Measuring fish and their physical habitats: Versatile 2-D and 3-D video techniques with user-friendly software

Neuswanger, Jason R., Alaska Cooperative Fish and Wildlife Research Unit, Department of Biology and Wildlife, University of Alaska Fairbanks, Fairbanks, AK 99775, USA. Email: jason@troutnut.com. Phone: (917) 628-1747. Fax: (907) 474-7872. Present address: Warnell School of Forestry and Natural Resources, University of Georgia, Athens, GA 30602, USA.

Wipfli, Mark S., U.S. Geological Survey, Alaska Cooperative Fish and Wildlife Research Unit, Institute of Arctic Biology, University of Alaska Fairbanks, Fairbanks, AK 99775, USA. Email: mwipfli@alaska.edu.

Rosenberger, Amanda E., U.S. Geological Survey, Missouri Cooperative Fish and Wildlife Research Unit, Department of Fisheries and Wildlife, University of Missouri, Columbia, MO 65211, USA. Email: rosenbergera@missouri.edu.

Hughes, Nicholas F. Institute of Arctic Biology, University of Alaska Fairbanks, Fairbanks, AK 99775, USA.

Abstract

Applications of video in fisheries research range from simple biodiversity surveys to 3-dimensional (3-D) measurement of complex swimming, schooling, feeding, and territorial behaviors. However, researchers lack a transparently developed, easy-to-use, general-purpose tool for 3-D video measurement and event logging. Thus, we developed a new measurement system, with freely available, user-friendly software, easily obtained hardware, and flexible underlying mathematical methods capable of high precision and accuracy. The software, VidSync, allows users to efficiently record, organize, and navigate complex 2-D or 3-D measurements of fish and their physical habitats. Laboratory tests showed sub-millimeter accuracy in length measurements of 50.8-mm targets at close range, with increasing errors (mostly < 1 %) at longer range and for longer targets. A field test on juvenile Chinook Salmon feeding behavior in Alaska streams found that individuals within aggregations avoided the immediate proximity of their competitors, out to a distance of 1.0 to 2.9 body lengths. This system makes 3-D video measurement a practical tool for laboratory and field studies of aquatic or terrestrial animal behavior and ecology.

Introduction

Video-based methods to observe and measure animals and their behavior have diverse applications in fish research (Shortis et al. 2009), and they are especially useful for species sensitive to handling or difficult to capture (Ellender et al. 2012). The use of calibrated multi-camera systems for measurement, a process known as videogrammetry, enables, in environments with sufficient water clarity, precise determination of three-dimensional (3-D) positions, lengths, velocities, and more complex quantities that provide insights into locomotion (Hughes and Kelly 1996a; Butail and Paley 2012), habitat use (Laurel and Brown 2006; Fischer et al. 2007; Tullos and Walter 2015), and social and predatory behaviors (Hughes et al. 2003; Mussi et al. 2005; Piccolo et al. 2007; Uglem et al. 2009; Neuswanger 2014; Vivancos and Closs 2015). Videogrammetry can provide more precision and less bias than direct visual estimation, even by skilled observers (Harvey et al. 2001). Video also offers qualitative advantages over direct observation for analyzing behavior: (1) ambiguous behaviors, such as territorial conflicts in which the winner is unclear, can be viewed repeatedly and by multiple observers to assure consistent interpretations; (2) recordings can be re-analyzed from a new perspective as new questions arise; (3) observers can measure the simultaneous actions of many interacting subjects (e.g., shoaling fish) instead of a single focal animal; and (4) fleeting events (e.g., prey capture maneuvers) can be interpreted in slow motion or frame-by-frame.

The recent proliferation of inexpensive, waterproof action cameras has made high-definition, underwater video footage easier to acquire than ever before (Struthers et al. 2015), but the value of this footage as data depends on our ability to derive biologically meaningful information and measurements from video frames. Tracking algorithms that automate the

digitization of 3-D positions may require highly conspicuous subjects or even artificial reflective targets (e.g., Hendrick 2008), whereas many fish in the wild effectively blend in with their visually complex habitats. Furthermore, automated measurement may fail to record complex behaviors that require interpretation of fine-scale position or movement, such as fin posture to interpret aggression or mouth movements to interpret foraging maneuver outcomes. For these reasons, manual digitization remains important for many video analyses, and we seek to maximize the ease and efficiency of navigating events on video and recording, organizing, and retrieving measurements.

To facilitate manual digitization of video, software can minimize the steps required to record each measurement (e.g., by measuring directly on video clips, instead of requiring the export of still frames to another program) and implement mathematical methods compatible with a wide variety of measurement tasks. In contrast to this ideal, most published methods for videogrammetry in fish research were designed for specific tasks, with restrictive assumptions that suggest they were not intended for general application. For example, they may require cameras with parallel optical axes (Boisclair 1992; Petrell et al. 1997), or subjects with visible shadows against a flat surface (Laurel et al. 2005), or subjects presenting a dorsal view to the cameras (Dunbrack 2006). Hughes and Kelly (1996b) developed a mathematical method suitable for flexible applications, except that accuracy declined when fish were not positioned within a specific calibrated region relative to the cameras. None of the aforementioned methods were published with software that combines the measurement algorithms with a video player for efficient analysis. The only previous system we know to meet this criterion is a commercial software suite by SeaGIS ® (www.seagis.com.au), but its price is an impediment to some users

(Whitehead 2014), and its proprietary source code and mathematical methods are not fully transparent.

We developed an open-source Mac application called VidSync that provides a broadly applicable videogrammetry method integrated into modern video playback software in a freely available and transparent package. Its mathematical methods are compatible with a broad range of aquatic, terrestrial, or laboratory applications such as measuring through aquarium walls, filming with any number of cameras (event logging and 2-D measurement with one camera, or 3-D with two or more cameras), and use of cameras at right angles (e.g., top view and side view) in addition to the more typical side-by-side "stereo" camera configuration. Within the VidSync program, users can synchronize, calibrate, and navigate videos with detailed playback controls (e.g., frame stepping, slow motion, instant replay), record measurements into an organized hierarchy of objects (e.g., fish) and events (e.g., foraging maneuvers, length measurements), and use visual feedback (measurements overlaid on the video, and a magnified preview of the measured region) to guide precise input and visualize, retrieve, or modify existing data. The VidSync website (www.vidsync.org) contains a more comprehensive description of program features, a user manual, calibration hardware designs, and a field protocol.

VidSync has provided 2-D and 3-D measurements in a variety of fish research settings. It was developed to meet the needs of an *in situ* study of the drift-feeding behavior (Neuswanger et al. 2014), territoriality (Neuswanger 2014), and growth rates (Perry 2012) of juvenile Chinook Salmon (*Oncorhynchus tshawytscha*) in the Chena River, Alaska. Vivancos (2015) similarly measured the 3-D space-use behaviors of juvenile Roundhead Galaxiids (*Galaxias anomalus*) and Brown Trout (*Salmo trutta*) in New Zealand. Tullos and Walter (2015) investigated the response of juvenile Coho Salmon (*Oncorhynchus kisutch*) to hydraulic variability in an outdoor

experimental stream channel. Schoen (2015) used the 2-D capabilities of VidSync with four cameras positioned above quadrants of a large circular tank to measure the reaction distances of yearling Chinook Salmon to herring prey. On a larger scale, Whitehead (2014) used VidSync to quantify the avoidance of divers by Whale Sharks (*Rhincodon typus*) in the Seychelles. Although these applications indicate the system's potential, the mathematics underlying these measurements have not yet been fully described nor their performance formally tested.

In this paper, we (1) describe the novel synthesis of mathematical methods used for videogrammetry in VidSync; (2) test the system's accuracy and precision in an artificial setting; and (3) test the system's speed and utility on a fish research question requiring extensive fine-scale spatial data. Specifically, we examined the hypotheses that, within aggregations of drift-feeding juvenile Chinook Salmon, (a) each fish maintains a distance around itself wherein its nearest neighbor is less likely to be found than would be expected by chance; and (b) this region is elongated along the upstream-downstream axis, as would be expected if fish respond to shadow competition (Elliott 2002) by avoiding feeding directly downstream of competitors that deplete the drifting prey supply.

Mathematics of 3-D measurement

The VidSync software incorporates a novel combination of mathematical techniques based on the principle that one can triangulate a 3-D position from two or more known lines of sight. This section, summarized by Fig. 1, describes the steps required to correct footage for optical distortion, project clicks on the screen into 3-D lines of sight, and find the approximate intersection of those lines. VidSync users are not required to understand its background

calculations, but they are described here for transparency, to ensure repeatability, and to demonstrate the rationale for the steps involved in processing a video.

Correcting non-linear optical distortion

Optical imperfections in camera lenses and underwater housings distort the recorded image in ways that cause errors in 3-D reconstruction if not corrected. Wide-angle lenses common in underwater work exhibit radially symmetric "barrel" distortion, in which the image appears to bulge outward relative to a point near the image center called the center of distortion (Fig. 2*a*). This point may be offset from the image center by slight misalignments among the many lens and housing elements, causing asymmetric radial and tangential distortion effects known as decentering distortion.

To correct for both radial and decentering distortion, VidSync uses the Brown-Conrady model (Brown 1966) expanded to thirteen parameters: the center of distortion (u_0, v_0), seven coefficients for radial (k_1 through k_7) distortion, and four for decentering (p_1 through p_4) distortion. Let (u_d, v_d) represent the measured (distorted) pixel coordinates of an image point, as measured from the bottom left corner of the image. Define new coordinates, centered about the center of distortion, as $\bar{u} = u_d - u_0$ and $\bar{v} = v_d - v_0$. Letting $r = \sqrt{\bar{u}^2 + \bar{v}^2}$, the model calculates undistorted coordinates (u_u, v_u) as:

$$\begin{aligned} u_u &= u_0 + \bar{u}(1 + k_1 r^2 + k_2 r^4 + k_3 r^6 + k_4 r^8 + k_5 r^{10} + k_6 r^{12} + k_7 r^{14}) \\ &\quad + [p_1(r^2 + 2\bar{u}^2) + 2p_2\bar{u}\bar{v}][1 + p_3 r^2 + p_4 r^4] \\ v_u &= v_0 + \bar{v}(1 + k_1 r^2 + k_2 r^4 + k_3 r^6 + k_4 r^8 + k_5 r^{10} + k_6 r^{12} + k_7 r^{14}) \\ &\quad + [2p_1\bar{u}\bar{v} + p_2(r^2 + 2\bar{v}^2)][1 + p_3 r^2 + p_4 r^4] \end{aligned} \quad (1)$$

Distortion parameters for each camera are estimated from footage of a chessboard pattern, from which VidSync automatically extracts the distorted images of several straight lines, called plumblines. To obtain the distortion parameters that best straighten the plumblines in the corrected image, VidSync uses the downhill simplex method (Nelder and Mead 1965) to minimize a cost function defined as the sum, over all straightened plumblines, of the squared residuals from an orthogonal regression through each plumbline.

