Robust Methods for the Analysis of Images and Videos for Fisheries Stock Assessment: Summary of a Workshop

ISBN
978-0-309-31469-5

88 pages
7 x 10
PAPERBACK (2014)

Maureen Mellody, Rapporteur; Committee on Applied and Theoretical Statistics; Board on Mathematical Sciences and Their Applications; Division on Engineering and Physical Sciences; National Research Council

ROBUST METHODS FOR THE ANALYSIS OF Images and Videos for Fisheries Stock Assessment

SUMMARY OF A WORKSHOP

Maureen Mellody, *Rapporteur*

Committee on Applied and Theoretical Statistics

Board on Mathematical Sciences and Their Applications

Division on Engineering and Physical Sciences

NATIONAL RESEARCH COUNCIL
OF THE NATIONAL ACADEMIES

THE NATIONAL ACADEMIES PRESS
Washington, D.C.
www.nap.edu

THE NATIONAL ACADEMIES PRESS 500 Fifth Street, NW Washington, DC 20001

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine.

This study was supported by Contract No. WC 133R-11-CQ-0048 TO #6 between the National Academy of Sciences and the National Oceanic and Atmospheric Administration. Any opinions, findings, or conclusions expressed in this publication are those of the author(s) and do not necessarily reflect the views of the organizations or agencies that provided support for the project.

International Standard Book Number-13: 978-0-309-31469-5
International Standard Book Number-10: 0-309-31469-0

This report is available in limited quantities from

Board on Mathematical Sciences and Their Applications
500 Fifth Street NW
Washington, DC 20001
bmsa@nas.edu
http://www.nas.edu/bmsa

Additional copies of this report are available for sale from the National Academies Press, 500 Fifth Street NW, Keck 360, Washington, DC 20001; (800) 624-6242 or (202) 334-3313; http://www.nap.edu/.

Printed in the United States of America

THE NATIONAL ACADEMIES

Advisers to the Nation on Science, Engineering, and Medicine

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Ralph J. Cicerone is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. C. D. Mote, Jr., is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Victor J. Dzau is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Ralph J. Cicerone and Dr. C. D. Mote, Jr., are chair and vice chair, respectively, of the National Research Council.

www.national-academies.org

PLANNING COMMITTEE ON ROBUST METHODS FOR THE ANALYSIS OF IMAGES AND VIDEOS FOR FISHERIES STOCK ASSESSMENT: A WORKSHOP

RAMA CHELLAPPA, University of Maryland, College Park, *Chair*
RUZENA BAJCSY, University of California, Berkeley
LISE GETOOR, University of California, Santa Cruz
ALFRED HERO III, University of Michigan
ANTHONY HOOGS, Kitware, Inc.
DAVID KRIEGMAN, University of California, San Diego
RICHARD LEAHY, University of Southern California
FEI-FEI LI, Stanford University
GUILLERMO SAPIRO, Duke University
LANCE WALLER, Emory University

Staff

MICHELLE K. SCHWALBE, Program Officer
RODNEY N. HOWARD, Administrative Assistant

COMMITTEE ON APPLIED AND THEORETICAL STATISTICS

CONSTANTINE GATSONIS, Brown University, *Chair*
MONTSERRAT (MONTSE) FUENTES, North Carolina State University
ALFRED O. HERO III, University of Michigan
DAVID M. HIGDON, Los Alamos National Laboratory
IAIN JOHNSTONE, Stanford University
ROBERT E. KASS, Carnegie Mellon University
JOHN LAFFERTY, University of Chicago
XIHONG LIN, Harvard University
SHARON-LISE T. NORMAND, Harvard University
GIOVANNI PARMIGIANI, Harvard University
RAGHU RAMAKRISHNAN, Microsoft
ERNEST SEGLIE, Office of the Secretary of Defense (retired)
LANCE WALLER, Emory University
EUGENE WONG, University of California, Berkeley

Staff

MICHELLE K. SCHWALBE, Director
RODNEY N. HOWARD, Administrative Assistant

Acknowledgment of Reviewers

This report has been reviewed in draft form by individuals chosen for their diverse perspectives and technical expertise, in accordance with procedures approved by the National Research Council's Report Review Committee. The purpose of this independent review is to provide candid and critical comments that will assist the institution in making its published report as sound as possible and to ensure that the report meets institutional standards for objectivity, evidence, and responsiveness to the study charge. The review comments and draft manuscript remain confidential to protect the integrity of the deliberative process. We wish to thank the following individuals for their review of this report:

Margrit Betke, Boston University,
Jitendra Malik, University of California, Berkeley,
Pietro Perona, California Institute of Technology, and
Hanumant Singh, Woods Hole Oceanographic Institution.

Although the reviewers listed above have provided many constructive comments and suggestions, they were not asked to endorse the views presented at the workshop, nor did they see the final draft of the workshop summary before its release. The review of this workshop summary was overseen by Andrew Solow, Woods Hole Oceanographic Institution. Appointed by the NRC, he was responsible for making certain that an independent examination of this workshop summary was carried out in accordance with institutional procedures and that all review comments were carefully considered. Responsibility for the final content of this summary rests entirely with the author and the institution.

Contents

1

Introduction

The National Marine Fisheries Service (NMFS; informally known as "NOAA Fisheries") of the National Oceanic and Atmospheric Administration (NOAA) is responsible for the stewardship of the nation's living marine resources and their habitat. As part of this charge, NOAA Fisheries conducts stock assessments of the abundance and composition of fish stocks in several bodies of water. The use of images and videos, when accompanied by appropriate statistical analyses of the inferred data, is of increasing importance for estimating the abundance of species and their age distributions. NOAA Fisheries is actively seeking to improve the quality and reliability of data from still and stereo-video imagery, and, more generally, to automate more of the stock assessment process. In particular, NOAA Fisheries is interested in identifying promising directions for advancing its analytical capabilities, including opportunities to leverage capabilities from other fields (such as the use of machine learning in pattern recognition).

The accuracy and efficiency of fisheries stock assessments are limited in large part by data collection tools and techniques. At present, stock assessments rely heavily on human data-gathering and analysis. Automatic means of fish stock assessments are appealing because they offer the potential to improve efficiency and reduce human workload and perhaps develop higher-fidelity measurements. However, automatic counting or characterization remains a complex and difficult task because of numerous factors: many species move about during observations, individuals often look very similar, some species blend in with their background, lighting can be variable, and the correlation between measurable features and desired features (such as age or gender) may be weak. These complexities are com-

pounded by data collection techniques that may involve trawling or the collection of images via a camera on a moving platform.

WORKSHOP OVERVIEW

A workshop was developed to enable experts from diverse communities to share perspectives about the most efficient path toward improved automation of visual information for fisheries stock assessments and to discuss both near-term (3 to 5 years) and long-term goals that can be achieved through modest research and development efforts. On May 16-17, 2014, the National Research Council's (NRC's) Committee on Applied and Theoretical Statistics convened a workshop to discuss analysis techniques for images and videos for fisheries stock assessment. To conduct the workshop, a planning committee was established to refine the workshop topics, identify speakers, and plan the workshop agenda. The workshop was held at the National Academy of Sciences building in Washington, D.C., and was sponsored by NOAA. Approximately 40 participants, including speakers, members of the parent committee, invited guests, and members of the public, participated in the 2-day workshop. The workshop was also webcast live, with approximately 10 people participating remotely via webcast. A complete statement of task is shown in Box 1.1.

This report has been prepared by the workshop rapporteur as a factual summary of what occurred at the workshop. The planning committee's role was limited to planning and convening the workshop. The views contained in the report are those of individual workshop participants and do not necessarily represent the views of all workshop participants, the planning committee, or the NRC.

In addition to the workshop summary provided here, materials related to the workshop can be found online at the website of the Board on Mathematical Sciences and their Applications (http://www.nas.edu/bmsa), including the agenda, speaker presentations, archived webcasts of the presentations and discussions, and other background materials.

ORGANIZATION OF THIS REPORT

Subsequent chapters of this report summarize the workshop presentations and discussion, following the organization of the workshop. Chapter 2 sets the stage for current practice and future needs in fisheries stock assessment. Chapter 3 focuses on multi-modal sensing. Chapter 4 discusses methods of image processing and detection. Chapter 5 focuses on multi-object tracking. Chapter 6 discusses shape and motion analysis. Chapter 7 describes methods of identification and classification, and Chapter 8 summarizes lessons learned and strategies moving forward

BOX 1.1
Statement of Task

An ad hoc committee will plan and conduct a public workshop that will examine the frontiers in methodology for examining image, video, and possibly other sensor data related to the following tasks of importance to the National Marine Fisheries Service (NMFS):

- Automatic counting or characterization of fish as they pass through a trawl against a semi-static background.
- Interpreting video (e.g., identifying the species, counting individuals, characterizing their size distribution) from a stationary camera that views fish against the bottom of a body of water.
- Automatic interpretation (counting and characterizing) of individual snapshot images taken from a remotely operated moving camera.
- Automatic counting and characterization of fish in videos against a natural background.

NMFS will provide the committee with information about its current capabilities for collecting and analyzing images and video for these tasks. Based on that input, the committee will organize a workshop that will feature invited presentations and discussions involving participants from diverse fields to address the following topics:

- Identify promising directions for advancing NMFS's analytical capabilities for the tasks listed above, including opportunities to leverage capabilities from other fields.
- Share perspectives about the most efficient path toward more automation of fisheries stock assessments, identifying goals that might be achieved through 3-5 years of modest R&D investment and goals that should be considered longer term.

One or more rapporteurs who are not members of the committee will be designated to prepare an individually authored or co-authored summary of the presentations and discussions at the event.

from the workshop. Finally, Appendix A lists the registered workshop participants, Appendix B shows the workshop agenda, and Appendix C defines acronyms used in this report.

2

Setting the Stage

The first session of the workshop set the stage by discussing current approaches and potential future options for fisheries stock assessments. Some useful references for the opening session, as suggested by the workshop planning committee, include the following: Armstrong et al., 2006; Beijbom et al., 2012; Cadima, 2003; Cappo et al., 2006; Chen et al., 2006; Clarke et al., 2009; Kimura and Somerton, 2006; Mace et al., 2001; Mallet and Pelletier, 2014; NOAA Fisheries, 2012; Sale, 1997; Shortis et al., 2013; Spampinato et al., 2008, 2010; Sparre and Venema, 1992; Western Pacific Regional Fishery Management Council, 2004; and Williams et al., 2010.

Rama Chellappa (University of Maryland, College Park; chair, workshop planning committee) and Ned Cyr (NOAA Fisheries) opened the workshop and introduced the speakers of the first session: Benjamin Richards (NOAA Fisheries), Allan Hicks (NOAA Fisheries), Ruzena Bajcsy (University of California, Berkeley), and Steven Thompson (Simon Fraser University). In addition, the summaries of two later keynote speakers are included in this chapter: Demetri Terzopoulos (University of California, Los Angeles) and Concetto Spampinato (Università di Catania, Italy) spoke about computer graphics simulations of groups of fish and ongoing large-scale underwater video collection and identification of reef fish, respectively.

TYPES OF DATA USED IN FISHERY STOCK ASSESSMENTS

Allan Hicks, NOAA Fisheries

Allan Hicks began by defining fishery stock assessment models as "demographic analyses designed to determine the effects of fishing on fish populations

and to evaluate the potential consequences of alternative harvest policies" (Methot and Wetzel, 2013). In other words, assessment models are used to assimilate data and provide advice for fisheries management. The assessment models also characterize uncertainty and project into the future. Hicks stated that many types of data are used as input to the assessment models, including catch (amount and type), abundance (survey and fishery catch rates), and biological information (age, size, and maturity). The data are then input into a population model. The model may also use external information, such as climate and environmental observations. The model returns current and future projections of abundance and mortality. Hicks observed that a key step is fitting the model to data by minimizing the differences between observations and predictions.

Abundance data, said Hicks, may be the most informative type of data. Most fish abundance data are relative and provide information about changes from previous observations. While relative abundance provides information about trends in the fisheries populations, it does not provide information about total absolute biomass, the absolute mass of a given species in a particular area or fishery. Hicks explained that measurement of total absolute biomass requires that the following criteria be met:

- Complete spatial coverage of the stock's range is needed.
- All potential sample sites within each stratum have a known probability of being selected. Hicks noted that habitat variability can hinder the selection of some sites.
- All fish at each selected site have a known probability of being detected. Ambiguous sample areas (e.g., stationary cameras) and unusual fish behavior (e.g., avoidance or attraction to measurement devices) can skew the probability of detection.

Hicks pointed out that biological data, such as length, weight, and age observations, are also important measures for the estimation of growth, recruitment, selectivity, and mortality rates. Maturity data (i.e., age of the fish) helps with understanding and measuring the spawning potential of the fish population. Ecological and environmental relationships can also be inferred. A participant noted that with computer vision, biomass may be able to be directly calculated by modeling the volume of fish, rather than relying on length-to-weight ratios.

Data are typically collected from two different sources, Hicks explained:

1. *Fishery-dependent data.* These data are not scientific surveys; rather, they are derived from fishermen targeting a certain stock. Hicks noted that these ad hoc methods are not the optimal way to collect data, but that the data are easy to obtain. Fishery-dependent data may include measurements of

retained catch, discarded catch, fishing effort, catch-per-unit-effort (an indirect measure of abundance), and biological information. Fishermen are relying more on image and video data, and electronic monitoring is becoming popular.

2. *Fishery-independent data.* These data are from scientifically designed surveys for collecting biological and abundance data. Hicks provided examples of a number of fishery-independent surveys, such as

    - *Capture surveys.* Fish are caught and measured to provide abundance and biological data, such as a Bering Sea trawler that collects bottom species from hundreds of sites per year. Capture surveys are typically fatal to the fish.
    - *Acoustic surveys.* Fish are found using an echosounder. Some capture is needed to benchmark the species' composition and size. Acoustic surveys provide abundance data and some biological data.
    - *Visual and advanced surveys.* In most cases, fish are observed without causing mortality. This can include scuba, camera drops, and the use of underwater vehicles.

Hicks stressed that other data can also be collected aside from abundance and biological data. Fish can be tagged to see if they return to an area; one can make visual observations of habitat; and one can make environmental observations, such as sea surface temperature.

Hicks explained that images and videos can assist in data collection and improve stock assessments due to the following:

- Fish mortality can be decreased with the increased use of video and images.
- Habitat information can be observed.
- Species that are typically not retained in a capture survey (such as very small fish) can be identified.
- Analysis speed is increased.
- Shapes and patterns can be recognized and classified.

Hicks concluded by noting that stock assessments consist of heterogeneous data from a variety of sources, and fishery-independent surveys, particularly bottom trawl surveys, are a key component to stock assessments. While relative indices are useful, absolute indices would be a big improvement.[1] Finally, image and video

[1] Relative indices provide measures of a fish population compared against the populations of other species of fish in the region, while absolute indices provide fixed measures of a fish population.

analysis will be increasingly useful tools to assist in the collection and analysis of fishery data.

OVERVIEW OF SAMPLING IN SPACE AND TIME

Steven Thompson, Simon Fraser University

A sample, Steven Thompson explained, is an observation in space and time that is made when one is interested in certain properties of a population but can only observe a portion of that population. He noted that populations usually have spatial and temporal structure that moves or changes in time, and those changes may not be predictable. Designs for sampling can progress dynamically through time and space. Detection and observation may not be ideal, said Thompson; for example, a collection net's results depend on its mesh size. Thompson explained further that a population is not fixed. Rather, it can be considered a stochastic process that evolves in time, and the sampling process is also stochastic.

Thompson explained the spatial-temporal population model, which is used to assess the effectiveness of various sampling designs. The model includes the following:

- Clustering, mixing, and migration;
- Movements within and among groups; and
- Insertions and deletions of objects: birth and death processes and immigration and emigration processes.

The sampling design, Thompson said, is the procedure for selecting units to include in the sample. In the case of a fisheries stock assessment, the sampling unit may be a fish, but more often it is the path of trawl or of video, sonar, or other imagery. The acquisition process is the process by which units are selected into the sample, and the attrition process is the process by which units are removed from the sample. In a conventional design, the procedure through which samples are selected does not depend on the variable(s) of interest. In adaptive design, however, the procedure depends on observed values of the variable(s) of interest as well as other, auxiliary variables. Thompson explained that through inferences from sample data, one can make estimates of abundance.

Thompson said that there are two primary methods of approaching inference from sample data: through a design-based approach or a model-based approach. In a design-based approach, the values of the variable(s) of interest are considered fixed but unknown. In a model-based approach, the population variables are considered to be random variables. Thompson noted that model-based approaches can be computationally complex; the most practical implementation is a computational

Bayesian approach using Markov chain Monte Carlo and other methods. He explained that there can be a tension between communities that use a design-based approach and those that use a model-based approach; however, his own work has encompassed both methods. Thompson provided an example in catch-per-unit-effort data: ideally, if there are more fish, the catch-per-unit-effort will increase. However, catch alone may not be related to actual abundance. Commercial fisherman essentially create their own sampling designs that are (1) of unequal probability (e.g., fishermen tend to go to areas where they have had previous success or where they are familiar with the underwater terrain); (2) adaptive (e.g., fishermen may continue to focus in a small geographic region if they are having success or move further if they are not having success); and (3) non-ignorable, so that the design needs to be accounted for in estimation (e.g., fishermen use additional cues from their environment). Thompson suggested that information about the fishermen's sampling design be included in modeling abundance.

Thompson concluded by stating that the science of sampling involves understanding how a sample is selected. Sampling designs rely on inference, experiments, and interventions, and the choice of sampling design can strongly affect the resulting data.

In a later discussion, several participants commented on the importance of a strong collaboration between data collectors and data analysts to make decisions about research experiments. Data analysts can provide information about how best to record data; this is particularly critical in situations in which detection probabilities are changing in time or where adaptive sampling is used.

NOAA FISHERIES STRATEGIC INITIATIVE ON AUTOMATED IMAGE ANALYSIS

Benjamin Richards, NOAA Fisheries

Benjamin Richards chairs the NOAA Fisheries Strategic Initiative on Automated Image Analysis. Another NOAA initiative has been established to examine the related topic of sampling in untrawlable habitats. From both initiatives, NOAA seeks to obtain better estimates of abundance to improve its estimates of fish populations and associated stock assessments.

