3D DATA GENERATION USING LOW-COST CROSS-VIEW IMAGES

R. Qin[a,b*], S. Song[a], X. Huang[a]

[a] Dept. of Civil, Environmental and Geodetic Engineering, the Ohio State University (OSU), 2070 Neil Ave., Columbus, OH, USA
[b] Dept. of Electrical Computer Engineering, OSU, 2015 Neil Ave., Columbus, OH, USA
qin.324@osu.edu

Commission II, WG II/2

KEY WORDS: Cross-view images, Multi-view stereo (MVS) matching, 3D geo-registration, 3D meshing

ABSTRACT:

3D data generation often requires expensive data collection such as aerial photogrammetric or LiDAR flight. In cases such data are unavailable, for example, areas of interest inaccessible from aerial platforms, alternative sources to be considered can be quite heterogeneous and come in the form of different accuracy, resolution and views, which challenge the standard data processing workflows. Assuming only overview satellite and ground-level go-pro images are available, which we call cross-view data due to the significant view differences, this paper introduces a framework from our project, consisting of a few novel algorithms that convert such challenging dataset to 3D textured mesh models containing both top and façade features. The necessary methods include 3D point cloud generation from satellite overview images and ground-level images, geo-registration and meshing. We firstly introduce the problems and discuss the potential challenges and introduce our proposed methods to address these challenges. Finally, we practice our proposed framework on a dataset consisting of twelve satellite images and 150k video frames acquired through a vehicle-mounted Go-pro camera and demonstrate the reconstruction results. We have also compared our results with results generated from an intuitive processing pipeline that involves typical geo-registration and meshing methods.

1. INTRODUCTION

1.1 Introduction

City-scale data generation often requires expensive data collection such as aerial photogrammetric or LiDAR flight (Haala and Cavegn, 2016; Schwarz, 2010). Depending on the required accuracy, resolution, the efforts in data collection and processing can exponentially grow. Alternative and low-cost data sources are of particular interest for wide-area 3D modelling (Bosch et al., 2016): Satellite sensors running 24/7 offer overview images covering large regions for a single scan, which comparatively come with lower cost than aerial flights and do not require physical access to the area of interest (Qin, 2016). On the other hand, there exist a large number of street-view images coming either from crowdsourcing platforms or collected using relatively cheap equipment (e.g. video frames from low-cost cameras) that provides high-resolution information of object facades. Both the overview and the street-view data are complementary to each other and their view differences being approximately 90° forms cross-view dataset, a combined use of which may yield a low-cost solution for city-scale 3D modelling. This paper describes the attempt to address this challenging task by proposing an automated framework to convert the satellite overview and street-view video frames to complete 3D textured mesh models that contain both top and side view features.

The available commercial satellite images often have 0.3-0.5 meter GSD (ground sampling distance) and ground-level images taking from street-view easily reaches a GSD of a few millimeters. With significantly different resolution, the resulting 3D geometry may be associated with different uncertainties, which adds additional challenges to the mesh modelling task. In sum, to utilize the overview satellite images and street-view

*corresponding author

images for 3D mesh model reconstruction, major challenges include the following:

1) The quality of 3D output separately generated from satellite images and street-view images are scene-specific and may differ in terms of completeness and accuracy.

2) Due to the large view differences, the overview and street-view dataset may share very limited region in common, and additionally the 3D output from the street-view dataset may come with no geo-referencing information and may contain non-rigid topographic distortions (e.g. trajectories drift or distortions due to inaccurate interior/exterior orientation estimation), which further add challenges in 3D geo-registration of the dataset.

3) The combined 3D point clouds are from two sources with different resolution, uncertainty and radiometric properties of textures, obtaining visually consistent textured meshes can be extremely challenging.

We introduce in our proposed framework three major contributions to address the above-mentioned challenges, these being: 1) we introduce a monocular video-frame based 3D reconstruction pipeline to achieve the minimal geometric distortion by leveraging the speed and accuracy in a photogrammetric reconstruction pipeline; 2) we introduce a novel cross-view geo-registration algorithm that takes point clouds generated from satellite multi-view stereo (MVS) images and from street-view videos, to co-register the street-view point clouds to the overview point clouds; 3) we extend the existing mesh approaches to accommodate point clouds with images coming from different cameras. The rest of the paper is organized as follows: Section 2 introduces related works and considerations of cross-view data processing; Section 3 introduces our methodology and contributions to the processing pipeline; Section 4 describes the experiment dataset and the results of the 3D reconstruction Section 5 concludes this paper by discussing our planned future works.

