USF Tampa Graduate Theses and Dissertations

USF Graduate Theses and Dissertations

October 2018

Application of Image Recognition Technology to Foraminiferal Assemblage Analyses

Christian Helmut Gfatter
University of South Florida, cgfatter@mail.usf.edu

Follow this and additional works at: https://digitalcommons.usf.edu/etd

Part of the Geology Commons

Application of Image Recognition Technology to Foraminiferal Assemblage Analyses

by

Christian Helmut Gfatter

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science
College of Marine Science
University of South Florida

Major Professor: Pamela Hallock Muller, Ph.D.
Albert C. Hine, Ph.D.
Ryan P. Moyer, Ph.D.

Date of Approval:
October 10, 2018

Keywords: benthic, classification, FORAM Index, identification, VisualSpreadsheet™

Acknowledgments

Table of Contents

List of Tables

List of Figures

Abstract

Analyses of foraminiferal assemblages involve time-consuming microscopic assessment of sediment samples. Image recognition software, which systematically matches features within sample images against an image library, is widely used in contexts ranging from law enforcement to medical research. At present, scientific applications such as identification of specimens in plankton samples utilize flow-through systems in which samples are suspended in liquid and pass through a beam of light where the images are captured using transmitted light. Identification of foraminifers generally utilizes reflected light, because most shells are relatively opaque.

My goal was to design and test a protocol to directly image foraminiferal specimens using reflected light and then apply recognition software to those images. A library of high quality digital images was established by photographing foraminifers identified conventionally from sediment samples from the west Florida shelf. Recognition software, VisualSpreadsheet™ by Fluid Imaging Technologies, Inc., was then trained to improve automated assemblage counts and those results were compared to results from direct visual assessment. The auto-classification feature produced composite accuracies of foraminiferal groups in the range of 60–70% compared to traditional visual identification by a researcher using a stereo-microscope. Site SC34, the source of images for the original image library, had an initial accuracy of 75% that was improved slightly through an alteration to one of the software classes, but composite accuracy plateaued at

60% with the updated filters. Thus, image acquisition advancements and further development of image recognition software will be required to improve automated or semi-automated foraminiferal classifications. However, other potential applications were noted. For example, an advantage of acquiring digital images of entire samples or subsamples is the ability to collect quantitative data such as diameter and length, allowing size-frequency assessments of foraminiferal populations while possibly automating grain-size analyses without requiring separate processing. In addition, data files of library and sample specimens can be readily shared with other researchers.

1. Introduction

Since the advent of computers, untold kinds of arduous calculations and laborious tasks have been automated. Computer scientists understood that anything that could be transformed into a digital format of zeros and ones could be managed and manipulated. Early limitations with respect to graphical representations primarily revolved around file size and computational speed. The performance of a computer's underlying hardware, namely improvements in microprocessor speed and increases in random-access memory, along with greater storage capacity, created explosive growth in the industry with an enormous rise in computer applications. Hardware is constantly being improved, while the writing and adaptations of computer software often limits potential applications. However, applications that have been developed by the private sector or during government-sponsored research can often be adapted for uses in other areas. Thus, opportunities abound to adapt existing technologies and software to advance scientific studies.

Scientists in a wide range of fields envision applications for computer technology and often generate proof-of-concept papers and prototypes in an effort to advance their work or field. Foraminiferal assemblage analysis is a scientific application that has been recognized for its potential to benefit from the adaptation of image-analysis technology. In 1994, an image-analysis based system for the classification of planktic foraminifers was prototyped (Liu et al. 1994, Yu et al. 1996). Because the number of planktic foraminiferal species is fairly limited, such an application seemed viable. However, limitations such as the need for multiple images of the

same specimen, specific algorithms for this particular set of organisms, as well as programming time and effort, impeded the advancement of this application. Other limitations of this approach, including the use of scanning electron microcopy (SEM) and the use of pristine specimens, were noted by Ranaweera et al. (2009a).

Benthic foraminiferal species are much more numerous and diverse than planktic species, so an approach similar to that of Liu et al. (1994) and Yu et al. (1996) would require enormous resources. Moreover, most specimens in a typical sediment sample are imperfect (Álvarez et al. 2014). Computer-assisted identification keys are already available, but they are an aid to direct visual identification. After picking foraminifers from sediment samples, the specimens must be visually identified by researchers. The identification process is subjective and requires time and skill (Ranaweera et al. 2009b), but certain applications such as the FORAM Index (Hallock et al. 2003, Hallock 2012) may not require species-level identification.

Fluid Imaging Technologies, Inc. has developed an instrument, the FlowCam® VS Series and recently introduced 8000 Series, which integrates the use of digital imaging technology in combination with image recognition software to identify and analyze microscopic particles in a suspended fluid. Introduced in 1999, the device is used to identify plankton, sometimes to the species-level, and its software, VisualSpreadsheet™, collects over 30 parameters on each particle imaged (Fluid Imaging Technologies 2017). Particle analyses are generally performed on translucent organisms, as the camera uses back lighting to capture digital images that are of sufficient resolution for the software to analyze.

Unfortunately, opaque particles such as foraminiferal tests pose identification challenges for the FlowCam due to the lack of front illumination. The methodology to bypass the apparatus

and use Audio Video Interleave (AVI) files to utilize the VisualSpreadsheet software had been developed by researchers at the University of Colorado on static images of diatoms (as discussed with Mr. Nelson, personal communication 2016). Therefore, to take advantage of the image recognition capabilities of VisualSpreadsheet, I needed to acquire digital images of benthic foraminifers. This was accomplished using a microscope camera. Then, after the static images were collected, the images were converted to the AVI multimedia container file format so VisualSpreadsheet could process them. My goal was to address technical issues posed by foraminiferal tests and to assess the precision and accuracy of the software when applied to foraminiferal assemblage analyses.

There were five main objectives of this study:

1) test the precision of the software in its ability to identify the same particles in repeated test runs;

2) assess the "training" aspect of the software's image recognition capability by determining the number of runs needing supplementary quality control to acquire a stable level of accuracy in identifying known foraminiferal specimens, and at what degree of accuracy [e.g., morphotype (milioline, agglutinate, or hyaline), family, genus, or species];

3) compare "trained" software identifications of foraminiferal assemblage samples to the same samples identified by a researcher;

4) test whether trained software can identify foraminiferal specimens directly from sediment samples, and with what degrees of accuracy and precision; and

3

5) determine which metrics provided by the software best facilitate identification of the

foraminiferal taxa and if custom metrics can increase accuracy.

2. Methods

An outline of the approach is provided in Table 2.1. Components of the methodology along with recommendations are discussed in further detail in the following sections.

Table 2.1: Procedure for utilizing VisualSpreadsheet software to support foraminiferal assemblage analyses

1)	**Catalog data**
	a) Collect, ID, and photograph specimens to be used for the software libraries. Existing software libraries may be used if they pertain to the analysis.
	b) Collect, ID, and photograph specimens.
2)	**Prepare data for use by VisualSpreadsheet**
	a) Convert digital photographs of a single collection (e.g., specimens of a sample site, all specimens of a study group to be measured, etc.) to an AVI file.
	b) Determine optimal context.
	c) Generate a list file for each set of data.
	d) Refine list file by deleting undesirable images.
3)	**Set up software libraries**
	a) Define master libraries (e.g., of a species, genus, group, etc.) and keep a copy of core master libraries.
	b) Define more constrained libraries (e.g., species of a certain size).
	c) Create filters based on libraries of interest.
4)	**Classify the data**
	a) Define a classification and build classes with filters (using any list file).
	b) Execute auto classification and then refine classification by adding, deleting, modifying, and/or reordering classes. Create and modify filters as needed. Repeat until best classification is obtained.
	c) Define classification template.
	d) For each list file, define a classification based on a template, apply the classification, make adjustments (e.g., refine results, update a library, etc.), and save it.
5)	**Export the data for further analyses**
	a) Data from a list file, classification data, class data, and data from a library may be exported.

2.1. Imaging Technology

Digital images of foraminifers were taken using a Zeiss Stemi 2000-C microscope equipped with a Zeiss AxioCam MRc5 camera along with its AxioVision software, versions 4.4 and 4.8. The microscope was set to a magnification of 40x for all images. Images of larger specimens were also taken at a magnification of 20x for later size assessments. A Zeiss KL 1500 LCD with a bifurcated light guide was initially set to 3000K and the second lowest brightness setting. Illumination of specimens to be photographed was made by adjusting the light guides, with occasional minor adjustments made using the aperture control.

Digital images were saved in the Tagged Image File Format (TIF or TIFF) because such files store the digital data in a lossless compression format. This is preferred for archiving purposes as images can be reconstructed to their full, original quality, and no loss of information occurs when lossless files are saved, converted, or manipulated. The software was calibrated once and the white balance was set using the interactive feature with a white color card. Other settings are noted in Table 2.2.

Table 2.2: AxioVision software settings for the AxioCam MRc5 camera

Setting	Parameter	Notes
90.0 ms	Exposure time	Adjust tab, Exposure
150 %	Exposure percentage	Adjust tab, Exposure
1.14	Cyan 🔵 ◀ ▶ 🔴 Red	Adjust tab, White Balance, ☑ Show Channels
1.12	Magenta 🟣 ◀ ▶ 🟢 Green	Adjust tab, White Balance, ☑ Show Channels
0.74	Yellow 🟡 ◀ ▶ 🔵 Blue	Adjust tab, White Balance, ☑ Show Channels
0.6	Saturation	Adjust tab, Color Saturation
● RGB	Camera mode radio button	Frame tab
2584 x 1936 quality standard color	Camera mode dropdown	Frame tab
2584 x 1936	Frame size	Frame tab
Rotated 180 deg	Image Orientation dropdown	General tab, Image Orientation
☑ checked	Enhance Color	General tab, Filter Operations
☐ unchecked	Convert to 8-bit	When saving image as TIF file
☑ checked	Apply display mappings	When saving image as TIF file
0	Compression percentage	When saving image as TIF file

The source image files were then collected into Audio Video Interleave (AVI) multimedia container file format so that the images could be processed by VisualSpreadsheet™. This was accomplished using the image-manipulation program, VideoPad. Another image processing program, ImageJ, was tested, but the desired results for VisualSpreadsheet were not obtained. The general process was to isolate the foraminiferal specimens from a sample, identify and photograph them, convert a set of images to an AVI file, process an AVI file for foraminiferal images, and then review the resulting set of data.

2.1.1. Audio Video Interleave File Frame Rate for VisualSpreadsheet

When creating Audio Video Interleave (AVI) multimedia container files, several settings can be applied. After creating a number of preliminary AVI files and then processing them in

VisualSpreadsheet, I found that settings pertaining to Frame Rate were essential to controlling the output from VisualSpreadsheet. The Frame Rate should be set to 10 Frames per Second (FPS). Alternative FPS values caused VisualSpreadsheet to provide repetitive images of the same specimen. Although manual processing to remove redundant images from the output can be performed, unnecessary work can be avoided by adhering to the aforementioned 10 FPS recommendation. Constant frame rate should also be selected when provided as an option.

2.1.2. VisualSpreadsheet Setup and Context Settings

Two basic context files, one for images at 40x magnification and another for images at 20x magnification, were prepared from images containing scale bars to establish the ratio of micrometers (µm) per pixel. As the majority of images were at 40x magnification, this associated file served as the starting context file for the work herein and the other was reserved for possible future work with images at 20x magnification. The primary settings that were adjusted are found in the "Capture" tab and the "Filter" tab of the "Context" dialog box (Figure 2.1).
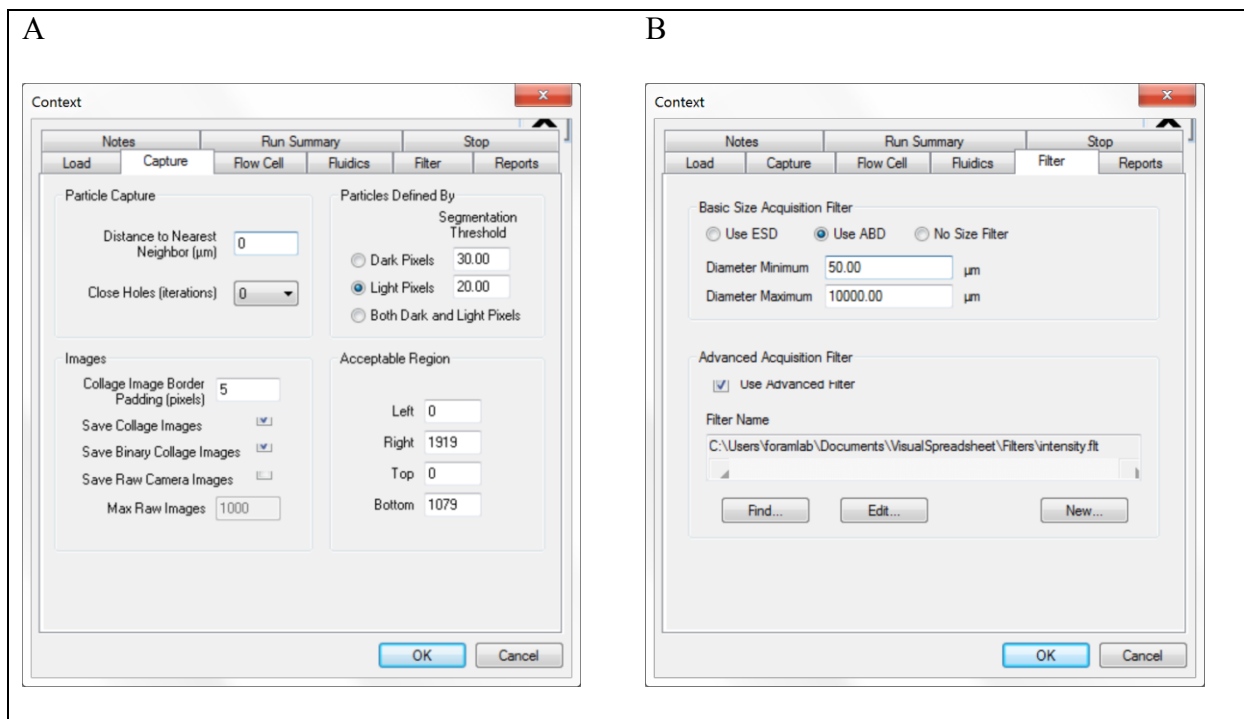
Figure 2.1: VisualSpreadsheet software Context dialog box for **A**) Capture tab and **B**) Filter tab

When processing images, the VisualSpreadsheet software defines a particle by evaluating the differences in contrast the particle has in comparison to the background on which the particle resides. The controls that govern contrast are referred to as "Segmentation Threshold" (ST). The initial library-sample AVI was processed numerous times using different Context settings, including variations to ST, to determine which parameters and settings would best capture the foraminiferal images. "Light Pixels" in the "Particles Defined By" section was the most prominent. Settings for "Dark Pixels" reduced the desired output when using "Both Dark and Light Pixels". When used alone, settings for "Dark Pixels" did not produce meaningful output because the particles were defined by reflected light. Although higher ST values for "Light Pixels" allowed the software to capture nearly all of the 637 specimen images in the AVI file, the

value of 20 increased the captured specimen area with minimal obfuscation of the result set from the desired ideal.

As displayed in Figure 2.2, VisualSpreadsheet captures an initial set of boundaries for each particle and then refines it based on algorithms that utilize the Context parameters. If the "Light Pixels" parameter is too low, initial particle boundaries may be too large and the outline may expand into areas beyond the particle because the differential between light and dark pixels is insufficient. On the other hand, if the "Light Pixels" parameter is too high, portions of a certain particle may not be captured by the software. Figure 2.3 illustrates the
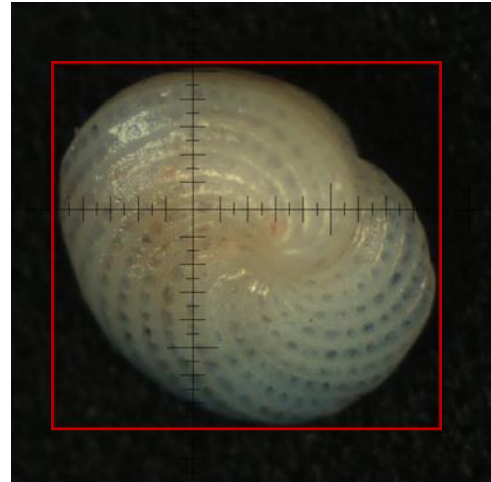
Figure 2.2: Determination of initial boundaries (shown in red) by VisualSpreadsheet (of SC34a60-2.tif during file processing of LibSC34.avi)

influence of a "Light Pixels" setting that is too low (left), the favored value (center), and one that is too high (right). Note that additional parameters affect the quantity and quality of output and may impact the effectiveness of other chosen settings, including "Light Pixels". Nonetheless, the ST value of 20 for "Light Pixels" provided the best results, even after integrating variations to additional parameters.
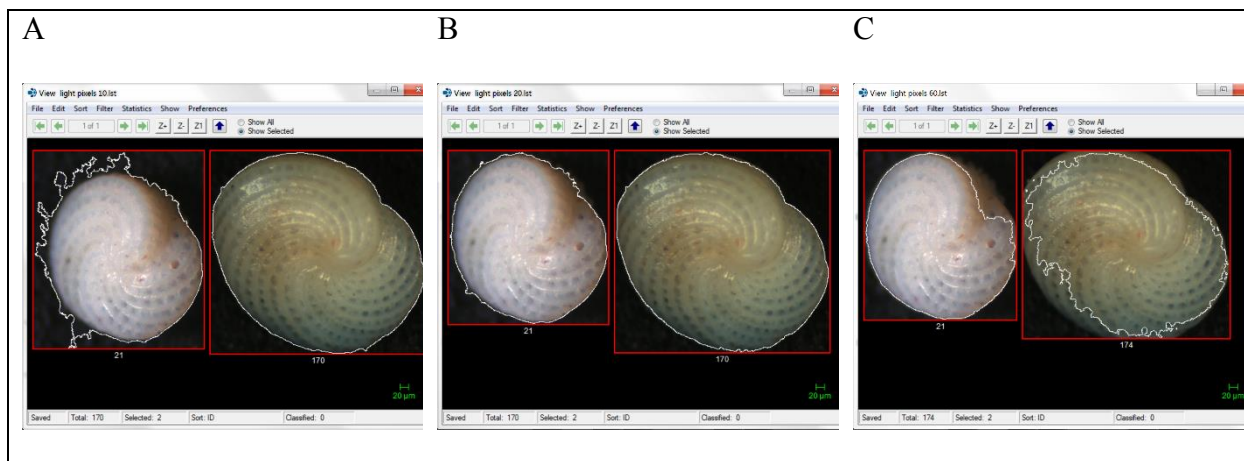
Figure 2.3: Effect on particle boundaries when altering Segmentation Threshold of Light Pixels **A)** set to 10 (too low), **B)** set to 20 (optimal), and **C)** set to 60 (too high)

"Distance to Nearest Neighbor" was set to 0 µm because only one specimen per image was in the AVI file created from the library sample. This parameter may need to be adjusted when processing AVI files created from samples containing more than one specimen per image. Within this study, this parameter setting provided satisfactory results and was not changed. Context files can be saved, so specific settings can be applied as needed.

