LIMNOLOGY
and
OCEANOGRAPHY: METHODS

Limnol. Oceanogr.: Methods 10, 2012, 278–288

Distance maps to estimate cell volume from two-dimensional plankton images

*Emily A. Moberg and Heidi M. Sosik**
Biology Department, Woods Hole Oceanographic Institution, Woods Hole, MA 02543

Abstract

We describe and evaluate an algorithm that uses a distance map to automatically calculate the biovolume of a planktonic organism from its two-dimensional boundary. Compared with existing approaches, this algorithm dramatically increases the speed and accuracy of biomass estimates from plankton images, and is thus especially suited for use with automated cell imaging technologies that produce large quantities of data. The algorithm operates on a two-dimensional image processed to identify organism boundaries. First, the distance of each interior pixel to the nearest boundary is calculated; next these same distances are assumed to apply for projection in the third dimension; and finally the resulting volume is adjusted by a multiplicative factor assuming locally circular cross-sections in the third dimension. Other cross-sectional shape factors can be applied as needed. In this way, the simple, computationally efficient, volume calculation can be refined to include taxon-specific shape information if available. We show that compared to traditional manual microscopic analysis, the distance map algorithm is unbiased and accurate (mean difference = –0.25%, standard deviation = 17%) for a range of cell morphologies, including those with concave boundaries that deviate from simple geometric shapes and whose volumes are not well represented by a solid of revolution around a single axis. Automated calculation of cell volumes can now be implemented with a combination of this new distance map algorithm for complex shapes and the solid of revolution approach for simple shapes, with an automated decision criterion to choose the appropriate approach for each image.

The need for automated methodologies to determine phytoplankton biovolumes has long been established in the literature, with motivation typically to reduce analysis time for microscopists quantifying taxon-specific biomass (e.g., Albertano et al. 1997; Carpentier et al. 1999; Estep et al. 1986; Krambeck et al. 1981; Sieracki et al. 1998; Sieracki et al. 1989). Semi-automated methods (i.e., wherein a computer interface mediates the entry of data or on-screen measurement with image analysis tools) are available and facilitate faster analysis (Carpentier et al. 1999; Congestri et al. 2000); however, for very large data sets, such as those generated by new automated cell imaging technologies, semiautomated methods are almost as prohibitively time consuming as traditional microscopy. For microplankton, the advent of technologies such as FlowCAM (Sieracki et al. 1998) and Imaging FlowCytobot (Olson and Sosik 2007) provides the capability to produce thousands of cell images per hour and long term data sets with > 10^8 images. This challenge has led to automated methods for taxonomic identification of images (e.g., Sosik and Olson 2007), but geometric analysis for accurate estimation of cell biovolumes from two-dimensional images is still needed.

Existing methodologies for cell volume estimation from two-dimensional images have been largely restricted to simple shapes such as spheres or cylinders or composites of those simple shapes (e.g., Hillebrand et al. 1999; Sun and Liu 2003). Sieracki et al. (1989) provided an important advance by proposing automated methodology based on creating a solid of revolution around a major axis (Fig. 1). This approach is appropriate for cells with convex boundaries or those whose boundary is intersected only twice by any line perpendicular to the axis of rotation. It is unsuitable, however, for complex plankton shapes such as those that are concave with multiple branching points associated with appendages or chain morphology.

A wider range of shapes can be accommodated by semiautomated methods and traditional microscopy because the

*Corresponding author: E-mail: hsosik@whoi.edu

Acknowledgments

We thank R. Olson, A. Shalapyonok, E.T. Crockford, E. Peacock, and members of the MVCO Operations Team, especially J. Fredericks, J. Sisson, and H. Popenoe, for making it possible to collect the image data used in this research. This research was supported by grants (to HMS) from the Gordon and Betty Moore Foundation and NASA's Ocean Biology and Biogeochemistry program, and a Woods Hole Oceanographic Institution Summer Student Fellow award (to EAM).

DOI 10.4319/lom.2012.10.278

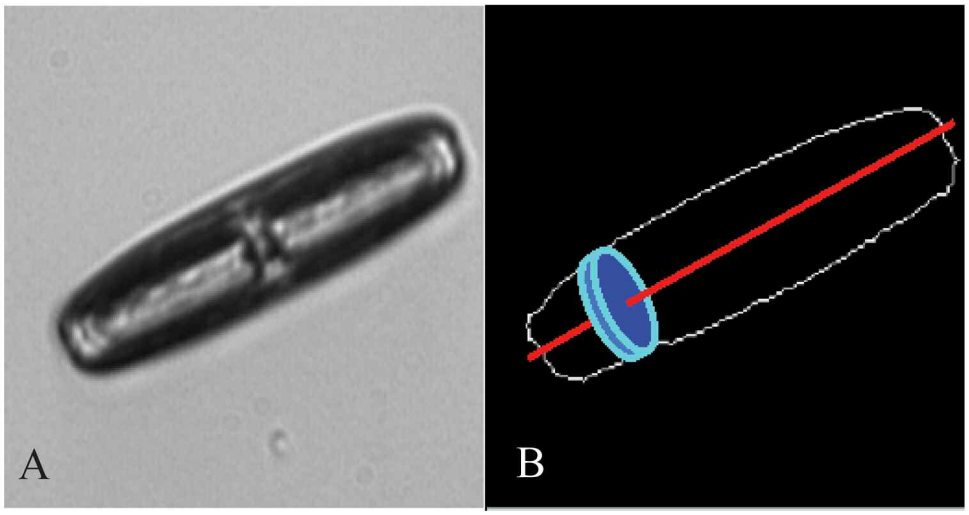

Fig. 1. (A) Original gray-scale image, and (B) associated target boundary, with major axis (red line) and one slice of a solid of revolution illustrated (blue pixel-thick disk projecting in the obscured "into the page" dimension).

technician can use knowledge about taxonomy of the organism of interest and infer appropriate geometric formulae for either the whole cell or parts of the cell (Hillebrand et al. 1999; Sun and Liu 2003). Such informed analysis is likely to provide accurate estimates of biovolume, especially by reducing uncertainty regarding shape and size in the dimension perpendicular to the viewing plane. As this strategy requires user input for each image, however, it is not suitable for rapid assays of large numbers of images.

