

Classifying Sonar Images: Can a Computer-Driven Process Identify Eels?

ANNA-MARIA MUELLER*

Aquacoustics, Inc., Post Office Box 1473, Sterling, Alaska 99672, USA

TIM MULLIGAN

*Department of Fisheries and Oceans, Pacific Biological Station,
3190 Hammond Bay Road, Nanaimo, British Columbia V9T 6N7, Canada*

PETER K. WITHLER

Pacific Eumetrics, Ltd., 1533 Arbutus Drive, Nanoose Bay, British Columbia V9R 5K6, Canada

Abstract.—The purpose of this study was to explore the extent to which a computer-driven process can be used to classify sonar images. The data we present come from a feasibility study for a hydroacoustic monitoring system aimed at the automatic detection of downstream-migrating adult American eels *Anguilla rostrata* in the intake canal of a small hydroelectric station. The images were collected by a dual-frequency identification sonar with sufficient resolution to show the distinct shape and swimming motion of eels, and thus to allow confident visual identification. The goal was to find a set of image processing, tracking, and pattern recognition techniques that would reproduce the results of the visual classification. Of the three classification methods that we tested with our example data set, neural network analysis had the lowest misclassification rate for eels (7% of the eels being misclassified as debris) and the second-lowest misclassification rate for debris (5% of the debris being misclassified as eels). Discriminant function analysis misclassified 12% of the eels as debris and 4% of the debris as eels. A *K*-nearest-neighbor analysis initially provided the poorest results (17% misclassified eels and 12% misclassified debris). However, after applying an algebraic correction, *K*-nearest-neighbor analysis yielded an accurate estimate of the number of eels in the data set. We discuss the value of flagging cases of uncertain classification, how image processing and feature selection can affect the results, and how the numeric ratio of the targets present determines what error rates are acceptable. We conclude that, depending on the application, different degrees of automation may be achieved, ranging from a relatively high degree of human supervision in the classification of all potential targets to a fully automated process that requires only periodic quality control and adjustments of the classification model.

A recent breakthrough in fisheries acoustics has come with the development of DIDSON, a high-frequency multibeam sonar that uses an acoustic lens system to focus the beam (Belcher et al. 2002). The high-resolution images generated by DIDSON provide not only information on the location and movement of the target, but also on its physical size, shape, and orientation (Burwen et al. 2007b). However, despite its increasing use in fisheries applications, computer-driven analysis of DIDSON data is still in its infancy (Maxwell and Gove 2004; Tiffan et al. 2004; Burwen et al. 2007a). Often, visual review of the data has been the most effective form of processing. This is due to the fact that the human brain is extremely well equipped to process video imagery in a way that allows us to detect movement and recognize objects of different size,

shape, and brightness. However, the enormous amount of labor involved in reviewing DIDSON data visually has become a major obstacle for medium- to long-term monitoring projects. Visual review of large amounts of data are not only costly, but fatigue can also become a significant factor in the quality of the analysis. Furthermore, visual review can be subjective and therefore more difficult to standardize. Finally, some applications require analysis results in real time or near real time. All these factors have led to an increasing demand for computer-driven analysis.

For the example presented here, we use data from a feasibility study aimed at developing a hydroacoustic monitoring system for the automatic detection of downstream-migrating adult American eels *Anguilla rostrata* in the Kennebec River, Maine. Over the past two decades, a decline in the abundance of the American eel has been observed in parts of North America (Casselmann 2003; USFWS 2007). This has raised concern about injuries and mortalities that may occur when silver-phase eels pass through hydroelec-

* Corresponding author: am@aquacoustics.com

Received February 11, 2008; accepted June 6, 2008
Published online January 5, 2009

tric turbines. As a result, the Federal Energy Regulatory Commission has recently included measures to protect eels as part of the licensing requirements for hydroelectric facilities. The data presented here were collected in the intake canal of the Anson Project, a small hydroelectric facility with five turbine generator units. The goal was to determine whether a monitoring system could be developed that would automatically detect downstream-migrating eels. Such an automatic detection system could then be used to initiate the opening of waste gates that would allow eels to bypass the turbines.

Playback of the DIDSON images in movie mode allowed confident visual identification of eels out to a maximum range of 20 m. The fact that the DIDSON images had sufficient resolution to show the unique shape and swimming motion of American eels made this data set a good candidate for the development of a classification algorithm. (An animated DIDSON movie clip showing the typical eel shape and swimming motion is available at www.aquacoustics.com/proj_eels.html.) One of the challenges we faced was that the classification algorithm needs to be able to positively identify eels but reject a large number of debris targets. Note that we use the term "debris" for all targets other than eels; it includes not only macroscopic debris (e.g., leaves, sticks, or grass) but also small fish of unknown species, entrained air, and other unidentifiable targets.

The objective of our study was to find a set of image processing, tracking, and pattern recognition techniques that can reproduce the visual classification of the data. Our approach concentrated on using image processing in conjunction with tracking algorithms to detect targets and extract DIDSON image features that would quantify target size, shape, motion, and echo intensity patterns and thus capture the difference between American eels and debris.