In this study, size restrictions of specimens were primarily addressed by the "Basic Size Acquisition Filter", specifically by setting the "Diameter Minimum" to 50 µm. Images of particles less than the threshold of 50 µm were discarded, deemed unlikely to be foraminifers or too small to consider. An "Advanced Size Acquisition Filter" can also be customized to exploit one or more available fields to constrain the software's output. This feature may be particularly useful if limits applied to these fields remove unwanted results. After examining some preliminary output, I determined that particles with low Intensity values were not foraminifers. The Intensity value ranges from 0 to 255, so a minimum (50) and a maximum (255) were specified to reject additional unwanted particles. The VisualSpreadsheet settings utilized on the

11

library sample are noted in Table 2.3 along with the settings applied for the "Advanced Size Acquisition Filter", highlighted in gray as they were considered supplementary.

Table 2.3: VisualSpreadsheet software Context settings

Setting	Parameter	Notes
0 μm	Distance to Nearest Neighbor (μm)	Capture tab, Particle Capture
0	Close Holes (iterations)	Capture tab, Particle Capture
● Light Pixels	radio button	Capture tab, Particles Defined By
20	Segmentation Threshold *for Light Pixels*	Capture tab, Particles Defined By
● Use ABD		Filter tab, Basic Size Acquisition Filter
50 μm	Diameter Minimum	Filter tab, Basic Size Acquisition Filter
10000 μm	Diameter Maximum	Filter tab, Basic Size Acquisition Filter
☑ checked	Use Advanced Filter	Filter tab, Advanced Size Acquisition Filter
Intensity	Currently Using *for Available Fields*	Filter tab, Edit…
50	Filter Min	Filter tab, Edit…
255	Filter Max	Filter tab, Edit…

2.2. Image Library

Foraminiferal assemblage data used to assess the practicality, precision, and accuracy of the software were from sediment samples collected on the Florida Springs Coast by Dr. Carlson and others of the Florida Fish and Wildlife Research Institute (FWRI) in summer 2013. Samples were processed using standard methods for assessment of foraminiferal assemblages in coastal samples (e.g., Hallock et al. 2003, Carnahan et al. 2009). Foraminifers were identified to the genus-level or species using descriptions and plates in Bock et al. (1971) and Poag (2015). Data on sediment grain size was determined using standard sieving procedures described by Folk

(1980). Software accuracy to identify foraminiferal categories was compared to grain-size statistics of the source sample sites to determine the effect of sediment texture.

Images of foraminifers, which were picked and identified from specific sites with a diversity of unique species, were used to construct the initial image library of foraminifers found in the Springs Coast region. Additional images were added for species not represented in the initial samples. Each image was individually focused and several orientations of each specimen were photographed. For libraries created to optimize phytoplankton classification, Camoying and Yniguez (2016) recommended that 10–15 characteristic images be taken. Two to four images of each benthic foraminifer picked from the selected sites, and data on these library images, were compiled to assist in the assessment of the software. The source specimens for the library were glued one per cell on a labeled micropaleontological faunal slide for future reference (Figure 2.4).
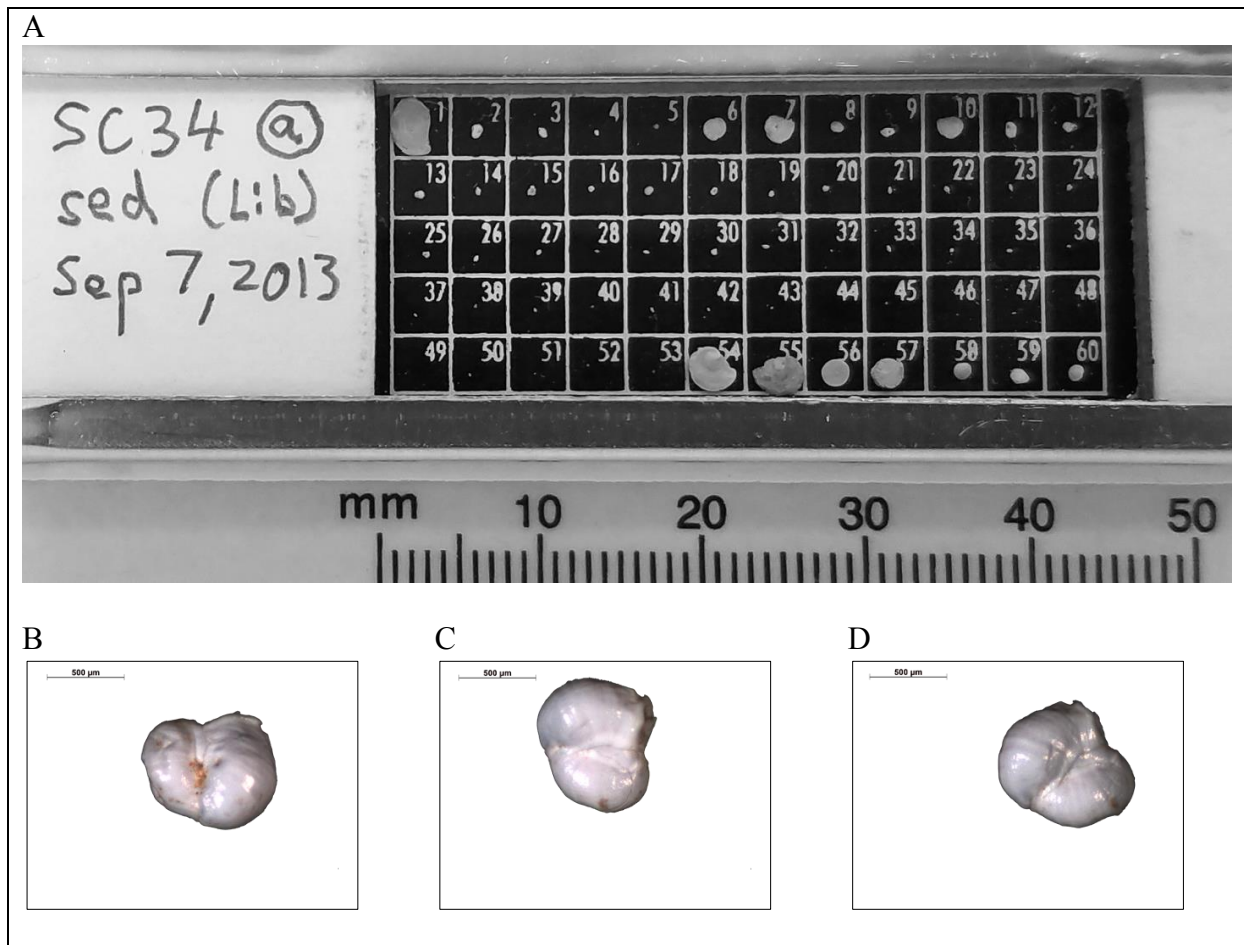
Figure 2.4: **A**) Micropaleontological faunal slide of 60 foraminifers stored for the library;
B-D) Three images of a single foraminifer stored on a micropaleontological faunal slide

Images of the library specimens were compiled into an AVI file, then loaded into the VisualSpreadsheet software application. Since each frame was processed for multiple images, all images captured by the software were quality controlled by rejecting non-foraminiferal images. Foraminiferal images were then identified based on visual identification performed prior to photography. Order, Family, Genus and Species were assigned to each foraminiferal image, as was functional group (e.g., A: symbiont bearing / larger Miliolida, B: smaller Miliolida, C: Rotaliida and other, and D: stress tolerant), specific shape (e.g., globular, linear, planispiral, or

14

trochospiral), and surface texture (e.g., agglutinated, coarse, smooth, ornamented, porcelaneous, hyaline / semi-transparent) categories.

The library used by the software to identify images grew as additional images from the sample sites were processed and added to the library. As the library was appended, the sites that contributed images to that version of the library were noted.

2.2.1. Master Image Library

A master library with a collection of images was created for each species identified in the library sample, such as images of *Archaias angulatus* (Fichtel & Moll, 1798). The list of foraminiferal species, with authors, is presented in Appendix A. The primary factors used by VisualSpreadsheet to distinguish particles are derivations of particle size (e.g., area, diameter, length, width, and so forth). Libraries with a subset of images can further delineate uniformity and uniqueness within a collection (e.g., small *A. angulatus* specimens, medium *A. angulatus* specimens, and large *A. angulatus* specimens) and were derived from the associated master library when favorable. Examples of *A. angulatus* libraries are shown in Figure B.1 of Appendix B: VisualSpreadsheet Image Libraries.

2.2.2. Group Image Library

A group library is a collection of associated images created to represent a foraminiferal category (e.g., all images of specimens categorized as Large Benthic Foraminifers). As previously noted, the software used factors based on size to distinguish particles. Therefore, a master group library was also subdivided as needed into several size-based libraries.

2.3. Sample Analysis

Images of foraminifers to be analyzed using the Springs Coast library were taken from onshore to offshore transects comprised of three to eight samples (Table 2.4). The standard method was to pick 200 or more foraminiferal specimens from each sample as described by Hallock et al. (2003). The protocol for Springs Coast samples that includes the image acquisition step is found in Appendix C. The target quantity (i.e., 200 or more) was chosen to provide a buffer for VisualSpreadsheet so 150–200 foraminiferal images per sample could be captured. One or more specimens of a similar size were then glued in a cell of a micropaleontological faunal slide, identified, and photographed. The area of a single cell used for multiple specimens did not exceed the amount of space needed to take a single image of that cell. All photographs were taken as previously described in the Image Library section, also at a magnification of 40x, with additional images at a lower magnification for specimens that were larger than the standard image area. Each image of one or more specimens was individually focused and a single orientation was photographed. Specimens from a specific sample site were stored together on labeled micropaleontological faunal slides for future reference (Figure 2.5). The label noted the sample site code, a letter designation with "a" being the first slide in the sequence, and the date the sample was taken.

16
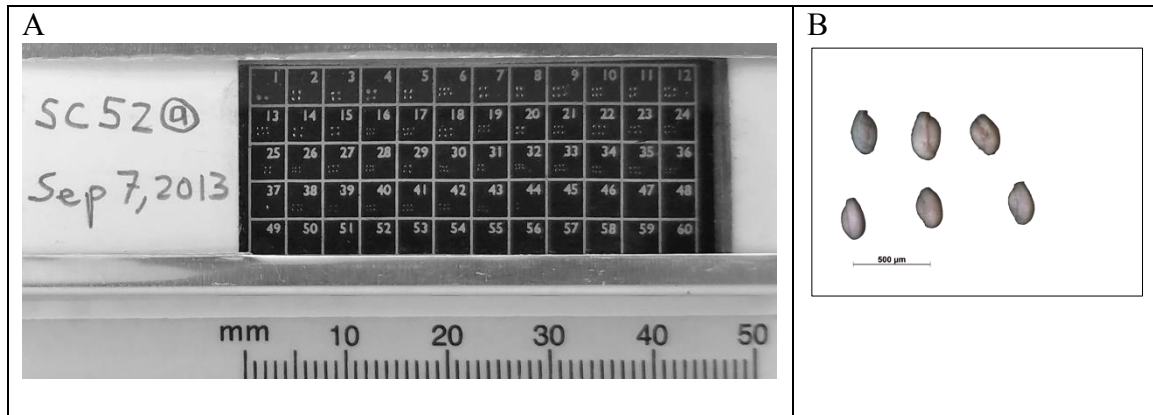
Figure 2.5: **A**) Micropaleontological faunal slide of specimens stored from a sample site; **B**) Enlarged view of a single cell with multiple specimens

One AVI file per sample site was generated from the images, and subsequently loaded into the VisualSpreadsheet software application. Because each frame was processed for multiple images, all images captured by the software for each sample site were quality controlled by rejecting non-foraminiferal images. In addition to superfluous images, some foraminiferal images were not captured by the software.

Sample sites listed in Table 2.4 were used for this study.

Table 2.4: Initial Springs Coast sample sites

Site ID	Transect #	Latitude	Longitude
SC034	1	28.321	-82.780
SC040	1	28.316	-82.803
SC046	1	28.324	-82.858
SC052	1	28.322	-82.886
SC088	2	28.491	-82.723
SC094	2	28.503	-82.775
SC100	2	28.491	-82.811
SC106	2	28.497	-82.865
SC135	3	28.655	-82.701
SC141	3	28.664	-82.742
SC147	3	28.662	-82.778
SC159	3	28.664	-82.860
SC165	3	28.677	-82.920
SC177	3	28.674	-82.997

Several site samples were treated with trichloroethylene as described by Murray (2006) to separate lighter foraminiferal specimens from a majority of the heavier sediment via floating in the dense trichloroethylene bath. The lighter particles, including foraminifers, were separated from the trichloroethylene via repeated decanting through a filter paper. The filter paper and the beaker containing the heavier sediment residual were then left to dry. Thereafter, the residue on the filter paper was collected for later analysis and the sediment residual was also examined.

A reference guide of the most abundant foraminiferal genera found in the Springs Coast (using Poag 2015) was used to help validate the results of the image recognition software. A list of common Springs Coast foraminiferal taxa is shown in Table 2.5.

Table 2.5: Foraminiferal categories and taxa

Group	Shape	Surface Texture	Order	Family	*Genus*	*Species*
A	planispiral	porcelaneous	Miliolida	Soritidae	*Archaias*	*angulatus*
	planispiral	porcelaneous	Miliolida	Soritidae	*Broeckina*	*orbitolitoides*
	planispiral	porcelaneous	Miliolida	Peneroplidae	*Laevipeneroplis*	*spp.*
	planispiral	porcelaneous	Miliolida	Peneroplidae	*Peneroplis*	*pertusus*
B	milioline	porcelaneous	Miliolida	Hauerinidae	*Affinetrina*	*planciana*
	milioline	porcelaneous	Miliolida	Hauerinidae	*Articulina*	*pacifica*
	planispiral	porcelaneous	Miliolida	Cornuspiridae	*Cornuspira*	*spp.*
	milioline	porcelaneous	Miliolida	Hauerinidae	*Cycloforina*	*sidebottomi*
	milioline	porcelaneous	Miliolida	Hauerinidae	*Flintinoides*	*labiosa*
	milioline	porcelaneous	Miliolida	Hauerinidae	*Miliolinella*	*spp.*
	milioline	porcelaneous	Miliolida	Hauerinidae	*Pseudotriloculina*	*spp.*
	milioline	porcelaneous	Miliolida	Hauerinidae	*Pyrgo*	*spp.*
	milioline	porcelaneous	Miliolida	Hauerinidae	*Quinqueloculina*	*spp.*
	milioline	porcelaneous	Miliolida	Spiroloculinidae	*Spiroloculina*	*spp.*
	milioline	porcelaneous	Miliolida	Hauerinidae	*Triloculina*	*spp.*
	milioline	porcelaneous	Miliolida	Hauerinidae	*Vertebrasigmoilina*	*mexicana*
	milioline	porcelaneous	Miliolida	Fischerinidae	*Wiesnerella*	*auriculata*
C Rotaliida						
	trochospiral	hyaline	Rotaliida	Cibicididae	*Cibicides*	*spp.*
	trochospiral	hyaline	Rotaliida	Anomalinidae	*Cibicidoides*	*spp.*
	trochospiral	hyaline	Rotaliida	Discorbidae	*Discorbis*	*spp.*
	trochospiral	hyaline	Rotaliida	Eponididae	*Eponides*	*spp.*
	linear	hyaline	Rotaliida	Bolivinitidae	*Fursenkoina*	*mexicana*
	planispiral	hyaline	Rotaliida	Epistomariidae	*Palmerinella*	*palmerae*
	trochospiral	hyaline	Rotaliida	Glabratellidae	*Planoglabratella*	*spp.*
	discoid	hyaline	Rotaliida	Planorbulinidae	*Planorbulina*	*mediterranensis*
	trochospiral	hyaline	Rotaliida	Rosalinidae	*Rosalina*	*spp.*
Agglutinates	linear	agglutinated	Lituolida	Prolixoplectidae	*Karrerulina*	*apicularis*
	linear	agglutinated	Textulariida	Textulariidae	*Textularia*	*spp.*
D	trochospiral	hyaline	Rotaliida	Rotaliidae	*Ammonia*	*spp.*
	planispiral	hyaline	Rotaliida	Nonionidae	*Astrononion*	*stelligerum*
	linear	agglutinated	Textulariida	Textulariidae	*Bigenerina*	*spp.*
	linear	hyaline	Rotaliida	Bolivinitidae	*Bolivina*	*spp.*
	planispiral	hyaline	Rotaliida	Elphidiidae	*Cribroelphidium*	*poeyanum*
	planispiral	hyaline	Rotaliida	Elphidiidae	*Elphidium*	*spp.*
	planispiral	hyaline	Rotaliida	Nonionidae	*Haynesina*	*spp.*
	planispiral	hyaline	Rotaliida	Nonionidae	*Nonionoides*	*grateloupii*

3. Data Analysis

To meet the five objectives of this project, the following protocols and analyses were employed:

1) To test the consistency of the software, the means and standard deviations of the prepackaged measurements were collected from runs of the same unshuffled library sample from two different AVI files where images were in the same order, as well as from three runs of AVI files with shuffled images. Results from the tests of the library sample were recorded. The library sample was then processed with the established context settings and retained as a VisualSpreadsheet™ list file. Thereafter, primary image libraries of foraminiferal species, genera, and categories were prepared. The library sample was processed using an initial species library of *Archaias angulatus*, a routinely identifiable and one of the more distinctive foraminifers found in the Springs Coast. This test library contained quality images of small to large specimens where extra-large (i.e., specimen extended beyond boundary) and inferior (e.g., broken or abnormally shaped) specimens were omitted. The library sample was filtered with the test library to help identify the *A. angulatus* specimens. Trials of the value-based and statistically-based filters were repeated several times to determine precision.

