DOI: 10.1111/2041-210X.12915

APPLICATION

AutoMorph: Accelerating morphometrics with automated 2D and 3D image processing and shape extraction

Allison Y. Hsiang ^{1,2} 💿 🍴	Kaylea Nelson ¹	Leanne E. Elder ¹	Elizabeth C. Sil	b ert^{3,4}
Sara S. Kahanamoku ^{1,5}	Janet E. Burke ¹	Abigail Kelly ⁶	Yusu Liu ⁷ Pin	celli M. Hull ¹

¹Department of Geology and Geophysics, Yale University, New Haven, CT, USA; ²Department of Bioinformatics and Genetics, Swedish Museum of Natural History, Stockholm, Sweden; ³Harvard Society of Fellows, Harvard University, Cambridge, MA, USA; ⁴Department of Earth and Planetary Sciences, Harvard University, Cambridge, MA, USA; ⁵Department of Integrative Biology and Museum of Paleontology, University of California, Berkeley, CA, USA; ⁶Smithsonian Tropical Research Institute, Balboa, Panama and ⁷Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

Correspondence Pincelli M. Hull Email: pincelli.hull@vale.edu

DNI8

Funding information American Chemical Society Petroleum Research Fund, Grant/Award Number: 55837-

Handling Editor: Robert Freckleton

Abstract

- 1. Large-scale, comparative studies of morphological variation are rare due to the time-intensive nature of shape quantification. This data gap is important to address, as intraspecific and interspecific morphological variation underpins and reflects ecological and evolutionary processes.
- 2. Here, we detail a novel software package, *AutoMorph*, for high-throughput object and shape extraction. *AutoMorph* can batch image many types of organisms (e.g. foraminifera, molluscs and fish teeth), allowing for rapid generation of assemblagescale morphological data.
- 3. We used *AutoMorph* to image and generate 2D and 3D morphological data for >100,000 marine microfossils in about a year. Our collaborators have used *AutoMorph* to process >12,000 patellogastropod shells and >50,000 fish teeth.
- 4. AutoMorph allows users to rapidly produce large amounts of morphological data, facilitating community-scale evolutionary and ecological studies. To hasten the adoption of automated approaches, we have made AutoMorph freely available and open source. AutoMorph runs on all UNIX-like systems; future versions will run across all platforms.

KEYWORDS

automated data extraction, bivalves, foraminifera, geometric morphometrics, ichthyoliths, macrofossils, microfossils, patellogastropoda, photogrammetry, virtual palaeontology

1 | INTRODUCTION

In spite of their potential utility, 3D-morphological approaches have yet to be widely applied to evolutionary studies of intraspecific and interspecific variation, community assemblages and comparative phylogenetics (Cunningham, Rahman, Lautenschlager, Rayfield, & Donoghue, 2014). This is primarily because such studies traditionally use laborious landmark and outline-based geometric morphometric techniques, extracting a specific subset of morphological information (i.e. a set of coordinates placed on or along specific features) (Bookstein, 1997; Slice, 2007; Zelditch, 2004). As a result, data-dense geometric morphometric time series of microevolution (e.g. Pearson & Ezard, 2014), comparative morphological analyses using 3D data or biomechanical simulations (e.g. Allen, Bates, Li, & Hutchinson, 2013; Balanoff, Bever, Rowe, & Norell, 2013) and community-scale investigations of morphological dynamics (as proposed by Dietl, 2013; Polly et al., 2011) are relatively rare.

Here, we describe AutoMorph, a novel software package for batchprocessing images of thousands of specimens per day by extracting 2D and 3D data that can be directly used in morphometric and geometric morphometric analyses. Our software takes advantage of image stacks taken at different focal planes using serial focus light microscopy to cheaply and rapidly generate external 3D shells of objects (Falkingham, 2012, 2013; Petti et al., 2008; Sutton, Rahman, & Garwood, 2014). Although our in-house pipeline focuses on foraminifera and serial focus light microscopy, AutoMorph can readily batch process other fossils and objects (e.g. fish teeth, shells, bones) photographed using serial focusing techniques.