Methods

Data collection.—The data used in this study consisted of 23 h of images recorded over the 10-d period September 25–October 4, 2005, in the intake canal of the Anson hydroelectric project on the Kennebec River. Data were collected nightly from 1730 to 0800 hours the following morning to include both night and twilight hours, when out-migrating American eels are most active. For the purpose of this study, we further narrowed our analysis to files that had been recorded in high-frequency mode (1.8 MHz) and in which we had seen at least one eel during visual inspection. The high-frequency mode uses an array of 96 beams spaced 0.3° apart and thus provides a higher cross-range resolution than the low-frequency mode

which forms 48 beams spaced 0.6° apart. We chose to analyze only high-frequency images collected over a 2–12-m range window because their higher resolution enabled us to estimate target size and shape more accurately than would have been possible with the low-frequency images collected at a longer range.

The DIDSON system was deployed from a pole mount that was bolted to the side wall of the intake canal and aimed across the canal, perpendicular to the flow. To sample as much of the vertical cross section of the canal as possible, the lens was positioned approximately 30 cm below the water surface, 40 cm away from the side wall and tilted 8° down from horizontal, which placed the top edge of the beam close to the water surface while covering most of the water column in the middle section of the canal. To maximize the coverage in the upstream–downstream dimension, the beam array was oriented with the 29° axis close to horizontal and the 14° axis close to vertical. Data were acquired at a frame rate of 7 frames/s.

Visual review.—The DIDSON image files were visually reviewed with Sound Metrics DIDSON Viewer software. The data were screened in echogram mode where targets detected on a given frame are plotted along a vertical line that is analogous to a "ping" in conventional echograms. The echograms were constructed using the following settings: process 16 beams and display the average intensity over the minimum threshold. The display intensity and threshold were set to optimize the image for adult American eels. The same settings were used for all of the files we reviewed. Every track that spanned a similar number of frames and range interval as the tracks shown in Figure 1 was considered a potential eel and further reviewed in movie mode. A target was classified as an eel when the DIDSON movie showed that it had the size (75 cm or greater), characteristic shape, and swimming motion of a large adult eel. In addition, a subsample of tracks that appeared clearly different from eels on the echogram was also reviewed in movie mode, which confirmed that these targets were not eels. Furthermore, all tracks that were subsequently classified as eels by one of the classification algorithms, but not visually on the echogram, were reinspected in movie mode.

Data processing.—Intensity corrections, image processing, and computation of frame-based features were performed in Hydra, an acoustic data processing software package that is being developed by Pacific Eumetrics, Ltd. Data were processed in the following sequence:

To correct for the loss in sound intensity due to absorption and geometric spreading, a time-varied gain was applied using

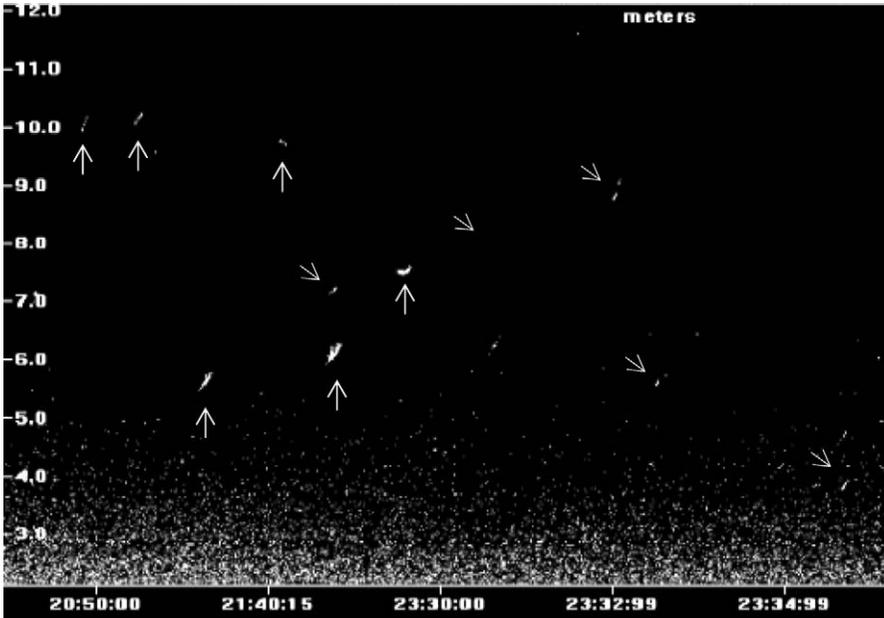


FIGURE 1.—Visual review of echograms in a Sound Metrics DIDSON Viewer. The targets indicated by arrows were considered potential eels and further reviewed in movie mode. Vertical arrows indicate targets that the movie clip confirmed to be eels, slanted arrows targets that the movie clip confirmed to be debris.

$$2\alpha R + N\log_{10}R,$$

where α is the absorption coefficient, N is the spreading coefficient, and R is the range in meters. We chose a value of 1 for α , based on the frequency, and a value of 30 for N , based on the visual appearance of the image. These settings provided an approximate compensation for the loss of intensity with range. Arguably, the value for N should be between 20, which would be appropriate for targets that fully cover individual beams, and 40, which would be appropriate for point source targets. At the range at which these data were collected, eels fully cover individual beams in the horizontal dimension but are more like point source targets in the vertical dimension.

In addition, Hydra also compensated for a reduction in intensity toward the outer beams of the array, to a maximum of 10 dB for the outermost beams. This correction was applied in the form of a smooth function which was fit to the azimuthal beam pattern measured by the manufacturer.

Next, a series of image processing steps were performed on each DIDSON frame. These are all standard techniques and are described in more detail in Seul et al. (2001).

A 5×5 averaging convolution was applied. The goal of the convolution was to help join pixels that belong to a given target into a contiguous object, while

at the same time preserving target shape and preventing adjacent targets from being merged into one object.