2) After a classification by the software was recorded, the percentage correct for a category was calculated by comparing the software's assessment to that of the same specimens identified by a researcher.

3) To assess the "training" capability of the software, the initial auto-classification output by the software for the functional group category was optimized for the first sample site, SC34. The accuracy of assigned taxa was calculated by comparing the number of correct assessments made by the software to the number of specimens classified. The auto classification of functional groups was then performed on all sample sites using the initial configuration, and the composite foraminiferal group accuracy was calculated and recorded. Additional images were then added to various software libraries, new filters were created, and classifications were updated. Then, the auto classification was reevaluated using the revisions. A progression log of each run through the software was kept. A stable level of accuracy for the functional group category was defined as the average of three sequentially recorded percentages where the last one in the group was marked by negligible change in the percentage. This tolerance threshold was initially set at +/- five percentage points with each step. The degree of accuracy was considered to be the lowest level of successful identification.

4) A trichloroethylene-treated sample, where a majority of the sediment had been removed, was used to assess the software's ability to directly distinguish foraminiferal specimens from a sample. Tabulation of the results from auto classification by the software was compared to manual modifications of the classified images along with manual classification of the unclassified ones.

5) The measurements of key metrics provided by VisualSpreadsheet from the list file of library-sample LibSC34 were sorted in ascending and descending order. For each metric, foraminiferal specimens in the lowest and highest deciles were identified, tallied by functional group, and recorded. Results were reported in a heat map format to highlight metrics that could be implemented in filters to assist in the identification process.

4. Results

4.1. Preliminary Results

Data collection of foraminiferal assemblages from the Springs Coast used sediment samples from three transects (Figure 4.1). The procedures of data collection for these sets utilized Rose Bengal staining (Walton 1952, Murray and Bowser 2000) to determine how many foraminifers were alive at the time of sampling. However, after reviewing these data (Table 4.1), the initial sets were reprocessed without Rose Bengal staining and future sets were not stained because there were very few specimens that indicated that they were alive at the time of collection and the stain interfered with color data collected by the software.

Table 4.1: Foraminiferal counts and estimated number of foraminifers alive at collection time from the application of Rose Bengal staining.

Site ID	Transect #	Specimens	"Live" Count	Percentage
SC034	1	203	2	1.0
SC040	1	283	4	1.4
SC046	1	212	18	8.5
SC088	2	276	5	1.8
SC100	2	231	3	1.3
SC135	3	283	1	0.4
SC141	3	271	1	0.4
Total		**1759**	**34**	**1.9**
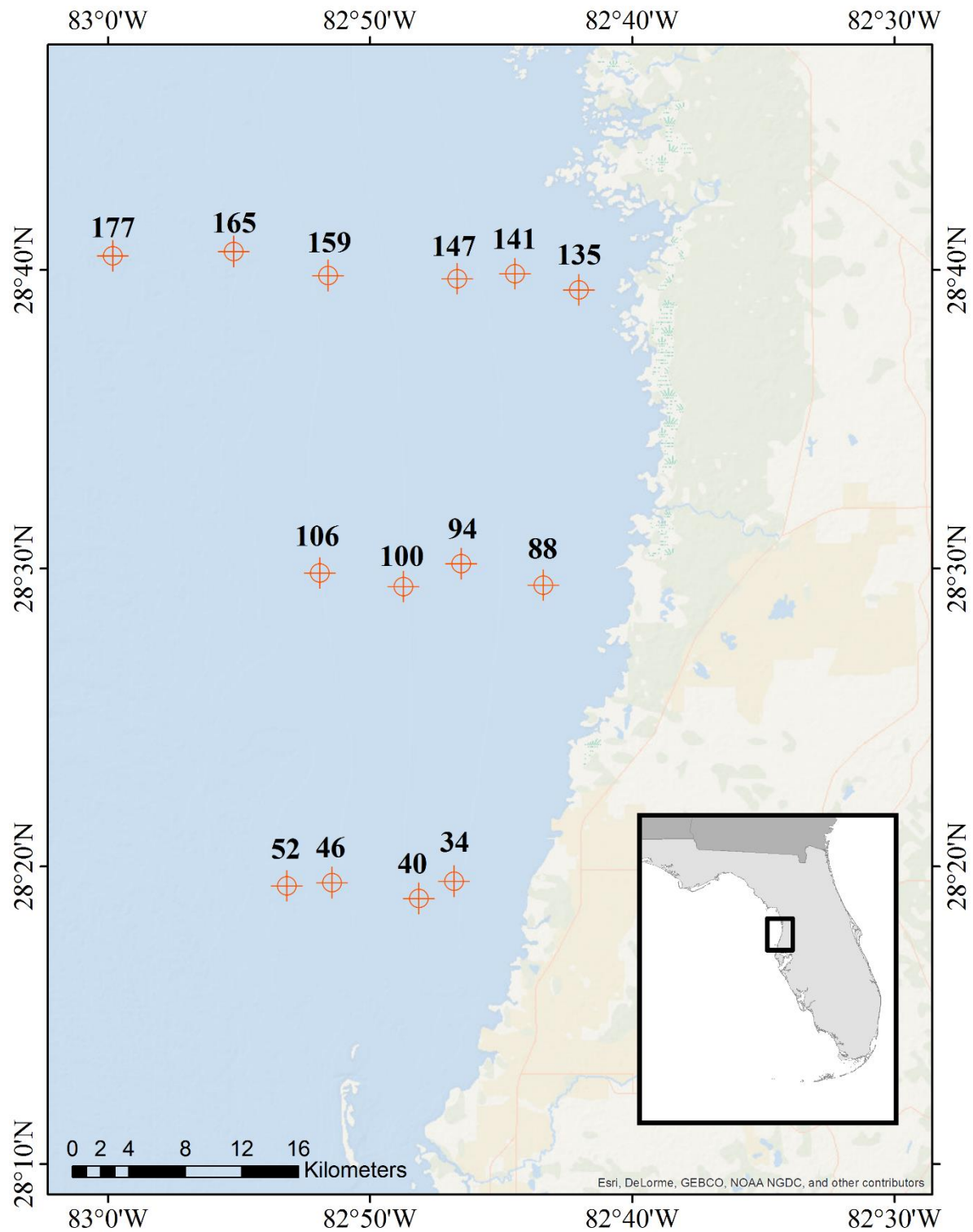
Figure 4.1: Transects and sample sites off the Florida Springs Coast

4.2. Software Precision

The library-sample AVI contained 637 camera images. By using the aforementioned procedure and Context noted in Table 2.3, VisualSpreadsheet™ captured 635 images. One image was that of a superfluous particle, which was consequently deleted, leaving a total of 634 images. Therefore, only three particles were not captured. Two of the three uncaptured camera images had been marked as "unidentified" and the remaining one had been identified as *Miliolinella circularis* (Bornemann, 1855). No further manual modifications were made, so the retained VisualSpreadsheet list file accordingly contained 634 images.

A second library-sample AVI file with the same images in the same order was processed as above and captured the same exact images. Three additional library-sample AVI files with the same images but in different orders were also processed using the said Context. VisualSpreadsheet captured 636, 606, and 632 images respectively. For the second of those AVI files, approximately 5% fewer particles were captured using the Context optimized for the original file.

The percentage differences from the mean values of each parameter collected from the original AVI file for each of the files with alternative image orders are listed in Table 4.2 and comparisons of the standard deviation values are listed in Table 4.3. For the file with "Order 1", the Elongation and Fiber Curl parameters had differences in mean values >20% from those of the original file. For the file with "Order 2", twelve parameters had differences in mean values >20% with the greatest changes also in the Elongation and Fiber Curl parameters. There were no parameters that had differences in mean values >20% for the file with "Order 3", though Fiber Curl was 16%.

Table 4.2: Difference in mean values from original order of images in AVI file

VisualSpreadsheet Statistic	Mean	Order 1	Order 2	Order 3
Area (ABD)	67190.24	3%	-15%	-1%
Area (Filled)	67618.46	2%	-9%	-1%
Aspect Ratio	0.68	n/d	-4%	n/d
Capture X	790.44	0%	1%	n/d
Capture Y	381.07	-1%	2%	0%
Circle Fit	0.47	9%	-26%	-2%
Circularity	0.65	5%	-26%	-3%
Circularity (Hu)	0.89	1%	-4%	n/d
Compactness	1.74	-11%	84%	3%
Convex Perimeter	758.33	2%	-5%	0%
Convexity	0.97	n/d	-4%	-1%
D[3 2](ABD)	645.91	-1%	-4%	0%
D[3 2](ESD)	708.13	-1%	3%	0%
D[4 3](ABD)	865.11	0%	-4%	0%
D[4 3](ESD)	965.09	-1%	2%	0%
Diameter (ABD)	222.17	2%	-9%	-1%
Diameter (ESD)	239.65	2%	-5%	0%
Diameter (FD)	222.7	2%	-8%	-1%
Edge Gradient	55.03	-6%	-19%	-4%
Elapsed Time	31.67	0%	4%	0%
Elongation	2.95	-21%	162%	7%
Fiber Curl	0.19	-42%	232%	16%
Fiber Straightness	0.98	3%	-22%	-2%
Geodesic Aspect Ratio	0.52	6%	-38%	-4%
Geodesic Length	319.37	-7%	42%	3%
Geodesic Thickness	139.18	7%	-30%	-3%
Image Height	295.16	2%	-4%	0%
Image Width	300.59	2%	-5%	0%
Intensity	119.16	-3%	9%	1%
Length	279.58	2%	-4%	0%
Particles Per Chain	1	n/d	n/d	n/d
Perimeter	917.09	-3%	20%	1%
Roughness	1.2	-3%	23%	2%
Sigma Intensity	30.53	9%	-13%	-1%
Sphere Complement	0	n/d	n/d	n/d
Sphere Count	0	n/d	n/d	n/d
Sphere Unknown	0	n/d	n/d	n/d
Sphere Volume	0	n/d	n/d	n/d
Sum Intensity	1.12E+07	1%	-11%	-1%
Symmetry	0.75	5%	-17%	-1%
Transparency	0.08	-13%	38%	
Volume (ABD)	2.89E+07	2%	-18%	-1%
Volume (ESD)	3.68E+07	1%	-1%	0%
Width	190.12	2%	-7%	0%

n/d = no difference

Table 4.3: Difference in standard deviation values from original order of images in AVI file

VisualSpreadsheet Statistic	Standard Deviation	Order 1	Order 2	Order 3
Area (ABD)	1.58E+05	1%	-11%	-1%
Area (Filled)	1.59E+05	n/d	n/d	n/d
Aspect Ratio	0.14	n/d	7%	n/d
Capture X	270.6	0%	2%	0%
Capture Y	196.16	-1%	3%	0%
Circle Fit	0.2	-5%	15%	5%
Circularity	0.15	-27%	47%	7%
Circularity (Hu)	0.08	n/d	38%	n/d
Compactness	0.98	-56%	292%	n/d
Convex Perimeter	643.47	0%	1%	0%
Convexity	0.04	-25%	100%	n/d
D[3 2](ABD)	376.28	0%	-4%	0%
D[3 2](ESD)	426.57	0%	1%	0%
D[4 3](ABD)	316.12	1%	-6%	-1%
D[4 3](ESD)	364.91	1%	-3%	0%
Diameter (ABD)	190.23	0%	-6%	0%
Diameter (ESD)	204.76	0%	1%	0%
Diameter (FD)	191.05	0%	-1%	0%
Edge Gradient	24.68	-33%	6%	-2%
Elapsed Time	18.35	0%	1%	-1%
Elongation	3.24	-53%	276%	0%
Fiber Curl	0.38	-47%	108%	5%
Fiber Straightness	0.26	-19%	15%	n/d
Geodesic Aspect Ratio	0.27	-11%	-4%	n/d
Geodesic Length	292.52	-11%	81%	3%
Geodesic Thickness	135.57	-1%	-8%	0%
Image Height	205.66	0%	1%	0%
Image Width	244.68	0%	1%	0%
Intensity	26.84	0%	0%	n/d
Length	225.91	0%	1%	0%
Particles Per Chain	0	n/d	n/d	n/d
Perimeter	780.76	-5%	48%	2%
Roughness	0.21	-52%	133%	5%
Sigma Intensity	9.45	6%	-10%	0%
Sphere Complement	0	n/d	n/d	n/d
Sphere Count	0	n/d	n/d	n/d
Sphere Unknown	0	n/d	n/d	n/d
Sphere Volume	0	n/d	n/d	n/d
Sum Intensity	2.37E+07	0%	-7%	n/d
Symmetry	0.14	-29%	50%	7%
Transparency	0.04	n/d	75%	n/d
Volume (ABD)	1.12E+08	1%	-16%	-2%
Volume (ESD)	1.51E+08	-1%	1%	n/d
Width	167.63	0%	0%	0%

n/d = no difference

27

A total of 74 images from 34 *Archaias angulatus* specimens were contained in the library-sample AVI. All of these images were captured by VisualSpreadsheet, but only 45 were selected as images to represent *A. angulatus* in the test library named "*Archaias angulatus* - LMS", where "LMS" indicates that images of large, medium, and small specimens constitute the file.

4.2.1. Testing the Value-Based Filter on Library Particles

Applying the value-based filter of the *Archaias angulatus* test library, 113 images were selected to represent *A. angulatus* from the total of 634 in the list file. The selection was cleared and the steps were repeated. The same 113 images were selected each of the ten times the filter was applied to the complete list. A total of 49 *A. angulatus* were recognized from the 113 images (i.e., 43%). These 49 specimens were isolated and their particle properties collected. After sorting the output on a property such as "Area (ABD)", data from the 49 specimens selected were identical to previous data recorded from them. Reported summary statistics, including mean and standard deviation, were also consistent.

Using the size-based library for large specimens that contained six images, those exact six images were selected from all 634 images in the list file (i.e., 100%). Two images of degraded large *A. angulatus* specimens, excluded from the library for large specimens, were not selected. Also, two images of large *Broeckina orbitolitoides* (Hofker, 1930) were not selected by the software.

Applying the size-based library for medium specimens with 21 images, 23 of 634 images were selected. One image of a *Flintinoides labiosa* (d'Orbigny, 1839) was recognized among them; therefore, 22 of 23 were correct (i.e., 96%).

28

Lastly, filtering was carried out using the size-based library for small specimens, which consisted of 18 images. From the 634 images in the list file, 44 were selected of which I identified 19 as *A. angulatus* (i.e., 43%).

Fourteen images of extra-large *A. angulatus* specimens, twelve of which constitute the "*Archaias angulatus* - X" library, were correctly ignored when applying any of the four abovementioned libraries as a value-based filter.

4.2.2. Testing the Statistically-Based Filter on Library Particles

Applying the statistically-based filter of the *Archaias angulatus* test library, 394 images were selected from the total of 634 in the list file. The selection was cleared, the steps were repeated, and the 394 particles were again selected. A total of 62 *A. angulatus* were recognized from the 394 particles (i.e., 16%). These 62 specimens were isolated and their particle properties collected. After sorting the output on a property such as "Area (ABD)", data from the 49 specimens selected were identical. Summary statistics, including mean and standard deviation, were also alike.

Applying the statistically-based filter with the test library divided into three libraries for large-, medium-, and small-sized specimens returned six accurate *A. angulatus* images of six "large" images (i.e., 100%), 40 accurate of 98 images (i.e., 41%) of medium-sized specimens, and 24 accurate of 257 images (i.e., 9%) of small-sized specimens. Although collectively the three filters selected 361 images of which 70 were identified as *A. angulatus*, the same images were sometimes selected with both the medium- and small-sized libraries.

4.2.3. Comparison of Filter Performance on Library Particles

The results of applying either the value-based filter or the statistically-based filter on

libraries of the most abundant species to the entire library-sample list file of 634 images are

shown in Table 4.4. Either of the two methods selected more images of that species than the

number of correct images; however, due to the algorithm applied by the statistical method, many

more possible images were selected when it was applied.

Table 4.4: Number of images selected from the library-sample list file using species library as a filter

VisualSpreadsheet Library	Library Images	Value Selected	Statistically Selected
Archaias angulatus	74	285	622
Flintinoides labiosa	80	346	577
Triloculina trigonula	51	186	541
Quinqueloculina candeiana	46	182	572
Quinqueloculina linneiana	42	120	520
Pseudotriloculina rotunda	36	162	511
Miliolinella circularis	32	178	415
Cycloforina sidebottomi	19	36	131
Quinqueloculina crassa	18	47	374
Triloculina bermudezi	18	50	232
Haynesina germanica	18	41	320
Quinqueloculina impressa	15	30	291
Quinqueloculina poeyana	15	113	458

The results of applying the filter as value based or statistically based for the most

abundant species found and further partitioned by size are shown in Table 4.5. Detailed results

for counts <200 are shown in parentheses. For example, of the 113 images selected using the

value-based filter on "*Archaias angulatus* - LMS", 64 were not images of *A. angulatus* while

four were images of *A. angulatus* in the list file but not in that library.