Distortion corrections are applied to each measurement in the background, without altering the image on screen. Therefore, when overlaying some results of 3-D calculations on the screen, it is necessary to re-distort their coordinates to overlay the distorted image, using the inverse of the distortion model. No closed-form inverse is known for the Brown-Conrady distortion model (Mallon and Whelan 2004), so it is instead found numerically using Newton's Method as implemented in the "gnewton" solver of the GNU Scientific Library (www.gnu.org/software/gsl/).

From 2-D screen coordinates to 3-D lines of sight

The first step of the 3-D calibration process is to establish the mapping between each screen's pixel coordinate system and the pair of known planes in a 3-D coordinate system shared among all cameras. This requires filming a "calibration frame," which consists of known points called nodes arranged in grids in two parallel planes. Different cameras may view different nodes in each plane, or even different planes perpendicular to those from other cameras (i.e., a "top view" camera may view different planes than a "side view" camera), provided the positions of all nodes on all planes of the physical frame are known in the same 3-D coordinate system. The position of the calibration frame during the calibration defines the 3-D coordinate system used throughout the video. The orientation, origin, and scaling of those coordinates can be adjusted

arbitrarily; however, this explanation adopts the convention that the front and back frame faces both lie in the x-z plane in 3-D, and the bottom left point on the front surface grid is the origin (0, 0, 0). The front and back calibration frame faces are located in the planes y=0 and y=d, where d is the separation between the faces.

To perform a calibration, the user inputs the real-world (x, z) coordinates for the dots on each face of the calibration frame and then clicks on each dot on the screen to establish corresponding screen coordinates in pixels (u_d, v_d). VidSync corrects these points for non-linear distortion to obtain undistorted screen coordinates (u_u, v_u). Having established correspondences between (x, z) and (u_u, v_u) coordinates for each node on one planar face of the calibration frame, VidSync estimates a homography (or projective transformation), represented by a 3x3 matrix H, that converts any undistorted screen coordinates (u_u, v_u) into (x, z) coordinates in that planar face (Fig. 3). The homographies operate on homogeneous coordinates, meaning screen coordinates are represented as (u_u, v_u, 1). Calibration frame plane coordinates (x, z) are recovered from the product H.(u_u, v_u, 1) by factoring out a scalar w such that the third element of that product is 1:

$$w\begin{pmatrix} x \\ z \\ 1 \end{pmatrix} = H\begin{pmatrix} u_u \\ v_u \\ 1 \end{pmatrix} \quad (2)$$

H is estimated using the normalized Direct Linear Transformation (DLT) algorithm as described by Hartley and Zisserman (2004 Algorithm 4.2). The calculation requires at least four point correspondences, preferably more, in which case the points define an over-determined linear system to which the DLT algorithm provides a least squares solution. The transformation's inverse H^{-1} is also calculated for the purpose of converting world coordinates back into screen coordinates when overlaying on-screen feedback, and for estimating reprojection errors, which are described later.

183 For each camera, homographies are calculated for front (H_f) and back (H_b) faces of the
184 calibration frame. In the usual case when the back frame face is viewed through the transparent
185 front face, an additional correction to the calculation of H_b is required to account for refraction
186 through the front face of the apparent positions of the back face points (Appendix A); VidSync
187 handles this automatically, given a user selection of the material type and thickness. To obtain a
188 3-D line of sight, the two homographies H_f and H_b convert each point in screen coordinates (u_u,
189 v_u) into two 3-D points—one on each face of the frame: (x_f, 0, z_f) and (x_b, d, z_b). These two points
190 define a line of sight from the camera through the measured object.

191 ***Calculating 3-D measurements, camera positions, and error indices***

192 VidSync calculates 3-D positions by estimating the intersections of lines of sight defined
193 by screen clicks (Fig. 4). Random errors prevent these lines from intersecting exactly, so we can
194 only estimate their closest point of approach (CPA). To this end, VidSync uses a geometrically
195 intuitive linear method, the results of which are refined by a more accurate iterative method in
196 cases where no refractive interface (such as an aquarium wall) separates the cameras or their
197 housings from the subjects of measurement.

198 The linear method's position estimate is the CPA of the lines of sight. For two lines from
199 two cameras, the CPA is the midpoint of the shortest possible line segment that connects the two
200 lines. For any number of lines, let q_i represent the first point on line i, let $I_{3\times3}$ represent the 3-by-
201 3 identity matrix, and let r_i be the unit vector along line i. Superscript T denotes the transpose.
202 The CPA (x, y, z) of any number of lines is

(3) $$CPA = \left(\sum_i (I_{3\times3} - r_i r_i^T)\right)^{-1} \left(\sum_i (I_{3\times3} - r_i r_i^T) q_i\right)$$

From the CPA, a useful index of error is calculated, the mean distance from the CPA to the lines from which it was calculated, which we term the point-line distance or PLD error:

(4) $$PLD\ error = \sum_i \|(CPA - q_i) \times (CPA - q_i - r_i)\|_2$$

An iterative method is used to refine measurements because linear triangulation methods such as the CPA are not optimal estimates of 3-D intersections (Hartley and Zisserman 2004). Instead, assuming normally distributed errors, the maximum likelihood estimate of a 3-D position is obtained by (1) constraining the lines of sight to perfectly intersect; (2) reprojecting candidate 3-D points back onto the screen; and (3) iteratively minimizing the distance between the input screen points and the reprojected screen points, which is termed the reprojection error.

In our two-plane geometric method, candidate 3-D points are reprojected into screen coordinates by finding where a line between the candidate point and the camera itself intersects the front calibration frame face plane ($y = 0$) and converting those front frame face coordinates back into undistorted screen coordinates using the inverse homography H_f^{-1}. The position of the camera is calculated as the CPA of several lines-of-sight from the camera, which are found by projecting the input screen positions of the back frame nodes during calibration onto both faces of the frame using H_f and H_b.

Let s_i be the undistorted screen coordinates of an input point in camera i, and let s_i' be the reprojected screen coordinates of the 3-D point (x, y, z) in that camera. Using the result from the linear triangulation method as a starting point to speed convergence, VidSync uses the iterative downhill simplex method (Nelder and Mead 1965) to estimate the 3-D point that minimizes the sum of squared reprojection errors across all n_c cameras. The reprojection error

223 reported for 3-D measurements by VidSync is the root mean square of the reprojection errors in
224 each camera view:

(5) $$RMS\ reprojection\ error = \sqrt{\frac{1}{n_c}\sum_i \left(\left\|s_i - s_i^{'}\ (x,y,z)\right\|_2\right)^2}$$

225 Methods used to test the system

226 *Cameras and calibration hardware tested*

227 We used a pair of Sony ® HDR-SR12 digital video cameras inside Ikelite ® #6038.94
228 underwater housings with Zen Underwater ® WAVP-80 wide-angle dome ports. The housings
229 were bolted 33 cm apart on a 55-cm length of 2-inch (5.08-cm) aluminum angle beam. These
230 cameras recorded video in 1080i resolution on internal hard drives in AVCHD format, which
231 was transcoded to Apple Intermediate Codec upon downloading to a computer, deinterlaced by
232 interpolation to 1080p, and compressed into the final .mov files in the H.264 codec with a 4
233 MB/s bitrate (about 30 GB per camera for 2 hours of footage) using Apple Compressor 3.
234 The calibration frame (visible in Fig. 3) was a clear box made of 3/8-inch (0.9525-cm)
235 Lexan ® sheeting bonded together by IPS Weld-On ® #3 polycarbonate adhesive. The front and
236 back surfaces each held a 0.4-m by 0.3-m grid of 0.95-cm diameter dots (drilled and filled with
237 black silicone sealant) spaced at 0.1-m intervals. The checkerboard used for distortion correction
238 was a ½-inch (1.27-cm) black-and-white checkered pattern large enough to completely fill the
239 screen in each camera when placed 10 cm in front of its dome port. This test of these calibration
240 devices led to several suggested improvements in a new design described at
241 www.vidsync.org/Hardware.

Pool test of precision and accuracy

We tested our hardware system and VidSync with 1,010 measurements of objects of known length in the University of Alaska Fairbanks swimming pool (Table 1). We examined the effects of various factors on precision and accuracy at the intended working distance of our hardware (0.2 to 1.0 m) and at greater distances. To observe how the distance between the cameras and calibration frame affects accuracy, we calculated all measurements using two separate calibrations, one centered at a distance of 0.6 m from the cameras ("Calibration A") and the other at 0.9 m ("Calibration B"). Calibration A was better centered within the intended working distance of our hardware system, so we used it for all analyses shown here, except that a row of results from Calibration B is included in Table 1 to show how accuracy at longer distances can be improved by calibrating at longer distances.

We used sections of the distortion correction chessboard in four different lengths as measurement targets to be held in front of our stationary camera system. The grid's precise design and sharp corners provided unambiguous endpoints and dimensions. Measurements based on Calibration A were grouped by their estimated distance d from the midpoint between the cameras, resulting in four measurement distance categories: (1) closer to the cameras than the front face of the frame, 0.142 m $\leq d <$ 0.389 m; (2) within the "calibrated range" between the front and back of the frame, 0.389 m $\leq d <$ 0.828 m; (3) close behind the frame, 0.828 m $\leq d <$ 2.000 m; and (4) far behind the frame, 2.000 m $\leq d \leq$ 7.058 m.

Field test: investigating Chinook Salmon space-use strategies

We tested our hypotheses about competitor spacing and shadow competition in juvenile Chinook Salmon using video footage from the 5th-order Chena River (median summer flow 20

264 m^3/s) in the Yukon River drainage in interior Alaska. We filmed five groups of feeding juveniles
265 at different sites representing a range of physical conditions and fish sizes. Site details are
266 described in Neuswanger et al. (2014), in which we used VidSync to manually track the foraging
267 activity of every visible individual throughout a 5- to 20-minute period from each video. At each
268 site, our stereo camera pair was placed at a stationary position such that fish were feeding within
269 1 m of both cameras most of the time. Fish resumed normal feeding behavior within 10 minutes
270 of camera placement, and battery life allowed cameras to record approximately 90 minutes of
271 undisturbed behavior.

