Head in the Point Clouds – Feet on the Ground

Jörg REKITTKE[1] and Yazid NINSALAM[1]

[1]National University of Singapore, Singapore · rekittke@nus.edu.sg

Abstract

There is no need for soothsayers to predict the future of digital technology for the purpose of site-specific landscape architectural terrain analysis, measurement, and documentation. It will be handheld, respectively mobile and light, broadly affordable, and will allow digital landscape capture in the form of three-dimensional data progressively in high precision, high density, and geo-referenced manner. Landscape data of that ilk, which emanate from high-capacity depth cameras as well as terrestrial laser scanners, can be made available with high visual clarity and aesthetic allure. A common result of the capture technique is their interim, respectively final transformation into point clouds, contingent upon their way of generation, purpose and application. Point clouds of every occurrence and scale at present become the most relevant digital slug as well as virtual modeling clay of the landscape architectural design discipline. With our heads in the point clouds, operating as research-oriented landscape architects in the context of academic design projects, we insist upon the necessity of working in situ and putting our feet on the ground. This axiom enables us to do fieldwork in undocumented or data-poor mega-urban environments, often under the urban canopy and in urban canyons where remote sensing technology fails to act. We experiment with all sorts of independent measuring and urban terrain data gathering technology that can be carried in the field. We continue to advocate the creative utilisation of inexpensive, ordinary tools and off-the-shelf technology. Having said that, the grasp at more sophisticated, more expensive, and more flamboyant technology remained appealing during the past years of our research. When such costly state-of-the-art technology became available to our university department, we took them to harsh megacity environments and tested their capacities and limits. This move signified a review of the theses and methods inherent to our general concept of Grassroots GIS – inclusive surprises and reassurances. For example, working in crowded informal narrow city layouts makes operations with laser-beam emitting terrain scanners Gordian. While the use of unmanned aerial vehicles carrying powerful cameras does not always have to end admissibly. These are the moments when the reliability of mature and frugal technology, as well as improvisation and tinkering skills, may steal the show.

1 Urban Landscape Strata

We came equipped with a variety of portable tools during our recent concerted on-site experimentation – a handheld Global Navigation Satellite System (GNSS), miscellaneous digital cameras, a terrestrial laser scanner, as well as a multiplicity of carrying accessories and mounting devices inclusive of a copter for image gathering, had been conducted in the urban river landscape of the Ciliwung River, situated in the central part of the Indonesian

Wissen Hayek, U., Fricker, P. & Buhmann, E. (Eds.) (2014): Peer Reviewed Proceedings of Digital Landscape Architecture 2014 at ETH Zurich.
ISBN 978-3-87907-530-0.

megacity of Jakarta. Back at the home university, the gathered data were used for the generation of precise and highly detailed spatial models and other three-dimensional copies (point clouds) of the multi-layered landscape. All applied tools and methods conduce to the provision of visible, understandable and designable spatial data in the context of academic studios in urban landscape design. The specific set-up of the described experimentation – focussing on diverse vertical levels of urban landscape recording rather than the horizontal, planar coverage of a site – can be understood as a freehanded reminiscence of the American documentary short film Powers of Ten (1968, rereleased in 1977), written and directed by RAY and CHARLES EAMES (Fig. 1).

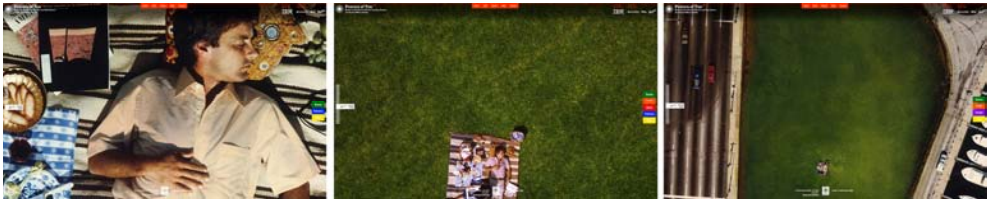

Fig. 1: Three zoom steps from the film Powers of Ten (1977). Left: Camera distance 100 = 1 m. Mid: Camera distance 101 = 10 m. Right: Camera distance 102 = 100 m (Source: Eames Office).

