Mediterranean Archaeology and Archaeometry, Vol. 16, No 5 (2016), pp. 11-17

DOI: 10.5281/zenodo.204962

PHOTOGRAMMETRIC AUTOMATION: IS IT WORTH?

A. Georgopoulos

Laboratory of Photogrammetry, School of Rural & Surveying Engineering,
National Technical University of Athens
(drag@central.ntua.gr)

Received:20/01/2016
Accepted: 20/08/2016

ABSTRACT

This paper focuses on the evaluation of automated image based techniques which are used lately in order to produce three dimensional digital models of cultural heritage objects. This implies a doubt as to how accurate and reliable are the products of these automated algorithms and how efficient they are for providing the necessary 3D material for the virtual environments. The implementation of this innovative approach has gradually become very popular in the field of Cultural Heritage during the last five years. Non-specialists found a way to easily produce 3D reconstructions just by taking a few images and using Structure-from-Motion and Multiview Stereo algorithms implemented by commercial or open source software. However, this fact led to debatable results, as a lot of ambiguities are lurking hidden in the "happiness" of automation. In order to assess the metric performance of these algorithms an innovative metric evaluation strategy has been designed. A very accurately measured test field, set up mainly for calibrating digital cameras, was used as an object of well-known ground truth. It was established that the accuracy of the resulting 3D reconstruction depends on the spatial analysis and the general quality of the original digital images, on the careful selection of the parameters provided by the software, the properties of the object itself and the computing power available.

KEYWORDS: Virtual Museums, 3D digitizing, Artifacts, Image Based Methods, Accuracy.

1. INTRODUCTION

Nowadays museums are definitely going virtual (Huhtamo 2002; Sylaiou et al. 2009) together with many other Cultural Heritage activities. This fact means that two dimensional or indeed three-dimensional digitization of artifacts, but also Cultural Heritage objects in general, is a fast emerging necessity. For many decades, geomatic engineers and photogrammetrists were developing methods and techniques for the 3D digitization of objects, including those belonging to the cultural heritage domain (Georgopoulos & Ioannidis 2005; Roncella et al. 2011; Manfrendini & Galassi 2013, Evgenikou & Georgopoulos 2015, Papadaki et al. 2015). This development was decisively enhanced after the fusion of computer vision and photogrammetric scientific fields which started in the late 80's. This led to the possibilities to include 3D visualizations in virtual environments, like virtual tours, virtual museums, serious games applications etc. However, this frenzy for the easily available automation has caused a two-fold problem. These methods became widely available and are used by a wide spectrum of people who ignore the inevitable pitfalls and, in the unlikely event that they discover some of them; they are not well equipped to correct or compensate them.

This paper focuses exactly on this last point, i.e. on how accurate and reliable are the products of these automated algorithms and how efficient they are for providing the necessary and correct 3D material for the virtual environments.

2. 3D DIGITIZING

In order to cover the ever increasing need for fast, low cost 3D digitization and effective visualization of large number of artifacts the use of terrestrial laser scanners boomed during the past 15-20 years. In addition scanners of different technologies, like e.g. structured light scanners, were also employed. However, these instruments were and still are expensive, bulky and very often not suitable for the job. One needed a multitude of such devices in order to be able to confront all possible applications.

In a few words, there were active and passive methods available. Scanners, with a light source are considered as active devices. The terrestrial laser scanners are categorized into Time-of-Flight (ToF), pulse or phase shift and triangulation scanners using one or two digital cameras. They vary in accuracy and applicability. Their accuracy ranges from a few millimetres for the pulse and phase shift ones, to fractions of millimetres for the triangulation scanners. However, they vary a lot in range possibilities from a few kilometres to a few metres.

Photogrammetry, on the other hand, has been active since almost a decade after the advent of photography (1860) and has been established in the 70's as a robust, albeit expensive and time consuming, method for cultural heritage documentation. It is in 1964 that the International Society for Photogrammetry and remote Sensing (ISPRS. www.isprs.org) and the International Council for Monuments and Sites (ICOMOS, www.icomos.org) have formed CIPA-Heritage Documentation (Comité Internationale de Photogrammetrie Architecturale, cipa.icomos.org), a common International Scientific Committee to foster the implementation of modern technologies to cultural heritage documentation.