272 For the present analysis, we began with the same VidSync Document files used by
273 Neuswanger et al. (2014). These files already contained digitized calibrations, fork length
274 measurements from six to 38 separate individuals, and measurements of several points on the
275 water's surface and water velocity tracers. Using the velocity tracers to indicate the downstream
276 direction and the water's surface to indicate the vertical direction, we calculated a transformation
277 to rotate the x, y, and z axes of the measurements into alignment with the true downstream, cross-
278 stream, and vertical directions, respectively. Appendix B describes the procedure for calculating
279 these "stream coordinates."

280 We measured how much time one observer took to digitize new fish position
281 measurements at 2.5-minute intervals throughout the full 85 to 97.5-minute duration of
282 undisturbed behavior in each video. At each time point (frame), we measured the tip of the snout
283 of every fish that was visible in both camera views before proceeding to the next time point
284 using a customizable frame stepping button set to 2.5 minutes. Upon reaching the end of the
285 video, we exported all measurements as XML files.

We used Wolfram Mathematica® 10 to import VidSync's XML output, convert measurements into stream-based coordinates, and calculate the position, relative to each fish, of its nearest neighbor in the same frame, measured from snout to snout. Nearest-neighbor positions and distances (NNDs) were combined across all frames from each video.

We compared observed fish positions against a null hypothesis that fish would be distributed randomly throughout the visible volume occupied by the group. To simulate this random distribution, we pooled the measured fish positions from all frames into a single data set for each video, then fit this cloud of points with a 3-D smooth kernel distribution using a rectangular kernel with a bandwidth of 3 cm. Graphical examination showed that drawing random variates from this distribution produced a point cloud very similar in outer extent and large-scale structure to the actual fish data, but with the fine-scale structure randomized. We then created 100 random frames per observation frame, with the number of fish per random frame selected by random sampling with replacement from the counts of fish in observation frames, and the positions of fish within each random frame drawn from the smooth kernel distribution. We calculated nearest-neighbor distances in these random frames by the same method we used for observation frames. To plot the probability density of nearest-neighbor distances for comparison between actual fish positions and the null hypothesis of random fine-scale positions, we estimated smooth kernel distributions from each set of nearest neighbor distances using Gaussian kernels with automatic bandwidth selection using the Silverman method. These can be interpreted as smoothed histograms.

Results

Distortion correction

Distortion corrections applied to the pool test video reduced the root mean square distortion, which represents the typical distance between a point in a straightened plumbline and an orthogonal regression through that plumbline, by 83.3 % to 0.26 pixels per point for the left camera, and by 88.2 % to 0.16 pixels per point for the right camera. Because the uncorrected variation includes random errors in chessboard corner locations, these results indicate a near-complete elimination of systematic distortion, which is visually evident by comparing the barrel distortion in Fig. 2*b* to the corrected grid in Fig. 2*d*. Parameter estimates and the calculated point corrections were similar across several images of the chessboard at different distances, provided the board was close enough to fill the screen.

To diagnose any uncorrected effects of radial distortion on length measurements, we constructed plots of absolute error against the maximum distance of each measurement's endpoints from the center of distortion in either camera. The absence of a clear increase in absolute error for measurements near the edge or center of the screen suggests that the current model adequately mitigates distortion (Fig. 5*a*).

Accuracy and precision of length measurements

Our hardware was configured to measure small objects close to the cameras, via our choice of camera separation, calibration frame dimensions, and the position of Calibration A. In our most direct test of this application, 618 length measurements of 50.8-mm targets within 2 m of the cameras had mean absolute errors < 0.5 mm—less than 1 % of target length. For all target

lengths, accuracy (absolute errors) and precision (variance) decreased as distance from the cameras increased (Fig. 6). At all distances, measurements of longer objects were less accurate and precise than measurements of shorter objects, but most remained within 1 % of true target lengths (Table 1). When we recalculated all measurements using Calibration B, accuracy was improved at long distances but reduced slightly in the region closest to the cameras (see the mean abs % error in Table 1), indicating an advantage to calibrating at a distance close to the intended working distance. We found no negative effect of measuring lengths at oblique angles of up to 50 degrees from the cameras (Fig. 5*b*).

Length measurements of actual fish were less precise than measurements of our chessboard, because they included more sources of uncertainty, including variation in the straight-line distance between a fish's head and tail fork as its body flexes during swimming. In a test of 10 repeated measurements of three juvenile Chinook Salmon 0.5 m from the cameras, we measured fork lengths (mean ± sd) of 54.5 ± 1.6 mm (2.9 %), 57.3 ± 1.5 mm (2.6 %), and 54.8 ± 0.8 mm (1.5 %). These contrast with a standard deviation of only 0.23 mm (0.45 %) for an artificial target of similar length, measured at similar distances, in our pool test (Table 1).

Diagnostic "error" measures

We used Spearman rank correlation tests to compare the actual error in 3-D measurements against the two "error" measures provided for each 3-D point by VidSync, the RMS reprojection error and the point-line distance (PLD) error. Real absolute errors in the 618 length measurements of 50.8-mm targets in our pool test were significantly ($p < 0.001$) but very weakly correlated with both the PLD error (Spearman's $\rho = 0.24$) and the RMS reprojection error ($\rho = 0.13$). The two error measures were significantly ($p < 0.00001$) but weakly ($\rho = 0.18$) correlated with each other. These weak correlations reflect the purpose of these measures

as tools to diagnose data entry mistakes or calibration problems, not to quantify actual errors in 3-D measurements. In this regard, they were helpful; examining points with the highest RMS reprojection errors revealed several points for which the target (a chessboard corner) had been poorly located or was moving slightly during measurement.

Chinook Salmon field test

Efficiency of data processing in VidSync

In our previous analysis of the same videos (Neuswanger et al. 2014), it took less than an hour per video to digitize calibrations in VidSync and record the measurements needed to convert results into stream-based coordinates. For the present analysis, we recorded 2,696 new 3-D positions of fish heads from a total of 186 video frames, with one observer manually digitizing 394 measurements per hour of time spent using VidSync. Measurement rate varies based on task complexity. For example, repeated measurements tracking a single fish's position in consecutive frames can be manually digitized quickly, whereas foraging attempt outcomes are recorded more slowly because of the time required to locate and interpret relevant observations.

Space-use patterns of juvenile Chinook Salmon

In all five videos, juvenile Chinook Salmon maintained greater distances from their nearest neighbors than would be expected under the null hypothesis of random distribution within the visible volume occupied by their group (Fig. 7). The radius of the sphere within which neighbors were less common than would be expected under the null hypothesis ranged from 4.6 to 14.7 cm, or 1.0 to 2.9 times the mean fork length of fish at their respective sites.

The nearest neighbor of a fish was located to its side more often than directly upstream or downstream. This pattern is visible as an elongation of the inner contours of the probability

distributions in Fig. 8 along the upstream-downstream axis, and by the lobes of higher probability density (darker shading) in lateral positions.

Discussion

VidSync provided 3-D measurements with high precision and accuracy—generally within 1% of the true length of measured objects (Table 1). We demonstrated its capacity to quickly process large quantities of data by recording 2,696 position measurements for a juvenile Chinook Salmon space-use analysis at a rate of 394 measurements per hour. This analysis produced biological insights that (1) fish avoided the immediate vicinity of their neighbors out to a radius ranging between 4.6 and 14.7 cm, or 1.0 to 2.9 body lengths; and (2) this avoided region was elongated along the upstream-downstream axis, consistent with behavioral responses to the depletion of drifting prey by upstream competitors in shadow competition (Elliott 2002). However, other factors such as visual distraction could also deter drift-feeding fish from feeding directly downstream of their neighbors.

Measurement error

Absolute errors in length measurement increased as the distance from the cameras increased, and as the length of the target increased (Table 1). The increase with distance is intuitive, but it is less obvious why error increases with target length. Harvey et al. (2010) noted that, “It has not been unequivocally demonstrated whether error is absolute (i.e., constant irrespective of the length of the object) or relative to the length of the object being measured.” We suggest that different sources of error scale in different ways, only some of which depend on the length of the measured object. Some errors result from random factors specific to each point

measurement, especially when the target is visually ambiguous (e.g., the fork of a translucent fish tail). Similar uncertainty can arise from motion blur, camouflage, a high-contrast background, turbidity, poor lighting, image noise, limited image resolution, video interlacing, or occlusion by closer objects. However, these random errors should not logically scale with the length of the object.

Each system is also subject to systematic errors. Inevitable imperfections in the calibration frame, arising from both its physical construction and its digitization in VidSync, result in a reconstructed 3-D space that is slightly warped compared to the real space it is meant to represent. Uncorrected components of non-linear distortion may have a similar effect. Other systematic errors are more situational; for example, misalignment of the cameras in between calibration and measurement can warp the reconstructed space. Another potential systematic error arises if the cameras or target objects are moving. When video clips are synchronized to the nearest frame, they are still out of sync by up to one-half the duration of a frame, averaging one-quarter frame. In video shot at 30 frames per second, the average position error in one camera is equivalent to the distance the object moved in one-quarter frame, or 1/120 s. This motion-dependent error is termed motion parallax (Harvey and Shortis 1996) or synchronization error (Hughes and Kelly 1996b). These systematic errors, particularly those related to the calibration, explain why absolute length errors increase with target length. Consider measuring a 100-mm fish and a 200-mm fish at the same location in an imperfectly reconstructed 3-D space, which is slightly stretched compared to real space such that the 100-mm fish is measured as 101 mm. The front and back halves of the 200-mm fish would each measure as 101 mm, giving a total length of 202 mm – twice the absolute error as for the shorter fish, but a similar percentage error. Although the errors in our test system were small, they were clearly target-length-dependent

(Table 1), suggesting that they were caused more by systematic than random errors. This understanding emphasizes the importance of constructing the calibration frame with precision and digitizing it carefully.

Both random and systematic errors increase with distance. Random errors in screen coordinates cause uncertainty in the angle of the 3-D line of sight, which corresponds to a small spatial uncertainty close to the cameras, and a much larger one far away. Also, the lines of sight from multiple cameras converge at a narrower angle for more distant targets, so small angular uncertainty in each line of sight leads to a larger uncertainty in their intersection than it does for nearby targets. Finally, systematic errors associated with imperfections in the calibration frame should also scale with distance outside the frame, because small imperfections will be extrapolated outward into larger ones.