Table 4.5: Number of images selected from the library-sample list file using a size-based species library as a filter (X = extra-large, L = large, M = medium, and S = small); in parentheses, a negative number or zero was the amount incorrect, and a subsequent positive number was the additional quantity correct (i.e., correct images not in the library)

VisualSpreadsheet Library	Library Images	Value Selected	Statistically Selected
Archaias angulatus - XLMS	57	$142_{(-79, +6)}$	$524_{(-451, +16)}$
Archaias angulatus - X	12	$12_{(0)}$	$12_{(0)}$
Archaias angulatus - LMS	45	$113_{(-64, +4)}$	$394_{(-332, +17)}$
Archaias angulatus - L	6	$6_{(0)}$	$6_{(0)}$
Archaias angulatus - M	21	$23_{(-1, +1)}$	$98_{(-58, +19)}$
Archaias angulatus - S	18	$44_{(-25, +1)}$	$257_{(-233, +6)}$
Flintinoides labiosa - L	58	$134_{(-76)}$	570
Flintinoides labiosa - M	16	$43_{(-27)}$	328
Flintinoides labiosa - S	6	$6_{(0)}$	$7_{(-1)}$
Triloculina trigonula - L	20	$29_{(-9)}$	261
Triloculina trigonula - M	15	$26_{(-11)}$	397
Triloculina trigonula - S	16	$22_{(-6)}$	$109_{(-93)}$
Quinqueloculina candeiana - L	17	$21_{(-4)}$	$61_{(-44)}$
Quinqueloculina candeiana - M	19	$19_{(0)}$	272
Quinqueloculina candeiana - S	10	$21_{(-11)}$	221
Quinqueloculina linneiana - L	8	$10_{(-2)}$	$51_{(-43)}$
Quinqueloculina linneiana - M	18	$27_{(-9)}$	$77_{(-59)}$
Quinqueloculina linneiana - S	16	$24_{(-8)}$	207
Pseudotriloculina rotunda - M	17	$22_{(-5)}$	$134_{(-117)}$
Pseudotriloculina rotunda - S	19	$74_{(-55)}$	444
Miliolinella circularis - M	24	$40_{(-16)}$	206
Miliolinella circularis - S	8	$10_{(-2)}$	$39_{(-31)}$

Using the value-based filter, four libraries produced an exact match of images: extra-large *A. angulatus*, large *A. angulatus*, small *Flintinoides labiosa*, and medium *Quinqueloculina candeiana* (d'Orbigny, 1839). Aside from the small sets of extra-large and large *A. angulatus*, more images than contained in the given library were selected using the statistically-based filter. Using the set of small *F. labiosa* in this way, only one image of a different species was selected.

Irregularly shaped *A. angulatus* of comparable size to those in the extra-large and large libraries were not selected when using those filters.

4.3. Image Library Assessment

The 14 Springs Coast site sample AVIs were processed using the aforementioned procedure and Context noted in Table 2.3 into VisualSpreadsheet and saved as list files. The files were each manually reviewed. Images not of a specimen were placed into one of three designated libraries for assorted particles greater than 50 µm in diameter—borders (white ink on the slide), fragments, or other—and subsequently deleted from the list. The images of four specimens in the AVI file were captured as multiple VisualSpreadsheet images, so the largest one of each was kept while the others were deleted. More than 95% of the specimens were captured with the settings, with ≈ 4% disregarded due to the constraint on the minimum diameter and <1% indeterminable through various settings. The results are summarized in Table 4.6.

Table 4.6: Synopsis of processed AVI files of site samples into VisualSpreadsheet list files

AVI		VisualSpreadsheet				
Site ID	Specimens	Total # of Images	Irrelevant Images	Captured	Missed	Diameter 25–50 μm
SC034	206	188	7	181	25	24
SC040	202	194	7	187	15	15
SC046	205	193	2	191	14	13
SC052	225	199	0	199	24	22
SC088	238	238	2	236	2	2
SC094	208	204	5	199	9	9
SC100	212	211	3	208	4	4
SC106	244	245	4	241	3	2
SC135	232	227	2	225	7	7
SC141	202	201	0	201	1	1
SC147	222	217	0	217	5	4
SC159	215	208	8	200	15	2
SC165	205	194	1	193	12	10
SC177	203	198	5	193	10	9
Total	3019	2917	46	2871	148	124

The constraint on the minimum diameter was reduced in 5 μm increments until no more missed specimens were found in the list file. With each iteration, more irrelevant images were present in the output. Once the diameter was reduced to a minimum 25 μm, the remaining specimens, if captured in the list file, were not recognizable. Although the total number of captured specimens increased by 124 (i.e., 84% of the missed specimens), specimens 25–50 μm in diameter were not easily recognizable from the images and the additional manual processing required would not produce a substantially improved list file (i.e., a larger set of identifiable images).

4.3.1. Auto Classification of Genera

A classification within VisualSpreadsheet consists of one or more classes with each class defined using one or more filters. Within this study, class refers to a class defined in the software

and not a taxonomic class. Except for the *Haynesina* software class, which used a single statistical filter, a VisualSpreadsheet class for a genus was created from multiple statistical filters derived from genus libraries differentiated by size. For example, the *Quinqueloculina* software class (class Q) screened for particles that matched the small, medium, or large *Quinqueloculina* statistical filters.

4.3.1.1. Classifying One Genus

Of the 87 *Quinqueloculina* specimens in the SC34 sample site list file, which contained 181 images, 85 were accurately classified using a software class built with three size-based statistical library filters for the genus. While this represented 98% of the available *Quinqueloculina* identified at site SC34, an additional 68 specimens were inaccurately categorized in this software class. This reduced the overall accuracy to 56%. *Quinqueloculina* specimens consisted of nearly half (48%) of the sample and the genus was the most frequently classified inaccurately when applying an auto classification of *Haynesina*, *Triloculina*, *Flintinoides*, *Miliolinella*, or *Archaias*. Similarly, a single software class for each of these other five genera, defined with associated filters, captured a majority (≥70%) of that specific genus (Table 4.7). However, the overall accuracy for each genus was <17%.

Table 4.7: Auto-classification totals using a single class

Class	Site SC34 Images	Total	# Accurate	% Accuracy	Genus %
Q - *Quinqueloculina*	87	153	85	56	98
H - *Haynesina*	20	85	14	16	70
T - *Triloculina*	17	147	16	11	94
F - *Flintinoides*	11	139	11	8	100
M - *Miliolinella*	11	128	10	8	91
A - *Archaias*	6	39	5	13	83
Other genera	30				
Total	181				

More than 50% of the *Haynesina*, *Triloculina*, *Flintinoides*, *Miliolinella*, and *Archaias* images captured were aggregated into the *Quinqueloculina* software class when applying auto classification as shown in Table 4.8.

Table 4.8: Auto-classification totals of genera using software class Q - *Quinqueloculina*

Genus	Site SC34 Specimens	Images Missed	Images Captured	Class Q - *Quinqueloculina*
Auto classified				153
Quinqueloculina	104	17	87	85
Haynesina	23	3	20	11
Triloculina	19	2	17	16
Flintinoides	12	1	11	9
Miliolinella	12	1	11	10
Archaias	6	0	6	4
Other genera	30	1	29	18
Total	206	25	181	

4.3.1.2. Sequential Identification of Multiple Genera

As illustrated in the previous section, many specimens were inadvertently classified. When applying the auto-classification feature to a classification using multiple software classes, each is applied in sequence such that once an image is assigned, it will not be reassigned to a

subsequent software class. Although particles can be manually moved or removed, the order of the software classes influences the auto-classification result. The outcomes of four different arrangements are shown in Table 4.9. The first classification scheme, QHTFMA, listed genera in the order of most abundant (*Quinqueloculina*) to least abundant (*Archaias*); the second arrangement, AMFTHQ, listed genera from least to most abundant; the third arrangement, HMTQFA, positioned *Haynesina* of group D first, followed by genera of group C, and then *Archaias* of group A; and the fourth order, AFQTMH, was the reverse of the third. In all four scenarios, 164 particles were classified.

Applying auto classification on these defined classifications, the software class listed first produced the same results as when the classification solely used that software class such that class Q, class H, and class A had a total of 153, 85, and 39 specimens respectively with the same accuracy. The classification using the order QHTFMA had the most correct with 88 of 164 specimens (54%), but only three specimens correct beyond software class Q - *Quinqueloculina*, the first one listed, which had 85 correct. When switching the order of the first two software classes in this scheme so that class H preceded class Q, only 79 versus 88 specimens were correct. The other three classification arrangements each had an accuracy <20%.

Table 4.9: Auto-classification totals of genera using six software classes in four orders

Genus	Site SC34	Class order QHTFMA	Class order AMFTHQ	Class order HMTQFA	Class order AFQTMH
Q - *Quinqueloculina*	104	153$_{(85)}$	6$_{(6)}$	8$_{(7)}$	22$_{(20)}$
H - *Haynesina*	23	8$_{(3)}$	2$_{(0)}$	85$_{(14)}$	2$_{(0)}$
T - *Triloculina*	19	2$_{(0)}$	7$_{(0)}$	17$_{(0)}$	1$_{(0)}$
F - *Flintinoides*	12	1$_{(0)}$	17$_{(2)}$	1$_{(0)}$	100$_{(6)}$
M - *Miliolinella*	12	0$_{(0)}$	93$_{(5)}$	53$_{(1)}$	0$_{(0)}$
A - *Archaias*	6	0$_{(0)}$	39$_{(5)}$	0$_{(0)}$	39$_{(5)}$
Total	176	164$_{(88)}$	164$_{(18)}$	164$_{(22)}$	164$_{(31)}$

Number in parentheses was quantity correct.

4.3.2. Auto Classification of Categories

4.3.2.1. Hyaline Surface Texture

A single class was created using a statistically-based filter produced with 38 specimens exhibiting a hyaline or glassy, translucent surface texture from library-sample LibSC34. The distinctive feature is conspicuous to researchers, but auto classification using the hyaline software class accurately assessed only about one quarter of the specimens from the SC34 site sample as having a hyaline surface texture. Moreover, the auto classification excluded nearly one-third (15 of the 47) hyaline specimens.

4.3.2.2. Functional Group

Four software classes were created from filters analogous to the functional groups of the FORAM Index (Hallock et al. 2003, Hallock 2012) outlined in Table 2.5. Groups A, B, C, and D correspondingly represent "Large Benthic", "Smaller Miliolid", "Other Smaller", and "Stress Tolerant" foraminifers. The initial filters for each software class were based on image libraries derived from library-sample LibSC34. Software classes A and B had three filters designed for small, medium, and large specimens whereas software classes C and D only had two filters designed for small and medium specimens. The dialog box for class A is illustrated in Figure 4.2, where particles selected should match any of the listed filters.
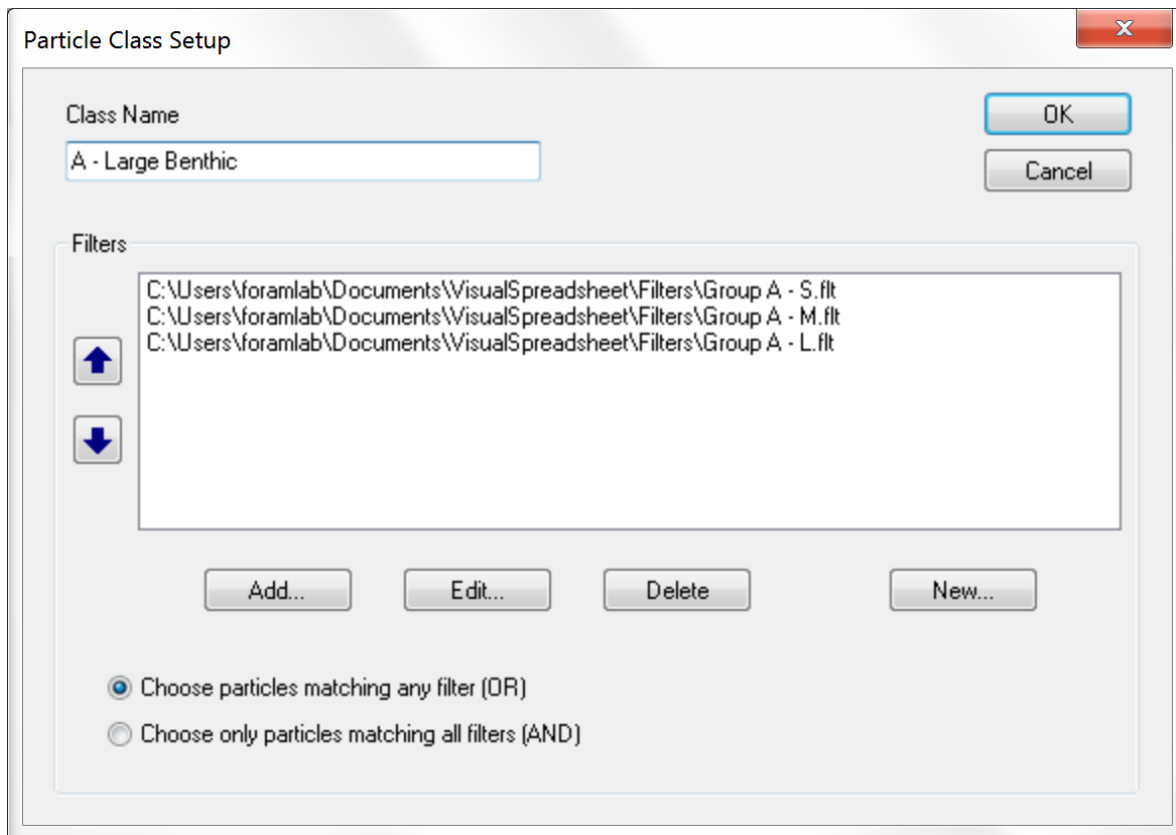
Figure 4.2: Particle Class Setup dialog box for class A with its initial three filters

The first classification scheme, BDCA, listed groups in order of category abundance found in the SC34 site sample. Group B included 128 of the 181 images and was listed first followed by groups D, C, and A with 39, 8, and 6 images respectively. The second order, ABCD, followed the presentation order of the functional groups in Table 2.5 where the larger symbiont-bearing taxa are in group A. The last order, DCBA, was the opposite of the second. In total, 161 particles were classified for all three arrangements as shown in Table 4.10. The composite accuracy was 80% for software class order BDCA, 75% for order ABCD, and 39% for order DCBA. Although the composite accuracy of software class order BDCA was the highest, that of order ABCD was comparable and it also classified specimens in all four software

classes. Since other sample sites will have different abundances of the four groups, software

class order ABCD was selected as the model template. Also, the initial statistical filters of group

C had the fewest available images, so auto classification could have placed more images in that

software class as images were added and updated filters were employed.

Table 4.10: Auto-classification totals of foraminiferal groups using four classes in three orders

Group	Site SC34	Class Order BDCA	Class Order ABCD	Class Order DBCA
A - Large Benthic Foraminifers	6	0	19	0
B - Smaller Miliolid Foraminifers	127	145	126	41
C - Other Smaller Foraminifers	9	0	1	0
D - Stress Tolerant Foraminifers	39	16	15	120
Total	181	161	161	161

The composite results of SC34 using software class order ABCD were evaluated to the

group level as shown in Table 4.11. Smaller Miliolid Foraminifers were the most abundant in the

SC34 site sample and the filters utilized in class B accurately classified 108 specimens.

Foraminifers in Group D were the next most abundant and its corresponding software class,

processed last by the auto classification, had an accuracy of 60%. The other two groups

represented approximately 8% of the specimens captured as images and only four were properly

placed. Twenty specimens were not classified.

Table 4.11: Auto-classification accuracy of foraminiferal groups using class order ABCD

Group	Site SC34	Classified	# Accurate	% Accuracy
A - Large Benthic Foraminifers	6	19	4	21
B - Smaller Miliolid Foraminifers	127	126	108	86
C - Other Smaller Foraminifers	9	1	0	0
D - Stress Tolerant Foraminifers	39	15	9	60
Unspecified	0	0		
Total	181	161	121	75

The auto-classification results from all sites are shown in Table 4.12 with a composite accuracy >65%. Site SC52 had the highest accuracy at 91%. Site SC165 had the lowest and was the only one with an accuracy <50%. The percentage of unclassified specimens was greater than that of SC34 (11%), the first sample site, for all other sites except SC165. The overall percentage of unclassified specimens exceeded 20% using the initial set of software classes. As tallied from the manual identifications, the list files contained 168 images of specimens from group A, 1573 from group B, 271 from group C, and 834 from group D.

Fewer than 1% of all images processed by auto classification (i.e., 25 specimens) were from taxa not assigned to a functional group as listed in Table 2.5. The manual identifications included eleven genera not originally listed; and the specimens were later reassessed, sometimes by correcting the assigned taxa or revising Table 2.5. In a few instances, the auto classification placed those specimens in the software class associated with the group in which they were expected to be assigned. For example, a total of five specimens of *Broeckina orbitolitoides*, a species geographically found in and near the Caribbean Sea of the Atlantic Ocean, were originally identified incorrectly in three site samples. Similarities in external features of *B. orbitolitoides* to phylogenetically distinct species can lead to misidentification (Holzmann et al. 2001). One specimen from the SC40 sample and two of three specimens from the SC94 sample were classified by the software as group A. The image from sample site SC40 was later added to one of the libraries.