The image background was removed with a dynamic mean background, which was estimated with an algorithm that smoothes the data over successive frames, similar to the dynamic background subtraction implemented in the Sound Metrics DIDSON Viewer software. After subtracting the dynamic mean background from the image, a threshold of 1 machine unit (DIDSON machine units range from 0 to 255, corresponding to 0–90 dB) was applied. We refer to the resulting image as the foreground image.

A preliminary mask was generated from the foreground image by setting pixels above threshold to 1 and pixels below threshold to 0. The preliminary mask was further processed with a K-Fill filter to remove “salt and pepper” noise (Seul et al. 2001). The resulting mask was used to detect targets in the foreground image. The mask defines the regions of the image (i.e., pixels or pixel clusters) that belong to preliminary targets. A minimum target area threshold of 200 cm² was imposed to eliminate isolated pixels and very small pixel clusters, many of which reflect random variation around the mean background or artifacts of the background removal process. Further processing of the foreground image included only clusters of pixels that exceeded the area threshold and were therefore considered potential targets.

Next, Hydra computed a series of frame-based features. The term “frame-based features” is used to refer to characteristics of the target image on a given frame, as opposed to “track-based features,” which are properties of the collection of images that make up a target track. The frame-based features included (1) the area (cm^2) of a given target image; (2) the border length (cm) of a given target image; (3) the mean intensity (dB) of the pixels within a given target image; (4) the SD of the intensity of the pixels within a given target image; (5) the compactness of the target image shape, defined by $C = B^2/4\pi A$, where C is the compactness, B is the border length, and A is the area of the target image (this definition results in a circle having the smallest value obtainable, 1, while other shapes have compactness values >1); (6) the coefficient of variation of the intensity of the pixels within a given target image (CVI), defined as SD/mean ; (7) the target image centroid in the X -coordinate, which is oriented approximately parallel to the flow (positive X pointing downstream); and (8) the target centroid in range, which we refer to as the R -coordinate.

The data were then imported into SonarData Echoview (version 4.20) for target tracking with an α - β tracking algorithm (Blackman 1986). As a target moves through the beam, its image will be captured on several successive frames. The tracking algorithm attempts to group the images generated by the same target. It tests for systematic movement using the positional information from the centroids in X and R that have been calculated by Hydra in the previous step. Random clusters of pixels do not exhibit systematic movement and will therefore not form tracks. Images of targets that move systematically are grouped into tracks and assigned a track identification number.

For each track, the following set of track-based features was computed using the R programming language: (1) the mean area (the mean of the frame-based area values over the entire track); (2) the mean border length (the mean of the frame-based border length values over the entire track); (3) the mean CVI (the mean of the frame-based CVI values over the entire track); (4) the mean compactness (the mean of the frame-based compactness values over the entire track); and (5) X -velocity, which is the velocity of the target in the X direction (i.e., approximately parallel to the flow). The X -velocity is estimated with a nonparametric smoothing function (`supsmu` in the R programming language) to smooth the frame versus X -centroid data. The distance traveled by the target along this smoothed trajectory through the beam is divided by the time between the first and last frame in the track.

Feature selection.—To select features that are useful for classifying objects into categories, we used three diagnostic tools: synchronized feature echograms, histograms, and scatterplots. For displaying synchronized feature echograms, we reimported the data into Echoview (version 4.20) in the form of a single-target file in which each record represents a target image on a given frame. The single-target file incorporates the frame-based feature values that were computed previously. We then displayed a series of echograms, one for each of the features. Scrolling through sets of synchronized feature echograms, we examined feature patterns within tracks, among tracks, and with respect to range. Close examination of echograms synchronized with the original DIDSON movie helped reveal relationships between individual features and the corresponding DIDSON images.

Classification.—The type of classification procedures we used are referred to as “supervised learning” (Ripley 1996; Duda et al. 2001). These procedures require a training data set that is used to construct the mathematical model that will classify each observation. In addition, they use a test data set whose members were not included in the training data to test the performance of the model. To avoid having multiple samples from the same target, only track-based data can be used for classifying targets. In terms of statistics, the classification data are assumed to be independent and identically distributed.

Classification algorithms are designed to assign samples to their true category. Samples are considered points plotted in multidimensional feature space. If the features capture differences between the categories, then the points from separate categories tend to be located in different regions of this space. If none of these regions overlap, perfect classification can be achieved; by contrast, partially overlapping regions will result in the misclassification of some samples. Classification algorithms attempt to define region boundaries that maximize correct classification.

We tested three different classification methods (Duda et al. 2001): (1) quadratic discriminant function analysis (as implemented in the programming language S-Plus); (2) neural network (we used a three-layer, feed-forward model in which the input layer had four nodes, the middle [hidden] layer had 15 nodes, and the output layer had one node), parameterization being done by adjusting the network weights using the back-propagation training algorithm (with leave-one-out cross validation as the stopping criterion); and (3) K -nearest neighbor, with $K = 5$ (as implemented in the programming language R [Hornik 2007]).

The data set used for the classification contained a total of 187 samples: 57 American eels and 130 debris.

The performance of the discriminant function analysis and the neural network were assessed by cross validation in a leave-one-out process that used 186 samples (without replacement) for training and the remaining sample for testing. The process was repeated for each of the 187 samples in the data set (i.e., each of the 187 samples became the test sample once).