2. RELATED WORKS AND CONSIDERATIONS

2.1 Related Works

The uses of multi-source 3D data have been attempted for different purposes, such as for localization, geo-registration, image synthesis and complete model generation (Gruen et al., 2013; Lin et al., 2015; Regmi and Borji, 2018). For example, (Gruen et al., 2013) utilized a combination of UAV (Unmanned Aerial Vehicles) images and mobile LiDAR (Light Detection and Ranging) for 3D model generation, where the geo-registrations are performed by manually measured ground control points (GCP) on the LiDAR data, followed by a Bundle Adjustment of the UAV images. All were performed following a surveying-grade process and thus minimal considerations were given to address topographical distortions.

Correlating the satellite overview and ground view images is extremely challenging, because the areas in common can sometimes be barely the ground and even less. There are two types of approaches that aim to address relevant tasks, such as 1) cross-view images localization (Lin et al., 2013; Lin et al., 2015; Tian et al., 2017) and 2) cross-view image synthesis (Regmi and Borji, 2018). Since the traditionally feature-based matching methods fail in cross-view data, the major technical approaches for cross-view data instead learn deep representations between cross-view data, with various strategies for learning scene-level descriptors used to match cross-view data, combing learned semantics and geometric transformation. A few earlier works also explored the use of manually crafted features for such a task (Castaldo et al., 2015; Lin et al., 2013). Most of the existing methods exploring 3D data co-registration, requires a certain portion of common regions and the transformation are often assumed to be simple models such as similarity or rigid transformations (Gruen and Akca, 2005; Rusinkiewicz and Levoy, 2001). Thus exploring methods for registering wide-area, cross-view dataset potentially with complex geometric distortions are particularly of interest.

Meshing point clouds seems to be standard practice with many applicable algorithms available. However, for image-based point clouds, meshing requires the use of the visibility information between the view and each point (Labatut et al., 2009; Tran and Davis, 2006) which sometimes are not easily available for multi-source data as first of all, they may share different camera model, and second of all, standard software packages generating point clouds from images do not offer such visibility information. As a result, a standard practice of using multi-source image-based point clouds only takes point-cloud based meshing methods (Kazhdan et al., 2006) which are designed for very dense point clouds and do not necessarily work well for point clouds with the level of uncertainty and complexity as the image-based point clouds.

2.2 General Considerations

Despite the aforementioned challenges, we consider the problem of turning the MVS satellite images and street-view Go-pro data to be approachable if scenario-specific information and intermediate results of the stereo reconstruction pipeline are accessible. To achieve, we have the following three considerations:

1) Street-view video frames taking alongside the street do not offer an optimal camera network, thus it is possible the results of the 3D reconstruction contains geometric distortion, for example, trajectory drifts, or topographic distortion due to the incorrectly estimated interior/exterior orientations, which will further add challenges to the geo-registration, we therefore consider to optimize our photogrammetric reconstruction workflow by considering self-calibration for each incremental reconstruction to minimize the potential trajectory drift.

2) We observed that in an urban environment, the boundary of objects from the satellite point clouds, e.g. buildings, might coincide well with the boundary produced by projecting the façade point clouds to the ground; therefore it can be seen as a view-invariant feature for co-registering the satellite point clouds and ground-view point clouds.

3) Meshing methods will unlikely to work well on the combined point clouds (from satellite and street-view point clouds) without the use of visibility information. Although theoretically possible, re-implementing a meshing algorithm considering different camera models can be painstakingly trivial. We consider the satellite point clouds to associated with an orthophoto under a parallel projection, thus the visibility can be easily computed and incorporated into an image-based meshing (Labatut et al., 2009) and texture mapping pipeline (Waechter et al., 2014).

