



LOW COST 3D SCANNING USING OFF-THE-SHELF VIDEO GAMING PERIPHERALS

Peter L. Falkingham¹

¹-Department of Comparative Biomedical Sciences, Structure & Motion Laboratory, Royal Veterinary College, London, UK and Department of Ecology Evolutionary Biology, Division of Biology and Medicine, Brown University, USA
Email: pfalkingham@rvc.ac.uk

ABSTRACT

Digitization of specimens is becoming an ever more important part of palaeontology, both for archival and research purposes. The advent of mainstream hardware containing depth sensors and RGB cameras, used primarily for interacting with video games, in conjunction with an open platform used by developers, has led to an abundance of highly affordable technology with which to digitize specimens. Here, the Microsoft® Kinect™ is used to digitize specimens of varying sizes in order to demonstrate the potential applications of the technology to palaeontologists. The resulting digital models are compared with models produced using photogrammetry. Although the Kinect™ generally records morphology at a lower resolution, and thus captures less detail than photogrammetric techniques, it offers advantages in speed of data acquisition, and generation of a completed mesh in real time at the point of data collection. Whilst it is therefore limited in archival applications, the ease of use and low cost, driven by strong market competition, make this technology an enticing alternative for studies where rapid digitization of general morphology is desired.

Keywords: Scanning; 3D; digital; model; laser; Kinect; digitization, photogrammetry

RESUMO [in Portuguese]

A digitalização de espécimes está a tornar-se numa parte cada vez mais importante da paleontologia, tanto para fins de arquivo como de pesquisa. O aparecimento de equipamento convencional contendo sensores de profundidade e câmaras fotográficas RGB, utilizados principalmente para interagir com jogos de vídeo, em conjunto com plataformas abertas de acesso livre desenvolvidas por programadores levou a um manancial de tecnologia altamente acessível para a digitalização de espécimes. Com o objectivo de demonstrar as potenciais aplicações desta tecnologia para paleontólogos, o Microsoft® Kinect™ é aqui usado para digitalizar os espécimes de vários tamanhos. Os modelos digitais resultantes são comparados com modelos produzidos utilizando fotogrametria.

Embora o Kinetic™ geralmente registe morfologia numa resolução menor e, portanto, capta menos detalhes do que as técnicas de fotogrametria, oferece vantagens na velocidade de aquisição de dados e geração da malha 3D concluída em tempo real durante a recolha de dados. Apesar de limitado em aplicações de arquivamento, a facilidade de uso e o baixo custo, impulsionados pela forte concorrência no mercado, estão a tornar esta tecnologia numa alternativa atraente para os estudos em que se deseje uma rápida digitalização de morfologia.

How to cite this article: Falkingham, Peter L., 2013. Low Cost 3D Scanning using off the shelf video gaming peripherals. Journal of Paleontological Techniques, 11: 1-9.



Copyright (c) 2013 by Peter Falkingham. This work is made available under the terms of the Creative Commons Attribution 3.0 Unported License, <http://creativecommons.org/licenses/by-sa/3.0/>.

INTRODUCTION

The ability to digitally capture the 3D morphology of a specimen has revolutionised palaeontology over recent years. Working in the virtual realm permits investigators to section, profile, manipulate and colour a specimen in ways that would otherwise be difficult or impossible, and often destructive, if applied to the real fossils. The use of digital models has facilitated a wide range of research into areas including locomotion, feeding, body mass calculations, documentation, conservation, hydro- and aerodynamics, and many others (Anderson et al., 2011; Bates et al., 2012; Bates and Falkingham, 2012; Bates et al., 2009; Falkingham et al., 2009; Farlow et al., 2012; Gidmark et al., 2013; Hutchinson et al., 2011; Panagiotopoulou et al., 2011; Rayfield, 2007; Sellers et al., 2012). Equally importantly, digitization of specimens in the internet age has enabled an unprecedented level of data sharing and collaboration, exemplified by online repositories such as Digimorph (www.digimorph.org), and online journals such as this, which enable digital models to accompany publications as electronic supplementary material (see appendices).

Until recently, however, digitization of fossils remained the purview of those with access to expensive hardware such as computed tomography (CT) and laser scanners, or expensive software-based photogrammetric solutions. While CT machines remain a requirement for internal morphology, methods with which to digitize external morphology have reached such a low cost that they have become available to anyone. Photogrammetric techniques can now be employed with a basic consumer camera, desktop PC, and free software (Falkingham, 2012).

One of the major developments in photogrammetric software most recently has been incorporating the GPU, or Graphics Processing Unit, found in many modern desktop and laptop computers - particularly in machines built for running computer games. The GPU contains many cores, far exceeding the number of cores found on the CPU, or Central Processing Unit (the 'processor' of the computer). While GPU cores are more specialized than those on the CPU, they can be used to dramatically speed up 3D applications (providing the software is written to take advantage of the GPU). Importantly, unlike more industrial hardware such as laser scanners, GPUs form part of a major consumer driven market - the video game industry - which drives prices down and processing power

up at exceptional rates. Coupled with the falling cost of rapid-prototype machines (3D printers), there is now also a growing consumer demand for 3D digitization, which means that palaeontologists can take advantage of software and hardware developments at prices determined by aggressive market forces.

