Automated tracking of fish in trawls using the DIDSON (Dual frequency IDentification SONar)

Nils Olav Handegard and Kresimir Williams

Handegard, N. O., and Williams, K. 2008. Automated tracking of fish in trawls using the DIDSON (Dual frequency IDentification SONar). – ICES Journal of Marine Science. 65: 636–644.

An application for the automated tracking of dual-frequency, identification sonar (DIDSON) data was developed and tested on fish observations taken in midwater trawls. The process incorporates target detection, multiple target tracking, and the extraction of behaviour information such as target speed and direction from the track data. The automatic tracker was evaluated using three test datasets with different target sizes, observation ranges, and densities. The targets in the datasets were tracked manually and with the automated tracker, using the manual-tracking results as the standard for estimating the performance of the automated tracking process. In the first and third dataset, where the targets were smaller and less dense, the automated tracking performed well, correctly identifying 74% and 57% of targets, respectively, and associating targets into tracks with $<$10% error compared with the manually tracked data. In the second dataset, where targets were dense and appeared large owing to the shorter observation range, 45% of targets were correctly identified, and the track error rate was 21%. Target speed and direction, derived from the tracking data, agreed well between the manual and automatic methods for all three test cases. Automated tracking represents a useful technique for processing DIDSON data, and a valuable alternative to time-consuming, manual data-processing, when used in appropriate conditions.

Keywords: DIDSON, multiple-target tracking, trawl-behaviour observation.

Received 13 June 2007; accepted 14 January 2008; advance access publication 11 March 2008.

N. O. Handegard: Institute for Marine Research, PO Box 1870 Nordnes, 5817 Bergen, Norway. K. Williams: Alaska Fisheries Science Center, National Marine Fisheries Service, 7600 Sand Point Way NE, Seattle, WA 98115, USA. Correspondence to N. O. Handegard: tel: +47 55 236803; fax: +47 55 236830; e-mail: nils.olav.handegard@imr.no.

Introduction

Observation of fish during the trawl-capture process facilitates a better understanding of how behaviour influences catch (Engås, 1994). The relationship between behaviour and catch is important in two contexts; in the management of commercial harvests using trawl gears, and in marine surveys where trawls are used as sampling tools. Therefore, the development of a methodology for fish observation has broad application within fisheries research. Technological advances are continually providing scientists with new equipment and methods for observing these traditionally difficult environments (see the reviews by Urqhuart and Stewart, 1993; Graham *et al.*, 2004).

Directly observing and quantifying fish behaviour within a trawl environment is challenging. Monitoring the large volume of the trawl space with sufficiently high resolution to isolate individual organisms requires technologically sophisticated optical and acoustic instrumentation. Here, we present a method to track individual fish within the trawl using a dual frequency identification sonar (DIDSON; Sound Metrics Corp., Seattle, WA, USA). Although this paper is based on fish behaviour observed inside a trawl, the tracking method is general and can be used for any behavioural observations collected with the DIDSON sonar.

In fisheries research and management, the DIDSON has been established as an instrument platform for observing and enumerating fish passage in rivers (Moursund *et al.*, 2003; Holmes *et al.*, 2005), investigating fish behaviour around static fishing gears such as pots (Rose *et al.*, 2005), and observing fish movements within dynamic fishing gears such as trawls (C. S. Rose, pers. comm.).

The DIDSON is a multibeam sonar that uses an acoustic lens to form the individual beams (Belcher, 2002). Owing to its high operating frequency (1.1–1.8 MHz), the DIDSON provides greater resolution than traditional sonars at ranges much greater than can be accomplished by standard optical devices, especially in turbid and low-light environments (Mueller *et al.*, 2006), effectively bridging the gap between these approaches. In addition, eliminating the need for artificial lighting reduces the potential for biased observations. Although reliable species and size identification of individual targets remains challenging using the DIDSON, the system is very useful in observing and quantifying fish movements. Indeed, DIDSON data are well suited for tracking applications, because the data essentially consist of range and angular deviation.

Target tracking is a well-established field within engineering (see Blackman and Popoli, 1999, for an overview), and the techniques have been effectively applied to fisheries using a split-beam echosounder (Handegard *et al.*, 2005; Handegard, 2007). Multiple-target-tracking (MTT) algorithms are most useful in applications monitoring fish movement, where the tracking system needs to handle several targets within the field of view simultaneously. In trawls, high fish densities are encountered

frequently, requiring well-constructed analyses to establish individual tracks.

Manually tracking fish targets has been effective in some DIDSON projects (C. S. Rose, pers. comm.), but it is time-consuming when large volumes of data need to be processed. Here, we present an automated procedure for target tracking, and compare the results with tracks derived manually. We present a general validation framework for evaluating automated tracking, and give specific examples of fish movement inside a midwater trawl.

Material and methods

We break this section into five parts. First, we present a technical introduction to the DIDSON instrument. Next, we describe test datasets, and follow this with an explanation of the tracking process, both target detection (TD) and multiple target tracking (MTT). Finally, we present the methods for evaluating the efficacy of the automatic tracking. The algorithm is coded in Matlab™, and the source code and test data are available on request from the authors. The Matlab *m*-files are licenced under open source (GNU Public licence), ensuring that further development benefits all users of the code.