In parallel, the computer vision community slowly emerged from the field of computer science and it started merging with the above disciplines in the 90's. This alternative for 3D digitizing, which emerged faster during the recent years supported by the increase in computing power, is the automated implementation of classic photogrammetric algorithms. This development provided interested parties with a rather cheaper, faster and perhaps more reliable solution to the problem.

Just by examining the keywords of the publications in the related conferences and journals during the last 15 years, one may establish a distinct trend in the number of publications, the methods employed and, ultimately, in the shift in scientific interest from laser scanning to image based, i.e. automated photogrammetric, techniques. The diagram in Figure 1 shows the change in number of related papers during the past 15 years.

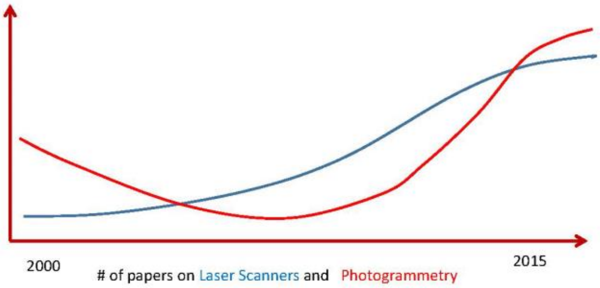

Figure 1. Number of scientific papers on Laser Scanning and Photogrammetry

3. IMAGE BASED 3D CAPTURE

For the three dimensional documentation of an object, large or small, the acquisition or production of a point cloud is necessary. Preferably this point cloud should also include colour, i.e. texture, information. Nowadays, the most common way to acquire such a point cloud is still by means of a terrestrial laser scanner. As mentioned above, these instruments are in essence active sensors and do not necessarily record texture information with the position of the points.

The alternative is, of course, photogrammetric techniques. Conventional photogrammetric methodology involves acquisition of overlapping pairs of images, in order to enable stereoscopic viewing (Figure 2). The process implies a series of processes to take into account the camera geometry, the relative geometry of the two pictures of each pair and the connection of the images to the world reference system. These processes are called, *interior orientation*, *relative orientation* and *absolute orientation*, or *image triangulation* or *bundle adjustment* in the case of more than two images. The final stage was the surface reconstruction using specialized instruments to enable stereoscopic viewing.

Computer vision has contributed to the automation of the image based techniques. The algorithms developed process a sequence of digital images with large overlap and produce all the above information as the conventional photogrammetric technique, but in addition the dense point clouds which also convey texture information as they have been extracted from the digital images.

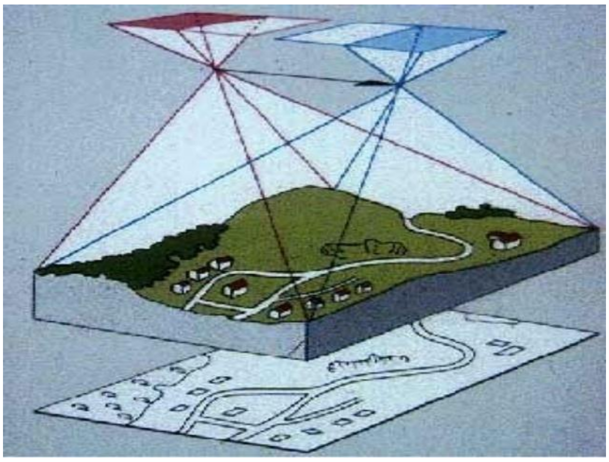

Figure 2: The conventional Photogrammetric process (http://www.stereoscopy.com/faq/aerial.html)

These algorithms are known as Structure-from-Motion (SfM) and Multi-View Stereo (MVS) algorithms and they actually determine a large number of points on the images mainly using the SIFT algorithm (Lowe, 2004), find their correspondences and thus reconstruct the object imaged and align the images, i.e. the bundles of rays. These processes are actually the classic photogrammetric processes, which combine overlapping images in order to allow for stereoscopic imaging and -usually manual- 3D reconstruction of the objects imaged in the overlapping part. So interior orientation is now called intrinsic camera parameters, relative orientation is equivalent to image alignment and stereoviewing is replaced by dense image matching. The outcome is again the surface 3D reconstruction, only in a shorter time.