Our tests suggested potential ways to improve the precision and accuracy beyond the values reported here. Foremost, our calibration frame (Fig. 3) could have been improved by using markers with easy-to-locate exact centers, and by using a wider and taller grid of nodes, with less spacing between the front and back faces, to maximize the screen coverage of the frame during calibration and reduce the need for extrapolation outside the calibrated grid. These suggestions influenced the new, recommended frame design (www.vidsync.org/Hardware). Accuracy also could have been improved by placing the cameras farther apart to widen the angle at which their lines of sight converge; however, our system was constrained by the biological, project-specific need to film in tight spaces in logjams with fish very close to the cameras.

Comparison to other videogrammetry methods

Here we compare VidSync with its mathematical predecessor, the method of Hughes and Kelly (1996b), and with the commercial SeaGIS ® (www.seagis.com.au) videogrammetry

439 software for Windows. We do not draw comparisons with the many other videogrammetric
440 methods designed for specific, narrower applications, nor with methods focused on the distinctly
441 different challenges of automatic tracking.

442 Hughes and Kelly (1996b) presented mathematical methods that have proven broadly
443 useful in studies of fish behavior (e.g., Hughes et al. 2003; Piccolo et al. 2007; Uglem et al.
444 2009). They introduced the concept of projecting screen coordinates onto two planes in world
445 space and intersecting the lines of sight defined by points in the front and back planes. Compared
446 with more common methods that implicitly assume light travels in a straight line from the
447 subject to the camera housings, the two-plane method is especially versatile for fish research
448 because it is compatible with filming through air-water interfaces such as the side of a tank, and
449 it is easily applied to systems of more than two cameras to cover larger viewing areas. VidSync
450 retains this two-plane concept but uses different mathematical techniques for other tasks, most
451 importantly using Direct Linear Transformation to map screen points onto the calibration planes.
452 This advance enables accurate measurement anywhere within the joint field-of-view of two or
453 more cameras, whereas the previous method, based on polynomial interpolation, had
454 substantially reduced accuracy when extrapolating measurements outside the region of the screen
455 occupied by the calibration frame during the calibration. Hughes and Kelly (1996b) reported
456 mean errors in locating 3-D points of 4.7 mm with a standard deviation of 2.7 mm, larger than
457 the errors reported here for most measurement tasks (Table 1), although their test methods were
458 not described in enough detail for direct comparison.

459 We know of only one other general-purpose system for videogrammetry with standalone
460 software comparable to VidSync—the commercial SeaGIS® software suite that includes their
461 CAL™ calibration program and EventMeasure Stereo™ measurement programs, which are

462 mathematically based on a bundle adjustment method (Granshaw 1980). In a recent test (Harvey
463 et al. 2010), this system's accuracy and precision were very close to those of VidSync. The mean
464 absolute error was 0.5 mm for measurements of a 50.5-mm-long target within 1 to 3 m from the
465 cameras, close to our mean absolute error of 0.37 mm for a 50.8-mm-long target within 0.828 to
466 2 m from the cameras. Although their other tests were not directly comparable to ours, they
467 summarized their results as accurate to approximately 1% of the true length of the measured
468 object, similar to our results.

469 Given the similarly high precision and accuracy of VidSync and the SeaGIS ® methods,
470 the most practically important differences between the systems are in their transparency, costs,
471 capabilities, and user interface features. VidSync is freely available, but it requires a Mac
472 computer and a calibration frame that can be built in-house, or with help from a sign printer, for
473 less than $300 USD. The SeaGIS ® products, according to their May 2015 price sheet
474 (www.seagis.com.au/SeaGIS%20Prices%202015_05.pdf), cost $8,895 AUD ($6,351 USD) for
475 the combination of products comparable to a VidSync system: academic/research licenses for
476 CAL™ and EventMeasure™, and their least expensive calibration hardware (website accessed
477 and currency converted on November 18, 2015). Both systems function well with side-by-side
478 stereo camera systems placed in the water with their subjects, but VidSync is additionally
479 compatible with filming through the sides of an aquarium, and with laboratory setups involving
480 more than two cameras (for example, a jointly calibrated row of four cameras, in which only any
481 two adjacent cameras have overlapping fields-of-view). Differences in software features are
482 extensive and may be explored by the reader on the websites of each system; however, two
483 primary differences are that VidSync offers more options for fine playback control (e.g., slow

motion, instant replay), and it has a more flexible system for the organization and retrieval of measurements.

Applications

Various 3-D videogrammetric methods have been used in ecological research for remote length measurement (Petrell et al. 1997; Shieh and Petrell 1998), biomass estimation (Lines et al. 2001), habitat mapping (Shortis et al. 2007b), abundance surveys (Williams et al. 2010), mapping foraging behaviors (Hughes et al. 2003; Piccolo et al. 2007; Piccolo et al. 2008); and for studying the kinematics of swimming maneuvers (Hughes and Kelly 1996b; Butail and Paley 2012), octopus grasping (Yekutieli et al. 2007), and insect flight (Hedrick 2008; Ardekani et al. 2013). VidSync is compatible with any such application, provided that the water is not too dark or turbid to observe targets clearly on video, and that the number of desired measurements does not require automated object tracking.

Many of the above-described, past applications of videogrammetry directly involved the developers of the measurement methods used. This suggests that biologists who did not have the substantial time and technical expertise required to cost-effectively develop videogrammetry systems themselves, or at least close access to such a developer, may have avoided pursuing research topics that required large quantities of precise, 3-D spatial data. We believe the methods presented here will make such studies more tractable by providing freely available, user-friendly videogrammetry software with high precision, accuracy, and versatility.

Acknowledgments

This work was supported by the Arctic-Yukon-Kuskokwim Sustainable Salmon Initiative, the Institute of Arctic Biology, Alaska EPSCoR NSF award #OIA-1208927 and the state of Alaska, and the Department of Biology and College of Natural Sciences and Mathematics at the University of Alaska Fairbanks. Lon Kelly contributed to the early development of the mathematical methods underlying VidSync. Megan Perry, Aurélian Vivancos, Darren Whitehead, and Erik Schoen tested early prototypes of VidSync and provided valuable feedback. David Neuswanger, Milo Adkison, and two anonymous reviewers helpfully critiqued the manuscript. Fish work was conducted under IACUC protocols #134754-1 and #175627-1. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

References

Ardekani, R., Biyani, A., Dalton, J., Saltz, J., Arbeitman, M., Tower, J., Nuzhdin, S., and Tavare, S. 2013. Three-dimensional tracking and behaviour monitoring of multiple fruit flies. J. Roy. Soc., Interface 10(78): 20120547.

Boisclair, D. 1992. An evaluation of the stereocinematographic method to estimate fish swimming speed. Can. J. Fish. Aquat. Sci. 49(3): 523-531.

Brown, D. 1966. Decentering distortion of lenses. Photogramm. Eng. 32(3): 444-462.

Butail, S., and Paley, D. 2012. Three-dimensional reconstruction of the fast-start swimming kinematics of densely schooling fish. J. Roy. Soc., Interface 9(66): 77-88.

Dunbrack, R. 2006. In situ measurement of fish body length using perspective-based remote stereo-video. Fish Res 82(1-3): 327-331.

Ellender, B., Becker, A., Weyl, O., and Swartz, E. 2012. Underwater video analysis as a non-destructive alternative to electrofishing for sampling imperilled headwater stream fishes. Aquat. Conserv. 22(1): 58-65.

Elliott, J.M. 2002. Shadow competition in wild juvenile sea-trout. J. Fish Bio. 61: 1268–1281.

536 Fischer, P., Weber, A., Heine, G., and Weber, H. 2007. Habitat structure and fish: assessing the
537 role of habitat complexity for fish using a small, semiportable, 3-D underwater observatory.
538 Limnol. Oceanogr. Methods 5: 250-262.
539
540 Granshaw, S. 1980. Bundle adjustment methods in engineering photogrammetry. Photogramm.
541 Rec. 10(56): 181-207.
542
543 Hartley, R., and Zisserman, A. 2004. Multiple view geometry in computer vision. Cambridge
544 University Press, Cambridge.
545
546 Harvey, E., Fletcher, D., and Shortis, M. 2001. A comparison of the precision and accuracy of
547 estimates of reef-fish lengths determined visually by divers with estimates produced by a stereo-
548 video system. Fish. Bull. 99(1): 63-71.
549
550 Harvey, E., and Shortis, M. 1996. A system for stereo-video measurement of sub-tidal
551 organisms. Mar. Technol. Soc. J. 29(4): 10-22.
552
553 Harvey, E., Goetze, J., Mclaren, B., Langlois, T., and Shortis, M. 2010. Influence of range, angle
554 of view, image resolution and image compression on underwater stereo-video measurements:
555 high-definition and broadcast-resolution video cameras compared. Mar. Technol. Soc. J. 44(1):
556 75-85.
557

558 Hedrick, T. 2008. Software techniques for two- and three-dimensional kinematic measurements
559 of biological and biomimetic systems. Bioinspiration Biomimetics 3(3): 034001.
560
561 Hughes, N. F., Hayes, J. W., Shearer, K. A., and Young, R. G. 2003. Testing a model of drift-
562 feeding using three-dimensional videography of wild Brown Trout, *Salmo trutta*, in a New
563 Zealand river. Can. J. Fish. Aquat. Sci. 60(12): 1462-1476.
564
565 Hughes, N. F., and Kelly, L. H. 1996a. A hydrodynamic model for estimating the energetic cost
566 of swimming maneuvers from a description of their geometry and dynamics. Can. J. Fish. Aquat.
567 Sci. 53(11): 2484-2493.
568
569 Hughes, N. F., and Kelly, L. H. 1996b. New techniques for 3-D video tracking of fish swimming
570 movements in still or flowing water. Can. J. Fish. Aquat. Sci. 53(11): 2473-2483.
571
572 Laurel, B., and Brown, J. 2006. Influence of cruising and ambush predators on 3-dimensional
573 habitat use in age 0 juvenile Atlantic cod *Gadus morhua*. J. Exp. Mar. Biol. Ecol. 329(1): 34-46.
574
575 Laurel, B., Laurel, C., Brown, J., and Gregory, R. 2005. A new technique to gather 3-D spatial
576 information using a single camera. J. Fish. Biol. 66(2): 429-441.
577
578 Lines, J., Tillett, R., Ross, L., Chan, D., Hockaday, S., and Mcfarlane, N. 2001. An automatic
579 image-based system for estimating the mass of free-swimming fish. Comput. Electron. Agr.
580 31(2): 151-168.

581

582 Mallon, J., and Whelan, P. F. 2004. Precise radial un-distortion of images. Pattern Recognition,
583 2004. ICPR 2004. Proceedings of the 17th International Conference on Pattern Recognition 1:
584 18-21.