2.3 The Proposed Data Generation Pipeline

To sum, our proposed data generation pipeline considers three major components. As shown in Figure 1, which includes separate 3D data generation (for MVS satellite images and ground-level video frames), geo-registration and meshing.
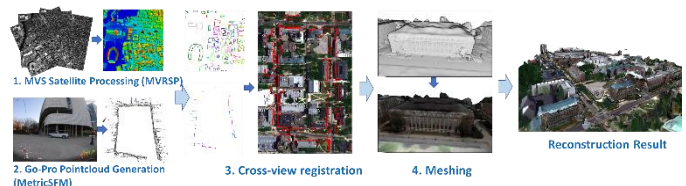

Figure 1. The general workflow of our processing pipeline.

where MVRSP (based on (Qin, 2017)) and MetricSFM are respectively our developed system for processing the satellite data and ground-level video frames. The geo-registration and meshing methods will be introduced in Section 3.

3. METHODLOGY

Following the above-mentioned pipeline (section 2.3, Figure 1), we briefly introduce our proposed methods in processing the cross-view data for 3D reconstruction.

3.1 Multi-view Stereo (MVS) Satellite Processing

The MVS satellite processing follows methods in (Qin, 2017; Qin, 2016), which takes a pair-wise reconstruction followed by a DSM (Digital Surface Model) fusion. The core matching algorithm uses a hierarchical Semi-Global Matching (Hirschmüller, 2008) with modifications to accommodate large-format images (Qin, 2014). We consider taking more than two images to obtain sufficient redundancies for 3D reconstruction. The satellite images are selected by following the approach of (Qin, 2019) based on the available images and their metadata from the digital globe (DG) (DigitalGlobe, 2020), and the images consist of both WorldviewI/II images (data will be introduced in Section 4). The readers may refer to specific details of the reconstruction in (Qin, 2017; Qin, 2019; Qin, 2016).

3.2 Point Cloud Generation from Go-Pro Video Frames

We take the standard structure-from-motion / photogrammetry reconstruction pipeline (Cernea, 2015), and implemented a few

strategies borrowed from the SLAM (Simultaneous Localization and Mapping) community (Mur-Artal and Tardós, 2017), which includes feature extraction and matching under the assumption of a continuous trajectory. To allow the optimal output accuracy, we performed bundle adjustment with self-calibration with every 10 images added into the incremental orientation, this leverages the speed and accuracy in terms of computing a very large number of images (in the scale of 150K images). We call our system MetricSFM. This system fully utilizes the trajectory consistency information in the video sequence to improve the speed and accuracy of structure from motion. More specifically, for feature detection, we use the ORB feature detector (Rublee et al., 2011) to extract key points considering its efficiency and performance. For feature matching, we take the velocity model presuming that neighbouring images travel with a constant speed, our method firstly matches each keypoint p_i on the current image i to two neighbouring images $i-1$ and the next image $i+1$, and generate the best matches p_{i-1} and p_{i+1} through minimal descriptor distances. The velocity model assumes the flows in the image space to be constant within a threshold (we take 20 pixels), and takes this as the constraint and the run the matches again to generate more matches to ensure sufficient observations especially for cases where repeated patterns present (walls/windows) for relative orientation and bundle adjustment. The dense matching were performed using standard matching pipeline in the open-source software OpenMVS (Cernea, 2015).

3.3 Geo-registration of the Overview and Street-view Point Clouds

Given point clouds generated from both the overview and the street-view, we take a three-step approach which computes the alignment on the building boundaries derived from both data sources, being: 1) Building Detection, 2) Individual Building Boundary Matching, and 3) Global Building Boundary Adjustment.

Building Detection: We extract both overview and street-view point clouds. The overview building boundaries are extracted using a well-developed morphological top-hat method (Qin and Fang, 2014; Vincent, 1993) applied on the DSM, with NDVI (Normalized Difference Vegetation Index) (Carlson and Ripley, 1997) for removing the trees from the binary masks. The street-view buildings are detected using a rather heuristic approach: we first separate the façade points from the ground points determined by normal vectors. And these points are further segmented using a region growing method(Tremeau and Borel, 1997) and those with very high projected density (to the ground) are determined as buildings.

Individual Building Boundary Matching: We use a heuristic registration algorithm that perform a targeted exhaustive search in the rotation space given a determined scale (either by GPS observations or given by a few known points). For each rotation hypothesis we compute the smallest, which are performed for each possible pair.