DIDSON system description

The DIDSON operates at two frequencies, providing a choice between high resolution—low range (1.8 MHz, $<$20 m)—and lower resolution—longer range (1.1 MHZ, $<$60 m). The beam array is 96 beams at the high-frequency setting, with equivalent beam angles of 0.3° by 12°. This array of beams is assembled into a sector along the narrow dimension of the beams to constitute a 29° × 12° field of view. The data are structured in a polar-coordinate system, the location of each data cell being defined by its beam identity, or its angle deviation relative to the central axis of the transducer, and the range along that beam. The range is subsampled into 256 equal-range bins, resulting in a high-frequency frame resolution of 96 × 256. The lower-frequency setting uses wider individual beams reducing the frame resolution to 48 × 256 over the 29° view. The raw data consist of a matrix of eight bit values (0–255), each representing the return amplitude ranging from 0 to 90 dB with 0.35 dB resolution (Belcher, 2002).

As the beams are arranged in a line, fish movement in the axis orthogonal to the multibeam "fan" is undetectable. The backscatter display provides the illusion of three dimensions when viewed because of its high resolution and the effect of shadowing. This feature can aid in the gross identification of fish, but cannot be taken advantage of in the method presented here, limiting the analysis of the data to two dimensions.

Test data

The DIDSON was deployed over several research and survey cruises conducted by National Marine Fisheries Service scientists in the Gulf of Alaska and the Bering Sea. Three test datasets were chosen to evaluate the automated tracking process, each set representing distinct target sizes and densities (Figure 1; Table 1). For each test dataset, targets were manually detected and associated into tracks using a manual-tracking application developed for the purpose. This application allowed tracking of individual targets by selecting the positions of each visually identified target on the computer display across successive time-steps.

The DIDSON data files were read using a Matlab function provided by Sound Metrics Corp. The target-detection process was run with raw data, uncorrected for acoustic-transmission losses. The DIDSON software can be used to apply a range-dependent compensation for transmission losses, but working with uncompensated data allows the use of a single threshold value for separating targets from background noise, i.e. the noise is not amplified with increasing range.

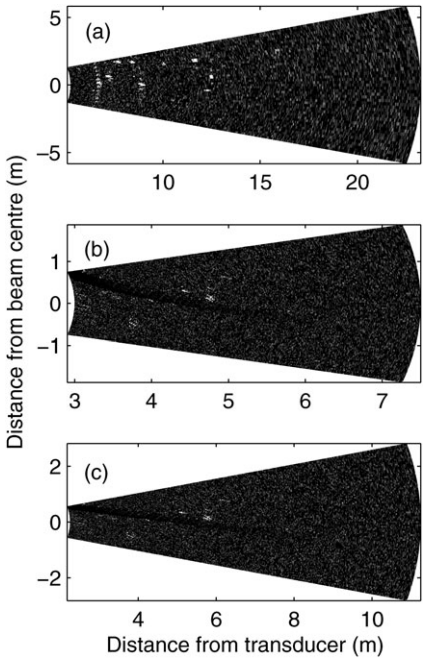

Figure 1. Selected frames from DIDSON test datasets used to evaluate automated tracking. One data frame is shown for each of test datasets 1–3 in panels (a)–(c), respectively. In test datasets 1 and 3, the DIDSON unit was mounted to the bottom panel of an acoustic-abundance survey midwater trawl, and in dataset 2 the unit was mounted within a commercial midwater pollock trawl. Observations were collected aboard research and chartered commercial vessels operating in the Bering Sea and the Gulf of Alaska. Note the different scale ranges.

Target detection

Image-processing techniques are highly suited for processing DIDSON data, because individual sample data (ping) for each beam are assembled into frames, which can be considered an analogue to an image frame in video data. Standard image-processing methods are used, consisting of mean subtraction, filtering,

Table 1. Characteristics of DIDSON test datasets collected in trawls used to evaluate tracker performance.

Parameter	Test dataset		
	1	2	3
DIDSON range (m)	5–23	3–7.5	2–11
Frame rate	6	8	7
Total frames processed	184	174	409
Operating frequency	1.1	1.8	1.8
Mean fish length (cm)	47[a]	44[b]	24[a]
Mean density (targets per frame)	3.29	5.22	1.91

[a]Estimated from catch.
[b]Estimated from DIDSON observations.

thresholding, and finally target isolation and position estimation. All these processes require user-input parameters which influence the target-isolation performance.

Mean subtraction removes the static elements from a series of data frames. For each frame, a mean value for each pixel is taken across several surrounding frames and subtracted from the frame data. Although this technique is most powerful with stationary DIDSON deployments, we utilized it to remove the visible trawl elements in the data. The length of the mean subtraction window was set effectively to reduce the intensity of the static elements of the data while minimizing bias against slow-moving targets. This step was necessary to prevent the tracking process from incorporating the trawl signature into fish tracks.