The film is itself an adaptation of the book Cosmic View – The Universe in 40 Jumps, by Kees BOEKE (1957) which depicts the relative scale of the Universe in factors of ten. Starting at a picnic by the lakeside in Chicago, the Eames film leads the viewer to the outer edges of the universe. Returning to Earth, the zoom ends inside a proton of a carbon atom within a DNA molecule in a white blood cell. Our own analytical documentary zoom (Fig. 2) is comparatively unassuming and fragmented, ranging from an inside view of the river (underwater) up to an operating altitude of about thirty meters in the air.

Fig. 2: Three strata of our documentary zoom (2013). Left: Close-range object capture. Mid: 3D scan at eye level. Right: Pole-mounted photography above the urban canopy.

Our experiment distinguishes a multitude of specific recording strata, respectively survey methods, eight of them presented in the paper at hand: 1) Photography in the river; 2) Photography on the river; 3) Close-range object capture (at the riverbank); 4) Close-range spatial capture (at the riverbank); 5) 3D scan at eye level; 6) 3D scan under the urban canopy; 7) Pole-mounted photography above the urban canopy; 8) Airborne overview survey. This spectrum of methods allows a widely complete overview of an exemplary location in a landscape architectural fieldwork operation. In our experiment, the center of reference is fixed – represented by our rubber boat, like the picnic blanket in Powers of Ten. The chosen site sample conduces to the finding and testing of the most suitable tools and techniques for the capture of the different layers and viewing angles of the urban landscape reality.

2 Landscape Puzzle Pieces

In the field, we had to be able to carry all our equipment at once. The diverse tools were used for different pieces of the final landscape puzzle we assembled. Our technological arsenal for the fieldwork in 2013 felt rather heavy and was not inexpensive, compared with earlier missions (Fig. 3). As main pieces of equipment, we carried the newest GoPro HD Hero 3 action camera Outdoor Edition as well as its forerunner Hero 2 model; a powerful remote controlled DJI Phantom quadcopter which can carry GoPro action cameras; a FARO Focus3D high-speed 3D Laser Scanner; a TRIMBLE GeoXH Handheld GNSS for geo-referencing, as well as a PrimeSense 3D Depth Sensor/Camera, with shared roots to a motion sensing input device for computer games. Additionally, for mounting and operation of the diverse tools, a good deal of accessories and spare parts were necessary.

Fig. 3: Technological arsenal (2013). Visible are the GoPro action cameras, the DJI Phantom quadcopter, the FARO Focus3D high-speed 3D Laser Scanner and the TRIMBLE GeoXH Handheld Global Navigation Satellite System.

For post-processing we used established and proven equipment and software components (REKITTKE et al. 2013). Most software applied for the described work is free. The main software used for this paper includes VisualSfM version 0.5.22 and CMPMVS version 6.0. VisualSfM is a graphics user interface (GUI) application of Structure from Motion (SfM) (WU 2011), based upon the measured correspondence of image features, inferring camera poses from a selection of respective photos (WANG 2011). The output of VisualSfM, a point cloud, can be further processed with the help of CMPMVS, multi-view reconstruction software, which generates a textured mesh and reconstructs the surface of the final 3D model (HAVLENA et al. 2010). For the processing and editing of unstructured 3D meshes we use the open source software MeshLab version 1.3.2, a tool developed with the support of the 3D-CoForm project.

2.1 Photography in the river

The GoPro Outdoor Edition came with a protective waterproof housing for the camera. We fixed one as bait at the end of a KAMKOP GoPro fiberglass telescopic mast (Fig. 4). The camera was dipped into the Ciliwung River waters, with camera settings set to time-lapse mode and capture 11 megapixels resolution images at 1 second interval. 165 photos were selected and uploaded into the VisualSfM software for postprocessing. We were not short of images but due to the obvious lack of feature matching from the images captured, we can sensibly conclude that other means of sensors would be needed to obtain data within the depths of the river.

Fig. 4: Photography in the river. The underwater photos could not be transformed into a model but the image material represents our only authentic attempt at documenting the river from inside.

