Brawn and Technology under the Urban Canopy

Joerg REKITTKE, Philip PAAR and Yazid NINSALAM

1 Base of Operations

Considering the growing variety of electronic tools being deployed by city and geo designers and researchers in fieldwork operations, we distinguish two main groups of users. The first group we refer to as ‘aviators’. Namely: research teams who profit directly from a large range of high-end technology – satellites and their data, all sorts of flying super-technology and respective precise aerial imagery. The second group we refer to as ‘foot soldiers’ (REKITTKE, PAAR & NINSALAM, 2012). At this point we feel up to put forward the alternative term ‘earthlings’, due to political correctness and peer scrupulosity. Developing our research and fieldwork practice as data gathering and 3D modelling earthlings in the third year of a related research project, we continue in improving and optimizing our light, portable and inconspicuous equipment. We stand by our claim – despite all the theoretically available remote sensing technology – that in order to build complete and highly detailed digital 3D models of complex terrain and urban territory, direct contact to ground and detail remains indispensable. Compared to the sophisticated machines and efficient methods in remote sensing, the craft of the earthling appears to be more intricate and laborious – but it can deliver unique results, which we subsume under the term of 'Grassroots GIS' (REKITTKE & PAAR, 2010). By now, we can be certain that the scientific and methodological niche of the GeoDesign earthling is located under the *urban canopy* and in *urban canyons*, where remote sensing technology is proven to be blind and ineffective. In this paper, we meticulously describe our most recent advances and let downs in the development of a method of on-site data and image gathering. This evidently creates a new kind of high precision three-dimensional puzzle, which makes a widely inaccessible and undocumented piece of *under the urban canopy* terrain visible, understandable and designable. The work we present focuses on the space defined by the Ciliwung River in the Kampung Melayu-Bukit Duri district, an informal urban segment in Jakarta, Indonesia. We bring ourselves, together with students, to this area in the course of the research module Landscape Ecology in the 'Future Cities Laboratory' (REKITTKE & GIROT, 2012), hosted and managed by the Singapore ETH Centre for Global Environmental Sustainability (CAIRNS et al., 2012). With our feet in the sludge, waste and faeces of the Ciliwung River, the help of a lot of brawn and some new tools, which are typically used by free fallers, bungee jumpers and other contemporaries tired of life, we also experimented with toy drones, featuring inbuilt cameras and representing the approaching armada of flying devices for the purpose of image data gathering in science, research and design.

2 Earthlings at Work

The described experiments have been conducted during a fieldwork campaign over a period of six days and invariably focussed the Ciliwung River and its edges. Our mission

Buhmann, E., Ervin, S. M. & Pietsch, M. (Eds.) (2013): Peer Review Proceedings of Digital Landscape Architecture 2013 at Anhalt University of Applied Sciences. ISBN 978-3-87907-527-0.

was to document, model and design the course of the river between 6°13'28. 07"S, 106°51'49. 35"E and 6°13'2. 23"S, 106°51'28. 75"E – a distance of 2.86 km. Initially we conducted our data gathering process by walking, picture taking and filming along the banks. This proved to be incredibly tedious when our feet sank too deep into the sludge, and finally came to a halt when the banks became too steep to continue. But to this point we could successfully deploy some of our newest equipment (Fig. 1) and give some new methodological ideas a try.

Fig. 1: Technological arsenal (2012): Multiple sets of the GoPro HD HERO2 action camera Outdoor Edition; Parrot AR Drones 2.0 inclusive accessories and spare parts; Leica Disto D8 laser meters; Garmin GPS-MAP 60CSx handheld tracking devices; Apple iPhones and iPads for drone flight control; additional DSLR cameras; camera poles etc. (Photo: Rekittke)