In 2001, Richards said, NOAA developed its fisheries stock assessment improvement plan (Mace et al., 2001). That plan identified accurate and precise estimates of species-specific, size-structured abundance as a main impediment to stock assessment. In other words, the accuracy and precision of output estimates are directly linked to the quality of the input data. Richards stressed that optical technologies have many advantages: they are fishery-independent (i.e., they are not influenced by market drivers or other variables that can bias fishery-dependent estimates), non-

invasive (i.e., they can be used on overfished stocks or in protected areas without additional impact), efficient, and accurate. However, the use of optical technologies results in extremely large data sets, too large to be examined solely by human analysis. Millions of images collected in the span of a few days would take humans months or years to examine. He emphasized the need to reduce the burden on the human in image analysis, as well as the need to reduce the subjectivity associated with human data analysis. Richards later noted that human observers, in general, do not miss many fish, but they have a tendency to over-identify objects as fish (more false positives). Different human observers may also have divergent opinions on the identity of the same individual and may be more attuned to different species based on level of interests, expertise, or past experience. Human observers also can make subjective decisions. Presently, algorithms, while they tend to be more consistent among samples, also tend to miss fish (more false negatives) and misidentify fish.

Richards described a 2010 NOAA workshop on automated image analysis. The workshop specifically recommended increasing interdisciplinary collaboration between the marine research and computer vision communities, creating an international working group for the automated analysis of marine species, developing a database of commonly encountered fish that is accessible to the user community, and optimizing the allocation of resources and automation.

Richards explained that image data sets can be broken into categories: still versus video, mono versus stereo, static versus dynamic backgrounds, and natural versus artificial lighting. He then provided specific examples of the types of data that NOAA examines:

1. *Towed-diver benthic surveys.* Richards said that towed-diver benthic surveys are a simple example of work conducted by the Pacific Islands Fisheries Science Center. In this case, a diver moves through the water at 0.5 knots, and a still camera captures a downward-facing image every 15 seconds in standard lighting conditions. Coral Point Count (or similar software) is used to distribute points, and humans then classify the benthic habitat at these points. The mission is conducted once or twice per year, with approximately 60,000 images collected, and some 6 million images have been archived. Five human analysts study these images.
2. *Habitat Mapping Camera System (HabCam).* HabCam is a towed camera for benthic surveys, primarily targeting sea scallops, benthic invertebrates, and benthic fish by the Northeast Fisheries Science Center. A camera sled is towed at 5 to 7 knots at 1 to 3 m above the bottom. The sled contains stereo, digital, still cameras, using standard lighting to obtain 6 frames per second. Some 15 million images are in the archive. Approximately 10 ana-

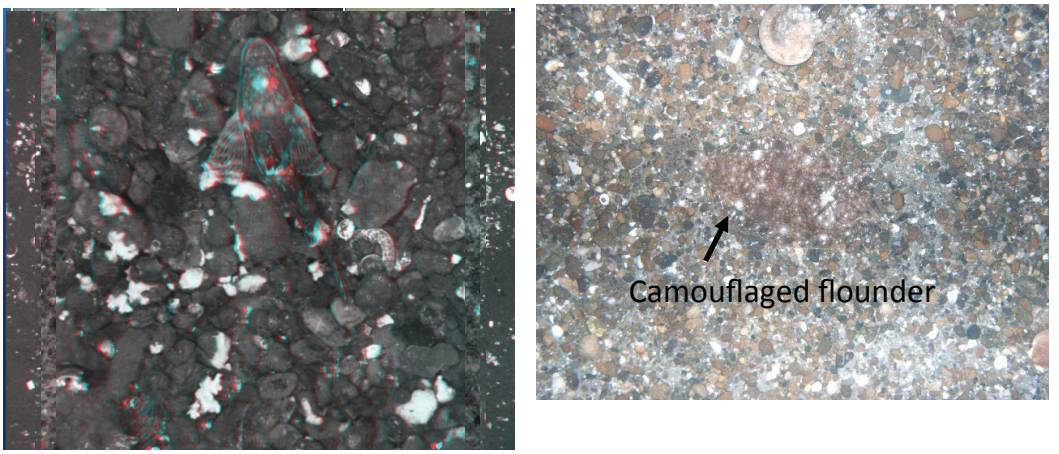

FIGURE 2.1 Sample images captured by HabCam. Both images contain benthic fish. SOURCE: Courtesy of the HabCam Group.

lysts study these images, and NOAA is leveraging crowdsourcing[2] to help with the analysis. Examples of images captured by HabCam are shown in Figure 2.1.

3. *SeaBED Autonomous Underwater Vehicle (AUV).* An AUV is used for surveys of demersal fish[3] by the Northwest Fisheries Science Center. The SeaBED AUV travels 3 m above the seafloor at 0.5 knots. It collects approximately 100,000 images per year, with a stored archive of 350,000 images. SeaBED AUV also has a video feed that produces around 100 hours of video per year.
4. *Cam-Trawl.* Cam-Trawl, a combined stereo camera and trawl system, is used by the Alaska Fisheries Science Center to sample pollock stocks. The trawl is used as an aggregating device to bring fish before the camera. The camera is side-facing (relative to the trawl), and a homogenous static background is used to ease the fish segmentation and measurement activities. Cam-Trawl acquires 3 million to 4 million images per year, with an archive of 8.2 million images. A sample Cam-Trawl image is shown in Figure 2.2.
5. *QuadCam.* QuadCam, a stereo camera platform used by the Southeast Fisheries Science Center to study reef fish, looks for fish against a complicated coral reef background. Fish come in and out of the image frames,

[2] For more information, see the Seafloor Explorer website at http://www.seafloorexplorer.org/, accessed June 6, 2014.

[3] Demersal fish live and feed at or near the bottom of the ocean.

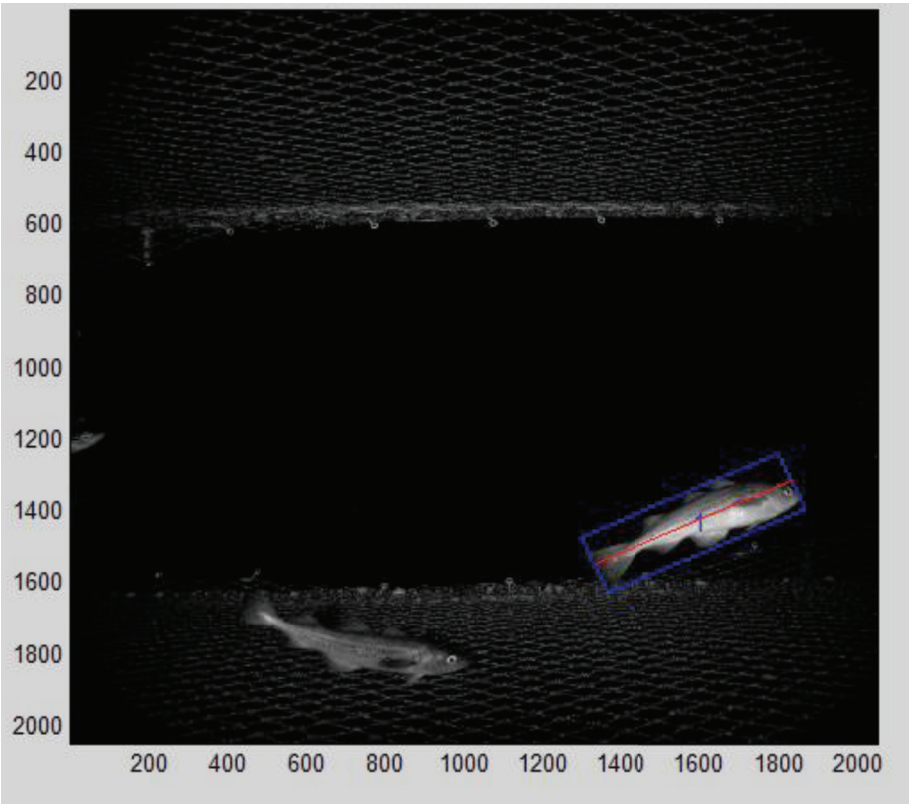

FIGURE 2.2 Sample image from Cam-Trawl. SOURCE: Cam-Trawl system, Alaska Fisheries Science Center, NOAA, courtesy of Kresimir Williams.

with varying levels of abundance and occlusion, and the ambient light conditions are constantly changing. A rosette of four stereo camera pairs takes images at 1.2 frames per second, resulting in 13.7 million image pairs per year and a large archive of 83 million images.

6. *Baited Remote Underwater Video Station (BRUVS).* BRUVS is a stereo camera system used by the Pacific Islands Fisheries Science Center as well as the University of Western Australia and others. Like QuadCam, it also targets reef fish. It is a small, easy-to-use, and fairly inexpensive system that uses off-the-shelf commercial cameras.
7. *Bottom Camera Bait Station (BotCam).* BotCam is similar to BRUVS and is used by the Pacific Islands Fisheries Science Center and University of Hawai'i to target deepwater bottom fish using ambient lighting at distances of up to 250 m. BotCam uses analog cameras that are targeted for light-gathering capability. Richards indicated that NOAA is transitioning

BotCam to a new digital camera system, which should improve automated analysis options. BotCam produces 100 hours of video per year.

Richards explained that NOAA maintains a website[4] to collect images from different technology platforms and make them publicly available to research partners.

Richards briefly described the following main challenges with image analysis in fisheries stock assessment:

- Species identification,
- Unclassified targets,
- Occlusion,
- Cryptic or non-moving targets,
- Complicated, moving backgrounds,
- Fish that enter and reenter the frame,
- Catchability, and
- Scaling to absolute abundance.

Richards posited that a worthwhile goal is to develop a toolbox—a collection of open-source tools to automate image and video analysis—that could be made readily available to the public for research and general use. A participant later noted that any open-source toolbox would need to be maintained and tested, and it would need to be transitioned to a company or open-source association; such maintenance is unlikely to occur in academia.

In response to a later question, Richards explained that NOAA funds automated-image-analysis projects in three ways: (1) requests for proposals developed through working groups, (2) direct funding of projects through work on a strategic initiative, and (3) small business innovation research grants. Several participants suggested that NOAA advertise these programs more widely in the computer vision community to bring in new participants who may not be aware of these opportunities.

OVERVIEW OF COMPUTER VISION

Ruzena Bajcsy, University of California, Berkeley

Ruzena Bajcsy explained that she would not discuss computer vision as a whole, but instead would focus on the specific computer vision challenges posed by fisheries stock assessment. She indicated that because the fisheries community seems

[4] For more information, see the NOAA Fisheries Strategic Initiative on Automated Image Analysis website at http://marineresearchpartners.com/nmfs_aiasi/Home.html, accessed June 6, 2014.

well aware of the existing, standard computer vision technologies, this workshop could help provide additional information about new and novel ideas that can be applied to the specific problems in fisheries. She listed some fisheries-specific challenges, including the following:

- Light intensity (bright versus dark),
- Water clarity (clear versus murky),
- Background in images (homogeneous versus heterogeneous),
- Contrast (low versus high),
- Camera movement (stationary versus moving), and
- Assemblage type (shallow versus deepwater).

Bajcsy noted that many of these challenges result from poor signal-to-noise ratios. One option, studied by Ben Recht (University of California, Berkeley), is to frame denoising as an optimization problem. Bajcsy noted that signal-to-noise ratios can also be improved by including multiple cameras, as described below, a method that is more feasible now that cameras are less expensive.

Bajcsy then stated the goals of fisheries image analysis:

1. Segment the fish into individual components and recognize the categories. Bajcsy noted that the difficulty of segmentation depends largely on the signal-to-noise ratio.
2. Measure the body mass of each fish. To do this, one must first compute the volume.
3. Compute the fish mortality.
4. Compute the maximum sustainable yield.

Bajcsy explained that a fisheries stock is overfished when its cumulative biomass (measured in task 2) has fallen to a level below that which can produce the maximum sustainable yield. She also noted that there is a need to monitor the relationship of the fish mortality (task 3) and the level of total biomass (measured in task 2) at the maximum sustainable yield. Bajcsy stated that tasks 2, 3, and 4 have relationships that change as a function of time, and she noted that computer vision can be used to classify different species using the outline of the fish and standard machine learning technology for classification.

One method of improving the signal-to-noise ratio, Bajcsy noted, is through the use of a camera array (such as a 4 by 4 camera array) at a fixed length to a target. With a priori knowledge of the array construction, one can quickly compute dimensional information from sets of images. This is a novel way to implement computer vision, as most systems are limited to two or three cameras. While the

camera array has advantages in speed and quality, it is limited in size and must be moved along the ocean floor.

Bajcsy then pointed to a recent paper discussing feature extraction for segmentation and identification on fast feature pyramids for object detection (Dollar et al., 2014). To improve speed, this research suggests approximating multi-resolution image features by extrapolating from nearby scales rather than computing them directly. Bajcsy said that by selecting a desired resolution, one can leave out background information and focus on the feature(s) of interest. For example, Carson et al. (2002) used a joint color-texture-shape feature representation as "blobs" for segmentation and recognition. A similar paper in 2013 (Lee et al., 2013), Bajcsy said, takes advantage of the fact that some features do not change in time.

Bajcsy then listed a number of useful machine learning techniques that may be helpful in this community, including the following:

- Maximum likelihood estimation,
- Multivariate Gaussian distributions,
- Linear regression,
- Logistic regression, optimization support vector machines (SVMs),
- SVM non-parametric methods,
- Nearest-neighbor clustering,
- Decision trees,
- Neural networks,
- Unsupervised learning,
- Mode seeking, and
- Dimensionality reduction using principal component analysis.

Bajcsy noted that stereo camera systems and other techniques now provide the ability to generate multi-dimensional data (both three- and four-dimensional data). From these data, one can compute a measure of biomass. A superquadric representation[5] gives a volumetric representation of an object to provide an assessment of biomass. The superquadric representation can be combined with other, more general, transformations in a systematic way to model other specific behaviors, such as twisting or bending.

Bajcsy concluded by stating that there is a clear need for the fisheries and computer vision communities to collaborate for mutual benefit. She suggested that the analyses of fishery data be framed as a food security issue, not just an ecological issue, to highlight its importance. She also suggested including the environment

[5] Superquadrics are equations that define geometric shapes; they are similar to the equations that describe ellipsoids, but the squaring operation is replaced by an arbitrary power. This technique is commonly used in computer graphics.

in studies, not just fish: the data set becomes richer and potentially interesting to more communities. Other participants noted the importance of distributing data to a wider audience; casting a wider net will bring in more interested people.

SIMULATING FISH AND OTHER SWIMMERS

Demetri Terzopoulos, University of California, Los Angeles

Demetri Terzopoulos explained that his work (with his former Ph.D. student Xiaoyuan Tu) on simulating fish movement was first published nearly 20 years ago and focused on reverse engineering real-life swimming examples. The artificial life approach, according to Terzopoulos, yields lifelike, autonomous agents through the comprehensive modeling of animals. This includes not only conventional computer graphics models of the shape and appearance of animals, but also modeling the functionality of the animal (including biomechanics, perception, motor control, behavior, learning, and cognition).

Terzopoulos explained that each artificial fish's components were developed in a bottom-up approach. The artificial fish agents are independent and make their own decisions in response to their environment. He later clarified that the artificial fish model has randomness included in it; fish randomly explore if no other stimulus (mating, feeding) is in place. The artificial fish model consists of the following elements, shown schematically in Figure 2.3:

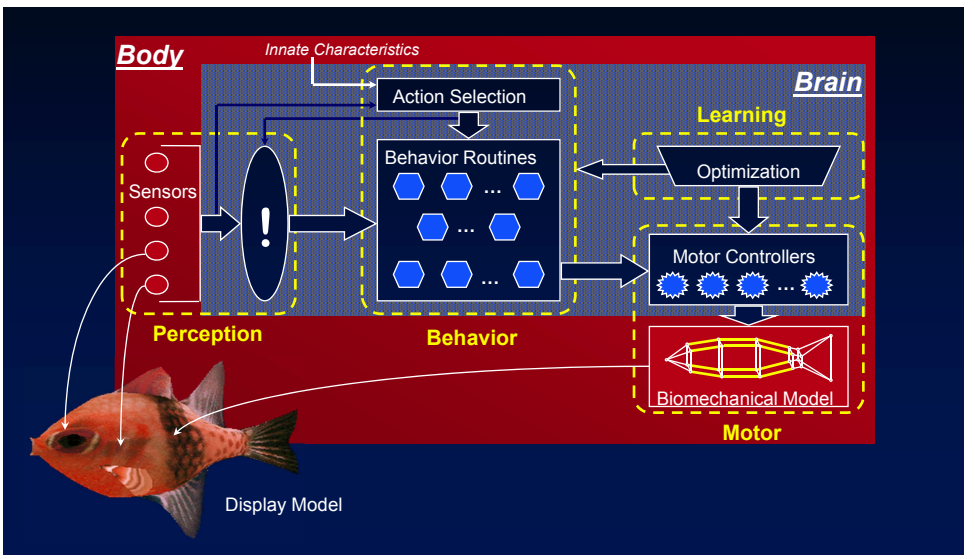

FIGURE 2.3 Schematic diagram of the components of the artificial fish model. SOURCE: Courtesy of Professor Demetri Terzopoulos, University of California, Los Angeles.

- *Display model.* This includes geometric and appearance representations. Terzopoulos indicated that the various fish display models were created using image-based modeling.
- *Body model.* This includes a biomechanical submodel, sensors (such as eyes and lateral lines), and a brain.
- *Brain model.* This includes a motor center to drive the biomechanical model, a perception center that interprets sensory information, a behavior center that ties percepts to the appropriate actions, and a learning center that enables the fish to learn from its experiences.
- *Biomechanical submodel.* Terzopoulos indicated that this was a simple physics-based model, yet it was capable of synthesizing visually realistic fish locomotion. The three-dimensional (3D) model consists of 23 lumped masses (particles) and 91 viscoelastic elements, 12 of which are contractile muscles. The model is mathematically characterized by a system of differential equations whose numerical time integration simulates the fish's motion.
- *Learning center.* Terzopoulos indicated that fish locomotion tends to be energy efficient, so the locomotion learning problem can be solved using an optimization strategy, from which the natural rhythmic caudal fin beating pattern emerges.
- *Perception center.* This models the capabilities and limitations of the animal's sensory apparati. Terzopoulos indicated that this consists of a sensorimotor perception system that includes both a stabilization module and a foveation module.
- *Behavior center.* Here, a set of behavior routines is organized in a loose hierarchy. Low-level behavior routines form a substrate supporting higher-level behaviors. The behavior models consist of three components: innate characteristics (such as gender, preferences, and capabilities), mental state (such as fear, hunger, and libido), and action selection. Action selection is prioritized by the level of perceived danger; for instance, first, a fish would avoid collisions, next, it would avoid predators, then, it would find food, and finally, it would find a mate.

