Chapter 23 Photogrammetry for 3D Mapping in Urban Areas



Bo Wu

Abstract Photogrammetry is the technology for obtaining 3D geometric information from photographs or images. This chapter describes the fundamental knowledge and latest advances in photogrammetry for 3D mapping in urban areas. First, the key fundamental techniques in photogrammetry for deriving 3D information from imagery are presented. Then, the latest advances in photogrammetry for 3D mapping in urban areas, including structure-from-motion (SfM), multi-view stereo (MVS), and integrated 3D mapping from multiple-source data, are described and discussed. Examples of using photogrammetry for 3D mapping and modeling in urban applications are presented. Finally, concluding remarks and future outlooks are addressed.

23.1 Introduction

Photogrammetry is the science and technology for obtaining reliable 3D geometric and physical information about objects and the environment from photographic images (ASPRS 1998). Practically, photogrammetry allows 3D measurements of geometric information of objects (e.g., positions, orientations, shapes, and sizes) from photographs.

Photogrammetry has a long history and can be dated back to the 1850s (Konecny 1985). In its earlier stage, the main purpose of photogrammetry was map generation from aerial photographs. Since the 1960s, the emerging of satellite and close-range imaging and measurements has facilitated the application of photogrammetry to various areas, such as 3D mapping and modeling, industrial inspection, architecture, robotics, civil engineering, and hazard monitoring. Advances in photogrammetry had been insignificant over the past 50 years until the recent decade. The latest advances from the photogrammetry and computer vision communities, such as aerial oblique photogrammetry, structure-from-motion (SfM) and multi-view stereo (MVS), and

B. Wu (🖂)

Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong, China e-mail: bo.wu@polyu.edu.hk

integrated 3D mapping, have facilitated the development of photogrammetry towards a more automatic solution for 3D mapping and modeling, with better quality, even for challenging cases such as in urban areas.

This chapter first describes the key fundamental knowledge for obtaining 3D information from images through photogrammetry. Then, the latest advances in photogrammetry for 3D mapping in urban areas, including SfM, MVS, and integrated 3D mapping from multiple-source data, are described and discussed. Examples of using photogrammetry for 3D mapping and modeling in Hong Kong and other typical urban areas are presented. Finally, summary remarks are given and future outlooks are discussed.

23.2 Fundamentals of Photogrammetry

The following describes the fundamental techniques for obtaining 3D information from images via photogrammetry, including image orientation, bundle adjustment, and image matching.

23.2.1 Image Orientation

Image orientation is the procedure of recovering the positional and orientation information of the optical ray when the image is collected. Image orientation includes two consecutive steps: interior orientation (IO) and exterior orientation (EO).

IO defines the transformation from the pixel coordinates measured on the image to the image-space coordinates referring to the focal plane. Taking a traditional aerial image as an example, typically, there are four to eight fiducial marks distributed in the corners and along the edges of the image. Their pixel coordinates can be directly measured on the image. Also, the coordinates of these fiducial marks in the image-space coordinate system are usually known. They can be used to determine the principal point (x_0 , y_0) in the image-space coordinate system. They can also be used to derive a 2D transformation model between the image-space coordinates and the image measurements, and then the 2D transformation model can be used to transform any other pixel coordinates measured on the image to the image-space coordinates.

The coordinates of the principal point (x_0, y_0) and the principal distance (or focal length) f are the intrinsic parameters of the camera. The camera intrinsic parameters normally do not change. However, there are usually distortions existing on images, such as lens distortions, different pixel spacing, and stretching or shrinkage of the images. They have to be calibrated before using the images for 3D mapping. Errors in these parameters will lead to errors in the IO process and the subsequent 3D measurement. These parameters and distortions can be calibrated using a particular control field with calibration targets precisely measured by a total station or

differential GPS. They can also be computed during the 3D mapping task through self-calibration approaches (Wu 2017).

EO defines the transformation from the image-space coordinates to the 3D object space coordinates, which can be formulated using the following co-linearity equations (Wang 1998):

$$x - x_0 = -f \frac{m_{11}(X - X_S) + m_{12}(Y - Y_S) + m_{13}(Z - Z_S)}{m_{31}(X - X_S) + m_{32}(Y - Y_S) + m_{33}(Z - Z_S)}$$

$$y - y_0 = -f \frac{m_{21}(X - X_S) + m_{22}(Y - Y_S) + m_{23}(Z - Z_S)}{m_{31}(X - X_S) + m_{32}(Y - Y_S) + m_{33}(Z - Z_S)}$$
(23.1)

The co-linearity equations connect a point (x, y) on the image and its corresponding position (X, Y, Z) in the 3D object space. (X_S, Y_S, Z_S) represent the coordinates of the camera perspective center in the object space when the image is taken. m_{ij} are the components of a rotation matrix, which is derived from three rotation angles $(\varphi, \omega, \kappa)$ of the camera frame referring to the object space. These six parameters—three positions (X_S, Y_S, Z_S) and three rotation angles $(\varphi, \omega, \kappa)$ —are called EO parameters.