One such recent development is in using depth sensors, designed for interacting with computer games without a traditional controller, to scan and digitize objects and environments. Such sensors first came into being (at least in the mainstream) with the Microsoft Kinect™ for Xbox 360®, released for the gaming console in 2010. This was later followed by the Microsoft Kinect™ for Windows®, and the Asus® Xtion Pro, which possess similar specifications but are designed for developers and a personal computer environment.

Although developed for computer games, the low cost depth sensor was quickly 'hacked' for other uses, particularly in the fields of robotics (Stowers et al., 2011) and rehabilitation (Chang et al., 2011). Of pertinence here is that one of these uses included real time mapping of the environment, and tracking of the sensor (Newcombe et al., 2011). Despite initially being developed for robots navigating environments, the 3D mapping functionality can be used to create digital 3D models of objects or environments, and software applications designed for this purpose are now available (Izadi et al., 2011). These applications are possible because of the recent increases in consumer GPUs; being able to record a point cloud from the depth sensor, and in real time mesh that point cloud to produce a 3D model.

In this paper, I aim to demonstrate the applicability of these gaming peripherals and associated software packages for scanning and digitising palaeontological specimens.

METHODS

Hardware

Throughout this paper the Microsoft® Kinect™ for Windows® was used (available for around \$150). The Kinect™ sensor incorporates an RGB camera, recording at 640x480 pixels, and a depth sensor which uses an infrared laser. Alternatives to this include the original Kinect™ for Xbox 360®, or the Asus® Xtion PRO, both available for a similar price to the Kinect™ for Windows®, but with minor differences. Both use an open source driver to communicate with the PC, making them compatible with Windows, Unix, and Mac operating systems, while the Kinect™ for Windows® requires a Windows PC.

The Kinect™ for Windows® has a shorter minimum range compared with the other options (40cm vs 80 cm) enabling closer scanning, while the Xtion PRO benefits from being powered by the USB cable, thereby offering greater mobility if attached to a laptop or tablet PC. It is worth noting that because the Kinect™ for Xbox® has been predominantly developed for, and sold to, the computer game market, devices are available for very low cost (<\$80 at the time of writing). The sensor was connected to a Laptop running Windows 7, containing an Intel® 8-core CPU (2.20 GHz) and an Nvidia GTX 560M GPU.

Software

There are currently several software options available for using the Kinect™ or Xtion PRO as a 3D scanner, including both commercial and non-commercial programs. Most programs will work equally with either hardware system. The two major programs that are relatively mature at this stage are ReconstructMe (<http://reconstructme.net/>), a commercial code with an additional non-commercial license available, and KinFu, part of the Point Cloud Library (<http://pointclouds.org/>), available as part of the pre-release v1.7 source.

For this paper, the non-commercial console version of ReconstructMe was used. In order to capture an object or environment in 3D, the sensor is held in the hand and moved slowly over the specimen to be digitised. The depth sensor records xyz coordinates of points within the scene which are meshed in real time (producing a solid digital model rather than a cloud of points). If movements of the sensor are made too quickly, and the scene changes too much for the software to track, recording will pause and the user must return the sensor to the last known position (as shown on screen). This loss of tracking can be particularly troublesome when first starting to use the sensor, but experience in moving the sensor by hand will result in more stable data acquisition.

Specimens

In order to illustrate the abilities - and weaknesses - of this technology for palaeontology, three specimens of varying size and complexity were digitized: A block containing dinosaur tracks (block ~30 cm across, tracks ~10 cm in length), an Elephant tooth (~30 cm), and a mounted skeleton of a Pronghorn (*Antilocapra americana*) (~1.2 m length). The specimens were sourced from the teaching and research collections at Brown University. Both the Elephant tooth and the

track block were digitized laid on a table, and so only one side was recorded.

Additional comparative data

For comparison of the results of scanning with the Kinect™ with more commonly used digital acquisition techniques, photogrammetric models were produced of all specimens. Photographs were taken with a Sony Nex-6 and 16-50mm lens. All photos were taken at 16mm focal length and a resolution of 4912 x 3264. The number of photographs was arbitrarily chosen in each case so as to maximize coverage of the specimens, whilst maintaining a high resolution of the specimen. The photographs were processed using VisualSFM (Wu, 2007, 2011; Wu et al., 2011) for the sparse reconstruction, and CMVS/PMVS (Furukawa et al., 2010a, b; Furukawa and Ponce, 2010) through the VisualSFM application for the dense reconstruction. This was carried out on the same laptop used for scanning with the Kinect™ sensor.