Global Building Boundary Adjustment: A local and pair-wise segment match is unlikely to provide a global solution and a "winner takes all" strategy may result in many outliers. Therefore, we consider a smooth constraint that penalizes those determined transformation parameters of neighbouring segments to be different, to achieve a more consistent set of transformation parameters for point clouds of a trajectory. The transformation parameters associated with each segment are discretized and thus the global adjustment can be performed as a labelling problem solved by a classic graph-cut formulations (shown in Equation (1)) (Boykov et al., 2001), in which the cost term $C(S_i, T_i)$ is

defined as for each segment, the registration error given a transformation hypothesis T, and the smoothness term $P_{S_i,S_j}(T_i, T_j)$ defines difference between transformation parameters (angular difference and translation difference). The goal is to find for each street-view point cloud segment S_i, the transformation parameters T_i out of a set of hypotheses $\mathcal{T}$, such that the energy defined in Equation (1) is minimized.

$$E(\mathcal{T}) = \sum_i C(S_i, T_i) + \sum_{i,j} P_{S_i,S_j}(T_i, T_j), \qquad (1)$$

Once the transformation parameters are determined, we use the transformation parameters to re-adjust the images to update the poses of the images for the purpose of 2D-3D geometric consistencies and the following texture mapping.

3.4. Meshing and Texture Mapping of the Cross-view Data

Meshing algorithms for image-based 3D dataset requires visibility information for surfaces and such information can be difficult to obtain for 3D point clouds generated from images with different views, sources, resolutions and uncertainties. Considering it is technically trivial, we therefore propose a meshing algorithm that regards the satellite point clouds to be associated with orthophoto under a parallel projection. With this assumption, we have modified the existing image-based point cloud meshing (Labatut et al., 2009) and texture mapping method (Waechter et al., 2014): 1) we extended state-of-the-art image-based surface reconstruction method by incorporating geometric information produced by satellite images to create wide-area surface model. 2) We extended a texture mapping method to accommodate images acquired from different sensors, i.e. side-view perspective images and satellite images.

3.4.1. Meshing

The base method (Labatut, et al. 2009) takes the constructed Delaunay tetrahedrons from the point clouds as the input to determine the surface. These tetrahedrons can be viewed as a connected graph, in which the tetrahedrons are the notes and shared/common faces are edges. Our method extends from this base algorithm by incorporating point clouds generated from the satellite images.

Our meshing pipeline builds meshes generated from street-view images and satellite point clouds consists of three steps:

1. We form Delaunay tetrahedrons using the combined point cloud set from the satellite and street-view based point clouds.
2. We take the visibility information from the MetricSFM pipeline and build visibility information using the parallel projection by associating the satellite point clouds with orthophoto views.
3. Solve minimum *s-t* (source-sink, acyclic) cut for labelling problem and extract surface following the method of (Labatut et al., 2009)

Figure 2 shows the surface model built from street-view only data with base method (Labatut et al., 2009) and the surface model built from both street-view and satellite images with ours, which shows that although the satellite point cloud based meshes present relatively coarse information on the roofs, it completes the street-view based meshes which are visually more informative.
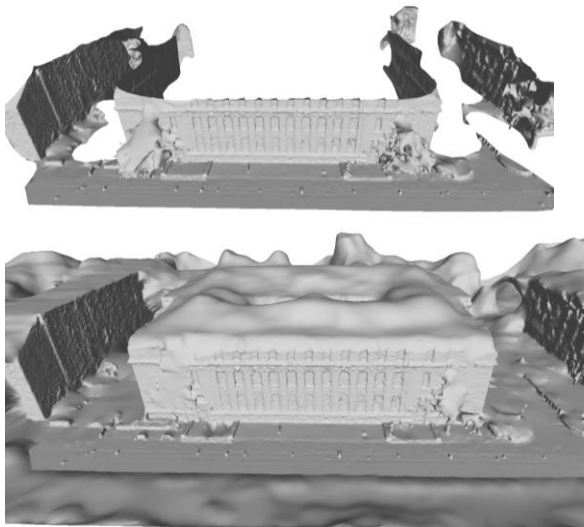
Figure 2. Reconstructed mesh models. Top: with only street-view images; bottom, with the combined satellite and street-view data.