The development of these automated algorithms has been greatly boosted by the increase in computer power on one hand and by the relevant computer vision algorithms, such as interest operators, feature detection, matching etc. on the other. Perhaps the most well-known operator is SIFT. Scale Invariant Feature Transform (SIFT) is an image descriptor for image-based matching and recognition developed by Lowe (1999, 2004). This descriptor as well as related image descriptors are used for a large number of purposes in computer vision related to point matching between different views of a 3-D scene and view-based object recognition. The SIFT descriptor is invariant to translations, rotations and scaling transformations in the image domain and robust to moderate perspective transformations and illumination variations. Experimentally, the SIFT descriptor has been proven to be very useful in practice for image matching and object recognition under real-world conditions.

In its original formulation, the SIFT descriptor comprised a method for detecting interest points from a grey-level image at which statistics of local gradient directions of image intensities were accumulated to give a summarizing description of the local image structures in a local neighbourhood around each interest point, with the intention that this descriptor should be used for matching corresponding interest points between different images. Later, the SIFT descriptor has also been applied at dense grids (dense SIFT) which have been shown to lead to better performance for tasks such as object categorization, texture classification, image alignment and biometrics. The SIFT descriptor has also been extended from grey-level to colour images and from 2-D spatial images to 2+1-D spatio-temporal video (http://www.scholarpedia.org/article/Scale_Invariant_Feature_Transform).

On the other hand Structure-from-Motion (SfM) is a range imaging technique; it refers to the process of estimating three-dimensional structures from two-dimensional image sequences which may be coupled with local motion signals, i.e. images. It is studied in the fields of computer vision and visual perception, but actually originates from the field of photogrammetry. In biological vision, SfM refers to the phenomenon by which humans (and other living creatures) can recover 3D structure from the projected 2D (retinal) motion field of a moving object or scene. Today there are a lot of software available implementing these algorithms. There are quite a few commercial software, like e.g. Photoscan, Pix4D, Acute3D, 3DFZephyr, etc. At the same time there are web services available applying the above principles, like e.g. ARC3D, 123DCatch and now Recap, Memento and Photosynth. These services require an internet connection and the user agrees that the uploaded images become property of the service. Finally there are numerous freeware pieces of software like e.g. VisualSfM, MicMac, SfM Toolbox for Matlab, Structure and Motion Toolkit, Bundler - Structure from Motion for unordered photo collections etc.

This automation has made the 3D reconstruction widely available and popular, as implementing photogrammetry has always been a complicated, time consuming and rather costly action. This "democratization", however, has been escorted with a lot of scepticism, as the masses were not able to control the results whatsoever and evaluate them in terms of accuracy and reliability. Hence questions arose as to whether these algorithms are producing usable and efficient 3D reconstructions. Moreover a question of trusting these algorithms is often asked by photogrammetric experts.

The implementation of this innovative approach became very popular also in the field of Cultural Heritage. Non-specialists found a way to easily produce 3D reconstructions just by taking a few images. In most cases where fast and "light" 3D models are required just for visualization purposes there are no major problems. If, however, accurate and "heavy" 3D models are required for metric purposes, serious problems tend to arise. Hence, this development led to debatable results, as a lot of ambiguities are lurking hidden in the happiness of automation.

On the other hand the software developers are reluctant to publish the exact algorithms they have used for their software and this is understandable to a certain extent. Consequently the exact way of processing the raw data is not known a priori. This causes uncertainty as to the accuracy of the final products and also for the ability of the algorithms to reconstruct a geometrically reliable model. In general the quality of the final product depends on the quality of the images, i.e. on the presence of noise and on the pixel size of the sensor), on the overlap percentage, on the properties of the object itself and, of course, on the choice of the various parameters for the execution of the various procedures of the software, which are constrained by the characteristics of the computing unit, i.e. RAM, processor etc.

4. EVALUATION OF AUTOMATION

It becomes obvious that the necessity arises to evaluate the performance of these algorithms in terms of reliability and accuracy. In the relevant literature, few reports can be found on that subject (Roncella et al. 2011; Oliensis 2000; Xiang & Cheong 2003; Seitz et al. 2006; Skarlatos & Kyparissi 2012). In order to provide an answer to the above issues an accuracy test was designed and performed with several implementation examples. They make use of popular commercial SfM/MVS software.