Table 4.12: Preliminary composite auto classification of foraminiferal functional groups

Site ID	List File	Classified	# Accurate	% Accuracy	Unclassified	% Unclassified
SC034	181	161	121	75	20	11
SC040	187	162	99	61	25	13
SC046	191	127	75	59	64	34
SC052	199	152	138	91	47	24
SC088	236	200	157	79	36	15
SC094	199	170	117	69	29	15
SC100	208	178	112	63	30	14
SC106	241	207	117	57	34	14
SC135	225	152	79	52	73	32
SC141	201	178	119	67	23	11
SC147	217	185	120	65	32	15
SC159	200	99	66	67	101	51
SC165	193	172	81	47	21	11
SC177	193	138	91	66	55	28
Total	2871	2281	1492	65.4	590	20.6

The most specimens not classified by the software were from group D as presented in Table 4.13. However, for site SC159, a high percentage of *Archaias angulatus* were unclassified, which should have been designated as group A. Further examination of those images revealed a number of very large, partially captured, and degraded specimens as illustrated by the particle edge trace of the selected images in Figure 4.3. For example, image number 24 (middle row, second image from left to right in Figure 4.3) did not fully encompass the specimen thereby portraying it as less circular than its actual shape.
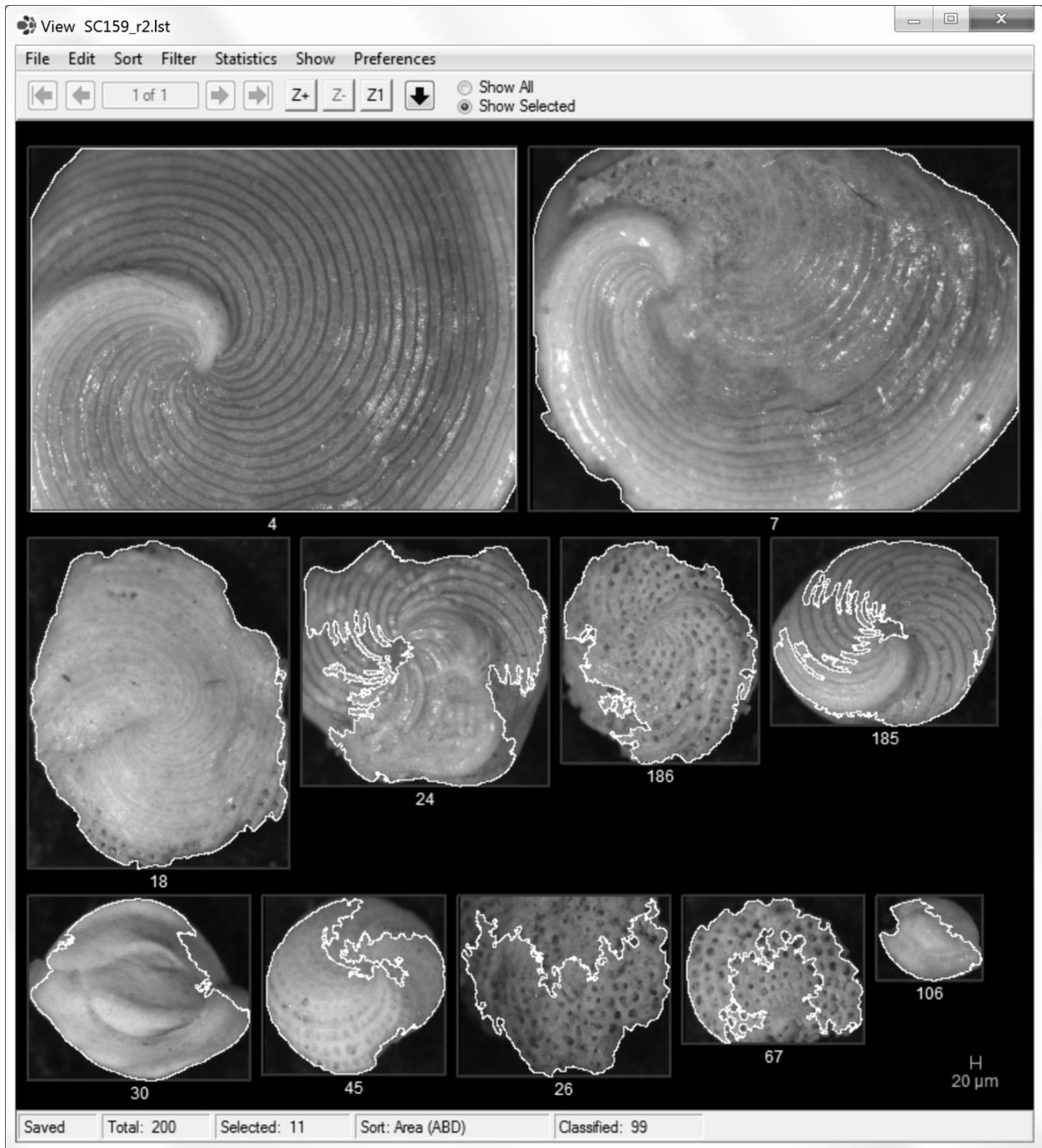
Figure 4.3: Selected unclassified images from sample site SC159

Table 4.13: Distribution of unclassified specimens

Site ID	Group A	Group B	Group C	Group D	Unassigned	Total
SC034	2	5	3	10	0	20
SC040	1	14	2	8	0	25
SC046	1	12	0	50	1	64
SC052	0	12	0	35	0	47
SC088	0	29	3	4	0	36
SC094	4	13	3	9	0	29
SC100	2	12	6	10	0	30
SC106	3	7	7	17	0	34
SC135	0	4	4	65	0	73
SC141	0	7	3	12	1	23
SC147	0	12	8	12	0	32
SC159	79	14	3	5	0	101
SC165	0	3	4	13	1	21
SC177	3	10	12	29	1	55
Total	95	154	58	279	4	590

Since group A typically had a lower proportion of specimens in most samples, an evaluation of the auto classification was performed by modifying software class A. By removing the filter for small specimens in class A, accuracy results of SC34 improved from 75% to 81% (Table 4.14).

Table 4.14: Auto-classification accuracy of foraminiferal groups following a revision of class A by removing filter for small specimens of Large Benthic Foraminifers

Group	Site SC34	Classified	# Accurate	% Accuracy
A - Large Benthic Foraminifers[1]	6	5	4	80
B - Smaller Miliolid Foraminifers	127	140	118	84
C - Other Smaller Foraminifers	9	1	0	0
D - Stress Tolerant Foraminifers	39	15	9	60
Unspecified	0	0		
Total	181	161	131	81

[1] Class was only comprised of filters for medium and large specimens regarding Large Benthic Foraminifers

The adjustment did not change the number of images classified for site SC34, but the improvement justified processing all sample sites with the modification to determine the impact

on the overall accuracy. Collectively, the accuracy percentage increased from 65% to 71% and

nine fewer images were classified as shown in Table 4.15. Except for site SC52, which had no

change, all sites increased in accuracy. Using a significance level of 5%, a 10,000 iteration

permutation test on the differences in accuracy was performed. The resulting *p*-value of 0.099

did not reject the null hypothesis that the improvement was by chance. Although not significantly

better, this classification arrangement with the aforementioned software classes and filters that

excluded the filter in class A for small specimens of Large Benthic Foraminifers formed the

baseline for further analysis in this study.

Table 4.15: Composite foraminiferal group auto-classification comparison between including and excluding the filter for small specimens of Large Benthic Foraminifers (LBFs) in class A

| Site ID | Including filter for small LBFs | | | Excluding filter for small LBFs | | |
	Classified	# Accurate	% Accuracy	Classified	# Accurate	% Accuracy
SC034	161	121	75	161	131	81
SC040	162	99	61	162	108	67
SC046	127	75	59	124	80	65
SC052	152	138	91	152	138	91
SC088	200	157	79	200	171	86
SC094	170	117	69	170	127	75
SC100	178	112	63	178	118	66
SC106	207	117	57	207	122	59
SC135	152	79	52	148	94	64
SC141	178	119	67	178	142	80
SC147	185	120	65	184	131	71
SC159	99	66	67	98	67	68
SC165	172	81	47	172	84	49
SC177	138	91	66	138	99	72
Total	2281	1492	65.4	2272	1612	71.0

The accuracy that excluded the filter for small LBFs was compared to grain-size statistics

of the source sample sites in Appendix D. A correlation analysis was performed to determine if

accuracy was influenced by either grain-size or cumulative grain-size percentages. As shown in

Table 4.16, there was no significant correlation between accuracy and either grain-size or cumulative grain size.

Table 4.16: Correlation between foraminiferal group accuracy (excluding filter for small LBFs) and grain-size percentages

	Accuracy (excluding filter for small LBFs)	
Grain size	Separate	Cumulative
> 2 mm	-0.06	-0.06
> 1 mm	0.10	-0.01
> 0.5 mm	-0.05	-0.03
> 0.25 mm	-0.23	-0.11
> 0.125 mm	0.28	0.18
> 0.063 mm	-0.10	0.19
< 0.063 mm	-0.19	0.22

4.4. Evaluation of "Trained" Software Identifications

As previously described in section 4.3.2.2, the initial statistical filters were built from library images from the LibSC34 library-sample. Details on the original VisualSpreadsheet classification template are provided in Appendix E. Additional images were added to libraries from sample sites as outlined in Table 4.17, and then new filters were built. Next, new classifications were created from an updated template that utilized the revised filters, and sites were reassessed using the auto-classification feature. Once a given software library had more than 60 images, no further images were added from succeeding revisions and its associated filter remained unchanged. Marginal images of specimens for each group were also collected in libraries to create supplementary filters for inclusion in later trials.

The first revision included images from sample site SC34. The second revision included additional images from sample sites SC40 and SC46. As those sample sites provided sufficient

images of distinctly large specimens in comparison to typical medium-sized ones belonging to

group D, a new software library and an associated statistical filter were created. Since the

software class for group D was last in the series, this additional filter for large group D

specimens was included from this point forward because it would not impact the auto

classification of the other classes. The third revision added images from sample sites SC52 as

well as filters for class C and class D based on marginal images.

Table 4.17: Quantity of images in software library used to build statistical filters

Filter	Initial	Revision 1	Revision 2	Revision 3
Group A - S	14	18	20	n/c
Group A - M	14	n/c	17	n/c
Group A - L	8	n/c	10	n/c
Group A - X	12	13	15	n/c
Group B - S	60	100	n/c	n/c
Group B - M	46	57	62	n/c
Group B - L	34	n/c	39	n/c
Group C - S	6	9	17	19
Group C - M	2	3	4	n/c
Group D - S	12	35	85	n/c
Group D - M	18	23	49	51
Group D - L	n/a	n/a	15	17
Group A - (m)	9	10	n/c	n/c
Group B - (m)	n/a	4	21	23
Group C - (m)	n/a	3	4	n/c
Group D - (m)	n/a	10	50	58

(m) = marginal images; n/a = not applicable; n/c = no change

In addition to applying the revised filters, a final modification to the classification

structure was tested. This modification used the latest filter revisions and appended a second

software class of group A specimens that included filters for small, marginal, and extra-large

specimens. As shown in Table 4.18, the percentage classified increased with each modification,

but the accuracy declined, plateauing ≈60%.

Table 4.18: Composite foraminiferal group auto-classification percentage accuracy using revised filters

Auto Classification	% Classified	% Accuracy	Net % Accuracy
Original	79.1	71.0	56.1
Modification 1	79.6	68.9	54.8
Modification 2	87.2	60.7	52.9
Modification 3	87.4	60.7	53.1
Modification 4	89.3	60.9	54.4

4.5. Identifying Foraminiferal Specimens Directly From a Sediment Sample

A sample from site SC46 was treated with trichloroethylene (TCE) to remove the majority of the sediment and concentrate the foraminiferal specimens. Particles from the float portion of the separated sample were placed on micropaleontological faunal slides and photographed. One hundred camera images were taken of the particles using a standard focus (i.e., each image was not focused separately) and converted into an AVI file. The images were then processed by VisualSpreadsheet into a list file. The original auto classification using four classes corresponding to foraminiferal groups was then applied to the list. Only the assignment of foraminifers to one of the VisualSpreadsheet classes was considered. That is, distinguishing an image as a foraminiferal specimen from a general particle (e.g., fragment, sediment, non-foraminifers, etc.) was the purpose, but the original classification was used for continuity and to separate the sets of filters.

As presented in Table 4.19, a total of 536 "particle" images were captured from the 100 photographs. Approximately 40–45% of the total were identified as foraminiferal specimens, but some of the 232 images recognized had poor particle boundaries (e.g., partial trace of a specimen

or inclusion of extraneous nearby sediment). Approximately 75% of the classified images (i.e., 188 of 248) were correctly identified as foraminiferal specimens.

Table 4.19: Auto classification of foraminifers from a trichloroethylene-treated sample from site SC46

Class	Images	Specimens[1]	Rejects
A - Large Benthic Foraminifers	44	41	3
B - Smaller Miliolid Foraminifers	136	110	26
C - Other Smaller Foraminifers	5	4	1
D - Stress Tolerant Foraminifers	63	33	30
Classified	248	188	60
Not classified	288	44	244
Total	536	232	304

[1] Specimens were not checked for accurate classification

4.6. Software Metrics That Facilitate Identification

The following 24 VisualSpreadsheet metrics, described in the FlowCam® Manual (Fluid Imaging Technologies 2014b), were evaluated for their capability to help screen images for genus and functional group: Area (Filled), Aspect Ratio, Circle Fit, Circularity, Circularity (Hu), Convex Perimeter, Convexity, Diameter (ABD), Edge Gradient, Elongation, Fiber Curl, Fiber Straightness, Geodesic Length, Geodesic Thickness, Intensity, Length, Perimeter, Roughness, Sigma Intensity, Sum Intensity, Symmetry, Transparency, Volume (ABD), and Width. The metrics Compactness and Geodesic Aspect Ratio were excluded because those measurements are the inverse of Circularity and Elongation respectively. In addition, Area (ABD), Diameter (ESD), Diameter (FD), and Volume (ESD) were not reported due to their similarity to comparable metrics. The data from the VisualSpreadsheet list file of library-sample LibSC34 containing 634 images were used for the analysis.

The quantity of images belonging to each functional group from the first decile (i.e., 63) of each metric, when sorted from smallest to largest and largest to smallest, were recorded in Table 4.20 and Table 4.21 respectively. Metrics with extreme numbers (i.e., near 0 or 63) could indicate a threshold or a range of values that could be applied in filters to aid identification of specimens or a functional group.

For example, as shown in Table 4.20, 58 of 63 specimens that had a small Circularity (Hu) measurement belonged to the Smaller Miliolid group and those Circularity (Hu) values ranged from 0.57 to 0.77. In contrast, 335 specimens had a measurement of zero for Fiber Curl, so small values of this metric may not provide an indication of a Smaller Miliolid although 59 of 63 specimens were listed for that group in Table 4.20.

Table 4.20: Functional group quantity in LibSC34 list file from the first decile of each metric's smallest measurements

	Images in the sample	Area (Filled)	Aspect Ratio	Circle Fit	Circularity	Circularity (Hu)	Convex Perimeter	Convexity	Diameter (ABD)	Edge Gradient	Elongation	Fiber Curl	Fiber Straightness	Geodesic Length	Geodesic Thickness	Intensity	Length	Perimeter	Roughness	Sigma Intensity	Sum Intensity	Symmetry	Transparency	Volume (ABD)	Width
Large Benthic Foraminifers	76	0	1	4	10	1	0	2	0	14	10	0	13	0	1	7	0	0	12	13	0	8	33	0	0
Smaller Miliolid Foraminifers	463	39	57	52	36	58	38	34	39	39	49	59	31	46	28	22	33	42	48	21	33	32	28	39	45
Other Smaller Foraminifers	10	4	1	0	0	0	5	2	3	0	0	1	0	1	4	2	7	3	0	3	3	1	0	3	1
Stress Tolerant Foraminifers	58	15	2	5	15	1	15	24	16	9	2	1	18	9	27	30	16	12	2	24	24	21	2	16	12

Color scale is relative to value where maximum is darkest

Similarly, as shown in Table 4.21, 49 of 63 specimens that had a large Diameter (ABD) measurement belonged to the Large Benthic group and those Diameter (ABD) values ranged

from 373 to 1159. Furthermore, 39 of those 49 specimens had a Diameter (ABD) measurement >540. Although zero specimens in the Large Benthic group were listed with those having the smallest Diameter (ABD), a filter that excluded specimens with diameter less than a certain measurement may also exclude some in the group.

Table 4.21: Functional group quantity in LibSC34 list file from the first decile of each metric's largest measurements

	Images in the sample	Area (Filled)	Aspect Ratio	Circle Fit	Circularity	Circularity (Hu)	Convex Perimeter	Convexity	Diameter (ABD)	Edge Gradient	Elongation	Fiber Curl	Fiber Straightness	Geodesic Length	Geodesic Thickness	Intensity	Length	Perimeter	Roughness	Sigma Intensity	Sum Intensity	Symmetry	Transparency	Volume (ABD)	Width
Large Benthic Foraminifers	76	49	24	21	11	33	47	11	49	4	10	13	8	43	46	7	45	46	12	5	50	10	1	49	49
Smaller Miliolid Foraminifers	463	14	25	34	48	23	16	48	14	53	36	31	53	20	17	54	18	17	32	55	13	53	47	14	14
Other Smaller Foraminifers	10	0	3	2	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Stress Tolerant Foraminifers	58	0	9	4	3	4	0	1	0	1	15	18	1	0	0	0	0	0	17	0	0	0	13	0	0

Color scale is relative to value where maximum is darkest

In summary, Tables Table 4.20 and Table 4.21 identify metrics that deserve further review if considered for use in filters to aid identification or separation of specimens or functional group, but the metrics should be carefully scrutinized to be applied effectively.

5. Discussion

Digital imaging technologies combined with image recognition software programs are being applied to areas of scientific study ranging from nanotechnology to space exploration. In marine research, innovative imaging applications are similarly diverse, covering nanoscale structures to satellite imagery. So why, given that foraminifers are the most abundant shell-producing organisms in the ocean and foraminiferal applications range from cytology to historical geology, has image recognition not found wide-ranging use in micropaleontological research? The potential of image recognition technology to support the identification of foraminifers has been recognized for more than two decades. As previously mentioned, an image-analysis based system for the classification of planktic foraminifers was prototyped (Liu et al. 1994, Yu et al. 1996), but the potential has not been realized (Ranaweera et al. 2009a, 2009b).

My research was motivated by seeing how systems like the FlowCam® VS Series have been used in marine applications to study community structures of plankton (See et al. 2005, García-Muñoz et al. 2013) and to determine organism size (Álvarez et al. 2011, Spaulding et al. 2012). In a preliminary trial of the instrument with a sediment sample containing foraminiferal specimens, images were too dark because the FlowCam was designed to illuminate particles from behind as they passed through the flow cell to be digitally photographed. Since the system was not able to capture opaque images with sufficient clarity to distinguish specimens, digital

51

images were taken with a microscope camera. Thereafter, the methodology developed by researchers at the University of Colorado (as discussed with Mr. Nelson, personal communication 2016) was implemented where collections of images were converted to Audio Video Interleave (AVI) files and later processed by the VisualSpreadsheet™ software.

