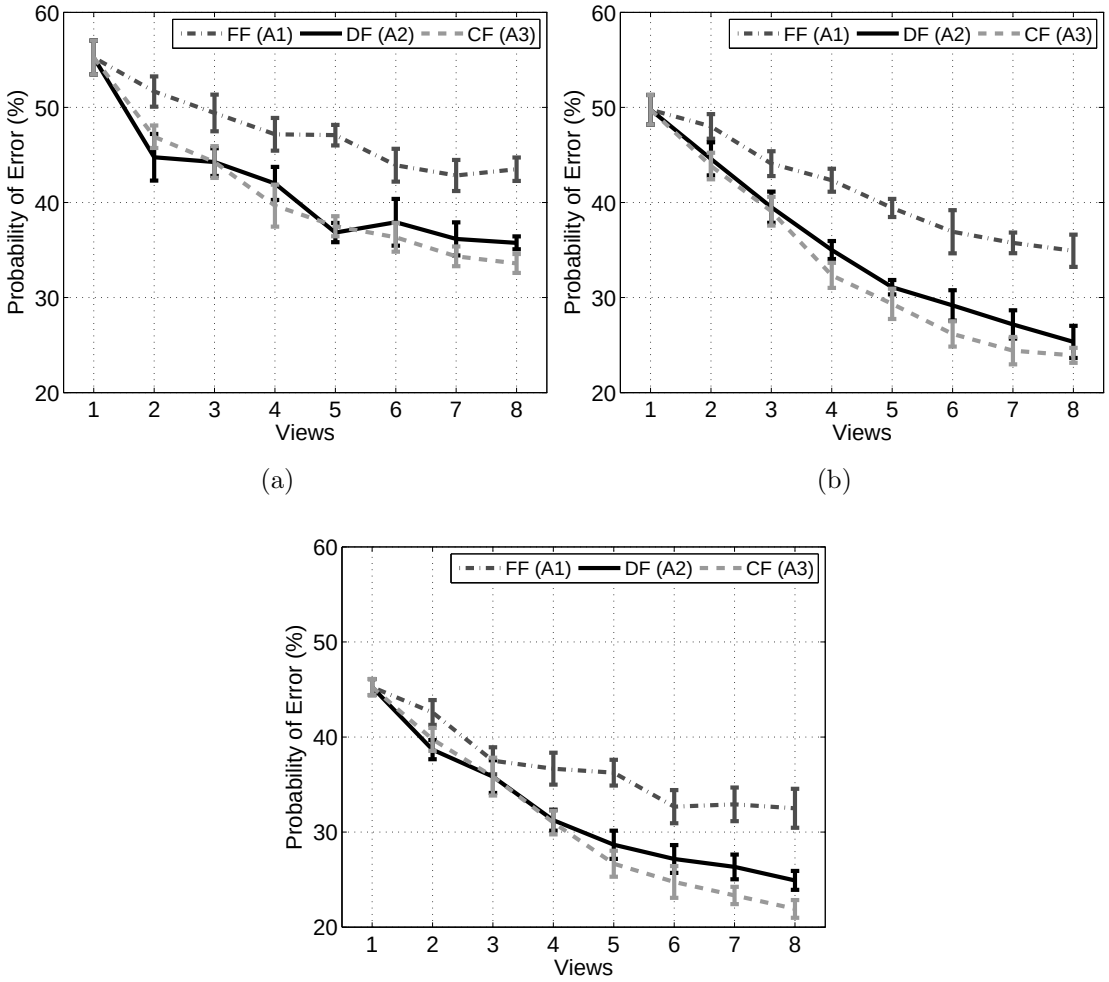
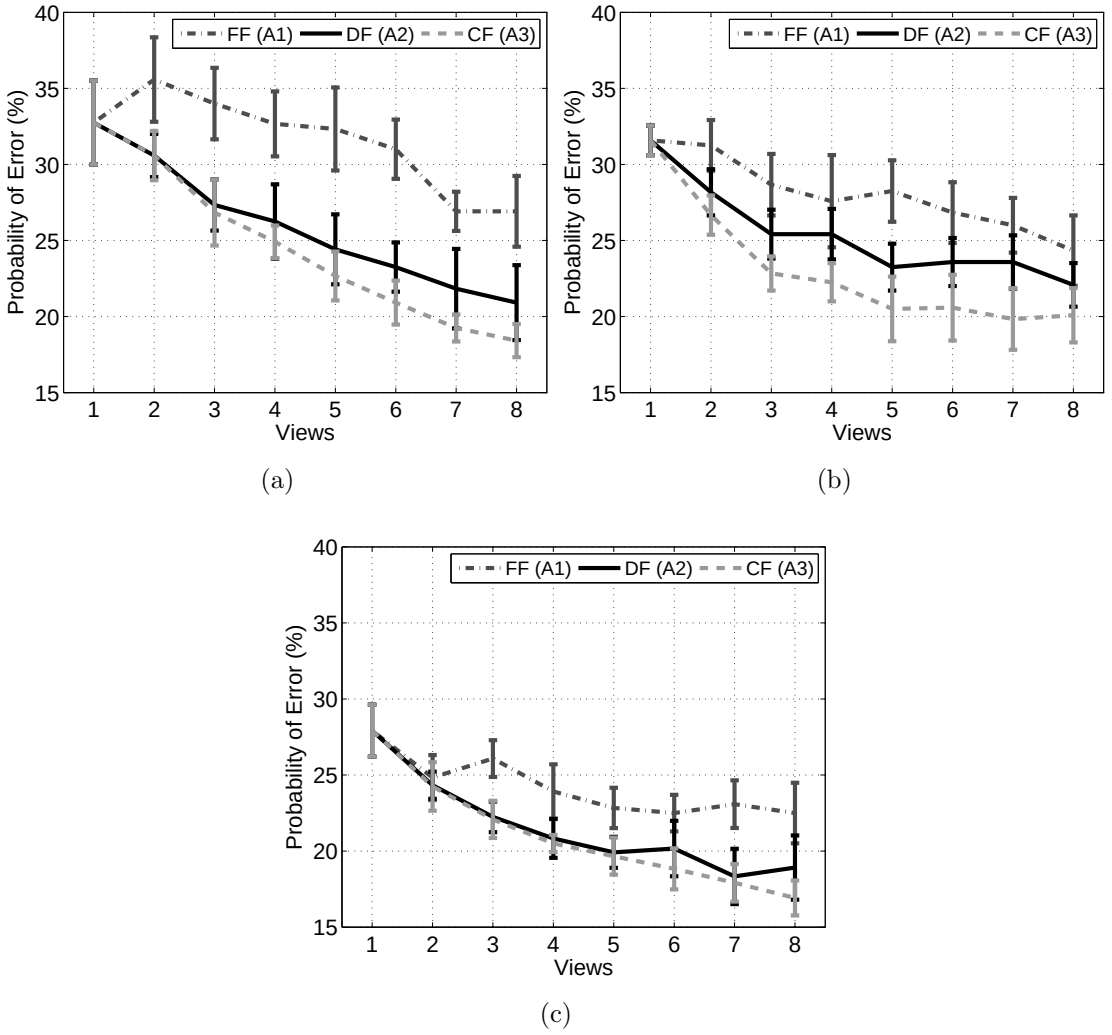


Figure C.5: Probability of error on D_6 vs. the number of views used for classification, all three classification algorithms, and the DCT (a), db4 (b) and Haar (c) feature spaces.



(a) (b)

(c)

Figure C.6: Probability of error on D_7 vs. the number of views used for classification, all three classification algorithms, and the DCT (a), db4 (b) and Haar (c) feature spaces.

