

**VITECH INNOVATIVE
RESEARCH AND CONSULTING**

6280 DOULTON AVE, RICHMOND, B.C., CANADA V7C 4Y4

TEL: 1-604-241-5810

EMAIL: LDING@APEXLINK.CA

**PSC 2018 SOUTHERN FUND
PROJECT CLOSURE REPORT**

**Automating In-Season Salmon Species
Composition Estimation at Mission using
Imaging Sonar**

**PREPARED FOR
PACIFIC SALMON COMMISSION**

MARCH, 2019

TABLE OF CONTENTS

1	EXECUTIVE SUMMARY	3
2	INTRODUCTION	5
3	METHODS	5
3.1	Technical Background Overview	5
3.2	Variability in Length Measurements	6
3.3	Derivation of Fish Characteristics from DIDSON Data.....	7
3.4	Estimation Methods for Species Composition	8
3.5	Performance Evaluation of Estimation Methods	9
4	RESULTS AND DISCUSSION	9
4.1	Measurement of Individual Fish Length	9
4.2	Direct Comparison of Measured Fish Length with Ground Truth Data	10
4.3	Optimal Estimation of Live Fish Length	12
4.4	Distributions of Feature Variables in Mono-specific Data	14
4.5	Performance Evaluation of the EM Algorithm in Estimation of Species Composition...	16
4.6	Software Support for EM Estimation of Species Composition	17
4.7	Current Status and Follow-up	18
5	RECOMMENDATIONS	20
6	ACKNOWLEDGEMENTS	20
7	APPENDICES	21
7.1	Region-Based Image Segmentation	21
7.2	Expectation Maximization.....	21
8	REFERENCES	24

1 EXECUTIVE SUMMARY

High-frequency image sonar such as DIDSON and ARIS deployed at Mission and upstream locations yields high resolution images of individual fish, which provide rich information about the shape, size and behavior of individual fish, thus offering a promising technology for estimating compositions of salmon species in the region.

Species estimation methods can be model-based such as a mixture model with the Expectation Maximization (EM) algorithm, or based on classification of individual tracks using machine learning approaches. In previous projects funded by the Southern Fund (2007-2008, 2009-2010), we have investigated the second approach by calculating a set of feature variables for individual tracked targets, and applying the Discriminant Function Analysis (DFA) algorithm to classify each individual fish, thus obtaining an estimate of the species composition.

In the current project we focused on the model-based approach. Since fish length appears to be one of most obvious characteristics that would differentiate fish species, we have made a significant effort to automate the measurement of individual fish length by using an image segmentation algorithm. We have currently fully integrated the image segmentation algorithm into the tracking software we built before (IntelliHAT), which allows automatic measurement of the length of every tracked target on individual frames, in addition to target position and backscatter strength. The length measured over frames has great variability due to highly variable backscatter from individual fish and interference from adjacent fish. For this reason, we derived fish length from a sequence of measurements by tracking individual targets, instead of single measurements. To assess the performance of the fish length estimator, we have also compared with some 'ground truth' data, which is fish length measured manually by a human operator on images (via a computer mouse), and was provided by the Hydroacoustics Group of the Pacific Salmon Commission. We found that the difference between the measured length and the ground truth was 2cm on average, with a standard deviation of 12-20cm, suggesting that there was little systematic bias.

In the 2010 SEF project, we built a software suite (Fish Analysis and Composition Estimation, or FACE) which implements functions for feature variable calculation, and various machine learning algorithms. In this project, we have added the EM module to FACE which allows the user to provide initial model parameters and generate an estimate of class proportions at multiple classes. This streamlines the estimation procedure: generating track data and length measurement data from the tracking software, which are then fed to FACE for species composition estimation. The EM module also supports PSC' manually measured fish length data, which is currently used in the field in species composition estimation.

To assess the performance of the EM algorithm, we recollected a set of monospecific data (to the best of our judgement), and constructed different compositions of species. Because of a limited amount of the monospecific data, we used bootstrapping to generate different subsets of data from the same pool of data, thus obtaining a distribution of composition estimates. For the case of two species, sockeye and pink, the estimated compositions, based solely on fish length data, have a bias of less than 2% from the true value, when the two classes have similar

proportions (40%-70%), but the bias increases as the difference between the proportions increases (for example, at pink proportion = 10% or 90%, the bias is as large as 9%).

In summary, in this project we have enabled automatic measurement of fish length from DIDSON image data. Such length measurement data alone can be used to estimate species compositions with the EM algorithm, and combined with other characteristics, may offer better estimates. We have also integrated the EM algorithm into the existing software, allowing streamlined operations. The same measurement function and procedure are expected to be compatible with PSC's ARIS image data. It is also possible that this approach can be applied to similar problems where DIDSON or ARIS systems are in place.

2 INTRODUCTION

Applications of high-resolution imaging sonar such as the dual-frequency identification sonar (DIDSON) and ARIS have provided a unique opportunity to tackle the problem of species identification, by providing rich information about the shape, size and behavior of individual fish. In our previous work (Vitech 2008, 2010), we extracted fish morphometric and behavioral information from DIDSON image data, and used machine learning algorithms to classify each individual fish and thus obtain an estimate of species composition. We have also developed a standalone software application, streamlining the estimation procedure and incorporating various statistical and computational tools.

In this project, we focused on improving the accuracy and robustness of automatically measured fish length based on sonar images, since fish length appears to be the most influential factor in fish species discrimination. We applied an image morphological analysis approach to enhance the performance in length measurement. In addition, the fish length data alone can be used to determine species proportions via a mixture model, as currently implemented on R by PSC for manually measured fish length data. The mixture model has been added to the existing software package.

This report will first review the methods used in the project, including the image segmentation algorithm for automatic length measurement. Then we present the results of applying these methods to some historical DIDSON data. We will also compare automatic length measurement data with the corresponding manually measured data (by a human operator on images) provided by PSC. Next, we evaluate the performance of the mixture model in estimating species compositions from our automatic measurement data, by constructing mixture data at predetermined compositions from monospecific data. Finally we make some recommendations for future work.

3 METHODS

3.1 Technical Background Overview

Our approach to species composition estimation is to analyze fish morphology and behavior characteristics extracted from DIDSON sonar image data. However, before we can extract fish morphological and behavioral data, we need to be able to track individual fish, and we have developed a software tool (IntelliHAT) for this purpose (Vitech 2008). Our current approach involves the following phases:

1. Object detection: on each frame of image data, pixels in tight clusters with higher intensity are identified as potential candidates for targets (fish). In the meantime, the morphological characteristics of the potential candidates are estimated and recorded.
2. Object tracking: those candidates are fed to a multi-target tracking algorithm that tracks the candidates over time. Tracking links the same candidates in different frames and filters out spontaneous and isolated candidates.

- Morphological analysis: morphological characteristics of individual tracked targets are analyzed and optimal estimates are derived.

To perform species identification, we construct a set of feature variables (descriptors) from morphometric and behavioral characteristics derived from DIDSON data using the approach described above. Once a set of feature variables is selected, we apply machine learning methods to classify the observations and obtain an estimate of species composition.