Terzopoulos showed examples of synthesized fish motion that exhibited different behaviors, such as foraging, avoiding predators, and schooling. Schooling, Terzopoulos noted, is a distributed local model and does not require a leader; the behavior of each fish is guided by maintaining a certain distance to nearby animals and following the fish directly in front of it. A participant in the audience noted that schools can behave on a more macroscopic level: if a predator appears, the school coalesces into a ball. The fish in the middle are less likely to be caught, so

all the fish go to the middle. The simulation here does include simple predator-induced schooling behavior, and Terzopoulos responded that the behavior model is extensible and can incorporate more complex behaviors.

Terzopoulos explained that artificial fish were a popular topic in the late 1990s, and he referenced a best-selling book by Richard Dawkins (1996) that reviewed the artificial fish model and the journal *Artificial Life,* in which a technical paper on the model was published. He also noted the popularity of the virtual fish tank exhibit at the Boston Museum of Science, the Submarine Virtual Reality Theater, and popular screensavers showing animated fish.

Terzopoulos emphasized that the work on artificial fish was decades old and that a more complex model could be built today. His current models of human swimming include all the bones and almost all the relevant skeletal muscles in the body, along with a complex, 3D finite element model to simulate soft tissue behavior subject to the contractions of the embedded muscle actuators and induced water-pressure forces. The simulation of human motion consists of three interleaved simulators: rigid/articulated bone simulation, soft tissue simulation, and simulation of the surrounding water using computational fluid dynamics.

A participant asked if Terzopoulos studied inverse kinematics from a physiological point of view. He responded that he has forged relationships with biomechanics researchers who study human motion; for example, he was recently contacted by medical school researchers who were interested in neck motor control problems.

A participant noted that turbulence can be hard to model. Salmon may be able to swim upstream with greater than 100 percent efficiency by taking advantage of turbulence. Terzopoulos acknowledged that new literature exists on fast swimmers, such as tuna and salmon, who exploit turbulence and higher-order effects.

The discussion concluded with a workshop participant noting that the behavior of fishermen is complex and adaptive and may be much harder to model than fish.

THE FISH4KNOWLEDGE PROJECT: AUTOMATED UNDERWATER VIDEO ANALYSIS FOR FISH POPULATION MONITORING

Concetto Spampinato, Università di Catania (Italy)

Concetto Spampinato described the Fish4Knowledge (F4K) project, a large-scale underwater video collection and processing effort to enable dynamic browsing and presentation of massive amounts of marine data. F4K, which was a 3-year program funded at a level of 2 million euros, encompassed a variety of difficult environmental factors and differing user needs.

The F4K experimental design consisted of nine static cameras that filmed for 12

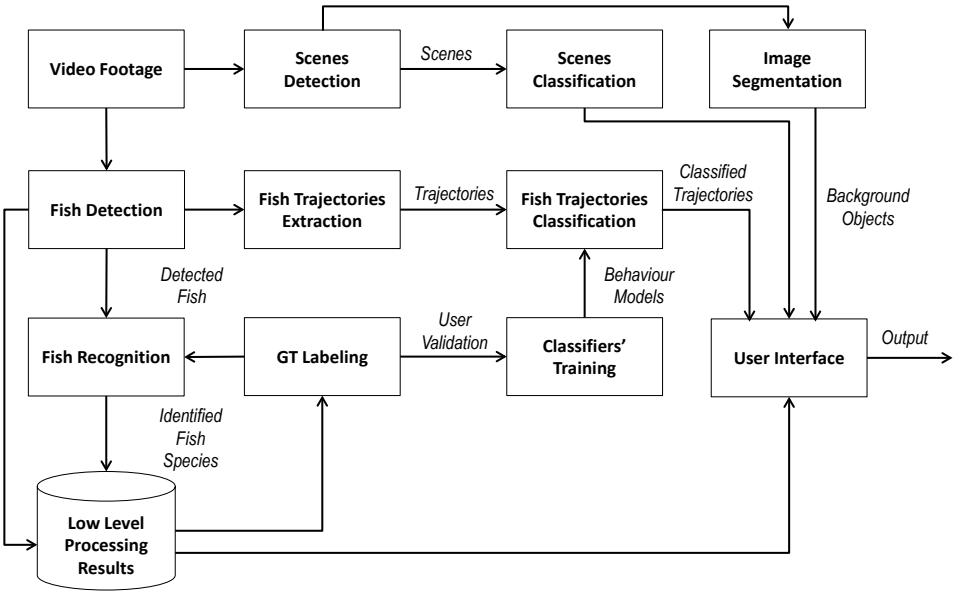

FIGURE 2.4 Schematic of the Fish4Knowledge video analysis system. SOURCE: Concetto Spampinato, Università di Catania (Italy).

hours per day for 3 years, resulting in several terabytes of data. Spampinato stated that within the F4K an automatic approach for fish detection and species classification was developed. As a result, from 3 years of videos, F4K identified a total of 1.55 billion fish, half of which were identified by species. Ninety-nine percent of the fish observed belonged to 1 of 23 species of fish.

F4K used high-performance computing facilities for processing. An interface allowed the user to filter the data by time of day, week of the year, year, location, or camera. F4K provided some information about relative abundance. Spampinato explained that all aspects of F4K remain publicly available, including the source code,[6] user interface,[7] and data.[8]

Spampinato described the video analysis system in detail and presented a schematic of the analysis system (shown in Figure 2.4). Annotated images were

[6] For more information, see SourceForge, "Fish4Knowledge Project," http://sourceforge.net/projects/fish4knowledgesourcecode/, accessed June 16, 2014.

[7] The Fish4Knowledge user interface is at http://f4k.project.cwi.nl, accessed June 16, 2014.

[8] For more information, seeFish4Knowledge, "Fish4Knowledge Video Sample Download Page," http://groups.inf.ed.ac.uk/f4k/F4KDATASAMPLES/INTERFACE/DATASAMPLES/search.php, accessed June 16, 2014.

needed for testing the video analysis methods, and Spampinato indicated it was difficult to obtain volunteers for this large task. Instead, F4K developed an online game, among other interfaces for annotation collection, to encourage users to select and identify fish. With a large number of users, he reported that the quality of the annotations was high.

For fish detection in videos, F4K applied background modeling methods. F4K used neighborhood samples to model the background, instead of more traditional methods.

F4K initially used kernel density estimation[9] for background and foreground modeling; Spampinato indicated that this method provided the best results among several tested techniques (Spampinato et al., 2014). He noted, however, that the kernel density estimation approach is slow, able to analyze about 1.5 frames per second. The ViBE[10] approach worked more quickly but was susceptible to false positives from changes in the light intensity; because of its increased speed, however, it was selected in F4K's system. To reduce the rate of false positives, a post-processing module was developed with rejection algorithms to assign an object a probability estimating the likelihood of that object being a fish. With this amendment, F4K was able to exploit the faster speed of the ViBE approach. To improve fish detection performance, both intraframe (boundary complexity, boundary color contrast, etc.) and interframe (motion on boundary, motion homogeneity, etc.) features were used in a naïve Bayes classifier[11] (Spampinato and Palazzo, 2012).

After the fish were detected, they were classified by their species. A feature vector with 69 features (with metrics to describe color, boundaries, and texture) was used in a balance-guaranteed optimization tree[12] (Huang et al., 2012). The classifier was applied to about 30,000 detections, and results were confined to the 23 most popular fish species. With rejection algorithms applied, the classification accuracy was about 65 percent.

Spampinato pointed out that the F4K system had over a billion detections; that amount of information can be exploited to improve detection and classification in the future.

F4K annotated two data sets: one data set consisted of 20 million images that were labeled as either having a fish or not having a fish; a second data set consisted of 2 million images annotated with labels of the 23 most common fish species.

[9] Kernel density estimation is a smoothing function that non-parametrically estimates the probability density function of a random variable.

[10] For more information, see ViBE Corporation, "Welcome to ViBE," http://www.vibeinmotion.com/Home.aspx, accessed June 16, 2014.

[11] A naïve Bayes classifier assumes that the presence of one feature is not related to the presence of any other feature.

[12] A balance-guaranteed optimization tree is a decision tree that selects a subset of features at each node for object classification.

Spampinato explained that they then conducted semi-supervised learning: they first trained a classifier, and the most likely results were put into the training set for retraining.

Spampinato concluded by describing two new projects under development:

1. *AQUACAM.* This is a program to introduce the F4K technology into the Caribbean to study biomass, marine protected areas, and preservation. Unlike F4K, this project will have a stereo-based approach for fish biomass, while it will conduct automatic species recognition on the most common Caribbean species, and provide a user interface for queries and results, as in F4K.
2. *UNDERSEE.* This project will investigate how to automatically adapt to a change in domain using semantics-guided computer vision approaches.

3

Multi-Modal Sensing

The second session of the workshop provided an overview of multi-modal sensing and discussed some of the key challenges associated with various types of sensor data (such as image, video, lidar, hyperspectral, and stereo with motion) and their integration. Useful references for this session's topics, as suggested by the workshop program committee, are Atrey et al., 2010, and Kunz and Singh, 2013. The session was chaired by Nicholas Makris (Massachusetts Institute of Technology). Presentations were made in this session by Dvora Hart (Northeast Fisheries Science Center), Jules Jaffe (University of California, San Diego), Nicholas Makris, and Fraser Dalgleish (Florida Atlantic University).

FISHERIES PERSPECTIVE OF MULTI-MODAL SENSING

Dvora Hart, Northeast Fisheries Science Center

Dvora Hart began by discussing conventional surveys that are conducted via traditional fishing gear. The resulting time series enables one to gauge trends, measure length, and obtain physical samples. She noted that single data points are not particularly helpful, as fish tend to aggregate; any single measurement gives little information about the number of individuals living throughout a region of interest because it can be too high or too low depending on whether an aggregated mass of fish is observed. Hart showed results from traditional surveys of two populations, Gulf of Maine cod and Georges Bank sea scallops. Two distinct trends have been

observed in those populations. The Gulf of Maine cod have been experiencing an exponential decay of their population. In contrast, the Georges Bank sea scallops are a recovering stock. A participant asked if the declining cod populations were likely due to warming of the ocean temperatures in that area, overfishing, or some other variable. Hart responded that overfishing was the primary cause; the cod population began its decline before the recent warming trends. She noted that while warming may not be the proximate cause of the cod's decline, it can make the population more vulnerable to other stressors such as overfishing.

While conventional surveys provide information about population trends, they do not directly provide absolute size. Hart described how to estimate absolute size from catch (which can be counted) and natural mortality (which is approximately known). Total mortality can be estimated by comparing year-to-year data; fishing mortality is total mortality minus natural mortality; and catch is fishing mortality times fishable biomass—i.e., biomass is catch divided by fishing mortality rate. However, the estimate of natural mortality is highly uncertain and can vary in time, leading to a corresponding uncertainty in total mortality. Ironically, as fishing mortality rates approach natural mortality rates (a desired outcome), absolute biomass and fishing mortality estimates become even more uncertain. Hart emphasized the difficulty in obtaining accurate absolute scale, as well as its importance to fisheries stock assessment.

Advanced sensors, Hart explained, have the potential to provide a direct measurement of absolute scale. They also can be used in complex habitats, such as reefs, where trawling cannot be done, and are a non-lethal method of sampling. Hart provided an example from the Northeast Fisheries Science Center of the HabCam-towed camera system. This system covers a large area in a short amount of time (approximately 50,000 km^2 in 3 weeks of ship time). More than 2 million images were collected per week using stereo cameras, a side-scan sonar, and a sensor package. The side-scan sonar observed evidence of fishing activity (such as dredging or bottom fishing), and aggregating those observations with other sensor data could be potentially very valuable, said Hart. About 150 paired tows with HabCam and dredging have been conducted in order to calibrate the dredge to an absolute scale. Hart noted that dredge surveys are estimated to have an efficiency of 40 percent in sand and 25 percent in gravel or other rough bottom conditions. HabCam can conduct full-scale resource surveys to track and observe fish populations, providing orders of magnitude more data than a conventional survey. Abundance and biomass are then modeled and estimated.

Hart then described the use of HabCam to determine the biomass of scallops. Six million photographs were taken by HabCam, and 1 percent of those photos were analyzed by hand. The manual annotation of images took about 5 weeks, and applying the models took another week. Hart emphasized that automation could significantly reduce both costs and labor.

Hart also described challenges associated with estimating finfish. A finfish is only present in 1 out of every 100 to 1,000 images (1 to 2 orders of magnitude less than the prevalence of scallops), so finding them manually can be very difficult; automation would thus be very useful for this task. She described a specific prototype study of yellowtail flounder. In a 2010 survey, more than 150,000 images were examined, in which a total of 250 yellowtail were observed. In 2012, more than 83,000 images were examined, in which only 19 yellowtail were observed—a 7-fold reduction in prevalence. (A sample image is shown in Figure 3.1.) Hart emphasized that in both instances the images were taken in a sandy area, the habitat preferred by yellowtail flounder—in other words, these densities are likely to be an upper bound. The decrease in population from 2010 to 2012 may be the result of temperature effects; increased bottom water temperatures may be causing populations to decline, said Hart.

Hart concluded by reiterating that absolute scale is difficult to obtain, particularly when fishing mortality is at an appropriately low level. However, advanced sensing technology can help in estimating absolute scale, and the automation of

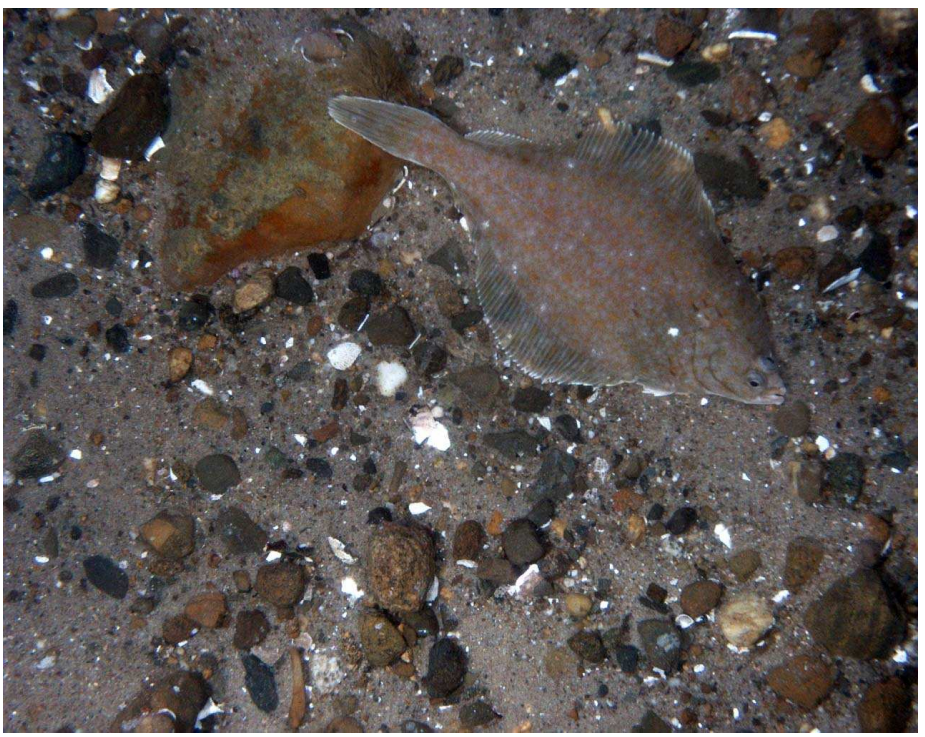

FIGURE 3.1 Image in which a yellowtail is observed. SOURCE: Courtesy of the HabCam Group.

image processing is critical to the analysis of large numbers of images. She clarified in a later discussion session that in the near future full automation is unlikely. However, she emphasized the importance of reducing the number of manual annotations required.

In a discussion session, a participant stated that many fish species are morphologically similar but exhibit different swimming and behavior patterns; he posited that behavior modeling could help in species identification. Large-scale movement patterns can also help with species identification. Another participant suggested that modeling the territorial behavior of fish might help avoid overcounting.

SYNERGISTIC ACOUSTIC AND OPTIC OBSERVATION AND ESTIMATION

Jules Jaffe, University of California, San Diego

Jules Jaffe explained that he was interested in observing different marine life using both acoustic and optic observations. He explained the general differences between acoustic and optic observation using the information in Table 3.1.

Jaffe explained that he used optics and acoustics to study euphausiids, mesopelagics, and diverse zooplankton in less than 500 m of water. The goals of the project are to obtain simultaneous target strength (i.e., the strength of the reflected sonar pulse, measured in decibels) and in situ identification, develop multibeam sonar systems, measure animal activity as a function of time of day, and monitor animal behavior. In the 1990s, Jaffe constructed a multibeam system for three-dimensional (3D) tracking of targets. Known as FishTV, the sonar system consisted of an 8-by-8 multibeam system, a source in the 400-500 kHz range, resolution on the order of 2 degrees by 2 degrees by 1 cm, with four images per second. Jaffe also described the Optical and Acoustical Submersible Imaging System (OASIS), which calibrated sonar with images. OASIS could identify targets as a function of body length, and it was used to study plankton as a function of size. OASIS observed size-dependent

TABLE 3.1 Comparisons Between Optic and Acoustic Observations of Marine Life

	Optics	Acoustics
What is there?	Excellent	Poor; usually requires net tows or a priori knowledge
How much is there?	Size dependent: • small (<mm): good • large (>mm): poor	Echo-counting: excellent Echo-integration: medium
What is it doing?	In situ behavioral observations never done	Can track some behavior

migration patterns: smaller plankton would migrate to the surface first at night (De Robertis et al., 2000).

Jaffe also examined mesopelagics, noting that these are understudied. He explained that immature myctophids have an air bladder that reflects acoustic energy well; in adulthood, the animals do not retain the swim bladder, and their acoustic reflection profile is smaller as a result. Jaffe described an observing system called OmniCam used to study mesopelagic swarms seen via acoustic backscatter. OmniCam has six wide-angle cameras for full environmental coverage and simultaneously records video, light spectrum, 3D orientation, depth, and temperature.

Jaffe also described some more recent work in acousto-optic imaging of plankton. A Multiple-Aspect Acoustic Zooplankton (MA-ZOOPS) system was fitted with an optical imaging system (MA-ZOOPS-O) to combine acoustic reflectivity with optical imagery. The system first needed to be cross-calibrated and was then used to study marine snow[1] reflectivity. Other work involves examining the target length and the duration of its acoustic reflectivity to obtain more information about the organism (Roberts et al., 2009).