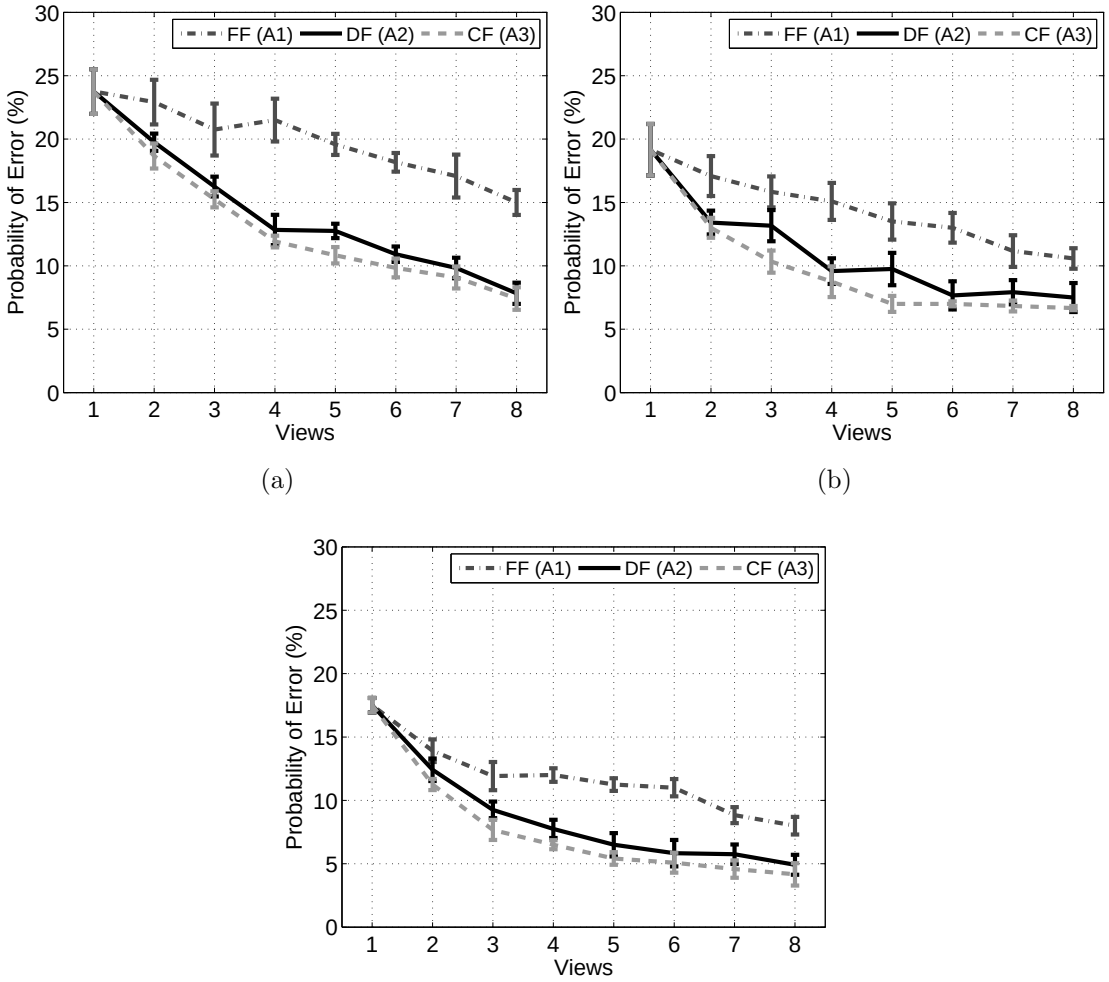
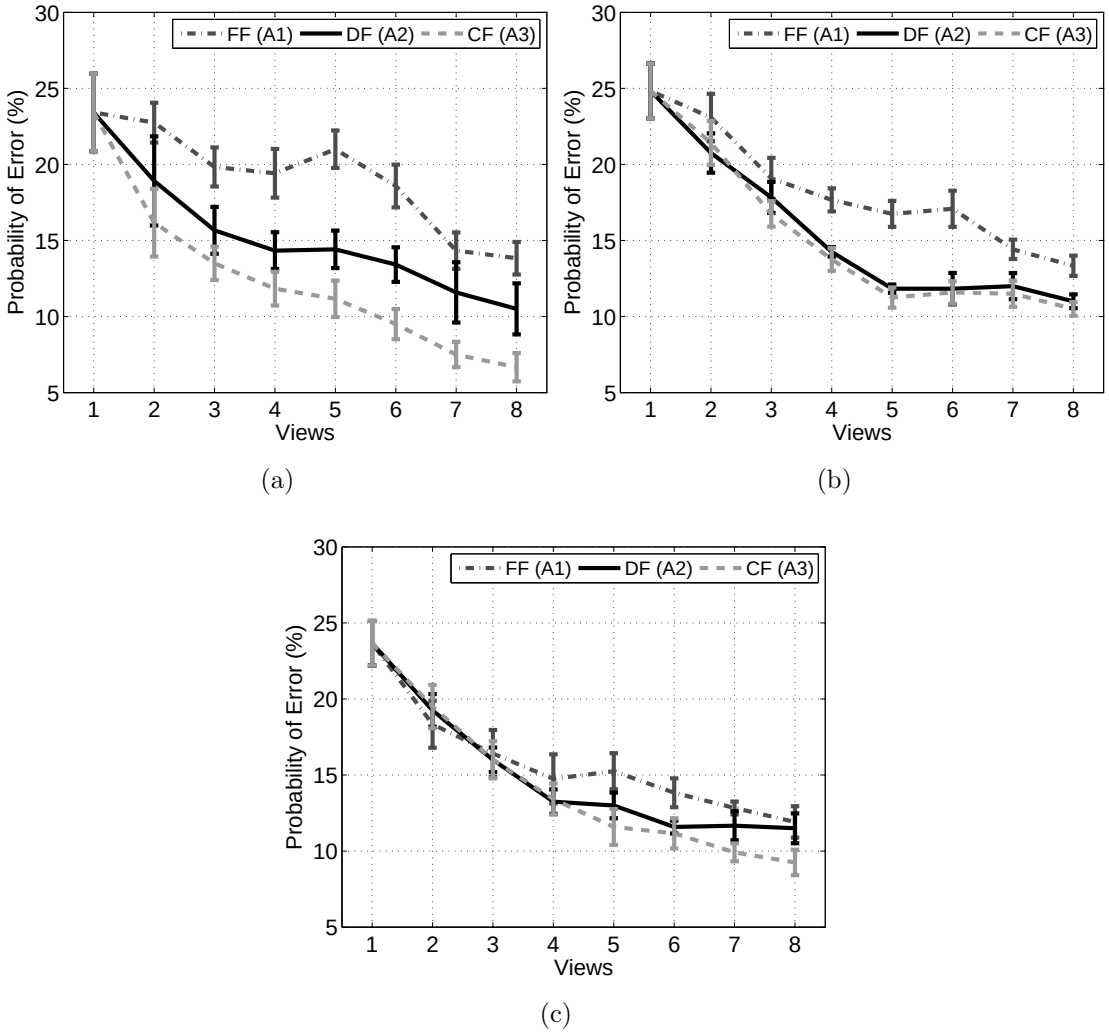


Figure C.7: Probability of error on D_9 vs. the number of views used for classification, all three classification algorithms, and the DCT (a), db4 (b) and Haar (c) feature spaces.



(a) (b)

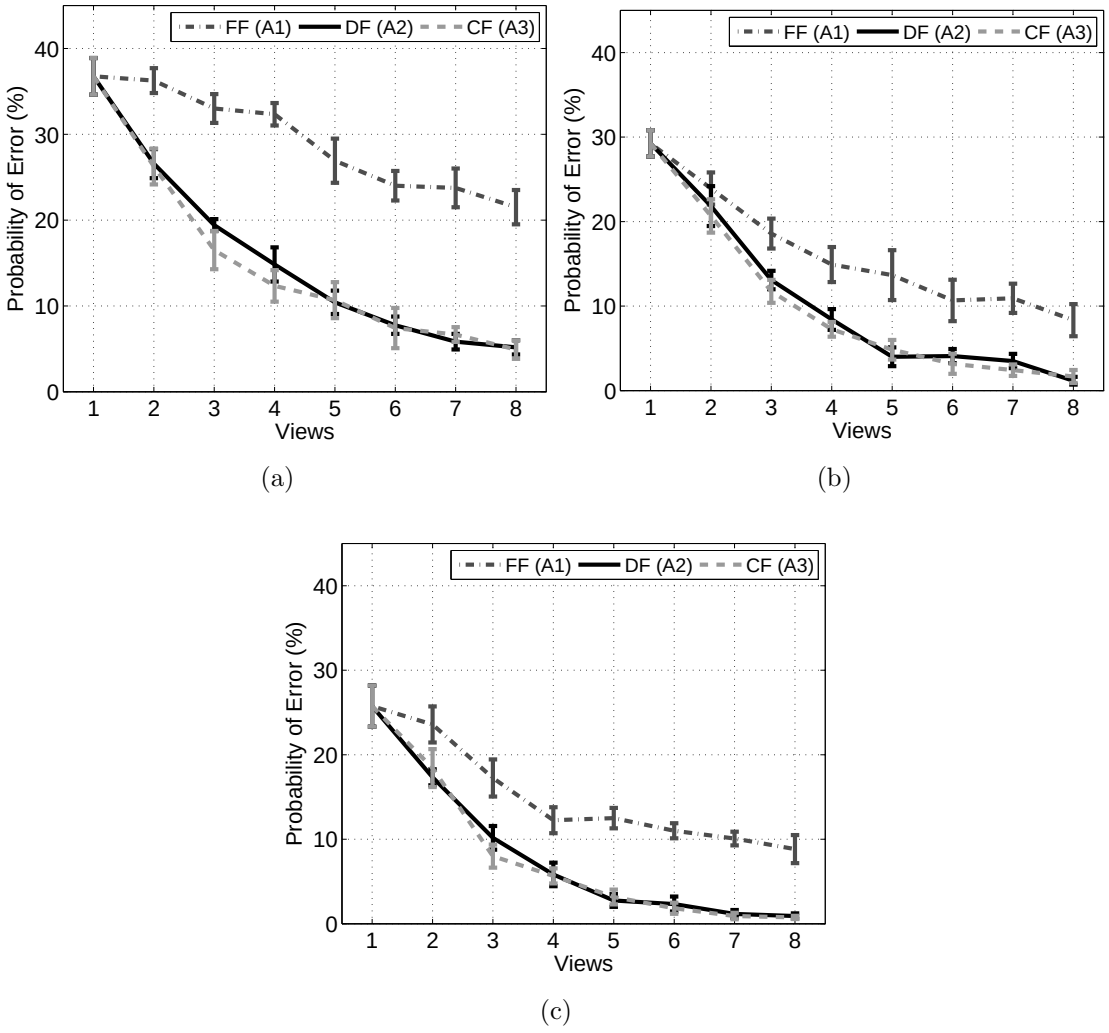
(c)

Figure C.8: Probability of error on D_{10} vs. the number of views used for classification, all three classification algorithms, and the DCT (a), db4 (b) and Haar (c) feature spaces.

Appendix D

Additional error curves for the random geometry

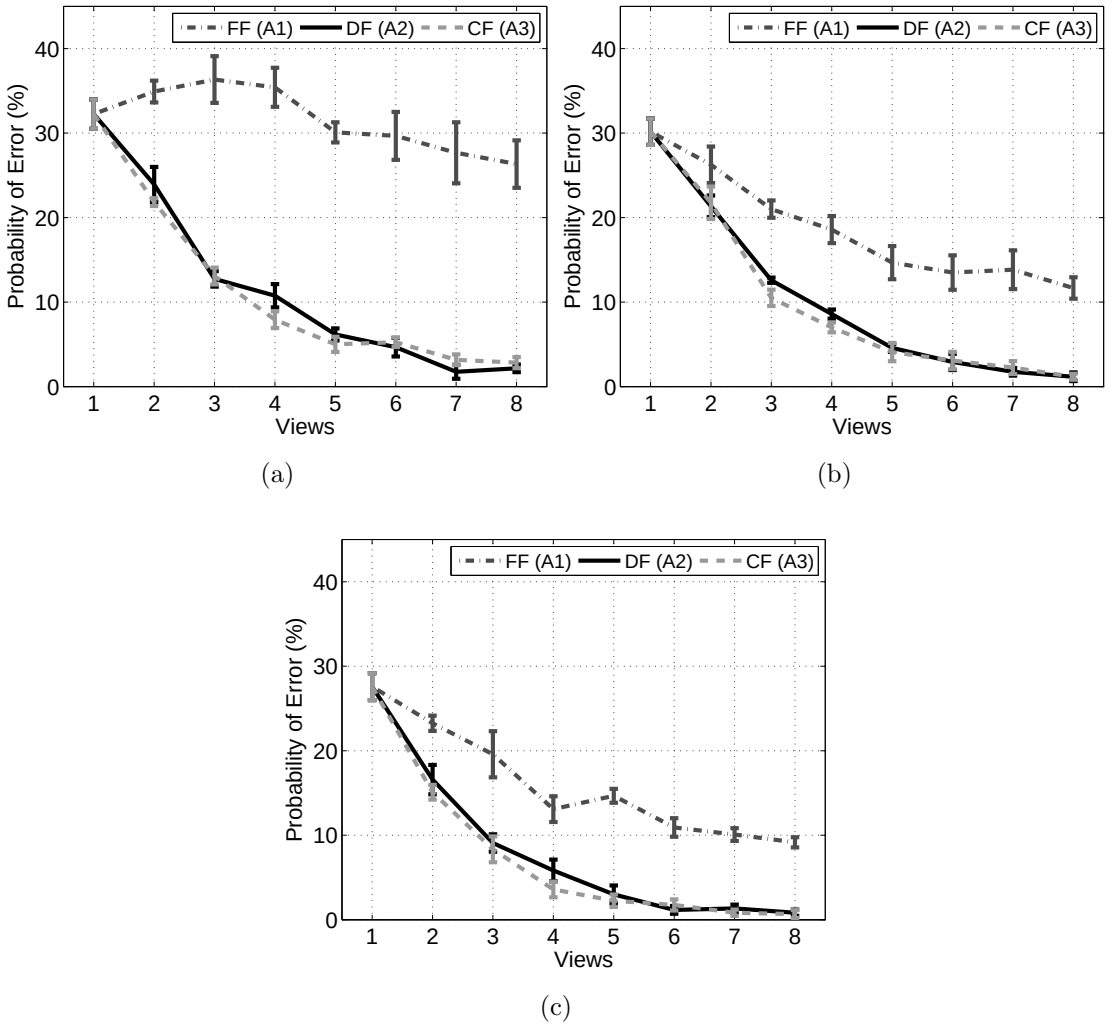
Error curves for the additional data sets defined in Table 4.3 are shown in the following figures for all three feature spaces and classification algorithms defined in Chapter 4. Error curves in this section are computed by treating the view geometry as being random, as defined in Section 4.5.3, and therefore the angular sampling frequency and aperture vary between training and testing. The aperture is not restricted by the experimental setup defined in Section 4.2) and as a result views are sampled from the full 360^os around the fish.