Earthlings roam as backpackers, they have to be able to carry all their equipment in the field. The technological arsenal of the fieldwork at hand felt considerable. We carried along three sets of the GoPro HD HERO2 action camera Outdoor Edition – video cameras fitted in waterproof housings and attached to 3-way pivots on a telescopic mast; three iPad-controlled Parrot AR Drones 2.0 – remote controlled by the gratis AR.Free Flight App – inclusive a good deal of accessories and spare parts. These inexpensive plaything quadrocopters, equipped with a built-in on-board video camera for wireless video recording and transmission via a picture resolution of 1280 x 720 dpi, enabled us to also visually illuminate completely inaccessible or hidden spots of the site. In addition we carried several Leica Disto D8 laser meters as well as Garmin GPS-MAP 60CSx handheld tracking devices as well as Apple iPhones and iPads for further navigation and measuring applications. For our photographic work, we would always carry high-capacity DSLR cameras. Two purchases turned out to be unexpectedly useful for our mission. Firstly, a six meter long telescopic camera mast, model Kamkop GoPro, made of GFK and weighing only 850 grams, allowed us to generate unique image material from bird's eye view (Fig. 2). The long pole almost beats flying devices concerning precision and ease of handling. Secondly, an offhanded purchased rubber boat, a Challenger 4 Inflatable – the mother of all dinghies, finally brought us onto the river. The originally supplied plastic oars broke during the first of six trips down the river but the craft with the capacity of four people, sustaining a load of up to 400 kg, transformed bogged earthlings into successful river rafters. All our items are consumer grade, easily accessible and comparatively low cost.

Fig. 2: The six meters long telescopic mast Kamkop GoPro allows for work under as well as above the urban canopy (Photo: Paar)

3 Terrain Capture Techniques

We tested miscellaneous terrain capture techniques under the hypothesis that the combination and overlap of picture material generated by independent, multiple camera sources would, after post-processing, result in better quality models of the Ciliwung River than it would be feasible with the material generated by a single camera path. For the multiple sources method we used: a) a Canon DSLR camera for high-quality close-range shots of accessible segments along the river; b) a single GoPro HD HERO2 action camera, attached to the end of the Kamkop GoPro telescopic mast, in order to capture images through a *swaying* motion across river and terrain (Fig. 3); c) the built-in front camera of the Parrot AR Drone 2.0 (Fig. 4), delivering a sequence of pictures from the captured video; d) a set of three GoPro HD HERO2 action cameras, mounted on a pole for three-directional simultaneous dolly shots, carried down the river in a rubber boat mission (Fig. 5).

Fig. 3: *Swaying* a GoPro HD HERO2 action camera, attached to the Kamkop GoPro telescopic mast, in order to capture river and terrain

Fig. 4: Flying the Parrot AR Drone 2.0 over river and terrain – under the canopy

Fig. 5: Three GoPro HD HERO2 action cameras, mounted on a pole for three-directional simultaneous dolly shots – on the river

For the single camera path method we merely used the above mentioned set of three pole and boat mounted GoPro HD HERO2 action cameras, adjusted for three-directional simultaneous shots.

4 Postprocessing of Image Material

To come to the point, our hypothesis – the more the better – crystallised being not axiomatically true. We will try to explain this conclusion in conjunction with the explanation of our postprocessing methods. The hardware we use for postprocessing represents common standards in computer gamer circles – Grassroot GIS asks for affordable off-the-shelf articles: a laptop with a 2.9Ghz 3rd Generation Intel Core i7-3920XM Processor, 32GB RAM running 64-Bit Windows 7. All software we apply for the described work is free. For the postprocessing of former fieldwork material we gained experience with the web-based software Autodesk 123D Catch that only allows for a maximum upload of less than one hundred photos at once. For the described mission we experimented with processing of up to three thousand photos in one session, which can be done with a combination of the free programmes VisualSfM version 0.5.20 and CMPMVS version 4.0. VisualSfM is a graphics user interface (GUI) application of Structure from Motion (SfM) (WU, 2011). The system is based upon the measured correspondence of image features, such as points, edges and lines – inferring camera poses from a selection of respective photos (WANG, 2011). Subsequently the measured camera poses are processed by CMPMVS, a multi-view reconstruction software, which generates a textured mesh and re-constructs the surface of the final 3D model (HAVLENA et al., 2010). VisualSfM attends to the needs of both basic and advanced users, the application runs by executing the supplied scripts and parameters. The resulting product of VisualSfM – a point cloud – then can be processed with the help of CMPMVS. The standard workflow (Fig. 6) for VisualSfM and CMPMVS can be abstracted as follows:

1) Import of all the images into the SfM workspace as a readable image file in jpg-format.
2) Feature detection and full pairwise image matching, this generates a set of measured image features and correspondence (WU, 2007).
3) Incremental reconstruction, this will start the multicore bundle adjustment, which calculates the 3D point positions and camera parameters from the set of measured image features and correspondence (WU et al., 2011)
4) Dense reconstruction by using a multi-view stereo (MVS) application – CMPMVS – this performs the surface reconstruction from the point clouds (HAVLENA et al., 2010).