3.2 Variability in Length Measurements

As mentioned above, a fish can be tracked and its length can be measured automatically on individual frames using the region-based segmentation algorithm. Length measurements for a single fish may vary greatly over its trajectory, as shown in Fig. 1 (upper-left), for a number of reasons: 1) fish sometimes swim in tight aggregations, making it difficult to separate adjacent fish and measure their length unambiguously; 2) fish often flex as they move across the acoustic view, making their visibility greatly variable over time; 3) fish are less visible if their body is more aligned with the acoustic beam; 4) background interference at some locations may be strong. Despite these limitations, we still could derive a reasonable estimate of fish length from a sequence of measurements over frames, as long as local fish density does not always stay high. Moreover, for estimation of fish species, we are more interested in collecting statistical distributions of fish length, than measuring each individual fish with high precision.

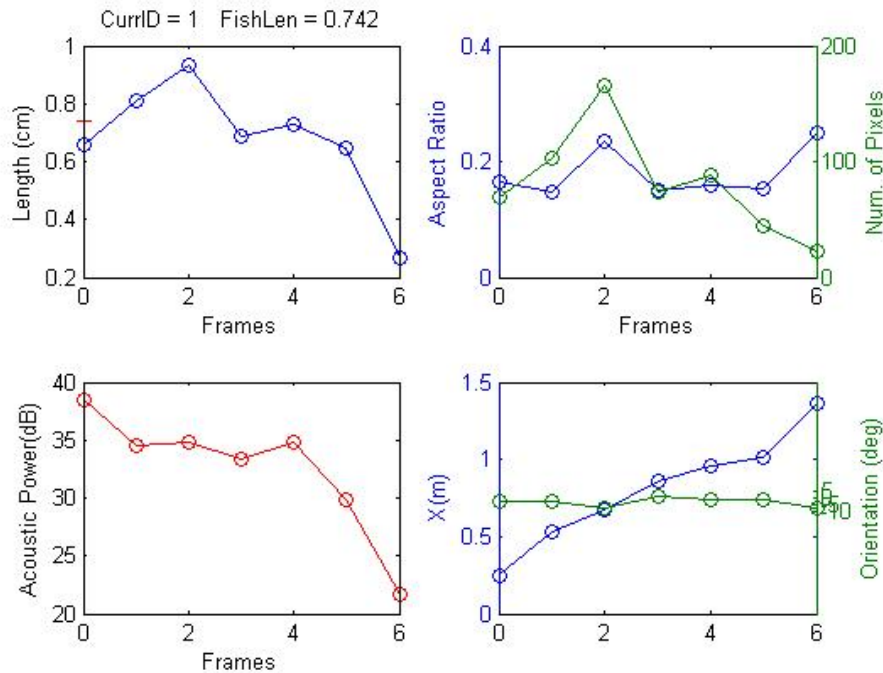


Fig. 1.: Measurements of length and other parameters for a live fish as a function of time (frames) over. Upper-left: fish length; Lower-left: average acoustic backscatter; Upper-right: body aspect ratio and number of pixels associated with the fish; Lower-right: across-beam position (X) and orientation relative to the acoustic beam (zero degree means fish body is perpendicular to the acoustic beam).

Now the question is which measurement represents the best estimate of the fish length, given such variability of the measurements. Intuitively, it is reasonable to assume that the length measurement is most reliable when the fish image is brightest, but high brightness (resulting from strong acoustic backscatter) is often associated with the swimbladder of a fish, not necessarily with the entire fish body. It may also be more likely to have better measurements if the body is perpendicular to the acoustic beam axis. Also, if the image patch associated with a target is more elongated, we can assume that the target is more likely a fish.

Therefore, when calculating the fish length, we also calculated simultaneously a number of parameters as shown in Fig. 1, including average acoustic power (lower-left), aspect ratio and number of pixels in a region (upper-right), and orientation (lower-right). An optimal estimate of the true fish length can be derived from the observed length, with the assistance of these simultaneously measured parameters. We may model the observed fish length as

$$L = L_0 + a_1AR + a_2Pwr + a_3Ori + a_4Np$$

Where

- L_0 : true length, which is also what we want to estimate;
- AR: aspect ratio of the image patch associated with an identified fish. A smaller aspect ratio indicates a more elongated target. See the section below for more details.
- Pwr: averaged acoustic backscatter for the pixels associated a fish.
- Ori: angle of a fish relative to the acoustic beam. See the section below for more details.
- Np: number of pixels in the image patch associated with a fish.

Our objective is to use this model to estimate fish length from measurement data. We can use a set of selected measurement data with corresponding ground truth data (L_0) available (fish length directly measured by hand on the image using a computer mouse) to train the model. The coefficients, a_1 to a_4 , can be obtained using linear regression. In application, we use the estimated coefficients obtained from the linear regression training, and measurement data for each individual target, to compute an estimate of the target length (L_0) at each frame in the track. Then we take the mean of the estimated length at each frame to represent the length of the target.

3.3 Derivation of Fish Characteristics from DIDSON Data

In addition to fish length, we can also derive other characteristics (referred to feature variables) of individual fish from DIDSON data for species classification. The technical details have been described in Vitech (2008 and 2010). Here we review a number of feature variables, including those that have been used before:

- *Fish length*, as described in the preceding section
- *Path-averaged speed*, which is the length of a target trajectory divided by the time it spans. This represents how fast a fish can swim on average, regardless of the direction.
- *Cross-beam velocity* (speed and direction), which provides an indicator of the upstream/downstream direction for migrating salmon.

- *Aspect ratio of the imaged fish*, which is defined as the ratio of standard deviations of the pixel positions in two perpendicular directions, one of which is the principal component. We use principal component analysis to determine the two directions automatically. In our definition, the smaller the aspect ratio is, the more elongated the imaged fish is.
- *Orientation*: this is the orientation of the principal component of the imaged fish which typically aligns with the head-to-tail direction. The more perpendicular to the acoustic beam the fish body is, the stronger backscatter is.
- *Acoustic backscatter*, which is defined as the total acoustic intensity in an ‘imaged’ fish, normalized by the number of the pixels associated with the fish. This parameter is measured at the same time as fish length is measured. Note that the acoustic intensity is not calibrated, but compensated for spherical spreading of sound.
- *Local density*, which measures the fish density in the neighborhood of a fish. Local density is the target density in the neighbor of a fish. This is also a dynamic parameter varying as the fish moves. It is calculated in the following procedure:
 1. Track a video file and generate a corresponding track file.
 2. For each target (fish) in the track file, examine each frame in which the fish appears. Use an unsupervised clustering algorithm (Batchelor and Wilkins’ Algorithm) to find natural clusters in the target positions on the current frame. Calculate the maximum spatial extent in each cluster, and define local density as the number of targets in the cluster divided by the spatial extent.
 3. Repeat the calculation for each frame where the target exists. Find the maximum value to represent the local density of the target.

3.4 Estimation Methods for Species Composition

The ultimate goal of this project is to estimate species composition based on DIDSON sonar observations. Estimation methods can be categorized into two groups. One is based on species classification, in which individual observations are analyzed to determine which species they may have originated from. Species composition is then estimated by counting the number of observations classified to each species and performing appropriate corrections. The other group of methods is model-based, where species composition is derived directly from a set of observations based on a model, without having to classify individual observations. In this project, we focus on the methods in the first group.

We have employed the first approach in our previous project (Vitech 2008 and 2010). In this project, we explore a model-based approach: Expectation Maximization (EM). This algorithm is suited for cases where the available data set is incomplete. For example, a set of feature variable data such as fish length is known to have been drawn from a number of species (e.g. sockeye and pink), but we lack the information regarding which sample is drawn from which group (labelling). The EM algorithm models the complete data set as the observed data (length) and the unobservable data (labeling samples with a species). It takes the expectation (average) of the probability function of the complete data (based on a statistical model such as the normal distribution), averaging out the unobservable data. Then based on Bayes’ theory, it adjusts model parameters, including prior probabilities, to maximize the expectation, thus providing an estimate of the group proportions. See the appendix section for more details.