The time taken to generate the models with both methods was noted (Table 1) for comparison. Data acquisition was defined as the time taken to move the Kinect™ over the specimen and record the digital model, or the total time between first and last photograph in the case of the photogrammetric models. Data processing time was zero for the Kinect™ as ending the scan results in a complete 3D model almost instantaneously. For the photogrammetry, data processing time included the time taken for feature detection, feature matching, sparse reconstruction, and dense reconstruction. Note that in all three cases of photogrammetry, the dense reconstruction accounted for over 90% of the total time. Both photogrammetric and Kinect™ models were then cleaned of extraneous points using the cropping tool in CloudCompare v2.4 (<http://www.danielgm.net/cc/>), though the time taken to do this (1-2 minutes in each case) was not recorded as the same process, a simple crop, was used for both methods.

Models were compared using the cloud/mesh distance tool in CloudCompare, after scaling the photogrammetric models accordingly (because photogrammetry is a scale-less method, an object of known size, such as a scale bar, must be included in the final model such that the resultant point cloud can be correctly scaled). The output mesh produced by the Kinect™ and ReconstructMe was compared directly to the dense photogrammetric point cloud, rather than meshing the point cloud first. This method of comparison was chosen for two reasons, firstly, producing a mesh from

the dense output of VisualSFM and PMVS/CMVS is a process highly dependent on user inputted variables, and as such meshes can vary in quality, and in relation to the raw data, depending on workflow. Secondly, generating meshes for objects such as the Pronghorn skeleton are difficult because of the proximity

of features such as the ribs, and often require clean-up in post-processing, adding a second subjective source of error (see discussion). That the output from the Kinect™ is in meshed format, as opposed to a point cloud, will be addressed in the discussion section.

Table 1

Specimen	Digitising Technique	Data acquisition time (mins)	Data Processing time (mins)	Number of Vertices
Elephant molar (lateral)	Kinect	01:05	00:00	493583
	Photogrammetry	1:27 (42 pictures)	42:50	990199
Pronghorn	Kinect	03:56	00:00	523008
	Photogrammetry	1:16 (40 pictures)	47:03	281611
Dinosaur Track	Kinect	01:42	00:00	511294
	Photogrammetry	1:00 (29 pictures)	57:35	840095

RESULTS

The outputs of the Kinect™ sensor and ReconstructMe are presented in Figures 1-3, alongside photogrammetric models and comparison data (also see appendices). The resolution of detail in resultant models is considerably lower in the Kinect™ models than the photogrammetric models, as is particularly evident in the track and elephant tooth, where the Kinect™ models clearly show a lack of finer features compared to the photogrammetric point cloud. Resolution of the Kinect is on the order of 5-10 mm in the best case, while the photogrammetric methods can resolve details at the millimetre scale in this instance. The Kinect™ models are natively scaled correctly however, and as such can be measured directly in any 3D modelling package. This is in contrast to the photogrammetric method where models must be scaled by the user according to an object of known size within the final 3D model. The block containing dinosaur tracks proved to be at the lower limits of usefulness in digitizing with the Kinect™, with details poorly resolved (Figure 1). The model produced using the Kinect™ appears smoothed in relation to the physical specimen or photogrammetric model.

However, despite this smoothing, the model is reasonably accurate, the largest inaccuracies occurring around the tips of digits, where the relief becomes smaller than the 5 mm resolution of the Kinect™. In these areas, the photogrammetric point cloud and Kinect™ mesh differ by ~5 mm, though the majority of the surface of the block differs only by 1-2 mm. As with the track block, the Elephant tooth was also more smoothed in the Kinect™ mesh, though the overall morphology was well captured (Figure 2). The primary location of difference between the Kinect™ and photogrammetric point cloud was on the tooth roots, where the tips were not recorded by the Kinect™ resulting in 16 mm maximum difference between methods. Smaller features such as the cracks were recorded by photogrammetry, but lost entirely in the Kinect™ model. Again however, most of the model maintained just a few mm difference to the photogrammetric point cloud.

The mounted Pronghorn skeleton offered a different challenge, as the general morphology was much larger and more complex than that of either the track block or Elephant tooth. The vertebrae were poorly differentiated by the