3.4.2. Texture Mapping using overview and street-view images

Our texture mapping framework is based on Waechter's work (Waechter et al., 2014), which has been well practiced and widely used in many famous open source projects, e.g. TexRecon (Waechter et al., 2014) , OpenMVS (Cernea, 2015), etc. We consider that the street-view images are perspective and the satellite orthophoto is in parallel projection. Given these many images serving as potential source of texture, the first step is to pick a best image for each triangle face: by rendering faces onto images applying perspective and parallel projection respectively, in the meantime using depth buffer to determine the nearest faces, every visible view is assigned a weight to be associated with a face. Then the best view selection problem can be solved using a belief propagation algorithm.

Seamless texture fusion is also taken into account in our method, follows the base algorithm (Waechter, et al. 2014), our pipeline adjusts the color from different sources and fuses the seams between patches by using color balance and Poisson blending. Figure. 3 shows an example result of our multi-source texture mapping method on the produced mesh surface using our mesh reconstruction pipeline. The readers may refer more details of the method in (Song and Qin, 2020)

Figure 3. An example of the textured meshes reconstructed from our propose pipeline using both satellite and street-view data.

4. RESULTS

4.1. Data Description

We take the Ohio State University (OSU) Columbus Campus as our test site, of which we have collected twelve overlapping satellite images consisting of WorldView-I and WorldView-II images. These images selectively form 31 pairs used for the reconstruction, and many of these images are not from the same year thus creating challenges for the reconstruction. Table 1. provides an overview of the first 10 pairs used from the acquired images: not all of these pairs forms in-track stereo, while the large redundancy does provide the advantage in producing more accurate surface model. Figure 4. shows the generated digital surface model. The achieved RMSE (root-mean-squared-error) is 1.26 meters evaluated through LiDAR point clouds, and the RMSE reached 0.60 meters by excluding changed buildings, rivers and trees.

	Intersection angle (degree)	Sun difference angle (degree)	Time difference (days)
1	7.54	0.03	0
2	28.53	0.08	0
3	26.08	19.17	93
4	10.46	4.98	378
5	11.82	9.92	619
6	17.92	5.01	378
7	17.88	6.36	304
8	9.65	21.80	3024
9	8.66	21.78	3024
10	11.70	26.16	1350

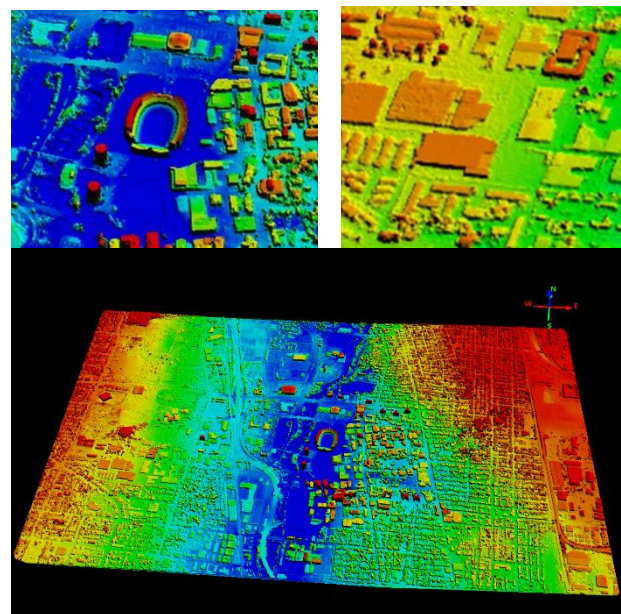Table 1. Examples of Metadata of pairs used for satellite-based 3D reconstruction.

Figure 4. The generated Digital Surface Models of the OSU campus using our satellite data processing pipeline. Top-view shows enlarged views.

We have also collected approximately 300 GB of Go-pro videos covering a trajectory equivalent to 33 km, and the reconstruction for the street-view images take 150k frames (with a resolution of 1500×2000 pixels per frame) out of these videos. Figure 5. shows the reconstructed point clouds of approximately two thirds of the region. The pose estimation time takes approximately 20 hours and dense matching takes 4hrs in a normal i-7 desktop computer.

Figure 5. Dense reconstruction using our MetricSFM for two thirds of the campus region, totalling 7 billion color points.

4.2 Experiment Results

We demonstrate that the resulting geometry shows completeness in terms of the rooftop and façade information (for places where street-view images are available). Figure 6 provides an overview of the registered point clouds and a comparison showing the mis-registration using a typical point cloud based algorithm (Rusinkiewicz and Levoy, 2001). Our co-registration achieves an RMSE of 1.44 m in error, which are reasonable considering that the satellite point clouds have a resolution of 0.5 m.