2 | MATERIALS AND METHODS

AutoMorph consists of four modules—segment, focus, run2dmorph and run3dmorph—which together comprise a complete dataextraction pipeline from images to 2D and 3D measurement data (Figure 1). The general pipeline is as follows:

- Objects are arranged on a dark background and imaged, producing z-stack images and a 2D extended-depth-of-focus (EDF) image (i.e. a composite image of the in-focus sections from different focal depths, reducing blur across a surface that crosses multiple focal planes) of the entire slide (Figure 1.1 and 1.2).
- Segment identifies all objects in raw z-stack images, extracting and cropping each object such that the z-stack of each object is saved in its own subdirectory (Figure 1.2 to 1.3).
- **3.** *Focus* outputs an EDF image for each object identified by *segment* using focus stacking implemented in either a commercial software package (Zerene Stacker LLC, 2016) or the open-source software *FIJI* (Schindelin et al., 2012) (Figure 1.3).
- **4.** *Run2dmorph* applies a series of image filters to each 2D EDF image outputted by *focus* and returns an object outline (*x*,*y*-coordinates in CSV format) and 2D shape measurements including area, perimeter and aspect ratio (Figure 1.4).
- Run3dmorph extracts a 3D mesh of each object, along with estimates of surface area and enclosed volume, using the z-stacks isolated by segment (Figure 1.4).



FIGURE 1 Overview of highthroughput imaging workflow: sample preparation, imaging, image processing, 2D and 3D shape extraction, and data storage

AutoMorph is written in Python and is currently compatible only with Python 2.7. Although Python 3 is the most up-to-date version of Python, we have written AutoMorph in Python 2.7 due to its large user base and community. Future Python 3 compatibility for AutoMorph is planned. AutoMorph is under active development and available on GitHub (https://github.com/HullLab/AutoMorph).

2.1 | Segment: Image segmentation and labelling

Segment takes as input z-stack images (JPEG/TIFF, including BigTIFFs) of objects arranged on a dark background. Segment extracts all objects from the original slide and saves individual, labelled z-stacks for all identified objects (Figures 2 and 3). Segment is called using the following command:

segment <control file>

The second argument is the path to a control file, a text file containing the user settings that control *segment*'s behaviour. Detailed information about *segment*'s parameters can be found in the Data S1 and in the *AutoMorph* tutorials on GitHub. The most important parameters for controlling *segment*'s ability to find individual objects are as follows: (1) the *threshold* parameter, which is used when converting the source image to black and white in order to identify objects; and (2) the *maximum_size* and *minimum_size* parameters, which stipulate the size range in pixels within which identified objects should fall. This allows users to easily eliminate background noise. AutoMorph allows users to test multiple *threshold* and size values before running the complete segmentation process by specifying *sample* mode in the control file. *Sample* mode returns low-resolution JPEGs of the original input image with red boxes drawn around all objects identified by *segment* (Figure 2.1). One JPEG per *threshold* and size value combination is generated. This allows users to identify the best *threshold* and size values for their sample before committing to the full analysis, saving computation time and resources.

2.2 | Focus: Image stack re-focusing

Focus takes the z-stacks of each object identified by *segment* and produces a labelled EDF image using either *Zerene Stacker* (proprietary) or *FIJI* (free; Figure 2.2 to 2.3). Based on our experience in-house, the EDFs generated by *Zerene Stacker* are generally of higher quality. *Focus* is called via the following command:

focus <directory>

where <directory> is the path to the "final" directory generated by *segment*, which contains the individual z-stack subdirectories for all the objects identified. *Focus* can be controlled by a configuration file, which must be located in the called <directory>. In the absence of a configuration file, *focus* runs using a default configuration file in the installation directory. *Focus* also accepts additional arguments, such as *--reset*, which restores the input directory to its original, pre-focused state. These optional arguments, along with the details regarding the configuration file, are described in the tutorial files found on GitHub.