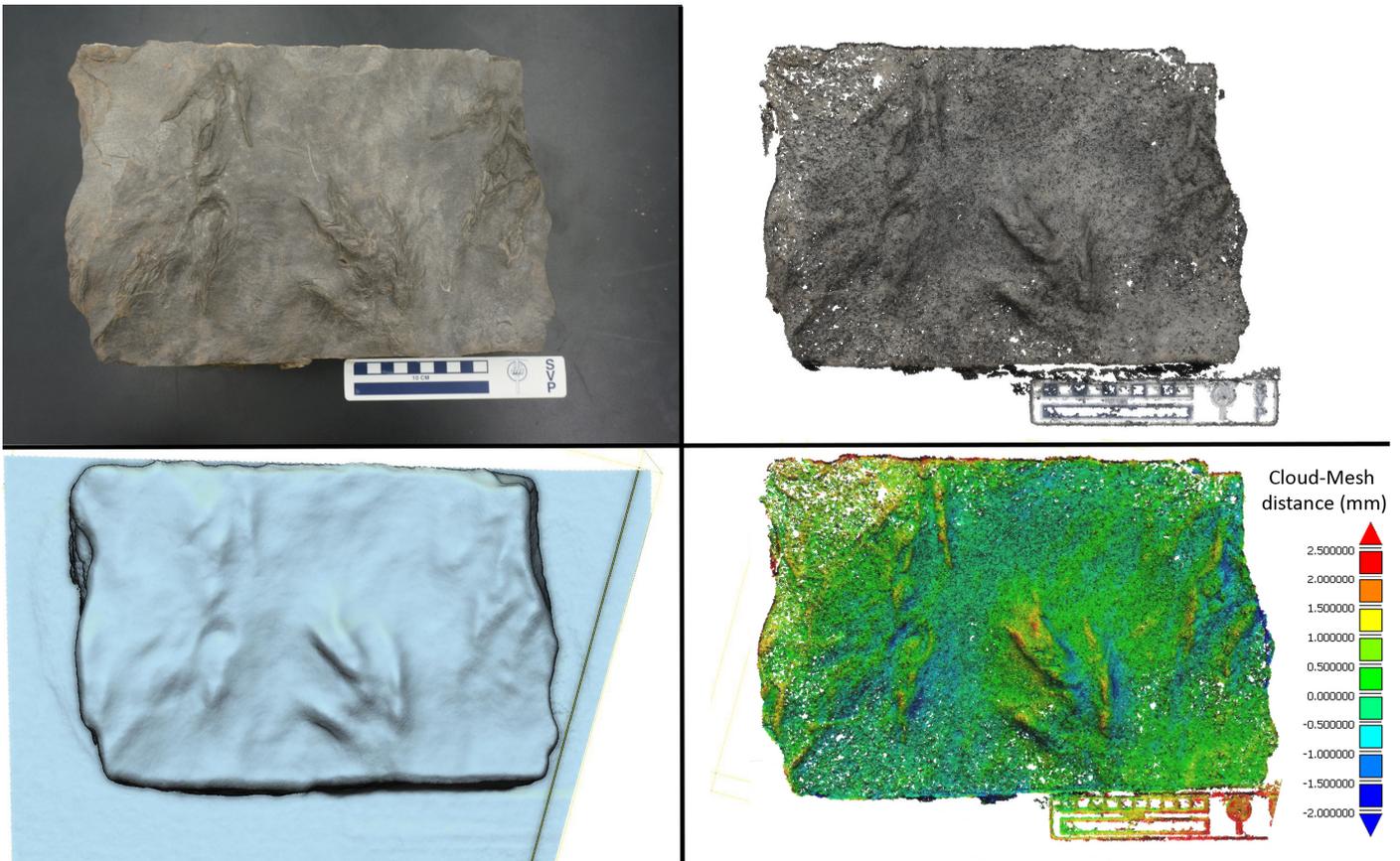


Figure 1 - Top left, the fossil track digitised. Top right, the photogrammetric point cloud produced using VisualSFM. Bottom left, the 3D model generated using the Kinect™ and ReconstructMe. Bottom right, the result of calculating the cloud-mesh distances. The models are generally very correspondent (mostly within +/- 1.5mm), though the Kinect™ mesh is clearly lacking the finer details.