Our objective is a fully automated approach that can operate on an image regardless of a priori knowledge of biological taxon or target shape, yet still provide acceptable accuracy for biovolume. Computational efficiency of the algorithm is also of concern, as the methodology is intended for processing larger amounts of data than used in prior studies. Last, the approach needs to be highly reproducible and free of anomalous results for unexpected cell shapes, as no user interaction is expected to be available to detect such problems. Though it may be applicable for other image types, the methodology we describe was specifically targeted for phytoplankton and protozoa in the microplankton size range (20-200 µm). Furthermore, given the effectiveness and computational efficiency of the Sieracki et al. (1989) rotational method for convex shapes, the new distance map algorithm described here is designed to be applied only for cells and chains of cells that fail to meet the criteria for rotation about a single axis. For a range of cell types that might be encountered in natural plankton communities, we propose that the rotation and distance map algorithms be used in combination, with automated criteria to choose the best algorithm for each image.

Materials and procedures

The distance map algorithm was developed and evaluated with data collected by Imaging FlowCytobot (Olson and Sosik 2007; Sosik and Olson 2007) at the Martha's Vineyard Coastal Observatory (http://www.whoi.edu/mvco). However, the general approach outlined here could be applied to any phytoplankton (or other microplankton) images that are available digitally. Imaging FlowCytobot is an imaging-in-flow cytometer (Olson and Sosik 2007), which uses video technology to collect 8-bit gray-scale images through a 10× objective during a 1-µs exposure. As in conventional flow cytometry, particles are hydrodynamically focused in a flowing stream of seawater such that they pass one at a time within the focal depth of the collection optics. In Imaging FlowCytobot, image acquisition is triggered by fluorescence generated when pigmented cells pass through a laser beam (perpendicular to the flow and upstream of the image site). As such, this system produces in-focus images typically of single cells (or chains of cells) with approximately 1 µm resolution and at rates that can exceed 10 Hz. The instrument is also automated and submersible so large quantities of image data (order 1000s h^{-1} for months to years) necessitate automated processing.

Preprocessing of images to ensure proper edge detection and identification of closed boundaries was done according to the approaches previously described by Sosik and Olson (2007). MATLAB (The MathWorks) was used to perform all calculations and image processing, but other languages would be equally well suited; a MATLAB function is provided along with pseudo-code in Appendix Table 1 to facilitate implementation in a variety of programming environments. To use this methodology, the basic requirements are digital, binary boundaries of phytoplankton (or other targets of interest in images), sufficient storage space for those images, and a computing language capable of processing those images.

To assist with algorithm assessment, a selected set of images encompassing a range of complex plankton morphologies were inspected and measured for manual determination of

volume. A few images were rejected from inclusion in this test set if adequate manual analysis was not possible (e.g., noisy boundaries, detritus stuck to cells, etc.). Boundary images were displayed, and measurements of dimensions were recorded with the *imtools* user interface in MATLAB. A minimal number of measurements were recorded to mimic conventional microscopic measurement techniques. Biovolumes were then estimated from the formulae recommended for manual microscopy by Hillebrand et al. (1999). Inherent in these volume calculations is the assumption that the obscured dimension has symmetry with the image plane along one axis or a set of component axes in cases with multiple comprising shapes (e.g., radiating cylinders as in Fig. 2).

Distance map biovolume algorithm

To describe the distance map biovolume algorithm, we first define some relevant terms. The "target" is a cell, chain, or colony of cells, or other object of interest within an image. Our objective is to estimate the volume, or "biovolume" for the application to organisms, of this target. Each target is enclosed by a "boundary" which is a closed polygon one pixel wide (Fig. 2). In general, more than one target each with associated boundary can be contained within a single image. Each would be treated independently in the following volume calculations. For a given target, the "image plane" is the two-dimensional plane wherein all the pixels lie (i.e., plane of the page). The "obscured dimension" is defined as the direction orthogonal to the image plane and is equivalent to the into-the-page dimension for the image. The "centerline" is the line, curve, or set of line segments whose points are interior to the shape and equidistant from the closest boundaries; if a shape, or portion of a shape, is conceived as a solid of revolution, this line would be the axis of rotation. The centerline applies locally to a region of a target that is convex; for a branched target such as a cross, the centerline is comprised of the long axes of each of the four branches (Fig. 2). A "transect" refers to a line segment orthogonal to the centerline (in the image plane), which is one pixel wide and has endpoints at the adjacent boundaries.

The cornerstone of this method is the idea of a "distance map." The distance map is a matrix the size of the image in pixels. Each element in this matrix contains the distance of that pixel/element from the nearest boundary. A pixel on the boundary has a distance of 0; a pixel midway between two boundaries has the value of half the distance from boundary to boundary.

We derive the distance map **D** for a boundary image by calculating the Euclidean distance of each pixel to every boundary point, and then assigning each element $d_{i,j}$ of **D** as the minimum from the resulting set of distances for the corresponding pixel [*i,j*] (Fig. 3). Because the boundary has a value of zero, which is indexed incorrectly for later calculations, we add 1 to every pixel location. The resulting overall distance map (which also includes distances outside the target image) is masked by the target to keep only those values within the interior (i.e., pixels outside the closed boundary are reassigned a distance of zero) (Fig. 3). The distance map is thus a matrix with dimensions [*i,j*] equal to the height and width of the image in pixels; only the pixels corresponding to the interior of the target have positive values associated with distances from the target boundary.

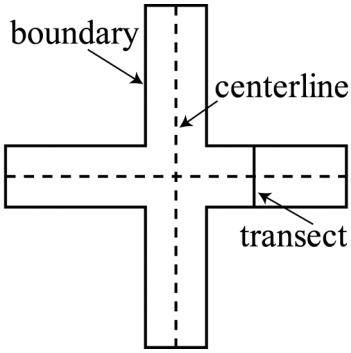

Fig. 2. Illustration of boundary, centerline, and an example transect for a hypothetical concave target.

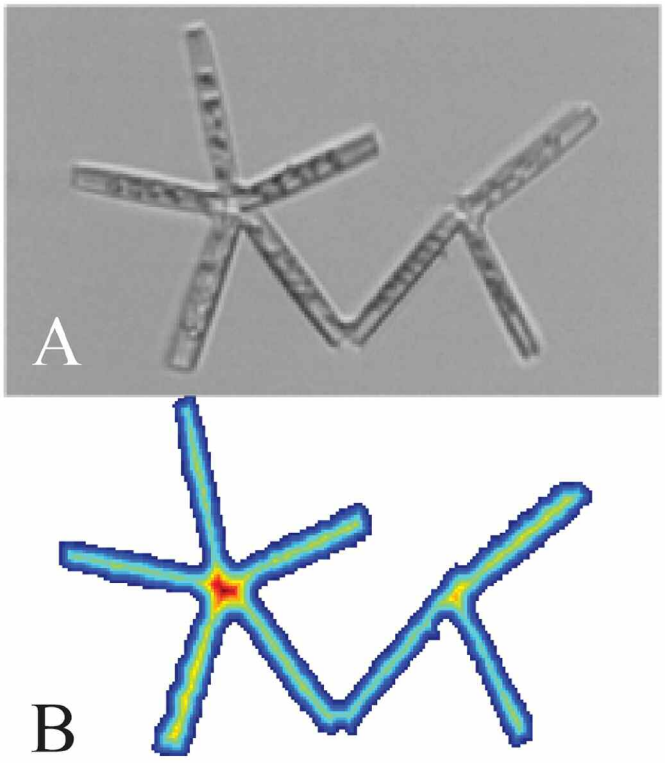

Fig. 3. (A) Original image and (B) associated distance map for a diatom chain (genus *Thalassionema*). In the distance map, values range from 0 pixels (dark blue) to 15 pixels (red).