FIGURE 2 Image segmentation via *segment*. (2.1) *sample* mode outputs a low-resolution JPEG of the entire slide with red boxes drawn around every identified object (two example close-up regions shown to emphasize red lines). (2.2) In *final* mode, *segment* returns labelled image stacks for every unique object. (2.3) *focus* refocuses every object using *Zerene Stacker* or *FIJI* to produce a single-best 2D object EDF (extended-depth-of-focus) image





FIGURE 3 Example label, colour coded by information type. All unique objects are given an object number. Printed on every object image (including layers) is the full sample information needed to uniquely identify the fossil (including catalog number [linked to a full sample database entry], site name and sieve size fraction, slide number, image magnification, and z-slice or EDF identification) and to use the image output for research (e.g. information pertaining to image acquisition and processing, and full sample size). We propose this label as a working standard for high-throughput image products

2.3 | Run2dmorph: 2D measurements and outlines

Run2dmorph performs batch extraction of object outlines, x,ycoordinates and basic measurements of 2D shape. 2D shape parameters and outline coordinates are extracted from black-to-white thresholded images produced by a series of image filters (Figure 4) on the EDF images outputted by *focus*. More details about *run2dmorph*'s image filters can be found in the Data S1. Run2dmorph is called using the following command:

run2dmorph <control file>

Run2dmorph's parameters are set via a control file. These parameters include controls for the various image filters, such as *disk_ size_opening* (which controls the size of the disc used to perform morphological opening) and *contrast_adjustment* (which controls the strength of contrast enhancement). Details about all parameters can be found in the *AutoMorph* tutorials on GitHub. *Run2dmorph* includes a useful option for determining the appropriateness of the threshold and image filter parameters chosen by the user: by turning on the *save_intermediates* option, the output of each filter (equivalent to the panels in Figure 4) is saved at each step. This allows users to determine exactly how each filter performs and adjust parameters as necessary.

All objects are measured for area, perimeter, minor axis length, major axis length, eccentricity, aspect ratio and rugosity. The rugosity measure quantifies the surface texture along an object edge by comparing the perimeter length of an unsmoothed versus smoothed outline. *Run2dmorph* outputs a CSV file containing all aforementioned measures, along with object identifiers (e.g. the catalog # and object #). A CSV file containing all extracted x,y-coordinates (downsampled by default to 100 points; this number can be changed by the user) is also generated.

2.4 | Run3dmorph: 3D meshes and volume/ surface area estimation

Run3dmorph performs 3D wireframe extraction and volume/surface area estimation. Run3dmorph is described in detail in Hsiang, Elder,

and Hull (2016) and is designed to function with the file structure outputted by *focus*. In brief, *run3dmorph* generates a height maps for each object to extract a semi-3D mesh of the visible upper half of the object (Figure 5). Surface area and volume of the mesh are then calculated exactly for the visible half. For the bottom half, surface area and volume are estimated using a series of idealized shapes (i.e. conical, cylindrical and domed). The idealized bottom half shapes were chosen specifically for application to foraminifera and may not be applicable to other organisms. However, *run3dmorph* separately outputs the surface area and volume measurements for both (1) the visible half alone and (2) the combined visible half and estimated bottom half. This gives users the flexibility to use the measurements of the top half of the object in combination with their own appropriate assumptions about the bottom half of the object (e.g. doubling the top measurements for laterally symmetric objects).

Run3dmorph outputs the 3D mesh data as x,y,z-coordinates, Wavefront OBJ files and OFF (Object File Format) files. 3D PDFs, which allow the user to quickly view the extracted mesh in comparison with the original object, are optionally outputted by the program. The command for *run3dmorph* is as follows:

run3dmorph <control file>

As with *segment* and *run2dmorph*, parameters are passed into *run3dmorph* via a control file. Details about *run3dmorph*'s parameters can be found in the *run3dmorph* tutorial on GitHub.