(a) (b)

(c)

Figure D.1: Probability of error on D_2 vs. the number of random views used for classification, all three classification algorithms, and the DCT (a), db4 (b) and Haar (c) feature spaces.

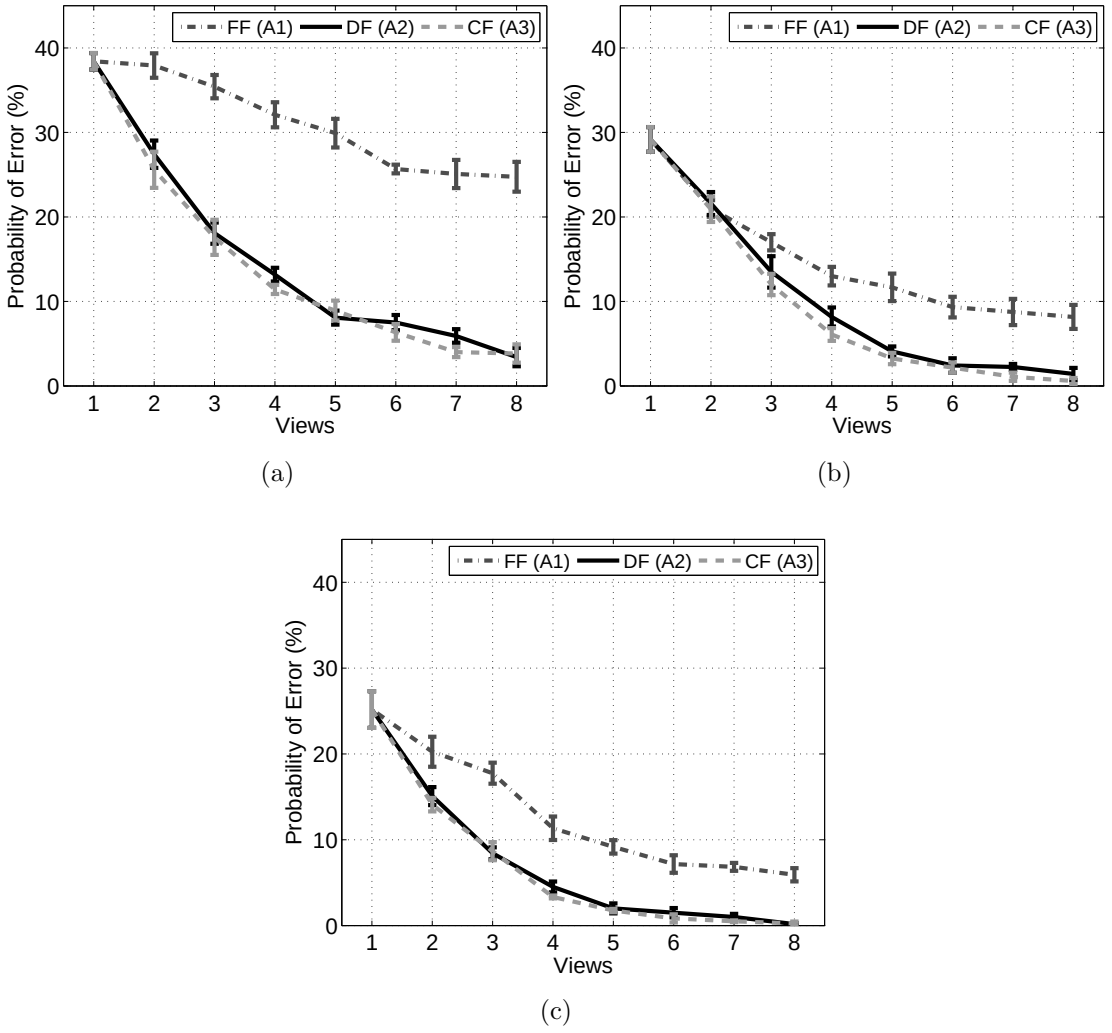


(a)

(b)

(c)

Figure D.2: Probability of error on D_3 vs. the number of random views used for classification, all three classification algorithms, and the DCT (a), db4 (b) and Haar (c) feature spaces.



(a)

(b)

(c)

Figure D.3: Probability of error on D_4 vs. the number of random views used for classification, all three classification algorithms, and the DCT (a), db4 (b) and Haar (c) feature spaces.

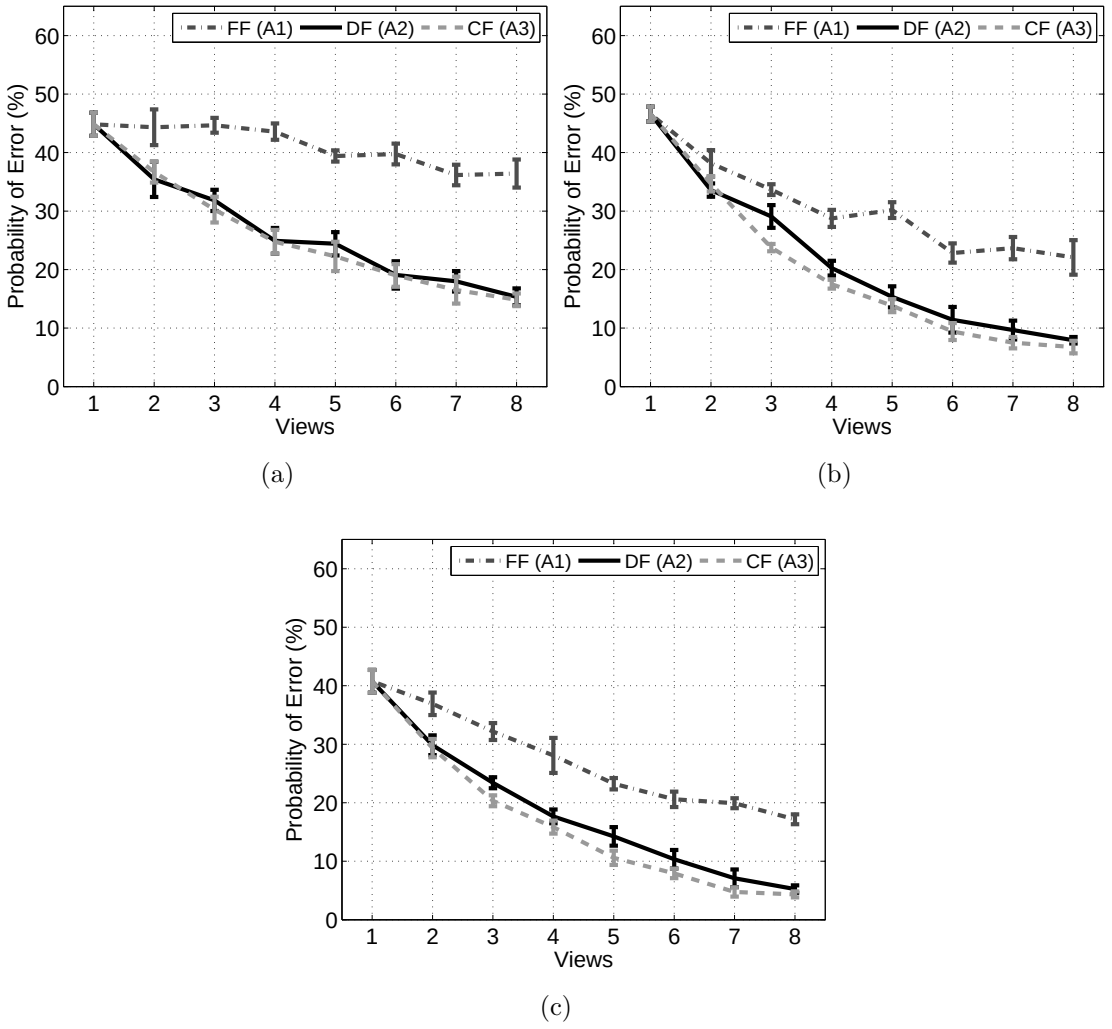
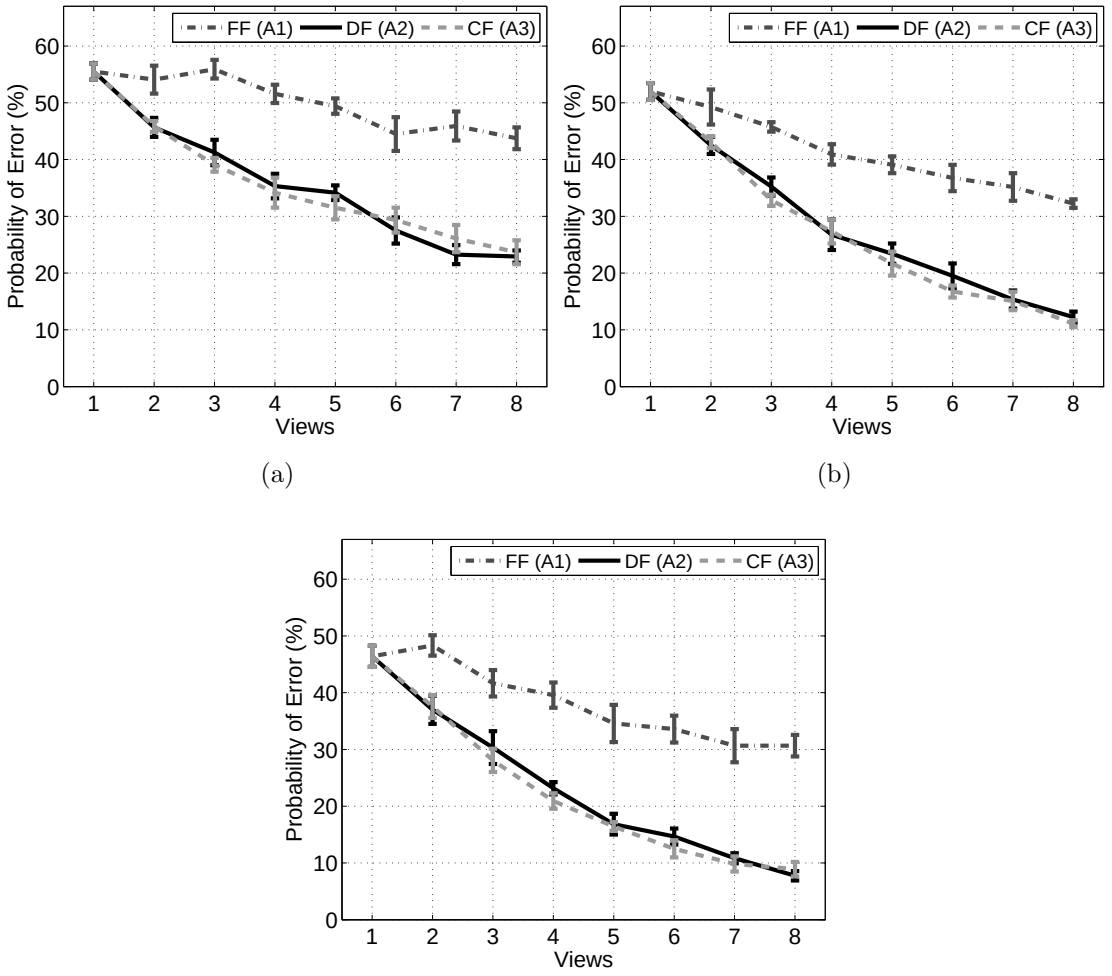