We then assume that each pixel's distance from a boundary is equal to half the total thickness of the object in the obscured dimension. That is, each distance $d_{i,j}$ from the boundary is projected orthogonally both upwards and downwards in the obscured dimension (Fig. 4).

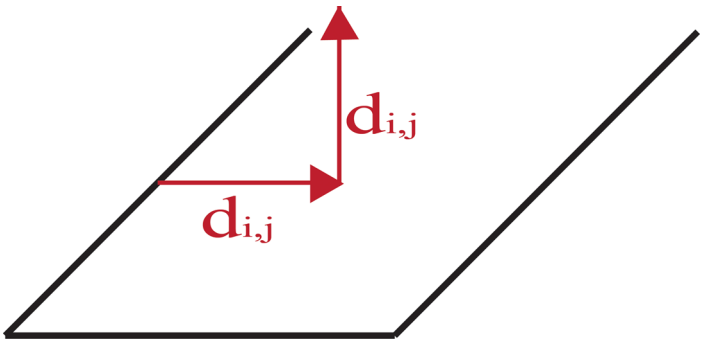

Fig. 4. Illustration of distance map projection into three-dimensional space. For the pixel located at the head of the horizontal arrow, its distance $d_{i,j}$ from the nearest boundary (black line at base of horizontal arrow) in the image plane is projected into the obscured (third) dimension. This projection is made both above (as depicted here) as well as below the image plane.

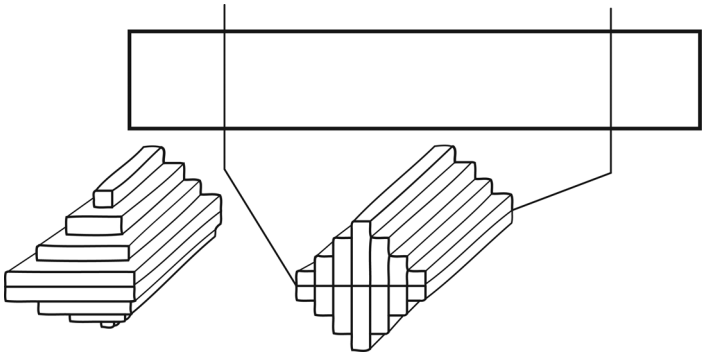

Fig. 5. Depiction of the 3D "step-pyramid" that results from a rectangular boundary once its distance map is projected into the obscured dimension. Note that the center section of the target has a different projection (right pyramid) from the distal end sections (left pyramid).

Because we are dealing with discrete pixels, the resulting overall shape is now a step pyramid, where the step size is one pixel. Each pixel is now at the center of a "bar" with a cross section of one pixel, oriented perpendicular to the image plane and extending twice the distance of that pixel from its nearest boundary (i.e., the height of the bar is $2d_{i,j}$; Fig. 5). We can then estimate the overall volume of the solid formed from these discrete "bars" as the sum of twice the distance map, D:

$$V_1 = 2\sum_i \sum_j \mathbf{D} \quad (1)$$

where V_1 is estimated target volume in pixel cubes.

This calculated V_1 provides a first order approximation of the space filled by a plankton cell or chain (Fig. 6), but we achieve a shape that more closely mimics those typically found in nature by applying a correction to convert the pyramidal cross-section (in the obscured direction) to a circular one; this effectively provides the same result as creating a solid-of-revolution about the local centerline. We derive this correction factor in two steps.

First, consider that, in the limit for infinitesimally small pixels, geometrically, the pyramid cross-section approaches a simple diamond (Fig. 7A). V_1 can be adjusted to reflect true diamond cross-sections with a multiplicative factor that accounts for the effect of pixel resolution. Consider a transect of length x; this would correspond to a transect whose central element, $d_{i,j,}$ is $x/2$. If x is an even number, the volume achieved by adding the bars across the transect is

$$\frac{x^2}{2} + x \quad (2)$$

and if x is odd, it is

$$\frac{x^2}{2} + x + \frac{1}{2} \quad (3)$$

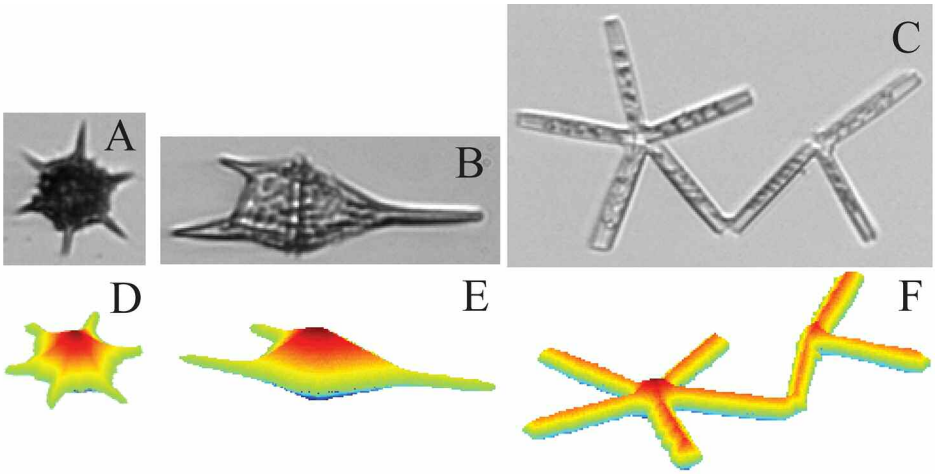

Fig. 6. (A, B, C) Example original images of plankton and (D, E, F) their distance map representations in "step-pyramid" form. (A, D) *Dityocha* sp. (B, E) *Ceratium* sp. (C, F) *Thalassionema* sp. In these 3-D distance map representations, yellow to red colors indicate larger distance above the image plane, green to blue colors, a larger distance below the image plane.