585

586 Mussi, M., Mcfarland, W., and Domenici, P. 2005. Visual cues eliciting the feeding reaction of a
587 planktivorous fish swimming in a current. J. Exp. Biol. 208(Pt 5): 831-842.

588

589 Nelder, J., and Mead, R. 1965. A simplex method for function minimization. Comput. J. 7: 308-
590 313.

591

592 Neuswanger, J.R. 2014. New 3-D video methods reveal novel territorial drift-feeding behaviors
593 that help explain environmental correlates of Chena River Chinook Salmon productivity. Ph.D.
594 dissertation, University of Alaska Fairbanks, Fairbanks, AK.

595

596 Neuswanger, J.R., Wipfli, M.S., Rosenberger, A.E., Hughes, N.F. 2014. Mechanisms of drift-
597 feeding behavior in juvenile Chinook Salmon and the role of inedible debris in a clear-water
598 Alaskan stream. Environ. Biol. Fish. 97(5): 489–503.

599

600 Perry, M.T. 2012. Growth of juvenile Chinook Salmon (*Oncorhynchus tshawytscha*) in an
601 interior Alaska river: responses to supplemental feeding and temperature. M.S. thesis, University
602 of Alaska Fairbanks, Fairbanks, AK.

603

Petrell, R., Shi, X., Ward, R., Naiberg, A., and Savage, C. 1997. Determining fish size and swimming speed in cages and tanks using simple video techniques. Aquacult Eng 16(1-2): 63-84.

Piccolo, J. J., Hughes, N. F., and Bryant, M. D. 2007. The effects of water depth on prey detection and capture by juvenile Coho Salmon and steelhead. Ecol Freshw Fish 16(3): 432-441.

Piccolo, J. J., Hughes, N. F., and Bryant, M. D. 2008. Water velocity influences prey detection and capture by drift-feeding juvenile Coho Salmon (*Oncorhychus kisutch*) and steelhead (*Oncorhynchus mykiss irideus*). Can. J. Fish. Aquat. Sci. 65(2): 266-275.

Schoen, E.R. 2015. Linking individual-level foraging interactions of piscivores to food-web dynamics in pelagic systems. Ph.D. dissertation, University of Washingtion, Seattle, WA.

Shieh, A., and Petrell, R. 1998. Measurement of fish size in Atlantic salmon (*Salmo salar* L.) cages using stereographic video techniques. Aquacult. Eng. 17(1): 29-43.

Shortis, M., Harvey, E., and Abdo, D. 2009. A review of underwater stereo-image measurement for marine biology and ecology applications. Oceanogr. Mar. Biol. 47: 257-292.

Shortis, M., Harvey, E., and Seager, J. 2007a. A review of the status and trends in underwater videometric measurement. SPIE Conference 6491 Videometrics IX: 26.

627

628 Shortis, M., Seager, J., Williams, A., Barker, B., and Sherlock, M. 2007b. A towed body stereo-
629 video system for deep water benthic habitat surveys. Eighth Conf. Optical 150-157.

630

631 Struthers, D.P., Danylchuk, A.J., Wilson, A.D.M., Cooke, S.J. 2015. Action cameras: bringing
632 aquatic and fisheries research into view. Fisheries 40(10): 502–512.

633

634 Tullos, D., and Walter, C. 2015. Fish use of turbulence around wood in winter: physical
635 experiments on hydraulic variability and habitat selection by juvenile Coho Salmon,
636 *Oncorhynchus kisutch*. Environ. Biol. Fish 98(5): 1339–1353.

637

638 Uglem, I., Kjorsvik, E., Gruven, K., and Lamberg, A. 2009. Behavioural variation in cultivated
639 juvenile Atlantic cod (*Gadus morhua* L.) in relation to stocking density and size disparity. Appl.
640 Anim. Behav. Sci. 117(3-4): 201-209.

641

642 Vivancos, A., and Closs, G.P. 2015. Quantification and comparison of individual space-use
643 strategies in foraging drift-feeding fish using fine-scale, multidimensional movement analysis.
644 Can. J. Fish. Aquat. Sci. 72(11): 1760–1768.

645

646 Whitehead, D.A. 2014. Establishing a quantifiable model of Whale Shark avoidance behaviours
647 to anthropogenic impacts in tourism encounters to inform management actions. M.S. thesis,
648 University of Hertfordshire, United Kingdom.

649

650 Williams, K., Rooper, C., and Towler, R. 2010. Use of stereo camera systems for assessment of
651 rockfish abundance in untrawlable areas and for recording pollock behavior during midwater
652 trawls. Fish. Bull. 108(3): 352-362.
653
654 Yekutieli, Y., Mitelman, R., Hochner, B., and Flash, T. 2007. Analyzing octopus movements
655 using three-dimensional reconstruction. J. Neurophysiol. 98(3): 1775-1790.
656
657

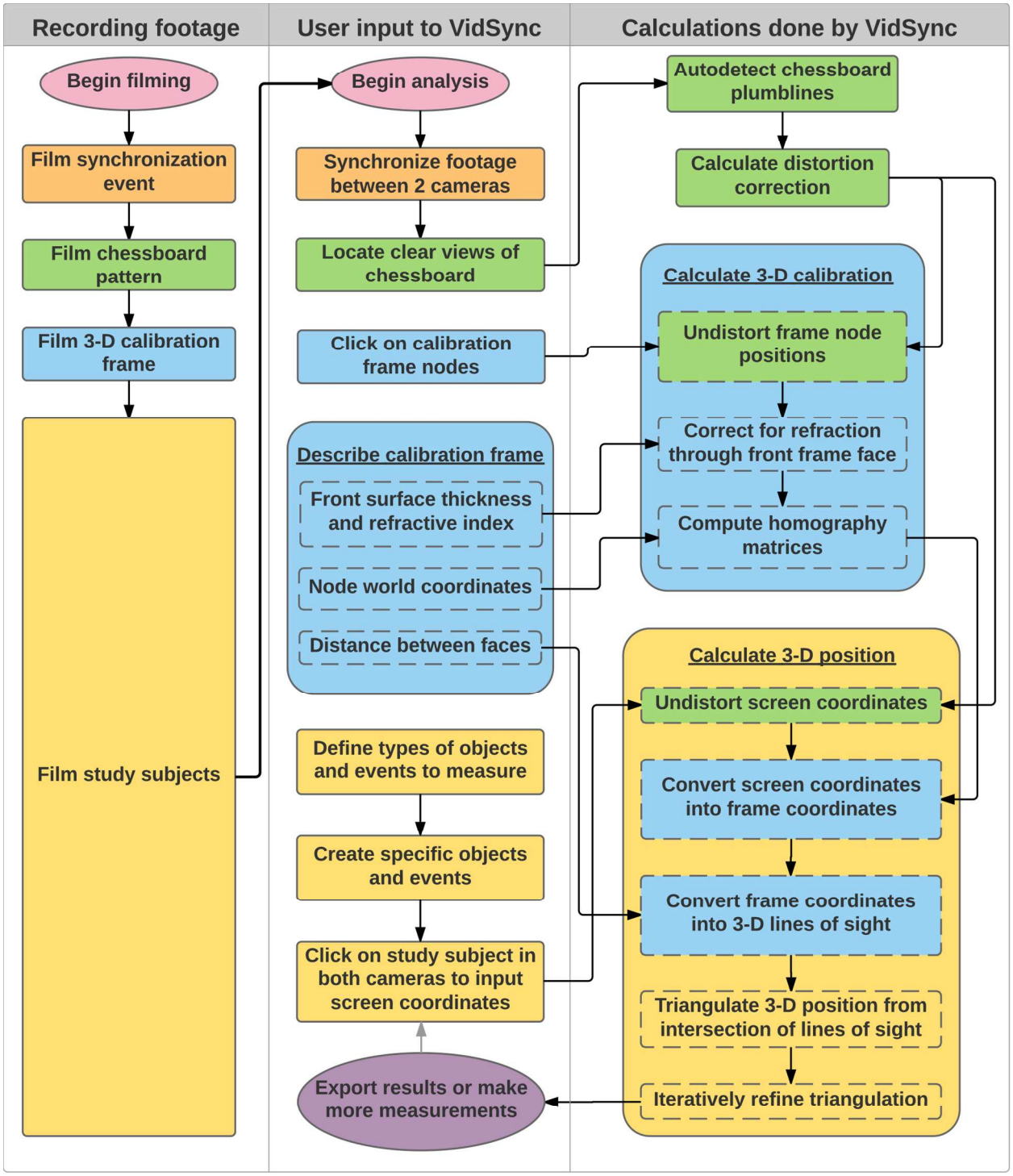
Tables

Table 1. Summary of 1,010 measurements of objects of 4 known lengths at range of distances.

True length (mm)	Distance *d* from cameras (m)	n	Calibration A				Calibration B
			Mean error (mm)	Mean abs error (mm)	Std dev (mm)	Mean abs % error	Mean abs % error
50.8	$0.142 \leq d < 0.389$	119	-0.02	0.16	0.19	0.3 %	0.4 %
	$0.389 \leq d < 0.828$	371	0.13	0.22	0.23	0.4 %	0.5 %
	$0.828 \leq d < 2.000$	128	0.36	0.42	0.42	0.8 %	0.7 %
	$2.000 \leq d < 7.058$	70	0.99	1.05	1.07	2.1 %	1.4 %
152.4	$0.142 \leq d < 0.389$	30	-0.07	0.45	0.50	0.3 %	0.4 %
	$0.389 \leq d < 0.828$	31	0.27	0.59	0.62	0.4 %	0.4 %
	$0.828 \leq d < 2.000$	31	0.95	0.97	0.65	0.6 %	0.6 %
	$2.000 \leq d < 7.058$	30	1.67	1.80	1.47	1.2 %	0.6 %
381	$0.389 \leq d < 0.828$	33	1.65	1.65	0.39	0.4 %	0.6 %
	$0.828 \leq d < 2.000$	44	1.96	2.48	1.89	0.7 %	0.7 %
	$2.000 \leq d < 7.058$	30	6.41	6.41	2.73	1.7 %	0.6 %
596.9	$0.389 \leq d < 0.828$	31	0.94	0.97	0.72	0.2 %	0.1 %
	$0.828 \leq d < 2.000$	31	3.27	3.27	2.43	0.6 %	0.3 %
	$2.000 \leq d < 7.058$	31	6.73	6.80	4.13	1.1 %	0.4 %