For *K*-nearest neighbor, to get unbiased results we used equal sample sizes of American eels and debris in the training data set. Otherwise, the higher density of debris samples in feature space would result in biasing the nearest neighbors toward debris and thus increasing the probability that American eels would be misclassified as debris. We chose 45 samples (without replacement) from both categories for the training set. This left a sufficient number of eels (12) and debris (85) samples in the test set.

The classification results are summarized in “confusion matrices” showing how a method assigned samples to the two categories (Duda et al. 2001). For the *K*-nearest-neighbor analysis, we also show an algebraic correction of misclassifications, which used the confusion matrix to solve algebraically for the estimated true population. First, we resampled the data 1,000 times, choosing 45 samples without replacement from each category to be the training set and keeping the remaining samples as the test set. Next, we calculated the confusion matrices for these 1,000 cases and then the resulting mean and SD for each matrix element. The confusion matrix of the means was then used to construct the corresponding set of simultaneous equations:

$$PE = \alpha_1 TE + \alpha_4 TD$$

and

$$PD = \alpha_3 TD + \alpha_2 TE,$$

where PE is the mean predicted number of American eels, PD is the mean predicted number of pieces of debris, TE is the true number of eels, TD is the true number of pieces of debris, α_1 is the mean proportion of eels correctly classified, α_2 is the mean proportion of eels incorrectly classified as debris, α_3 is the mean proportion of debris correctly classified, and α_4 is the mean proportion of debris incorrectly classified as eels.

These equations are solved to estimate TE, that is,

$$TE = \frac{\alpha_3 PE - \alpha_4 PD}{\alpha_1 \alpha_3 - \alpha_4 \alpha_2}.$$

Results

The data set contained little noise and produced good DIDSON images, which enabled the image

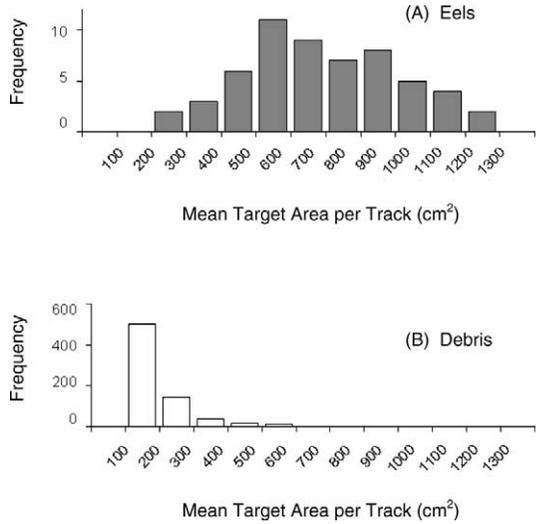


FIGURE 2.—Histograms of the mean target area per track for (A) American eels and (B) debris. Although there is some overlap in the distributions, on average eels have a larger target area than debris.

processing and tracking algorithms to perform well. Examination of the track data showed that for all eels the mean area per track was larger than 250 cm² (Figure 2), while for the majority of debris targets it was smaller. We therefore raised the mean area threshold to 250 cm², which eliminated 600 out of 730 debris targets. The truncated data set, which comprised 130 noneel and 57 eel targets, was used to generate both the training and test data sets for the classification algorithms. Figure 2 demonstrates that for debris targets the mean area per track was approximately exponentially distributed; whereas, for eels the mean area showed a unimodal distribution. We anticipated that this difference in distribution may make discriminant function analysis (which assumes that the data are multivariate normally distributed) less accurate than classification algorithms that do not assume a specific type of distribution.

The feature echograms (Figure 3) provided an excellent synopsis of the data that can be screened faster than the movie. A dynamic link between the echogram and the movie display allowed a quick review of selected movie clips when interesting targets were encountered on the echogram. The same basic concept has been implemented in the echogram mode of the Sound Metrics DIDSON Viewer. The feature echograms expanded this concept one step further by incorporating image information (such as the target area or compactness) in the form of color.

Based on our examination of the DIDSON movie

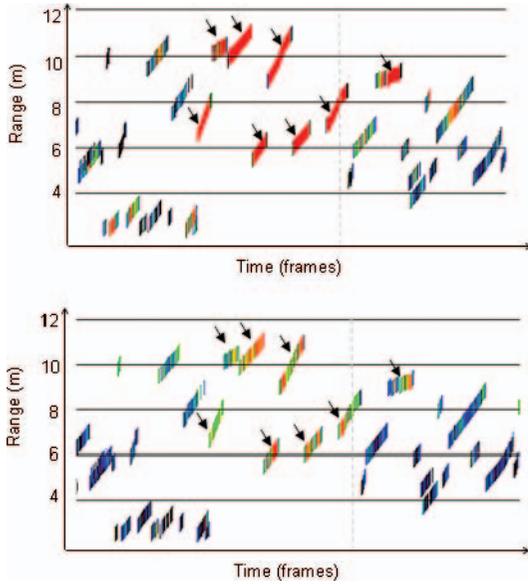


FIGURE 3.—Two Echoview echograms derived from DIDSON data, color coded by frame-based area (top panel) and frame-based compactness (bottom panel), with warmer colors representing higher values. The echograms are condensed to show only the frames in which targets were detected. Each echogram represents 675 frames, or 1.6 min of the 6 min of data collected. The height (range interval) of the echo traces was chosen arbitrarily to make the traces and their colors easy to see. Shown are eight echo traces generated by American eels (indicated by arrows) and 27 generated by debris. Note that the area and compactness values tend to be higher for eels than for debris. The dashed gray line indicates the DIDSON frame shown in Figure 4.