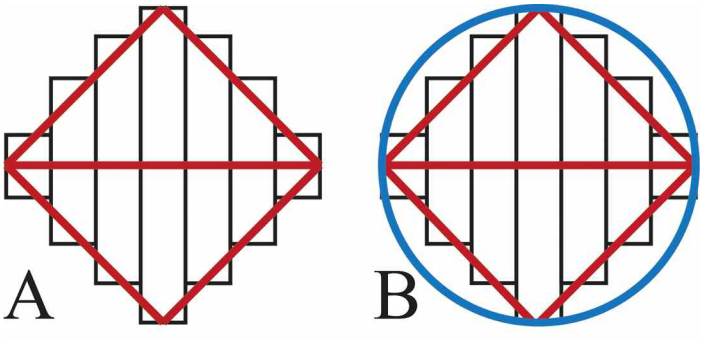

Fig. 7. Schematic of the transitions (corrections) from "step-pyramid" (A) to a diamond, and then (B) to the circumscribing circle that is used as the final approximated shape in the obscured dimension (i.e., orthogonal to the image plane).

Because we are equally likely to encounter odd and even values of x, the expected value of this volumetric formula is

$$\frac{x^2}{2} + x + \frac{1}{4} \tag{4}$$

The volume, if calculated for the inscribed diamond, is

$$\frac{x^2}{2} \tag{5}$$

The multiplicative correction factor for a given transect is simply then the quotient of the geometrical volume (Eq. 5) over the summative volume (Eq. 4):

$$c = \frac{\frac{x^2}{2}}{\frac{x^2}{2} + x + \frac{1}{4}} = \frac{x^2}{x^2 + 2x + \frac{1}{2}} \tag{6}$$

This is simple when applied to an individual transect, but more complicated to apply to real shapes. Rather than slow computation time by calculating each transect individually (which also gets complicated near intersection regions, where transects may not be well defined), our goal is to derive a representative x that we can use for the whole image. To do this, we derived a relationship between the mean value of all elements in **D** and a representative transect length, x_r, for that distance map. Consider a hypothetical $\mathbf{D_h}$ for a transect with length x_h. By analogy with Eq. 4, the distance map sum in this case is

$$\sum_i \sum_j \mathbf{D} = \frac{x_h^2}{4} + \frac{x_h}{2} + \frac{1}{8} \tag{7}$$

[Note that the factor of 2 difference between Eq. 7 and Eq. 4 results because Eq. 4 includes both the upper and lower portion of the step-pyramid (i.e., the factor of 2 in Eq. 1).] The mean value of all elements in **D** can then be determined by dividing the sum (Eq. 7) by the length of the centerline (x_h):

$$mean(\mathbf{D}) = \frac{\frac{x_h^2}{4} + \frac{x_h}{2} + \frac{1}{8}}{x_h} = \frac{x_h}{4} + \frac{1}{2} + \frac{1}{8x_h} \tag{8}$$

For values of x_h greater than 3, the nonlinear term $1/(8x_h)$ becomes negligible, and the equation can be rearranged to solve for x_h:

$$x_h = 4(mean(\mathbf{D})) - 2 \tag{9}$$

For the general case of **D** with variable transect lengths, this x solution (Eq. 9) can be interpreted as the representative transect length, x_r, in the associated pyramidal volume. x_r determined this way can be inserted into Eq. 6 to provide an operational version of the first correction factor.

The resulting correction factor, c_1, adjusts the pyramidal cross-section to an interpolated diamond (Fig. 7A). We then calculate a second correction factor to convert the diamond cross-section to a circular one. Our goal is to replace the diamond with a circle that circumscribes it (Fig. 7B). From the areas of the inscribed diamond (A_d) and the circumscribing circle (A_c) of diameter and diagonal length $2r$, we can derive the second multiplicative correction c_2:

$$c_2 = \frac{A_c}{A_d} = \frac{\pi r^2}{2r^2} = \frac{\pi}{2} \tag{10}$$

The final target volume, V, is thus:

$$V = c_1 c_2 V_1 = \left(\frac{x_r^2}{x_r^2 + 2x_r + \frac{1}{2}} \right) \pi \sum_i \sum_j \mathbf{D} \tag{11}$$

where

$$x_r = 4(mean(\mathbf{D})) - 2 \tag{12}$$

Algorithm criterion

This distance map algorithm is suited to concave, complex shapes, and may under-represent shapes such as those with small rectangular cross-sections (*see* "Assessment"). The previously described rotational method of Sieracki et al. (1989) (hereafter the "Sieracki method"), on the other hand, is accurate and computationally efficient for simple shapes and fails for more complex ones. Consequently, for image data sets with variable shapes, both algorithms are needed for most accurate and efficient volume computations. Our implementation of the Sieracki method is faster than the distance map algorithm, so using it for the abundant, small, near-circular plankton decreases overall computing time for typical natural samples. Selecting the appropriate algorithm for each image is then a critical step. We use the ratio of convex area to overall area of the target (in the image plane) as the parameter for differentiation (Fig. 8). Because we found that analyzing a complex shape with the Sieracki method has the potential to result in higher errors, we

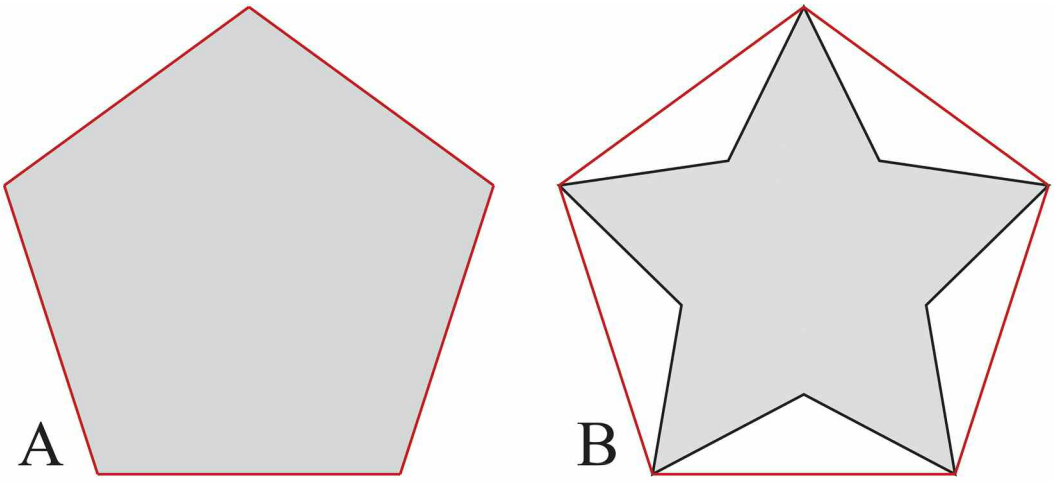

Fig. 8. (A) An example convex shape and its convex boundary (both red). The ratio of the convex area to the actual area (as depicted in gray) is 1. (B) An example concave shape (black) and its convex boundary (red). The ratio of the convex area (the white and gray areas) to the actual area (gray) is > 1.

empirically chose a threshold of 1.2, such that it is more likely for targets with nearly circular boundaries to be selected for application of the distance map algorithm than for complex shapes to have the Sieracki method applied. We note that, due to edge detection noise, even generally convex shapes empirically exhibit ratios > 1 with our image processing.