3.5 Performance Evaluation of Estimation Methods

After an estimation method is selected, an important step is to evaluate its performances in terms of classification errors and overall estimation bias and variance. A straightforward evaluation approach is to collect a set of monospecific samples, independent of those used for training purposes, and test the output of an estimation method against some ‘ground truth’ data. However, the set of monospecific samples at our disposal is often finite and has to be used in both training and testing. One evaluation approach in the case of a limited number of training data is the bootstrap method, which is commonly used in pattern classification (Han and Kamber, 2005). This method generates a set of training data, by randomly and uniformly sampling an available data set with replacement (i.e. the same data may be selected more than once). For a data set of size N , if we select N data points with replacement, there will be roughly 63% of the original data selected, while the rest (37%) will be left out. Those selected will be used for training and those left out used for testing. This procedure is repeated for a number of times, leading to a statistical distribution of the outputs, from which the performances can be evaluated in a statistical sense. Although there are other evaluation approaches recommended for pattern classification such as cross-validation, we feel the bootstrap method is appropriate within the scope of this project.

4 RESULTS AND DISCUSSION

4.1 Measurement of Individual Fish Length

In our previous work (Vitech 2008), we used a boxing approach to measure fish length in an image: first, for each tracked fish in a frame, we used a simple automatic thresholding algorithm (Gonzalez and Woods, 2002) to determine automatically an intensity threshold in a fixed size box surrounding the fish. Pixels in this box with intensity above the threshold were classified as fish while the other pixels were considered background. The fish length was then calculated from the selected pixels.

The most significant challenge to the above approach is that fish sometimes swim in tight aggregations, resulting in multiple fish in a box and thus making it difficult to measure unambiguously. In order to overcome this problem, we developed an alternative approach known as Region Growing. First, on each frame, we identify a few ‘hot spots’ where echo intensity is distinguished from the background. This information is available in the object detection step. Pixels in these hot spots are used as seeds to grow regions from the seeds.

For region growing, we select from the original image, pixels surrounding the seeds whose intensity is within a threshold different from the intensity at the seeds. This threshold must be relaxed enough to allow sufficient pixels in for size measurement. Note that depending on how the threshold is selected, some additional noise (individual isolated pixels) will be included. To minimize noise, we apply a connectivity algorithm to filter out the noise. This algorithm reconstructs regions around the seeds by including pixels that are connected with the seeds, thus eliminating those isolated pixels (See the appendix section for more details). Figure 2 shows an example of region-based segmentation, where three targets are identified. The color (other than blue) patches are regions identified by the algorithm.

The positions of the pixels in these reconstructed regions are fed to a length measurement algorithm. This length measurement task is performed in the object detection phase. The object detection data, including the length data, are fed to the tracking algorithm. This allows us to follow individual targets and measure their length over frames, leading to a time series of length for each tracked fish.

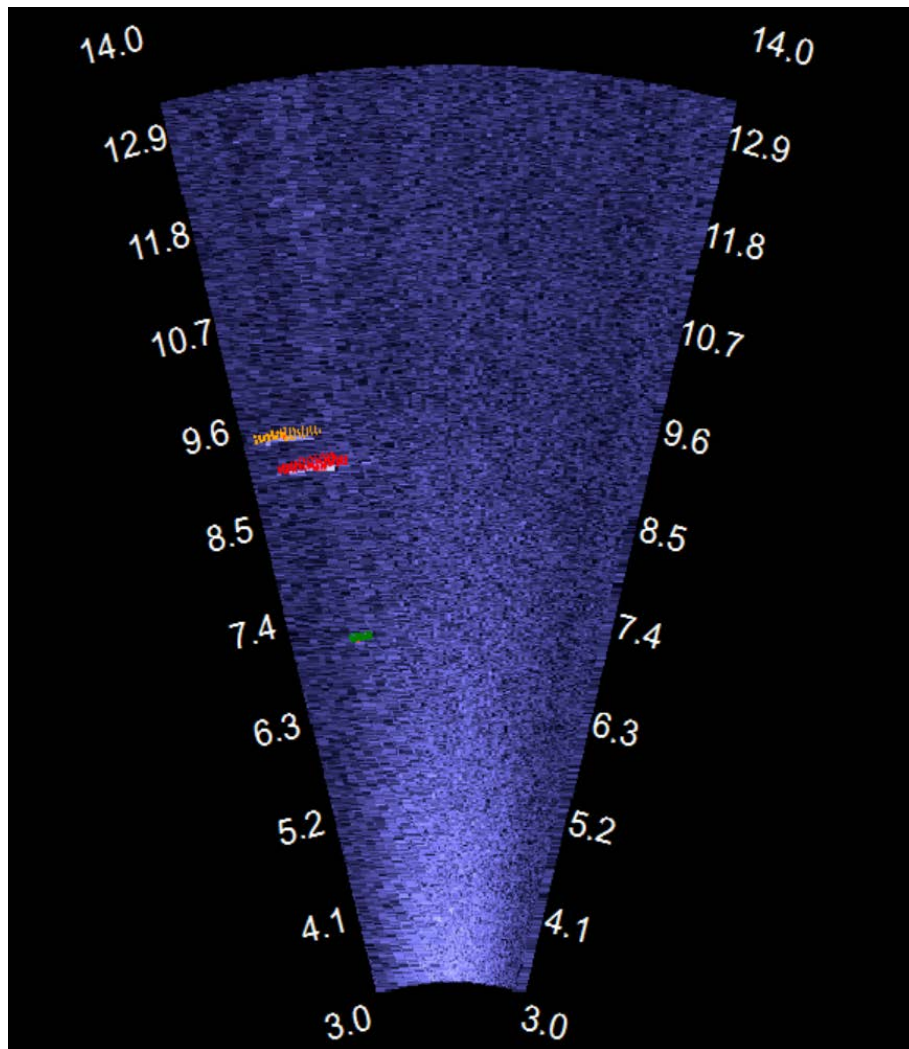


Fig. 2.: Example of image segmentation. Pixels in colors other than blue indicates regions associated with targets (see the text for more explanation).

4.2 Direct Comparison of Measured Fish Length with Ground Truth Data

For clarification, by ground truth data, we do not mean measurement data of the physical length of a fish. Instead, we mean fish length measured manually by a human operator on images (via a computer mouse). The Hydroacoustics Group of the Pacific Salmon Commission has collected a considerable amount of human measured fish length data from DIDSON

images. We will use some of these data as ground truth to evaluate the corresponding measurements obtained with our algorithm described in Section 3.

In the ground truth data provided by PSC, the frame number and range of a fish on an image were recorded together with the measured length. To proceed with direct comparison, we inspected individual fish in the ground truth data and made sure the fish was observed clearly on images at the specified frame and range. Then we inspected corresponding track data and identified the track that best matched the fish identified in the ground truth, given the frame number and range. With the identified track ID, we can find all the measured parameters for the track. Since this part of the work is labor intensive, here we use only a subset of the ground truth data.

It should be pointed out that human perception of the existence of a fish can be quite different from what a computer algorithm ‘sees’. In our algorithm, we use a set of fixed target detection and tracking parameters to process all frames in a set of image files. If we could fine-tune those parameters for each frame, it would be possible to detect the same targets as the ground truth. However, this would not be practical because we could have to keep an indefinite set of parameters.