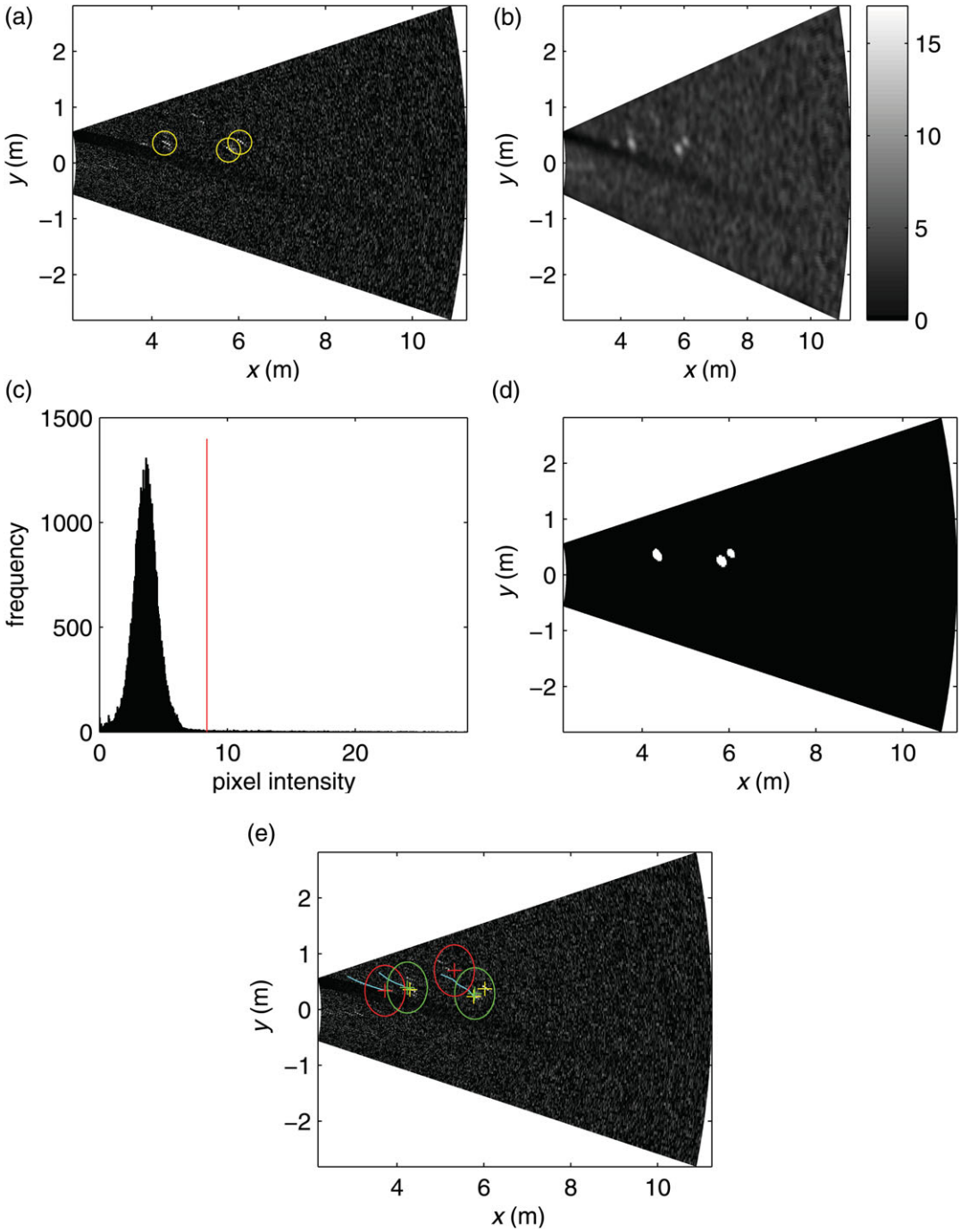

Figure 2. Automated tracking of targets from DIDSON data. Panels (a)–(d) show the separate steps of the target-detection process: (a) raw data image with detected targets (yellow circles), (b) range-dependent filtering, (c) intensity histogram and threshold (red vertical line), (d) threshold application, and (e) data association (tracking). In (e), the blue lines represent live tracks, the yellow "+" signs denote targets detected, the green and red "+" signs the predictions from the previous time-step, and the green and red circles represent the gates. The green and red colours, respectively, indicate whether or not a successful association between a prediction and observation has taken place.

Low-pass filtering is used to reduce high-frequency noise in the data frame (Figures 2a and b). The true spatial dimension of the image pixels increases with range because of beam spreading, so a target of constant size occupies a decreasing number of pixels as it moves away from the transducer face. Additionally, targets often appear fragmented at short ranges owing to target or platform movement, or both factors. Both these issues require the use of a range-dependent filtering technique.

Standard image-convolution-filtering techniques are used, but the size of the filter kernel is dynamically matched to the true spatial dimension, i.e. the filter kernel encompasses more beams nearer the transducer (Appendix A). After filtering, the data were thresholded. When setting the threshold for a set of data frames, the density of targets and the signal distribution should be considered (Figure 2c). A connected components-labelling algorithm (Haralick and Shapiro, 1992) was applied to the frame following thresholding, to isolate contiguous pixel regions representing individual targets (Figure 2d). These regions were accepted as valid targets if the number of pixels occupied was within an acceptable range of values, separating very large regions likely consisting of several overlapping targets, or very small regions that were likely not to be valid targets.

To assign an intensity-weighted single position to a group of contiguous pixels, a spatial mean of the pixel intensity values within the contiguous region was taken. The position of the detected target was

$$\begin{bmatrix} \theta \\ r \end{bmatrix} = \frac{\sum_i \begin{bmatrix} \theta_i \\ r_i \end{bmatrix} I_i^p}{\sum_i I_i^p}, \tag{1}$$

where θ_i and r_i are the angle and range values defining the location of pixel i in the accepted region, I_i the intensity for that pixel, and p a power parameter. The power parameter determines how the intensity of the pixels should be treated in the estimation. If $p = 0$, the mean spatial position of the associated pixels is taken as an estimate, disregarding the corresponding intensity. If $p = 1$, the pixel positions are linearly weighted by the pixel intensity, but for $p > 1$, relatively more weight is placed on the most intense pixel. For a very large value, this is, in effect, the position of the most intense pixel.