Assessment

Initial evaluation of candidate approaches

We begin with a brief discussion of why the distance map algorithm was chosen over other candidate approaches. As mentioned above, rotation/integration about one axis (as implemented by Sieracki et al. 1989) was rejected as it is not suitable (highly inaccurate) for complex, and especially concave, shapes. A related approach that breaks the image into segments before rotation, as was manually done by Congestri et al. (2000), was considered more carefully, as it circumvents the issues of rotating a boundary about a single non-representative axis. This is accomplished by identifying multiple axes each through a different component of the overall shape. To explore automating this approach, we investigated a skeletonization technique (to identify the centerlines of segments to be broken apart). In initial tests, however, we found these approaches to be computationally intensive, sensitive to edge noise/irregularities, and unsuitable for use with boundaries having complex morphometries, especially shapes with many intersections that were not easily defined by the skeleton limbs. Furthermore, errors of only a couple of pixels could result in rotation around the wrong axis and > 100% volume errors. The use of skeletons in image processing is still being developed, but the primary focus is apparently for image recognition or medical imaging (e.g., Chen et al. 2009; Nemeth et al. 2010), with results that are not immediately portable to automated processing of plankton images with the accuracy, reliability, and computational efficiency we desire.

Another approach that we considered involved identifying a set of simple geometric shapes (e.g., rectangle, circle) that together would represent the overall boundary, and then applying geometric formulae accordingly, similar to the method often used in manual microscopic analyses (Carpentier et al. 1999; Hillebrand et al. 1999; Sun and Liu 2003). Preliminary tests, however, suggested this approach was computationally intensive and prone to errors when automated. Furthermore, automatically detecting sub-shapes to comprise a two-dimensional shape does not help constrain the three-dimensional projection (e.g., a circle can be projected as a sphere or a disc of indeterminate thickness).

The distance map algorithm developed here is computationally simpler and better suited for application to a wide variety of complex shapes than the other methodologies considered.

Sources of error

There are several potential sources of inaccuracy in volume estimates from our distance map algorithm. First, the boundary upon which the method is applied may not accurately represent the boundary of the organism in nature; this is largely a function of the quality of the original image and its post-processing, and as such, we assume the user knows about and is in control of this source of error. Second, simplifications involved in deriving the algorithm correction factors could result in quantifiable volume errors. Third, the inherent assumption about the relationship between the obscured dimension and the image plane may not be accurate for some types of targets (e.g., certain plankton forms with flattened discoid shapes, triangular cross-sections, etc.). It should be noted that all volume estimates from single plane microscopic and image analysis methods require some assumptions about this obscured dimension, whether it is from prior (expert) knowledge of assumed shape based on taxonomy or otherwise. The distance map algorithm is particularly well suited to those targets that are imaged in the plane of largest area, because that plane con-

tains the most information about overall target shape. Because the second and third errors are within the control of this specific framework, they will be addressed further.

Edge effects comprise one type of error. The calculations for transects (Eqs. 2-6) apply only when the boundaries parallel to the centerline are closest; at the distal margins, the orthogonal boundaries to the centerline are closest, which results in a pinching effect at the edges as they taper (Fig. 5). If the actual edges of the target are sharply cylindrical (flat edges in planes through the obscured dimension), the algorithm will underestimate in that local region. In principle, this effect could be reduced with additional correction factors applied to target margins. We note, however, that the biological reality of pinched versus sharp edges varies by species. As we have developed it here, the distance map algorithm, through the aforementioned "edge effects," assumes that phytoplankton have chamfered, rather than right-angled edges. Electron micrographs of cleaned diatom frustules (Round et al. 1990) show that the assumption of such chamfered edges is reflective of the morphology of many species, including some discoid genera, such as *Coscinodiscus*. However, it should be noted that some genera, such as *Eupodiscus*, *Sheshukovia*, and *Acinocyclus*, do have pronounced right-angled edges, and traditional geometrical formulations that assume such edges (e.g., cylinder) could better represent their volumes.

In considering these edge effects, it is possible to quantify relative errors that might be expected for organisms with right angled edges. Consider a rectangular boundary of height x and length L. The area affected corresponds to the area nearest the distal boundaries extending inward $x/2$ on either end (total area impacted by tapering = x^2); the rest of the shape (unaffected area = $xL - x^2$) will have its volume approximated as strictly cylindrical. The effect of edge tapering on the overall volume calculation can be determined via summation of pixels. In the edge regions, the volume can be obtained as

$$4\sum_{i=0}^{i=x/2}[(i^2+i)+(x-2i)(i+1)] \tag{13}$$

where i is the index of the pixels between the boundaries that define the width of the target. The volume in the center can be computed as

$$(L-x)\left(\frac{x^2}{4}+\frac{x}{2}\right) \tag{14}$$

From Eqs. 13 and 14, we can compute the expected difference in total volume from that for the equivalent shape without tapered edges (i.e., the "center region" extends the full length of L). For x values larger than 8 pixels (non-linearities occur below this limit, and below which the methodology was not designed; this corresponds to ~2 µm for images from Imaging FlowCytobot), length-to-width ratios > 6.6 produce errors < 5% (defined as abs[1 – $V_{flat}/V_{tapered}$]), while length-to-width ratios > 3.5 produce errors < 10%. As the ratio gets closer to one, errors approach ~30%.

Application to real phytoplankton shapes

Our objective is to apply the distance map algorithm to complex phytoplankton shapes whose boundaries may be noisy and irregular; so we explored its performance on a variety of images spanning a number of taxonomic groups. The test set was chosen specifically to reflect the complex types of shapes that can occur, and is not representative of the relative abundance of shapes within our dataset from natural assemblages at the Martha's Vineyard Coastal Observatory. The image test set was analyzed to determine volumes calculated manually and algorithmically. As mentioned above, the manually derived volumes assume that the obscured dimension has symmetry with the image plane along one axis or a set of component axes for cases with multiple comprising shapes. In this way, the manual calculations are consistent with the major assumption in the distance map algorithm. Plankton shapes were defined as locally cylindrical, spherical, or spheroid.

Taxa such as the branching chain diatom *Thalassionema* (Fig. 9A) were manually represented as separate rods and hand-measured and algorithmic estimates of volume were very similar (e.g., <1% difference for the example in Fig. 9A).