REVEALING FISH POPULATION AND BEHAVIOR WITH OCEAN ACOUSTIC WAVEGUIDE REMOTE SENSING

Nicholas Makris, Massachusetts Institute of Technology

Nicholas Makris explained that Ocean Acoustic Waveguide Remote Sensing (OAWRS) is a technology that uses the ocean as a waveguide. The sound source is in the audible range (usually 400 to 2000 Hz), and sound waves are emitted at the source for about 1 second in pulses that repeat roughly every minute. The resulting images can be concatenated into a "movie" that goes on for hours. Makris explained that this technology can be used to identify fish populations over very wide areas and has found very large fish shoals containing hundreds of millions of fish (Makris et al., 2006, 2009). OAWRS provides horizontal information only; if vertical information is desired, OAWRS must be complemented with conventional echosounding technology. Also, OAWRS provides species information by spectral analysis, where resonant swim bladder response can be used to identify species remotely (Jagannathan et al., 2009; Gong et al., 2010; Jain et al., 2013). For ground truth species classification, Makris said, supplemental trawling and catching is required if other behavioral clues are not sufficient. Makris explained that heavy

[1] Marine snow is a shower of organic material falling from upper layers of water to the deep ocean. For more information, see NOAA, "What Is Marine Snow?," http://oceanservice.noaa.gov/facts/marinesnow.html, accessed July 9, 2014.

equipment is needed for both the acoustic source and the receiver, but the equipment is no more cumbersome than typical trawl gear.

Makris provided an example of acoustic surveys of the Mid-Atlantic Bight in 2003 (Makris et al., 2006) and the Gulf of Maine and Georges Bank in 2006 (Makris et al., 2009). These were fairly extensive efforts, similar to those used in typical trawl surveys, consisting of an OAWRS source ship, OAWRS receiver ship, and one or two ships using conventional echo sounding and trawl. (The OAWRS system has also been used from a single ship that also has conventional echo sounding.) Each of the studies discovered many large herring shoals not found using NOAA's conventional techniques. The 2006 study determined that spawning herring shoals were forming overnight: during the day, fish would stay near the bottom, and near sunset they would start to develop small formations that would coalesce into a shoal. Once the shoal reached a critical size, it would quickly rise in population and extent to form vast shoals of more than 250 million fish. This is shown in Figure 3.2.

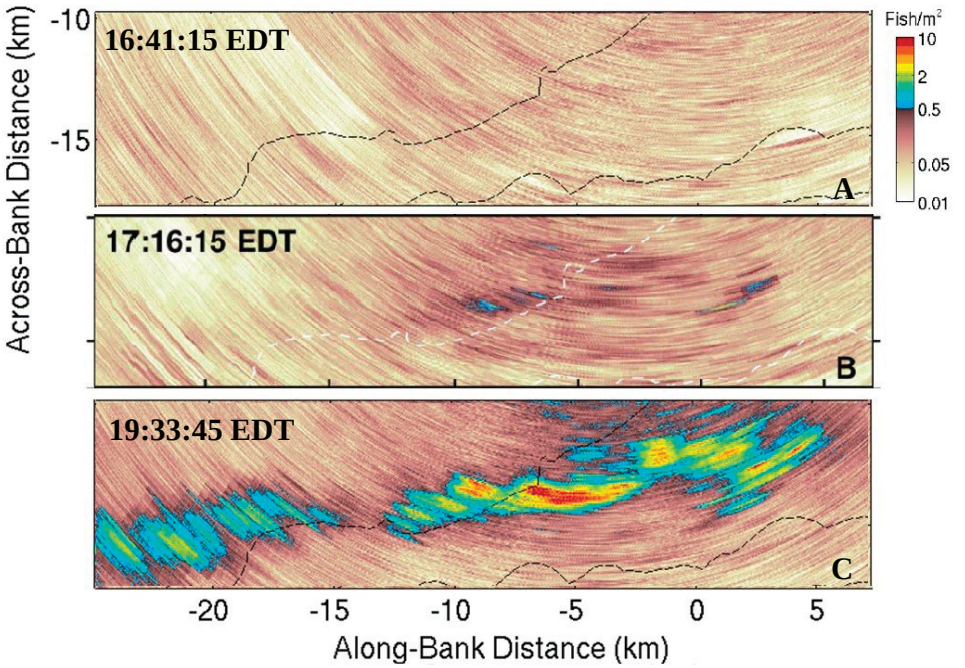

FIGURE 3.2 Critical population density of herring triggers rapid formation of a vast oceanic shoal. SOURCE: N.C. Makris, P. Ratilal, S. Jagannathan, Z. Gong, M. Andrews, I. Bertsatos, O.R. Godo, R. Nero, and J.M. Jech, 2009, Critical population density triggers rapid formation of vast oceanic fish shoals, *Science* 323:1734-1737. Reprinted with permission from AAAS.

Makris explained that OAWRS has also been used to conduct herring stock assessment by compiling images of shoals during a spawning period. The maximum population density values of each point each day were integrated, leading to an overall abundance estimate. This technique was applied to data from the Gulf of Maine, resulting in a population estimate consistent with NOAA's estimates using other techniques. The data indicate that a single herring likely only goes through the shoaling process once before spawning, said Makris (Makris et al., 2009; Gong et al., 2010). OAWRS technology has also been applied to examine Atlantic cod in the Gulf of Maine. Cod shoals were only discovered in a 0.5 km region, which is very small relative to historic cod shoal regions (Jain et al., 2013). Makris also explained that computer vision can be used to estimate velocity fields from time-varying density images (Jagannathan et al., 2011). The velocity field shows how a shoal is moving.

During the discussion period, a participant pointed out that the session focused on ways optics could inform acoustics to identify targets and asked if the converse could be true: Are there ways in which acoustics could inform the classification of optical targets?

Several participants noted that, in general, acoustics can describe behavior on a larger scale and set the context, enabling the researcher to conduct finer-scale sampling. For instance, Makris pointed out that acoustics can inform where in the ocean would be the best location for an optical sensor. Another participant noted that while humans tend to explore with optics, light attenuates rapidly and can have limited function. In contrast, audible sound can travel nearly around the world, and long electromagnetic wavelengths can penetrate water.

SEAFLOOR LASER IMAGING TECHNIQUES

Fraser Dalgleish, Florida Atlantic University

Fraser Dalgleish explained that he pairs emerging laser technologies with spatially and temporally resolved methods of imaging. Dalgleish listed three rules to improve underwater imaging:

1. Maximize photon density on each resolution element.
2. Maximize sensing of resolution element photons.
3. Minimize path radiance effects.

He explained that his systems are somewhat unconventional in that the laser source and receiver are located on different platforms.

Dalgleish first described the most common legacy system in laser-based imaging, known as the Laser Line Scanner (LLS), patented in 1973. LLS is a long-range

imaging system that has been successfully deployed on a variety of platforms, including helicopters, towed systems, and submersibles. Its range is 4 to 5 attenuation lengths, and it provides high contrast. However, it is physically very large (making its deployment expensive), has many moving parts, and has significant power requirements.

Dalgleish then described the multiple field of view (MFOV) prototype, which is a combination sonar/laser system. With no moving parts, MFOV has a compact design and can provide resolution of less than 0.5 cm in clear water at a depth of 9 m, which is sufficient for the identification of many benthic species.

He then showed results combining an LLS with a high-repetition rate pulsed laser (lidar) (Dalgleish et al., 2009, 2011; Dalgleish and Caimi, 2011; Caimi and Dalgleish, 2010). The lidar pulses 1 million times per second and scans 200 to 300 target lines per second. This is a smaller system in which the source and receiver are almost collocated, and backscatter is not a concern. A trial system was deployed in December 2013 that towed a source-receiver system at a depth of 70 m, about 11 m from the seafloor. The system was able to image objects on the seafloor at centimeter resolution. However, the laser uses a significant amount of power (up to 10 W); also, it is not eye-safe for humans and may not be eye-safe for marine life.

Other systems are being investigated that would not require such high levels of power. A simple solution, Dalgleish said, would be to move the laser source closer to the target; the swath size is sacrificed, but the common volume is reduced, minimizing path radiance effects. The detector can then be opened wider to allow for a larger angular field of view. With such a system, the distance from the target to the detector can increase to nearly 200 m in the open ocean. This is known as the distributed serial laser imaging concept. A prototype has been recently developed and built that uses low-power, eye-safe lasers and has more flexibility, as there is no need for alignment between the source and receiver. The prototype has been used in test conditions in a tank; in clear water, the receiver can be 11 m from the target, while the transmitter distance varies from 11 to 5 m. The prototype was also successfully deployed in the open ocean, under various realistic conditions. Dalgleish did note that many forms of marine life are attracted to the laser light, which can complicate the testing.

Dalgleish concluded by stating that serial laser imaging techniques offer improved image contrast and range of operation relative to cameras. Field tests have shown that distributed serial laser systems can provide images under a wide range of conditions with relatively simple, compact hardware. Future systems can be adaptive and used to track and identify targets and cover large scene volumes.

4

Image Processing and Detection

The third session of the workshop, chaired by Chuck Stewart (Rensselaer Polytechnic Institute), discussed issues related to image processing, such as imaging platforms, color and illumination correction, segmentation, recognition, and species detection. Some useful references for this session's topics, as suggested by the workshop program committee, include Dawkins et al., 2013; Kaeli and Singh, in review; Tolimieri et al., 2008; Treibitz et al., 2012; and Singh et al., 2007. Presentations were made by Clay Kunz (Google), Hanumant Singh (Woods Hole Oceanographic Institution), Ruzena Bajcsy (University of California, Berkeley), and Chuck Stewart.

COMPUTER VISION UNDERWATER

Clay Kunz, Google

Clay Kunz began by stating that platforms to study fish populations are a solved research problem: there are now a variety of platforms (such as AUVs, diver-carried rigs, towed cameras, remotely operated vehicles, and variable-ballast floats) that can be used for a given application. He explained that the variety of platforms enables the production of significant amounts of video or still imagery—as much as 100 terabytes of image data. Many different research groups are examining this data, leading to a lack of coherence in the community about research methods. The community would benefit, Kunz said, from the automated semantic analysis of image content. However, he posited that techniques used to recognize and classify human faces do not apply well to the classification of fish for two possible reasons:

1. Computer vision makes invalid assumptions underwater.
2. Locally valid corrections may not generalize to other image groups.

One invalid assumption that can be made in computer vision, explained Kunz, is the geometric calibration of the camera when placed underwater. However, he stated that current lens distortion models can compensate adequately for refraction underwater (Treibitz et al., 2012). With reasonable calibration, one can develop 3D imagery from stereo cameras, develop structure from motion, and fuse mono or stereo imagery with bathymetry to image a fish against the seafloor.

A second class of invalid assumptions in computer vision, said Kunz, is in radiometric calibration, and these issues are more complicated. He said that certain radiometric assumptions in computer vision may no longer be valid, including the following:

- Underwater scattering and absorption,
- Non-diffuse light sources,
- Non-Lambertian surfaces,[1] and
- Fluorescence and other adaptations.

To overcome radiometric calibration issues, Kunz suggested choosing the best platform to obtain physical proximity to the scene, using cameras with high dynamic range, using training data, and controlling the light sources as well as possible. He noted that radiometric correction techniques do not necessarily translate well across all underwater environments.

Kunz concluded by noting that today, the perception of fisheries work is that the problems are applied rather than theoretical and do not result in Ph.D. research or scholarly publications. He suggested this mind-set be changed. He also pointed out that with cloud computing, storage is now inexpensive, so data sets can be made more widely accessible. In addition, new types of sensors and higher-fidelity cameras are changing rapidly and in useful ways.

IMAGE UNDERSTANDING UNDERWATER

Hanumant Singh, Woods Hole Oceanographic Institution

Hanumant Singh began his presentation by emphasizing the need for vertical integration of all components that analyze underwater images: sensors, platforms, algorithms, and applications. He indicated that illumination-corrected,

[1] A Lambertian surface is one in which the apparent brightness of the surface is independent of the viewing angle; a non-Lambertian surface brightness does vary by angle.

high-resolution imagery is now readily available. Similarly, a number of vehicles and platforms exist from which one can obtain high-quality data. Singh therefore hypothesized that algorithms provide the greatest opportunity for improvements to underwater image analysis, particularly algorithms in the areas of segmentation and classification.

In Singh's view, fisheries-independent stock assessments have the following challenges: spatial (how to examine large areas and generalize the counting done in small areas), temporal (how to examine the same areas the following week or year), dynamics and biases (how to avoid counting fish multiple times due to fish attraction and/or avoidance), and processing time (real-time feedback).

Singh described a specific example in the classification of rockfish. He indicated that rockfish identification and analysis is (or should be) a tractable problem, as rockfish are red and provide high contrast against their surroundings. He pointed out that when imaging rockfish one is also imaging the habitat—i.e., the background against which the fish is imaged. Habitat can be considered a conditional probability: certain types of habitat, such as rubble, mud, or sand, are more likely to result in an image with a fish. In the rockfish example, while their red color is relatively easy to find, they can be confused with other fish species or with elements in the background. This complexity is demonstrated in the images in Figure 4.1. While certain fish are attracted to a camera, AUV, or trawl, and other fish avoid the hardware, rockfish move slowly enough that avoidance is not a significant concern. Singh noted the value of computer-assisted systems, rather than fully automated systems, because even a partial identification would help the fisheries community to measure abundance and classify species.

FIGURE 4.1 Images showing the challenges associated with identifying images of rockfish, even though rockfish are a bright red. The top row shows rockfish images. The middle row show images of other species. The bottom row shows images of background objects. SOURCE: Courtesy of H. Singh, N. Loomis; copyright Woods Hole Oceanographic Institution.

Singh explained that camouflage can affect both human and automated classification. He described work in real-time segmentation of images to find camouflaged marine life, looking specifically at the classification of octopuses and skates. He noted that in counting camouflaged animals a certain number of false positives will need to be tolerated.

Light-field cameras, Singh said, have the potential to help with seeing through turbid water, provide 3D reconstructions, and provide high dynamic range. Calibration may be a concern for light-field cameras because of differences with the air/water interface, although Singh believes this is a solvable problem.

UNDERWATER TELE-IMMERSION: POTENTIAL AND CHALLENGES

Ruzena Bajcsy, University of California, Berkeley

Ruzena Bajcsy explained that current computer vision technology allows real-time, 360-degree capture of the surrounding world. Tele-immersion is the digital capture and representation of the surrounding world. Her vision, she explained, is to develop a Skype of higher order, one in which people could communicate and interact in the same environment when not co-located, and she posed the question of whether such a system would be useful to fisheries scientists. To enable tele-immersion, she explained, three systems are necessary: capture (computer vision), communication (networking to bring two people into the same environment in real time), and display (3D, real-time rendering of the environment).

Bajcsy described a 45-camera system developed at the University of California, Berkeley, to provide a field of view that encompasses a complete human body. Her research group conducted a test exercise in which two people (one at the University of California, Berkeley, and one at the University of Illinois, Urbana-Champaign) danced together. She then suggested extending this concept of tele-immersion to the underwater environment, in which fishermen can interact with the environment and the fish. There may also be teaching and analysis applications to fisheries tele-immersion.

Bajcsy pointed out that challenges associated with tele-immersion systems include the following:

- Large-scale camera systems distributed over networks (though she noted that camera technology is rapidly improving);
- Real-time processing (3D reconstruction and rendering);
- Configuration and coordination of 3D/4D services, including calibration;
- Software synchronization;
- Portability;

- Ease of system maintenance; and
- Real-time operating system services on multi-core architectures.

Challenges in networking include the following:

- Large bandwidth demands,
- Real-time communication needs,
- Multi-stream coordination,
- Synchronizing multiple streams, and
- Different scales.

Challenges in display and rendering include the following:

- Immersive display of 3D video. Bajcsy indicated that reconstruction should be improved for more lifelike rendering and improved scalability. In the dancing example, the dancers' bodies rendered well, but there was insufficient detail in rendering their faces.
- Merging 3D video with a collaborative virtual workspace.
- Synchronizing virtual workspaces to enable face-to-face virtual conversations.

Bajcsy explained that tele-immersion can have a number of applications, including in the geosciences, in archaeology (to capture and digitally reconstruct information across different institutions as archaeologists remove objects), and in health care (medical data visualization). She noted that a user can interact with himself in the tele-immersive environment via pre-recording.

Bajcsy explained that she analyzes data to determine the minimum number of observations needed to describe a system (she contrasted this to Terzopoulos, who modeled and synthesized fish via a large number of individual bones and muscles). She provided a specific example in which the motion of 12 subjects was captured via cameras and sound recordings. The subjects were asked to do different, simple activities (such as sitting or bending over) while Bajcsy measured the most energetic joints and used that information to classify different activities. She noted that this type of measurement system has medical applications; for instance, one can measure the efficacy of different interventions on the range of movement of patients with muscular dystrophy. Bajcsy provided a second example in which muscle strength is measured with a single electromyography sensor. She said that this system could have medical applications for the aging; as an aging person loses strength in a particular muscle or joint, a device could be customized to supply additional force needed to reinstate some of the original function.

UNDERWATER IMAGING AND DETECTION

Chuck Stewart, Rensselaer Polytechnic Institute

Chuck Stewart explained that HabCam (described by both Dvora Hart and Benjamin Richards) returns 500,000 images per day, most of them uninteresting—in other words, most of the images do not have any marine life in them. A key in underwater imaging and detection is to identify those images that do contain something of interest.

Light intensity is non-uniform on the seafloor, explained Stewart, both because light sources are not uniform and also because light attenuation depends on wavelength. Image illumination can be estimated, and light maps can be "learned" at multiple altitudes above the seafloor, developing a 3D lookup table based on the angle of the HabCam sensor. The only light source is the one carried by HabCam; the lookup tables, therefore, can restore uniformity to the images.

Stewart then described the algorithm used to detect sea scallops in HabCam images, which consists of the following steps:

- *Preprocess.* Color is replaced with the likelihood of color, which enhances uncommon colors.
- *Detect candidate regions that are likely to contain scallops.* Four independent extractors from image processing are used in concert.
- *Extract large feature vectors.* The feature vectors used are about 3,800 units long and include such information as color, edge properties, and texture.
- *Classify.* Stewart explained that multiple AdaBoost classifiers[2] were used to determine whether an object is a brown scallop, white scallop, dead scallop, sand dollar, or clam. The substrate (sand, mud, or gravel) is hand-specified, and that information is used to determine the type of classifier to use.
- *Make a final detection decision.*

Stewart explained that no one classifier is clearly the best in a particular environment, although they all are efficient and provide reasonably accurate results. Overall detection rates vary from about 60 percent to about 95 percent, depending on the type of training and the substrate. Stewart noted that this success rate is consistent with human classifications from images. Automatic determination of the substrate will lead to automatic selection of classifiers.