Multiple-target tracking

MTT algorithms are standard tools (for a general introduction, see Blackman and Popoli, 1999). The aim of MTT is to associate single targets from the target-detection process to tracks, and to estimate the position and velocity (the "state") based on the associated observation and track history.

The TD process derives detected targets represented in polar coordinates, but the position and velocity estimates are more useful in Cartesian coordinates. Therefore, throughout the tracking procedure, it is necessary to transfer position data between these two systems, which are defined as "observation space" and "state space", respectively. These concepts and their conversions are presented in detail in Appendix B. There may be errors associated with both systems, particularly measurement errors and model errors associated with the assumption of constant velocity of the targets. Here, we give an overview of the different steps of the MTT used in this work, consisting of gating, target association, track estimation, prediction, and track maintenance.

Gating limits the targets that are considered for association with an existing track. Target association is the actual consolidation of targets into tracks. The input to this process is the detected targets and predictions from existing tracks. Predictions are based on an assumption of constant velocity between time-steps, i.e. the estimated position and velocity from the previous time-step is used to predict the position for the current one. This position is converted to polar coordinates, and a gate distance d_{ij} [see Appendix C, Equation (10)] between observation i and prediction j is calculated. This distance is a measure that takes both observation error and fish behaviour into account. The gating simply flags the pairs that are unrealistic.

The next step associates observations with the predictions, given that there are some observations inside the gate. This process is graphically represented in Figure 2e. If several targets fall within the gate, the closest target in terms of d_{ij} is associated with the existing track. If a single target falls within the gates of several tracks, i.e. that several tracks potentially could incorporate the same observation, the global-nearest-neighbour method (Blackman and Popoli, 1999) is used to associate the target to the most appropriate track. This is solved by the Bertsekas auction algorithm (Bertsekas, 1990).

After the observations and predictions are associated, they are used to calculate the estimated state, consisting of position and velocity information. This is achieved using an alpha–beta, fixed-gain filter (Blackman and Popoli, 1999). The alpha and beta parameters are constants that determine how much weight should be placed on the predicted target position, based on track history, as opposed to the observed target position. The alpha parameter acts on the predicted and observed positions in the current time-step, and the beta parameter weights the velocity prediction for estimating the position of the target in the following time-step, formally defined as

$$\begin{bmatrix} \hat{x} \\ \hat{y} \end{bmatrix} = \begin{bmatrix} \tilde{x} \\ \tilde{y} \end{bmatrix} + \alpha \left\{ \begin{bmatrix} x_y \\ y_y \end{bmatrix} - \begin{bmatrix} \tilde{x} \\ \tilde{y} \end{bmatrix} \right\} \tag{2}$$

$$\begin{bmatrix} \hat{\dot{x}} \\ \hat{\dot{y}} \end{bmatrix} = \begin{bmatrix} \tilde{\dot{x}} \\ \tilde{\dot{y}} \end{bmatrix} + \frac{\beta}{T} \left\{ \begin{bmatrix} x_y \\ y_y \end{bmatrix} - \begin{bmatrix} \tilde{x} \\ \tilde{y} \end{bmatrix} \right\}, \tag{3}$$

where x_y and y_y are the observations mapped to Cartesian coordinates, where the "$\wedge$" and "$\sim$" symbols denote the estimate and prediction, respectively, the dot over the variables the time derivative (i.e. the velocity component), and x and y the position in Cartesian space (see Appendix B).

Targets that have not been assigned to any track are used to initiate new tracks, with an assumption of an initial velocity (e.g. the water flow within the trawl relative to the sonar). Tracks that are not associated to any targets over a predetermined maximum number of frames are ended. At this point, all non-closed tracks have an estimate of position and velocity. By assuming constant velocity, the position of a target can be predicted for the next data frame. This prediction is the starting point of the MTT process for the following time-step.

The alpha–beta filter is used in data association only. Once targets are associated into tracks, several methods may be used to extract information to describe the behaviour of individual fish or groups of fish. The true movement of the target may be estimated by applying a linear regression, a smoothing spline, or another standard model to the track data, reducing the influence of observation errors on the velocity estimates. Spline estimates

have worked well for split-beam acoustic data (Handegard *et al.*, 2005). Here, the target positions and velocities were estimated by fitting a smoothing spline to the associated target-position estimates. Velocities were decomposed into speed and direction components by converting the Cartesian velocity vectors to polar coordinates. This provided the basis for estimating the mean speed and direction per track. Mean direction was calculated using circular-statistics methods (Zar, 1999).

Evaluation of automated tracking

An important part of any automated tracking analysis is to measure its performance. Here, we present a method of evaluation by comparing it with the manual-tracking data, which are generally considered to be more accurate. The TD and MTT components were evaluated independently using a set of performance metrics.

The first step in comparing the TD algorithm with manual detections is to pair targets detected by the two methods. This is accomplished by calculating the distance between the sets of targets in each frame, similar to the gating procedure presented above. If the distance is within an acceptable threshold, set at 0.4 m for the comparisons here, the automatic- and manual-process targets are considered to be the same. "Missed" targets are true targets not detected by the TD process, and "false" targets are erroneous detections by the TD process that do not represent valid targets, with the assumption that the manual detections are without error. The paired, missed, and false targets are then counted for each dataset.