2.2 Photography on the river

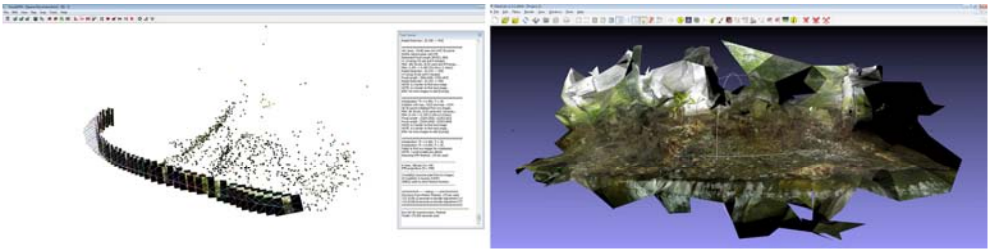

Fig. 5: Left: 60 photos were extracted from a GoPro HD Hero 2 video recording of the bank side. It was used to reconstruct the specific stretch of the bank in VisualSfM. Right: Textured surface model of the bank visualised in MeshLab.

For this part of the puzzle – the lower river edge – we reverted to the rubber boat missions of the penultimate fieldwork trip. Captured via a single camera path method, we used three GoPro HD Hero 2 action cameras, mounted on a pole for three-directional simultaneous dolly shots (REKITTKE et al. 2013). The visual results (Fig. 5) were not ideal – we conducted experiments without any demand for perfection – but still represented the most coherent close-range view and river model of the whole river length (research segment), that could be generated in the frame of our research project.

2.3 Close-range object capture

Common handheld DSLR cameras serve as quick and successful imaging devices for object capture in the field (REKITTKE et al. 2012). A comparable visual quality and final model quality of objects can only be achieved by sensing devices like 3D scanners or depth cameras (see chapters 3.5 and 3.6). As a sample, we took 237 photos of the equipmentfilled boat – the fixed reference object – with a Canon EOS 55D DSLR camera and processed them in VisualSfM. The final textured surface model was exported from MeshLab (Fig. 6).

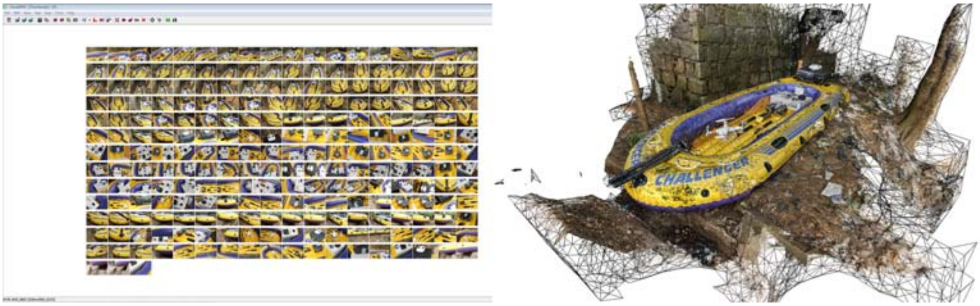

Fig. 6: Left: First step of the VisualSfM workflow – feature detection from a set of 237 photos as input. Right: Final textured surface model, generated in MeshLab.

2.4 Close-range spatial capture

For the capture of the proximate space beyond the reference object (boat) we tested the GoPro HD Hero 3 camera, set to capture images at time-lapse mode of 1 second interval. With work concerning any optical device, it is indispensable to take the lighting conditions into consideration during the documentation process. The glare of sunlight, intense lighting contrast as well as shadow effects can compromise the quality and usability of captured images. We brought home successful material as well as outcomes of the mentioned difficulties (Fig. 7). In one instance, a large number of 171 quality photos from the GoPro HD Hero 3 camera, processed in VisualSfM, resulted in a rather fragmentary model because of inconsistent scene luminance during the capture process.

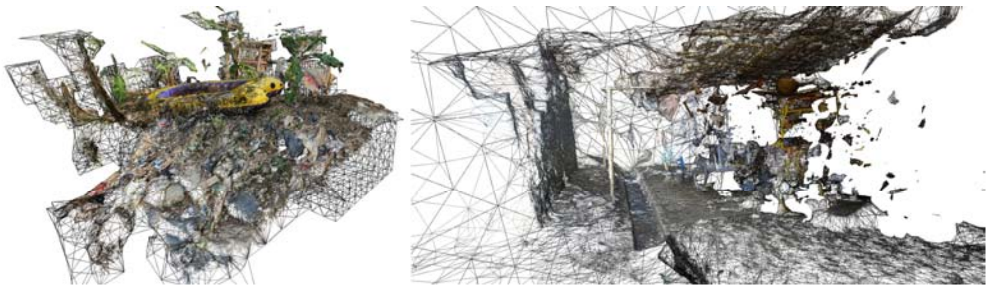

Fig. 7: Left: Expedient capture and modeling of spatial context beyond the reference object. Right: Inconsistent scene luminance, amplified by an improvised awning, led to a highly fragmented model.