These steps allowed me to take advantage of and to evaluate the image recognition capabilities of VisualSpreadsheet. Yet, many challenges, including issues posed by foraminiferal tests associated with image acquisition and specimen orientation as well as other technical problems, were encountered during the research. Those difficulties along with recommendations are discussed in further detail in the following sections.

5.1. Primary Objectives of the Thesis

5.1.1. Precision and Replicability

My first objective was to test the precision of the software in its ability to identify the same particles in repeated test runs. The software identified the same particles in repeated runs and provided identical data as long as the context settings were unchanged and the images were in the same order. Because the background was calibrated using the first 32 images, altering the order of images taken from different micropaleontological slides affected the quantity of images recognized and the area captured. Modifications to the context settings are needed to maximize the output, and consistency is aided by clear distinctions between the particle and the background. As a result, a uniform background supports consistency. Image acquisition issues are discussed in more detail in section 5.2.

5.1.2. Training to Achieve Stable Results

The second objective was to determine the number of runs needing supplementary quality control to acquire a stable level of accuracy in identifying known foraminiferal specimens. My hypothesis was, if additional images were added to the initial image libraries of the image recognition software, then its ability to classify foraminiferal specimens through updated filters would improve as measured by an increase in the composite foraminiferal group auto-classification accuracy. The null hypothesis was that there would be no significant difference in the composite accuracy. A stable level of accuracy was established at 70% and, with subsequent filter modifications, accuracy plateaued at 60%. Because there was no improvement to the initial accuracy, the hypothesis was not supported.

I also evaluated the classification process at multiple ranks of categorization from specific taxon to genus-level to functional group. At the most specific rank, using *Archaias angulatus* as the species, more identifications were correct when *A. angulatus* was noticeably distinguishable from other specimens by its substantially larger size. The smaller *A. angulatus* were selected along with many other similarly sized and shaped specimens. When evaluating *A. angulatus* within foraminiferal assemblages, the characteristic chamber arrangement did not result in identification of *A. angulatus*.

At the genus-level, *Quinqueloculina* spp. were the most abundant, representing 39% of all specimens in the 14 site samples. Although abundances of *Quinqueloculina* were variable and some specimens were not captured, their prevalence required placing associated categories (i.e., software classes "Q - *Quinqueloculina*" and "B - Smaller Miliolid Foraminifers") before others to obtain relatively favorable classification results. Furthermore, even though a stable level of

accuracy was reached at 60% for the composite foraminiferal group auto classification, species distribution in an assemblage affected the results. For example, at site SC52, there were 152 captured images of which 121 of 125 (i.e., 97%) from group B were accurately classified. The 91% overall accuracy reached at site SC52 (Table 4.12 and Table 4.15) was influenced by dominance of one group in the assemblage. Researchers, therefore, may be able to achieve enhanced semi-automated classification results by applying *a prior* knowledge or expectations.

Although high levels of accuracy (i.e., 80–90%) were not expected at more detailed levels of classification, improvements to the accuracy of early software tests were sought by first determining an appropriate baseline. The degree of accuracy was established at foraminiferal functional group with a preliminary auto-classification accuracy of 65%. The auto-classification results using software classes with filters based on specimens exhibiting a hyaline surface texture (≈25%) and genus-level partitions (≈55% maximum) were less than the minimum preliminary goal of 60%. Although results from the library sample using species-level partitions of larger *A. angulatus* specimens were acceptable, the majority (>90%) of foraminiferal images available were small (area-based diameter <115,000 μm^2). Overall, the results of taxa selected using the statistically-based filters (Table 4.4 and Table 4.5) were not sufficient to examine species-level auto classification. The evaluation of the technology and specific challenges are discussed further in sections 5.6 and 5.7 respectively.

I noticed that *A. angulatus*, a member of group A, was proportionately underrepresented in most samples. After adjusting the filters for group A, a trial on site SC34 improved the accuracy from 75% to 81%. The adjustment was then applied to all samples, and the composite accuracy was 71%, which was not significantly better than the original 65% ($\alpha = 0.05$). Two

subsequent modifications to the filters lowered the composite accuracy to 61% and two additional modifications to the filters did not change it, so 2–4 modifications were needed to acquire a stable level of accuracy. Importantly, revised statistical filters were a function of newly acquired images, so no modifications may have been needed if the original set of filters contained a sufficient number images. Although no images were rejected beyond the initial quality control step to remove non-foraminiferal images, manual reviews and corrections would still be needed. Nonetheless, the composite accuracies of foraminiferal groups (60–70%) did not reach the 80% threshold of a satisfactory level of accuracy.

5.1.3. Comparison of Software to Researcher Identifications

The third objective of the thesis was to compare auto-classification identifications of foraminiferal assemblage samples to results from microscopic analysis of the same samples by a researcher. Although the manual identifications were regarded as the standard, the software did reveal five misidentifications and several transcription errors. Fewer than 1% of researcher identifications were not confirmed as correct during either quality control after processing AVI files or validation of the auto-classification results. The auto-classification feature of VisualSpreadsheet produced composite accuracies of foraminiferal functional groups in the range of 60–70% compared to traditional visual identification by a researcher using a stereo-microscope.

5.1.4. Direct Identification from a Sediment Sample

The fourth objective of my project was to test whether trained software could identify foraminiferal specimens directly from a sediment subsample, and with what degrees of accuracy

and precision. From 100 photographs, 536 images were captured and 248 were identified as foraminiferal specimens. Of those, approximately 75% (i.e., 188) were correctly identified, while 44 additional foraminiferal specimens were manually recognized among the 288 images not classified by the software.

Because a standard focus was used when photographing the images for this trial to emulate bulk processing, image clarity was compromised. However, the added effort to optimize the depth of field for each photograph can be problematic in a sample with highly variable particle sizes, as diameters of foraminiferal tests can range from <100 µm to a centimeter or more. One strategy to address that challenge would be to assess similar size ranges, but that requires extra effort of sieving and of then standardizing results from each size fraction.

Furthermore, only a small sample from the float portion of the trichloroethylene procedure (Murray 2006) that contained lighter foraminiferal specimens was processed. Because only 40–45% of this aliquot were identified as foraminiferal specimens, a sample of the sediment residual was not examined.

Distinguishing foraminiferal specimens from other particles directly from a sample of the float portion was possible, but the level of precision was very limited so further tests were not conducted. Additional identification challenges are discussed in sections 5.3 and 5.7, but direct analyses is foreseeable if an enhancement to capture internal features as mentioned in section 5.5 is realized.

5.1.5. Software Metrics

The final objective of this study was to determine which metrics provided by the software could facilitate identification of the foraminiferal taxa and if customized metrics could increase

accuracy. Several morphologic parameters could be postulated to support identification, including size-related parameters, circularity, and symmetry. For example, smaller values of circularity (Hu), as well as larger values for fiber straightness and symmetry, were associated with smaller Miliolida. However, specimens of this group were also the most abundant in the test sample. Another morphologic parameter, fiber curl, had a measurement of zero for a majority (>50%) of the images.

Size-related parameters such as area, diameter, volume, and width are typically associated with larger benthic foraminifers such as *A. angulatus*, so these metrics could be used to filter for these foraminifers. However, juvenile specimens of such taxa would not be recognized by size and no metric distinguished them from similar-sized specimens. That is, the software easily partitioned specimens by size, but there were no accompanying metrics that would complement foraminiferal identification or further categorization. Because no software metrics were recognized that definitively identified foraminiferal taxa, custom metrics that combined parameters were not created.

5.2. Acquisition of Consistent Images

Because the FlowCam did not capture images that provided sufficient clarity to distinguish foraminiferal specimens in a preliminary trial, this challenge was addressed by taking digital images with a microscope camera. The individually photographed images provided considerably more detail than the automated ones generated by the instrument. However, differences in the lighting and inconsistency in the black background provided by micropaleontological faunal slides induced noticeable variability. The lighting issue could be

addressed by normalizing the directed light and dedicating the equipment to the task. Another approach to improve consistency would be to photograph specimens from all of the samples at once rather than after each sample was picked and catalogued. However, this approach would only provide consistency within an individual project, and not between applications, projects, or research teams. An automated imaging instrument should use direct lighting and capture the perspectives needed for identification.

The most problematic issue for the software in assessing camera images was the contrast of the background. The surface of a standard micropaleontological faunal slide was coarse when magnified, so the boundary of a specimen sometimes was not sufficiently distinct. As a consequence, the software occasionally did not correctly trace the edge of some specimens, either including portions of the slide or paring parts of the specimen. The captured outline of darker or hyaline specimens was most adversely affected. Such specimens might have been better photographed using a white background, but utilizing multiple backgrounds could further complicate the process. Wet surfaces from the process of transferring specimens onto a slide sometimes caused added glare, but this did not have a noticeable effect on the software's ability to delineate the contour of specimens.

To attain superior images, a microscope with camera and lighting equipment should be dedicated to the procedure. A set of 40 images using several representative specimens of the category of interest should be photographed on a black surface and processed in VisualSpreadsheet until the preferred background is identified. This background should then be used consistently. If relying on micropaleontological faunal slides, use a single batch for consistency. If working with darker or hyaline specimens, consideration of a white background

should be given explicitly for them. Also, wet surfaces should be allowed to dry. Finally, the AVI files, in addition to having approximately 40 or more images, should preferably contain concurrently developed images (e.g., photographs taken on a single micropaleontological faunal slide). Alternatively, a more consistent photographic surface must be developed and utilized for this application.

5.3.   Specimen Orientation for Diagnostic Criteria

Identification of foraminifers often requires viewing specimens from several perspectives to determine the morphology of the test (e.g., milioline, planispiral, trochospiral, etc.), type of aperture (e.g., terminal slit, with bifid tooth, etc.), ornamentation and other external features. For the initial library of images, 3–4 views of the dorsal and ventral sides, and the apertures of specimens were photographed. Careful handling was often required to position specimens for a view of the aperture, which accounted for most of the time and effort to photograph the initial specimens. However, only dorsal views were amassed in the software libraries because that orientation often had the most diagnostic criteria, a consistent view was desired for image recognition algorithms, and site samples were photographed with a single orientation.

As a consequence of the limited perspectives, a large number of benthic foraminiferal specimens were not uniquely identifiable. Therefore, the most comprehensive identification should be made by researchers at the point specimens are photographed. In addition, if multiple specimens are to be photographed together, the specimens should be of a comparable size. For example, it is beneficial to cluster similar specimens whenever possible (e.g., multiple specimens of *A. angulatus*).

5.4. Transformation and Organization of Image Files

Creating AVI files from images taken by the microscope camera provided enhanced details for VisualSpreadsheet to process and classify the foraminifers in an assemblage. An important specification of the AVI file for use with VisualSpreadsheet was the frame rate. An AVI file with a constant frame rate of ten frames per second produced the desired results, in which camera images were analyzed for distinct particles and specimens without redundant output. Early in the project, processing a sequence of TIF files directly was attempted but did not generate the anticipated set of specimens. Since the software calibrated the background from the first 32 images, the order of the images in the AVI file was relevant to the context settings, which was most noticeable when images were collected from several micropaleontological faunal slides.

Camera images were saved as uniquely named TIF files using a convention that began with the site identification code (e.g., SC34) and included a cross reference to the storage location on the associated micropaleontological faunal slide. All digital images pertaining to a collection site were stored together in a file folder and backed up. This facilitated the creation of AVI files with VideoPad. Since the VisualSpreadsheet software manages images separately, it is not required to replicate collections of images, and AVI files can serve as a mechanism to organize specific sets of images for analyses. For example, images of Large Benthic Foraminifers would not have to be copied to another external folder because images already in one or more VisualSpreadsheet list files can be collected into a VisualSpreadsheet library. In addition, a set of external images could be gathered into an AVI file to be processed by VisualSpreadsheet and then classified.

Retrospectively, it would have been beneficial to pad encoded file name sequences with zero characters (e.g., "0001"). For example, I recommend using file names such as "SC34a0001.tif" to "SC34a0100.tif" rather than "SC34a1.tif" to "SC34a100.tif". To maintain consistency, file names should avoid encoded categorical or taxonomic references. However, storing replicated images of similarly identified specimens (e.g., pristine images of *Flintinoides labiosa* used for identification purposes) in a designated directory may provide future advantages. Based upon resolution of technical issues experienced with the initial set of AVI files, I recommend that AVI files be created with a constant frame rate of ten frames per second.

5.5. Creation and Management of Image Libraries

Camoying and Yniguez (2016) recommended that 10–15 characteristic images per taxon be taken for phytoplankton libraries created to optimize classification. A range of 40–100 representative images was suggested for libraries to be utilized with the advanced software package for VisualSpreadsheet mentioned in section 5.8. These quantities were based on flow cytometry. The images I used were collected using a microscope camera. The advantage of this technique was to orient specimens before photographing the image. The software algorithms used to identify images were based on a single orientation and did not include internal characteristics. Without a standard orientation, multiple libraries of a given specimen can be used to create "unique and uniform" image libraries that may improve automated identification (Fluid Imaging Technologies 2014a). Capturing internal features of images to improve automated classification is actively being considered by Fluid Imaging Technologies, Inc. (personal communication 2018). If this enhancement is realized so that foraminiferal features such as

number of chambers and chamber arrangement are captured, superior automated classification results could be attained through image library management. Alternatively, flow cytometry technology could improve to the point where multiple perspectives of a particle are photographed.

Benthic foraminiferal libraries can also serve researchers as a guide in the process of classification. For semi-automated identification utilizing a microscope camera, initial libraries should preferably contain a minimum of 10 images with a target of 40 or more images. Those images should include one taxon with a single, standard orientation of the most recognizable perspective. Supplementary orientations should be photographed and kept as a reference, but should not be mutually incorporated into a morphology-based software library used for automated classifications because uniformity is an important factor for image recognition. Unfortunately, in the case of foraminifera, multiple views are typically required for species-level identification, with such views including both spiral and umbilical sides, where appropriate, as well as characteristics of the aperture

I recommend that images for the software libraries use a uniform magnification when possible. For specimens that exceed the image boundaries, separate processing in a similar manner using a magnification that captures the entire outline of the largest specimen should be performed.

5.6. Evaluation of Image Recognition on Benthic Foraminifers

After converting the library sample to a VisualSpreadsheet list file, the first goal was to see if the software could identify a foraminiferal species. A test library of *Archaias angulatus*

with images of large, medium, and small specimens was created to evaluate the software's ability to select them from the entire list file. Applying the filter "Like Library Particles", whether value based or statistically based, selected more images than the number of *A. angulatus* images in the list file. To attain more uniformity (as recommended by Mr. Nelson, personal communication 2017), the library was partitioned by specimen size. Selected images based on the libraries of medium- and large-sized *A. angulatus* adequately distinguished the limited number of larger specimens in the sample. Smaller *A. angulatus* were not adequately distinguished from many other specimens of similar size. The spiral growth pattern and chambers typically seen in *A. angulatus* specimens were observable, which made the images easily identifiable to myself and other researchers, but no software metrics were associated with these characteristics. Results of applying a filter using libraries of other species, both collectively and based on size, were similar. Results from value-based filters were not compared to those from statistically-based filters. Value-based filters are associated with direct measurements of particles and more narrowly interpreted. Statistically-based filters, on the other hand, employ a model particle derived from a user provided set and then they score the similarity of each particle in the entire set. Tolerances can be applied to available fields as well as to filter score.

Genus-level libraries, as well as libraries for several categories such as functional group, were created to determine if VisualSpreadsheet could accurately classify more general foraminiferal groupings. Not surprisingly, the most abundant genus, *Quinqueloculina*, had the highest auto-classification quantity and accuracy. When combined with other genera, the large number of specimens selected earliest as software class Q - *Quinqueloculina* limited the opportunity for VisualSpreadsheet to accurately classify other genera in the classification

63

scheme. Because auto classification processed each software class sequentially, rearranging software class Q - *Quinqueloculina* from the first position reduced accuracy.

The composite auto classification of functional groups peaked in accuracy at about 71% on the original set of software classes that excluded the filter for small specimens of LBFs in class A. Modifications of the classification scheme, primarily with new filters based on libraries with additional images, did not improve the results of the software as accuracy leveled off at approximately 60%. Therefore, as indicated by the actual reduction in accuracy level, the software was not able to be "trained" in this method to improve its ability to classify benthic foraminiferal assemblages.

5.7. Challenges in Analyses of Benthic Foraminifers

The surface texture of foraminifers with hyaline tests, noticeable in the images, was difficult for the software to classify. With respect to Large Benthic Foraminifers (which were all porcelaneous in this sample set), a number of specimens from site SC159 were not classified because the captured area was incomplete or abnormally shaped. Not classifying degraded specimens could be advantageous by not skewing the data collected. Furthermore, smaller specimens that are degraded and not identifiable should be discarded. Also, exclusions present an opportunity to differentiate foraminiferal tests. For example, the quantity of larger tests could be compared to that of degraded ones in a spatial and temporal analyses.

A major challenge confronting automated image identification of benthic foraminifers is the large number of extant species (Ranaweera et al. 2009a, 2009b). Another problem is misidentification of species, even by researchers, when external features are similar, as illustrated

previously in this paper with regard to *Broeckina orbitolitoides*. However, image recognition technology can be combined with researcher expertise to catalogue foraminiferal assemblages and collect detailed information for further analyses (Gfatter 2016).

5.8. Developments in Technology

The Classifier Advanced, a newer software package for VisualSpreadsheet, overcomes the limitation of sequential processing by the auto-classification feature. The recommended minimum number of images for this feature to work well is also approximately 40 images per library. However, a preliminary test using functional group libraries did not indicate that this feature would considerably improve identification of foraminiferal assemblages. Nonetheless, the inclusion of machine-learning methods and the use of training sets in the software package present advancements in particle classification. Additional comparison of this innovative software package was beyond the scope of this thesis project.