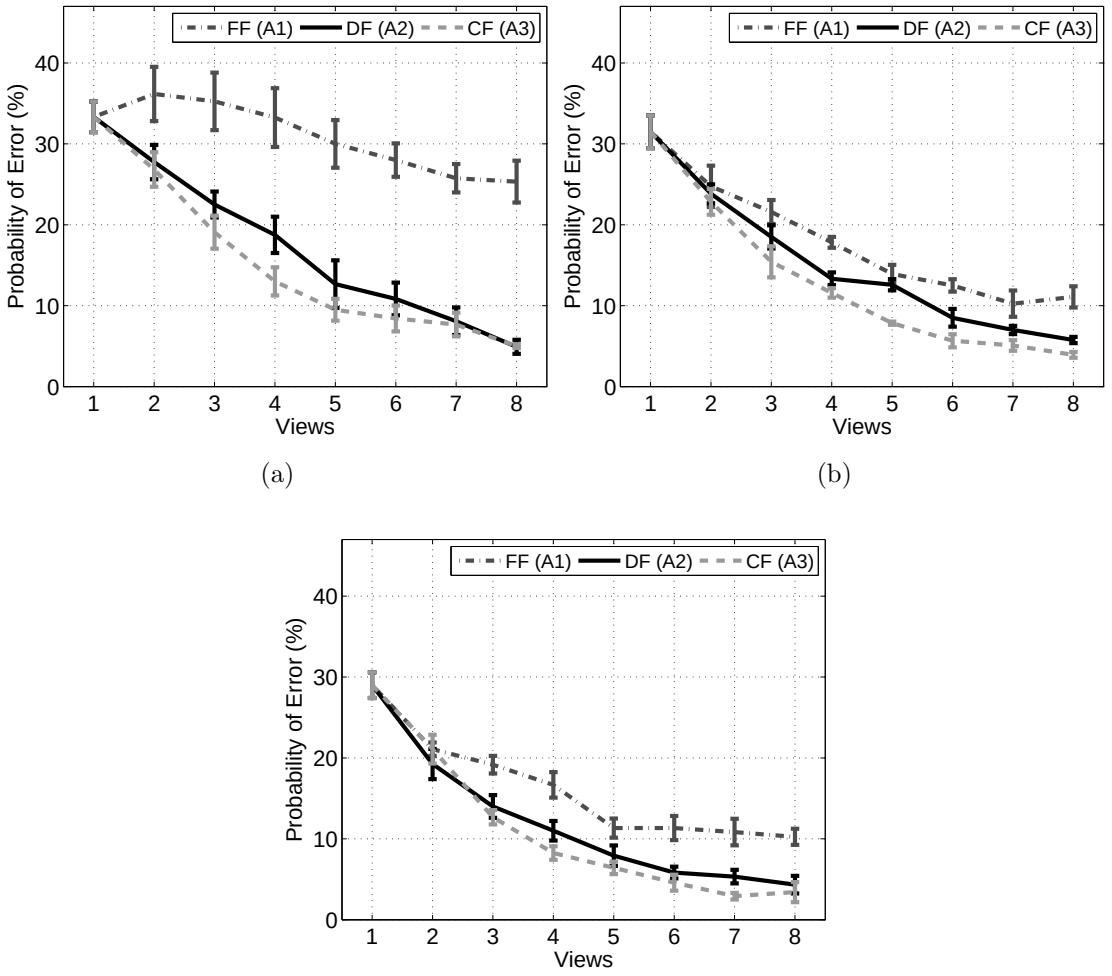
Figure D.4: Probability of error on D_5 vs. the number of random views used for classification, all three classification algorithms, and the DCT (a), db4 (b) and Haar (c) feature spaces.



(a) (b)

(c)

Figure D.5: Probability of error on D_6 vs. the number of random views used for classification, all three classification algorithms, and the DCT (a), db4 (b) and Haar (c) feature spaces.



(a) (b)

(c)

Figure D.6: Probability of error on D_7 vs. the number of random views used for classification, all three classification algorithms, and the DCT (a), db4 (b) and Haar (c) feature spaces.

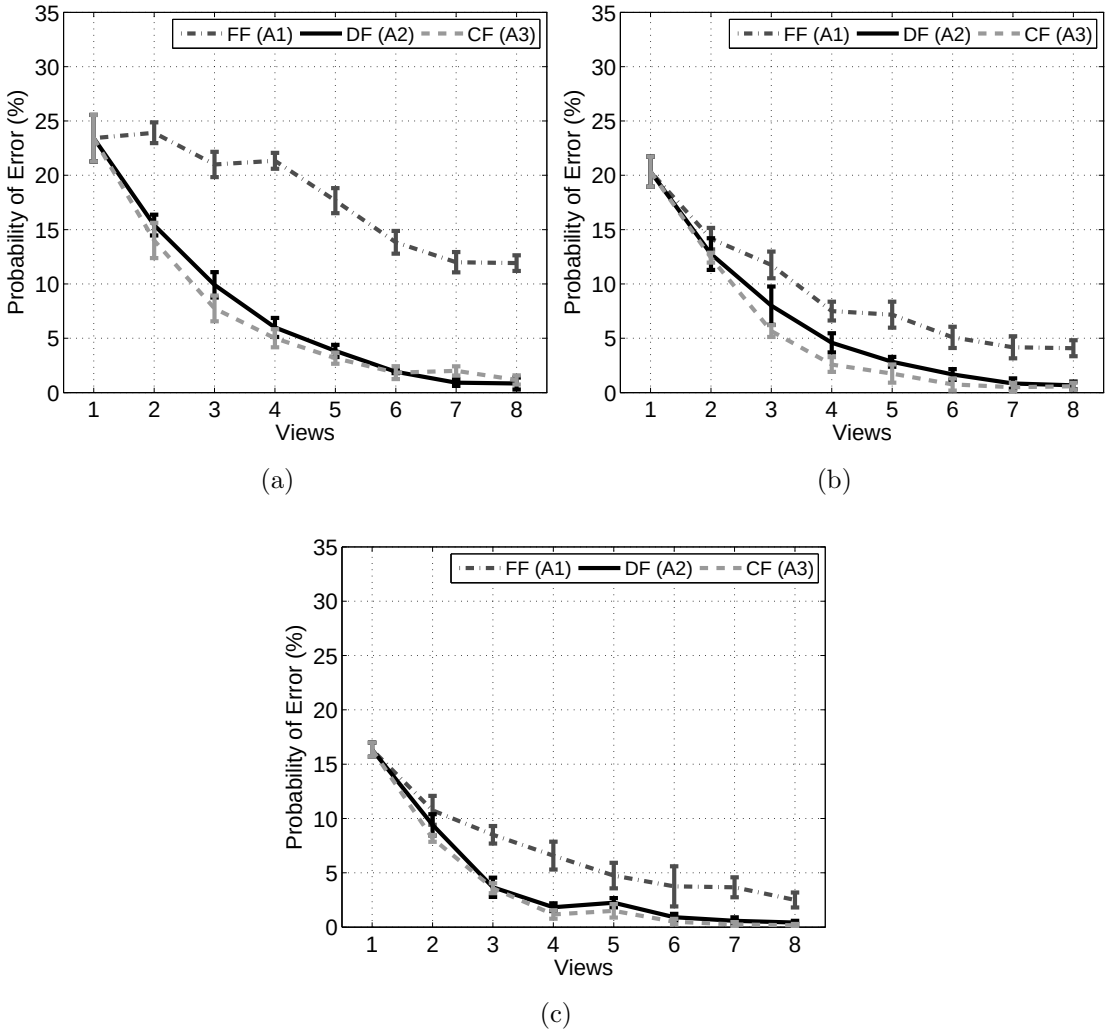


Figure D.7: Probability of error on D_9 vs. the number of random views used for classification, all three classification algorithms, and the DCT (a), db4 (b) and Haar (c) feature spaces.