Stewart concluded by noting that individual animals on land can be identified in a surprising number of species, including the elephant, jaguar, rhinoceros, seal,

[2] AdaBoost, short for adaptive boosting, is a method of machine learning in which a number of weak, inaccurate classifiers are combined to make a more highly accurate prediction rule (Freund and Shapire, 1997).

whale shark, wild dog, and nautilus. He suggested that the techniques used for recognizing individuals in such cases may be useful in the underwater context as well. He recognized, though, that computer vision techniques cannot be adapted to the underwater fish environment without considering a number of important, and potentially messy, technical details. He also emphasized the use of contests as powerful tools to motivate the study of well-defined problems with clear data sets.

5

Multi-Object Tracking

The fourth workshop session focused on multi-object tracking, including information such as extracting species-specific characteristics, minimizing double counting, and species-specific parameterization. Several of the presentations addressed domain areas distinct from traditional areas of fisheries research—including image analysis of bats, human crowds, and bees—to draw similarities between fields and examine areas in which capabilities can be leveraged from other fields to fisheries research. Some useful references for this session's topics, as suggested by the workshop program committee, include Schell and Linder, 2006; Schell et al., 2004; and Yilmaz et al., 2006. The session was chaired by Mubarak Shah (University of Central Florida), with presentations by Margrit Betke (Boston University), Mubarak Shah, Jules Jaffe (University of California, San Diego), and Ashok Veeraraghavan (Rice University).

MULTI-OBJECT MULTI-VIEW TRACKING

Margrit Betke, Boston University

Margrit Betke explained that she works with computer vision applied to the three-dimensional (3D) multi-object tracking and analysis of bat and bird flight. This is a collaborative effort that includes experts from geography, engineering, computer science, and biology. Betke specifically studies the Brazilian free-tailed bat, whose population may have dropped from 150 million to 11 million in the past 50 years (Betke et al., 2008). She noted that the Brazilian free-tailed bat acts as

a pest control service, with each bat eating as many as 114 corn earworm moths in a single night, so there are economic as well as environmental reasons for desiring the bat to thrive.

Betke explained that she studies bats using thermal infrared cameras with a high spatial resolution at 131 frames per second. Using a three-camera system, she is able to develop 3D reconstructions of flight behavior. Her team has developed a protocol and calibration for multi-camera videography (Theriault et al., 2014). A calibration device with landmarks, which are easily identifiable because of their differing temperatures, is used to calibrate the space in which the bats are moving. After calibration, the two main elements to develop an object's trajectory are detection and tracking. Tracking can be further divided into both position estimation (which Betke indicated is fairly easy to accomplish) and data association/disambiguation (which Betke indicated is still quite difficult).

Betke explained that if bats are imaged at an observation distance of about 10 m, the resulting resolution is at least 10 pixels per bat. In the reported experiment, observation distances were chosen so that the reconstructed 3D positions had uncertainties less than 10 cm, the approximate length of a single bat; at a 10 m distance, the measured uncertainty in the 3D projection was 7.8 cm.

Betke explained that objects are tracked in 3D space with two-dimensional (2D) measurements through two possible methods of path reconstruction:

1. *Reconstruction-tracking method.* 3D positions are first reconstructed from multiple views, then a tracking approach is applied using feature-to-feature fusion. In other words, the correspondence of detected objects is found first across views, then across time.
2. *Tracking-reconstruction method.* 2D tracking in each view is applied independently, followed by the reconstruction of 3D trajectories through track-to-track fusion. In other words, the correspondence of detected objects is found first across time, then across views.

Betke explained that her group's tracking approach couples object detection with position estimation and data association (Wu et al., 2012). This method, essentially a multi-dimensional optimization problem, provides additional flexibility and scalability, is less sensitive to initialization, and does not constrain the pixel resolution of the target.

Betke then turned to the study of group behavior. Brazilian free-tailed bats, she explained, exhibit columning behavior when they emerge from their daytime roosts, which gives rise to behavioral questions about how bats fly together in the column (including questions such as distance between bats, relative positioning, flight speed variations, and avoidance). Her studies indicate that bats do not like to fly above one another, are adept at avoiding one another without needing to

change velocity, and tend to fly at roughly the same velocity regardless of their location within the column.

Betke also discussed applying trackers to pedestrian traffic, or appearance-based tracking (Wu et al., 2013; Bai et al., 2013). In appearance-based tracking, objects are divided into patches, and the motion of each patch is observed and monitored. This approach works for both stationary and moving cameras. Pedestrian data tend to contain a significant number of occlusions, making the task more difficult. Betke explained that she and her students also visited the New England Aquarium to record fish with their three-camera visible-light high-speed video-recording system; at this point they had merely conducted census studies, which were fairly accurate: the number of fish counted in each of the three camera views mostly matched the number of fish known to be in the tank.

Betke concluded by stressing the importance of computer vision in collecting and analyzing multi-camera video data sets. She stated that her software tool for planning field experiments (Theriault et al., 2014) might be helpful to the fisheries community. She emphasized the importance of estimating uncertainty in measurements for different target distances and angles between the optical axes of the cameras used, and indicated that the appearance-based techniques her group has developed for pedestrian tracking may be helpful with fisheries species classification.

CROWD TRACKING AND GROUP ACTION RECOGNITION

Mubarak Shah, University of Central Florida

Mubarak Shah explained that his work focuses on high-density, crowded scenes, such as festivals, marathons, marches, and other instances of high-density, moving people. The work encompasses counting, localization, tracking, and characterizing crowd behavior. Shah noted that counting is particularly important, as there is often a desire to know such information as the number of people that participate in rallies, how many people attend concerts, or the volume of commuting traffic. Counting requires fusing information from multiple sources, and there is demand for improved automation. However, the reliability of current methods is limited by low-resolution imagery, occlusion, foreshortening, and variations in perspectives.

Shah gave an overview of a framework for analyzing crowds (Idrees et al., 2013). Three steps—head detection, Fourier analysis, and interest-point-based counts—are used in combination to obtain patch counts. Head detection is used for counting because bodies are often obscured in crowd images. Shah explained that a human detection model is trained on an existing data set so it can recognize heads. Fourier analysis is used when the head size is too small; as crowds often contain

a periodic recurrence of heads, Fourier analysis can identify that periodicity, said Shah. Interest-point-based counting is primarily conducted using a Scale Invariant Feature Transform. The three methods are used in concert for enhanced accuracy.

Shah noted that, in general, overall crowd count is fairly accurate, but each person is not localized in this system. He noted that the system produces many false positives as well as some inconsistent scaling (in which people in the same region of the image are not the same approximate size). He explained his research group's method of finding bounding boxes, which is useful when localization is needed. That method consists of the following steps: detect a human (using a combination-of-parts model), apply scale and confidence information (information about the approximate size of a person and the confidence that detection has been made, respectively), apply a Markov random field model, iterate, and then apply global occlusion reasoning to make a final detection decision. Global occlusion reasoning, Shah explained, sets rules to help find and collect visible parts of the same person. A chain constraint is used to avoid degenerate solutions (e.g., when a head and legs are selected for the same bounding box, but the matching shoulders and abdomen are not selected for that bounding box). Shah indicated that this method of human detection compares favorably to state-of-the-art systems.

Shah then applied his methods to tracking individuals in a dense crowd (Idrees et al., 2014). There are two key elements to successfully tracking individuals:

- *Context.* A neighborhood motion concurrence model is used to understand context around an individual.
- *Prominence.* Some crowds have prominent individuals who are easy to track, referred to as "queens," who can be identified first.

Shah explained that crowd behaviors, such as bottlenecks, departures, lane formations, arches and rings, and blockings, can also be tracked and identified.

In a later discussion session, Shah noted that there appears to be a gap between what the fisheries community is doing in the areas of sensors, 3D, and metadata relative to what is traditionally done in computer vision, and, thus, there may be opportunities for future work and improvements in the approaches used for fisheries stock assessments.

TRACKING IN THE OCEAN, VEHICLES, AND FISH

Jules Jaffe, University of California, San Diego

Jules Jaffe began by stating that fish are highly maneuverable and capable of behavior that is not common in many other contexts. For instance, they turn quickly in the water with rapid acceleration rates, exerting forces on their bodies

many times the force of gravity. As a result, the required frame rates to track fish are quite high—on the order of hundreds of frames per second. While one might usually implement a Kalman filter to track sources such as these, Jaffe said, sudden movements of fish are beyond the capability of a Kalman filter, which is better suited to predictable, piecewise linear motion. A segmented track identifier may be better suited to more abrupt fish movements. Jaffe described a segmentation and fitting algorithm to develop a piecewise, least-squares fit of a track to a parametric motion model; he compared it conceptually to a spline (Schell et al., 2004; Schell and Linder, 2006). The two tracking methods were applied to data gathered via 3D sonar with several cameras.

Jaffe then described underwater vehicles under development in his laboratory. The mini-autonomous underwater explorer (mini-AUE) is a compact, self-balancing vehicle with a hydrophone and oceanographic sensors. It has a battery life of approximately 1 week and can change its vertical height. The methods of recovery (Global Positioning System [GPS], radio frequency beacon, and light-emitting diode strobe) are redundant, as are the mini-AUE's buoyancy control mechanisms (piston, burn wire with drop weight, and timed weight release). Jaffe's laboratory developed 20 mini-AUE devices for an underwater position tracking system. GPS-equipped reference buoys were stationed at the surface of the water; for each drifting mini-AUE, distance measurements were made by observing the acoustic time-of-flight from the mini-AUE to the beacons. A factor graph was used to make a state-space representation of the drifter moving in time, and it was solved using an iterative message-passing algorithm. A localization test was conducted in December 2011 using a floating hydrophone array and five reference buoys spread across a 2-km-diameter range. The localization test showed that the acoustic track matched the known GPS track with less than 2 m average error (0.1 percent of the range). A "mini-swarm" of mini-AUEs were tested in September 2013 (Mirza and Schurgers, 2009, 2012; Mirza et al., 2012, 2013; Yi et al., 2013).

Jaffe concluded by shifting to a brief discussion of his planned experiments with omni-directional stereo tracking. One hundred cameras will be inset into a 3D-printed scaffold set "cross-eyed" to obtain stereo information. He believes that this novel camera design may have useful underwater applications.

SHAPE- AND BEHAVIOR-ENCODED TRACKING

Ashok Veeraraghavan, Rice University

Ashok Veeraraghavan described three projects that are peripherally related to fisheries tracking: 2D tracking in cluttered environments, 3D tracking of multiple small targets, and small baseline tracking using light-field cameras. His research

focuses on bee tracking. Veeraraghavan explained that bees are a challenging topic of study: beehives are a highly cluttered environment, with many other bees in close proximity; bees are capable of rapid movements; and there are complex interactions between the bees. Conventional bee tracking has been conducted manually: the bees are hand-marked, which can be a labor-intensive process. Tracking insects automatically is accomplished by modeling behavior at three different levels: structural/anatomical, behavioral, and interactions. An anatomical model divides the bee into its three basic body parts (head, thorax, and abdomen) and models each as an ellipse. An anatomical model has the following constraints:

- The physical limits of body parts are consistent.
- Structural limitations are incorporated.
- Correlation is made among the orientation of body parts.
- Insects move in the direction of their head.

A behavioral model takes into account specific behaviors that bees typically exhibit. The modeling is done via a hidden Markov model, explained Veeraraghavan, although the states are manually defined by the scientists. An interaction model takes into account how bees move and interact in the hive. Veeraraghavan said that he and his coworkers specifically tracked foraging bees upon their return to the hive, observing their dance communication behaviors.

Veeraraghavan then turned to a discussion of 3D tracking of small, dim targets. The objective of the work was to observe the response of bees to visual stimuli. Bees were placed in a highly scripted environment (i.e., a staged room), where a few hundred bees were observed at a time and two fixed cameras recorded their motion. Bees were typically 20 to 50 m from a camera, each image frame contained an average of 6 to 8 bees, and each bee was represented by 5 to 10 pixels. The bees are considered a dim target because of the relatively few pixels per bee. Veeraraghavan explained that the tracking algorithms used were very simple: background subtraction to remove variations in the background, connected component analysis to link pixels belonging to the same target, and probabilistic data association. The points at which the bees turn in flight (i.e., the points of maximal 2D curvature) are used to corroborate data across the two cameras. Veeraraghavan said that no camera calibration was needed because the bees' unique trajectory curvature made it simple to correlate the cameras. Once correspondence is established, the data are triangulated to create a 3D flight path.

Veeraraghavan then turned to a discussion of light-field cameras and their application. Light-field cameras capture single-shot 3D and multi-view information by including an array of microlenses within one camera. These cameras are compact, lightweight, and easy to use. The main drawback to a light-field camera, Veeraraghavan said, is that it provides only low-resolution images. These cameras

record information about depth that can be extracted and exploited to refocus the image on different depths. Veeraraghavan used a light-field camera to image and track fish in a small fish tank to determine the feasibility of applying light-field cameras in water. He was able to demonstrate that 3D trajectories of fish targets can be extracted, even when one fish occludes another.

6

Shape and Motion Analysis

The fifth session of the workshop focused on current approaches used to classify fish and identify key parameters such as size, shape, texture, color, and motion. This workshop session also included examples from a diverse array of disciplines. Some useful references for this session's topics, as suggested by the workshop program committee, include Miller et al., 2013, 2014; Dryden and Mardia, 1998; Srivastava et al., 2011; Kurtek et al., 2012; Hsiao and Hebert, 2013; Gu et al., 2012; Zhu et al., 2008; Thayananthan et al., 2003; Belongie et al., 2002; Ochs et al., 2014; Ricco and Tomasi, 2012; and Sundaram et al., 2010. The session was chaired by Hui Cheng (SRI International) and included presentations by Elizabeth Clarke (NOAA Fisheries), Anuj Srivastava (Florida State University), Anthony Hoogs (Kitware), Hui Cheng, and Michael Miller (Johns Hopkins University).

FISH SIZE AND MORPHOLOGY

Elizabeth Clarke, NOAA Fisheries

Elizabeth Clarke stated that while it is vital to obtain accurate fish size in order to obtain corresponding age and biomass measurements, measuring fish is a time-consuming step. She explained that the weight, age, and life stage of individual fish can be estimated from the measurement of total fish length. Other morphometric

and meristic features[1] can be later measured for taxonomists, but these measurements are not done in the field, as they are too time-consuming.

Clarke explained that, currently, measurements are typically made by hand, on the deck of a vessel or in a laboratory. Automated measuring boards exist to help with on-deck measurements of large catches. While lasers also have been used from submersibles to obtain in situ measurements, stereo imaging cameras are currently considered state-of-the-art for underwater applications.

In addition to measuring length, on-deck sampling observers can measure weight, extract otoliths,[2] and obtain tissue and other biological samples. Clarke emphasized that some catch will always be needed to obtain these physical samples, no matter how well other measurements can be obtained via automated methods. In response to a later question, she indicated that real-time information would also be useful for on-deck applications, such as allowing fish catch to be quickly assessed on fishing vessels.

Clarke stated that fish come in a variety of shapes and sizes, with a corresponding variety in their visibility and ease of measurement in images. She explained that lasers had been the standard method of obtaining fish size in situ. In such a system, several lasers are mounted a known, fixed distance apart. However, it can be difficult to obtain accurate measurements via lasers: most fish are not on the bottom of the seafloor and are therefore not perpendicular to the camera. In addition, Clarke noted that fish tend to react to certain lasers: flat fish have been known to "chase" green lasers, for instance. Stereo cameras are now the de facto standard, Clarke said, although they must be carefully calibrated to correct for (1) optical distortions of the lenses and (2) epipolar geometry (i.e., the relative translation and rotation of the cameras). Commercial or "one-off" software is then used to examine images and obtain measurements. These measurements are done by hand, and Clarke emphasized the laborious nature of this task. She also explained that the morphometric and meristic measures are critical to identify fish species. Even with those observations, however, certain fish (vermillion and sunset rockfish, for example) can be distinguished only by their genetics and cannot be identified in-hand. Other fish can be very difficult to identify via photos, she said, as certain distinguishing details are not visible on an image. In discussion, a participant indicated that in an automated context, it may be easier to obtain surface area rather than length, and that may be a more robust measure to use for weight and age relationships.

Clarke concluded by noting that improved biomass estimates could corre-

[1] Morphometric and meristic features are quantitative measures of a fish, such as the number and size of fins, scales, jaw length, and eye-to-jaw ratio, that can be used by taxonomists to distinguish different species.

[2] The otolith is a structure in the inner ear of vertebrates that helps with balance and movement. In fish, markings on the otolith can be used to determine the animal's age.

spondingly improve assessment of fish stocks. She suggested that it would also be helpful to be able to distinguish geologic features for classifying habitat.

A ROLE FOR STATISTICAL SHAPE ANALYSIS IN FISHERIES STOCK ASSESSMENT

Anuj Srivastava, Florida State University

Anuj Srivastava discussed his research in the detection of the types, locations, and quantity of underwater mines. He noted that there are few mines in the images he obtained; most of what is imaged is considered clutter, such as fish or underwater debris. The first step in the detection process is to find regions of interest using a variety of machine learning algorithms. The second step is to look at a "big picture" to detect spatial patterns in the regions of interest and to model the terrain. Finally, centers of attention are examined to determine features such as shape and appearance, and the object is classified via an analysis of shape contours.

Srivastava explained that shape analysis of contours is complex and consists of two challenges:

1. *Extraction of contours.* Srivastava explained that extracting the boundaries of the segmentation may have to overcome complexities such as low image contrast, occlusion and clutter, and multiple sources.
2. *Statistical analysis of contour shapes.* Metrics for comparing shapes and models for capturing the statistical variability in the shapes are both needed, said Srivastava. A statistical model can enable the researcher to put a confidence value on any shape classification.

Srivastava explained that registration is the determination of correspondences across curves and the matching of points across curves. For every two shapes under comparison, the optimal registration must be found, and linear registration is usually suboptimal. In elastic shape analysis, registration and comparisons are performed jointly. Elastic shape analysis requires the computation of a metric to quantify the shape difference; it should be independent of the choice of comparison points and should not disturb points that are already well matched. Srivastava indicated that shape metrics are useful in clustering data, and a statistical model of the shapes can be developed with summary information on the population of shapes (information such as the mean, covariance, and principal components of the population). Srivastava said that statistical shape analysis has been successfully applied to the shape analysis of leaves. However, more tools are needed for statistical shape analysis, particularly the following:

1. *Partial shape analysis.* The full shape may not always be available.
2. *Moving beyond similarity invariance.* Objects observed from different viewing angles may have an altered or distorted perceived shape, Srivastava explained, and a similarity transformation would not be applicable. Other types of projections or transformations may be needed.
3. *Comparing populations.* Currently, individual shapes are compared to one another; Srivastava posited that it may be useful to compare population distributions, rather than merely individual shapes.