Therefore, there are cases where the ground truth indicates the existence of a fish but the algorithm does not ‘see’ the fish at the specified frame; the algorithm may detect the fish at previous or subsequent frames. In such cases, we choose the closest frame where the algorithm can detect a target that matches the fish in the ground truth data (typically within 3 frames from the frame number in the ground truth).

Figure 3 shows a distribution of the difference between ground truth length and corresponding measured length on the same frame (or at the closest frame) and at a similar range. Positive differences indicate that the measured length is greater than the ground truth. It can be seen that the measured length is more likely to overestimate the true length. The upper figure is the result from a set of data when the DIDSON sonar was configured at high frequency (HF), while the lower figure is from the low-frequency (LF) configuration. The mean difference at high frequency is around 7.9cm and standard deviation of 17.8cm, while the difference at low frequency has a mean of 21cm and standard deviation of 43cm. The larger bias and variability at low frequency are due mainly to lower beam resolutions at the LF configuration, and focus on greater ranges thus having a lower signal to noise ratio.

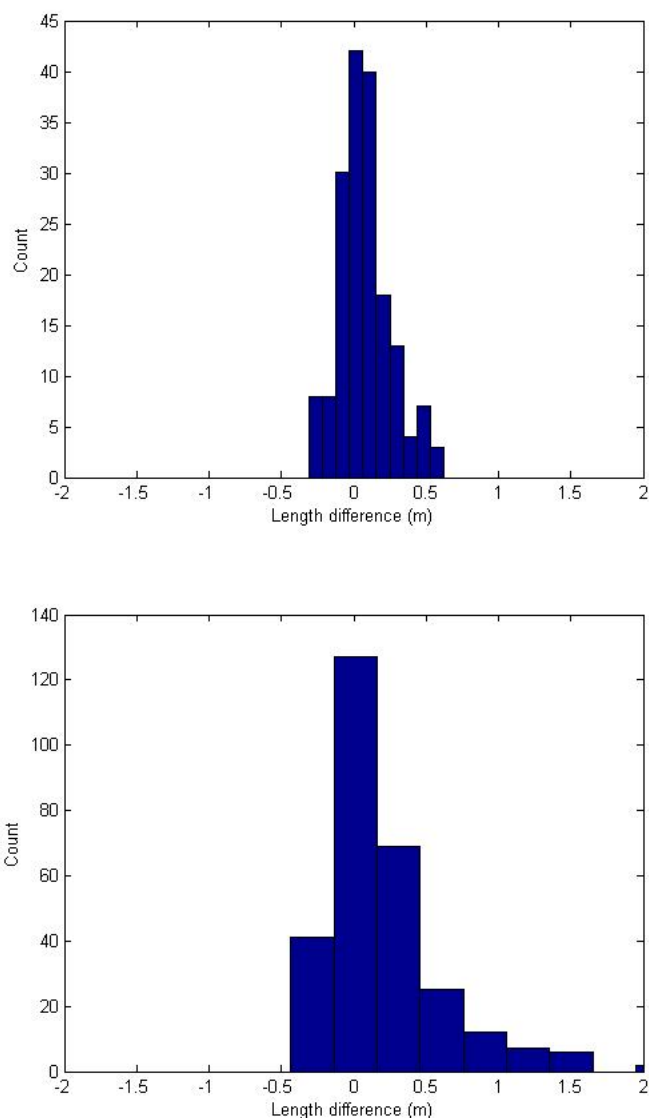


Fig. 3.: Histogram of difference between ground truth length and corresponding measured length on the same frame (or at the closest frame) and at similar range. Positive differences indicate that the measured length is greater than the ground truth. The upper plot is for data collected at the high frequency configuration of the imaging sonar, while the lower plot is at the lower frequency configuration.

4.3 Optimal Estimation of Live Fish Length

As can be seen from the above results, there exists great discrepancy in the direct comparison, mainly because some ground truth data were taken within tight clusters of fish or at a low SNR. Human vision is usually much better at separating fish in this case than our current algorithm (unless we resort to deep learning segmentation approaches). However, the same fish do not necessarily stay close to each other all the time, and there are times when they would move away from each other. Therefore, we may be able to obtain a better estimate of individual fish

length from a sequence of measurements along the trajectory of the fish, using the approach outlined in Section 3.3.

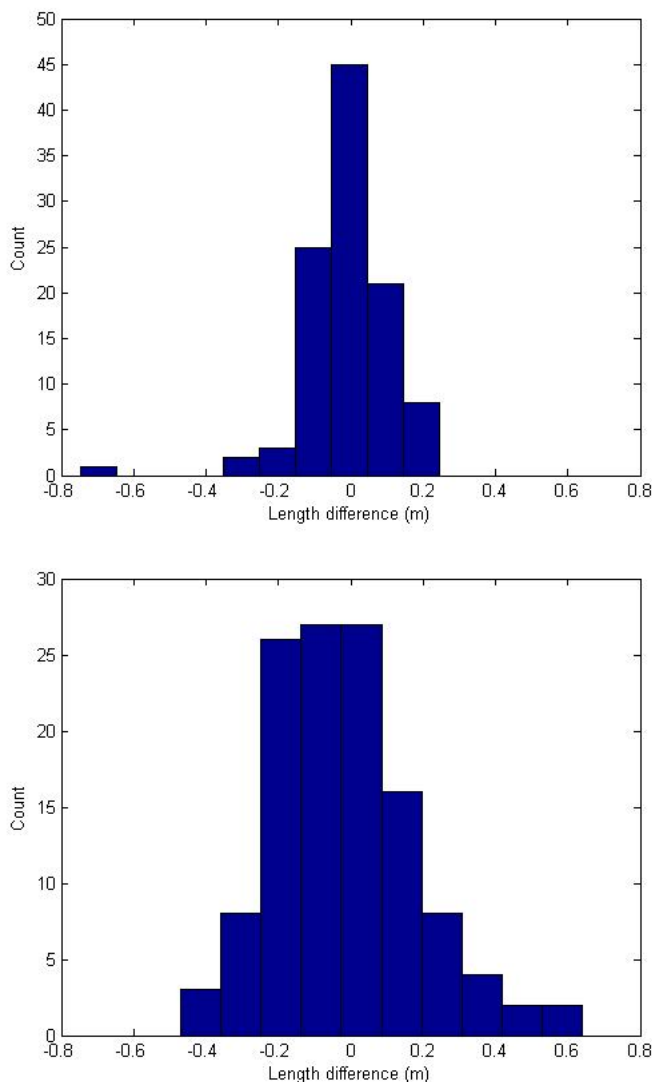


Fig. 4.: Histogram of difference between ground truth length and corresponding measured length based on linear regression of a sequence of measured parameters for each individual track. Positive differences indicate that the measured length is greater than the ground truth. The upper plot is for data collected at the high frequency configuration of the imaging sonar, while the lower plot is at the lower frequency configuration.

We used 40% to 50% of the ground truth samples (with both manually measured length and corresponding track data) to train the linear model, and the coefficients thus obtained were used to estimate fish length for the rest of the samples (see Section 3.3). The estimated length is then compared with the ground truth. Figure 4 shows the distribution of difference between the estimated length and the ground truth. The upper plot is for the HF case, where it can be seen that the distribution is much more concentrated around zero with a mean of -1.0cm with standard deviation of 12cm. The lower plot is for the LF case, where we can see the distribution is more symmetrically centered at zero, with a mean of -2cm and standard deviation of 20cm, much better than the result of direct comparison in Fig. 3.