To test the performance of the MTT process, the manually associated tracks ("true" tracks), are compared with the tracks obtained from the MTT algorithm ("auto" tracks). Two types of error are observed. The first type is called a splitting error (J_{split}), and occurs when the observations along a single true track have corresponding observations associated with more than one auto-track, meaning that the MTT algorithm failed to associate the targets along the true track one or more times. The J_{split} value is defined as

$$J_{\text{split}} = \frac{\sum_i C_i^s}{\sum_i (L_i - 1)}, \quad (4)$$

where C_i^s is the number of changes in the auto-track identifier along the true track i, and L_i is the length of the true track i. The second type of association error, termed a connect error (J_{connect}), is the inverse situation, where observations belonging to a single auto-track have corresponding observations along several true tracks. It is given by

$$J_{\text{connect}} = \frac{\sum_i C_i^c}{\sum_i (L_i - 1)}, \quad (5)$$

where C_i^c is the number of changes in the true-track identifier along auto-track i, and L_i is the length of auto-track i. Note that this measure is dependent on tracking-parameter settings. The same error definitions were used for testing the performance of the data-association algorithm for tracking from split-beam echosounders (see Figure 3 of Handegard *et al.*, 2005).

Results

The results from the methods evaluation are given here. As described above under Material and methods, the TD and MTT components of the tracking process are evaluated separately, along with the behaviour metrics derived from the track data. The results provide insight into the condition-specific limitations of the approach, and a general level of accuracy that can be expected. The parameter settings used for the different test datasets are given in Table 2. The parameters were set based on an iterative approach, evaluating the tracking performance after each trial.

Test dataset 2 had the lowest proportion of targets detected by both automated and manual approaches relative to the total manually detected targets (~25%). In this test dataset, larger fish observed over a narrow range interval, a larger pixel area is occupied by the targets than in the other datasets. Moreover, the target density in dataset 2 was the highest of the three datasets. With smaller fish, as encountered in dataset 3, there was a substantial incidence of missed targets. Dataset 1 had the highest success rate of the three trawl environments observed.

The poorest MTT performance was also observed with dataset 2, with the highest incidence of split and join errors (Table 3). The difference in mean track length and total track counts was the highest in this dataset, indicating many false associations, and a higher number of track splits relative to the other datasets. The poor MTT performance was probably a result of the large proportion of false targets detected, suggesting that the overall automated process accuracy is heavily dependent on the quality of the TD process. The automated processing performance of datasets 1 and 3 was a better comparison with the manual data.

One of the goals of target tracking in the context of fisheries science is to provide quantitative descriptors of fish behaviour. Aggregate level information, such as mean swimming speed or mean swimming direction relative to some reference, is useful in characterizing general fish behaviour, and can be used to compare different populations. In this application, the results of the automatic tracking compare favourably with the manual

Table 2. Parameters used in the tracking software for the three DIDSON example datasets.

Parameter	Test dataset			Unit
	1	2	3	
Mean subtraction window	10	10	20	Number of frames
Low-pass target size σ	0.10	0.06	0.06	m
Threshold	12.1	13.0	8.4	n/a
Minimum target size	5	5	5	Number of pixels
Maximum target size	500	500	500	Number of pixels
Power parameter P	4	4	4	n/a
[a]Gate parameter θ_G	1	1	1	degrees
[a]Gate parameter r_G	0,1	0.1	0.1	m
[a]Gate parameter x_G	0.54	0.48	0.45	m
[a]Gate parameter y_G	0.54	0.48	0.45	m
Filter parameter α	0.9	0.9	0.9	n/a
Filter parameter β	0.5	0.5	0.5	n/a
Spline smoothing λ	10	10	10	n/a
[b]Initial velocity $\dot{x_o}$	0	0	0	m s^{-1}
[b]Initial velocity $\dot{y_o}$	0	0	0	m s^{-1}
Missing pings	2	2	2	n/a

n/a, no unit (i.e. not applicable).
[a]Defined in Appendix C [Equation (11) and (12)].
[b]Defined in Appendix B [Equation (9)].

Table 3. Comparison of TD and data association using the DIDSON automated-tracking application with manual-tracking data on three example datasets collected from trawls.

Parameter	Test dataset		
	1	2	3
Matched targets	456	382	500
False targets	59	149	103
Missed targets	149	526	281
Total tracks (auto)	148	95	244
Total tracks (man)	137	148	208
Mean track length (auto)	5.6	7.9	7.3
Mean track length (man)	5.6	13.2	7.5
J_{split} errors	0.03	0.08	0.05
$J_{connect}$ errors	0.06	0.1	0.03

standard (Figure 3), a greater difference between methods being observed in mean speed estimates than in mean movement direction.

When comparing manual and automatic methods, we also have to consider the time investment. Manual tracking required an average of 3.5 s per frame to process, and with the typical data collection frame rate of 8 frames s^{-1}, 1 h of DIDSON data would translate to ~28 h of manual processing. In contrast, the automated method processed 2 frames s^{-1} using a computer with a 2 GHz processor, although some time was required for parameter settings where an iterative approach using subsets was needed.