The error sources hidden in the implementation of these automated algorithms are many and diverse. Camera geometry and optical deformations and instabilities of the photographic lens are often overlooked and this leads to unpredictable results. Although most of the available software provide for a camera calibration module, it is often not good enough to model the digital camera. Image quality is another source of ambiguities. Especially when compact cameras are used and the pixel size on the digital sensor is small, the algorithm may fail, or may lead to erroneous correspondences and unreliable reconstructions. Moreover the network geometry plays an important role for the final result. Images should be taken in a well-planned sequence in order to assist the algorithm to produce a complete reconstruction. Finally the ground control should also be carefully selected and introduced. This will give the desired scale and perhaps position to the 3D reconstruction, which most often is decisive for the quality of the final product. All the above error sources and their effects are investigated and discussed in detail.

In order to assess the metric performance of these algorithms an innovative metric evaluation strategy has been designed (Kremezi & Kristollari, 2014). A very accurately measured test field, set up for camera calibrations and available at the Laboratory of Photogrammetry of NTUA, was used as an object of ground truth (Figure 3). This test field is comprised of 16 vertical aluminium rods each bearing 3-4 targets and its dimensions are 6.5x2.65 m. The 3D coordinates of the test field's points are measured with specially developed geodetic techniques with an accuracy of 0.1mm every year.

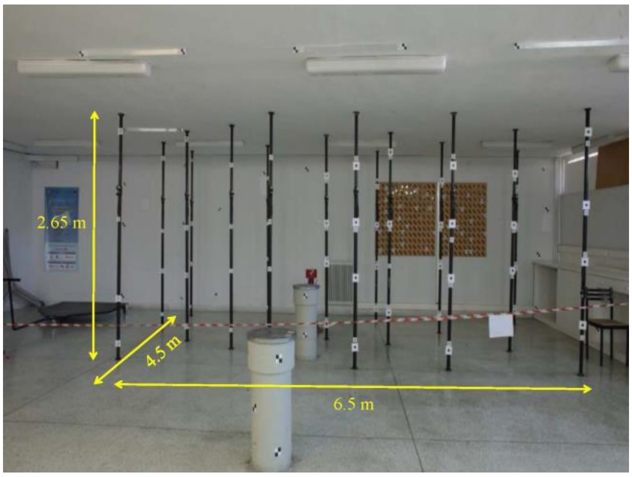

Figure 3: The test field of NTUA, Lab. of Photogrammetry

The test field was imaged with two different digital cameras. One was a full frame DSLR (Canon 1Ds MIII) with two fixed lenses (50mm and 24mm) and the other a digital compact camera (Sony Cybershot DSC-W530). In Table I the main characteristics of the two cameras are presented together with the properties of the two image sequences.

Table I. Camera and Image characteristics

	Canon MIII	Sony DSC-W530
Resolution (MP)	21	14
Sensor size (mm^2)	36x24	6.2x4.7
Pixel pitch (μm)	6.4	1.4
Aperture	f/8	f/2.7
Shutter speed (s^{-1})	2	30
ISO	400	800

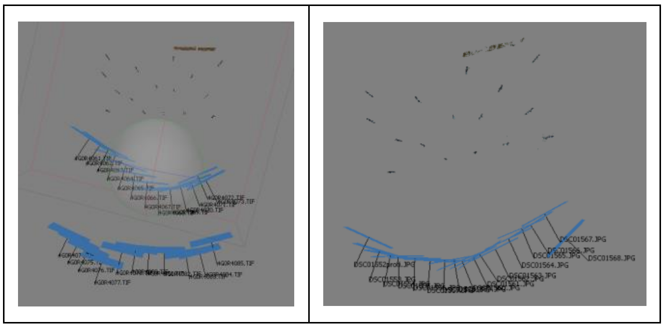

Figure 4: The distribution of the three sets of images with the Canon (left) and the Sony (right) digital cameras

The two sets of images (Figure 4) were processed separately within the software. For the satisfactory reconstruction of the object several attempts were necessary using different parameters of the software. Special masking of the vertical rods was needed for optimum results. The characteristics of the two reconstructions, i.e. one with all images from the DSLR and the second with those from the compact digital camera are presented in Table II.