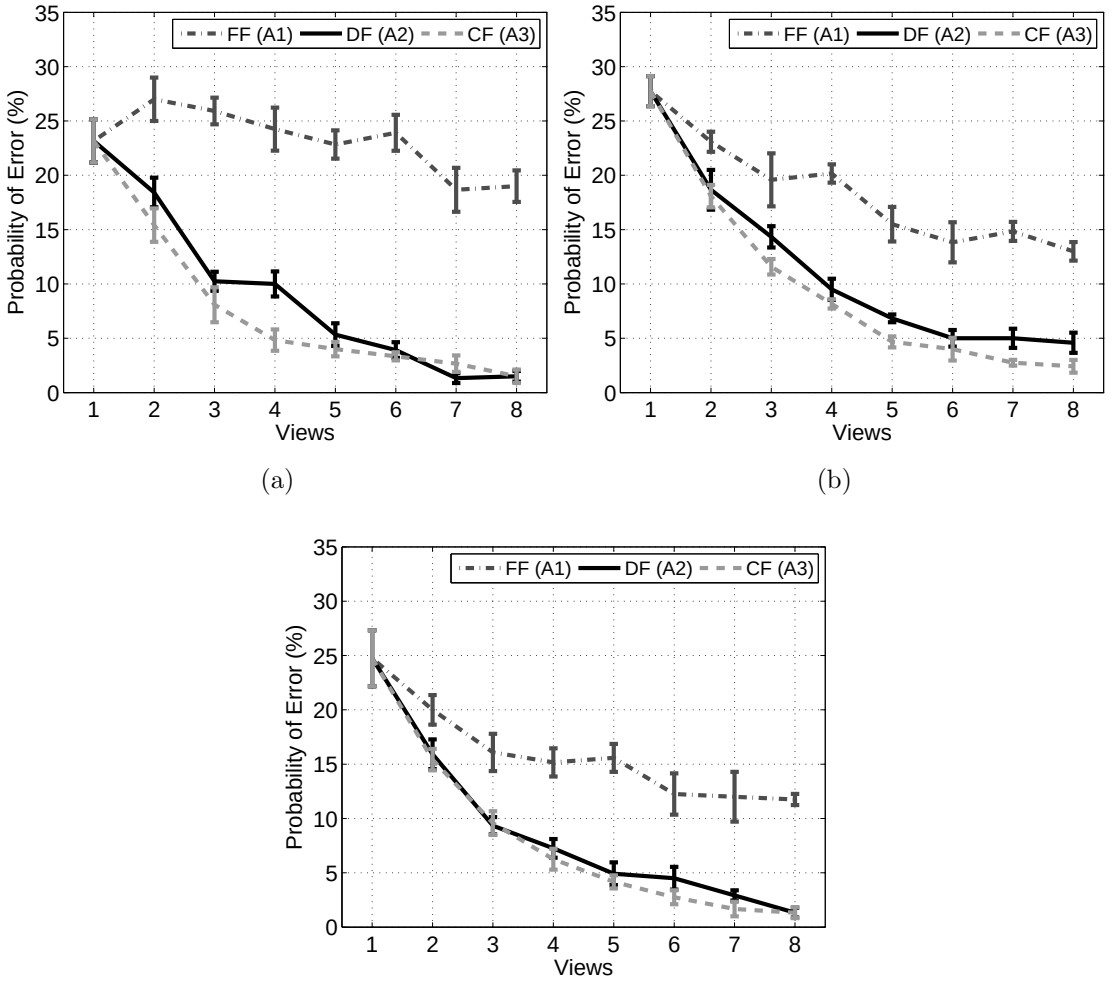


Figure D.8: Probability of error on D_{10} vs. the number of random views used for classification, all three classification algorithms, and the DCT (a), db4 (b) and Haar (c) feature spaces.

Bibliography

[1] S. A. Levin and J. Lubchenco, "Resilience, robustness, and marine ecosystem-based management," *BioScience*, vol. 58, no. 1, pp. 27–32, 2008. [Online]. Available: http://www.bioone.org/doi/abs/10.1641/B580107

[2] D. Pauly, V. Christensen, J. Dalsgaard, R. Froese, and F. Torres, "Fishing down marine food webs," *Science*, vol. 279, no. 5352, pp. 860–863, Feb. 1998.

[3] D. Pauly and M. L. Palomares, "Fishing down marine food web: It is far more pervasive than we thought," *Bulletin Of Marine Science*, vol. 76, no. 2, pp. 197–211, Mar. 2005.

[4] D. Pauly, V. Christensen, S. Guenette, T. J. Pitcher, U. R. Sumaila, C. J. Walters, R. Watson, and D. Zeller, "Towards sustainability in world fisheries," *Nature*, vol. 418, no. 6898, pp. 689–695, Aug. 2002. [Online]. Available: http://dx.doi.org/10.1038/nature01017

[5] L. B. Crowder, E. L. Hazen, N. Avissar, R. Bjorkland, C. Latanich, and M. B. Ogburn, "The impacts of fisheries on marine ecosystems and the transition to ecosystem-based management," *Annual Review Of Ecology Evolution And Systematics*, vol. 39, pp. 259–278, 2008.

[6] A. J. Richardson and D. S. Schoeman, "Climate impact on plankton ecosystems in the northeast atlantic," *Science*, vol. 305, no. 5690, pp. 1609–1612, Sep. 2004.

[7] D. Roemmich and J. McGowan, "Climatic warming and the decline of zooplankton in the california current," *Science*, vol. 207, no. 5202, pp. 1324–1326, 1995.

[8] G. E. Hofmann and S. D. Gaines, "New tools to meet new challenges: Emerging technologies for managing marine ecosystems for resilience," *Bioscience*, vol. 58, no. 1, pp. 43–52, Jan. 2008.

[9] P. Wiebe and M. Benfield, "From the hensen net twoard four-dimensional biological oceanography," *Prog. Oceanogr.*, vol. 1, pp. 130–138, Jan 2003.

[10] O. R. Godo, "What can technology offer the future fisheries scientist: Possibilities for obtaining better estimates of stock abundance by direct observations," *Journal of Northwest Atlantic Fishery Science*, no. 23, pp. 105–131, 1998.

[11] P. H. Wiebe, K. H. Burt, S. H. Boyd, and A. W. Morton, "Multiple opening-closing net and environmental sensing system for sampling zooplankton," *Journal Of Marine Research*, vol. 34, no. 3, pp. 313–326, 1976.

[12] M. C. Benfield, C. S. Davis, P. H. Wiebe, S. M. Gallager, R. G. Lough, and N. J. Copley, "Video plankton recorder estimates of copepod, pteropod and larvacean distributions from a stratified region of georges bank with comparative measurements from a mocness sampler," *Deep-Sea Research Part Ii-Topical Studies In Oceanography*, vol. 43, no. 7-8, pp. 1925–1945, 1996.

[13] J. Jaffe, M. Ohman, and A. Derobertis, "OASIS in the sea: measurement of the acoustic reflectivity of zooplankton with concurrent optical imaging," *Deep Sea Research Part II: Topical Studies in Oceanography*, vol. 45, pp. 1239–1253, 1998.

[14] E. A. Broughton and R. G. Lough, "A direct comparison of mocness and video plankton recorder zooplankton abundance estimates: Possible applications for augmenting net sampling with video systems." *Deep-Sea Research Part II Topical Studies in Oceanography*, vol. 53, no. 23-24, pp. 2789–2807, 2006.

[15] R. E. Thresher and J. S. Gunn, "Comparative analysis of visual census techniques for highly mobile, reef-associated piscivores (carangidae)," *Environmental Biology of Fishes*, vol. 17, no. 2, pp. 93–116, Oct. 1986.

[16] M. A. Samoilys and G. Carlos, "Determining methods of underwater visual census for estimating the abundance of coral reef fishes," *Environmental Biology Of Fishes*, vol. 57, no. 3, pp. 289–304, Mar. 2000.

[17] P. Carrera, J. H. Churnside, G. Boyra, V. Marques, C. Scalabrin, and A. Uriarte, "Comparison of airborne lidar with echosounders: a case study in the coastal atlantic waters of southern europe," *Ices Journal Of Marine Science*, vol. 63, no. 9, pp. 1736–1750, Nov. 2006.

[18] B. N. Tissot, W. W. Wakefield, M. A. Hixon, and J. E. R. Clemons, "Twenty years of fish habitat studies on heceta bank, oregon." *Alaska Sea Grant Report*, vol. 08-03, pp. 203–217, 2008.

[19] G. E. Rosenkranz, S. M. Gallager, R. W. Shepard, and M. Blakeslee, "Development of a high-speed, megapixel benthic imaging system for coastal

fisheries research in alaska," *Fisheries Research*, vol. 92, no. 2-3, pp. 340–344, Aug. 2008.

[20] C. S. Rose, A. W. Stoner, and K. Matteson, "Use of high-frequency imaging sonar to observe fish behaviour near baited fishing gears," *Fisheries Research*, vol. 76, no. 2, pp. 291–304, Nov. 2005.

[21] A. W. Herman, "Design and calibration of a new optical plankton counter capable of sizing small zooplankton," *Deep-Sea Research Part A-Oceanographic Research Papers*, vol. 39, no. 3-4A, pp. 395–415, Mar. 1992.

[22] W. M. Graham, D. L. Martin, and J. C. Martin, "In situ quantification and analysis of large jellyfish using a novel video profiler," *Marine Ecology-Progress Series*, vol. 254, pp. 129–140, 2003.

[23] C. S. Davis, F. T. Thwaites, S. M. Gallager, and Q. Hu, "A three-axis fast-tow digital video plankton recorder for rapid surveys of plankton taxa and hydrography," *Limnology And Oceanography-Methods*, vol. 3, pp. 59–74, Feb. 2005.

[24] D. M. Checkley, R. E. Davis, A. W. Herman, G. A. Jackson, B. Beanlands, and L. A. Regier, "Assessing plankton and other particles in situ with the solopc," *Limnology And Oceanography*, vol. 53, no. 5, pp. 2123–2136, Sep. 2008.