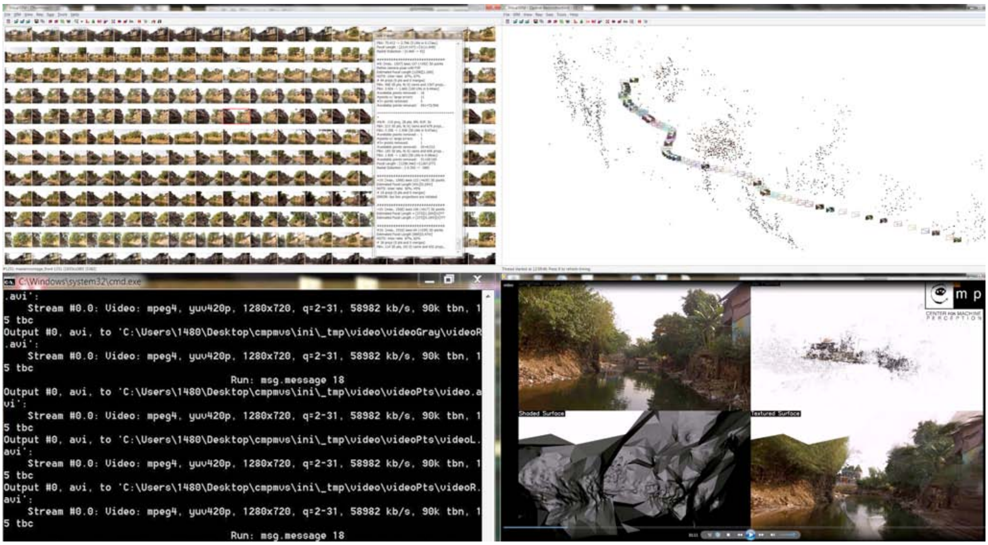

Fig. 6: VisualSfM workflow: Feature Detection and Bundle Adjustment (top l.+r.); Dense and Surface Reconstruction in CMPMVS and video output (bottom l.+r.)

The Autodesk software 123D Catch offers limited functionality concerning a detailed determination of the resulting model, but it is the best choice for basic users. Features that hitherto overtrump VisualSfM include the implementation of a reference scale allowing scaling of the entire 3D model, taking of reference measurements, animation of the 3D model in the software directly, and a complete 3D surface reconstruction pipeline. The academic project VisualSfM thrives on constructive feedback by an active user community that constantly seeks advice and makes suggestions on a Google Groups discussion thread. VisualSfM stores and processes the selected image material on a local folder. This bypasses any time consuming and often impractical web-based upload action and enables the user to check feature matching and sparse reconstruction processing already in the field. A critical issue during time-constrained field campaigns, because in semester-long academic design studios a return visit is often not an option. Main goal of our research is to achieve the best model quality possible. Therefore, VisualSfM currently turns out to be the optimal cost-free option (Fig. 7). It allows a flexible addition of selected photos to integrate and improve existing models, and a manually performed initialization of the automatic feature detecting process by the determination of a proper pair of photographs. A 'Pause' and 'Run' button facilitates the integration of further photos during the feature-matching phase. Such sophisticated fine-tuning is not possible in 123D Catch, it prompts the user to manually define at least three corresponding points on the sequential images and does not allow pausing a running reconstruction process.

Fig. 7: Screenshot of a 3D model generated via Autodesk 123D Catch (left); screenshot of the same model generated via VisualSfM (right)