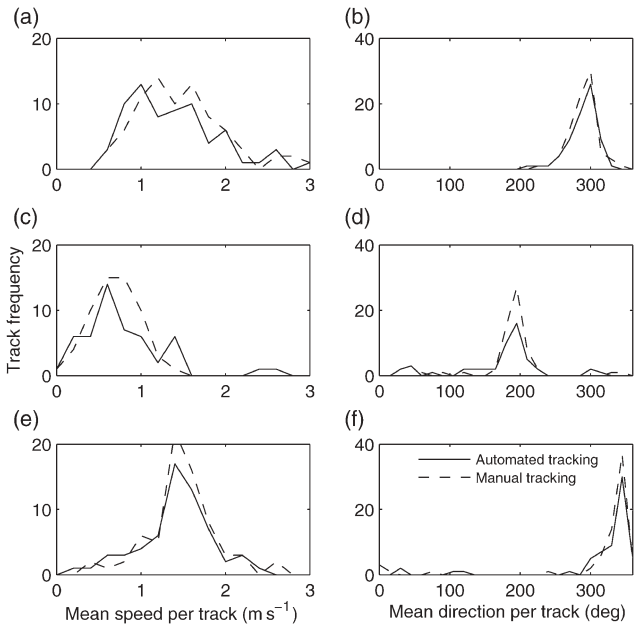

Figure 3. (a), (c), (e) Comparison of target speed and (b), (d), (f) direction estimates from the DIDSON data examples, using automated tracking (solid line) and manual tracking (dashed line). Panels (a) and (b) are dataset 1, (c) and (d) dataset 2, and (e) and (f) dataset 3.

Discussion

We have shown in this study that using an automated approach to tracking fish in a trawl can yield reliable results for the dataset and conditions with which we experimented. As with any such analysis, automated tracking requires a basic understanding of the underlying processes, which in turn provides a reasonable level of confidence in the results.

Tracking is a powerful method for quantifying behaviour. For example, when studying fish behaviour in trawls, it may be useful to measure the nearest approach of a fish to some object, such as a trawl panel. This can be accomplished using track data combined with an estimate of the trawl-panel position. Preferred positions of different species within the trawls, as well as their residence times in certain trawl sections, are useful in the design of trawl gears to reduce or improve selectivity (Albert *et al.*, 2003; Piasente *et al.*, 2004). Tracking methods can provide the means for measuring the interaction between fish and the trawl surfaces, allowing the objective categorization of certain patterns, such as active escapement or avoidance. The DIDSON provides an easily deployable platform for observing fish movement, and when combined with the tracking approach presented here, can yield valuable quantitative data.

Automatic TD varied in success across the test datasets, as would be expected given the different target environments. The performance of the automated tracking method depended in large part on the selection of input parameters. Determining optimal parameters for a certain data environment requires an iterative approach, the tracking performance being evaluated visually using a subset of data, to optimize the input parameters. When choosing the parameter values for the TD process, the threshold setting was most sensitive. In the MTT process, manipulating the gate-size settings exerted the greatest impact on the perceived visual-data association.

The greatest disagreement between manual and automated tracking was when a high density of large fish were observed at close range (test dataset 2). Larger fish targets tend to have variable target strength along their bodies, often producing strong returns from the anterior portion of the body and from the tail. This can result in fragmentation of targets into several additional "false" ones. Likewise, small targets such as the juvenile fish in test dataset 3 can be difficult to isolate from the background noise in the data, resulting in many smaller targets being missed by the TD process. In all datasets, we infer that acoustically weaker targets had a lesser probability of being detected, which can result in an overall size bias in the tracking results. This effect is also relevant for various levels of background noise, introducing biases for low signal-to-noise ratios. Therefore, conditions of high variability in target size, background noise, or intensity should be avoided if any general inferences are to be made about the data.

A critical part of the tracking process is estimating the track from the associated positions of a fish target. To describe fish motion from position data, we need to fit an appropriate model. The complexity of the model chosen depends on the range of motions anticipated, and on the questions asked of the data. Robust and simple models such as linear regression often work well where targets do not change their direction of movement very much. With the DIDSON-data examples, we used splines to model fish movement. This is a useful method because the level of smoothness can be adjusted. If the level of smoothness is

high, the fitted model will be closer to a straight line, and if the level of smoothing is low, complex behavioural patterns can be extracted. If abrupt changes or discontinuities in fish movement, such as a reaction to some object in the path of the fish, are being investigated, the spline model may be less useful because it assumes similar smoothing along the whole track. In such cases, a segmented regression may prove more suitable. Although panel-avoidance reactions are discontinuous, i.e. the fish abruptly change their course after perceiving or striking the trawl mesh, we were able to capture this behaviour using the spline method.

Despite some inequalities in the performance of the TD and MTT components across the three datasets used here, the higher-level estimates of average track speed and direction were similar between the methods in all the test datasets (Figure 3). Therefore, an ability to describe the behaviour of the different fish-size groups under different observation conditions in the three test datasets was not greatly reduced by applying automation, and the method did not impart considerable artefact or bias to these measures, compared with the manual set.

In conclusion, automated tracking is a useful tool for characterizing the behaviour of organisms, and can be particularly useful in quantifying fish movement within the trawl when using a DIDSON. Although overall performance based on the example datasets was encouraging, the confidence placed in automated tracking depends on the level of information sought from the data. In the test datasets presented here, automated tracking was useful in identifying behaviour patterns over aggregated species or size groups of organisms, rather than isolating individual fish behaviours from mixed target-track data. The best conditions for automated tracking are where homogeneous targets are observed in moderate density. When used under such conditions, automated tracking provides a promising method for extracting quantitative fish-behaviour information from observations made using the DIDSON system.