4.4 Distributions of Feature Variables in Mono-specific Data

We now revisit the mono-specific data collected in a previous project (Vitech 2008). These data were collected with the assistance of test fishing information, which helped to find time periods in which a single species dominates. Test fishing data were usually collected 20-km downstream of the sonar observation site. Within the distance, we do not expect species composition to change significantly. Even during these periods, there were occasionally local small fish or other salmon species in the data, and we carefully reviewed the image data to exclude fish that did not appear to belong to the species of interest. For example, local small fish look smaller and directionless in video playback of the data, and can be easily identified and excluded from pink or sockeye samples. Pink tend to swim tightly in groups of two or more, and when selecting pink samples, we focused on those targets in groups.

We used the same linear regression coefficients as used in Fig. 4 to process the mono-specific data and obtain distributions of the estimated length for sockeye and pink. Figure 5 shows the results, where it can be seen that the measured length of pink (middle plot), while peaked around 40cm, shows a longer tail to the right. This is probably due to the fact that pink tend to swim in groups, which more likely results in a larger estimated length than the true length if the fish get too close to each other. The distribution for sockeye (bottom plot) peaks between 50-60cm, and does not spread much to the right. Sockeye tend to swim separately (at least with a larger distance from others), making it less likely to measure adjacent fish as a single fish. As a reference, we also show the distribution for a dataset (top plot) that contains mainly local residence fish, where we can see the fish length distribution peaks at 20cm

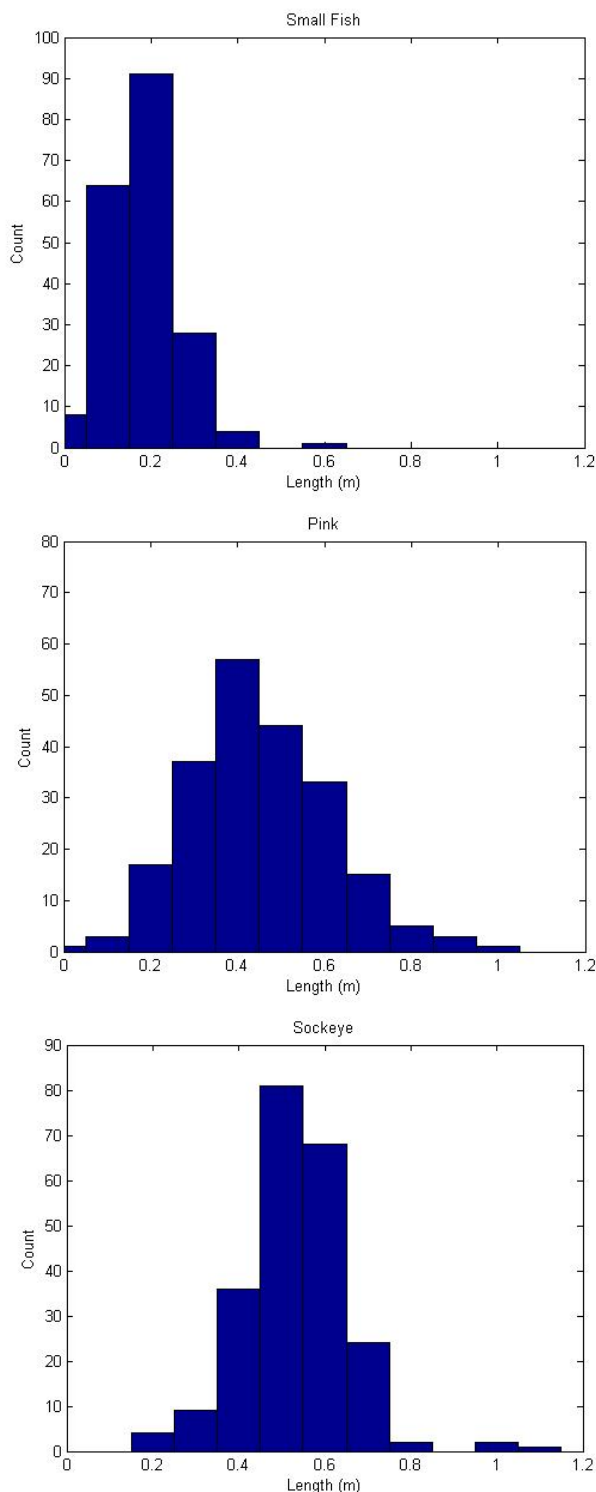


Fig. 5.: Histogram of mono-specific fish length measured for each individual track, using the same linear regression coefficients obtained with the ground truth data (see text for more details). The top plot is the distribution for local residence fish, the middle for pink, and the bottom for sockeye.

4.5 Performance Evaluation of the EM Algorithm in Estimation of Species Composition

Intuitively, the most obvious feature available from the image data that distinguishes the species of salmon is fish length. However, pink and sockeye are only different by about 10-20cm in length on average, and their distributions of length overlap significantly as seen in the length histograms in Fig. 5. It appears that if we want to rely solely on length for species composition estimation, a model-based approach, such as Expectation Maximization (EM), appears to be appropriate.

The ability of the EM algorithm to estimate species composition can be evaluated by using the monospecific data and the bootstrapping approach as described in Section 3.6. With this method, we can construct different species compositions by randomly selecting a desired number of samples from the monospecific data, and evaluate the accuracy of composition estimates derived from the EM model. We then repeat the sampling for a number of times to generate a distribution of composition estimates and compare with the true compositions.

As described in the appendix, the EM algorithm can simultaneously adjust model parameters (e.g. mean and covariance) and class proportions. That is, the EM algorithm can be used in unsupervised classification where training data are not readily available. However, since the statistical difference between the length of sockeye and pink is modest, allowing both model parameters and class proportions to be freely optimized does not generate good results. Therefore, we constrain the optimization to class proportions only, while fixing the model parameters based on some prior information such as mono-specific data. In other words, we take a supervised approach by training the model first with mono-specific data, and use the trained model to estimate class proportions.

We partition the available mono-specific samples into two exclusive datasets (about 50/50 partition): one for training which provides estimates of the model parameters, and the other for testing, which evaluates the performance of the EM model. We test the case of two classes: sockeye and pink, at different compositions. Table 1 shows the resulting estimates of pink proportion and the corresponding true value. For each composition, the EM is run for 100 different realizations. The first set of estimated compositions is based solely on fish length data, where it can be seen that the bias from the true value is very small (<2%) when the two classes have similar proportions (0.4-0.7), but the bias increases as the difference between the proportions increases (for example, at pink proportion = 0.1 or 0.9, the bias is as large as 9%).

In our application, there are three categories of characteristics that could be used to construct feature variables: 1) individual fish behavior such as swim speed and direction; 2) morphological data of individual fish such as length and shape; 3) behavioral characteristics of fish aggregations or schools. Here we show a case where fish aggregation (see Section 3) is also included in the EM model, in addition to fish length. The second set of estimates in Table 1 is the result of using the two variables. It can be seen that the bias is reduced at small and large pink proportion (0.1 and 0.9).