Each set of co-linearity equations represents a straight line that links an image point, the camera perspective center, and a 3D point in the object space. To determine the object point's 3D position, at least two straight lines are necessary to form an intersection. In other words, a pair of corresponding points measured on a stereo pair of images will be necessary to compute their corresponding 3D position in the object space. This process is called space intersection.

The EO parameters of each image can be measured by sensors (e.g., GPS and IMU) mounted on the same platform as the camera when it takes the image so that 3D measurements can be achieved by using at least two images together with their EO parameters. However, direct measurement of the EO parameters by the sensors will usually have errors and sometimes no direct measurement of the EO parameters will be provided. Therefore, in photogrammetry, the EO parameters are usually derived or improved in one of three ways: space resection, relative orientation (RO) followed by absolute orientation (AO), or simultaneous orientation through bundle adjustment.

Space resection is based on the above co-linearity equations. If three control points (their coordinates in the image-space and object space are known) are available, they offer six observations based on the co-linearity equations and provide a unique solution to the six EO parameters. Normally, more control points are used to calculate the EO parameters through the least-squares adjustment for improved accuracy. Usually, space resection is used to determine the EO parameters of a single image. For an image block, other methods are used as they require fewer control points.

RO is used to determine the internal relationship between two images. RO is able to generate a scale-free 3D model of the imaged scene within an arbitrary coordinate system. Before the 3D model obtained from RO can be used for actual measurement, it must be scaled, rotated, and translated to the actual coordinate system in object space. This is the procedure of AO. AO uses 3D transformations (e.g., 3D conformal transformation) to convert the model coordinates obtained by RO into real object coordinates. The RO and AO can be performed on a single stereo pair or on large image blocks.

23.2.2 Bundle Adjustment

Bundle adjustment (BA) is an alternative method to the above RO and AO procedures. Based on the principles of the co-linearity equations, an optical ray can be defined that starts from the image point, passes through the perspective center of the camera, and finally reaches the 3D point in the object space. This produces an observation based on the co-linearity equations. Giving some tie points matched on a stereo pair of images or multiple images, a bundle of optical rays determined by the tie points can link the images together, and subsequently link the image-space to the object space. In the ideal situation, the optical rays from the tie points on different images should exactly intersect at the same object point. However, this will usually not be true in the reality due to uncertainties and errors of different levels in the image orientation parameters. Therefore, BA is used to improve the image orientation parameters, from which the bundle of optical rays can intersect at the 3D point in the object space correctly.

BA is based on the least-squares principle. Usually, four types of observation equations can be formulated in a BA system, as listed in the following.

$$Av + B\Delta = f$$

$$v_x - I\Delta = f_x$$

$$A_c v_c + C\Delta_c = f_c$$

$$A_{ap} v_{ap} + D\Delta_{ap} = f_{ap}$$
(23.2)

The first observation equation is for the image measurements (tie points matched on the images), which is based on the co-linearity equations that connect the image measurements with their 3D coordinates. Δ is the vector of the unknown EO parameters. *A* is the matrix of observation coefficients. *B* is the matrix of parameter coefficients. *v* is the vector of residuals. The second observation equation is for the unknown EO parameters and the 3D object coordinates of the tie points to be calculated. The third observation equation is for constraints of the parameters. For instance, a stereo camera system with a fixed camera base can provide a constraint that the distance between the three positional EO parameters of the left image and those of the right image should equal to the length of the camera base. The fourth observation equation is for self-calibration, of which the additional parameters (e.g., principal distance, lens distortions) can be solved simultaneously in the BA system.

Based on the observation equations and provided with a small number of 3D control points and a large number of tie points matched on the images, BA is able to compute the unknown parameters and the 3D object coordinates of tie points simultaneously. BA is actually the simultaneous process of space resection and intersection as described previously. In the BA system, different weights can be assigned to

different types of observations based on their *a priori* precision or practical analysis, so that the contributions of different observations can be controlled. For example, observations with higher precision (less uncertainty) will be assigned with higher weights, so that they will contribute more and be adjusted less in the BA system. Observations with less knowledge (large uncertainties) will be assigned with lower weights so that they will contribute less and be adjusted more. BA is fully rigorous through corrections for systematic errors and provides abundant statistical information. The residuals of all parameters can be calculated and they can be used to evaluate the performance of BA.

23.2.3 Image Matching

Image matching is for identifying image correspondences in two or more images with overlapping coverages. The corresponding points on images represent the same point in the object space. They usually have similar appearances on different images. Generally, image matching is based on finding the similarities in grey levels of small local patches on images or matching an image patch with an image template. Image matching may be implemented on a pixel-by-pixel basis, known as dense matching, or by matching individual point or pattern features, which is called feature matching.