5 Less Input, more Model

The comparison between the postprocessing results of single and multiple source images for the reconstruction of terrain (modelling) along the Ciliwung River will illustrate the advisable balance between image input and model output. Before looking into the results we need to distinguish the different methods of image processing. We work with two methods, both based on manually selectable settings within the VisualSfM platform. The first method makes use of an *exhaustive pairwise matching* technique, the second method makes use of *sequential pairwise matching*. We fed the material of our multiple sources capture technique (see point 3) into VisualSfM by selecting the exhaustive pairwise matching option. This option attempts to match every single image and their features to all corresponding features of all available images within the supplied dataset. From video sequences one frame of every 30th frame was extracted, in order to have regular image coverage of the environment. Images from the Canon DSLR shot in RAW were downsized to match the resolution of the extracted video images. In our tests of the exhaustive pairwise matching option we processed a total of 543 images along a segment of the river (Fig. 8). For the processing of the material gained via our single camera path capture technique (see point 3) we selected the sequential pairwise matching option in VisualSfM. This option is optimised for the processing of images that had been taken in a continuous sequence (path). To compare the results of the sequential pairwise matching option with those of the exhaustive pairwise matching option, we selected – for the same river segment – a total of 100 sequential images from our HD resolution video footage, by extracting one frame of every 30th frame.

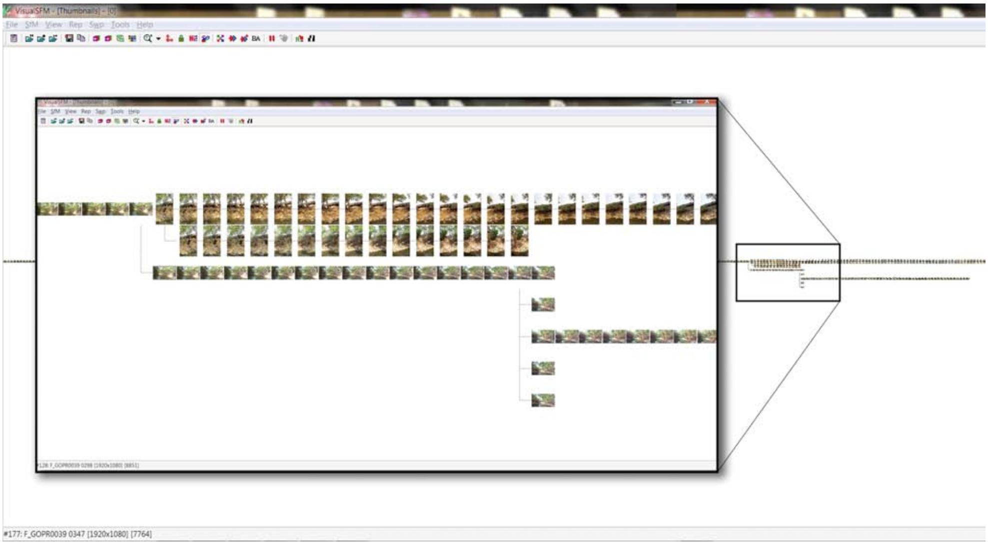

Fig. 8: VisualSfM *spanning tree*, indicating the corresponding matches between images from multiple sources. The highlighted window shows an instance, where 72 (of a total of 543) images matched, originating from the AR Drone 2.0 and Canon DSLR camera.

As a result we observed that in the reconstruction of the river corridor sequence, the sequential pairwise matching option outplays the exhaustive pairwise matching option of VisualSfM in both the quality of the resulting model and time taken to post process the images. Although some model sets successfully integrated images from various sources – the model quality is similar to that of the sequential pairwise matching operation as it includes the same main image source (GoPro HERO2 HD cameras), the most coherent models result from single source sequential images (Fig. 9). However the exhaustive pairwise matching option can improve segments of the geometry where the model lacks data due to insufficient image coverage. The quality of the model is determined by the quality and integrity of the images from the source. Images from multiple sources can be added successfully to a single sequential dataset as a model improver in place of using it as the primary model generating method. This is especially relevant for our fieldwork environments where physical constraints like low-hanging branches or other urban canopy may hinder the continuous filming of the entire river sequence and contribute to intermittent recording. An issue worth mentioning concerning the multiple sources input in the VisualSfM exhaustive pairwise matching workflow is the fact that the subsequent surface reconstruction procedure of the CMPMVS software is unable to generate geometry from image sources that do not share the same pixel dimensions.

Fig. 9: Three-dimensional puzzle of partial terrain models, positioned in Google Earth (top). Video capture (middle) of the river terrain *under the urban canopy*, and textured model of the same site (bottom).

On these grounds it is most advisable to use cameras with same specifications (and settings), when adding images to supplement the model generating process via exhaustive pairwise matching. Maybe the processing time proves crucial for the method of choice. For the same river segment sample, the completion of the feature matching process in VisualSfM, using the exhaustive pairwise matching option and a total of 543 images, took six days of uninterrupted computation. The sequential pairwise matching method, using 100 omnidirectional images, took three hours for the same task.