Srivastava also explained that an active contour model, in collaboration with statistical shape analysis, can be used to develop a shape. A priori information about the types of desired shapes can be given to the active contour model. He indicated that this method has proven useful when applied to an object of interest that is only several pixels in size.

Srivastava concluded by noting the importance of shape analysis to not only identify individual items, but also provide feedback to other elements in a larger detection system, such as the determination of regions of interest. In addition, several of these shape analysis methods have already been extended to include 3D objects and surfaces.

BEHAVIORAL ANALYSIS AND ACTION RECOGNITION

Anthony Hoogs, Kitware

Anthony Hoogs began by defining a paradigm in which a user can interact with a machine learning system to tell the system the type of information the user would like. He described a concept of vision algorithm generation by non-experts. The system contains a variety of potential algorithms, and it determines which algorithm(s) would best apply to a given input to obtain a desired output. The user supplies the content (i.e., an image or set of images), and the system provides the algorithm and parameters to obtain a result. Such a system would enable scientists, analysts, and other end users to extract useful content from imagery and video across heterogeneous scientific domains without requiring any additional expertise in analytic algorithms or visualization.

Hoogs explained that the normal process today calls for an expert to manually adjust between 5 and 25 algorithm parameters and choose the best values based on experience, intuition, and trial and error. However, even with a large investment of time, the optimal parameter and model configuration settings may never be found with manual tuning. In addition, the parameters do not then generalize, and each subsequent data stream requires this same, potentially lengthy, manual tuning procedure. Hoogs indicated that a better method would be to automatically

iterate the potential parameters and models until performance is acceptable. His approach is to cluster videos into similarity sets based on a set of variables, such as scene conditions or camera metadata, and learn the best configuration parameters for each set.

Hoogs also discussed automatic video surveillance. In an exemplar-based query, the user provides a video clip (such as a recording of a person bending over or performing some other specific activity) and asks the system to provide other instances of that same activity from a lengthier recording. In this surveillance application, the user provides feedback to tune the results. This system has been in development for 5 to 6 years, explained Hoogs, and it is in the process of transition to defense applications. The system first detects movement and then determines the object type (i.e., whether it is a vehicle, person, or other object). Tracks are examined at the global level (to obtain information such as the kinematic properties, object type, and how the object is changing in time). Many descriptors are computed, including track interval descriptors, motion descriptors, and relational descriptors. In low-resolution video, humans and other objects of interest may be only a few pixels in size, which adds to the challenge. Hoogs explained that statistical shape analysis proved to be the best tool for identification and tracking, and he showed results identifying instances of a specific activity, including finding people wearing backpacks and people doing cartwheels. He suggested that similar classifiers could be trained, with user input, to find fish similar to a selected one.

Hoogs concluded by emphasizing the importance of using a meta-learning approach to generate vision algorithms by the domain scientist, rather than by experts in computer vision. He stated that auto-tuning, combined with user interaction and feedback, can improve the results. He suggested that the community develop a broader biological data set suitable for developing and testing tracking algorithms. Such a set could then be used to identify the best trackers to suit the data. In a later discussion, Hoogs suggested that the underwater observing community may suffer from too much specificity in its data sets; he has observed that as many as 30 to 40 percent of research papers introduce new data sets, leading to data set competition instead of collaboration.

MULTI-CUE ENTITY DETECTION, TRACKING, AND CLASSIFICATION

Hui Cheng, SRI International

Hui Cheng began by describing a number of the challenges associated with shape, motion, and texture analysis in video for fisheries applications, such as the diverse life forms with variation in shape, size, movement, and texture as well as diverse environments with variation in lighting and background conditions. Cheng explained that one potential way to meet these challenges is through the integration

of multiple sensors and cues to cut across shape, motion, behavior, and texture. He stressed that SRI International has not investigated any underwater applications at this time.

Cheng explained that optical flow analysis, the traditional method of analyzing video, tends to blur boundaries when items are in motion. It also does not adequately address occlusion. He showed results using bilateral filtering (Xiao et al., 2006) that produces sharp, clear motion boundaries. He also said that with coarse-to-fine model-based image alignment, one high-resolution camera can be paired with lower-quality cameras without loss of quality in detection and tracking. This may contrast with the natural impulse to use the highest-resolution cameras possible.

Cheng described work in large-area aerial surveys to detect and track different types of moving vehicles. The goal is to be able to distinguish different vehicle types from one another, as well as identify and track specific vehicles (Cheng referred to the latter as "fingerprinting" the vehicle). Images are first scanned, and any objects found are indexed. The items are then matched, and the matched sets are verified to create a unique match or a shorter list of possible matches. Finally, the object is "fingerprinted," which is done in three dimensions, he explained, because there are so many possible two-dimensional (2D) renderings of each object, and 2D matching is too computationally expensive. Specific points on different vehicle models (such as corners and joints) are used as key feature locations. The key feature locations are then projected back onto the 2D image, and a matching matrix is obtained to describe the feature matching. Feature matching also provides information about what parts of the vehicle are occluded and what parts of the vehicle do not match well to the model.

Cheng then discussed the challenges associated with finding and tracking people in video imagery (Zhu et al., 2014). In a large-area survey, a person is very small on the image—typically only 6 to 8 pixels wide. In addition, the images tend to be low contrast contain sensor artifacts and incomplete information, and contain occlusions. To detect people in video, one looks for a spatial signature (size and contrast) as well as a motion signature (contrast difference, speed, and travel distance). In addition, one uses a multi-frame cue to detect changes from one frame to the next, along with measures of spatial and temporal saliency.

Cheng explained that these wide-area video analyses can also classify behaviors, or patterns of activity, using entity-centric event detection and recognition (Cheng et al., 2008). This has been applied to aerial and surveillance videos, and he noted that this type of tracking could also be applied to fish.

GEODESIC POSITIONING SYSTEMS AND HIGH-THROUGHPUT INFORMATIC BRAIN CLOUDS

Michael Miller, Johns Hopkins University

Michael Miller described the emerging discipline of computational anatomy, which applies the vernacular and formalism of computational linguistics to the quantitative analysis of biological shape (Miller and Grenander, 1998). A set of metrics is used to index and cluster biological shapes, and machine learning is then performed on those shapes. Biological shape, Miller explained, is developed by using geodesic positioning to obtain coordinates; those coordinates then describe the shape. As with computational linguistics, parsing is a critical element. In computational linguistics, parsing means dividing the data into smaller segments—that is, building correspondences between a target (such as a word) and some ontology (such as a grammar). In computational anatomy, positioning is the equivalent of parsing, and diffeomorphisms[3] are used for the transformation from the observed image and some template. Miller noted that he is particularly interested in applying computational anatomy to pediatric neurodevelopment.

Miller showed an example of the application of geodesic positioning and diffeomorphisms applied to temporal lobe dementia. Over 2 years, he imaged the temporal lobe of patients with dementia to observe the progression of the disease, and observed changes in the volumes of those lobes during that time (Miller et al., 2013, 2014).

He then explained that a geodesic positioning system provides information about both positioning and the geodesic coordinates. Geodesic coordinates describe the diffeomorphic shape; that is, the vector field at the origin determines the geodesic flow. With these coordinates, one can do machine learning and build classifiers to identify different diseases. Miller then described the application of such a system to high-throughput brain clouds. Johns Hopkins University has a database of more than 10,000 brain images, but there is no structured index associated with the images for searching them. Text-based searches are not terribly helpful, Miller pointed out, because high-quality, structured information about anatomical locations does not exist. Miller then explained that brain images can be parsed into an ontology: each brain image (consisting of 10 million or more pixels) can have its dimensional space reduced to 1,000 bits. Geodesic coordinates are then developed for each image, and machine learning is performed on the resulting coordinates. The vectors are then clustered to group images with similar characteristics. Miller showed some examples applying this technique to images of brains affected by cerebral palsy and noted that cerebral palsy has many different profiles.

[3] In mathematics, diffeomorphisms are smooth, differentiable, invertible maps that relate points on one manifold to those on another manifold, encoding all pointwise relationships.

7

Identification and Classification

The sixth workshop session focused on current approaches to identification and classification, with specific application to fisheries and other disciplines. Some useful references for this session's topics, as suggested by the workshop program committee, include Beijbom et al., 2012; Branson et al., 2014; Kumar et al., 2012; and Wah et al., 2014. David Jacobs (University of Maryland, College Park) chaired the session, with presentations made by David Kriegman (University of California, San Diego), David Jacobs, Serge Belongie (Cornell Tech), and Gunasekaran Seetharaman (Air Force Research Laboratory).

AUTOMATIC ANALYSIS OF BENTHIC REEF IMAGES

David Kriegman, University of California, San Diego

Coral coverage, David Kriegman stated, has declined enormously in the past 30 years: it has decreased by 80 percent in the Caribbean (Gardner et al., 2003), 50 percent in the Indo-Pacific (Bruno and Selig, 2007), and 50 percent in the Great Barrier Reef (De'ath et al., 2012). Corals, Kriegman explained, are simultaneously an animal, mineral, plant, and microbe: while they are technically animals, they have a calcium carbonate reef structure and contain a symbiotic relationship with different types of algae that inhabit them. Coral bleaching occurs when, due to heat stress, the coral will expel the algae inhabiting them. Coral can survive a bleaching event, and the algae can be repopulated if the temperatures stabilize, although the overall health and stability of the coral reef can suffer. In 2009, the Center for

Biological Diversity petitioned NOAA to list 83 coral species as threatened or endangered. Because abundance and trend data were virtually non-existent for most coral, however, only 23 species were considered for listing, and only 3 species were added to the endangered species list. This points to the need for additional information on coral to support its protection, said Kriegman.

In the past, coral coverage was determined by hand-counting in the field, which Kriegman explained is very time consuming. Now, more rapid data acquisition is possible via imaging, but the processing time by humans has increased, resulting in little net improvement in overall analysis time, although the photographic record is very valuable. He described a representative data survey of the Moorea Coral Reef Long-Term Ecological Research Site: 1,250 images were hand annotated with more than 250,000 annotations. This process took 6 to 9 months.

Kriegman listed the steps needed in an automatic image annotation (Beijbom et al., 2012):

- Preprocess. Resize and normalize the contrast space.
- Filter to obtain a feature vector for each image following Varma and Zisserman (2005).
- Map each pixel to a visual word.
- Create a histogram of words. This can be done at multiple scales (individual pixels or groups of pixels).
- Train a classifier on the histogram word counts. Annotations of the training data would form the class labels and be used to predict the annotation of new images.

Beijbom et al. (2012) applied this system to the Moorea labeled coral data. The system currently requires about 20 seconds per image, with a classification accuracy of 70 to 80 percent. Kriegman said that humans were also found to have an accuracy of around 80 percent in a NOAA Coral Reef Ecosystem Division study.

Kriegman then described a website, CoralNet,[1] where people can upload coral reef images, automatically annotate images, and view annotation statistics. Currently, CoralNet has 97 different data sets.

A retrospective analysis of Moorea data attempted to identify coral species. Kriegman explained that porites are the most pervasive coral genus in Moorea, the Moorea population of porites has declined precipitously in recent years: over one-third of porites was lost in one season. In 2010, the data became annotated at the species level, but the data before that was annotated only to the genus level. Beijbom trained his system on the 2010-2011 species-level hand annotations and

[1] For more information, see the CoralNet website at http://coralnet.ucsd.edu/, accessed June 13, 2014.

then looked at genus-level data from previous years to "fill in" the species-level data in prior years. Tests showed species identification to be 96 percent accurate (Beijbom et al., 2012). By examining both future and legacy data, Kriegman argued, one can help identify biological trends.

Kriegman also briefly explained that he seeks to improve the accuracy of identification by including fluorescence imaging. This may help in the identification of juvenile corals, which are typically difficult to find in a standard color image. He also described future work, which involves larger-scale use of automatic image annotation, leveraging CoralNet data, improving recognition techniques, and adding 3D reconstruction.

CLASSIFYING LEAVES USING SHAPE

David Jacobs, University of Maryland, College Park

David Jacobs explained that he uses species identification techniques to identify tree species. Leafsnap[2] is an application to help professional botanists and laypeople identify tree species in the field using only a smartphone. To use Leafsnap, the leaf needs to be photographed against a white background (such as a piece of paper); the image is then subjected to a shape-based, nearest-neighbor search. The classifier first identifies whether the image is viable for searching; this saves time and eliminates spurious queries. The classifier is trained on many examples of user-uploaded images. While the shape detection problem may seem easy, said Jacobs, high accuracy is difficult to achieve because shadows and non-uniform backgrounds can strongly affect the results.

To conduct the classification, Jacobs explained that the leaf image is first segmented on a pixel-by-pixel level. Each pixel has a 2D representation consisting of its saturation and value. (Jacobs noted that hue is not helpful in segmentation because the background generally has a greenish hue due to light reflected from leaves.) Post-processing of the image removes shadows and stems. (Jacobs noted that stems are not useful for classification because people tend to be inconsistent in how they pick the leaf off the tree.) An expectation-maximization algorithm is used to cluster the 2D representations and identify the shape. To classify the leaves, the curvature of the leaf is computed at a variety of scales; this captures both the overall leaf shape (coarse scale) along with information about serration (fine scale). Integral measures of curvature are used; Jacobs noted that this was less noisy than other, derivative-based measures. Essentially, a disk is moved around the boundary of the leaf, and the amount of leaf inside the disk (versus the amount

[2] For more information, see Leafsnap, "Leafsnap: An Electronic Field Guide," http://leafsnap.com/ accessed June 13, 2014.

of background) is recorded as the measure of curvature. A multi-scale histogram of curvature is built and used for classification. Jacobs stated that for a pool of 184 species of trees from the northeastern United States, the classifier returns the correct tree as the first choice 69 percent of the time, and the correct answer is in the top five results 93 percent of the time (Kumar et al., 2012).

He reported that Leafsnap was released in 2011 and has been downloaded more than 1 million times. He indicated an interest in expanding the leaf database beyond leaves in the northeastern United States. A similar product has been released in the United Kingdom, and he said that there is interest in developing databases for different animals and tropical plants. Jacobs also briefly described the efforts to classify dogs by breed. The system, Dogsnap, first detects the dog's face, then localizes three parts: the two eyes and the tip of the nose. Facial features are extracted and breed classification is conducted using these facial parts. Currently, the dog classification accuracy is at 60 to 70 percent. Jacobs is also working to classify birds in a developing application called Birdsnap.[3]

FINE-GRAINED VISUAL CATEGORIZATION WITH HUMANS IN THE LOOP

Serge Belongie, Cornell Tech

Serge Belongie described work that is part of a larger effort, known as Visipedia,[4] in which users can search a visual encyclopedia by image. Visipedia is designed to be the visual equivalent of Wikipedia. Visipedia is crowd-powered and relies on people to label training images. Belongie noted that some groups can be enthusiastic about a hobby related to a particular data set (such as birds, planes, hand tools, or bicycles). By engaging those groups with the data sets, text-to-image and image-to-article searches can be improved. He noted that standard classifications fail when working with a large number of related classes (such as birds); classifying a bird as an animal may not be difficult, but classifying a bird's species can be quite difficult. Difficulties arise because there are few training examples per class available, variation between classes is small, and the variation within a class is often still high. Belongie emphasized that Visipedia is a service, not a polished application.

Birds, Belongie noted, are a rich problem domain: they must be observed from a distance, move frequently, have many colors and parts, are difficult to segment,

[3] Birdsnap was under development at the time of the workshop but has since been released; for more information, see http://birdsnap.com/, accessed July 14, 2014.

[4] Visipedia is a joint project between Pietro Perona's Vision Group at Caltech and Serge Belongie's Vision Group at Cornell Tech (for more information, see "Visipedia Project," http://www.vision.caltech.edu/visipedia/, accessed June 16, 2014.

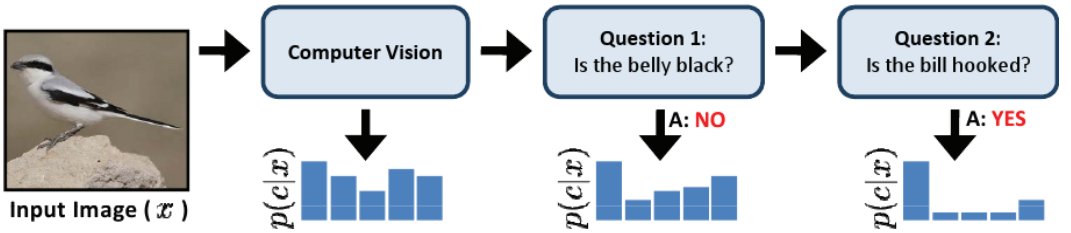

FIGURE 7.1 Schematic of the "Visual 20 Questions" game. SOURCE: S. Branson, C. Wah, F. Schroff, B. Babenko, P. Welinder, P. Perona, and S. Belongie, Visual recognition with humans in the loop, pp. 438-451 in *Computer Vision—ECCV 2010,* Lecture Notes in Computer Science, Volume 6314, Springer, Berlin Heidelberg.

and are not well described by boundaries. The Birds-200 data set[5] was the first developed by the CalTech Vision Group, and a new data set, CCUB NABirds 700,[6] is planned for release later this year and includes input from an ornithologist. Researchers ask crowdsourced volunteers to label the presence or absence of birds in images, draw bounding boxes around each bird, and label individual parts of the birds (such as breast patterns, crown color, beak shape, and wing color). Belongie explained that his group used Mechanical Turk[7] for their crowdsourcing.

Belongie said that his research group used basic, off-the-shelf computer vision methods; the novel element was adding an interactive, human component. He showed that refinement can take place by using a method referred to as "Visual 20 Questions" (see Figure 7.1), in which an image is processed using standard computer vision techniques. The results are then refined by asking questions designed to maximize the expected information gain.

Belongie briefly described perceptual embedding, in which similarity comparisons are made among existing images in a database. This avoids a weakness associated with the use of crowdsourcing: non-experts may not perform well in searching for an item for which there are not enumerated parts or for which the part names are obscure and unknown to the average user, he said. (Belongie used a mushroom as an example; most people do not know the formal names of the

[5] For more information, see "Caltech-UCSD Birds-200-2011," http://www.vision.caltech.edu/visipedia/CUB-200-2011.html, accessed June 16, 2014.