Acknowledgements

We are greatly indebted to Craig Rose of the Alaska Fisheries Science Center for advice and sample data, and to Ed Belcher from Sound Metrics Inc. for providing code for accessing DIDSON data with Matlab.

References

Albert, O. T., Harbitz, A., and Høines, Å. S. 2003. Greenland halibut observed by video in front of survey trawl: behaviour, escapement and spatial pattern. Journal of Sea Research, 50: 117–127.

Belcher, E. 2002. Dual-frequency Identification Sonar. DIDSON operation manual, 4.47.15. Sound Metrics, Seattle, WA. 44 pp.

Bertsekas, D. P. 1990. The auction algorithm for assignment and other network flow problems: tutorial. Interfaces, 20: 133–149.

Blackman, S. S., and Popoli, R. 1999. Design and Analysis of Modern Tracking Systems. Artech House, MA.

Engås, A. 1994. The effects of trawl performance and fish behaviour on the catching efficiency of demersal sampling trawls. *In* Marine Fish Behaviour, pp. 45–68. Fishing News Books, Oxford.

Graham, N., Jones, E. G., and Reid, D. G. 2004. Review of technological advances for the study of fish behaviour in relation to demersal fishing trawls. ICES Journal of Marine Science, 61: 1036–1043.

Handegard, N. O. 2007. Observing individual fish behaviour in fish aggregations: tracking in dense aggregations. Journal of the Acoustical Society of America, 122: 177–187.

Handegard, N. O., Patel, R., and Hjellvik, V. 2005. Tracking individual fish from a moving platform using a split-beam transducer. Journal of the Acoustical Society of America, 118: 2210–2223.

Haralick, R. M., and Shapiro, L. G. 1992. Computer and Robot Vision, 1. Addison-Wesley, Reading, MA. 672 pp.

Holmes, J. A., Cronkite, G. M. W., Enzenhofer, H. J., and Mulligan, T. J. 2005. Accuracy and precision of fish-count data from a "dual-frequency identification sonar" (DIDSON) imaging system. ICES Journal of Marine Science, 63: 543–555.

Moursund, R. A., Carlson, T. J., and Peters, R. D. 2003. A fisheries application of a dual-frequency, identification-sonar acoustic camera. ICES Journal of Marine Science, 60: 678–683.

Mueller, R. P., Brown, R. S., Hop, H., and Moulton, L. 2006. Video- and acoustic-camera techniques for studying fish under ice: a review and comparison. Reviews in Fish Biology and Fisheries, 16: 213–226.

Piasente, M., Knuckey, I. A., Eayrs, S., and McShane, P. E. 2004. *In situ* examination of the behaviour of fish in response to demersal trawl nets in an Australian trawl fishery. Marine and Freshwater Research, 55: 825–835.

Rose, C. S., Stoner, A. W., and Matteson, K. 2005. Use of high-frequency imaging sonar to observe fish behaviour near baited fishing gears. Fisheries Research, 76: 291–304.

Urqhuart, G. G., and Stewart, P. A. M. 1993. A review of techniques for the observation of fish behaviour in the sea. ICES Marine Science Symposia, 196: 135–139.

Zar, J. H. 1999. Biostatistical Analysis. Prentice Hall, NJ.

Appendix A

Range-dependent, low-pass filtering

Here, we define the range-dependent, low-pass filter used in the target-detection process. The weights in the 2D Gaussian filter are given by the bivariate distribution:

$$P(X_1, X_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{z}{2(1-\rho^2)}\right\},$$

where

$$z = \frac{(X_1-\mu_1)^2}{\sigma_1^2} - \frac{2\rho(X_1-\mu_1)(X_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(X_2-\mu_2)^2}{\sigma_2^2},$$

and

$$\rho = \mathrm{corr}(X_1, X_2) = \frac{\sigma_{12}}{\sigma_1\sigma_2}.$$

If we assume a circular target in state space, the bivariate distribution reduces to

$$P(x,y)_{e_g} = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(x-\mu_1)^2}{2\sigma^2}\right) \exp\left(-\frac{(x-\mu_2)^2}{2\sigma^2}\right), \quad (6)$$

where $[\mu_1\ \mu_2]$ is the position of the filtered pixel, $[x\ y]$ the raw-image positions, and $\sigma = \sigma_1 = \sigma_2$ a bandwidth parameter for the filter that should be set based on the size of the fish targets. This filter is in state space (e_g, see Appendix B), but the image to be filtered is in observation space (e_p). The filter conversion is given by

$$P(\theta, r)_{e_p} = \mathrm{H}P(x,y)_{e_g}\mathrm{H}^{\mathrm{T}},$$

where H is the linearized, coordinate-transformation matrix between state space and observation space [Equation (8), Appendix B], and T the matrix transpose. When ρ is zero, the filter may be separated because the functions of θ and r range separately, reducing the computational cost. We have made this assumption in our implementation of the method, but this can be changed if the targets are larger and hence less circular.

Appendix B

Observation space vs. state space

The idea of "state space" and "observation space" is widely used in physics. State space contains the variables of interest, whereas the observation space contains the observations. In this case, the state space includes the information of the fish behaviour, $\mathbf{x}_{ik}=[x\ y\ \dot{x}\ \dot{y}]^{\mathrm{T}}$, where x is the position away from the transducer face, y the position perpendicular to x across the beams, and $\dot{x}$ and $\dot{y}$ their respective velocities, at time t_k and track i. T is the matrix transpose. The observation space contains the observations from the sonar, i.e. the detected targets inside the echo beam represented by $\mathbf{y}_{jk}=[\theta\ r]^{\mathrm{T}}$, where θ is the angle of the beam and r the range along that beam, for target j at time t_k. This state space and observation space is denoted by e_g and e_p, respectively.