[25] D. V. Manov, G. C. Chang, and T. D. Dickey, "Methods for reducing biofouling of moored optical sensors," *Journal Of Atmospheric And Oceanic Technology*, vol. 21, no. 6, pp. 958–968, Jun. 2004.

[26] P. Fernandes, F. Gerlotto, D. Holliday, O. Nakken, and E. Simmonds, "Acoustic applications in fisheries science: the ices contribution," in *100 Years of Science under ICES: papers from a symposium held in Helsinki, 1-4 August 2000. ICES Marine Science Symposia.*, no. 215, 2002, pp. 483–492.

[27] K. G. Foote, "Linearity of fisheries acoustics, with addition theorems," *Journal Of The Acoustical Society Of America*, vol. 73, no. 6, pp. 1932–1940, 1983.

[28] D. Holliday, P. Donaghay, C. Greenlaw, M. D.E., M. McManus, S. J.M., and J. Miksis, "Advances in defining fine- and micro-scale pattern in marine plankton," *Aquatic Living Resources*, vol. 16 (3), pp. 131–136, 2003.

[29] M. Godlewska, M. Colon, L. Doroszczyk, B. Dlugoszewski, C. Verges, and J. Guillard, "Hydroacoustic measurements at two frequencies: 70 and 120 khz - consequences for fish stock estimation," *Fisheries Research*, vol. 96, no. 1, pp. 11–16, Feb. 2009.

[30] J. C. Coetzee, D. Merkle, C. L. de Moor, N. M. Twatwa, M. Barange, and D. S. Butterworth, "Refined estimates of south african pelagic fish biomass from hydro-acoustic surveys: quantifying the effects of target strength, signal attenuation and receiver saturation," *African Journal Of Marine Science*, vol. 30, no. 2, pp. 205–217, Sep. 2008.

[31] K. M. Boswell, M. P. Wilson, and J. H. Cowan, "A semiautomated approach to estimating fish size, abundance, and behavior from dual-frequency identification sonar (didson) data," *North American Journal Of Fisheries Management*, vol. 28, no. 3, pp. 799–807, Jun. 2008.

[32] V. M. Trenkel, V. Mazauric, and L. Berger, "The new fisheries multibeam echosounder me70: description and expected contribution to fisheries research," *Ices Journal Of Marine Science*, vol. 65, no. 4, pp. 645–655, May 2008.

[33] D. P. Gannon, "Passive acoustic techniques in fisheries science: A review and prospectus," *Transactions Of The American Fisheries Society*, vol. 137, no. 2, pp. 638–656, Mar. 2008.

[34] D. McNaught, "Acoustical determination of zooplankton distributions," in *Proc. 11th Conf. Great Lakes Res.*, 1968, pp. 76–84.

[35] D. Holliday, *Extracting biophysical information from the acoustic signatures of marine organisms.* Plenum Press, New York, 1977, ch. 10, pp. 619–625.

[36] D. Holliday, R. Pieper, and G. Kleppel, "Determination of zooplankton size and distribution with multi-frequency acoustic technology," *J. Cons. Int. Explor. Mer*, vol. 46, pp. 51–62, 1989.

[37] T. Stanton, D. Chu, P. Wiebe, L. Martin, and R. Eastwood, "Sound scattering by several zooplankton groups. I. Experimental determination of dominant scattering mechanisms," *J. Acoust. Soc. Am.*, vol. 103 (1), pp. 225–235, 1998.

[38] L. Martin, T. Stanton, P. Wiebe, and L. J.F., "Acoustic classification of zooplankton," *ICES J. Mar. Sci.*, vol. 53 (2), pp. 217–224, 1996.

[39] K. G. Foote, P. R. Atkins, D. T. I. Francis, and T. Knutsen, "Measuring echo spectra of marine organisms over a wide bandwidth," in *Proceedings of the International Conference on Underwater Acoustic Measurements: Technologies and Results*, 2005, pp. 501–508.

[40] G. A. Colombo, A. Benovic, A. Malej, D. Lucic, T. Makovec, V. Onofri, M. Acha, A. Madirolas, and H. Mianzan, "Acoustic survey of a jellyfish-dominated ecosystem (mljet island, croatia)," *Hydrobiologia*, vol. 616, pp. 99–111, Jan. 2009.

[41] P. Ayon, G. Swartzman, A. Bertrand, M. Gutierrez, and S. Bertrand, "Zooplankton and forage fish species off peru: Large-scale bottom-up forcing and local-scale depletion," *Progress In Oceanography*, vol. 79, no. 2-4, pp. 208–214, Oct. 2008.

[42] D. R. McKelvey and C. D. Wilson, "Discriminant classification of fish and zooplankton backscattering at 38 and 120 khz," *Transactions Of The American Fisheries Society*, vol. 135, no. 2, pp. 488–499, Mar. 2006.

[43] G. L. Lawson, P. H. Wiebe, T. K. Stanton, and C. J. Ashjian, "Euphausiid distribution along the western antarctic peninsula - part a: Development of robust multi-frequency acoustic techniques to identify euphausiid aggregations and quantify euphausild size, abundance, and biomass," *Deep-Sea Research Part Ii-Topical Studies In Oceanography*, vol. 55, no. 3-4, pp. 412–431, 2008.

[44] T. Stanton, D. Chu, and P. Wiebe, "Acoustic scattering characteristics of several zooplankton groups," *ICES J. Mar. Sci.*, vol. 53 (2), pp. 289–295, 1996.

[45] P. H. Wiebe, D. G. Mountain, T. K. Stanton, C. H. Greene, G. Lough, S. Kaartvedt, J. Dawson, and N. Copley, "Acoustical study of the spatial distribution of plankton on georges bank and the relationship between volume backscattering strength and the taxonomic composition of the plankton," *Deep-Sea Research Part Ii-Topical Studies In Oceanography*, vol. 43, no. 7-8, pp. 1971–&, 1996.

[46] J. D. Warren and P. H. Wiebe, "Accounting for biological and physical sources of acoustic backscatter improves estimates of zooplankton biomass," *Canadian Journal Of Fisheries And Aquatic Sciences*, vol. 65, no. 7, pp. 1321–1333, Jul. 2008.

[47] A. C. Lavery, P. H. Wiebe, T. K. Stanton, G. L. Lawson, M. C. Benfield, and N. Copley, "Determining dominant scatterers of sound in mixed zooplankton populations," *J. Acoust. Soc. Am.*, vol. 122, pp. 3304–3326., 2007.

[48] G. L. Lawson, P. H. Wiebe, C. J. Ashjian, D. Z. Chu, and T. K. Stanton, "Improved parameterization of antarctic krill target strength models," *Journal Of The Acoustical Society Of America*, vol. 119, no. 1, pp. 232–242, Jan. 2006.

[49] D. B. Reeder, J. M. Jech, and T. K. Stanton, "Broadband acoustic backscatter and high-resolution morphology of fish: Measurement and modeling," *The Journal of the Acoustical Society of America*, vol. 116, no. 4, pp. 2489–2489, 2004.

[50] B. Lundgren and J. R. Nielsen, "A method for the possible species discrimination of juvenile gadoids by broad-bandwidth backscattering spectra vs. angle of incidence," *ICES J. Mar. Sci.*, vol. 65, no. 4, pp. 581–593, May 2008.

[51] E. Rogers, G. Fleischer, P. Simpson, and G. Denny, "Broadband fish identification of laurentian great lakes fishes," *Geoscience and Remote Sensing Symposium, 2004. IGARSS '04. Proceedings. 2004 IEEE International*, vol. 2, pp. 1430–1434 vol.2, Sept. 2004.

[52] A. Aglen, *Sources of error in acoustic estimation of fish abundance.* In *Marine Fish Behaviour in Capture and Abundance Estimation*, A. Ferno and S. Olsen, Eds. Fishing News Books, Oxford, 1994.

[53] S. McClatchie, J. Alsop, Z. Ye, and R. F. Coombs, "Consequence of swimbladder model choice and fish orientation to target strength of three new zealand fish species," *ICES Journal of Marine Science*, vol. 53, pp. 847–862 (16), 1996.

[54] I. McQuinn and P. D. Winger, "Tilt angle and target strength: target tracking of atlantic cod (gadus morhua) during trawling," *ICES J. Mar. Sci.*, vol. 60 (3), pp. 575–583, 2003.

[55] J. Warren, T. Stanton, D. McGehee, and D. Chu, "Effect of animal orientation on acoustic estimates of zooplankton properties," *IEEE J. Ocean. Eng.*, vol. 27 (1), pp. 130–138, 2002.

[56] L. Traykovski, T. Stanton, P. Wiebe, and J. Lynch, "Model-based covariance mean variance classification techniques: Algorithm development and application to the acoustic classification of zooplankton," *IEEE J. Ocean. Eng.*, vol. 23 (4), pp. 344–364, 1998.

[57] D. McGehee, R. O'Driscoll, and L. Traykovski, "Effects of orientation on acoustic scattering from Antarctic krill at 120 kHz," *Deep-Sea Res. Part II-Topical Studies in Oceanography*, vol. 45 (7), pp. 1273–1294, 1998.

[58] J. S. Jaffe, "Using multiple-angle scattered sound to size fish swim bladders," *ICES J. Mar. Sci.*, vol. 63, no. 8, pp. 1397–1404, 2006.

[59] P. L. D. Roberts and J. S. Jaffe, "Multiple angle acoustic classification of zooplankton," *J. Acoust. Soc. Am.*, vol. 121, pp. 2060–2070, 2007.

[60] P. L. D. Roberts and J. S. Jaffe, "Classification of live, untethered zooplankton from observations of multiple-angle acoustic scatter," *Acoustical Society of America Journal*, vol. 124, pp. 796–802, 2008.

[61] B. Buelens, T. Pauly, R. Williams, and A. Sale, "Kernel methods for the detection and classification of fish schools in single-beam and multibeam acoustic data," *ICES J. Mar. Sci.*, p. fsp004, 2009.

[62] A. G. Cabreira, M. Tripode, and A. Madirolas, "Artificial neural networks for fish-species identification," *ICES J. Mar. Sci.*, p. fsp009, 2009.

[63] M. Azimi-Sadjadi, D. Yao, Q. Huang, and G. Dobeck, "Underwater target classification using wavelet packets and neural networks," *IEEE Trans. Neural Networks*, vol. 11 (3), pp. 784–794, MAY 2000.

[64] N. Dasgupta, P. Runkle, L. Couchman, and L. Carin, "Dual hidden Markov model for characterizing wavelet coefficients from multi-aspect scattering data," *Signal Processing*, vol. 81 (6), pp. 1303–1316, JUN 2001.