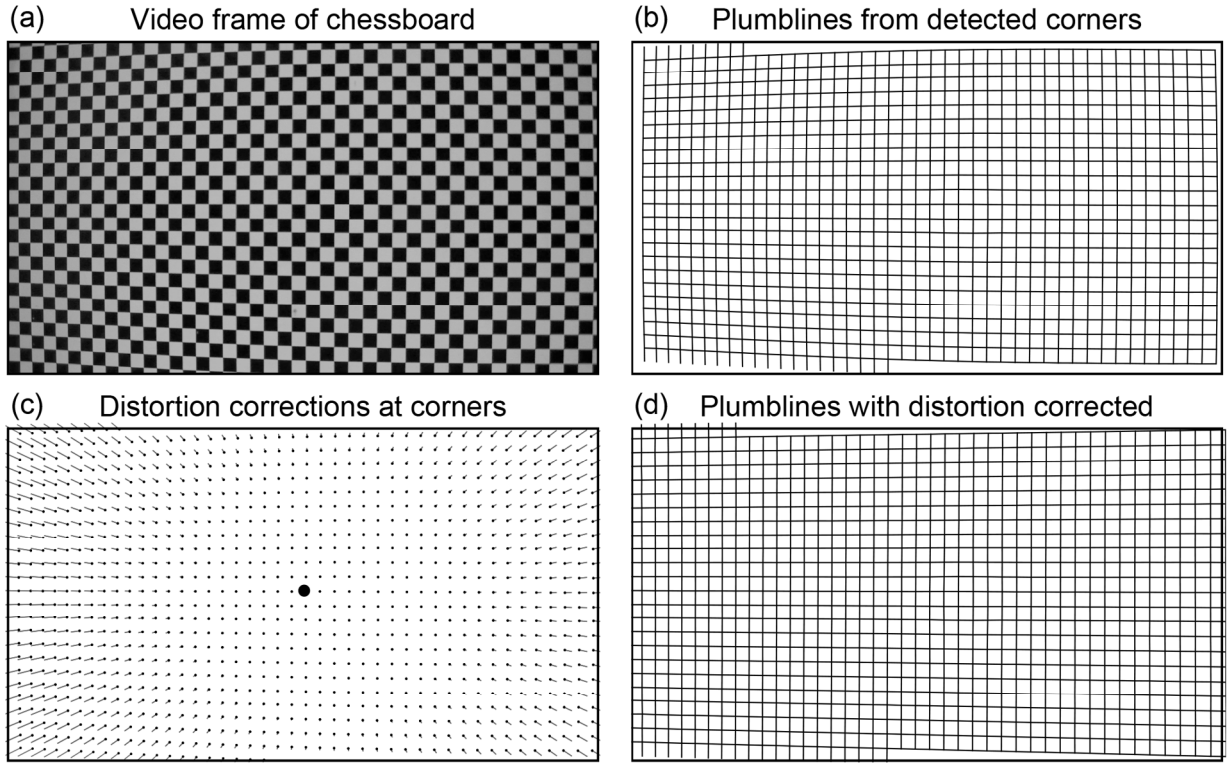
Note: Metrics of accuracy and precision included the mean error (measured length - true length), mean absolute error (absolute value of error), standard deviation of the measured length, and mean absolute error as a percentage of the true length. All metrics are shown for Calibration A. To show the effect of calibration distance on errors, the exact same measurements were recalculated for comparison using Calibration B, which was obtained with the calibration frame 0.3 m farther from the cameras than in Calibration A. No measurements are shown for the largest 2 targets at the smallest distance range because they did not fit within the field of view at that distance.

668 # Figures



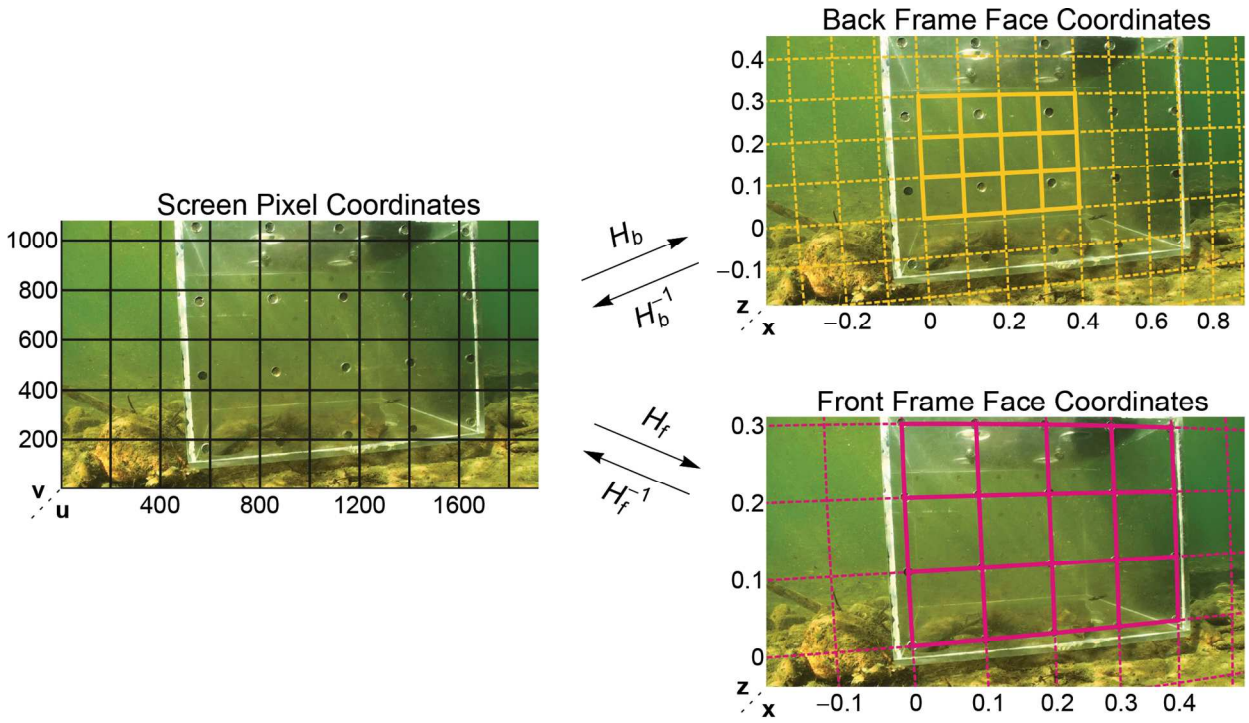
669

670 **Fig. 1.** Flowchart of the process of recording 3-D measurements from a stereo camera system in
671 VidSync. Colors indicate groups of related tasks such as calibration and distortion correction.



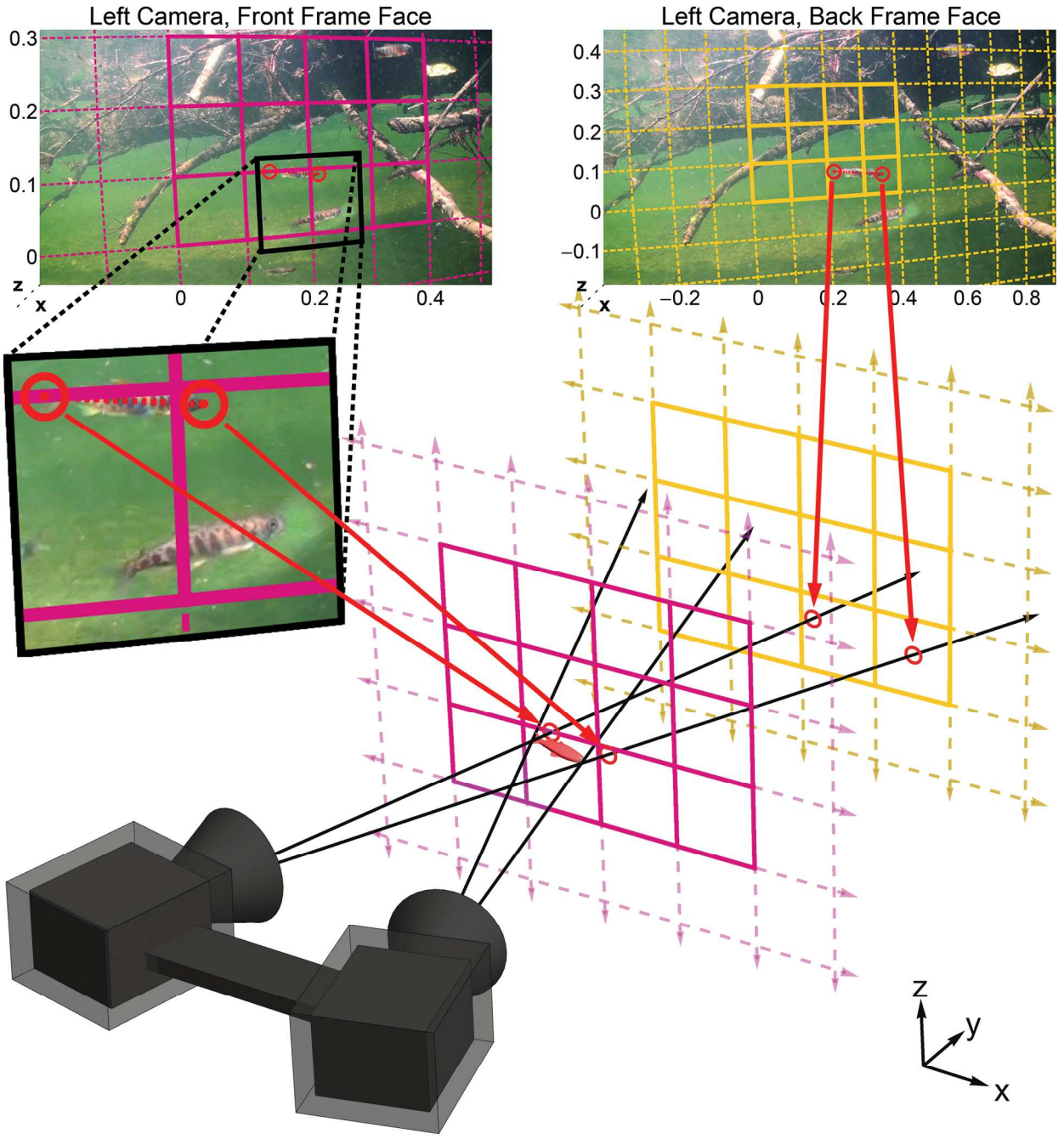
672

673 **Fig. 2.** Correcting non-linear distortion. (*a*) A sign printed with a chessboard pattern is filmed
674 close enough to fill the screen. (*b*) VidSync detects corners of the chessboard and arranges them
675 into plumblines for estimating the distortion model parameters. (*c*) Lines radiating from the
676 center of distortion (large black dot) show the magnitude and direction of distortion correction
677 from each detected chessboard corner. (*d*) Applying the correction to the original plumblines has
678 straightened them.