Table II. Reconstruction characteristics

	Canon DSLR	Sony Cybershot
Images	24	17
Sparse cloud	high	high
Dense cloud	high	high
# of points	1.090.864	311.177
# of triangles	2.642.816	1.719.997

When the 3D models were satisfactorily reconstructed, a number of control points were used to georeference them. The resulting 3D reconstructions were compared to the known values of the targets of the test field. The accuracy procedure was checked on a number of points not used for the referencing (check points) by comparing their coordinates taken off the reconstruction, i.e. determined by the software.

5. PRESENTATION OF RESULTS

The resulting differences have been statistically examined in order to draw conclusions as to the metric performance of the algorithm. In this case the commercial software Photoscan by Agisoft was used, which implements the SfM and MVS algorithms. In addition it provides to the user with certain tools in order to enhance the processing of the image sequences. The results of this thorough evaluation are presented in detail and conclusions are drawn for the metric abilities of the algorithm.

In Table III the concentrated results of the DSLR reconstruction are presented. It is obvious that in this case the software has reached the limits of the coordinate accuracy. It should be borne in mind that in this case the ground sampling distance (GSD), i.e. the pixel size on the object, was less than a millimeter.

Table III. Coordinate differences for the DSLR reconstruction

(m)	**X**	**Y**	**Z**	**Total**
		GCP's		
max	0.00041	0.00040	0.00081	0.00098
min	0.00003	0.00003	0.00003	0.00020
average	0.00021	0.00019	0.00037	0.00050
		Check Points		
max	0.00071	0.00132	0.00177	0.00194
min	0.00005	0.00010	0.00014	0.00053
average	0.00029	0.00056	0.00083	0.00113

Similarly in Table IV the results for the coordinate differences from the 3D reconstruction of the images of the compact camera are presented. In this case the GSD was 1.1mm.

Table IV. Coordinate differences for the compact reconstruction

(m)	X	Y	Z	Total
		GCP's		
max	0.00093	0.00114	0.00192	0.00206
min	0.00005	0.00003	0.00020	0.00048
average	0.00040	0.00037	0.00064	0.00095
		Check Points		
max	0.00368	0.00381	0.00854	0.00964
min	0.00007	0.00014	0.00047	0.00071
average	0.00170	0.00099	0.00295	0.00367

From the above results it becomes obvious that both reconstructions give acceptable results in this fully controlled environment. However a significant decrease of the accuracy achieved is observed in the case of the compact camera. For the GCP's the differences are double in size compared to those from the DSLR and in the case of the check points, which is more realistic, three times in most of the cases.

6. CONCLUDING REMARKS

It was established that the accuracy of the resulting 3D reconstruction depends on the spatial analysis and general quality of the original digital images, on the careful selection of the parameters provided by the software, the properties of the object itself and the computing power available.

The general conclusion is that the SfM-MVS software examined is perfectly capable of producing 3D reconstructions which are accurate but also aesthetically acceptable provided careful selection of parameters is performed. This can be very tricky in certain cases and needs deep knowledge of the algorithm and the significance of the parameters. Every case could be different; hence no general rule may apply. However, the final accuracy and reliability of the 3D reconstruction is dependent on the following parameters:

- ✓ The spatial resolution of the images (sensor element size) and their quality (focusing, signal-to-noise ratio SnR)
- ✓ The presence of occluded areas, i.e. areas that are not imaged on at least two pictures
- ✓ The technical characteristics of the computer, especially the available RAM
- ✓ The suitable selection of the various parameters offered by the software for the calculations and the implementation of the algorithms
- ✓ The properties of the object itself

In this case of the controlled experiment performed and described, the accuracy achieved was a result of many trials. The reconstructed model with the DSLR images presents an accuracy which is smaller than the GSD, because of the relatively large number of GCP's. In addition, the corresponding results from the compact camera are better than anticipated. This is a rather promising fact, giving the possibility to very simple and cheap cameras to achieve acceptable results, provided certain conditions are met.