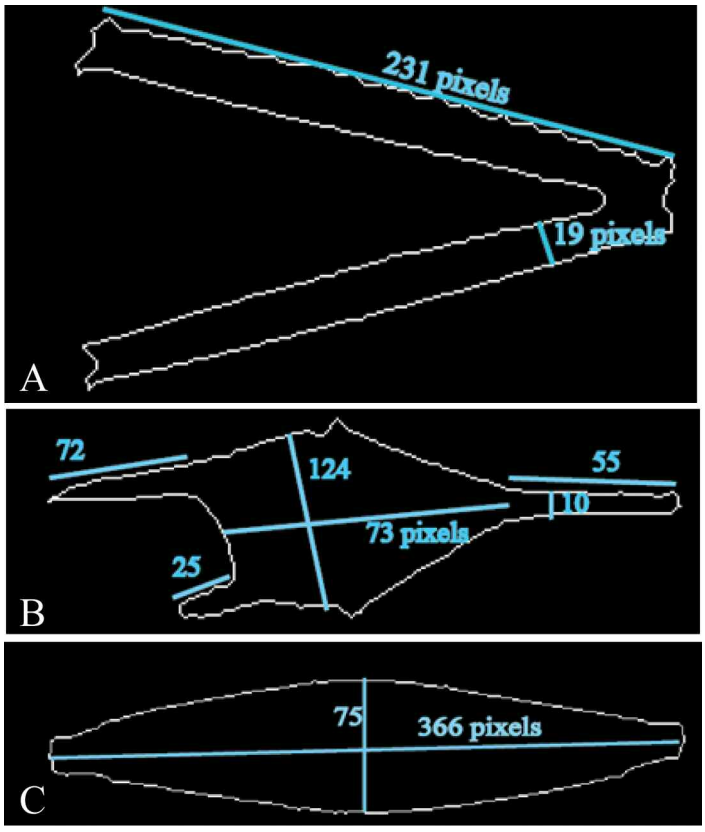

Fig. 9. Selected plankton boundaries with hand-measurements depicted. Local boundary shapes were chosen to be represented as spheres, spheroids, or cylinders (A) *Thalassionema* was represented as 2 cylinders, (B) *Ceratium* was represented as a central prolate spheroid and 3 protruding cylinders, and (C) *Pleurosigma* was represented as a prolate spheroid. Cells not shown to scale.

More complex forms were represented as combinations of different shapes, such as the case for the thecate dinoflagellate *Ceratium* (Fig. 9B) modeled as branching rods around a spheroidal center. *Ceratium* is an example of a planktonic form that requires many hand measurements and for which assigning a few shapes—like a prolate spheroid to the center—produces an overly simplified representation. In these cases, hand-measured and algorithmic volumes can exhibit larger differences (e.g., 17.2% for Fig. 9B), with the distance map estimate likely more accurately reflecting the true organism shape (see Fig. 6E). Algorithmic representation of volume that is substantively better than the manual approach can occur even for simpler shapes, such as the nearly convex diatom *Pleurosigma* (Fig. 9C) typically analyzed with the Sieracki method. In these cases, hand measurements can overestimate true volume, for instance, because a single prolate sphere cannot accommodate features such as tapered distal ends that are readily reproduced in a solid of revolution or distance map approach (hand estimate 17% higher for example in Fig. 9C).

To further quantify the performance of the combined distance map and solid of revolution approach across a broad range of taxa, a total of 50 images were analyzed both by hand and automatically (Fig. 10). This data set included some near-circular targets, which were analyzed with the Sieracki method if they fell below the convex-to-total area ratio cutoff of 1.2. The mean percent difference between the two methods was 0.25% (std. dev. 17%), indicating no systematic over- or underestimate by the algorithm relative to hand-measurements. The mean of the absolute value (i.e., excluding sign, which only indicates whether the hand-measured or algorithmically calculated volume was larger) was 13.6%. Furthermore, the percent difference is only weakly correlated with target volume (Fig. 11). Since algorithmic edge effect errors should be relatively larger for smaller targets, this result suggests this type of error is not a dominant source of uncertainty or bias in volume estimates.

Computational speed was monitored with the *profile* function in MATLAB. For a set of images of various taxa, the Sieracki method required 0.096 s for 18 images (~0.005 s/image), whereas the distance map algorithm took 0.7 s to process 57 images (~0.01 s/image) on a single 1.6 GHz processor; the images were chosen randomly and the convex area threshold was used to determine which method was used. The time dependence on image size was also investigated (Fig. 12). This analysis gives a rough estimate of performance time and emphasizes that images can be processed rapidly enough to be practical for large datasets. This will be especially the case with fast, multi-core, multi-processor computing systems since the analysis of many images is inherently highly parallelizable.

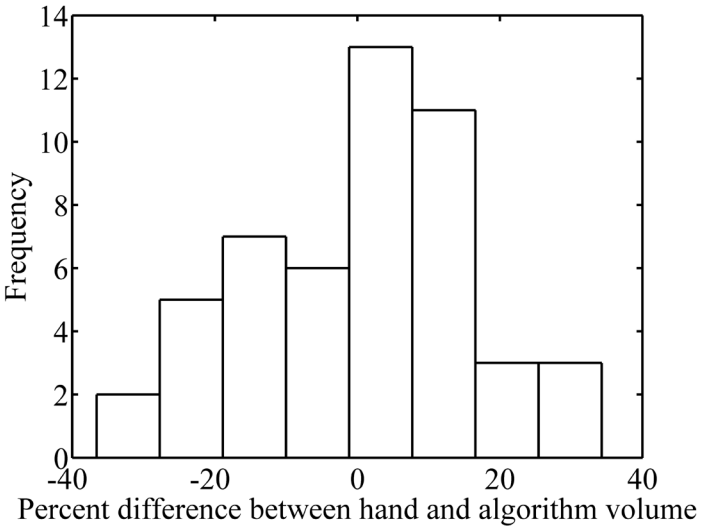

Fig. 10. Histogram of percent differences between the hand-measured and algorithm volumes. The mean value is –0.25%, with a standard deviation of 17%. Differences between the methods should not be interpreted as error in the algorithm because there are notable cases where the conventional hand-measurement approach has large uncertainty (see Fig. 9B, C for examples).