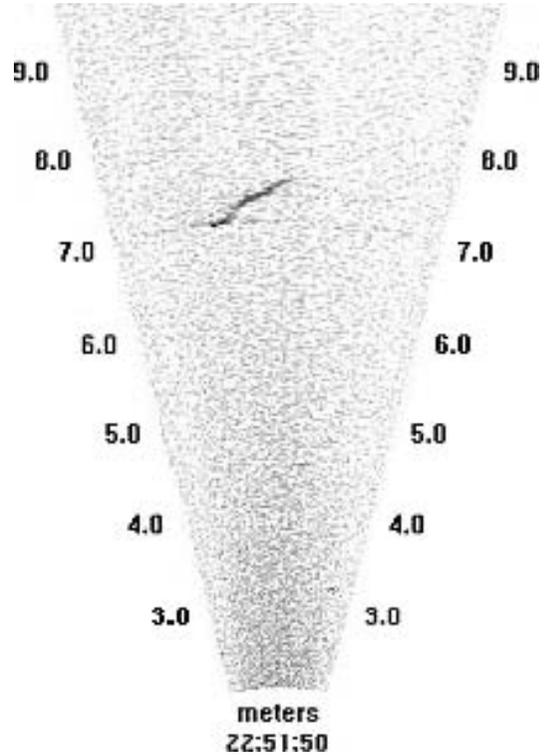


FIGURE 4.—A DIDSON image of an adult American eel. The frame shown corresponds to the one indicated by the dashed gray line on the echograms in Figure 3.

images (Figure 4), synchronized feature echograms (Figure 3), feature histograms (Figure 2), and scatterplots (Figure 5), we selected the following four features for the classification: (1) mean area per track, (2) mean compactness per track, (3) mean CVI per track, and (4) velocity in the upstream or downstream direction. For features 1, 2, and 4, American eels tend to have greater values than debris, whereas the CVI is typically greater for debris than for eels.

The classification results are summarized by confusion matrices (Table 1) showing how each method assigned samples to the two categories (Duda et al. 2001). For the discriminant function analysis, we included a third category, “unknown,” which was assigned to all samples that had a less than 60% probability of belonging to either the “eel” or “debris” category. Table 1 shows that of the 57 American eels, the discriminant function analysis classified 49 correctly as eels, 7 incorrectly as debris, and 1 as unknown. The total predicted number of eels was 54,

while the true number of eels was 57. Without the probability threshold, the samples that were assigned to the category unknown (one eel and two debris) would have been misclassified.

All of the misclassifications of debris as American eels were caused by “arcing,” an artifact of the

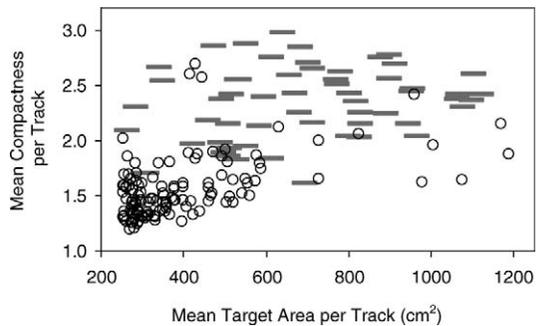


FIGURE 5.—Scatterplot of mean target compactness versus mean area of American eels (solid rectangles) and debris (open circles). The combination of compactness and area provides good but not complete separation of the two types of targets.

TABLE 1.—Confusion matrices of the results of the classification analyses of DIDSON images that were performed.

Variable	Predicted eels	Predicted debris	Predicted unknown	True total
Discriminant function analysis^a				
True eels	49	7	1	57
True debris	5	123	2	130
Predicted total	54	130	3	
Neural network analysis^a				
True eels	53	4		57
True debris	7	123		130
Predicted total	60	127		
K-nearest-neighbor analysis^b				
True eels	10	2		12
True debris	10	75		85
Predicted total	20	77		
K-nearest-neighbor analysis (mean [SD])^c				
True eels	11.3 (0.78)	0.7 (0.78)		12
True debris	10.7 (1.50)	74.3 (1.50)		85
Predicted total	22.0 (1.81)	75.0 (1.81)		

^a Represents the sum of 187 cross-validation estimates generated with a leave-one-out process.

^b Represents the result of applying a K-nearest-neighbor model trained on a set of 45 American eels and 45 pieces of debris to a test set of 12 eels and 85 pieces of debris.

^c Means and SDs of estimates obtained with K-nearest-neighbor analysis in which the data were resampled 1,000 times without replacement.

The results of the neural network analysis show that the misclassification rate for American eels is less than that obtained with discriminant function analysis, while the misclassification rate for debris is slightly higher (Table 2). The results for the K-nearest-neighbor analysis show much poorer classification than was obtained by the other two methods. The predicted number of eels (20) is nearly twice as high as the true number (12). Because the K-nearest-neighbor analysis produced the poorest estimate for the predicted number of eels, we used it as an example for the algebraic correction of estimates. Table 1 shows the mean and SD obtained for each matrix element after resampling 1,000 times, each time taking 45 samples without replacement from each category.

The standard deviations were small, indicating that the classification error was relatively consistent. Therefore, it was appropriate to use the mean values and treat the problem as a set of simultaneous equations that can be solved to give an estimated true category size (see Methods). The solution of the simultaneous equations is shown in Table 3, along with the original estimates. The poor estimates of the initial K-nearest-neighbor analysis improved dramatically—the corrected estimates predicted the true number of American eels and pieces of debris with much greater accuracy.