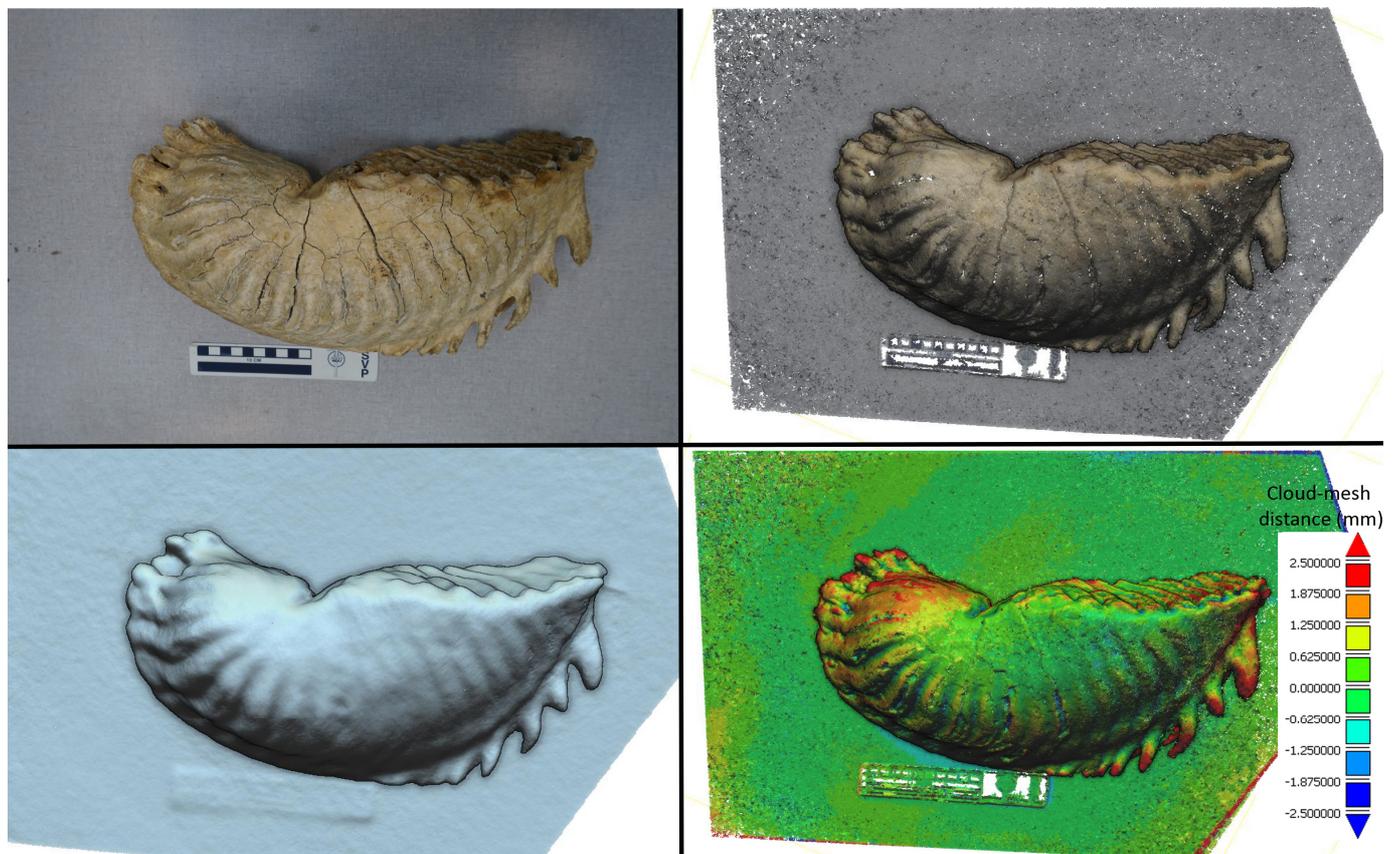


Figure 2 - Top left, the Elephant tooth used. Top right, the photogrammetric point cloud produced using VisualSFM. Bottom left, the 3D model generated using the Kinect™ and ReconstructMe. Bottom right, the result of calculating the cloud-mesh distances. Note that for the majority of the specimen, differences are limited to +/- 5 mm, but some of the more complex morphology, particularly the roots, is up to 16 mm different between the Kinect model and the photogrammetry model.