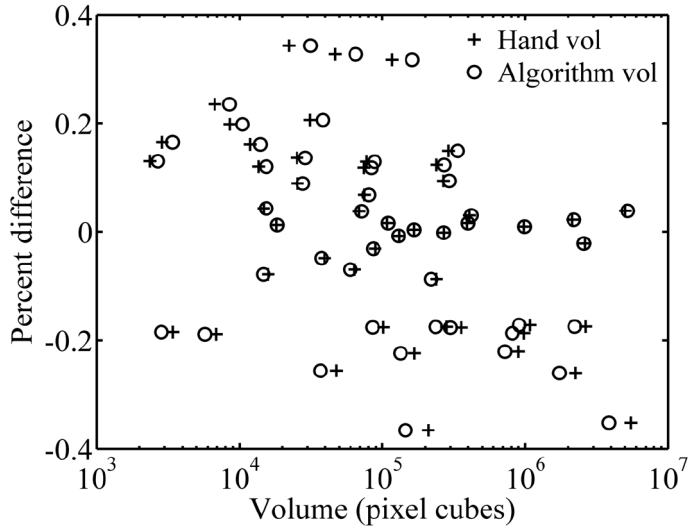

Fig. 11. Dependence of percent difference between algorithmic and hand-calculated volume on cell volume showing showing weak negative correlation (r^2 = 0.12 and 0.07 for the hand-calculated and algorithmic data, respectively; n = 50).

Discussion

Comparison with techniques typically applied manually for analysis of microscopy samples shows that the distance map biovolume algorithm provides unbiased estimates that are within errors expected from measurement uncertainty (Fig. 10). This conclusion applies to a wide range of planktonic forms, with the notable prospect of more accurate volume representations than are typically achieved with manual analysis of complex convex shapes. Furthermore, analytical speed is many orders of magnitude faster (~100 Hz for the algorithm v. minutes per cell for a manual analyst), making the algorithmic approach practical for new automated imaging systems that

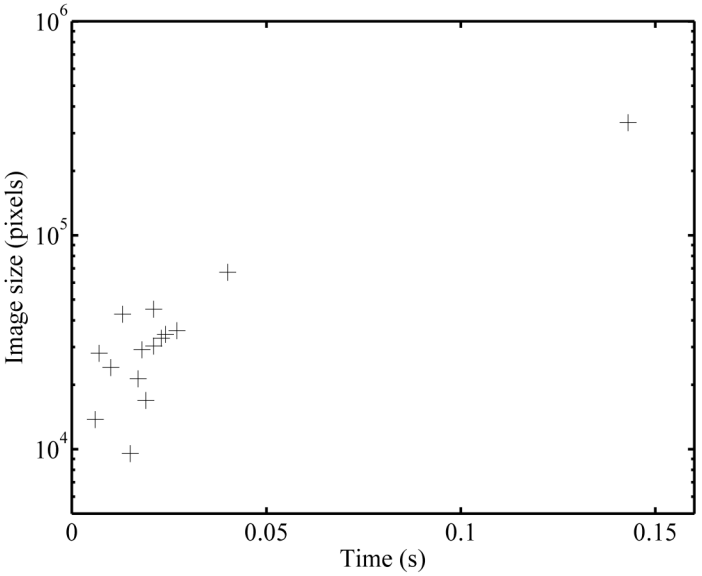

Fig. 12. Computing time for 15 images. Time refers to the amount of time spent in the distance-map or Sieracki algorithm (i.e., it excludes any pre-processing of images).

typically generate such large quantities of data (i.e., > 100 images per min) that manual analysis is prohibitive. As such, this new algorithm represents an important milestone toward automated analysis related to identification and characterization of plankton, for which there is increasing demand (Benfield et al. 2007; Culverhouse et al. 2006; Macleod et al. 2010; Sieracki et al. 2009).

Traditional manual microscopy does offer the potential advantage of being able to manipulate organisms to a better viewing position (Carpentier et al. 1999) or to see plankton in different orientations and thus infer the dimension that may be obscured in another view (e.g., the pervalvar dimension for diatoms that have settled valve-side up). Microscopy techniques that include adjusting focus from the upper to the lower face can also provide information about three-dimensional shape and dimensions. The two-dimensional images collected by systems such as Imaging FlowCytobot rarely contain information regarding that obscured dimension. In this situation, assuming radial symmetry along appropriate axes remains the most appropriate approach. Previous automated analysis based on this assumption was limited to convex, single-axis shapes (Sieracki et al. 1989). A major strength of our new distance map algorithm is that it enables automated application of this symmetry principal to a much wider range of plankton forms. For cells whose shapes can be described as, for example, spherical, cylindrical, or spheroidal/prolate ellipsoidal or any combinations of these, this assumption produces results that are mathematically equivalent to the geometrical formulae. Genera that fall into this category include *Leptocylindrus*, *Dictyocha*, *Guinardia*, and *Corethron* (Hillebrand et al. 1999; Round et al. 1990). However, for shapes in which the length in the obscured dimension differs significantly from those in the viewing plane, the symmetry assumption is flawed, and the errors associated with biovolume estimates are a function of the length in that obscured dimension. Examples of such shapes include gomphonemoids, discs, and elliptic prisms (shapes after Hillebrand et al. 1999). Genera that exhibit these types of shapes include *Chaetoceros*, *Coscinodiscus*, and *Licmophora* (Hillebrand et al. 1999; Round et al. 1990).

For algal shapes that fall into this second category, a further level of analysis is required to derive accurate biovolumes. It is not realistic for a computer algorithm to always lead to accurate inference about a dimension that is not sampled in an image. For instance, a disk-like *Thalassiosira* cell seen from the valve view and a sphere appear similarly circular in a two-dimensional boundary image. In future implementations, it may be possible to help constrain probable shape and size in the obscured dimension by adding taxonomic information, such as could be derived with automated image classification (Sosik and Olson 2007). With this information, taxon-specific correction factors can be applied. We noted that constraining geometric corrections will be more complex and error prone if view orientation is more variable than it is in systems such as Imaging FlowCytobot that measure cells along their most hydrodynamically resistant plane. A detailed understanding of which view of the organism is being seen for other imaging systems may be required before more explicit corrections can be reliably applied.

Comments and recommendations

With any biovolume calculation from two-dimensional measurements, simplifying assumptions must be made about the third or missing dimension. With this type of methodology (i.e., using images to derive volume), we are limited by having no access to one dimension except through published relations or assumptions about symmetry. The distance map algorithm provides a basis for volume calculations with the same simplicity of assumptions and lack of need for taxonomic knowledge as the Sieracki et al. (1989) solid of revolution approach, but it is applicable for a much wider range of morphologies. Furthermore, this framework can be extended by adding taxon-specific correction factors to improve accuracy for species whose morphologies differ significantly from the assumptions. Naturally, results from distance map volume estimation are only as good as the original image and image processing applied to derive target boundaries. If a boundary does not represent the dimensions of an organism, the algorithm will propagate those errors; thus accurate, consistent image capture and processing is a critical prior step. We expect the distance map algorithm to be especially useful in studies that require biovolume estimates from a large number of images and where taxonomic information may be limited.