True Pink Prop	Mean EM Est. (1)	STD Est. (1)	Mean EM Est. (2)	STD Est. (2)
0.1	0.19	0.097	0.16	0.079
0.2	0.29	0.111	0.23	0.082
0.3	0.34	0.118	0.31	0.079
0.4	0.42	0.112	0.40	0.094
0.5	0.52	0.122	0.49	0.079
0.6	0.61	0.117	0.58	0.081
0.7	0.70	0.117	0.67	0.105
0.8	0.77	0.116	0.78	0.104
0.9	0.82	0.106	0.88	0.085

Table 1: EM estimates of different pink proportions. The first set is based solely on fish length data, while the second set is obtained with information about fish aggregation also included. These results were obtained by running the model for 100 realizations of bootstrapped mono-specific samples.

4.6 Software Support for EM Estimation of Species Composition

In a previous project, we built a software suite (Fish Analysis and Composition Estimation, or FACE) which implements various machine learning algorithms. It has a simple main interface as shown (Fig. 6), and consists of four modules: generation and visualization of feature variables, classifier training, performance evaluation of a trained classifier, and application of a trained classifier to new data. See our previous report (Vitech 2010) for more details.

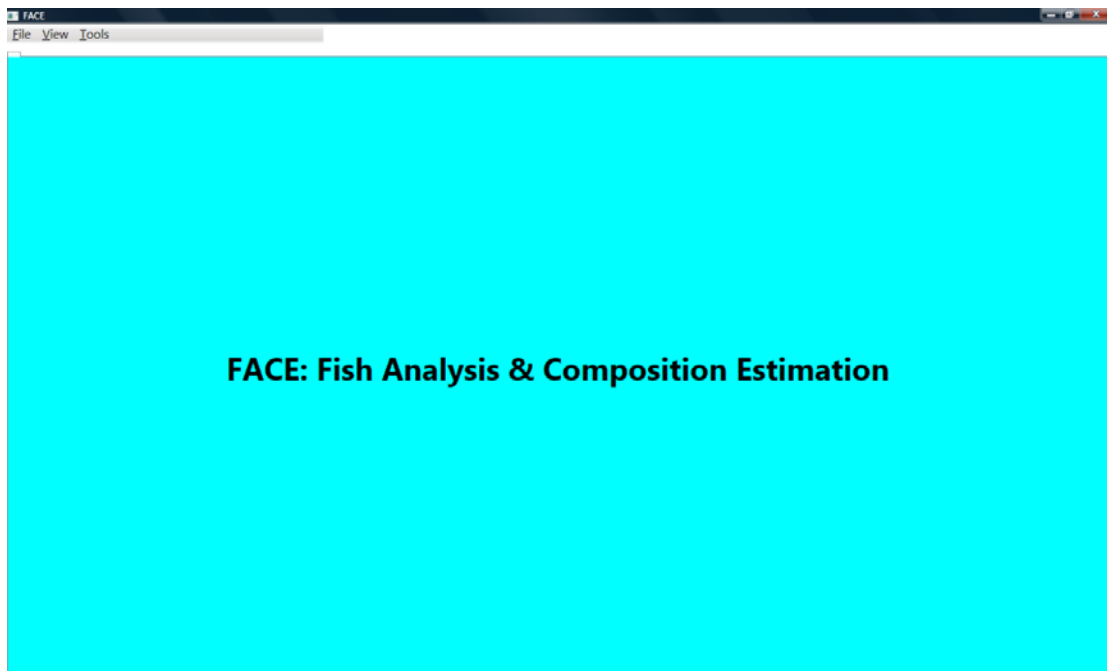


Fig. 6.: Interface of the software FACE.

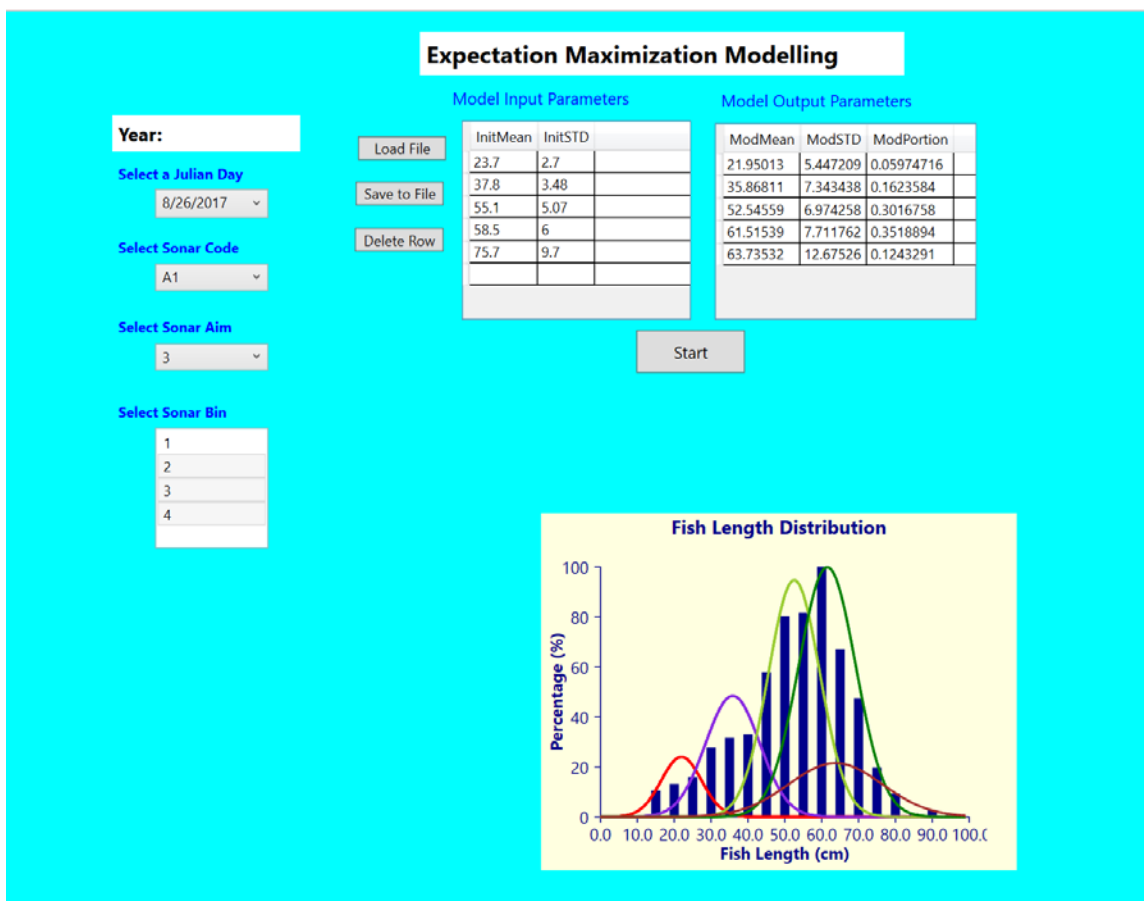


Fig. 7.: Interface for EM mixture modeling. User loads a length data file, and selects sonar aim and bins. Then the user enters initial model parameters in the input box, or loads a model configuration file. On the output, model parameters and proportions are shown corresponding to the input modes. Histogram of the data and model probability distributions at each mode are also shown.

In this project, we have added the EM module to FACE as shown in Fig. 7. This module allows the user to enter initial model parameters manually in the input textbox and save them a configuration file. The user can also load a previously saved configuration file. It supports fish length data created by the Pacific Salmon Commission. The data contains fish length manually measured by a user on sonar images and records the aim and bin (range) of the imaging sonar. The user can select data from a day, sonar aim, and bins, and run the model to estimate class proportions, and the results are shown in the output box.