[65] P. Bharadwaj, P. Runkle, L. Carin, J. Berrie, and J. Hughes, "Multiaspect classification of airborne targets via physics-based HMMs and matching pursuits," *IEEE Trans. Aero. and Elec. Sys.*, vol. 37 (2), pp. 595–606, APR 2001.

[66] P. Runkle, L. Nguyen, J. McClellan, and L. Carin, "Multi-aspect target detection for SAR imagery using hidden Markov models," *IEEE Trans. Geosci. Remote Sens.*, vol. 39 (1), pp. 46–55, JAN 2001.

[67] Z. X. Li, S. Papson, and R. M. Narayanan, "Data-level fusion of multilook inverse synthetic aperture radar images," *IEEE T. Geosci. Remote.*, vol. 46, no. 5, pp. 1394–1406, May 2008.

[68] N. Dasgupta, P. Runkle, L. Carin, L. Couchman, T. Yoder, J. Bucaro, and G. Dobeck, "Class-based target identification with multiaspect scattering data," *IEEE J. Ocean. Eng.*, vol. 28 (2), pp. 271–282, 2003.

[69] H. Liu, N. Dasgupta, and L. Carin, "Time-reversal imaging for wideband underwater target classification," *Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03). 2003 IEEE International Conference on*, vol. 5, pp. V–5–8 vol.5, April 2003.

[70] M. Robinson, M. Azimi-Sadjadi, and J. Salazar, "Multi-aspect target discrimination using hidden Markov models and neural networks," *IEEE Trans. Neural Networks*, vol. 16 (2), pp. 447–459, MAR 2005.

[71] S. Ji, X. Liao, and L. Carin, "Adaptive multiaspect target classification and detection with hidden Markov models," *IEEE Sens. J.*, vol. 5 (5), pp. 1035–1042, OCT 2005.

[72] P. Runkle, P. Bharadwaj, L. Couchman, and C. L., "Hidden Markov models for multiaspect target classification," *IEEE Trans. Signal Proc.*, vol. 47 (7), pp. 2035–2040, JUL 1999.

[73] P. Runkle, L. Carin, L. Couchman, J. Bucaro, and T. Yoder, "Multiaspect identification of submerged elastic targets via wave-based matching pursuits and hidden Markov models," *J. Acoust. Soc. Am.*, vol. 106 (2), pp. 605–616, AUG 1999.

[74] K. Ni, Y. Qi, and L. Carin, "Multi-aspect target classification and detection via the infinite hidden markov model," *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, vol. 2, pp. II–433–II–436, April 2007.

[75] J. Cartmill, M. Azimi-Sadjadi, and N. Wachowski, "Buried underwater object classification using a collaborative multi-aspect classifier," *Neural Networks, 2007. IJCNN 2007. International Joint Conference on*, pp. 1807–1812, Aug. 2007.

[76] C. Greenlaw, "Acoustic estimation of zooplankton populations," *Limnol. Oceanogr.*, vol. 24, pp. 226–242, 1979.

[77] D. McGehee, D. Demer, and J. Warren, "Zooplankton in the Ligurian Sea: Part I. Characterization of their dispersion, relative abundance and environment during summer 1999," *J. Plankton Res.*, vol. 26 (12), pp. 1409–1418, 2004.

[78] M. McManus, O. Cheriton, P. Drake, D. Holliday, C. Storlazzi, D. P.L., and C. Greenlaw, "Effects of physical processes on structure and transport of thin zooplankton layers in the coastal ocean," *Mar. Ecol. Prog. Ser.*, vol. 301, pp. 199–215, 2005.

[79] L. Goodman, "Acoustic scattering from ocean microstructure," *J. Geophys. Res., C*, vol. 95, pp. 11 557–11 573, Jul. 1990.

[80] K. G. Foote, "Rather high frequency sound scattering by swimbladdered fish," *J. Acoust. Soc. Am.*, vol. 78, no. 2, pp. 688–700, 1985.

[81] T. Stanton, D. Chu, and P. Wiebe, "Sound scattering by several zooplankton groups. II. Scattering models," *J. Acoust. Soc. Am.*, vol. 103 (1), pp. 236–253, 1998.

[82] T. Stanton and D. Chu, "Review and recommendations for the modelling of acoustic scattering by fluid-like elongated zooplankton: euphausiids and copepods," *ICES J. Mar. Sci.*, vol. 57 (4), pp. 793–807, 2000.

[83] D. Reeder and T. Stanton, "Acoustic scattering by axisymmetric finite-length bodies: An extension of a two-dimensional conformal mapping method," *J. Acoust. Soc. Am.*, vol. 116 (2), pp. 729–746, 2004.

[84] A. Lavery, T. Stanton, D. McGehee, and C. DZ, "Three-dimensional modeling of acoustic backscattering from fluid-like zooplankton," *J. Acoust. Soc. Am.*, vol. 111 (3), pp. 1197–1210, 2002.

[85] D. E. McGehee, M. Benfield, D. V. Holliday, and C. Greenlaw, "Advanced multifrequency inversion methods for classifying acoustic scatterers." [Online]. Available: http://zooplankton.lsu.edu/scattering_models/MultifreqInverseMethods.html

[86] B. Pei and M. Bao, "Multi-aspect radar target recognition method. Based on scattering centers and HMMs," *IEEE Trans. Aero. and Elec. Sys.*, vol. 41 (3), pp. 1067–1074, JUL 2005.

[87] Y. Dong, P. Runkle, L. Carin, R. Damarla, A. Sullivan, M. Ressler, and J. Sichina, "Multi-aspect detection of surface and shallow-buried unexploded ordnance via ultra-wideband synthetic aperture radar," *IEEE Trans. Geosci. Remote Sens.*, vol. 39 (6), pp. 1259–1270, JUN 2001.

[88] D. Li, M. Azimi-Sadjadi, and M. Robinson, "Comparison of different classification algorithms for underwater target discrimination," *IEEE Trans. Neural Networks*, vol. 15 (1), pp. 189–194, JAN 2004.

[89] A. Fleminger, J. D. Isaacs, and J. G. Wyllie, "Zooplankton biomass measurements from calcofi cruises of july 1955 to 1959 and remarks on comparison with results from october, january, and april cruises of 1955 to 1959," CalCOFI Atlas No. 21, Tech. Rep., 1974.

[90] P. Morse and K. Ingard, *Theoretical Acoustics*. Princeton University Press, 1968.

[91] J. Jaffe, E. Reuss, M. D., and G. Chandran, "Ftv: A sonar for tracking macrozooplankton in three dimensions," *Deep-Sea Res. I*, vol. 42, no. 8, pp. 1495–1512, 1995.

[92] A. Genin, J. S. Jaffe, R. Reef, C. Richter, and P. J. S. Franks, "Swimming against the flow: A mechnaism of zooplankton aggregation," *Science*, vol. 308, no. 5723, pp. 860–862, 2005.

[93] S. M. Kay, *Fundementals of Statistical Signal Processing: Detection Teory*. Prentice Hall, Upper Saddle River, New Jersey 07458, 1998, vol. 1.

[94] T. K. Moon and W. C. Stirling, *Mathematical methods and algorithms for signal processing*. Prentice Hall, 2000.

[95] R. Duda, P. Hart, and D. G. Stork, *Pattern Classification*. Wiley Interscience, 2000.

[96] C. M. Bishop, "Neural networks and their applications," *Review of Scientific Instruments*, vol. 65, no. 6, pp. 1803–1932, Jun 1994.

[97] I. Nabney, *Netlab: Algorithms for Pattern Recognition.* Springer, 2001.

[98] M. D. Ohman and B. E. Lavaniegos, "Comparative zooplankton sampling efficiency of a ring net and bongo net with comments on pooling of subsamples," CalCOFI Rep., Vol. 43, Tech. Rep., 2002.

[99] E. Brinton and J. G. Wyllie, "Distributional atlas of euphausiid growth stages off southern california, 1953-1956," CalCOFI Atlas, No. 24, Tech. Rep., 1976.

[100] P. H. Wiebe, C. H. Greene, T. K. Stanton, and J. Burczynski, "Sound scattering by live zooplankton and micronekton - empirical-studies with a dual-beam acoustical system," *Journal Of The Acoustical Society Of America*, vol. 88, no. 5, pp. 2346–2360, Nov. 1990.

[101] A. De Robertis, J. S. Jaffe, and M. D. Ohman, "Size-dependent visual predation risk and the timing of vertical migration in zooplankton limnol. oceanogr. 45, 838-1844." *Limnol. Oceanogr.*, vol. 45, pp. 838–1844, 2000.

[102] K. G. Foote, "Spheres for calibrating an 11-frequency acoustic measurement system," *Journal Du Conseil*, vol. 46, no. 3, pp. 284–286, 1990.

[103] J. G. Minonzio, C. Prada, D. Chambers, D. Clorennec, and M. Fink, "Characterization of subwavelength elastic cylinders with the decomposition of the time-reversal operator: Theory and experiment," *Journal Of The Acoustical Society Of America*, vol. 117, no. 2, pp. 789–798, Feb. 2005.

[104] D. Chu and T. Stanton, "Application of pulse compression techniques to broadband acoustic scattering by live individual zooplankton," *J. Acoust. Soc. Am.*, vol. 104 (1), pp. 39–55, 1998.

[105] L. Traykovski, R. O'Driscoll, and D. McGehee, "Effect of orientation on broadband acoustic scattering of Antarctic krill Euphausia superba: Implications for inverting zooplankton spectral acoustic signatures for angle of orientation," *J. Acoust. Soc. Am.*, vol. 104 (4), pp. 2121–2135, OCT 1998.

[106] O. A. Misund, "Underwater acoustics in marine fisheries and fisheries research," *Reviews in Fish Biology and Fisheries*, vol. 7, no. 1, pp. 1–34, Mar. 1997.

[107] J. Simmonds and D. MacLennan, *Fisheries Acoustics: Theory and Practice (Fish and Aquatic Resources).* Wiley-Blackwell; 2 edition, 2006.

[108] D. L. Burwen, P. A. Nealson, S. J. Fleischman, T. J. Mulligan, and J. K. Horne, "The complexity of narrowband echo envelopes as a function of fish side-aspect angle," *ICES J. Mar. Sci.*, vol. 64, no. 5, pp. 1066–1074, Jul. 2007.