679

680 **Fig. 3.** Screen and calibration frame coordinate systems. A single image is overlaid with the (*u*,
681 *v*) pixel coordinates in which input is received and the (*x*, *z*) world coordinates (in meters) in the
682 2-D planes of the front ($y = 0$) and back ($y = 0.439$) faces of the calibration frame. The
683 homographies calculated during this calibration step convert between these coordinate systems
684 as shown, and they remain valid for measurement throughout the video (note the identical grid
685 overlays in Fig. 4).

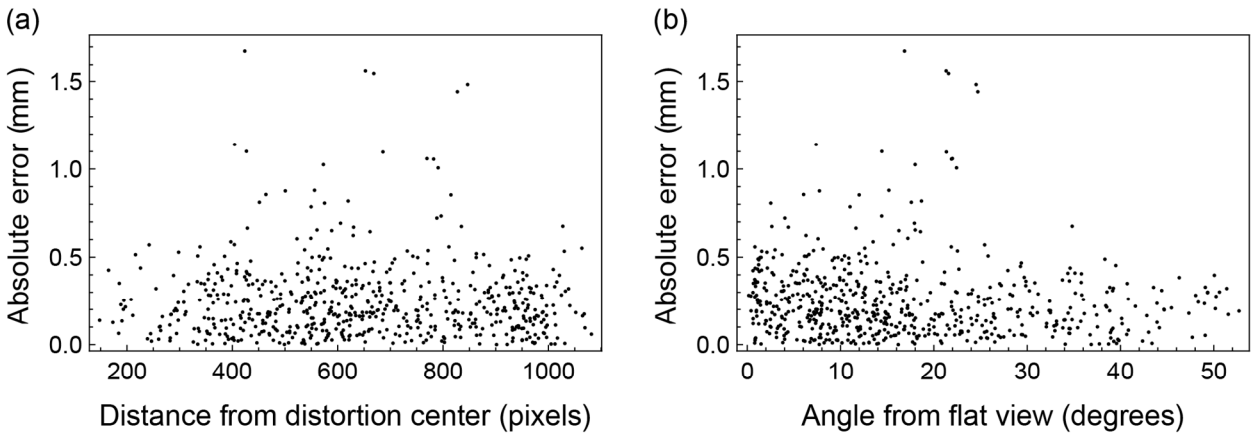


686

687 **Fig. 4.** Obtaining 3-D world coordinates to measure fish length. In the left camera, the user clicks
688 on the fish's head and tail. Those clicks (red circles) are expressed in (x, z) coordinates in the
689 planes of the front and back faces of the calibration frame, using the homographies described in
690 Fig. 3. Each of the two 2-D points (head and tail) is converted into two 3-D points using the
691 known y coordinates of the front and back frame faces. Mapped out in 3-D, these points define

692 the line of sight from the camera through the fish's head and tail. The 3-D positions of the head
693 and tail are measured as the estimated intersection of each line with the corresponding line from
694 the other camera. The fish's length is the Euclidean distance between its head and tail points.

695



696

697 **Fig. 5.** Relation between absolute error and (*a*) the maximum distance of one of the
698 measurement's endpoints from the center of distortion in that camera, and (*b*) the maximum
699 angle between the target and either of the cameras. Both plots use only data from 50.8-mm
700 targets within 2 m of the cameras, to reduce confounding effects of larger sources of error, such
701 as distance from the cameras.

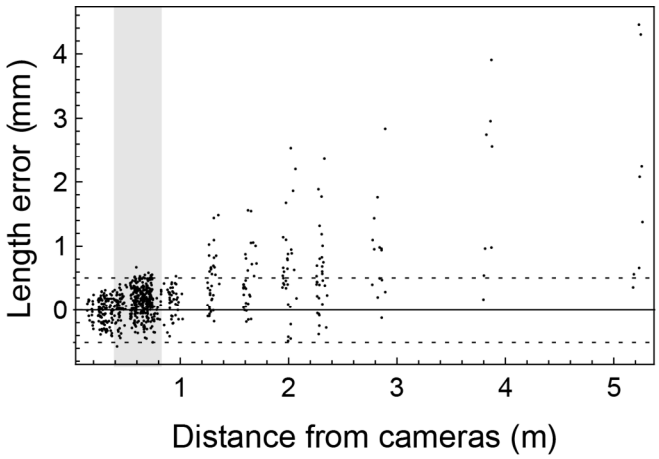
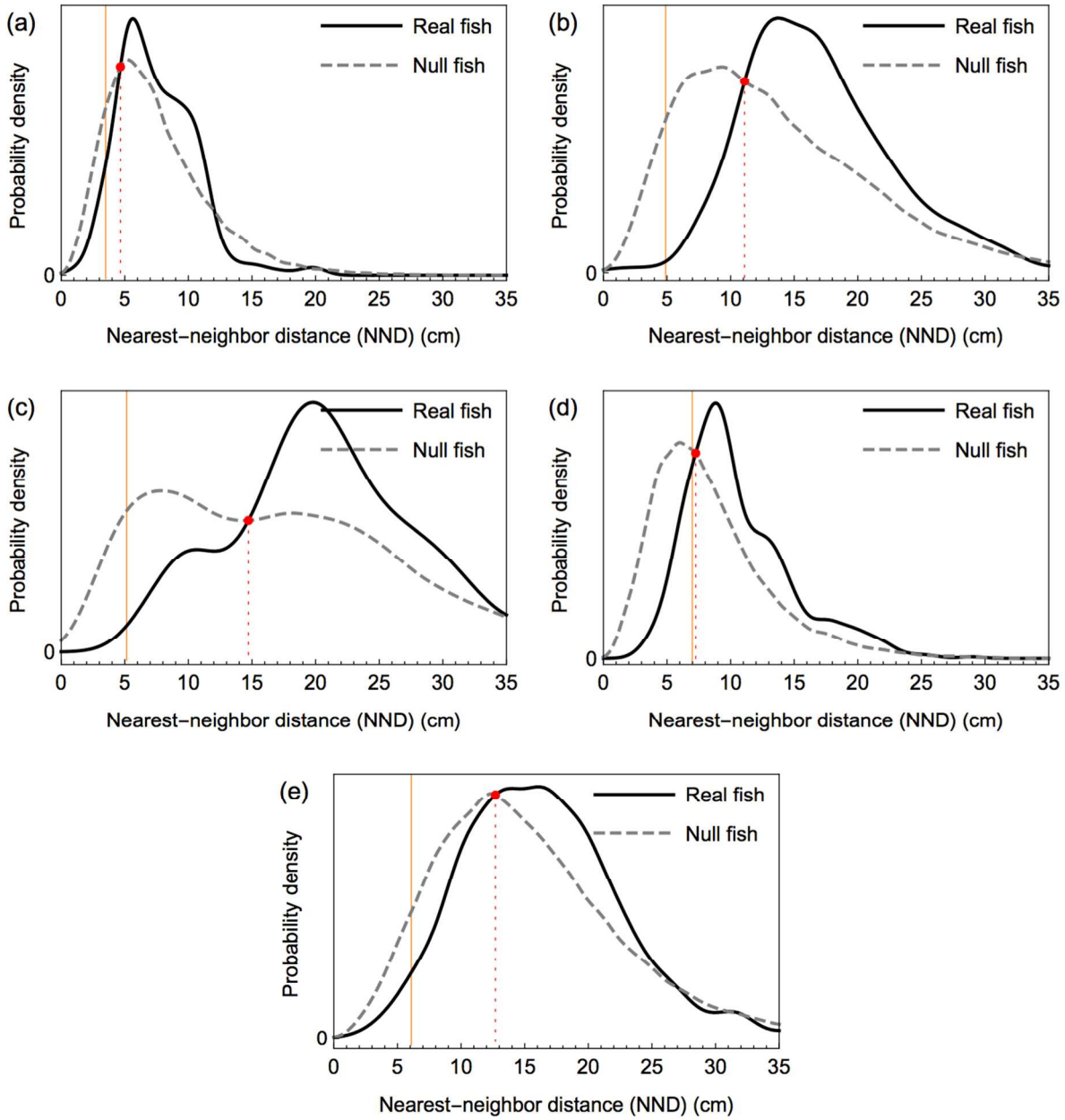


Fig. 6. Length errors (VidSync-measured length minus true length) in measuring a 50.8-mm object. Camera distance is measured from the midpoint of the length measurement to the midpoint between the cameras. The calibrated distance range, shaded in gray, is defined by the front and back plane positions of the calibration frame at the time of calibration. The dotted lines mark a threshold of 1 % error in the length measurement.