Discussion

We have presented a computer-driven process that classifies DIDSON images of eels and debris with high accuracy. While the parameters we have chosen are specific to the given application, the overall concept can also be applied to classifying DIDSON images of other objects. An overriding prerequisite, however, is

DIDSON image that is sometimes seen with highly reflective targets (Figure 6). One of the eels was misclassified because it was swimming upstream rather than downstream. All other misclassified eels were the result of the image processing steps having failed to merge all pixels belonging to the eel image, thus breaking it into multiple smaller fragments (Figure 7).

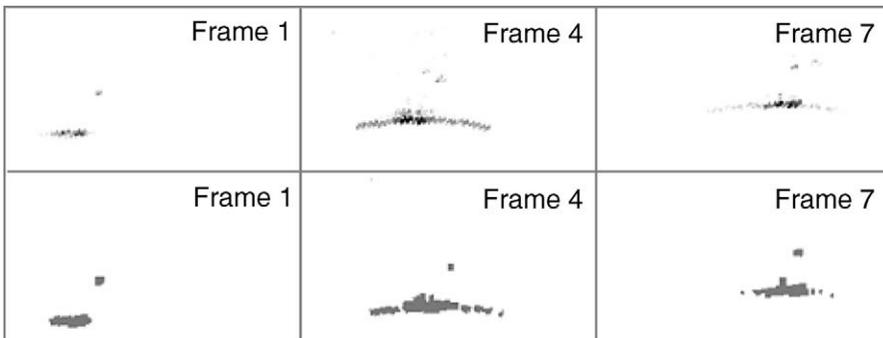


FIGURE 6.—Zoomed view of DIDSON images of debris over a series of frames. Shown are the images after background subtraction (top row) and the corresponding binary masks (bottom row). Note the shape of the debris in frame 1, the arcing effect as debris approaches the beam axis in frame 4, and the reduced arcing in frame 7. Arcing creates an elongated shape in the binary mask similar to that of an eel. A second, smaller debris target is visible at longer range close to the center of each frame; its shape remains unchanged because it does not reflect enough sound to trigger the arcing effect. A third small debris target is barely visible in frames 4 and 7 in the top row; this target does not show up in the binary mask because its area is too small to pass the threshold.

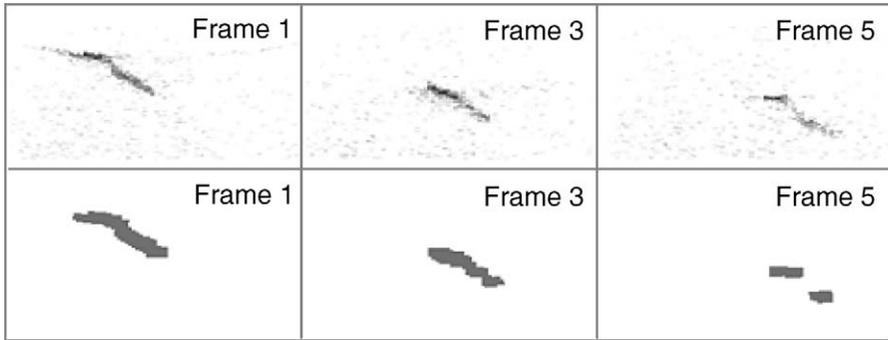


FIGURE 7.—Zoomed view of DIDSON images of an American eel over a series of frames. Shown are the images after background subtraction (top row) and the corresponding binary masks (bottom row). Note that in frame 5 the binary mask fails to join all pixels belonging to the eel, resulting in two separate fragments.

that the DIDSON images be of sufficient resolution and clarity to show the characteristics that distinguish these objects. Only in special cases—when a particular fish species has a unique size, shape, or swimming motion (as is the case with adult American eels)—can this technique be used to identify species.

Careful visual review of the data provides a good first assessment of the potential for classification by a computer-driven process. However, it is often easier to recognize a pattern visually than to translate the visual pattern recognition process into a formal algorithm. The eel images were easily recognized by eye but required considerable experimentation until we found a computer-driven process that performed well. While a visual observer will generally outperform computers in visual pattern recognition, there are cases where the reverse is true. One of the features that contributed significantly to the correct classification of eels was the object’s velocity; American eels were moving downstream slightly faster than debris, which drifted passively with the current. The difference in velocity was too small to be noticed by eye.

Our approach to computer-driven classification of DIDSON images follows four equally important steps: (1) image processing, (2) target tracking, (3) feature selection, and (4) classification. The result of each step

has a significant effect on the performance of subsequent steps, and therefore on the performance of the overall process. In our example, only one misclassification (an upstream-swimming eel) was unrelated to the processing method. All other misclassifications were due to either fragmentation of the target image or artificial elongation (arcing), both of which are artifacts that may possibly be reduced with more sophisticated image processing techniques.

In some instances, poor tracking performance may become a limiting factor. Tracking can become problematic when the algorithm fails to incorporate images that belong to the target or joins images from different targets. In either case, track features will be less characteristic of the target and thus diminish the chances of correct classification. Furthermore, tracks that are broken or incorrectly joined increase the error in the estimated number of targets, irrespective of classification. Depending on the requirements of the application and the effort involved, manual editing may be an acceptable solution. The preferred option, however, is to optimize the preceding image processing steps, in particular background subtraction and smoothing, which can improve tracking dramatically. An alternative approach would be to augment or replace the tracking algorithms with image processing techniques that can group pixels in space and time (H.