[109] K. G. Foote, "Effect of fish behavior on echo energy - the need for measurements of orientation distributions," *Journal Du Conseil*, vol. 39, no. 2, pp. 193–201, 1980.

[110] G. L. Lawson, M. Barange, and P. Freon, "Species identification of pelagic fish schools on the South African continental shelf using acoustic descriptors and ancillary information," *ICES J. Mar. Sci.*, vol. 58, no. 1, pp. 275–287, 2001.

[111] R. J. Kloser, T. Ryan, P. Sakov, A. Williams, and J. Koslow, "Species identification in deep water using multiple acoustic frequencies," *Canadian Journal of Fisheries and Aquatic Sciences*, vol. 59, pp. 1065–1077(13), June 2002.

[112] B. Pei and Z. Bao, "Multi-aspect radar target recognition method based on scattering centers and hmms classifiers," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 41, no. 3, pp. 1067–1074, July 2005.

[113] N. Ramani and P. H. Patrick, "Fish detection and identification using neural networks - some laboratory results," *IEEE J. Oceanic. Eng.*, vol. 17, no. 4, pp. 364–368, Oct. 1992.

[114] E. J. Simmonds, F. Armstrong, and P. J. Copland, "Species identification using wideband backscatter with neural network and discriminant analysis," *ICES J. Mar. Sci.*, vol. 53, no. 2, pp. 189–195, Apr. 1996.

[115] J. Ghosh, L. M. Deuser, and S. D. Beck, "A neural network based hybrid system for detection, characterization, and classification of short-duration oceanic signals," *IEEE J. Oceanic. Eng.*, vol. 17, no. 4, pp. 351–363, Oct. 1992.

[116] R. Gorman and T. Sejnowski, "Learned classification of sonar targets using a massively parallel network," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 36, no. 7, pp. 1135–1140, Jul 1988.

[117] C. I. H. Anderson, J. K. Horne, and J. Boyle, "Classifying multi-frequency fisheries acoustic data using a robust probabilistic classification technique," *J. Acoust. Soc. Am.*, vol. 121, no. 6, pp. EL230–EL237, Jun. 2007.

[118] R. A. Malkin and D. Alexandrou, "Acoustic classification of abyssopelagic animals," *IEEE J. Ocean. Eng.*, vol. 18, no. 1, pp. 63–72, Jan. 1993.

[119] B. Ayrulu and B. Barshan, "Reliability measure assignment to sonar for robust target differentiation," *Pattern Recogn.*, vol. 35, no. 6, pp. 1403–1419, Jun. 2002.

[120] J. Cartmill, N. Wachowski, and M. R. Azimi-Sadjadi, "Buried underwater object classification using a collaborative multiaspect classifier," *Oceanic Engineering, IEEE Journal of*, vol. 34, no. 1, pp. 32–44, Jan. 2009.

[121] D. Chu, J. Michaeljech, and A. Lavery, "Inference of geometrical and behavioural parameters of individual fish from echo-trace-analysis," *Deep Sea Research Part I: Oceanographic Research*, vol. 50, pp. 515–527, Apr. 2003.

[122] L. M. Deuser and D. Middleton, "Classification of underwater acoustic-signals .1. environmentally adaptive approach," *J. Acoust. Soc. Am.*, vol. 65, no. 2, pp. 438–443, 1979.

[123] L. M. Deuser, D. Middleton, T. D. Plemons, and J. K. Vaughan, "Classification of underwater acoustic-signals .2. experimental applications involving fish," *J. Acoust. Soc. Am.*, vol. 65, no. 2, pp. 444–455, 1979.

[124] S. J. Fleischman and D. L. Burwen, "Mixture models for the species apportionment of hydroacoustic data, with echo-envelope length as the discriminatory variable," *ICES J. Mar. Sci.*, vol. 60, no. 3, pp. 592–598, Jun. 2003.

[125] S. G. Conti and D. A. Demer, "Wide-bandwidth acoustical characterization of anchovy and sardine from reverberation measurements in an echoic tank," *ICES J. Mar. Sci.*, vol. 60, no. 3, pp. 617–624, Jun. 2003.

[126] M. E. Zakharia, F. Magand, F. Hetroit, and N. Diner, "Wideband sounder for fish species identification at sea," *ICES J. Mar. Sci.*, vol. 53, no. 2, pp. 203–208, Apr. 1996.

[127] C. H. Chen, J. D. Lee, and M. C. Lin, "Classification of underwater signals using wavelet transforms and neural networks," *Math. Comput. Model.*, vol. 27, no. 2, pp. 47–60, Jan. 1998.

[128] A. Pezeshki, M. Azimi-Sadjadi, L. Scharf, and M. Robinson, "A canonical correlation-based feature extraction method for underwater target classification," *Oceans '02 MTS/IEEE*, vol. 1, pp. 29–37 vol.1, Oct. 2002.

[129] A. Pezeshki, M. Azimi-Sadjadi, and L. Scharf, "Undersea target classification using canonical correlation analysis," *Oceanic Engineering, IEEE Journal of*, vol. 32, no. 4, pp. 948–955, Oct. 2007.

[130] M. Azimi-Sadjadi, D. Yao, A. Jamshidi, and G. Dobeck, "Underwater target classification in changing environments using an adaptive feature mapping," *IEEE Trans. Neural Networks*, vol. 13 (5), pp. 1099–1111, SEP 2002.

[131] B. Erkmen and T. Yildirim, "Improving classification performance of sonar targets by applying general regression neural network with pca," *Expert Systems with Applications*, vol. 35, no. 1-2, pp. 472–475, 2008.

[132] C. F. Gaumond, "Echo components for aspect-independent detection and classification," *IEEE J. Oceanic. Eng.*, vol. 24, no. 4, pp. 436–446, Oct. 1999.

[133] A. Kundu, G. C. Chen, and C. E. Persons, "Transient sonar signal classification using hidden markov-models and neural nets," *IEEE J. Oceanic. Eng.*, vol. 19, no. 1, pp. 87–99, Jan. 1994.

[134] F. B. Shin and D. H. Kil, "Full-spectrum signal processing using a classify-before-detect paradigm," *J. Acoust. Soc. Am.*, vol. 99, no. 4, pp. 2188–2197, Apr. 1996.

[135] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*, 2nd ed. Prentice-Hall, 1998.

[136] M. Piccardi, "Background subtraction techniques: a review," vol. 4, Oct. 2004, pp. 3099–3104 vol.4.

[137] R. C. Gonzales and R. E. Woods, *Digital Image Processing.* Prentice Hall, 2002.

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

[139] T. Joachims, "Making large-scale support vector machine learning practical," pp. 169–184, 1999.

[140] "The spider toolbox for matlab," 2006, version 1.71. [Online]. Available: http://www.kyb.mpg.de/bs/people/spider/main.html

[141] J. Zhu, S. Rosset, H. Zou, and T. Hastie, "Multiclass adaboost," Tech. Rep., 2005.

[142] R. E. Schapire, "The boosting approach to machine learning: An overview," in *MSRI Workshop on Nonlinear Estimation and Classification*, 2002.

[143] R. Collobert and S. Bengio, "Links between perceptrons, mlps and svms," in *In: Proceedings of ICML. (2004.* ACM Press, 2004.

[144] M. E. Tipping, "Sparse bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, 2001.

[145] B. Morris and M. Trivedi, "Learning, modeling, and classification of vehicle track patterns from live video," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 9, no. 3, pp. 425–437, Sept. 2008.

[146] C. Nadeau and Y. Bengio, "Inference for the generalization error," *Mach. Learn.*, vol. 52, no. 3, pp. 239–281, 2003.

[147] M. Markatou, H. Tian, S. Biswas, and G. Hripcsak, "Analysis of variance of cross-validation estimators of the generalization error," *J. Mach. Learn. Res.*, vol. 6, pp. 1127–1168, 2005.

[148] D. Knepp, "Antenna aperture effects on measurements of propagation through turbulence," *Antennas and Propagation, IEEE Transactions on*, vol. 23, no. 5, pp. 682–687, Sep 1975.

[149] J. H. Churnside, R. J. Hill, G. Conforti, and A. Consortini, "Aperture size and bandwidth requirements for measuring strong scintillation in the atmosphere," *Appl. Opt.*, vol. 28, no. 19, pp. 4126–4132, 1989. [Online]. Available: http://ao.osa.org/abstract.cfm?URI=ao-28-19-4126

[150] J. Oeschger and L. Goodman, "Acoustic scattering from a thermally driven buoyant plume," *The Journal of the Acoustical Society of America*, vol. 100, no. 3, pp. 1451–1462, 1996.

[151] H. Lee, "Resolution Trade-Off Analysis for Aperture Size and Signaling Bandwidth of Diffraction Tomography Based on Spatial-Frequency Spectral Coverage," *INTERNATIONAL JOURNAL OF IMAGING SYSTEMS AND TECHNOLOGY*, vol. 19, no. 1, pp. 1–4, 2009.

[152] D. Yao, M. Azimi-Sadjadi, A. Jamshidi, and G. Dobeck, "A study of effects of sonar bandwidth for underwater target classification," *IEEE J. Ocean. Eng.*, vol. 27 (3), pp. 619–627, JUL 2002.

[153] T. K. Stanton, D. B. Reeder, and J. M. Jech, "Inferring fish orientation from broadband-acoustic echoes," *Ices Journal Of Marine Science*, vol. 60, no. 3, pp. 524–531, Jun. 2003.

[154] D. Wipf and B. Rao, "Sparse bayesian learning for basis selection," *Signal Processing, IEEE Transactions on*, vol. 52, no. 8, pp. 2153–2164, Aug. 2004.

[155] S. R. Deans, *The Radon Transform and Some of Its Applications.* Krieger Publishing Company, 1983.

[156] D. Z. Chu and Z. Ye, "A phase-compensated distorted wave born approximation representation of the bistatic scattering by weakly scattering objects: Application to zooplankton," *J. Acoust. Soc. Am.*, vol. 106, no. 4, pp. 1732–1743, Oct. 1999.

[157] C. C. Ferguson, "Intersections of ellipsoids and planes of arbitrary orientation and position," *Mathematical Geology*, vol. 11, no. 3, pp. 329–336, Jun. 1979.

[158] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing.* New York: Dover, 1972, ch. 10, pp. 437–442.

[159] R. M. Mersereau and A. V. Oppenheim, "Digital reconstruction of multidimensional signals from their projections," *Proc. IEEE*, vol. 62, pp. 1319–1338, 1974.