2.5 | Example datasets, performance and demonstration of application and breadth

The Data S1 includes a description of an example dataset (a slide containing 1,150 microfossils from an Atlantic core top sample) that can be used to test *AutoMorph*, along with performance benchmarks. The Data S1 also contains examples of *AutoMorph*'s application to a wide variety of fossil and extant organisms, including patellogastropods, bivalves, ichthyoliths and foraminifera. Measurement error quantifications are also provided (Figure 6). In general, *AutoMorph* extracts 2D measurements with high fidelity



FIGURE 5 Pipeline overview for *run3dmorph*. *Run3dmorph* uses the z-stack images of each individual object identified by *segment* to generate height maps, which relate the greyscale value of each pixel in the height map to a real-world height. The height map is used to extract *x*,*y*,*z*-coordinates, which are then used to build a 3D mesh using Delaunay triangulation. The major data products of *run3dmorph* are as follows: a CSV of the extracted coordinates, a CSV containing volume and surface area measurements and estimates, OFF and OBJ representations of the mesh, and a 3D PDF containing the extracted mesh and associated metadata

and accuracy (3.69%-10.65% mean absolute percentage error, assuming hand measurements represent the "true" values). Much of this error may be the result of error introduced during hand

measurement (see Data S1). The 3D measurements, in contrast, are highly susceptible to error introduced by improper imaging and/or scaling (see Data S1). It is important to note, however, that CT scans



FIGURE 6 2D and 3D data verification using limpet shells, bivalve shells, ichthyoliths and foraminifera. (a) Major and minor axis length measurements for 139 individual fossil patellogastropods are compared between calliper (x-axis) and *AutoMorph* (y-axis) measurements. (b) Calliper (x-axis) vs. *AutoMorph* (y-axis) measurements of 33 Recent and subrecent bivalve shells, presented as the square area (calliper: length * width; *AutoMorph*: major axis length * minor axis length). (c) Major and minor axis length measurements for 60 ichthyolith specimens in three samples are compared between manual *ImageJ* (x-axis) and *AutoMorph* (y-axis) measurements. The dashed regression lines correspond to the sample with the same colour dots; the black regression line and R² value refers to all three samples combined. (d) Major and minor axis length measurements for 10 planktonic foraminiferal specimens compared between manual ocular micrometer (x-axis) and *AutoMorph* (y-axis) measurements. (e) Comparison of 3D volume and surface area from microCT scans (x-axis) and *AutoMorph* (y-axis) for 18 planktonic foraminifera, assuming a dome base. The 1-to-1 identity line is plotted in purple

(against which the AutoMorph 3D measurements were compared) are also susceptible to errors arising from warping, improper or poor calibration, or scanner quality and maintenance, which may further

compound the discrepancy observed. Furthermore, despite the high amount of absolute error, *AutoMorph* is able to accurately capture the *relative* difference in 3D volume ($R^2 = 0.8378$) and surface area

 $(R^2 = 0.7815)$ among specimens, again suggesting that the error observed is caused largely by scaling effects.

Finally, we reiterate that manual measurements may themselves be biased—which, in fact, serves as impetus for the development and use of automated measurement methods. The hand-vs.-*AutoMorph* measurements provided here are intended to inform users of the general accuracy level and broad applicability of *AutoMorph*. We encourage users to test measurement accuracy using methods such as high-resolution scanning or comparing *AutoMorph*'s results against measurements generated by landmark and outline-based geometric morphometric techniques, before deciding whether *AutoMorph* is suitable for their particular specimens and purposes. Thorough testing to determine error, bias and repeatability is particularly important when researchers are interested in using measurements that may be highly susceptible to variations in mounting and orientation, such as perimeter and roundness (Brombacher, Wilson, & Ezard, 2017).

2.6 | Future directions and conclusions

AutoMorph provides tools for extracting 2D and 3D coordinate data and basic size and shape measurements from photographic images. Although many bioinformatic tools and algorithms have been developed over the past few years (e.g. Boyer et al., 2011; Falkingham, 2012; Plyusnin, Evans, Karme, Gionis, & Jernvall, 2008; Smith & Strait, 2008; e.g. Kristensen, Parsons, Hallgrimsson, & Boyd, 2008; Shen, Farid, & McPeek, 2009), the number of individual fossils or taxa with data for common morphological traits is low due to the resourceintensive nature of morphological data collection (see Cunningham et al., 2014; Rowe & Frank, 2011). AutoMorph aims to bridge this gap by allowing for rapid extraction of object-specific images and 2D and 3D shape products. Indeed, in about a year, our group has generated images and 2D and 3D data for >100,000 marine microfossils using AutoMorph, of which >61,000 are planktonic foraminifera, comprising the largest single collection of such images and data to date (Elder, L. E., et al., unpubl. data). Other large volume imaging projects carried over the same period included imaging >12,000 patellogastropods (Kahanamoku, S. S., et al., unpubl. data) and >50,000 fish teeth (Sibert, E. C., unpubl. data) using AutoMorph.