TABLE 2.—Misclassification rates of the three types of classification analyses performed.

Analysis	Misclassified American eels (%) ^a	Misclassified debris (%) ^b
Discriminant function	12.3	3.8
Neural network	7.0	5.4
<i>K</i> -nearest neighbor	16.7	11.8

^a Eels identified as debris, as percentage of the true number of eels.

^b Debris identified as eels, as percentage of the true number of pieces of debris.

TABLE 3.—Means (SDs) of the number of American eels and pieces of debris predicted by *K*-nearest-neighbor analysis in which the data were resampled 1,000 times without replacement, before and after algebraic correction (see text).

Object	Predicted value		True value
	Before correction	After correction	
Eels	22.0 (1.81)	11.97 (2.20)	12
Debris	75.0 (1.81)	85.03 (2.20)	85

Balk, University of Oslo, personal communication). Unlike tracking algorithms, which are inherently predictive (i.e., they try to predict what happens on the next frame), the image processing approach makes direct use of the data that are provided over the entire series of relevant frames. In general, we found that tracking performance is less of an issue with DIDSON data than it is with split-beam data. Unlike split-beam sonar, which relies on easily corrupted phase measurements, DIDSON derives an accurate measurement of the target azimuth from the position of individual beams in the array. Because DIDSON provides two accurately measured coordinates (range and azimuth), its data can be tracked more accurately.

Good features capture differences between categories and should not be highly correlated with one another. A set of features useful for classification in fisheries applications may, for example, describe fish shape, size, location, and movement. Feature selection should be based on a combination of empirical observations, some knowledge of the targets that are present, and an understanding of DIDSON data. For example, with DIDSON data the location of the target is known in only two dimensions: range and azimuth. Without the third dimension, it would be impossible to correct echo intensity for the beam pattern effect, which is a function of the target's azimuth as well as its vertical position. Furthermore, because the spreading loss has not been well defined, the time-varied gain we have chosen did not compensate for all range-related changes in intensity. Therefore, we decided to use the CVI as a feature rather than the mean intensity. Because each target covered only a relatively small region of the beam, neither the beam pattern nor range appeared to have a significant effect on the coefficient of variation of intensity.

In addition to the standard feature histograms and scatterplots, we found synchronized echograms to be a useful tool for feature selection. With a series of echograms, one can see, at a glance, patterns in a variety of frame-based features; one can see how the values of each feature vary within tracks, with range, and with target category. As discussed earlier, patterns in range or within a track sometimes indicate anomalies (e.g., an inappropriate time-varied gain or a beam edge effect). The echograms of frame-based features can provide important clues to which summarization over the track (mean, SD, or other statistic) has the best chance of being a good track-based feature. Finally, a dynamic link between a selected area on a feature echogram and the corresponding looped movie display of DIDSON data provides a powerful visual environment for finding features that capture characteristics that are noticed by eyes in the DIDSON movie.

The key to successful classification is to choose features that differ among categories and then match the feature characteristics with an appropriate classification algorithm (Duda et al. 2001). For example, discriminant function analysis assumes that each feature is normally distributed within a category, whereas K -nearest-neighbor and neural network algorithms do not make any assumptions about the feature distributions. In spite of this limitation, discriminant function analysis may be an appropriate choice for our data since its classification accuracy appeared to be adequate and it can be implemented more easily. Neural networks are capable of describing very complicated and disjoint distributions in feature space. However, to get good classification accuracy in such cases may require a large training data set. If one is working with relatively rare objects (as we were with eels), then a large training set may not be possible. For all classification methods, it is important to maintain a large ratio of sample size to number of features. Our goal was to have the sample size for the smallest category in the training data set be at least 10 times greater than the number of features used by a classification model.

To obtain unbiased results from K -nearest-neighbor analysis, the categories in the training set must have an equal number of samples; otherwise, the classification results may be biased toward the category that has more samples in the training set. K -nearest-neighbor analysis measures the sample's distance (in feature space) to all other points in the training set and assigns a classification based on the known class of the K -nearest neighbors. If we had not reduced the number of debris samples in the training set to match the number of American eels, the eel samples would have been more likely to be misclassified as debris. This is due to the fact that many of their neighbors would have been debris samples solely because of the higher overall density of debris samples, not because of their distribution in feature space.

An important option with classification procedures that is sometimes overlooked is the possibility of identifying objects whose classification is marginal (Ripley 1996). Specifically, for our case with just two categories, a classification probability of 0.5 means that the tracked object is equally likely to belong to either category. It is possible to construct a posteriori probabilities for category membership for each of the classification methods we tested (Duda et al. 2001). The discriminant function analysis routine in S-Plus provides these probabilities as part of its results; therefore, we chose discriminant function analysis to illustrate this procedure. We set a probability threshold that must be exceeded for the target to be assigned to a

category. If the threshold is not exceeded, then the target is assigned to the category "unknown." The classification "unknown" can then be used to direct the labor-intensive manual review toward the uncertain cases instead of wasting time on data that can be classified correctly with an algorithm. This strategy combines the intelligent pattern recognition ability of the human brain with the superior speed and consistency of computer-driven processing. It is especially useful with large amounts of data that contain relatively few targets and when results are not required in real time.