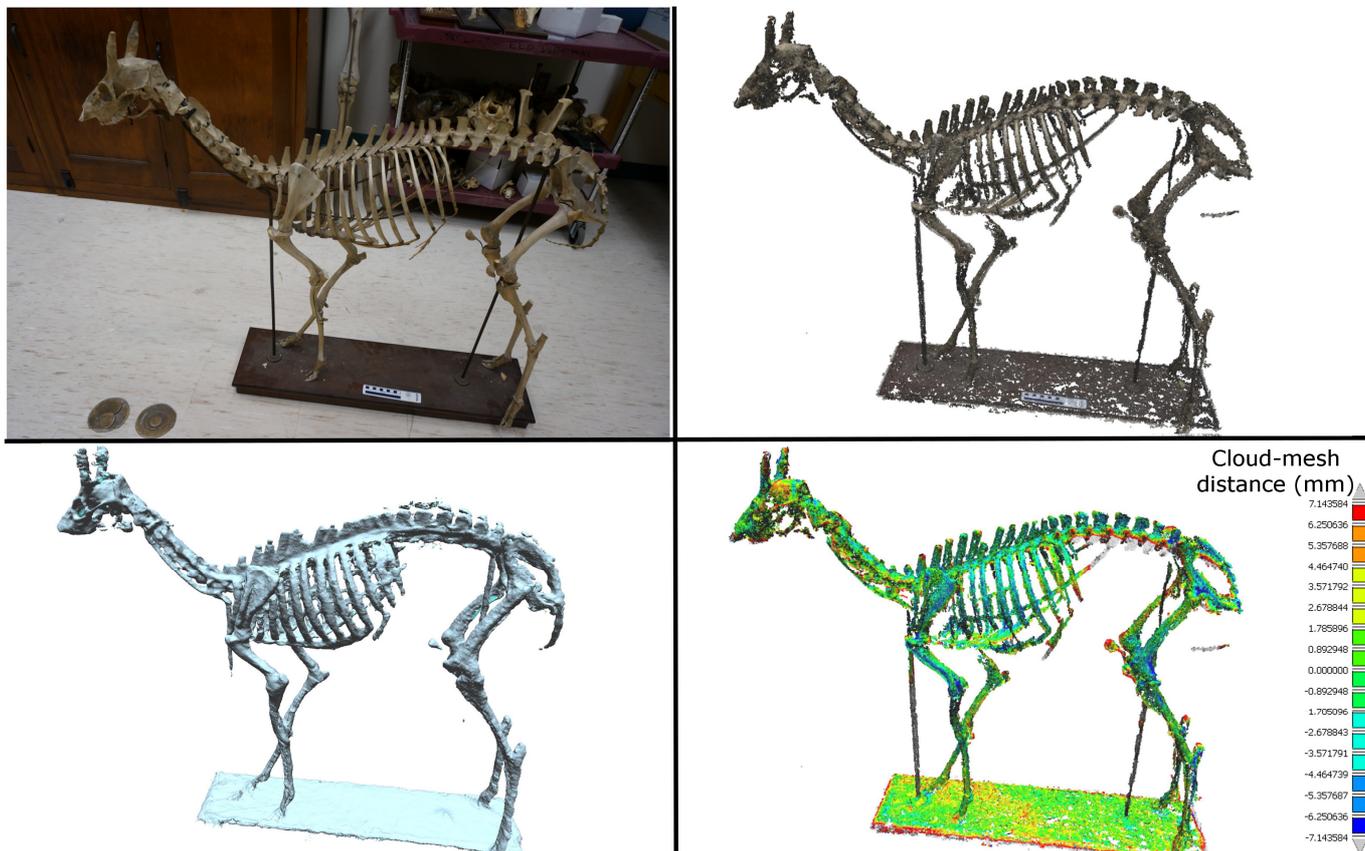


Figure 3 - Top Left image of the mounted Pronghorn skeleton. Top Right , the photogrammetric point cloud produced using VisualSFM. Bottom left, the 3D model generated using the Kinect™ and ReconstructMe. Bottom right, the cloud-mesh distances. Note that although the maximum and minimum distances seem very high, these values are generally limited to the poles between the legs which were too small to be recorded by the Kinect.

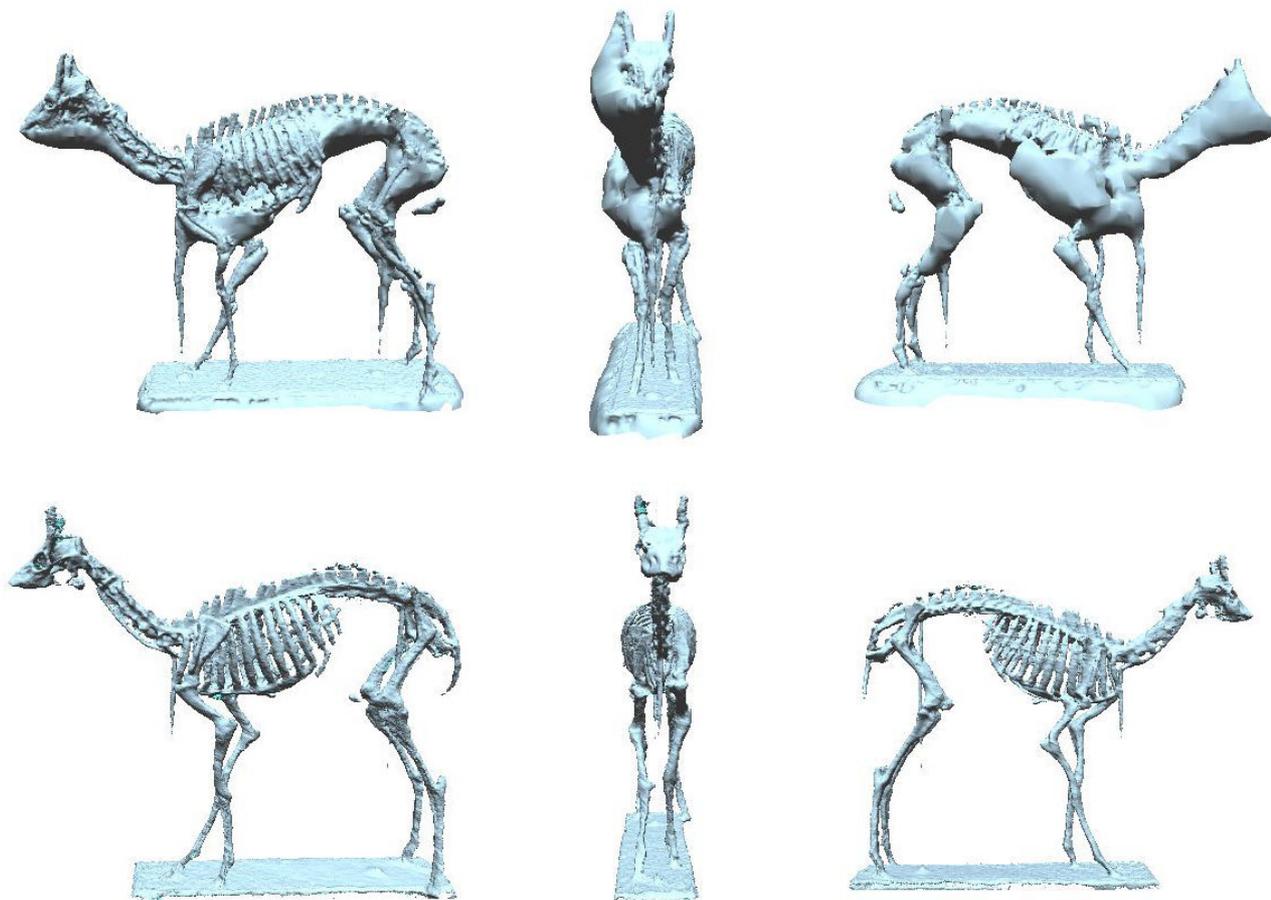


Figure 4 - Difficulties in meshing. Top, the photogrammetric point cloud meshed using the Poisson Surface Reconstruction feature in CloudCompare, Octree depth 10. Bottom, the mesh produced by the Kinect and ReconstructMe.