5.9. Research Considerations

Financial limitations are an important concern in many research projects. Several technologies were explored during this study because of availability. The zooSCAN device was tested, but it did not provide the quality of images sought and its associated software was not reviewed. The researchers at the University of Colorado utilized ImageJ, an image processing program, in their methodology, but ImageJ did not produce the AVI files sought after multiple attempts, even though the same steps were followed. VideoPad Video Editor, on the other hand, was able to process the images, but the program functioned very slowly on the computer linked

to the microscope-camera. Faster computing was suggested by exploring the possibility of executing the application on a parallel-processing platform, but the task was less of a multitasking computational issue. Procurement of a personal computer with adequate random access memory (RAM), read-write speed and capacity of the hard disk drive (HDD), graphics card, and processing power of the central processing unit reduced the time to process about 200–250 images into an AVI file from over an hour to approximately ten minutes. Hardware recommendations are ephemeral, but these components should be closely evaluated when considering the use of image processing software. In consideration of VisualSpreadsheet, a 64-bit operating system with 16 gigabytes (GB) of RAM and a 500 GB HDD performed well. Because of the expense of sophisticated instruments and software applications, exhibitions and demonstrations at conventions hosted by organizations such as the Geological Society of America are recommended as a venue to review current equipment and technologies.

5.10. Potential Application: An Example

Research applications that may benefit from the use of VisualSpreadsheet software include further exploration of the study conducted by Beckwith (2016), where the focus was on *Archaias angulatus*, or expansion of the data set in the exploratory work by Gfatter et al. (2016). For example, the sediment sample from site SC120 of the Springs Coast (latitude 28.568, longitude -82.759) contained 230 foraminiferal specimens of which 64 were identified as *A. angulatus*. There was also a single specimen identified as *Broeckina orbitolitoides*, another large benthic foraminifer (LBF), so 65 specimens from the LBF functional group were represented. The sample was processed as described in this study and 212 images were captured by

VisualSpreadsheet. Five of these were not foraminiferal images, so 207 of the 230 foraminiferal specimens were captured. All of the 64 *A. angulatus* images were captured as well as the single image of *B. orbitolitoides*. A classification was created for site SC120 with three software classes: "*Archaias angulatus*", "*Broeckina orbitolitoides*", and "non-LBFs". The auto classification was not implemented. Instead, the 64 *A. angulatus* images, the single *B. orbitolitoides* image, and the other 142 foraminiferal images were placed in their respective software class.

Images in this example were not quality controlled since levels of acceptance should be determined by the objective of the research. Nonetheless, *A. angulatus* data from the trial are presented in Appendix F.

6. Conclusions

1. The assessment of the auto-classification feature in the image recognition software, VisualSpreadsheet™, produced composite accuracies of foraminiferal groups in the range of 60–70% compared to identifications based upon traditional visual identification by a researcher using a stereo-microscope.

2. Image recognition technology intended to assist with the identification of foraminiferal assemblages required more than size (e.g., area, length, and width), morphology (e.g., circularity and symmetry), and other parameters available in the software.

3. Data collected using software such as VisualSpreadsheet could be especially useful in time-series population studies of foraminiferal species, where the specimens of interest are identified, organized, and processed.

4. Curated libraries with high quality images will become more beneficial as image recognition technology advances.

5. Acquisition of consistent images requires a reliable photographic surface and a background with a sharp contrast and minimal reflectance.

6. The most comprehensive identifications are made by researchers when photographed, but semi-automated classifications are aided with libraries of images having a single orientation showing the most diagnostic criteria.

7. Digital images should be saved in a lossless compression format so that their full, original quality will be preserved.

8. Audio Video Interleave multimedia container files created for processing by VisualSpreadsheet should have a constant frame rate of ten frames per second.

9. Image libraries utilized for automated classifications should contain 40–100 images using several representative specimens.

Literature Cited

Álvarez, E., Á. López-Urrutia, E. Nogueira, and S. Fraga. 2011. How to effectively sample the plankton size spectrum? A case study using FlowCAM. *Journal of Plankton Research* 33 (7):1119-1133.

Álvarez, E., M. Moyano, Á. López-Urrutia, E. Nogueira, and R. Scharek. 2014. Routine determination of plankton community composition and size structure: A comparison between FlowCAM and light microscopy. *Journal of Plankton Research* 36 (1):170-184.

Beckwith, S.T. 2016. Abundance of *Archaias angulatus* on the West Florida Coast indicates the influence of carbonate alkalinity over salinity. Master of Science Master's thesis, College of Marine Science, University of South Florida.

Bock, W.D., W.W. Hay, J.I. Jones, G.W. Lynts, S.L. Smith, and R.C. Wright. 1971. *A Symposium of Recent South Florida Foraminifera*. Miami: Miami Geological Society.

Camoying, M.G., and A.T. Yniguez. 2016. FlowCAM optimization: Attaining good quality images for higher taxonomic classification resolution of natural phytoplankton samples. *Limnology and Oceanography: Methods* 14 (5):305-314.

Carnahan, E.A., A.M. Hoare, P. Hallock, B.H. Lidz, and C.D. Reich. 2009. Foraminiferal assemblages in Biscayne Bay, Florida, USA: Responses to urban and agricultural influence in a subtropical estuary. *Marine Pollution Bulletin* 59 (8–12):221-233.

Fluid Imaging Technologies. 2014a. FlowCAM® image libraries for aquatic samples. www.fluidimaging.com.

Fluid Imaging Technologies. 2014b. FlowCAM® manual. www.fluidimaging.com.

Fluid Imaging Technologies. 2017. FlowCAM® VS-series specifications.
http://info.fluidimaging.com/hs-fs/hub/300163/file-2220046276-pdf/documents/Flowcam-VS-Specification.pdf?t=1493324581565.

Folk, R.L. 1980. *Petrology of Sedimentary Rocks*. Austin: Hemphill Publishing Company.

García-Muñoz, C., L.M. Lubián, C.M. García, A. Marrero-Díaz, P. Sangrà, and M. Vernet. 2013. A mesoscale study of phytoplankton assemblages around the South Shetland Islands (Antarctica). *Polar Biology* 36 (8):1107-1123.

Gfatter, C. 2016. Utilizing image recognition technology for foraminiferal assemblage analyses. 2016 GSA Annual Meeting, Denver, September 25, 2016. https://gsa.confex.com/gsa/2016AM/webprogram/Paper283336.html.

Gfatter, C., K. Amergian, S. Beckwith, and P. Hallock. 2016. Refugia for carbonate producing organisms in high carbon dioxide environmental conditions. 2016 AGU Fall Meeting, San Francisco, December 16, 2016. https://agu.confex.com/agu/fm16/meetingapp.cgi/Paper/164144.

Hallock, P. 2012. The FORAM Index revisited: Uses, challenges, and limitations. 12th International Coral Reef Symposium, Cairns, Australia, July 10, 2012. http://www.icrs2012.com/proceedings/manuscripts/ICRS2012_15F_2.pdf.

Hallock, P., B.H. Lidz, E.M. Cockey-Burkhard, and K.B. Donnelly. 2003. Foraminifera as bioindicators in coral reef assessment and monitoring: The FORAM Index. *Environmental Monitoring and Assessment* 81 (1):221-238.

Holzmann, M., J. Hohenegger, P. Hallock, W.E. Piller, and J. Pawlowski. 2001. Molecular phylogeny of large miliolid foraminifera (Soritacea Ehrenberg 1839). *Marine Micropaleontology* 43 (1):57-74.

Liu, S., M. Thonnat, and M. Berthod. 1994. Automatic classification of planktonic-foraminifera by a knowledge-based system. Tenth Conference on Artificial Intelligence for Applications, Proceedings, Los Alamitos. http://ieeexplore.ieee.org/document/323653.

Murray, J.W. 2006. *Ecology and Applications of Benthic Foraminifera*. Cambridge: Cambridge University Press.

Murray, J.W., and S.S. Bowser. 2000. Mortality, protoplasm decay rate, and reliability of staining techniques to recognize 'living' foraminifera: A review. *Journal of Foraminiferal Research* 30 (1):66-70.

Poag, C.W. 2015. *Benthic Foraminifera of the Gulf of Mexico: Distribution, Ecology, Paleoecology*. College Station: Texas A&M University Press.

Ranaweera, K., S. Bains, and D. Joseph. 2009a. Analysis of image-based classification of foraminiferal tests. *Marine Micropaleontology* 72 (1-2):60-65.

Ranaweera, K., A.P. Harrison, S. Bains, and D. Joseph. 2009b. Feasibility of computer-aided identification of foraminiferal tests. *Marine Micropaleontology* 72 (1-2):66-75.

See, J.H., L. Campbell, T.L. Richardson, J.L. Pinckney, R.J. Shen, and N.L. Guinasso. 2005. Combining new technologies for determination of phytoplankton community structure in the northern Gulf of Mexico. *Journal of Phycology* 41 (2):305-310.

Spaulding, S.A., D.H. Jewson, R.J. Bixby, H. Nelson, and D.M. McKnight. 2012. Automated measurement of diatom size. *Limnology and Oceanography: Methods* 10:882-890.

Walton, W.R. 1952. Techniques for recognition of living foraminifera. *Contributions from the Cushman Foundation for Foraminiferal Research* 3:56-60.

Yu, S., P. SaintMarc, M. Thonnat, and M. Berthod. 1996. Feasibility study of automatic identification of planktic foraminifera by computer vision. *Journal of Foraminiferal Research* 26 (2):113-123.

Appendices

Appendix A: Authors of Foraminiferal Species

The authors of the foraminifers discussed in the study are listed in Table A.1. Foraminifers only referenced by a filter created from their VisualSpreadsheet™ image library are highlighted in gray.

Table A.1: Foraminiferal species reference

Species	Author(s)	Year
Archaias angulatus	Fichtel & Moll	1798
Broeckina orbitolitoides	Hofker	1930
Flintinoides labiosa	d'Orbigny	1839
Miliolinella circularis	Bornemann	1855
Quinqueloculina candeiana	d'Orbigny	1839
Cycloforina sidebottomi	Rasheed	1971
Haynesina germanica	Ehrenberg	1840
Pseudotriloculina rotunda	d'Orbigny (in Schlumberger)	1893
Quinqueloculina crassa	d'Orbigny	1850
Quinqueloculina impressa	Reuss	1851
Quinqueloculina linneiana	d'Orbigny	1839
Quinqueloculina poeyana	d'Orbigny	1839
Triloculina bermudezi	Acosta	1940
Triloculina trigonula	Lamarck	1804

Appendix B: VisualSpreadsheet Image Libraries

        The VisualSpreadsheet™ libraries shown in the following tables (Table B.1 to Table B.4) are comprised of multiple files, so the ".flb" extension of the primary library file was excluded. The first seven species libraries shown in Table B.1 are further partitioned into size-based libraries shown in Table B.2. Experimental species libraries presented in Table B.1 are highlighted in dark gray. Tentative genus libraries without images listed in Table B.3 are also highlighted in dark gray. Lastly, the four functional group image libraries in Table B.4 are highlighted in color followed by data for the hyaline surface texture image library.

        Example VisualSpreadsheet libraries of *Archaias angulatus* images are illustrated in Figure B.1. The images in the size-based libraries on the right side of the figure are from the initial comprehensive set.
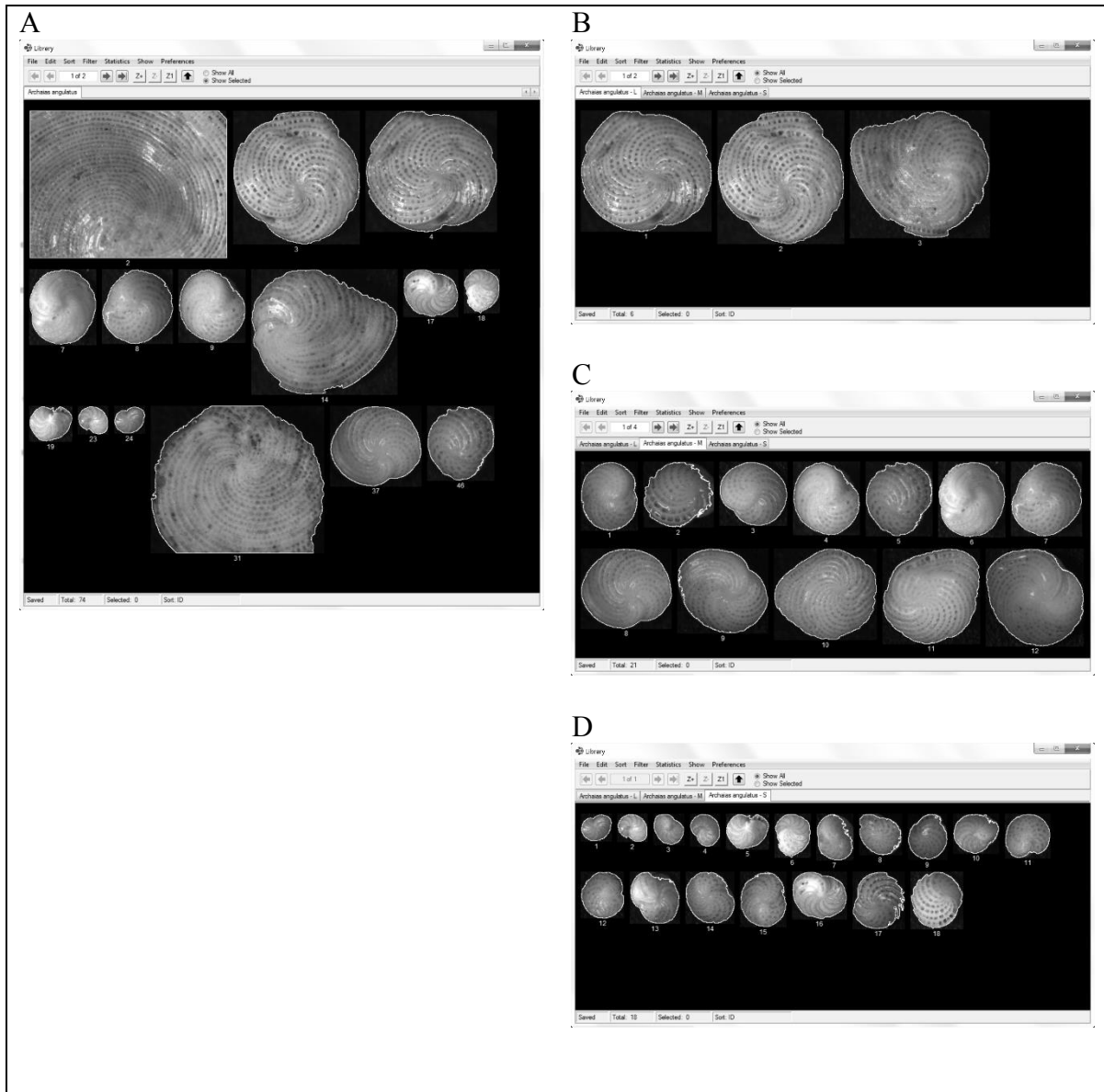
Figure B.1: **A)** Master VisualSpreadsheet library of *Archaias angulatus* images;
B-D) VisualSpreadsheet libraries of large, medium, and small specimens

Table B.1: Species image libraries

VisualSpreadsheet Library	Group	Specimens	Images
Archaias angulatus	A	34	74
Flintinoides labiosa	B	26	80
Triloculina trigonula	B	17	51
Quinqueloculina candeiana	B	15	46
Quinqueloculina linneiana	B	14	42
Pseudotriloculina rotunda	B	12	36
Miliolinella circularis	B	11	32
Cycloforina sidebottomi	B	6	19
Quinqueloculina crassa	B	6	18
Triloculina bermudezi	B	6	18
Haynesina germanica	D	6	18
Quinqueloculina impressa	B	5	15
Quinqueloculina poeyana	B	5	15
Cribroelphidium poeyanum	D	3	9
Palmerinella palmerae	C	3	10
Quinqueloculina bicostata	B	3	9
Triloculina oblonga	B	3	9
Ammonia tepida	D	2	7
Elphidium discoidale	D	2	6
Haynesina depressula	D	2	6
Quinqueloculina bosciana	B	2	7
Quinqueloculina costata	B	2	7
Quinqueloculina lamarckiana	B	2	6
Quinqueloculina striata	B	2	6
Quinqueloculina tipswordi	B	2	6
Triloculina variolata	B	2	6
Broeckina orbitolitoides	A	1	2
Elphidium advenum	D	1	3
Elphidium galvestonense	D	1	3
Miliolinella suborbicularis	B	1	3
Miliolinella subrotunda	B	1	2
Pyrgo williamsoni	B	1	3
Quinqueloculina bicarinata	B	1	3
Quinqueloculina laevigata	B	1	3
Wiesnerella auriculata	B	1	3

Table B.2: Species image libraries based on size parameter, Area Based Diameter (ABD)

VisualSpreadsheet Library	ABD Range (μm^2)		Images
Archaias angulatus - XLMS	10,000	1,100,000	57
Archaias angulatus - X	750,000	1,100,000	12
Archaias angulatus - LMS	10,000	750,000	45
Archaias angulatus - L	350,000	750,000	6
Archaias angulatus - M	100,000	350,000	21
Archaias angulatus - S	10,000	100,000	18
Flintinoides labiosa - L	25,000	115,000	58
Flintinoides labiosa - M	9,000	25,000	16
Flintinoides labiosa - S	2,000	9,000	6
Triloculina trigonula - L	27,000	100,000	20
Triloculina trigonula - M	11,000	27,000	15
Triloculina trigonula - S	2,000	11,000	16
Quinqueloculina candeiana - L	48,000	100,000	17
Quinqueloculina candeiana - M	26,000	48,000	19
Quinqueloculina candeiana - S	7,000	26,000	10
Quinqueloculina linneiana - L	30,000	100,000	8
Quinqueloculina linneiana - M	10,000	30,000	18
Quinqueloculina linneiana - S	2,000	10,000	16
Pseudotriloculina rotunda - M	50,000	115,000	17
Pseudotriloculina rotunda - S	10,000	50,000	19
Miliolinella circularis - M	10,000	50,000	24
Miliolinella circularis - S	2,000	10,000	8