REFERENCES

Evgenikou V., Georgopoulos, A., 2015. Investigating 3d reconstruction methods for small artifacts. ISPRS Archives, Volume XL-5/W4 Page(s) 101-108, WG V/4, CIPA - 3D-Arch 2015 – 3D Virtual Reconstruction and Visualization of Complex Architectures 25–27 February 2015, Avila, Spain. Editor(s): D. Gonzalez-Aguilera, F. Remondino, J. Boehm, T. Kersten, and T. Fuse. http://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XL-5-W4/101/2015/isprsarchives-XL-5-W4-101-2015.pdf

Georgopoulos, A., Ioannidis, Ch., 2005. 3D Visualization by Integration of Multisource Data for Monument Geometric Recording, in Recording, Modeling and Visualization of Cultural Heritage, Baltsavias et al. (eds), Taylor & Francis Group ISBN 0 415 39208 X, International Workshop, Ascona, 2005.

Huhtamo Erkki,2002. On the Origins of the Virtual Museum. Nobel Symposium (NS 120) Virtual Museums and Public Understanding of Science and Culture, May 26-29, 2002, Stockholm, Sweden.

Kremezi, M., Kristollari, V., 2014. Accuracy assessment of the Agisoft Photoscan software. Technical Report, Summer Course in Photogrammetry, Lab of Photogrammetry, NTUA (in Greek), unpublished.

Lowe, D. (1999). Object recognition from local scale-invariant features. Proc. 7th International Conference on Computer Vision (ICCV'99) (Corfu, Greece): 1150-1157.doi:10.1109/ICCV.1999.790410.

Lowe, D. 2004. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 60, 2 (2004), pp. 91-110, doi:10.1023/B:VISI.0000029664.99615.94

Manferdini A. M., Galassi M., 2013. Assessments for 3D reconstructions of cultural heritage using digital technologies. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XL-5/W1, 2013 3D-ARCH 2013 - 3D Virtual Reconstruction and Visualization of Complex Architectures, 25 – 26 February 2013, Trento, Italy.

Oliensis J., 2000. A Critique of Structure-from-Motion Algorithms. Computer Vision and Image Understanding 80, 172–214 (2000) doi:10.1006/cviu.2000.0869, http://www.idealibrary.com.

Papadaki A. I. , Agrafiotis P., Georgopoulos, A., S. Prignitz, 2015. Accurate 3d scanning of damaged ancient Greek inscriptions for revealing weathered letters. ISPRS Archives, Volume XL-5/W4, WG V/4 Page(s) 237-243, CIPA - 3D-Arch 2015 – 3D Virtual Reconstruction and Visualization of Complex Architectures 25–27 February 2015, Avila, Spain. Editor(s): D. Gonzalez-Aguilera, F. Remondino, J. Boehm, T. Kersten, and T. Fuse. http://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XL-5-W4/237/2015/isprsarchives-XL-5-W4-237-2015.pdf

Roncella R., Re C., Forlani G.,2011. Performance evaluation of a structure and motion strategy in Architecture and cultural heritage. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XXXVIII-5/W16, 2011 ISPRS Trento 2011 Workshop, 2-4 March 2011, Trento, Italy.

Seitz S. M., Curless B., Diebel J., Scharstein D., Szeliski R., 2006. A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms. Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference (Volume:1) 17-22 June 2006, pp.519-528, ISSN:1063-6919, Print ISBN: 0-7695-2597-0 DOI: 10.1109/CVPR.2006.19.

Skarlatos D., Kiparissi S., 2012. Comparison of Laser Scanning, Photogrammetry and SFM-MVS Pipeline Applied in Structures and Artificial Surfaces. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume I-3, 2012 XXII ISPRS Congress, 25 August – 01 September 2012, Melbourne, Australia.

Sylaiou S., Liarokapis F., Kotsakis K., Patias P., 2009. Virtual museums, a survey and some issues for consideration. Journal of Cultural Heritage 10 (2009) 520–528, doi:10.1016/j.culher.2009.03.003.

Xiang Tao, Cheong Loong-Fah, 2003. Understanding the Behavior of SFM Algorithms: A Geometric Approach, International Journal of Computer Vision, February 2003, Volume 51, Issue 2, pp 111-137.