2.5 3D scan at eye level

Our biggest gun in the Grassroots GIS research context as yet (REKITTKE & PAAR 2010), the FARO Focus3D scanner is almost too heavy to be carried in the informal urban environment that we work in. It uses phase shift technology in the measurement of distances between scanned objects to the sensor. The phase shift of the returning pulse and stored pulse determines the distance travelled by the laser beam. The sensor’s field of view extents to a range of 360° horizontal and 305° vertical. The resolution of the colour image captured stands at 70 megapixels and has a dynamic colour feature which automatically adapts to the brightness of the site. For our site sample, two sets of scans were captured, one at routine resolution (6.136 mm) and the other at maximum resolution (1.534 mm) within a 10 meter scan radius. We tested the maximum option to exhaust the capacity of the scanner. It captured 16 times more points and took 10 times longer than the routine setup. Apart from that, maximum resolution scans are only justified when a differentiation of minute subtleties is necessary, for example in archaeological or architectural survey projects. The FARO data are registered via the software FARO Scene. Target based registration composes partial scans to a whole (Fig. 8).

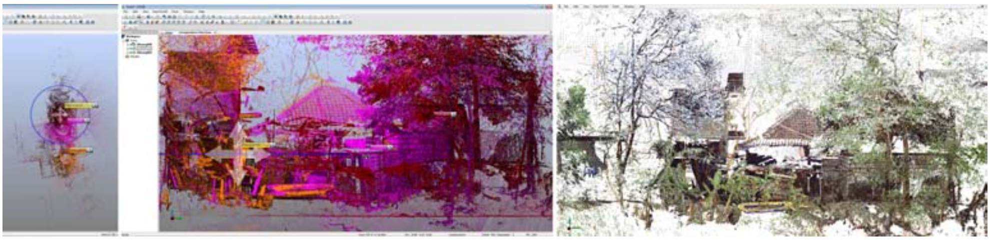

Fig. 8: Left & Mid: Manual registration (matching) via FARO Scene, merging the point cloud data of 3 separate scans (red, pink, orange). Right: The FARO 3D scan looks like a mélange of a precise drawing and a three-dimensional photography.

For final processing we then exported the registered point clouds to third party software programmes, namely Bentley Systems Pointools V8i or Rhinoceros 5.0. A number of tools can be used to clean the data and convert the point clouds into solid surfaces. This process also removes scan outliers, refines scan points, reduces redundancy and merges fractional point clouds. Terrestrial scanners have achieved impressive results and illustrate the kind of digital slug and virtual modeling clay future generations of designers will splash about.

2.6 3D scan under the urban canopy

Under the urban canopy – thick vegetation, roofs, architectural overhangs, superstructures, urban canyons – remote sensing technology remains to be blind and the use of photography for close range structure from motion capture can be limited. For this reason we brought a depth camera into the field, the PrimeSense Carmine 1.08 3D Sensor. This device allows a real-time dynamic capture of 3D colour images – even in the dark – and of audio information. All sensory information is tied to timing alignment and can be transferred via USB 2.0. The sensor’s field of view extents to a range of 58° horizontal, 45° vertical, and 70° diagonal. The 3D image is captured at 640 by 480 pixels resolution (VGA) while the colour image (RGB) is captured at 1280 by 960 pixels. The resulting point cloud has a spatial X/Y resolution of 3.5 mm at 2 m distance and Z at 0.3 cm at 1 m distance from the sensor. The sensor has a maximum image throughput of 60 fps and an operational range of 0.8 to 3.5 m. The user has to walk very slowly in order to ensure that enough spatial features are obtained within the field of view. We dynamically capture the point cloud data via FARO Scenect, a free point cloud software currently under development, which enables the user to view and capture objects and environments in real-time. The scans can be exported to different file formats for processing in third party applications.