For tracking purposes, it is necessary to change between these coordinate systems. To do so,

$$\mathbf{y}_{ik} = \mathbf{h}(\mathbf{x}_{ik}) \tag{7}$$

is defined. More specifically,

$$\begin{bmatrix}\theta\\ r\end{bmatrix} = \begin{bmatrix}h_1(x,y)\\ h_2(x,y)\end{bmatrix} = \begin{bmatrix}\arctan(y/x)\\ \sqrt{x^2+y^2}\end{bmatrix},$$

is defined. In addition, the linearized version of this mapping,

$$\mathrm{H} = \begin{bmatrix}\frac{\partial h_1}{\partial x} & \frac{\partial h_1}{\partial y}\\ \frac{\partial h_2}{\partial x} & \frac{\partial h_2}{\partial y}\end{bmatrix}, \tag{8}$$

is defined. This is used to map the gates (see Appendix C) and filters (see Appendix A) defined in e_g to observation coordinates e_p. The inverse mapping from observation space to state space is not directly possible because there is no information on velocity in observation space. This mapping is required for initiating new tracks, and we get around this by assuming an initial velocity

$$\mathbf{v}_0 = [\dot{x}_0\ \dot{y}_0]^{\mathrm{T}}. \tag{9}$$

Appendix C

Gating

To associate predictions and observations, the distance between the prediction and a subset of the total available targets that fall within the "gate" has to be computed for each track. First, for each ping k, we define

$$\boldsymbol{\varepsilon}_{ij} = \mathbf{h}(\tilde{\mathbf{x}}_i) - \mathbf{y}_{j,}$$

where $\mathbf{h}$ is the mapping between the state and the observations [Equation (7), see Appendix B], $\tilde{\mathbf{x}}_i$ the predicted state for track i from the last time-frame, and $\mathbf{y}_j$ is observation j. The distance measure between observation and prediction is defined by

$$d_{ij} = \boldsymbol{\varepsilon}_{ij}\mathrm{G}^{-1}\boldsymbol{\varepsilon}_{ij} \leq 1, \tag{10}$$

where G is a 2×2 element matrix defining the gate. This distance measure is used in gating and for data association. Distances >1 are deemed outside the gate and are not considered in the subsequent data association. The size of the gate is determined by two factors: (i) the potential observation error attributable to measurement limitations, and (ii) the target's behaviour, specifically its potential to change velocity within one time-step (time-lag between data frames).

Both fish behaviour and observation error affect how the gate should be set. If one assumes that there is no error in the prediction model, that velocity is constant, and that errors in θ and r are independent and random, i.e. no range or angle-dependent error in both variables, the gate can be defined simply as

$$\mathrm{G}_0 = \begin{bmatrix}\theta_G^2 & 0\\ 0 & r_G^2\end{bmatrix}. \tag{11}$$

where θ_G and r_G are parameters setting the gate size. These can, for example, be set to twice the standard deviation for θ and r, respectively. This gate is an ellipsis in observation space (see Figure A1, broken line), and is constant throughout the space. If we assume on the other hand a maximum deviance per frame from the straight-line swimming model in both directions as x_G and y_G, respectively, the gate could be defined as

$$\mathrm{G}'_b = \begin{bmatrix}x_G^2 & 0\\ 0 & y_G^2\end{bmatrix}. \tag{12}$$

The problem with this formulation is that it is based on the Cartesian coordinate system. The transformation from e_g to e_p coordinates is required. This is achieved by

$$G_b = \mathrm{H}\mathrm{G}'_b\mathrm{H}^{\mathrm{T}}, \tag{13}$$

where H is the linear coordinate transformation matrix [Equation (8), see Appendix B]. Closer to the transducer, the constant physical

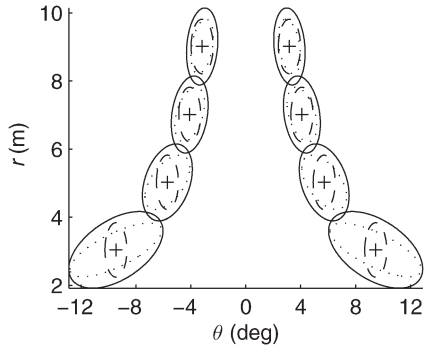

Figure A1. Gates G_o, G_b, and G are denoted with a dashed, dotted, and solid line, respectively. Note how G_b is rotated and scaled in e_p. Parameters are $x_G = 0.8$ m, $y_G = 0.12$ m, $\theta_G = 0.014$, and $r_G = 0.8$ m.

volume covers a larger proportion of the observation, i.e. H is dependent on r and θ space (see Figure A1, dashed line). Gates have the same properties as covariance matrices. They can be thought of as covariance matrices for prediction $\tilde{\mathbf{x}}$ and observation $\mathbf{y}$. The final gate should therefore be considered a covariance matrix of the innovation $\boldsymbol{\epsilon}$. If we assume that the behaviour is independent of the observation errors, the gate capable of handling both types of error is simply $G = G_0 + G_b$ (see Figure A1, solid line).

doi:10.1093/icesjms/fsn029