The classification models we have described are designed to assign each member of the test data set to the category that has the highest probability. However, it is often the estimated total number of targets within a category that is of interest and not necessarily the category to which a particular target belongs. If the classification error is consistent (i.e., the predictions have small variance), a corrected estimate of the total number of targets in a given category can be derived algebraically.

Unlike classification analysis, which assigns individual targets to categories, the correction process only estimates the total number of targets in each category. For example, if a target had a probability of 0.75 of belonging to the eel category and a probability of 0.25 of belonging to the debris category, classification analysis would classify it as an eel and add one to the number in that category. By contrast, the correction procedure would add 0.75 to the eel category and 0.25 to the debris category. Estimating the total number of members in a class requires only the corrected result. However, sometimes one is also interested in the classification of the individual targets (e.g., to derive spatial and temporal distributions for each category). Therefore, the classification of individual targets and the algebraic correction of the estimated total are both useful procedures. We chose K -nearest neighbor to demonstrate the correction procedure because its classification error was consistent and, of the three methods tested, it had the highest number of misclassifications, and therefore the greatest potential for improvement.

What classification error rates are acceptable depends not only on the purpose of the project but also on the ratio of targets that are present. Ripley (1996) describes the use of a loss function that evaluates the cost of failing to correctly classify an object against the cost of missing a potential category member. This loss function could be modified to include the effect of very unequal category sizes. For our project, a false-positive (debris misclassified as eel) rate of 10% would be sufficiently low if the debris : eel ratio is 1:100 but too

high if the ratio is on the order of 100:1. In the latter example (disregarding false negatives), 1,000 pieces of debris and 10 eels would result in 110 as the predicted number of eels; the estimated number of eels would be driven by the number of debris tracks rather than by the true number of eels.

Depending on the targets present and the quality of the data, different degrees of automation may be achieved. The spectrum of what is feasible ranges from a high degree of human supervision in the classification of all potential targets (in-depth visual review of uncertain cases only) to a fully automated process (which requires only periodic inspection of selected subsamples). For the detection of adult eels it appeared feasible to improve each of the four processing steps (image processing, target tracking, feature selection, classification) to a degree that makes the development of a highly automated procedure possible. Processing may include error estimation and the generation of a corrected estimate. If changes occur in the environment (e.g., the kind of targets that are present or their appearance or behavior), then the classification model may have to be adjusted. While software exists for each of these steps and we have been able to transfer data between the different software packages, any effort to automate or partially automate the process will require a more streamlined and integrated procedure.

Acknowledgments

We thank Madison Paper Industries for permission to publish the data. Don Degan (Aquacoustics) was instrumental in equipment setup, site selection, and data collection. We thank Brandon Kulik, Jesse Wechsler (Kleinschmidt Associates), and David Lovley (Madison Paper Industries) for logistical support and project maintenance. We also thank the staff at Myriax (formerly SonarData), in particular Matt Wilson and Briony Hutton, for their excellent support and the implementation of Echoview software features that made this project possible. We also very much appreciate the suggestions of three anonymous reviewers, as well as those of George Cronkite, John Holmes, Helge Balk, and others, which have greatly improved this manuscript.

References

- Belcher, E. O., W. Hanot, and J. Burch. 2002. Object identification with acoustic lenses. Pages 187–192 in *Proceedings of the 2002 International Symposium on Underwater Technology*. Institute of Electrical and Electronic Engineers, Piscataway, New Jersey.
- Blackman, S. S. 1986. *Multiple-target tracking with radar applications*. Artech House, Dedham, Maryland.
- Burwen, D. L., S. J. Fleischman, and J. D. Miller. 2007a. Evaluation of a dual-frequency imaging sonar for

- detecting and estimating the size of migrating salmon. Alaska Department of Fish and Game, Fishery Data Series 07-44, Anchorage.
- Burwen, D. L., P. A. Neelson, S. J. Fleischman, T. J. Mulligan, and J. K. Home. 2007b. The complexity of narrowband envelopes as a function of side aspect angle. *ICES Journal Marine Science* 64(5):1066–1074.
- Casselman, J. M. 2003. Dynamics of resources of the American eel, *Anguilla rostrata*: declining abundance in the 1990s. Pages 255–274 in K. Aida, K. Tsukamoto, and K. Yamauchi, editors. *Eel biology*. Springer-Verlag, Tokyo.
- Duda, R., P. Hart, and D. Stork. 2001. *Pattern classification*, 2nd edition. Wiley, New York.
- Hornik, K. 2007. The R FAQ. Available: <http://CRAN.R-project.org/doc/FAQ/R-FAQ.html>. (November 2007).
- Maxwell, S. L., and N. E. Gove. 2004. The feasibility of estimating migrating salmon passage rates in turbid rivers using a dual-frequency identification sonar (DIDSON) 2002. Alaska Department of Fish and Game, Regional Information Report 2A04–05, Anchorage.
- Ripley, B. 1996. *Pattern recognition and neural networks*. Cambridge University Press, Cambridge, UK.
- Seul, M., L. O’Gorman, and M. Sammon. 2001. *Practical algorithms for image analysis: description, examples, and code*. Cambridge University Press, Cambridge, UK.
- Tiffan, K. F., D. W. Rondorf, and J. J. Skalicky. 2004. Imaging fall Chinook salmon redds in the Columbia River with dual-frequency identification sonar. *North American Journal of Fisheries Management* 24:1421–1426.
- USFWS (U.S. Fish and Wildlife Service). 2007. 12-month finding on a petition to list the American eel as threatened or endangered. *Federal Register* 72: 22(2 April 2007):4967–4997.