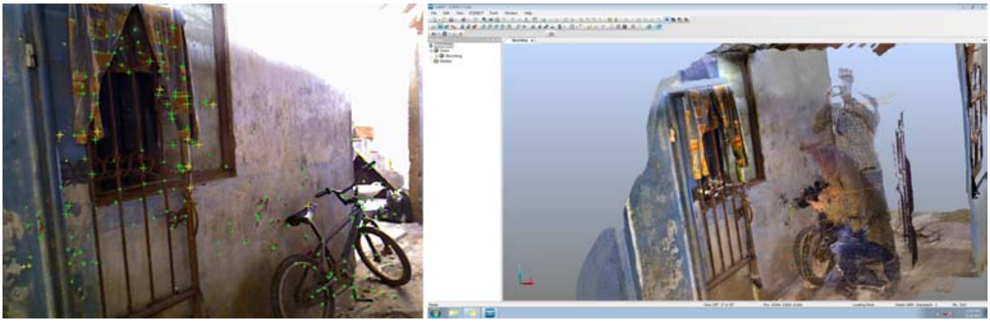

Fig. 9: Left: The camera's field of view. Green crosses indicate used features while tracking, yellow indicates unused but suitable features for tracking. Right: The scene in FARO Scenect. The person recorded at different stages moved through the site at time of scan, a feature that can be used creatively.

2.7 Pole-mounted photography above the urban canopy

Our inexpensive and well-tried six meter long KAMKOP GoPro telescopic camera mast, made of GFK and weighing only 850 grams, is in all likelihood the most successful aid for our approach thus far. By carrying and panning the pole-mounted GoPro HD Hero 3 Outdoor Edition, set to capture 11 megapixels resolution images at 10 photos per second and tilted 30° from the pole, we were able to reach heights up to 8.5 meters from the ground and generate amazing perspectives above the urban canopy (Fig. 10).

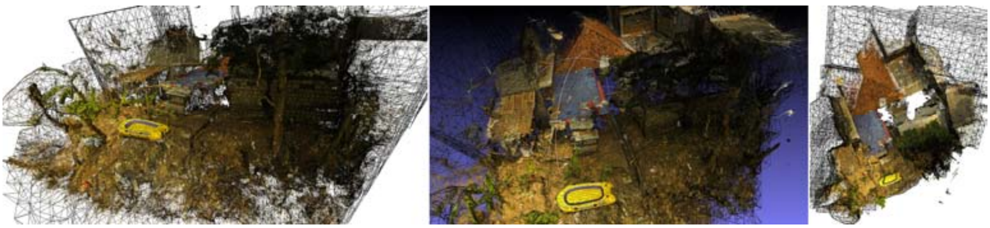

Fig. 10: Extensive coverage (863 images) of the site with the use of the KAMKOP telescopic mast enables an infinitely variable close-range bird's eye view. A total of 11372 756 vertices and 22744134 mesh faces were generated from a textured surface digital model of the bank visualised in MeshLab.

2.8 Airborne overview survey

For infinitely variable overviews from up to 50 meters altitude, we used a remote controlled DJI Phantom quadcopter which carried the GoPro action cameras.

Fig. 11: Left: The Phantom copter produced overviews from specific height intervals, evident as three main levels (approx. 10, 20 and 30 m). Mid & Right: Overview (30 m altitude) site model, visualised in MeshLab.

Such copters are powerful and flexible tools but bare two main risks: a) they may lose connection to the radio control network, and b) they may finally crash. In the densely populated city areas we work in, this constitutes a considerable peril. For this reason we started experimenting with balloons. They are safe as tools but are at the mercy of the wind.

The patchy visual results we show here (Fig. 11), are textured 3D models generated from the copter-gathered aerial photos. The original photos are high-grade and consistently cover the whole area in high resolution, better than any Google imagery.

3 Going where Google hasn’t

Our most relevant reference projects are those from teams going where Google hasn't (KOCH 2103), using handheld equipment and unorthodox methods to capture real-world environments. In 2013, a team from Harvard University travelled to India to document and analyse the processes involved in the Kumbh Mela festival in Allahabad. It is the world’s largest religious festival, occurring every twelve years, and drawing millions of visitors to a temporary tent city on the banks of the Ganges and Yamuna. In order to understand the urban processes, students of Rahul Mehrotra from Harvard GSD used handheld DSLR cameras, inclusive a high-capacity camera mounted to a kite which flew over the vast site (Ibid.). Meanwhile, in Cantagallo, an informal settlement in Lima, Peru, Jeff Warren from MIT Media Lab, Center for Future Civic Media, partnered with local residents to generate maps with the community. After using digital cameras with continuous mode shooting lofted by kites or balloons, the geographically referenced aerial images were overlaid on Google Maps, showing a resolution one hundred times higher than the existing Google imagery (WARREN 2011). A third example is the work of Jonathan Roberts, research program leader at the Australian Commonwealth Scientific and Industrial Research Organisation (CSIRO). He showed how a hand-held spring-mounted laser scanner (BOSSE et al. 2012) can be used to build a 3D map of the interior of the leaning Tower of Pisa – by walking freely and within a short time (BOWDLER 2013). There is a whole universe out there which has not been illuminated by global super technology and waits to be researched in detail – on foot.