(a) ICP　　　　　　　　(b) ours
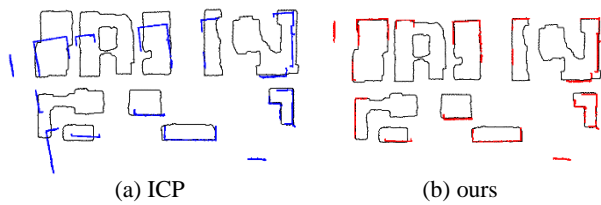
(c) Co-registered point clouds, point size of the satellite point clouds and street-view point clouds are appropriately adjusted to optimize the visualization in this figure.

Figure 6. Registration result of ICP (a) and our method (b) on the distorted street-view trajectory. (c) shows part of the registered street-view point clouds generated on 150k Go-Pro images.
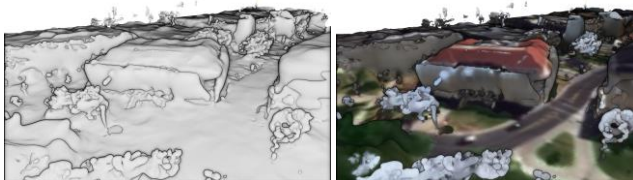
(a) Reconstructed mesh using Poisson reconstruction

(b) Reconstructed mesh using our reconstruction method
Figure 7. Left: shaded mesh model. Right: textured mesh model.

With the registered point clouds, we are able to generate the meshes using our proposed meshing pipeline introduced in section 3.4. Figure 7. shows the reconstructed meshes (shaded and textured) using our pipeline, and we have also included the results from a pure point cloud based meshing method, which visually demonstrates much worse results. In Figure 8, we have also included the reconstruction results of a relatively larger region using our reconstructed pipeline.

Figure 8. A screenshot of the generated textured mesh of the OSU campus area using our proposed pipeline, which includes information from the top-view and details on the facades.

5. CONCLUSIONS

In this paper, we report the results of our work that aims to perform 3D reconstruction from overview satellite and ground view images (called cross-view dataset). We present our processing framework (Figure 1.) that consists of three major components: 1) 3D reconstruction separately from the top-view satellite images and ground-level images; 2) Cross-view geo-registration between the satellite point clouds and street-view point clouds; 3) Meshing reconstruction based on the combined satellite and ground point clouds. In each of these components, we present our developed systems and on-going research efforts in addressing the potential challenges (introduced in Section 1.1), and the in-progress results. We demonstrate that our proposed pipeline is able to achieve visually more consistent textured meshes, in comparison to a standard and intuitive processing method. The proposed framework and the attempts for integrating satellite and street-view images and converting them to textured models can be of particular interest for data collection in areas where standard datasets such as aerial/UAV (unmanned aerial vehicle) photogrammetric/LiDAR flights. This work is ongoing and the current geo-registration procedure is rather computationally heavy, our future works include a focus on the registration algorithms and further optimizing individual modules of our processing pipeline and part of these modules will be made available once they are optimized for practical uses.

ACKNOWLEDGEMENT

REFERENCES

Bosch, M., Z. Kurtz, S. Hagstrom and M. Brown, 2016. A Multiple View Stereo Benchmark for Satellite Imagery. *Proceedings of the IEEE Applied Imagery Pattern Recognition (AIPR) Workshop, October 2016*,

Boykov, Y., O. Veksler and R. Zabih, 2001. Fast approximate energy minimization via graph cuts. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 23 (11), 1222-1239.

Carlson, T. N. and D. A. Ripley, 1997. On the relation between NDVI, fractional vegetation cover, and leaf area index. *Remote Sensing of Environment* 62 (3), 241-252.

Castaldo, F., A. Zamir, R. Angst, F. Palmieri and S. Savarese, 2015. Semantic cross-view matching. In: *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pp. 9-17.

Cernea, D., 2015. OpenMVS: Open Multiple View Stereovision. DigitalGlobe, 2020. Digital Globe Image Finder, https://discover.digitalglobe.com/. Accessed 2020.02.01

Gruen, A. and D. Akca, 2005. Least squares 3D surface and curve matching. *ISPRS Journal of Photogrammetry and Remote Sensing* 59 (3), 151-174.