[6] For more information, see Cornell Lab of Ornithology, "CCUB NABirds 700 Dataset: Backyard Bird Edition," http://birds.cornell.edu/nabirds, accessed June 16, 2014.

[7] For more information, see the Amazon Mechanical Turk website at https://www.mturk.com/mturk/welcome, accessed June 16, 2014. Mechanical Turk is also known as MTurk.

parts of a mushroom.) Instead, by searching via similarity, where the user is asked to identify images with similar shapes or patterns, one can find the correct match without the need for any specialized terminology.

TRACKING VEHICLES IN LARGE-SCALE AERIAL VIDEO OF URBAN AREAS

Gunasekaran S. Seetharaman, Air Force Research Laboratory

Gunasekaran Seetharaman explained that the Air Force Research Laboratory (AFRL) began examining very large images of urban areas in 2003 for the purposes of tracking. The key challenges were the required wide field of view, the number of pixels on targets of interest (vis-a-vis ground-sensed-dimension), and vast variations in intensities of objects over time, in addition to more obvious factors such as the large size and very large number of potential objects in the field of view. The images were very high resolution and resulted in large file sizes that were difficult to analyze to produce timely results. Today, many more sensors are involved in imaging urban areas; lessons learned from the earlier projects with large volume and velocity of data pushed AFRL toward onboard computing.

Seetharaman cautioned that AFRL is application-driven, and its efforts tend to be governed by a pragmatic balance between assured performance and sustainability at scale relative to innovative, cutting-edge solutions that are very specific to one context. While AFRL tends to develop sustainable and affordable products, some tactical products may not use the most optimal algorithm in the literature. He then described AFRL work that would be relevant to the fisheries community. The first project in this area is Net-Centric Exploitation Tools (NCET), a tool to identify a specific action that might occur, such as dropping a bag, in video surveillance footage. NCET integrates a large network of cameras and processors to accomplish this interpretation. Another AFRL project examines data fusion techniques for integrating information from multiple cameras and sensors and prioritizing information from the most relevant sensor.

Seetharaman indicated that these systems return a deluge of data, resulting from the following:

- Increased sensor resolution.
- Improved deployment methods, including the harvesting of data from deployed systems.
- New methods of data acquisition that result in increased resolution.
- Complex operational environments, such as when a sensor is deployed but not accessible.
- Multi-spectral and hyper-spectral imaging. Seetharaman noted that this may push the current onboard computing techniques forward.

Seetharaman described the principles of acquiring high-numerical-aperture imaging through an array of identical cameras with specifically derived relative positions between them. He shared AFRL's experience with prototyping a sensor named Angel Fire. Angel Fire obtains its wide field of view due to the camera array as well as the circular motion of the platform over an area persisting for hours.

He also referred to the Autonomous Real-Time Ground Ubiquitous Surveillance (ARGUS) system, which is an array of imaging sensors capable of producing gigapixel images. ARGUS can have a very large geographic area within the field of view, but it tends to be directionally limited. He described another gigapixel system (Cossairt et al., 2011) that is more omnidirectional. He also mentioned the need to incorporate multi-modal data and the collection of data from sensor networks.

Seetharaman described a method of discriminating between stationary and moving image features (i.e., separating a moving target from its background) that uses a flux tensor analysis. When combined with other techniques, the system can learn to classify in a variety of conditions, including a stationary or moving foreground, adaptively changing parameters, and dynamic backgrounds (Wang et al., 2014).

Seetharaman also described a suite of appearance-based tracking algorithms, including the following:

- *Likelihood of Features-based Tracker (LoFT)*, which is most suitable for low-frame-rate imagery, models the target with a set of color, texture, and shape feature descriptors, then computes the match likelihoods for each feature by comparing the target to a local search image through a sliding window. This method helps inject contextual information (i.e., other knowledge about the dynamics of the field of regard) into the tracking performance
- *Clustered Set of Structured Uniformly Sampled Features (CSURF) Tracker,* which computes and clusters descriptors to create a deformable, parts-based, predictive model that is robust to occlusion. In each scene, the parts-based model identifies the best matches, a target is localized, and the next scene is predicted.

Seetharaman concluded by noting the opportunity for using a network of federated sensors to allow for the collection of high-resolution data in remote and widespread areas. He emphasized that multi-target detection and tracking will remain a challenge. He noted that these flux tensors, when combined with split Gaussian mixtures methods, handle several of these problems in practical applications.

8

Strategies Going Forward

CONCLUDING REMARKS

In the final workshop session, each session chair was charged with highlighting some of the next steps that were discussed during each session's workshop presentations and discussions. Each session chair was asked to describe what the different disciplines (biology, computer vision, and visual analytics) should do for one another. Benjamin Richards suggested that the computer vision community provide a set of integrated software tools to extract biological data needed from existing image data; he also suggested that it develop a set of hierarchical classifiers. Mubarak Shah said that computer vision researchers should learn about the types of existing data and problems in the fisheries community. Chuck Stewart suggested that fisheries researchers identify a basic set of problems at the right scale and engage all user groups to solve it. David Jacobs emphasized the need for using real-world data, to see what can be done with imperfect tools. Hui Cheng suggested that the fisheries community prioritize a list of problems that need to be addressed.

The question was then posed to the larger audience. One participant proposed making the algorithms smart enough to understand when they do not apply well to the image under consideration; otherwise, the fisheries community may not realize when an algorithm does not match well to a given set of circumstances. Another participant emphasized the importance of dual novelty: a problem must be novel in and useful to both the fisheries community and the computer vision community for both communities to want to engage in research. Another participant suggested returning the level of uncertainty with a result, which is as important as the result

of a classifier; many current fisheries identification and classification tools do not provide a probability, only a result. Shah noted that the most common classifier, the support vector machine, provides a confidence interval.

A participant considered the potential utility of prizes and asked how these were used in the computer vision community. Computer vision does have a history of using prizes to motivate research in certain topics or on certain data sets. The cost of developing the tools to run a competition, someone noted, is far greater than the prize money itself. Another participant noted the challenge of creating the right data set in a competition; for example, if there is any bias in the data set, it can be exploited to obtain the best results, and then the results do not generalize to other data sets. Another participant cautioned that prizes do not draw a representative sampling of researchers, because established researchers may hesitate to risk their reputation in a competition.

Another participant suggested that the community focus on "low-hanging fruit"—that is, projects that would have a high likelihood of success without a large investment in resources. An example of this, the participant suggested, might be Research Experiences for Undergraduates projects.[1]

OTHER WORKSHOP THEMES

Several additional topics were discussed at the workshop on different occasions but not in the final concluding session. Other discussion items that were addressed by multiple speakers or participants during the course of the workshop include the following:

- *Partial automation.* Several speakers from the fisheries community, including Dvora Hart, Allan Hicks, Benjamin Richards, and Elizabeth Clarke, emphasized to the workshop participants that making manual measurements of fish abundance, size, and taxonomy is a laborious process that consumes human capital; the automation of such tasks would be very valuable to the fisheries community. Several participants, including Dvora Hart and Hanumant Singh, stated that even partial automation would be helpful—for example, the automatic separation of underwater images into those with fish and those with no fish, because subsequent processing need not be performed on the latter set.
- *Absolute abundance measurements.* Allan Hicks, Dvora Hart, and other participants from the fisheries community also stressed the importance of

[1] The National Science Foundation funds research opportunities for undergraduates through its Research Experiences for Undergraduates projects. For more information, see "NSF REU Program Overview Home Page," http://www.nsf.gov/crssprgm/reu/, accessed September 26, 2014.

tools and techniques to obtain accurate measures of absolute abundance, rather than the relative measures that are more commonly available today.

- *Data sets.* Several speakers and participants, including Clay Kunz and Anthony Hoogs, suggested increasing the visibility of the types of problems that currently exist in the fisheries community, because computer vision experts are unlikely to be aware of these data sets and research questions. Anthony Hoogs and Clay Kunz indicated some difficulty in the current access to fisheries data; there is no central repository for data from which computer vision and visual analytics experts can familiarize themselves with the types of data and types of problems in the fisheries community. In addition, Hoogs noted that lack of easy access to data can lead to competition, rather than collaboration, for data sets.
- *Tools.* Benjamin Richards and other participants stressed the need for an integrated kit of video and image processing tools that could be applied to the fisheries community. Other participants noted that many such tools are under development and are available in open source. However, maintaining such toolkits, some said, is a substantial task that would need to be formally addressed. Anthony Hoogs proposed the development of meta-learning computer vision tools to enable scientists who are not experts in computer vision or visual analytics to generate and apply different computer vision algorithms to their data.
- *Engaging the community.* Participants discussed novel ways to engage the user community in efforts to annotate images and track fish to add information to computer vision data training sets. Serge Belongie and Benjamin Richards both described efforts in crowdsourcing to hand annotate images, and Concetto Spampinato discussed an online game in which users identify fish in still images. In addition, concrete plans were made to hold a computer vision workshop on the topic of integrating computer vision and aquaculture operations in 2015.[2]

[2] See the website for the First Workshop on Automated Analysis of Video Data for Wildlife Surveillance, January 9, 2015, in Waikaloa Beach, Hawaii, at http://marineresearchpartners.com/avdws2015/Home.html, accessed September 22, 2014.

References

Armstrong, R., H. Singh, J. Torres, R. Nemeth, A. Can, C. Roman, R. Eustice, L. Riggs, and G. Garcia-Moliner. 2006. Characterizing the deep insular shelf coral reef habitat of the Hind Bank Marine Conservation District (US Virgin Islands) using the Seabed Autonomous Underwater Vehicle. *Continental Shelf Research* 26(2):194-205.

Atrey, P.K., M.A. Hossain, A. El Saddik, and M.S. Kankanhalli. 2010. Multimodal fusion for multimedia analysis: A survey. *Multimedia Systems* 16(6):345-379.

Bai, Q., Z. Wu, S. Sclaroff, M. Betke, and C. Monnier. 2013. Randomized ensemble tracking. Pp. 2040-2047 in *Proceedings of the IEEE International Conference on Computer Vision (ICCV).* http://ieeexplore.ieee.org/Xplore/home.jsp.

Beijbom, O., P.J. Edmunds, D.I. Kline, B.G. Mitchell, and D. Kriegman. 2012. Automated annotation of coral reef survey images. Pp. 1170-1177 in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* http://ieeexplore.ieee.org/Xplore/home.jsp.

Belongie S., J. Malik, and J. Puzicha. 2002. Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24(4):509-522.

Betke, M., D.E. Hirsh, N.C. Makris, G.F. McCracken, M. Procopio, N.I. Hristov, et al. 2008. Thermal imaging reveals significantly smaller Brazilian free-tailed bat colonies than previously estimated. *Journal of Mammalogy* 89(1):18-24.

Branson, S., C. Wah, F. Schroff, B. Babenko, P. Welinder, P. Perona, and S. Belongie. 2010. Visual recognition with humans in the loop. Pp. 438-451 in *Computer Vision—ECCV 2010,* Lecture Notes in Computer Science, Volume 6314. Springer, Berlin Heidelberg.

Branson, S., G. Horn, C. Wah, P. Perona, and S. Belongie. 2014. The ignorant led by the blind: A hybrid human-machine vision system for fine-grained categorization. *International Journal of Computer Vision (IJCV)* 108(1-2):3-29.

Bruno, J.F., and E.R. Selig. 2007. Regional decline of coral cover in the Indo-Pacific: Timing, extent, and subregional comparisons. *PLOS ONE* 2(8):e711.

Cadima, E.L. 2003. *Fish Stock Assessment Manual.* FAO Fisheries Technical Paper No. 393. Food and Agriculture Organization of the United Nations, Rome, Italy.

Caimi, F.M., and F.R. Dalgleish. 2010. Performance considerations for continuous-wave and pulsed laser line scan (LLS) imaging systems. *Journal of the European Optical Society-Rapid Publications* 5:10020S.

Cappo, M., E.S. Harvey, and M. Shortis. 2006. Counting and measuring fish with baited video techniques—An overview. Pp. 101-114 in *Australian Society for Fish Biology Workshop Proceedings*. http://www.asfb.org.au/pubs/.

Carson, C., S. Belongie, H. Greenspan, and J. Malik. 2002. Blobworld: Image Segmentation Using Expectation-Maximization and Its Application to Image Querying. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24(8):1026-1038.

Chen, H., S. Huang, and H.L. Bart. 2006. Taxonomy in fish species complexes: A role for multimedia information. Pp. 475-479 in *IEEE Eighth Workshop on Multimedia Signal Processing 2006*. http://ieeexplore.ieee.org/Xplore/home.jsp.

Cheng, H., C. Yang, F. Han, and H. Sawhney. 2008. HO2: A New Feature for Multi-Agent Event Detection and Recognition. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW08)*. http://www.sri.com/work/publications/ho2-new-feature-multi-agent-event-detection-and-recognition.

Clarke, M.E., N. Tolimieri, and H. Singh. 2009. Using the Seabed AUV to assess populations of groundfish in untrawlable areas. Pp. 357-372 in *The Future of Fisheries Science in North America, Fish and Fisheries Series, Volume 31* (R.J. Beamish and B.J. Rothschild, eds.). Springer Science + Business Media B.V., Dordrecht, Netherlands. http://www.springer.com/series/5973.

Cossairt, O.S., D. Miau, and S.K. Nayar. 2011. Gigapixel Computational Imaging. Paper presented at IEEE International Conference on Computational Photography. http://ieeexplore.ieee.org/Xplore/home.jsp.

Dalgleish, F.R., and F.M. Caimi. 2011. Laser line scanners for undersea imaging applications. Chapter in *Handbook of Optical and Laser Scanning* (G.E. Stutz and G.F. Marshall, eds.). CRC Press, Taylor and Francis Group, Boca Raton, Fla.

Dalgleish, F.R., F.M. Caimi, W.B. Britton, and C.F. Andren. 2009. Improved LLS imaging performance in scattering-dominant waters. *Proceedings of the SPIE*. Paper 7317. http://spie.org/x1848.xml.

Dalgleish, F.R., A.K. Vuorenkoski, G. Nootz, B. Ouyang, and F.M. Caimi. 2011. Experimental imaging performance evaluation for alternate configurations of undersea pulsed laser serial imagers. *Proceedings of the SPIE*. Paper 8030. http://spie.org/Publications/Proceedings/Paper/10.1117/12.888640.

Dawkins, M.D., C.V. Stewart, S. Gallager, and A. York. 2013. Automatic Scallop Detection in Benthic Environments. Paper presented at the IEEE Workshop on Applications of Computer Vision. http://ieeexplore.ieee.org/Xplore/home.jsp.

Dawkins, R. 1996. *Climbing Mount Improbable.* Norton Press, New York.

De'ath, G., K.E. Fabricius, H. Sweatman, and M. Puotinen. 2012. The 27-year decline of coral cover on the Great Barrier Reef and its causes. *Proceedings of the National Academy of Sciences* 109(44):17995-17999.

De Robertis, A., J.S. Jaffe, and M.D. Ohman. 2000. Size-dependent visual predation risk and the timing of vertical migration in zooplankton. *Limnology and Oceanography* 45(8):1838-1844.

Dollar, P., R. Appel, S. Belongie, and P. Perona. 2014. Fast feature pyramids for object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* In submission.

Dryden, I., and K. Mardia. 1998. *Statistical Shape Analysis.* John Wiley & Sons, Hoboken, N.J.

Freund, Y., and R.E. Shapire. 1997. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences* 55(1):119-139.

Gardner, T.A., I.M. Côté, J.A. Gill, A. Grant, and A.R. Watkinson. 2003. Long-term region-wide declines in Caribbean corals. *Science* 301(5635):958-960.

Gong, Z., M. Andrews, S. Jagannathan, R. Patel, J.M. Jech, N.C. Makris, and P. Ratilal. 2010. Low-frequency target strength and abundance of shoaling Atlantic herring *Clupea Harengus* in the Gulf of Maine during the Ocean Acoustic Waveguide Remote Sensing (OAWRS) 2006 experiment. *Journal of the Acoustical Society of America* 127:104-123.

Gu, S., Y. Zheng, and C. Tomasi. 2012. Twisted window search for efficient shape localization. Pp. 167-173 in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. http://ieeexplore.ieee.org/Xplore/home.jsp.

Hsiao, E., and M. Hebert. 2013. Gradient Networks: Explicit Shape Matching Without Extracting Edges. Paper presented at AAAI Conference on Artificial Intelligence. http://www.aaai.org/Conferences/AAAI/aaai.php.

Huang, P.X., B.J. Boom, and R.B. Fisher. 2012. Underwater live fish recognition using a balance-guaranteed optimized tree. Pp. 422-433 in *Lecture Notes in Computer Science Volume 7724: Eleventh Asian Conference on Computer Vision*. Springer, Berlin Heidelberg.

Idrees, H., I. Saleemi, C. Seibert, and M. Shah. 2013. Multi-source multi-scale counting in extremely dense crowd images. Pp. 2547-2554 in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. http://ieeexplore.ieee.org/Xplore/home.jsp.

Idrees, H., N. Warner, and M. Shah. 2014. Tracking in dense crowds using prominence and neighborhood motion concurrence. *Image and Vision Computing* 32(1):14-26.

Jagannathan, S., I. Bertsatos, D. Symonds, T. Chen, H. T. Nia, A. Jain, et al. 2009. Ocean Acoustic Waveguide Remote Sensing (OAWRS) of marine ecosystems. *Marine Ecology Progress Series* 395:137-160.

Jagannathan, S., B. Horn, P. Ratilal, and N. Makris. 2011. Force estimation and prediction from time-varying density images. *IEEE Transactions on Patterns Analysis and Machine Intelligence* 33(6):1132-1146.

Jain, A.D., A. Ignisca, D.H. Yi, P. Ratilal, and N.C. Makris. 2013. Feasibility of Ocean Acoustic Waveguide Remote Sensing (OAWRS) of Atlantic cod with seafloor scattering limitations. *Remote Sensing* 6:180-208.

Kaeli, J.W., and H. Singh. In review. Illumination and attenuation correction techniques for underwater robotic optical imaging platforms. *IEEE Journal of Oceanic Engineering,* http://www.whoi.edu/cms/files/kaeli_joe14_InReview_183164.pdf.

Kimura, D.K., and D.A. Somerton. 2006. Review of statistical aspects of survey sampling for marine fisheries. *Reviews in Fisheries Science* 14:245-283.

Kumar, N., P.N. Belhumeur, A. Biswas, D.W. Jacobs, W.J. Kress, I.C. Lopez, and J.V. Soares. 2012. Leafsnap: A computer vision system for automatic plant species identification. Pp. 502-516 in *Computer Vision–ECCV 2012*. Springer, Berlin Heidelberg.