Streamlined, high-throughput image processing and dataextraction methods are needed to facilitate a phenomics revolution on par with that of the genomic revolution. Here, we have provided tools to aid in the rapid extraction of morphological data from photographic images. Ongoing work to improve the accessibility and applicability of the *AutoMorph* code includes incorporating Python 3 and cross-platform compatibility, developing a graphical user interface and improving 3D height extraction for macroscopic objects.

ACKNOWLEDGEMENTS

This work was funded by the American Chemical Society PRF Grant #55837-DNI8 and supported in part by Yale University. We would like to thank Brian Dobbins for technical support and initial code development, and the two anonymous reviewers who provided comments and suggestions that greatly improved this manuscript.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR'S CONTRIBUTIONS

A.Y.H. wrote the *run2dmorph* and *run3dmorph* code, wrote the *AutoMorph* manuals and tutorials, generated data for the example coretop slides and drafted the manuscript. K.N. wrote the *focus* and *segment* code. L.E.E. imaged the example coretop slides, generated data for the example coretop slides and conducted initial beta testing. E.C.S. imaged and generated data for the ichthyolith samples and participated in beta testing. S.S.K. imaged and generated data for the patellogastropods and participated in beta testing. J.E.B. imaged and generated the foraminifera CT data. A.K. imaged and measured the bivalves. Y.L. wrote the initial code for *run2dmorph*. P.M.H. conceived of and designed the study, directed all code writing (including contributing to early versions of *segment* and *focus*) and wrote the first draft of the manuscript. All authors contributed to manuscript preparation.

DATA ACCESSIBILITY

All image data are available on Zenodo, including an example microfossil slide scan (https://doi.org/10.5281/zenodo.167557) and example images for the *AutoMorph* tutorials (https://doi.org/10.5281/ zenodo.167510). Data for the 2D and 3D measurement verifications are provided in Tables S1–S5 and on Zenodo (https://doi. org/10.5281/zenodo.1002627). More details about all datasets can be found in the Data S1.

ORCID

Allison Y. Hsiang ២ http://orcid.org/0000-0002-9384-8099

REFERENCES

- Allen, V., Bates, K. T., Li, Z. H., & Hutchinson, J. R. (2013). Linking the evolution of body shape and locomotor biomechanics in bird-line archosaurs. *Nature*, 497, 104–107. https://doi.org/10.1038/nature12059
- Balanoff, A. M., Bever, G. S., Rowe, T. B., & Norell, M. A. (2013). Evolutionary origins of the avian brain. *Nature*, 501, 93–97. https://doi.org/10.1038/ nature12424
- Bookstein, F. L. (1997). Landmark methods for forms without landmarks: Morphometrics of group differences in outline shape. *Medical Image Analysis*, 1, 225–243. https://doi.org/10.1016/ S1361-8415(97)85012-8
- Boyer, D. M., Lipman, Y., St Clair, E., Puente, J., Patel, B. A., Funkhouser, T., ... Daubechies, I. (2011). Algorithms to automatically quantify the geometric similarity of anatomical surfaces. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 18221–18226. https://doi.org/10.1073/pnas.1112822108