Gruen, A., X. Huang, R. Qin, T. Du, W. Fang, J. Boavida and A. Oliveira, 2013. Joint Processing of Uav Imagery and Terrestrial Mobile Mapping System Data for Very High Resolution City Modeling. *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 1 (2), 175-182.

Haala, N. and S. Cavegn, 2016. HIGH DENSITY AERIAL IMAGE MATCHING: STATE-OF-THE-ART AND FUTURE PROSPECTS. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences* 41

Hirschmüller, H., 2008. Stereo processing by semiglobal matching and mutual information. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30 (2), 328-341.

Kazhdan, M., M. Bolitho and H. Hoppe, 2006. Poisson surface reconstruction. In: *Proceedings of the fourth Eurographics symposium on Geometry processing*, pp.

Labatut, P., J. P. Pons and R. Keriven, 2009. Robust and efficient surface reconstruction from range data. In: *Computer graphics forum*, pp. 2275-2290.

Lin, T.-Y., S. Belongie and J. Hays, 2013. Cross-view image geolocalization. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 891-898.

Lin, T.-Y., Y. Cui, S. Belongie and J. Hays, 2015. Learning deep representations for ground-to-aerial geolocalization. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 5007-5015.

Mur-Artal, R. and J. D. Tardós, 2017. Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras. *IEEE transactions on robotics* 33 (5), 1255-1262.
Qin, R., 2017. Automated 3D recovery from very high resolution multi-view satellite images. In: *ASPRS (IGTF) annual Conference*, March 12-16, Baltimore, Maryland, USA, pp. 10.

Qin, R., 201. Change detection on LOD 2 building models with very high rsolution spaceborne stereo imagery. *ISPRS Journal of Photogrammetry and Remote Sensing* 96 (2014), 179-192.

Qin, R., 2019. A critical analysis of satellite stereo pairs for digital surface model generation and a matching quality prediction model. *ISPRS Journal of Photogrammetry and Remote Sensing* 154 139-150.

Qin, R., 2016. RPC Stereo Processor (RSP) –A software package for digital surface model and orthophoto generation from satellite stereo imagery. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* III (1), 77-82.

Qin, R. and W. Fang, 2014. A hierarchical building detection method for very high resolution remotely sensed images combined with DSM using graph cut optimization. *Photogrammetry Engineering and Remote Sensing* 80 (8), 37-48.

Regmi, K. and A. Borji, 2018. Cross-view image synthesis using conditional gans. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3501-3510.

Rublee, E., V. Rabaud, K. Konolige and G. Bradski, 2011. ORB: An efficient alternative to SIFT or SURF. In: *2011 International conference on computer vision*, pp. 2564-2571.

Rusinkiewicz, S. and M. Levoy, 2001. Efficient variants of the ICP algorithm. In: *3-D Digital Imaging and Modeling, 2001. Proceedings. Third International Conference on*, pp. 145-152.
Schwarz, B., 2010. LIDAR: Mapping the world in 3D. *Nature Photonics* 4 (7), 429.

Song, S. and R. Qin, 2020. Optimizing mesh reconstruction and texture mapping generated from a combined side-view and over-view imagery. *ISPRS Congress 2020 (this conference). ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*,

Tian, Y., C. Chen and M. Shah, 2017. Cross-view image matching for geo-localization in urban environments. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3608-3616.

Tran, S. and L. Davis, 2006. 3D surface reconstruction using graph cuts with surface constraints. In: *European conference on computer vision*, pp. 219-231.

Tremeau, A. and N. Borel, 1997. A region growing and merging algorithm to color segmentation. *Pattern recognition* 30 (7), 1191-1203.

Vincent, L., 1993. Morphological grayscale reconstruction in image analysis: applications and efficient algorithms. *IEEE Transactions on Image Processing* 2 (2), 176-201.

Waechter, M., N. Moehrle and M. Goesele, 2014. Let there be color! Large-scale texturing of 3D reconstructions. In: *European conference on computer vision*, pp. 836-850.

Waechter, M., N. Moehrle and M. Goesele, 2014. TexRecon – 3D Reconstruction Texturing, https://www.gcc.tu-darmstadt.de/home/proj/texrecon/index.en.jsp. Accessed 2020.02.01