4 Conclusion and Outlook

It seems to be a common denominator for all research teams which currently work with their heads in the point clouds and their feet on the ground, that they are able to generate three-dimensional documentary information which is precise and viable but inchoate. The interesting side of this common denominator is the intelligibleness that none of these teams – including us – seek a perfect portraiture of the real world but a usable representation of space and materiality for the purpose of understanding or change instead. The incompleteness and coarseness of the gathered spatial data still allows the abstraction that is indispensable for any design process. Progressing technology will lead to better results all the time but we may never lose track of design relevant essentials like aesthetics or artistic license – turning out not to see the wood for the trees.

References

BOEKE, K. (1957), Cosmic View: The Universe in 40 Jumps. John Day Company, New York.

BOSSE, M., ZLOT, R. & FLICK, P. (2012), Zebedee: Design of a Spring-Mounted 3-D Range Sensor with Application to Mobile Mapping'. IEEE Transactions on Robotics, 28 (5), 1104-1119.

BOWDLER, N. (2013), Tower of Pisa interior mapped in 3D. BBC News Technology, 17, September 2013. Available from: http://www.bbc.co.uk/news/technology-23856662 (accessed 15 December 2013).

EAMES, C. & EAMES, R. (1968 "Sketch" version / 1977 "Final" version), Powers of Ten. Short film. Running time 8 min ("Sketch" version), 9 min ("Final" version). IBM, USA.

HAVLENA, M., TORII, A. & PAJDLA, T. (2010), Efficient Structure from Motion by Graph Optimization. Computer Vision – ECCV 2010. Springer, Berlin/Heidelberg.

KOCH, K. (2013), Mapping a megacity's metabolism: Design School team finds lessons for urban planners at India's Kumbh Mela gathering, HARVARDgazette, February 5, 2013. Available from: http://news.harvard.edu/gazette/story/2013/02/mapping-a-megacitys metabolism/ (accessed 15 December 2013).

REKITTKE, J. & PAAR, P. (2010), Grassroots GIS – Digital Outdoor Designing Where the Streets Have No Name. In: Buhmann, E., Pietsch, M. & Kretzler, E. (Eds.), Reviewed Proc. of Digital Landscape Architecture 2010 at Anhalt University of Applied Sciences. Wichmann, Berlin/Offenbach, 69-78.

REKITTKE, J., PAAR, P. & NINSALAM, Y. (2012), Foot Soldiers of GeoDesign. In: Buhmann, E., Ervin, S. M. & Pietsch, M. (Eds.), Peer Reviewed Proceedings of Digital Landscape Architecture 2012 at Anhalt University of Applied Sciences. Wichmann, Berlin/Offenbach, 199-210.

REKITTKE, J., PAAR, P. & NINSALAM, Y. (2013), Brawn and Technology under the Urban Canopy. In: Buhmann, E., Ervin, S. M. & Pietsch, M. (Eds.), Peer Reviewed Proceedings of Digital Landscape Architecture 2013 at Anhalt University of Applied Sciences. Wichmann, Berlin/Offenbach, 12-21.

REKITTKE, J., PAAR, P., NINSALAM, Y. & LIN, E. (2013), Digital Reconnaissance. Journal of Landscape Architecture, 8 (1), 74-81.

WARREN, J. (2011), Grassroots Mapping. In: Design with the Other 90%: CITIES, Cooper-Hewitt, National Design Museum. Available from: http://www.designother90.org/solution/grassroots-mapping/ (accessed 15 December 2013).

WANG, Y.-F. (2011), A Comparison Study of Five 3D Modeling Systems Based on the SfM-Principles (Technical Report No. Visualsize Inc. TR 2011-01), Goleta, USA, 1-30.

WU, C. (2011), VisualSFM: A Visual Structure from Motion System. Available from: http://homes.cs.washington.edu/~ccwu/vsfm/ (accessed 27 November 2013).