Finally, we note that best results are likely to be obtained with images for which the imaging plane contains the largest cross-section of the target, as that is the plane that contains the most information about overall shape. Applying this type of

Appendix Table 1. Pseudo-code and code for MATLAB function

1. Calculate ratio of convex area to target area for the image.
2. Apply Sieracki et al. (1989) method to images that have a ratio below 1.2. Apply the following code to all other images.
3. For each image pixel: Calculate the Euclidean distances to every boundary pixel. Assign the minimum of this set of distances and to the pixel location in the distance map matrix.
4. Add 1 to every value in the distance map matrix.
5. Mask the distance map image such that any pixel outside the target boundaries is assigned a value of zero.
6. Calculate the mean of the non-zero values of D.
7. Calculate the representative distance x_r with Eq. 12.
8. Calculate corrected volume with Eq. 11.
9. Multiply by taxon specific correction factor, if applicable.
10. Convert volume in pixel cubes to physical dimensions with user-defined and calibrated conversion.

MATLAB function for pseudo-code steps 3-8

```
function [volume] = distmap_volume(boundary_image)
% [volume] = distmap_volume(boundary_image)
% distmap_volume(boundary_image) returns the volume in pixel cubes
% of a target with complex closed boundary shape indicated in the
% black and white 2D boundary_image (boundary pixels set to 1)
%
% volume is derived according to the distance map approach
% of Moberg and Sosik (2012), Limnology and Oceanography: Methods
%
% Emily A. Moberg and Heidi M. Sosik
% Woods Hole Oceanographic Institution, 2012
% calculate distance map
dist = bwdist(boundary_image);
dist = dist + 1;
% mask distance map image (all distances outside boundary set to NaN)
image_fill = imfill(boundary_image,'holes');
dist(image_fill = = 0) = NaN;
% find representative transect length
x = 4*nanmean(dist(:)) - 2;
% define cross-section correction factors
% pyramidal cross-section to interpolated diamond
c1 = (x^2)/(x^2 + 2*x + 1/2);
% diamond to circumscribing circle
c2 = pi/2;
% calculate final volume applying correction factors to distance map
volume = c1*c2*2*nansum(dist(:));
end
```

methodology to other viewing planes may result in much smaller volumes' being calculated. For microscopy involving cells settled on a surface or flow systems with geometry comparable to Imaging FlowCytobot (Olson and Sosik 2007), the condition of viewing the largest cross-section is typically met and the distance map algorithm provides capability to estimate biovolume rapidly and accurately for a wide range of organisms.

References

Albertano, P., D. D. Somma, and E. Capucci. 1997. Cynaobacterial picoplankton from the Central Baltic Sea: cell size classification by image-analyzed fluorescence microscopy. J. Plankton Res. 19:1405-1416 [doi:10.1093/plankt/19.10.1405].

Benfield, M. C., and others. 2007. RAPID: research on automated plankton identification. Oceanography 20:172-187 [doi:10.5670/oceanog.2007.63].

Carpentier, C., H. Ketelaars, A. Wagenvoort, and K. Pikaar-Schoonen. 1999. Rapid and versatile routine measurements of plankton biovolumes with BACCHUS. J. Plankton Res. 21:1877-1889 [doi:10.1093/plankt/21.10.1877].

Chen, D. Q., A. A. Farag, R. L. Falk, and G. W. Dryden. 2009. A variational framework for 3D colonic polyp visualization

in virtual colonoscopy. 2009 16th IEEE International Conference on Image Processing, Vols 1-6:2589-2592.

Congestri, R., R. Federici, and P. Albertano. 2000. Evaluating biomass of Baltic filamentous cyanobacteria by image analysis. Aquat. Microb. Ecol. 22:283-290 [doi:10.3354/ame022283].

Culverhouse, P. F., and others. 2006. Automatic image analysis of plankton: future perspectives. Mar. Ecol. Prog. Ser. 312:297-309 [doi:10.3354/meps312297].

Estep, K. W., F. Macintyre, E. Hjoerleifsson, and J. M. Sieburth. 1986. MacImage: a user-friendly image-analysis system for the accurate mensuration of marine organisms. Mar. Ecol. Prog. Ser. 33:243-253 [doi:10.3354/meps033243].

Hillebrand, H., C. D. Dürselen, D. Kirschtel, U. Pollingher, and T. Zohary. 1999. Biovolume calculation for pelagic and benthic microalgae. J. Phycol. 35:403-424 [doi:10.1046/j.1529-8817.1999.3520403.x].

Krambeck, C., H.-J. Krambeck, and J. Overbeck. 1981. Microcomputer-assisted biomass determination of plankton bacteria on scanning electron micrographs. Appl. Environ. Microbiol. 42:142-149.

Macleod, N., M. Benfield, and P. Culverhouse. 2010. Time to automate identification. Nature 467:154-155 [doi:10.1038/467154a].

Nemeth, G., P. Kardos, and K. Palagyi. 2010. Topology preserving 3D thinning algorithms using four and eight subfields. Image Analysis and Recognition Pt I, Proceedings 6111:316-325.

Olson, R. J., and H. M. Sosik. 2007. A submersible imaging-in-flow instrument to analyze nano- and microplankton: Imaging FlowCytobot. Limnol. Oceanogr. Methods 5:195-203 [doi:10.4319/lom.2007.5.195].

Round, F. E., R. M. Crawford, and D. G. Mann. 1990. The diatoms: biology and morphology of the genera. Cambridge Univ. Press.

Sieracki, C. K., M. E. Sieracki, and C. S. Yentsch. 1998. An imaging-in-flow system for automated analysis of marine microplankton. Mar. Ecol. Prog. Ser. 168:285-296 [doi:10.3354/meps168285].

Sieracki, M. E., C. L. Viles, and K. L. Webb. 1989. Algorithm to estimate cell biovolume using image analyzed microscopy. Cytometry 10:551-557 [doi:10.1002/cyto.990100510].

———, and others. 2009. Optical plankton imaging and analysis systems for ocean observation. OceanObs'09.

Sosik, H. M., and R. J. Olson. 2007. Automated taxonomic classification of phytoplankton sampled with imaging-in-flow cytometry. Limnol. Oceanogr. Methods 5:204-216 [doi:10.4319/lom.2007.5.204].

Sun, J., and D. Liu. 2003. Geometric models for calculating cell biovolume and surface area for phytoplankton. J. Plankton Res. 25:1331-1346 [doi:10.1093/plankt/fbg096].

Submitted 6 October 2011
Revised 28 February 2012
Accepted 6 March 2012