4.7 Current Status and Follow-up

As described in Section 3, species estimation methods can be model-based such as EM, or based on classification of individual tracks using machine learning approaches. We have investigated the second approach in our previous SEF projects (Vitech 2008, 2010), while in the current project we focused on the first approach. Since fish length appears to be one of most obvious characteristics that would differentiate fish species, we have applied an image

segmentation algorithm for automatic measurement of individual fish length, and the results are presented above.

The most significant challenge to automatic measurement of fish length is high density in a cluster of fish such that the spacing between fish is close to the inherent image resolution. For this reason, we derived fish length from a sequence of measurements by tracking individual targets, instead of single measurements. However, when the fish density is extremely high, such as shown in Fig. 8 in which pink passed at a rate of up to 30,000 per hour, our approach would start to break down by interpreting clusters of fish as single fish.

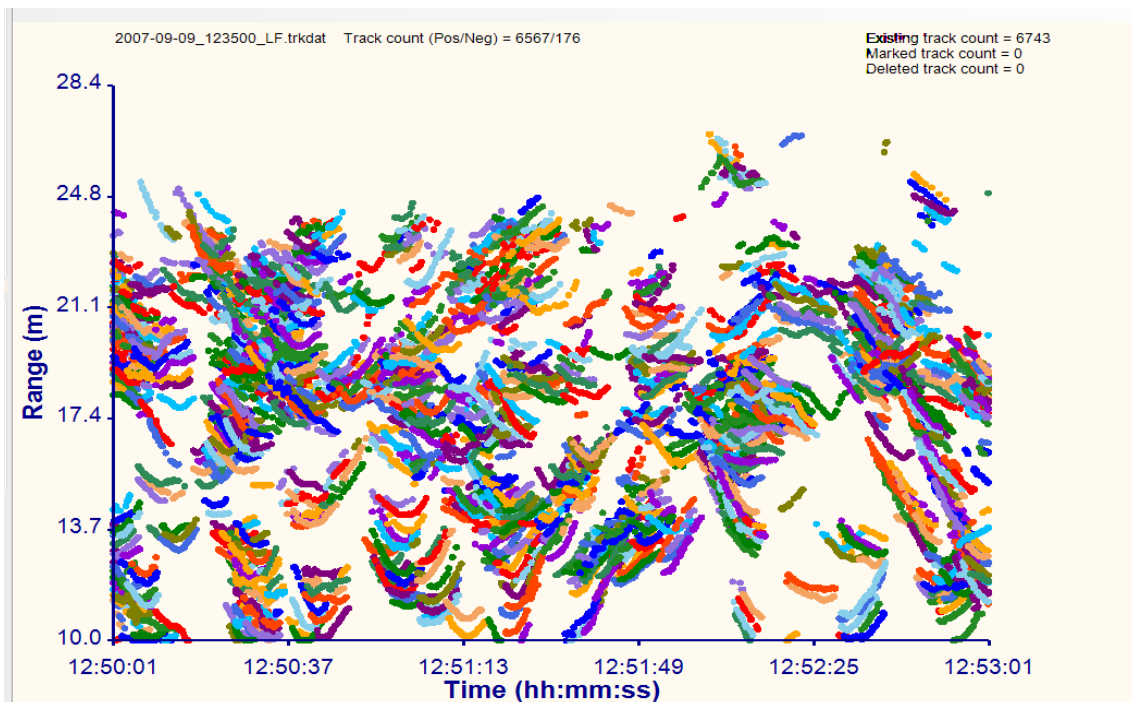


Fig. 8.: Extremely high passage of pink causes trouble for our current approach.

Another challenge is interference of noise, which may result from bubbles, debris, and milling or holding fish, but this is relatively easy to tackle by applying a filter. As described in Section 3, we first track moving targets in the image data and generate corresponding track files. Then a set of fish behavior parameters can be calculated from on the track data, and we further eliminate tracks whose parameters fail to pass certain criteria. These criteria include:

- Target velocity: Stationary or downstream targets would have different velocity from that of upstream targets of interest.
- Target orientation relative to the acoustic beam: If a fish is more parallel with the acoustic beam, the actual acoustic backscatter from the fish would be more vulnerable to noise interference;
- Target position: The closer to the beam axis a target is positioned, the more incident acoustic power the target would receive. Therefore, instead of classifying every tracked

fish in the entire field of view (FOV) of the imaging system, we may select only targets tracked in a smaller view around the center of the FOV, as they moved across the center line of sight. When a fish is close to the beam axis, or its body is perpendicular to the axis, the received backscatter is increased, leading to higher quality imaging data, and potentially better performances in species classification.

We have currently fully integrated the image segmentation algorithm into the tracking software, which allows automatic measurement of the length of every tracked target on individual frames, in addition to target position and backscatter strength. The calculation of feature variables has been implemented in FACE. The current plan is to generate track data and length measurement data from the tracking software, which are then fed to FACE for species composition estimation. We plan to continue to work with the Hydroacoustic Group of the Pacific Salmon Commission, on more data testing, including tests on ARIS data (a different version of image data), and come up with a reasonable set of track filtering parameters. After we come up with a set of optimal parameters, we can integrate more components into the tracking software.

5 RECOMMENDATIONS

One of the most critical prior parameters in the mixture model is the number of classes. In addition to the species of interest (sockeye and pink), there exist local resident fish (northern pikeminnow) in the early run of sockeye, and larger fish such as white sturgeons. So it may become necessary to introduce a third or fourth class in due time. For more robust estimation of species composition, we may want to be able to estimate the number of distinct classes first before applying the mixture model. Some unsupervised clustering algorithms, such as hierarchical algorithms (Theodoridis and Koutroumbas, 2003) have the ability to estimate the number of clusters based on a certain criterion. Such an approach also has an additional benefit that allows separation of local fish from sockeye in the early run of sockeye when pink are absent and helps to reduce human labor needed to perform manual estimates of local fish.

High fish passage occurs when pink return in large amounts. Our current approach becomes less effective when fish passage rate increases to 10,000-30,000 per hour. In this case, it would be necessary to find a method to determine whether one species completely dominates the underlying population, in which case we can assume that there is only one species.

6 ACKNOWLEDGEMENTS

We have collaborated with the Stock Monitoring Group at the Pacific Salmon Commission (PSC) in this project, who have provided the DIDSON data collected at Mission from the 2004-2007 management seasons, and valuable suggestions at various stages of this project. They have also kindly provided manually measured fish length data for performance assessment of the algorithms developed in this project.

7 APPENDICES

7.1 Region-Based Image Segmentation

Image segmentation is a process of partitioning an image into regions or objects. The level of partitioning depends on the problem being solved. Segmentation stops when the objects of interest in an application have been isolated. Segmentation algorithms are generally based on one of two properties of intensity values: discontinuity and similarity. For example, edge detection and linking is based on discontinuity. Thresholding and region growing are based on similarity

Region-based segmentation attempts to find regions directly. It typically starts from a set of seed points, or hot spots (e.g. those with highest intensity), and then appends to the seed points those neighboring pixels that have properties similar to the seed properties. This step is called region growth, and it stops when no more pixels satisfy the criteria for inclusion in that region. The selection criterion is an intensity threshold within which the difference between the intensity of selected pixels and the intensity at the seeds must fall. This threshold must be relaxed enough to allow sufficient pixels in for size measurement. Note that depending on how the threshold is selected, some additional noise (individual isolated pixels) will be included. To minimize noise, we apply a connectivity algorithm to filter out the noise. This algorithm reconstructs regions around the seeds by selecting pixels that are connected with the seeds, thus eliminating those isolated pixels.