Kinect™, but this is also true of the photogrammetric model, albeit to a lesser extent (Figure 3). There are some areas where the Kinect™ failed to resolve detail, particularly the tips of the snout and horns, and the transverse processes of the lumbar vertebrae. Conversely, the Kinect™ model successfully recorded more of the tail than did the photogrammetric reconstruction. The maximum difference between photogrammetric point cloud and Kinect™ mesh (300 mm) however is in the poles located between fore- and hind limbs, forming part of the mount. These poles were only 10 mm in diameter, and were not fully recorded by the Kinect™.

DISCUSSION

The data presented here makes it apparent that the Kinect™ can be a useful tool for palaeontologists. Although the sensor, and thus resulting model, lacks the resolution needed for smaller detailed scans, or for archival purposes (which should always be carried out at the highest resolution available), it is able to quickly capture the morphology of a specimen, and produce a digital model that requires only minimal post-processing. One area set to particularly benefit from using the Kinect™ sensor is in body mass and biomechanical simulations, where models are required of individual bones or complete mounted skeletons, but the sub-centimetre detail is unused, such as in convex-hulling specimens to estimate mass (Sellers et al., 2012).

One real advantage of using the Kinect™ sensor in this way is that the model is produced on-the-fly, including meshing, so any errors (e.g. holes in the data collection) can be corrected whilst scanning, contrary to photogrammetry where the quality of the model is not known until after all processing is complete, which can take some time and is algorithms used on photogrammetric point cloud data may fail, or require time-consuming input from the user. Of the samples used here, the pronghorn benefits from this immensely, as the Poisson Surface Reconstruction used to mesh the point cloud struggles to differentiate between close features, such as the ribs, as demonstrated in Figure 4. The meshed photogrammetric model could be improved dramatically by manually segmenting parts of the point cloud prior to surfacing, though obviously this would contribute considerably to the total time of the workflow.

A significant disadvantage to using ReconstructMe or the Point Cloud Library KinFu is that the scanning volume is limited, and must be set prior to scanning. Increasing that volume either decreases the resolution of the scan, or increases the graphics memory used. For the specimens used in this paper, this was not an issue, as the scanning volume could be set relatively small (< 1 m³) and a high resolution and low memory usage maintained. Fortunately, this limitation may be short-lived, as the pre-release version 7 of the Point Cloud Library contains an application 'kinfu_large_scale' which is able to record larger, dynamically sized environments (http://pointclouds.org/documentation/tutorials/using_kinfu_large_scale.php).

Another disadvantage is that the models produced in this paper by the Kinect™ lack any colour information. This is in contrast to photogrammetry, and some traditional laser scanners which can photo-texture their resulting models. However, in addition to the infrared depth sensor, the Kinect™ (and Asus Xtion Pro) also possesses RGB cameras capable of recording colour information. Although in the early stages of development, all of the relevant software packages discussed here are actively developing features to incorporate texture generation.

CONCLUSION

Whilst having some notable disadvantages over other scanning techniques (notably a low resolution, no texture recording, and a potentially limiting scanning volume), the Kinect™ offers a very low cost option for studies where general morphology, rather than small details are required in 3D. Aside from the cost, the Kinect™ sensor in conjunction with the freely available software is fast and provides processed data (meshed models) in real time, allowing checking of data at the point of acquisition. Although perhaps not as versatile to palaeontology as traditional laser scanning or photogrammetry, there are some tasks which would see immediate benefits in terms of ease-of use, particularly where the digitization of mounted skeletons is required. Importantly, this is a rapidly developing technology driven by a major global industry in which market forces drive innovation while simultaneously keeping prices as low as possible.

ACKNOWLEDGMENTS

I wish to thank the respective developers of all software referred to in this paper. This work was supported by a Marie Curie International outgoing Fellowship within the 7th European Framework Programme. I wish to thank

Christine Janis and Kristin Stover (Brown University) for help with access to specimens. I am grateful to Heinrich Mallison, Matteo Belvedere, and the editor for useful comments on improving the manuscript.