6 Homework

While the described fieldwork mission took quite a bit of brawn, it remained rewarding and evolved our understanding of the technology related challenges facing Grassroots GIS. The quality of image and data material is key and time is on our side. The craft of an earthling's fieldwork will continue to require rolling up their shirtsleeves, however precision, accuracy and dependability of available equipment will rapidly advance. Coherence as well as complexity of our technique and technology will have to be increased. The next generation of outdoor action cameras and camera carrying flying devices is in the stores yet. Our low-budget approach sets us certain financial limits but we begin to leer at portable 3D scanner systems with real-time 3D reconstruction capability and will have to decide if such technology will be part of our future baggage. These scanners are quickly becoming affordable and are paired with software development kits, which enable developers to create various new applications (ANDERSEN et al., 2012). Another future standard for the work in data poor environments will be made up by extremely precise portable satellite navigation systems with global coverage (GNSS) and enhancements by Differential Global Positioning Systems (DPGS). Eventually we will have to experiment with still expensive Light Detection And Ranging (LIDAR) technology, which is being integrated into standard GIS software (JORDAN, 2012). The trend points towards a conflation of GIS data sets and their corresponding image-based geometries. Conflation – in GIS – is defined as the process of combining geographic information from overlapping sources in order to retain accurate data, minimize redundancy, and reconcile data conflicts (LONGLEY et al., 2001). We undertake not to reduce our commitment in brawn but we might have to increase our investment in technology.

7 Acknowledgements

Special thanks go to Professor Christophe Girot, in his capacity as head of the research module "Landscape Ecology" of the Future Cities Laboratory, Singapore ETH Centre for Global Environmental Sustainability. Our work is embedded in the Future Cities Laboratory, in form of Design Research Studios as well as in form of PhD research. In the same way we would like to thank Professor Wong Yunn Chii, head of the NUS Department of Architecture, for his unorthodox support – in face of our out of the ordinary travel preferences.

References

Andersen, M. R., Jensen, T., Lisouski, P., Mortensen, A. K., Hansen, M. K., Gregersen, T. & Ahrendt, P. (2012), *Kinect Depth Sensor Evaluation for Computer Vision Applications* (Technical Report ECE-TR-6). Department of Engineering, Aarhus University, 1-39.

Cairns, S., Gurusamy, S., Prescott, M. & Theodoratos, N. (Eds.), *Rivers in Future Cities: The Case of Ciliwung, Jakarta, Indonesia*. Module VII: Landscape Ecology, FCL – Future Cities Laboratory Gazette, 12, Singapore ETH Centre for Global Environmental Sustainability.

Havlena, M., Torii, A. & Pajdla, T. (2010), *Efficient Structure from Motion by Graph Optimization*. In: Computer Vision – ECCV 2010. Berlin/Heidelberg, Springer-Verlag.

Jordan, L. (2012), *Integrating LiDAR and 3D into GIS*. Available from: http://www.youtube.com/watch?v=0USiCcsiFqE (accessed 25 November 2012).

Longley, P. A., Goodchild, M. F., Maguire, D. J. & Rhind, D. W. (2001), *Geographic Information Systems and Science*. John Wiley & Sons, 134.

Rekittke, J. & Girot, C. (2012), *Cali Ciliwung Experiment.* In: S.L.U.M. Lab. Sustainable Living Urban Model, Edition 7, 68-69.

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, Berlin/Offenbach, Wichmann Verlag, 199-210.

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. Berlin/Offenbach, Wichmann Verlag, 69-78.

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. (2007), *SiftGPU: A GPU implementation of Scale Invariant Feature Transform (SIFT)*. Available from: http://cs.unc.edu/~ccwu/siftgpu (accessed 1 November 2012).

Wu, C. (2011), *VisualSFM: A Visual Structure from Motion System.* Available from: http://homes.cs.washington.edu/~ccwu/vsfm/ (accessed 3 November 2012).

Wu, C., Agarwal, S., Curless, B. & Seitz, S. M. (2011), *Multicore Bundle Adjustment*. Presented at the IEEE Computer Vision and Pattern Recognition (CVPR), 3057-3064.