7.2 Expectation Maximization

General Description

The Expectation Maximization (EM) algorithm is suited for cases where the available data set is incomplete. For example, a set of feature variable data is known to have been drawn from G groups, but we lack the information regarding which sample is drawn from which group. We can, however, define a complete data set $\mathbf{y}_k = (\mathbf{x}_k, j_k)$, where j_k indicates that the k -th sample is drawn from the j -th group. Then we have (Theodoridis, and Koutroumbas, 2003)

$$p(\mathbf{x}_k, j_k; \boldsymbol{\theta}) = p(\mathbf{x}_k | j_k; \boldsymbol{\theta})P_{j_k},$$

where $\boldsymbol{\theta}$ is a set of unknown parameters in the probability distribution, and P_j is the prior probability which is also unknown but summed to one over all classes. Assuming mutual independence among the samples, we can define a log likelihood function as

$$L(\boldsymbol{\theta}) = \sum_{k=1}^N \ln(p(\mathbf{x}_k | j_k; \boldsymbol{\theta})P_{j_k}).$$

The EM algorithm first takes the expectation of the likelihood function (*E-Step*) over the unobserved data (j_k). That is,

$$\begin{aligned}
 Q(\boldsymbol{\theta}, \mathbf{P}) &= E \left\{ \sum_{k=1}^N \ln(p(\mathbf{x}_k | j_k; \boldsymbol{\theta}) P_{j_k}) \right\} = \sum_{k=1}^N E \left\{ \ln(p(\mathbf{x}_k | j_k; \boldsymbol{\theta}) P_{j_k}) \right\} \\
 &= \sum_{k=1}^N \sum_{j_k=1}^G P(j_k | \mathbf{x}_k; \boldsymbol{\theta}) \ln(p(\mathbf{x}_k | j_k; \boldsymbol{\theta}) P_{j_k})
 \end{aligned}$$

where $\boldsymbol{\Theta}=(\boldsymbol{\theta}, \mathbf{P})$ represents all the unknown parameters to be found. Since for each k , we sum up over all possible j and these are all the same for all k , we can simply drop the subscript k from j . The above equation becomes

$$Q(\boldsymbol{\Theta}) = \sum_{k=1}^N \sum_{j=1}^G P(j | \mathbf{x}_k; \boldsymbol{\Theta}) \ln(p(\mathbf{x}_k | j; \boldsymbol{\Theta}) P_j).$$

The posterior probability can be calculated based on Bayes' rule:

$$\begin{aligned}
 P(j | \mathbf{x}_k; \boldsymbol{\Theta}) &= \frac{p(\mathbf{x}_k | j; \boldsymbol{\Theta}) P_j}{p(\mathbf{x}_k; \boldsymbol{\Theta})} \\
 p(\mathbf{x}_k; \boldsymbol{\Theta}) &= \sum_{j=1}^G p(\mathbf{x}_k | j; \boldsymbol{\Theta}) P_j
 \end{aligned}$$

Then the algorithm maximizes (*M-Step*) the above function with respect to $\boldsymbol{\Theta}$. However, this step will lead to a set of nonlinear equations, which can be solved via an iterative approach.

Applying to Normally Distributed Feature Variables

Here we consider a special case where feature variables are normally distributed. The expectation function is then (ignoring constants):

$$Q(\boldsymbol{\Theta}) = \sum_{k=1}^N \sum_{j=1}^G P(j | \mathbf{x}_k; \boldsymbol{\Theta}) \left(-\frac{1}{2} \ln \mathbf{R} - \frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})' \mathbf{R}^{-1} (\mathbf{x} - \boldsymbol{\mu}) + \ln P_j \right)$$

Taking derivatives with respect to the unknown parameters and setting the derivatives to zeros, we have

$$\begin{aligned}
 P_j(t+1) &= \frac{1}{N} \sum_{k=1}^N P(j | \mathbf{x}_k; \Theta(t)) \\
 \boldsymbol{\mu}_j(t+1) &= \frac{\sum_{k=1}^N P(j | \mathbf{x}_k; \Theta(t)) \mathbf{x}_k}{\sum_{k=1}^N P(j | \mathbf{x}_k; \Theta(t))} \\
 \mathbf{R}_j(t+1) &= \frac{\sum_{k=1}^N P(j | \mathbf{x}_k; \Theta(t)) (\mathbf{x}_k - \boldsymbol{\mu}_j(t)) (\mathbf{x}_k - \boldsymbol{\mu}_j(t))'}{\sum_{k=1}^N P(j | \mathbf{x}_k; \Theta(t))}
 \end{aligned}$$

where $\Theta(t)$ is the parameter estimate at the current step. A constraint that the sum of P_j equals one has also been used in the maximization with a Lagrangian multiplier. All the next iteration estimates depend only on the posterior probability at the current iteration. The initial parameter inputs, $\Theta(0)$, can be provided by training data. A simple test can be used for convergence:

$$\frac{\|\Theta(t+1) - \Theta(t)\|}{\|\Theta(t)\|} < \varepsilon.$$

If the covariance matrix can be expressed as

$$\mathbf{R} = \begin{bmatrix} \sigma_1^2 & \dots & 0 \\ \cdot & \sigma_i^2 & \cdot \\ 0 & \dots & \sigma_M^2 \end{bmatrix}$$

Then the normal distribution can be expressed as

$$p(\mathbf{x}_k | j; \Theta) = \frac{1}{(2\pi)^{M/2} \left| \prod_{i=1}^M \sigma_{ij}^2 \right|^{1/2}} \exp\left(-\frac{1}{2} \sum_{i=1}^M \frac{(x_{ik} - \mu_{ij})^2}{\sigma_{ij}^2}\right)$$

and the expectation function as (ignoring constants)

$$Q(\Theta) = \sum_{k=1}^N \sum_{j=1}^G P(j | \mathbf{x}_k; \Theta) \left(-\frac{1}{2} \sum_{i=1}^M \ln \sigma_{ij}^2 - \frac{1}{2} \sum_{i=1}^M \frac{(x_{ik} - \mu_{ij})^2}{\sigma_{ij}^2} + \ln P_j \right).$$

The estimate of the covariance matrix can be reduced to

$$\sigma_{ij}^2(t+1) = \frac{\sum_{k=1}^N P(j | \mathbf{x}_k; \Theta(t)) (x_{ik} - \mu_{ij}(t))^2}{\sum_{k=1}^N P(j | \mathbf{x}_k; \Theta(t))}$$

8 REFERENCES

Gonzalez, R. C. and R. E. Woods, 2002: *Digital Image Processing, 2nd edition*, Prentice Hall.

Han, J. and M. Kamber, 2005: *Data Mining: Concepts and Techniques, 2nd edition*. Morgan Kaufmann.

Haralabous, J. and S. Georgakarakos, 1996: *Artificial Neural Networks as a Tool for Species Identification of Fish Schools*. ICES J. Mar. Sci., 53, pp. 173-18

Theodoridis, S. and K. Koutroumbas, 2003: *Pattern Recognition, 2nd edition*, Academic Press.

Vitech Innovative Research and Consulting, 2010: In-Season Estimation of Salmon Species Composition using DIDSON Sonar Image Data Collected at Mission. SEF 2009 Report submitted to Pacific Salmon Commission.

Vitech Innovative Research and Consulting, 2008: A feasibility study of using DIDSON imaging sonar to estimate species composition at Mission. SEF 2007 Report submitted to Pacific Salmon Commission.