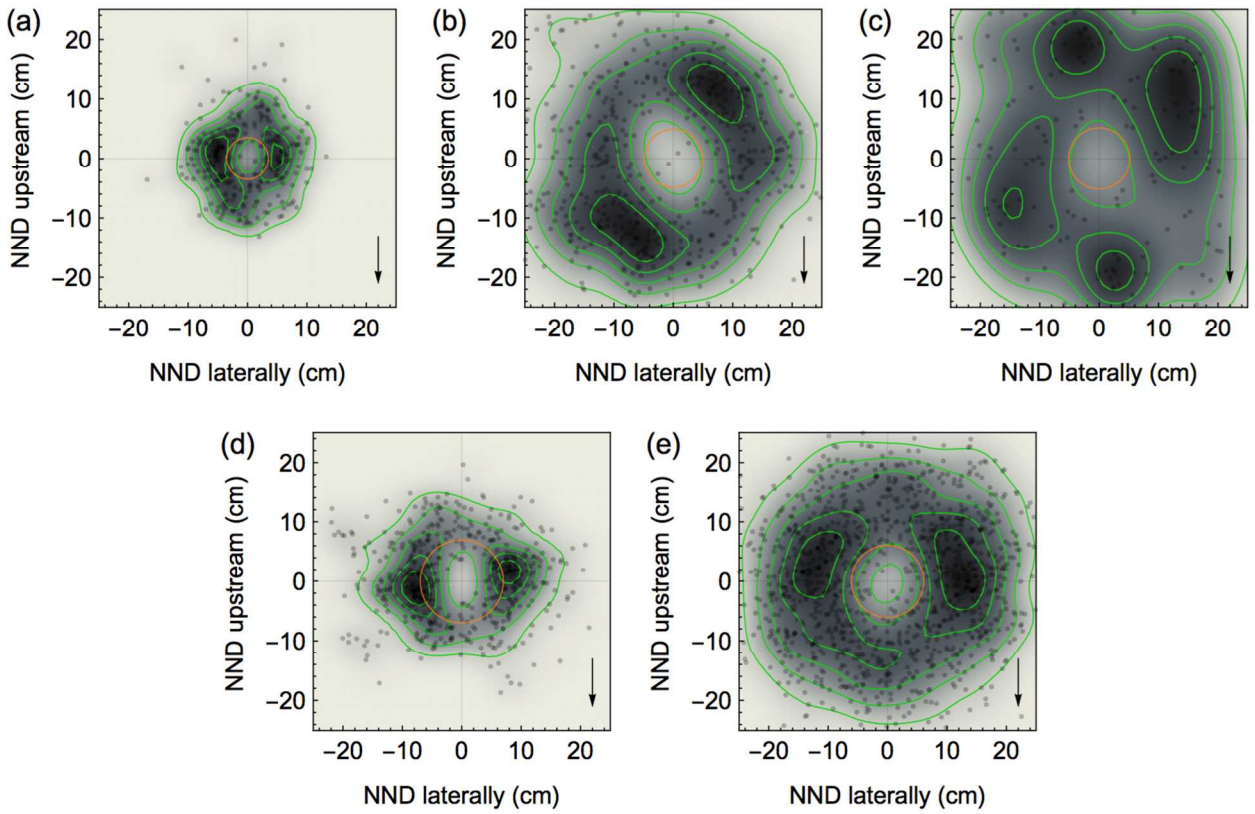


708

709 **Fig. 7.** Kernel-smoothed probability densities of the distance between any given fish and its
710 nearest neighbor (thick solid black line) and the expected distribution of this distance under a
711 null hypothesis in which fish are randomly distributed throughout the overall volume occupied
712 by the group (thick dashed gray line). The point at which these lines cross (red dot) indicates the
713 radius within which neighboring fish were less likely to be found than would be expected by
714 chance; in all cases, this was greater than the mean fork length of the fish (thin vertical orange

715 line). The five panels represent five sites filmed on (*a*) June 11, 2009, (*b*) June 28, 2010, (*c*) July

716 9, 2010, (*d*) August 14, 2009, and (*e*) September 15, 2010.



717

718 **Fig. 8.** Kernel-smoothed probability densities for the relative position of the nearest neighbor of
719 every fish observed. The position of each fish was set to (0, 0), and the relative position if its
720 nearest neighbor is shown by a gray dot. Positions were measured in 3-D and, without altering
721 the distance between fish, were rotated for display onto a horizontal plane passing through the
722 first fish and parallel to the water's surface. Dark shading indicates a high probability density of
723 finding neighbors in the shaded positions; some contours of this probability distribution are
724 outlined in green. Arrows on the bottom right indicate the direction of water flow. The radius of
725 the orange circle is the mean fork length of the fish. The five panels represent five sites filmed on
726 (a) June 11, 2009, (b) June 28, 2010, (c) July 9, 2010, (d) August 14, 2009, and (e) September
727 15, 2010.

Appendix A

Correcting refraction of the back plane points in a transparent calibration frame

Calibration frames with a transparent front face are appealing because of their potential precision and durability, but they introduce a small error that warrants correction. During calibration, light from the back surface passes through the front surface en route to the cameras, and it is refracted twice—as it enters and leaves that material—altering the apparent position of the points on the back face. These errors were on the order of 0.1 to 1 mm in our system, but importantly they are not random noise: their main effect is a slight apparent magnification of the entire back face, which substantially affects 3-D measurements. To eliminate this problem, consider a set of screen coordinates that were input by clicking on the refracted image of a back frame node during calibration. Because the frame is physically absent during later measurements, the calibration homographies should be calculated not with the real physical coordinates of the frame node's true position, but instead with its apparent position: the physical coordinates in the back frame face plane that would correspond to the same screen coordinates in the absence of the front face's refractive effect. For example, if a back frame node were physically located at (x, z) = (0.4, 0.3) meters, the correct homographies would map its screen coordinates not to (0.4, 0.3), but instead to its apparent position such as (0.4008, 0.3004).

This adjustment requires calculating the apparent position of a point B on the back frame plane, as viewed from a camera located at point C. A light ray traveling from B to C enters the front frame plane material at unknown point P_1 on the B side and exits at unknown point P_2 on the C side, so the full path of the light ray from B to C is $\vec{v}_1 + \vec{v}_2 + \vec{v}_3$, where $\vec{v}_1$ is a vector from B to P_1, $\vec{v}_2$ is from P_1 to P_2, and $\vec{v}_3$ is from P_3 to C. Let η_2 be the refractive index of the medium

through which $\vec{v}_2$ passes (the transparent frame material), while $\vec{v}_1$ and $\vec{v}_3$ pass through (usually the same) media such as water, with refractive indices η_1 and η_3.

Although VidSync performs this calculation with any coordinate orientation, assume for this explanation that the frame surfaces are parallel to the *x*-*z* plane, with known *y* coordinates. A unit vector normal to those planes is $\hat{n} = (0, 1, 0)$. Let subscripts *x*, *y*, and *z* denote their respective elements of the subscripted points. Having measured the thickness of the front frame material, P_{1y} and P_{2y} are known, and the unknowns are P_{1x}, P_{1z}, P_{2x}, and P_{2z}. These are calculated using Snell's law of refraction, which governs the angles (relative to the surface normal vector) at which light enters and leaves a surface. Let the ray coming from *B* enter the first interface at angle θ_1 from the normal and exit at θ_2. It enters the second interface at the same angle θ_2 (because the surfaces are parallel) and exits at θ_3, pointing toward *C*. These angles may be expressed in terms of the defined vectors as:

$$\theta_i = cos^{-1}\left(\frac{\vec{v}_i \cdot \hat{n}}{\|\vec{v}_i\|_2}\right) \quad (1)$$

These are used to write a system of four equations that depend on the four unknowns:

$$\begin{aligned} \eta_1 \sin\theta_1 &= \eta_2 \sin\theta_2 \\ \eta_2 \sin\theta_2 &= \eta_3 \sin\theta_3 \\ (\vec{v}_1 \times \hat{n}) \cdot \vec{v}_2 &= 0 \\ (\vec{v}_2 \times \hat{n}) \cdot \vec{v}_3 &= 0 \end{aligned} \quad (2)$$

The first two equations are the familiar form of Snell's law of refraction. The others specify that the light ray leaving each surface lies in the plane spanned by the normal vector and the ray that entered the surface (so the ray bends directly toward or away from the normal, rather than rotating around it).

VidSync solves this system for P_{1x}, P_{1z}, P_{2x}, and P_{2z} using a discretized version of the Hybrid algorithm for multidimensional root-finding, specifically the gsl_multiroot_fsolver_hybrids function of the GNU Scientific Library (www.gnu.org/software/gsl/). The points *C* and now-known P_2 define the camera's line of sight to the apparent position of the back frame point, which is recorded as the (*x*, *z*) coordinates at which that line passes through the *y* coordinate of the back frame plane. This apparent position is then used to calculate the calibration homography for the back frame surface.

VidSync users applying this correction need only specify the thickness of their front frame surface and refractive index of the medium (water or air) and frame material. Indices for several common materials are listed in the program. The correction can be disabled for users of wireframe-type calibration frames. Although the process described here is a type of refraction correction, it is specific to the described situation, and does not apply directly to the problem of correcting refraction through aquarium walls. However, analogous mathematics could be employed to extend VidSync for that purpose, and VidSync's two-plane calibration method is less sensitive to that problem than other common methods.

Appendix B

Calculating "stream coordinates"

A widely useful task in the study of stream fish behavior is the conversion of coordinates from those provided by VidSync, which are based on the orientation of the calibration frame relative to the cameras during calibration, into coordinates aligned with the true vertical direction and upstream-downstream axis of the region of interest. This technique is usable whenever cameras are placed in a stationary position to observe behavior, with the surface of the water visible and the current flowing in an approximately steady direction throughout the region of interest (i.e., not a swirling eddy).

To find the vertical direction, the first step is to measure at least four 3-D points on the water's surface, using distinctive cues like a twig poking through the surface or a fish striking floating prey to identify the same position on the surface in each camera. From each of the N points, subtract the mean of all N points and place the results as rows into an $N \times 3$ matrix A. Letting T denote the matrix transpose, perform a singular value decomposition on the matrix product $A^T A$. The right-singular vector corresponding to the smallest singular value of $A^T A$ is a unit vector normal to the plane of the water's surface, which we denote $\hat{n}_s$.

To find the downstream direction, we begin with calculating the mean current velocity vector within the region of interest. Individual velocity measurements are obtained by measuring two points along the path of a drifting item (either natural debris or an artificial tracer), subtracting the 3-D position of the upstream point from that of the downstream point, and dividing the result by the difference in time between the points. Averaging these vectors from several individual tracers produces a mean velocity vector $\vec{\omega}$. The projection of this vector onto

the plane of the water's surface is $\vec{\omega}_s = \vec{\omega} - (\vec{\omega} \cdot \hat{n}_s)\,\hat{n}_s$. This vector is normalized to calculate a unit vector pointing in the downstream direction, parallel to the plane of the surface, $\hat{n}_d = \vec{\omega}_s / \|\vec{\omega}_s\|_2$.

Finally, the unit vector in the "cross-stream" direction—perpendicular to the water velocity and parallel to the surface plane—is the cross product $\hat{n}_c = \hat{n}_s \times \hat{n}_d$. These three unit vectors comprise the rows of the 3 × 3 matrix *M*:

(1) $$M = \begin{pmatrix} \hat{n}_d \\ \hat{n}_c \\ \hat{n}_s \end{pmatrix}$$

Any 3-D position (*x, y, z*) in the calibration system coordinates provided by VidSync can then be converted into biologically meaningful "stream coordinates" (downstream x_s, cross-stream y_s, vertical z_s) coordinates by multiplying them by *M*:

(2) $$\begin{pmatrix} x_s \\ y_s \\ z_s \end{pmatrix} = M \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

The positive x_s direction is always downstream. The sign of the cross-stream and vertical directions may differ among videos depending on the calibration, surface, and velocity data, but it is easily determined by graphical inspection of results, and it can be flipped if necessary by multiplying $\hat{n}_c$ or $\hat{n}_s$ by -1 in equation (1).