REFERENCES CITED

- Andersen, A., Chapman, R. E., Dickman, J. & Hand, K. (2001).** Using rapid prototyping technology in vertebrate paleontology. *Journal of Vertebrate Paleontology*, 21(Suppl. 3), 28A.
- Anderson, P. S. L., J. A. Bright, P. G. Gill, C. Palmer, and E. J. Rayfield. 2011.** Models in palaeontological functional analysis. *Biology Letters*.
- Bates, K. T., R. B. J. Benson, and P. L. Falkingham. 2012.** A computational analysis of locomotor anatomy and body mass evolution in Allosauroidea (Dinosauria : Theropoda). *Paleobiology* 38:486-507.
- Bates, K. T., and P. L. Falkingham. 2012.** Estimating maximum bite performance in *Tyrannosaurus rex* using multi-body dynamics. *Biology Letters* 8:660-664.
- Bates, K. T., P. L. Falkingham, D. Hodgetts, J. O. Farlow, B. H. Breithaupt, M. O'Brien, N. A. Matthews, W. I. Sellers, and P. L. Manning. 2009.** Digital imaging and public engagement in palaeontology. *Geology Today* 25:134-139.
- Chang, Y. J., S. F. Chen, and J. D. Huang. 2011.** A Kinect-based system for physical rehabilitation: a pilot study for young adults with motor disabilities. *Research in developmental disabilities* 32:2566-70.
- Falkingham, P. L. 2012.** Acquisition of high resolution three-dimensional models using free, open-source, photogrammetric software. *Palaeontologia Electronica* 15:1T:15p.
- Falkingham, P. L., L. Margetts, I. Smith, and P. L. Manning. 2009.** Reinterpretation of palmate and semi-palmate (webbed) fossil tracks; insights from finite element modelling. *Palaeogeography, Palaeoclimatology, Palaeoecology* 271:69-76.
- Farlow, J. O., M. O'Brien, G. J. Kuban, B. F. Dattilo, K. T. Bates, P. L. Falkingham, L. Piñuela, A. Rose, A. Freels, C. Kumagai, C. Libben, J. Smith, and J. Whitcraft. 2012.** Dinosaur Tracksites of the Paluxy River Valley (Glen Rose Formation , Lower Cretaceous), Dinosaur Valley State Park , Somervell County , Texas Rastros de dinosaurios del valle del río Paluxy (Formación Glen Rose , Cretácico Inferior), Dinosaur Valley; pp. 41-69, Proceedings of the V International Symposium about Dinosaur Palaeontology and their Environment.
- Furukawa, Y., B. Curless, S. M. Seitz, and R. Szeliski. 2010a.** Clustering Views for Multi-View Stereo.
- Furukawa, Y., B. Curless, S. M. Seitz, and R. Szeliski. 2010b.** Towards Internet-scale Multi-view Stereo. Paper presented at the CVPR, 2010b.
- Furukawa, Y., and J. Ponce. 2010.** Accurate, Dense, and Robust Multi-View Stereopsis. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 32:1362-1376.
- Gidmark, N. J., N. Konow, E. Lopresti, and E. L. Brainerd. 2013.** Bite force is limited by the force-length relationship of skeletal muscle in black carp, *Mylopharyngodon piceus*. *Biology Letters* 9:20121181.
- Hutchinson, J. R., K. T. Bates, J. Molnar, V. Allen, and P. J. Makovicky. 2011.** A Computational Analysis of Limb and Body Dimensions in *Tyrannosaurus rex* with Implications for Locomotion, Ontogeny, and Growth. *PLoS ONE* 6:e26037.
- Izadi, S., D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe, P. Kohli, J. Shotton, S. Hodges, D. Freeman, A. Davison, and A. Fitzgibbon. 2011.** KinectFusion : Real-time 3D Reconstruction and Interaction Using a Moving Depth Camera. Proceedings of the 24th annual ACM symposium on User interface software and technology:559-568.
- Newcombe, R. A., S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. J. Davison, P. Kohli, J. Shotton, S. Hodges, and A. Fitzgibbon. 2011.** KinectFusion: Real-time dense surface mapping and tracking: In *Mixed and Augmented Reality (ISMAR), 2011 10th IEEE International Symposium on*, pp. 127-136. IEEE.

- Panagiotopoulou, O., S. D. Wilshin, E. J. Rayfield, S. J. Shefelbine, and J. R. Hutchinson. 2011.** What makes an accurate and reliable subject-specific finite element model? A case study of an elephant femur. *Journal of the Royal Society: Interface*.
- Rayfield, E. J. 2007.** Finite Element Analysis and Understanding the Biomechanics and Evolution of Living and Fossil Organisms. *Annual Review of Earth and Planetary Sciences* 35:541-576.
- Sellers, W. I., J. Hepworth-Bell, P. L. Falkingham, K. T. Bates, C. A. Brassey, V. M. Egerton, and P. L. Manning. 2012.** Minimum convex hull mass estimations of complete mounted skeletons. *Biology Letters* 8:842-845.
- Stowers, J., M. Hayes, and A. Bainbridge-Smith. 2011:** Altitude Control of a Quadrotor Helicopter Using Depth Map from Microsoft Kinect Sensor. Paper presented at the Proceedings of the 2011 IEEE International conference on Mechatronics, Istanbul, Turkey, 2011.
- Wu, C. 2007.** SiftGPU: A GPU implementation of scale invariant feature transform, SIFT. <http://cs.unc.edu/~ccwu/siftgpu>.
- Wu, C. 2011.** VisualSFM: A visual structure from motion system, Vol. 2012.
- Wu, C., S. Agarwal, B. Curless, and S. M. Seitz. 2011:** Multicore bundle adjustment. Paper presented at the Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, 2011.

Additional images and material can be downloaded at <http://www.jpaleontologicaltechniques.org/>