Kunz, C., and H. Singh. 2013. Map building fusing acoustic and visual information using autonomous underwater vehicles. *Journal of Field Robotics* 30(5):763-783.

Kurtek, S., A. Srivastava, E. Klassen, and Z. Ding. 2012. Statistical modeling of curves using shapes and related features. *Journal of American Statistical Association* 107(499):1152-1165.

Lee, Y.J., A.A. Efros, and M. Hebert. 2013. Style-aware mid-level representation for discovering visual connections in space and time. Pp. 1857-1864 in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. http://ieeexplore.ieee.org/Xplore/home.jsp.

Mace, P.M., N.W. Bartoo, A.B. Hollowed, P. Kleiber, R.D. Methot, S.A. Murawski, J.E. Powers, and G.P. Scott. 2001. *Marine Fisheries Stock Assessment Improvement Plan.* Report of the National Marine Fisheries Service National Task Force for Improving Fish Stock Assessments. NOAA Technical Memorandum NMFS-F/SPO-56. National Marine Fisheries Service, NOAA, Silver Spring, Md.

Makris, N.C., P. Ratilal, D. Symonds, S. Jagannathan, S. Lee, and R. Nero. 2006. Fish population and behavior revealed by instantaneous continental-shelf-scale imaging. *Science* 311:660-663.

Makris, N.C., P. Ratilal, S. Jagannathan, Z. Gong, M. Andrews, I. Bertsatos, O.R. Godo, R. Nero, and J.M. Jech. 2009. Critical population density triggers rapid formation of vast oceanic fish shoals. *Science* 323:1734-1737.

Mallet, D., and D. Pelletier. 2014. Underwater video techniques for observing coastal marine biodiversity: A review of sixty years of publications (1952-2012). *Fisheries Research* 154:44-62.

Methot, Jr., R.D., and C.R. Wetzel. 2013. Stock synthesis: A biological and statistical framework for fish stock assessment and fishery management. *Fisheries Research* 142:86-99.

Miller, M.I., and U. Grenander. 1998. Computational anatomy: An emerging discipline. *Quarterly of Applied Mathematics* LVI(4):617-694.

Miller, M.I., A.V. Faria, K. Oishi, and S. Mori. 2013. High throughput neuroinformatics. *Frontiers in Informatics* 7:31.

Miller, M.I., A. Trouve, and L. Younes. 2014. Diffeomorphometry and geodesic positioning systems in human anatomy. *Technology* 2(1):36.

Mirza, D., and C. Schurgers. 2009. Collaborative Tracking in Mobile Underwater Networks. Paper presented at ACM International Workshop on Underwater Networks (WUWNET09). http://dl.acm.org/.

Mirza, D., and C. Schurgers. 2012. On the performance of range-based Bayesian tracking. *IEEE Communications Letters* 16(7):1129-1132.

Mirza, D., C. Schurgers, and R. Kastner. 2012. Real-Time Collaborative Tracking for Underwater Networked Systems. Paper presented at ACM International Workshop on Underwater Networks (WUWNET12). http://dl.acm.org/.

Mirza, D., P. Roberts, J. Yi, C. Shurgers, R. Kastner, and J. Jaffe. 2013. Energy-Efficient Signaling Strategies for Tracking Mobile Underwater Vehicles. Paper presented at 2013 IEEE International Underwater Technology Symposium. http://ieeexplore.ieee.org/Xplore/home.jsp.

NOAA Fisheries. 2012. *Stock Assessment: The Core of Fisheries Science*. NOAA Office of Science Fact Sheet. February. http://www.st.nmfs.noaa.gov/Assets/science_program/AsmtFactSheet_Feb2012.pdf.

Ochs, P., J. Malik, and T. Brox. 2014. Segmentation of moving objects by long term video analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36(6):1187-1200.

Ricco, S., and C. Tomasi. 2012. Dense Lagrangian motion estimation with occlusions. Pp. 1800-1807 in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. http://ieeexplore.ieee.org/Xplore/home.jsp.

Roberts, P.L.D., J.S. Jaffe, and M.M. Trivedi. 2009. A multiview, multimodal fusion framework for classifying small marine animals with an opto-acoustic imaging system. Pp. 1-6 in *Proceedings of a Workshop on Applications of Computer Vision*. http://ieeexplore.ieee.org/Xplore/home.jsp.

Sale, P.F. 1997. Visual census of fishes: How well do we see what is there? *Proceedings of the 8th International Coral Reef Symposium* 2:1435-1440.

Schell, C., and S.P. Linder. 2006. Experimental evaluation of tracking algorithms used for the determination of fish behavioral statistics. *IEEE Journal of Oceanic Engineering* 31(3):672-684.

Schell, C., S.P. Linder, and J.R. Zeidler. 2004. Tracking highly maneuverable targets with unknown behavior. *Proceedings of the IEEE* 92(3):558-574.

Shortis, M.R., M. Ravanbakskh, F. Shaifat, E.S. Harvey, A. Mian, J.W. Seager, P.F. Culverhouse, D.E. Cline, and D.R. Edgington. 2013. A review of techniques for the identification and measurement of fish in underwater stereo-video image sequences. *Proceedings of SPIE 8791, Videometrics, Range Imaging, and Applications XII; and Automated Visual Inspection* 8791:87910G. doi:10.1117/12.2020941.

Singh, H., C. Roman, O. Pizarro, R. Eustice, and A. Can. 2007. Towards high resolution imaging from underwater vehicles. *International Journal of Robotics Research* 26(1):55-74.

Spampinato, C., and S. Palazzo. 2012. Enhancing Object Detection Performance by Integrating Motion Objectness and Perceptual Organization. Paper presented at the 21st International Conference on Pattern Recognition (ICPR). http://ieeexplore.ieee.org/Xplore/home.jsp.

Spampinato, C., Y.-H. Chen-Burger, G. Nadarajan, and R.B. Fisher. 2008. Detecting, tracking and counting fish in low quality unconstrained underwater videos. *Proceedings of the Third International Conference on Computer Vision Theory and Applications (VISAPP 08)*. http://www.visigrapp.org/BooksPublished.aspx.

Spampinato, C., D. Giordano, R. Di Salvo, Y.-H. Chen-Burger, R.B. Fisher, and G. Nadarajan. 2010. Automatic fish classification for underwater species behavior understanding. Pp. 45-50 in *Proceedings of the First ACM International Workshop on Analysis and Retrieval of Tracked Events and Motion in Imagery Streams*. http://dl.acm.org/.

Spampinato, C., S. Palazzo, and I. Kavasidis. 2014. A texton-based kernel density estimation approach for background modeling under extreme conditions. *Computer Vision and Image Understanding* 122:74-83.

Sparre, P., and S.C. Venema. 1992. *Introduction to Tropical Fish Stock Assessment*. Food and Agriculture Organization of the United Nations, Paris, France.

Srivastava, A., E. Klassen, S. Joshi, and I. Jermyn. 2011. Shape analysis of elastic curves in Euclidean spaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33(7):1415-1428.

Sundaram, N., T. Brox, and K. Keutzer. 2010. Dense point trajectories by GPU-accelerated large displacement optical flow. Pp. 438-451 in *Computer Vision—ECCV 2010,* Lecture Notes in Computer Science, Volume 6314. Springer, Berlin Heidelberg.

Thayananthan, A., B. Stenger, P.H.S. Torr, and R. Cipolla. 2003. Shape context and chamfer matching in cluttered scenes. Pp. 127-133 in *IEEE Proceedings of the 2003 Conference on Computer Vision and Pattern Recognition (CVPR)*. http://ieeexplore.ieee.org/Xplore/home.jsp.

Theriault, D.H., N.W. Fuller, B.E. Jackson, E. Bluhm, D. Evangelista, Z. Wu, M. Betke, and T.L. Hedrick. 2014. A protocol and calibration method for accurate multi-camera field videography. *Journal of Experimental Biology* 217:1843-1848.

Tolimieri, N., M.E. Clarke, H. Singh, and C. Goldfinger. 2008. Evaluating the SeaBED AUV for monitoring groundfish in untrawlable habitat. Pp. 129-141 in *Marine Habitat Mapping Technology for Alaska* (J.R. Reynolds and H.G. Greene, eds.). doi:10.4027/mhmta.2008.09.

Treibitz, T., Y.Y. Schechner, C. Kunz, and H. Singh. 2012. Flat refractive geometry. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34(1):51-65.

Varma, M., and A. Zisserman. 2005. A statistical approach to texture classification from single images. *International Journal of Computer Vision* 62(1/2):61-81.

Wah, C., G. Horn, S. Branson, S. Maji, P. Perona, and S. Belongie. 2014. Similarity comparisons for interactive fine-grained categorization. Pp. 859-866 in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. http://ieeexplore.ieee.org/Xplore/home.jsp.

Wang, R., F. Bunyak, G. Seetharaman, and K. Palaniappan. 2014. Static and Moving Object Detection Using Flux Tensor with Split Gaussian Models. Paper presented at the IEEE Workshop on Change Detection. http://ieeexplore.ieee.org/Xplore/home.jsp.

Western Pacific Regional Fishery Management Council. 2004. *Coral Reef Fish Stock Assessment Workshop. Interim Final Panel Report*. Honolulu, Hawaii.

Williams, K., C. Rooper, and J. Harms. 2010. *Report of the National Marine Fisheries Service Automated Image Processing Workshop*. NOAA Tech Memo NMFS-F/SPO-121. http://www.pifsc.noaa.gov/pubs/techpub_date.php.

Wu, Z., A. Thangali, S. Sclaroff, and M. Betke. 2012. Coupling detection and data association for multiple object tracking. Pp. 1948-1955 in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. http://ieeexplore.ieee.org/Xplore/home.jsp.

Wu, Z., J. Zhang, and M. Betke. 2013. Online Motion Agreement Tracking. Paper presented at 24th British Machine Vision Conference (BMVC). http://www.bmva.org/bmvc/2013/Papers/paper0063/paper0063.pdf.

Xiao, J., H. Cheng, H. Sawhney, C. Rao, and M. Isnardi. 2006. Bilateral filtering-based optical flow estimation with occlusion detection. Pp. 211-224 in *Lecture Notes in Computer Science Volume 3951: Computer Vision–European Conference on Computer Vision*. Springer, Berlin Heidelberg.

Yi, J., D. Mirza, C. Schurgers, and R. Kastner. 2013. Joint time synchronization and tracking for mobile underwater systems. Paper presented at the Eighth ACM International Conference on Underwater Networks and Systems (WUWNet13). http://dl.acm.org/.

Yilmaz, A., O. Javed, and M. Shah. 2006. Object tracking: A survey. *ACM Computing Surveys* 38(4).

Zhu, Q., L. Wang, Y. Wu, and J. Shi. 2008. Contour Context Selection for Object Detection: A Set-to-Set Contour Matching Approach. Paper presented at the Tenth European Conference on Computer Vision 2008 (ECCV 2008). http://www.cis.upenn.edu/~jshi/papers/contour-context-final.pdf.

Zhu, J., O. Javed, J. Liu, Q. Lu, H. Cheng, and H. Sawhney. 2014. Pedestrian detection in low-resolution imagery by learning multi-scale intrinsic motion structures (MIMS). Pp. 3510-3517 in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. http://ieeexplore.ieee.org/Xplore/home.jsp.

Appendixes

A

Registered Workshop Participants

Bajcsy, Ruzena – University of California, Berkeley
Belongie, Serge – Cornell Tech
Betke, Margrit – Boston University
Chellappa, Rama – University of Maryland, College Park
Cheng, Hui – SRI International
Clarke, Elizabeth – NOAA Fisheries
Cyr, Ned – National Oceanic and Atmospheric Administration (NOAA) Fisheries
Dalgleish, Fraser – Florida Atlantic University
Dreves, Harrison – National Research Council (NRC)
Glassman, Neal – NRC
Hart, Dvora – NOAA/Northeast Fisheries Science Center
Hero III, Alfred – University of Michigan
Hicks, Allan – NOAA Fisheries
Hoogs, Anthony – Kitware, Inc.
Howard, Rodney – NRC
Hunt, Stephanie – NOAA Fisheries
Jacobs, David – University of Maryland, College Park
Jaffe, Jules – University of California, San Diego
Kriegman, David – University of California, San Diego
Kunz, Clayton – Google, Inc.
Lapointe, George – Lapointe Consulting
Makris, Nicholas – Massachusetts Institute of Technology
Mellody, Maureen – NRC

Merrick, Richard – NOAA Fisheries
Methot, Rick – NOAA Fisheries
Michaels, William – NOAA Fisheries
Miller, Michael – Johns Hopkins University
Palaniappan, Kannappan – University of Missouri, Columbia
Richards, Benjamin – NOAA Fisheries
Roberts, Susan – NRC
Schwalbe, Michelle – NRC
Seetharaman, Guna – Air Force Research Laboratory
Shah, Mubarak – University of Central Florida
Singh, Hanumant – Woods Hole Oceanographic Institution
Spampinato, Concetto – Università de Catania, Italy
Srivastava, Anuj – Florida State University
Stahl, Jennifer – Alaska Department of Fish/Game
Stewart, Chuck – Rensselaer Polytechnic Institute
Terzopoulos, Demetri – University of California, Los Angeles
Thompson, Steven – Simon Fraser University
Veeraraghavan, Ashok – Rice University
Weidman, Scott – NRC

B

Workshop Agenda

DAY 1: MAY 16, 2014

8:00 a.m. **Welcome, Introductions, and Overview**

Opening Remarks and Meeting Overview
Rama Chellappa, University of Maryland, College Park, Workshop Planning Committee Chair
Richard Merrick, National Marine Fisheries Service

8:15 **Setting the Stage**
Session Chairs: Rick Methot, NOAA Fisheries
Rama Chellappa, University of Maryland, College Park, Workshop Planning Committee Chair

Overview of NMFS's Strategic Initiative
Benjamin Richards, Chair of the Strategic Initiative
Types of Data Used in Fishery Stock Assessments
Allan Hicks, NOAA
Overview of Computer Vision
Ruzena Bajcsy, University of California, Berkeley
Overview of Sampling in Space and Time
Steven Thompson, Simon Fraser University

Q&A and Open Discussion

10:15 **Overview of Multi-Modal Sensing**
Session Chair: Nicholas Makris, Massachusetts Institute of Technology

Fisheries Perspective of Multi-Modal Sensing
Dvora Hart, Northeast Fisheries Science Center
Synergistic Acoustic and Optic Observation and Estimation
Jules Jaffe, University of California, San Diego
Low Frequency Acoustic Imaging for Large Area Surveys
Nicholas Makris, Massachusetts Institute of Technology
Seafloor Laser Imaging Techniques
Fraser Dalgleish, Florida Atlantic University

Q&A and Open Discussion

12:15 p.m. **Lunch Keynote**
Demetri Terzopoulos, University of California, Los Angeles, to speak on artificial life simulations and the cross over with fisheries modeling

1:15 **Image Processing and Detection**
Session Chair: Chuck Stewart, Rensselaer Polytechnic Institute

Introduction to Fisheries Data Pre-Processing
Clay Kunz, Google
Underwater Robotic Platforms and Imaging
Hanumant Singh, Woods Hole Oceanographic Institution
Underwater Tele-Immersion: Potential and Challenges
Ruzena Bajcsy, University of California, Berkeley
Underwater Imaging and Detection
Chuck Stewart, Rensselaer Polytechnic Institute

Q&A and Open Discussion

3:30 **Multi-Object Tracking**
Session Chair: Mubarak Shah, University of Central Florida

Multi-Object Multi-View Tracking
Margrit Betke, Boston University
Crowd Tracking and Group Action Recognition
Mubarak Shah, University of Central Florida

Tracking in the Ocean, Vehicles and Fish
Jules Jaffe, University of California, San Diego
Shape and Behavior-Encoded Tracking
Ashok Veeraraghavan, Rice University

Q&A and Open Discussion

DAY 2: MAY 17, 2014

8:30 a.m. **Shape and Motion Analysis**
Session Chair: Hui Cheng, SRI International

Overview: How Sizing Is Currently Done in Fisheries (by hand and optically)
Elizabeth Clarke, NOAA Fisheries
Shape Analysis
Anuj Srivastava, Florida State University
Behavioral Analysis and Action Recognition
Anthony Hoogs, Kitware
DoD Work on Tracking and Classifying
Hui Cheng, SRI International
Shape and Pattern Theory into Medical Informatics
Michael Miller, Johns Hopkins University

Q&A and Open Discussion

10:45 **Identification and Classification**
Session Chair: David Jacobs, University of Maryland, College Park

Classifying Leaves Using Shape
David Jacobs, University of Maryland, College Park
Classifying Birds
Serge Belongie, Cornell Tech
Tracking Vehicles in Large-Scale Aerial Video of Urban Areas
Gunasekaran S. Seetharaman, Air Force Research Laboratory
Automatic Analysis of Benthic Reef Images
David Kriegman, University of California, San Diego

Q&A and Open Discussion

12:30 p.m. **Lunch Keynote**
Concetto Spampinato, Università di Catania (Italy), to talk about Fish4Knowledge

1:30 **Conclusions and Strategies Going Forward**

Panel Discussion:
Rama Chellappa, University of Maryland, College Park
Hui Cheng, SRI International
Ned Cyr, NOAA Fisheries
David Jacobs, University of Maryland, College Park
Nicholas Makris, Massachusetts Institute of Technology
Benjamin Richards, NOAA Fisheries
Mubarak Shah, University of Central Florida
Chuck Stewart, Rensselaer Polytechnic Institute

2:30 **Summary and Next Steps**

Comments from the Planning Committee and Sponsor
Rama Chellappa, University of Maryland, College Park, Workshop Planning Committee Chair
Ned Cyr, NOAA Fisheries

3:00 **Workshop Adjourns**

C

Acronyms

ARGUS	Autonomous Real-Time Ground Ubiquitous Surveillance
AUE	autonomous underwater explorer
AUV	autonomous underwater vehicle
BRUVS	Baited Remote Underwater Video Station
F4K	Fish4Knowledge
GPS	Global Positioning System
LLS	Laser Line Scanner
MA-ZOOPS	Multiple-Aspect Acoustic Zooplankton
MFOV	multiple field of view
NCET	Net-Centric Exploitation Tools
NMFS	National Marine Fisheries Service; informally known as "NOAA Fisheries"
NOAA	National Oceanic and Atmospheric Administration
NRC	National Research Council

OASIS	Optical and Acoustical Submersible Imaging System
OAWRS	Ocean Acoustic Waveguide Remote Sensing
SVM	support vector machine