- Brombacher, A., Wilson, P. A., & Ezard, T. H. G. (2017). Calibration of the repeatability of foraminifera test size and shape measures with recommendations for future use. *Marine Micropaleontology*, 133, 21–27. https://doi.org/10.1016/j.marmicro.2017.05.003
- Cunningham, J. A., Rahman, I. A., Lautenschlager, S., Rayfield, E. J., & Donoghue, P. C. J. (2014). A virtual world of paleontology. *Trends in Ecology & Evolution*, 29, 347–357. https://doi.org/10.1016/j. tree.2014.04.004
- Dietl, G. P. (2013). The great opportunity to view stasis with an ecological lens. *Palaeontology*, 56, 1239–1245. https://doi.org/10.1111/ pala.2013.56.issue-6
- Falkingham, P. L. (2012). Acquisition of high resolution three-dimensional models using free, open-source, photogrammetric software. *Palaeontologia Electronica*, 15, 1T.
- Falkingham, P. L. (2013). Low cost 3D scanning using off-the-shelf video gaming peripherals. *Journal of Paleontological Techniques*, 11, 1–9.
- Hsiang, A. Y., Elder, L. E., & Hull, P. M. (2016). Towards a morphological metric of assemblage dynamics in the fossil record: A test case using planktonic foraminifera. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371, 20159227. Retrieved from http://zerene systems.com/cms/stacker
- Kristensen, E., Parsons, T. E., Hallgrimsson, B., & Boyd, S. K. (2008). A novel 3-D image-based morphological method for phenotypic analysis. *IEEE Transactions on Biomedical Engineering*, 55, 2826–2831. https://doi. org/10.1109/TBME.2008.923106
- Pearson, P. N., & Ezard, T. H. G. (2014). Evolution and speciation in the Eocene planktonic foraminifer Turborotalia. *Paleobiology*, 40, 130–143. https://doi.org/10.1666/13004
- Petti, F. M., Avanzini, M., Belvedere, M., De Gasperi, M., Ferretti, P., Girardi, S., ... Tomasoni, R. (2008). Digital 3D modelling of dinosaur footprints by photogrammetry and laser scanning techniques: Integrated approach at the Coste dell'Anglone tracksite (Lower Jurassic, Southern Alps, Italy). Studi Trento di Scienze Naturali, Acta Geologica, 83, 303–315.
- Plyusnin, I., Evans, A. R., Karme, A., Gionis, A., & Jernvall, J. (2008). Automated 3D phenotype analysis using data mining. *PLoS ONE*, *3*, e1742. https://doi.org/10.1371/journal.pone.0001742
- Polly, P. D., Eronen, J. T., Fred, M., Dietl, G. P., Mosbrugger, V., & Scheidegger, C., ... Fortelius, M. (2011). History matters: Ecometrics and integrative

climate change biology. *Proceedings of the Royal Society B: Biological Sciences*, 278, 1131–1140. https://doi.org/10.1098/rspb.2010.2233

- Rowe, T., & Frank, L. R. (2011). The disappearing third dimension. Science, 331, 712–714. https://doi.org/10.1126/science.1202828
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., & Pietzsch, T., ... Cardona, A. (2012). Fiji: An open-source platform for biological-image analysis. *Nature Methods*, *9*, 676–682. https://doi. org/10.1038/nmeth.2019
- Shen, L., Farid, H., & McPeek, M. A. (2009). Modeling three-dimensional morphological structures using spherical harmonics. *Evolution*, 63, 1003–1016. https://doi.org/10.1111/evo.2009.63.issue-4
- Slice, D. E. (2007). Geometric morphometrics. Annual Review of Anthropology, 26, 261–281. https://doi.org/10.1146/annurev. anthro.34.081804.120613
- Smith, N. E., & Strait, S. G. (2008). PaleoView3D: From specimen to online digital model. *Palaeontologia Electronica*, 11, 11A.
- Sutton, M. D., Rahman, I. A., & Garwood, R. J. (2014). Techniques for virtual palaeontology. Hoboken, NJ: John Wiley & Sons, Ltd.
- Zelditch, M. (2004). Geometric morphometrics for biologists: A primer. Amsterdam: Elsevier Academic Press.
- Zerene Systems LLC. (2016). Zerene stacker [Computer software].

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Hsiang AY, Nelson K, Elder LE, et al. AutoMorph: Accelerating morphometrics with automated 2D and 3D image processing and shape extraction. *Methods Ecol* Evol. 2017;00:1–8. <u>https://doi.org/10.1111/2041-</u> 210X.12915