Table B.3: Genus image libraries

VisualSpreadsheet Library (partitions)	Group	Specimens	Available Images	Images
Quinqueloculina - large - medium - small	B	65	198	74 30 20 24
Archaias - large - medium - small	A	34	74	45 6 21 18
Triloculina - large - medium - small	B	29	87	68 18 15 35
Flintinoides - large - medium - small	B	26	80	53 35 12 6
Miliolinella - medium - small	B	13	37	37 24 13
Pseudotriloculina - medium - small	B	12	36	25 11 14
Haynesina	D	8	24	16
Cycloforina	B	6	19	
Elphidium	D	6	18	
Cribroelphidium	D	3	9	
Palmerinella	C	3	10	
Ammonia	D	2	7	
Broeckina	A	1	2	
Pyrgo	B	1	3	
Wiesnerella	B	1	3	

Table B.4: Group image libraries and hyaline surface texture image library

VisualSpreadsheet Library	Group	Specimens	Available Images	Images
Large Benthic Foraminifers	A	35	76	36
> 400,000 µm^2	- large			8
> 100,000 µm^2	- medium			14
> 20,000 µm^2	- small			14
Smaller Miliolid Foraminifers	B	153	463	140
> 50,000 µm^2	- large			34
> 20,000 µm^2	- medium			46
	- small			60
> 7,500 µm^2	- small-a			45
> 2,500 µm^2	- small-b			15
Other Smaller Foraminifers	C	3	10	8
> 20,000 µm^2	- medium			2
> 5,000 µm^2	- small			6
Stress Tolerant Foraminifers	D	19	58	30
> 7,500 µm^2	- medium			18
> 2,500 µm^2	- small			12
Hyaline Surface Texture		22	68	38

Appendix C: Protocol for Springs Coast Samples

    I.   **Procedures for Washing and Picking Sediments for Foraminifers**

Take a sample out of the freezer to thaw.
Each sample will be divided into 4 parts:

      1. For grain-size analysis and foraminifers counts

      2. For "live" foram assemblage analysis (stained)

      3. Return to freezer for later analysis of organic components

      4. Return to freezer in case an analysis needs to be redone

A) Remove several samples from the freezer and allow to thaw just enough so the sample can be subdivided. Divide with quartering tool, using spatula to put one half of a quarter portion onto a labeled filter paper, and the rest into a 63 μm mesh sieve. **Return samples to freezer** (for 2-3-4 above).
1) Weigh a labeled coffee filter then place the subsample on the labeled coffee filter in the hood to dry overnight.

Sediment Analysis

B) Start a blank data sheet for the sediment subsample.
2) Weigh dried subsample, recording weight of sample+filter.
3) Place dry subsample in beaker.
4) Weigh and record weight of the dry empty filter, subtract weight from sample+filter weight.
5) Add deionized water (DIW) to beaker with dry subsample and place in ultrasound (containing enough DIW to stabilize beaker) for 5 minutes to disaggregate.
6) Wash with DIW over a 63 μm mesh sieve to remove the mud.
7) Wash subsample with a squirt bottle into the labeled coffee filter and dry again in the hood.

C) The dry sand-sized fraction is then separated by grain size:
8) Weigh remaining sediment and record as "uncorrected sand fraction"
9) Weigh each of the sieves (2 mm – pan), recording weights on sieve mass section of data sheet.
10) Place subsample into the tower of sieves and set on shaker for 10 minutes at medium setting.
11) Reweigh each sieve with retained sediment, recording weights in the sediment+sieve section of data sheet.
    Weight of mud in the "pan" must be added to the weight of the mud washed out;
i.e., difference between weight of original dry subsample and the weight of washed subsample.
    Weight-percent of each size fraction is then calculated and the median phi size is determined.
12) Retain the sediment contained in the sieves (mud in pan can be discarded) and place into a whirl-pak bag labeled with the sample name and "**sediment portion**".

Picking and Archiving

D) The sediments in the 63 µm mesh sieve are cleaned & dried as previously described in step B and then are ready for picking, or the cleaned sediments in the whirl-pak bag noted in step C can be picked (sediment archive is finished).

- Thoroughly mix the dry subsample and take ≈ 1 gram of sediment from subsample.
- From this one gram, measure out increments of sediment of approximately 0.100 grams onto a pre-weighed tray and weigh, recording the weight of the sediment to be examined, i.e., sediment weight = tray+sediments - tray weight.
- Sprinkle this amount over a gridded tray, and examine under a stereomicroscope for foraminifers. Pick **all** foraminiferal specimens using a fine paintbrush onto a pre-weighed "holding" tray (e.g., tray having a brushed metal surface).
- Set aside picked sediment and repeat previous step until 200 foraminifers are collected, completing the picking of the current tray.
- Weigh the "holding" tray with the forams, recording the weight of forams picked; i.e., foram weight = tray+forams - "holding" tray weight.
  Weight-percent of forams is calculated: foram weight ÷ Σ(examined sediment weights); i.e., foram weight ÷ (foram weight + picked sediment weight).

E) Photograph foraminifers using the AxioCam software.

- Photograph multiple views of a foraminifer (4.0x magnification) for the image library. Foraminifer is then coated with water-soluble glue and placed onto the labeled slide.

  or

- Group foraminifers of similar size and glue them with water-soluble glue onto the labeled slide. Photograph foraminifers (4.0x magnification) for site samples.

The <u>goal is 150-200 foraminifers</u>; if this is not obtained in the first portion, the procedure is repeated with subsequent weighed subsamples until a total of 150-200 specimens are isolated or until the 1 gram of sediment is examined. If the 150-200 foraminifers are obtained and the sediment placed on the gridded tray has not been entirely examined, that measured amount of sediment **MUST** be continued to be picked for foraminifers until completed, no matter how many foraminifers over 150 are collected. Once the sample has been completely processed, place the whole gram into a whirl-pak bag labeled with the sample name and "**foram sample**".

II.   Trichloroethylene (C$_2$HCl$_3$) procedure for separating foraminifers from sediment

Utilize C$_2$HCl$_3$ from the bottle labeled "used" first, obtaining more from "new" bottle as needed.
* Perform work involving trichloroethylene **in a hood and wear latex gloves**!

1) Label one filter paper "float" with the sample ID & second filter paper "residual" with the sample ID.

2) Label two 300 ml beakers "float" & "residual" tall enough to contain the stem of a glass (or metal) funnel so it rests above the bottom of the beaker (Figure C.1), which will allow the trichloroethylene to collect there.

3) Fold and insert each filter paper into a funnel, and then place each into their respective beaker.

4) Place the dried sediment sample into a separate beaker.

Figure C.1: Two labeled beakers set up for the trichloroethylene procedure

5) **Put on latex gloves** and take the three beakers
to **the hood**. So the solution does not saturate the forams, work promptly and carefully. Pour some trichloroethylene into the beaker with the sample & gently swirl, and then decant solution with floating particles into the "float" apparatus. Repeat 2-3 times, ensuring no scum is left on the sides of the beaker. Decant remainder of the particles into the "residual" apparatus.

6) Pour excess C$_2$HCl$_3$ from the beakers into the bottle labeled "USED trichloroethylene" for future use. Let the beakers & filter paper dry before removing them from the hood. Float sample (F-sample) and residual sample (R-sample) are now ready to be picked (or stored).

7) Weigh and record both unpicked samples, subtracting their filter paper weight. Place F-sample on a gridded tray. Select random squares and pick each of them completely until 200 forams are collected, also picking the last square in its entirety. Place R-sample on a gridded tray. Select the same squares as the F-sample and pick each of them completely. Weigh and record both picked samples, subtracting their tray weight.

Appendix D: Grain-Size Analysis

Grain-size and cumulative grain-size percentages are presented in Table D.1 and Table D.2 respectively.

Table D.1: Grain-size analysis of the Florida Springs Coast sample sites

		Phi sizes						
		-1	**0**	**1**	**2**	**3**	**4**	**> 4**
Sample ID	**Sample wt. (g)**	**> 2 mm**	**> 1 mm**	**> 0.5 mm**	**> 0.25 mm**	**> 0.125 mm**	**> 0.063 mm**	**< 0.063 mm**
SC034	**7.407**	6.63%	9.25%	8.68%	6.25%	34.22%	17.96%	17.01%
SC040	**9.112**	4.81%	7.21%	5.21%	5.09%	50.26%	16.31%	11.11%
SC046	**7.350**	4.20%	2.23%	2.29%	3.63%	39.71%	23.90%	24.03%
SC052	**10.589**	0.00%	0.01%	0.12%	0.70%	73.90%	21.42%	3.85%
SC088	**13.721**	7.28%	11.73%	11.96%	8.63%	36.73%	19.73%	3.94%
SC094	**4.467**	2.53%	7.07%	20.17%	20.21%	38.93%	5.78%	5.31%
SC100	**6.446**	3.24%	5.41%	14.04%	22.39%	32.14%	10.81%	11.96%
SC106	**7.283**	3.94%	5.59%	11.99%	15.63%	31.18%	20.14%	11.53%
SC135	**9.745**	3.06%	3.86%	3.96%	5.41%	25.44%	45.22%	13.05%
SC141	**4.382**	10.57%	2.81%	6.16%	16.34%	24.40%	16.00%	23.73%
SC147	**3.255**	4.39%	6.64%	26.42%	27.53%	15.70%	6.79%	12.53%
SC159	**3.756**	31.10%	12.41%	22.26%	15.23%	6.68%	3.17%	9.16%
SC165	**7.403**	6.12%	3.50%	6.44%	13.37%	42.69%	18.11%	9.77%
SC177	**19.441**	0.61%	2.41%	4.58%	7.34%	51.77%	26.67%	6.62%

Median grain size for each sample is highlighted

Table D.2: Cumulative grain-size analysis of the Florida Springs Coast sample sites

		Phi sizes						
		-1	0	1	2	3	4	> 4
Sample ID	Sample wt. (g)	> 2 mm	> 1 mm	> 0.5 mm	> 0.25 mm	> 0.125 mm	> 0.063 mm	< 0.063 mm
SC034	7.407	6.63%	15.88%	24.56%	30.81%	65.03%	82.99%	100.0%
SC040	9.112	4.81%	12.02%	17.23%	22.32%	72.58%	88.89%	100.0%
SC046	7.350	4.20%	6.43%	8.72%	12.35%	52.06%	75.96%	100.0%
SC052	10.589	0.00%	0.01%	0.13%	0.83%	74.73%	96.15%	100.0%
SC088	13.721	7.28%	19.01%	30.97%	39.60%	76.33%	96.06%	100.0%
SC094	4.467	2.53%	9.60%	29.77%	49.98%	88.91%	94.69%	100.0%
SC100	6.446	3.24%	8.65%	22.69%	45.08%	77.22%	88.03%	100.0%
SC106	7.283	3.94%	9.53%	21.52%	37.15%	68.33%	88.47%	100.0%
SC135	9.745	3.06%	6.92%	10.88%	16.29%	41.73%	86.95%	100.0%
SC141	4.382	10.57%	13.38%	19.54%	35.88%	60.28%	76.28%	100.0%
SC147	3.255	4.39%	11.03%	37.45%	64.98%	80.68%	87.47%	100.0%
SC159	3.756	31.10%	43.51%	65.77%	81.00%	87.68%	90.85%	100.0%
SC165	7.403	6.12%	9.62%	16.06%	29.43%	72.12%	90.23%	100.0%
SC177	19.441	0.61%	3.02%	7.60%	14.94%	66.71%	93.38%	100.0%

Median grain size for each sample is highlighted

Appendix E: Original Classification for Testing VisualSpreadsheet Software

The outline for the VisualSpreadsheet™ classification template used to generate the initial results is presented in Table E.1. All initial images originated from specimens found in the site sample, LibSC34, designated solely for the original software libraries and associated filters.

Table E.1: VisualSpreadsheet classification template

Position	Class Name	Statistical Filters	Initial # of Images
1	A - Large Benthic	Group A - M Group A - L	14 8
2	B - Smaller Miliolid	Group B - S Group B - M Group B - L	60 46 34
3	C - Other Smaller	Group C - S Group C - M	6 2
4	D - Stress Tolerant	Group D - S Group D - M	12 18

Appendix F: Example of *Archaias angulatus* data from site SC120

      The *A. angulatus* images captured by VisualSpreadsheet™ and manually classified are illustrated in Figure F.1. Summary data are shown in Table F.1, but data for Sphere Complement, Sphere Count, Sphere Unknown, and Sphere Volume were omitted because all of those values were zero. Data for each specimen were also available. In this example, no images were removed from the c, including specimens that were not completely traced by the software. Researchers should specify the acceptable tolerances *a priori*.
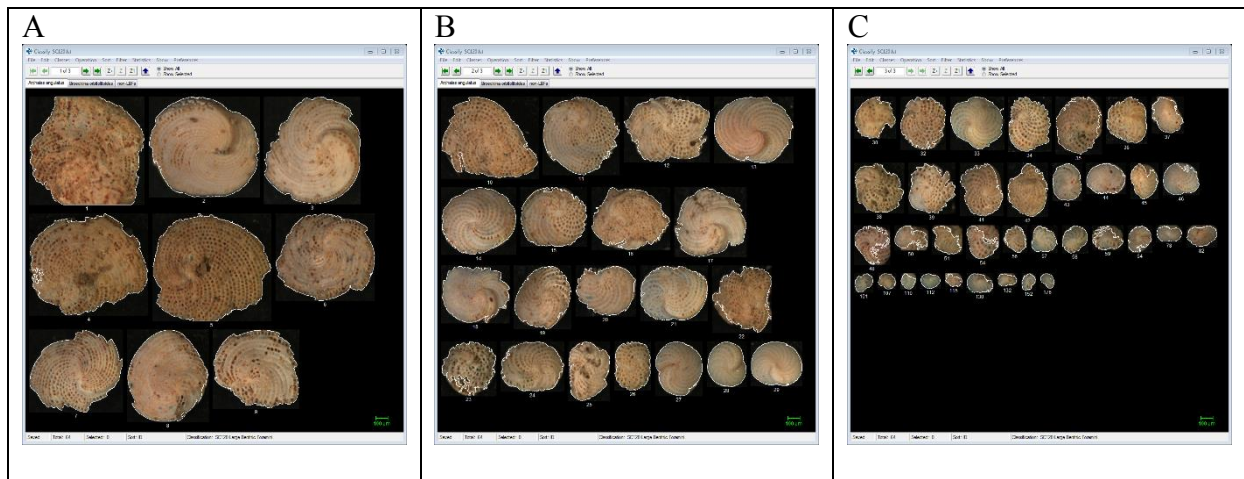

Figure F.1: **A-C**) *A. angulatus* images in the VisualSpreadsheet classification for site SC120

Table F.1: Summary data of *A. angulatus* in the VisualSpreadsheet classification for site SC120

List Name	SC120.lst				
Classification	SC120 Large Benthic Foraminifers				
Class	*Archaias angulatus*				
Count	64				
Particles / ml	6356				
Summary Stats	**Mean**	**Min**	**Max**	**StdDev**	**% CV**
Area (ABD)	1.51E+05	8328.17	6.28E+05	1.55E+05	102.65
Area (Filled)	1.58E+05	8437.75	6.73E+05	1.66E+05	104.63
Aspect Ratio	0.8	0.58	0.97	0.09	11.16
Capture X	552.38	251	1121	262.13	47.45
Capture Y	362.03	0	793	192.71	53.23
Circle Fit	0.53	0	0.8	0.21	40.48
Circularity	0.5	0.1	0.81	0.17	33.25
Circularity (Hu)	0.94	0.75	0.99	0.05	5.26
Compactness	2.44	1.24	10.36	1.57	64.35
Convex Perimeter	1306.86	357.12	3064.97	702.27	53.74
Convexity	0.94	0.78	0.99	0.05	4.99
D[3 2](ABD)	594.55	102.97	894.43	203.96	34.3
D[3 2](ESD)	636.59	111.99	973.75	223.12	35.05
D[4 3](ABD)	664.52	102.97	894.43	182.97	27.53
D[4 3](ESD)	714.79	111.99	973.75	203.51	28.47
Diameter (ABD)	384.73	102.97	894.43	210.62	54.75
Diameter (ESD)	414.27	111.99	973.75	223.54	53.96
Diameter (FD)	392.69	103.65	925.93	217.24	55.32
Edge Gradient	31.92	17.22	61.43	9.55	29.92
Elapsed Time	2.69	0	5.9	1.47	54.61
Elongation	5.36	1	30.52	5.04	94.04
Fiber Curl	0.65	0	2.91	0.56	86.23
Fiber Straightness	0.67	0.26	1.33	0.21	30.48
Geodesic Aspect Ratio	0.3	0.03	1	0.2	66.58
Geodesic Length	727.94	97.04	1995.16	394.98	54.26
Geodesic Thickness	184.8	34.23	529.56	121.52	65.76
Image Height	502.39	137	1080	255.88	50.93
Image Width	518.13	146	1190	280.96	54.23
Intensity	107.31	80.47	128.21	10.89	10.14
Length	452.89	127.77	1080.48	239.84	52.96
Particles Per Chain	1	1	1	0	0
Perimeter	1825.48	388.15	4582.47	943.68	51.69
Roughness	1.42	1.09	2.75	0.31	22.16
Sigma Intensity	19.5	12.5	26.7	3.5	17.93
Sum Intensity	2.41E+07	1.29E+06	9.43E+07	2.39E+07	99.04
Symmetry	0.65	0.02	0.9	0.19	28.31
Transparency	0.08	0.02	0.2	0.04	54.17
Volume (ABD)	5.99E+07	5.72E+05	3.75E+08	8.72E+07	145.58
Volume (ESD)	7.39E+07	7.35E+05	4.83E+08	1.09E+08	148.02
Width	367.36	88.9	887.78	207.24	56.41