In the photogrammetry and computer vision communities, much research has been done regarding image matching. A straightforward image matching method is the normalized cross-correlation (NCC) matching (Lhuillier and Quan 2002). NCC directly examines the level of similarity between two small image patches or local windows by calculating their cross-correlation score in terms of the grey levels. A significant development about feature point matching is the scale-invariant feature transform (SIFT) method (Lowe 2004) in the computer vision community. SIFT first detects feature points based on the local extrema in the scale space that are invariant to scale changes and distortions, and then matches the feature points according to the descriptors constructed based on their gradients in local regions. However, SIFT only provides sparse feature matching results. Semiglobal matching (SGM; Hirschmuller 2008) is another important development in dense image matching. SGM combines global and local methods for pixel-wise matching through optimization of an energy function. SGM is able to produce dense matching results; however, the global optimization strategy used in SGM may lead to an over-smoothing problem in 3D surface reconstruction.

Wu et al. (2011, 2012) presented a hierarchical image matching method, named self-adaptive triangulation-constrained matching (SATM). SATM includes a feature matching step followed by a dense matching step. It uses triangulations to constrain the matching of feature points and edges, of which the triangulations are dynamically updated along with the matching process by inserting the newly matched points and edges into the triangulations. Dense matching is conducted during the densification of the triangulations. In the matching propagation process, the most distinctive features are always successfully matched first; therefore, the densification of triangulations

self-adapts to the textural pattern on the image, and provides robust constraints for reliable feature matching and dense matching. Ye and Wu (2018) further extended the SATM algorithm by incorporating image segmentation into the image matching framework to solve the surface discontinuity problem for dense and reliable matching of images in urban areas. Figure 23.1 shows an example of the matching results using SATM and SGM for a stereo pair of aerial images for generating a digital surface model (DSM) in an urban area. As can be seen from the DSMs generated by SATM (Fig. 23.1b) and SGM (Fig. 23.1c), the former performs better than the latter in terms of feature preservation and recovery of building boundaries.



(a) A pair of aerial images with the matched results using SATM marked in red



(b) The generated DSM from SATM

(c) The generated DSM from SGM

Fig. 23.1 An example of the image matching algorithms SATM and SGM for DSM generation in urban areas

23.3 Advances in Photogrammetry for 3D Mapping in Urban Areas

Traditional photogrammetry has limited use for 3D mapping and modeling in urban areas (Oiao et al. 2010; Ye and Wu 2018). This is mainly due to the fact that traditional photogrammetry usually captures near-nadir images by cameras mounted on aircraft, and image matching in urban areas is particularly challenging. Most traditional photogrammetry systems require tremendous human labor to process images in urban areas, especially in metropolitan regions with tall buildings that are densely located. With the development of hardware and software in data acquisition and image processing in recent years, the image quality, automation degree, efficiency, and accuracy of photogrammetry have been boosted extensively in the past decade (Rupnik et al. 2015). The state-of-the-art oblique photogrammetry systems collect aerial oblique images in urban areas with high redundancy (e.g., with every ground point visible in over five or more images), which significantly improves the automatic image matching in urban areas and also provides information on building facades. Off-the-shelf solutions for 3D city modeling from aerial oblique images include two key steps: structure from motion (SfM) (Gerke et al. 2016) and multi-view stereo (MVS) (Galliani et al. 2015).

23.3.1 Structure from Motion and Multi-view Stereo

In the SfM method, feature points are used to obtain tie points between overlapped views of images automatically. For structured aerial images that are captured with designed flight plans, the connectivity between different images could be estimated accordingly. However, if the images are unordered, trying out all the possible image pairs is exhaustive for large datasets. Hence, image retrieval algorithms based on vocabulary trees (Gálvez-López and Tardos 2012) are used to find the putative image pairs that are similar and may have overlaps. After that, the initial orientation parameters are estimated and then refined by BA. BA approaches are typically divided into three categories in SfM, namely sequential, hierarchical, and global adjustment (Schonberger and Frahm 2016). Sequential adjustment methods start from a minimal image cluster (such as two or three well-connected images) and incrementally add new images to the existing clusters. The computation cost of this approach increases with each increment in reconstruction. Hence, a divide-and-conquer strategy can be adopted to reduce computation cost, which performs the BA hierarchically (Snavely et al. 2008). The scene graph is divided into several clusters first, and then these clusters are reconstructed individually. After that, these clusters are merged by a transformation with 7 degrees of freedom (DoF). Global methods normally estimate relative orientations of all the images at the same time, and estimate global rotation and translation separately (Toldo et al. 2015). However, it might be difficult for global

optimization algorithms to achieve convergence, requiring good initial estimations and robust outlier detection and removal.

The resulting image orientation parameters and the scene graph of SfM serve as the foundation for the MVS (Schonberger and Frahm 2016). However, the sparse point clouds obtained by BA do not contain any solid geometry about the scene. Hence, MVS algorithms are employed to turn oriented 2D images into dense 3D point clouds using multiple images (Musialski et al. 2013). An example of the widely adopted MVS algorithm in the photogrammetry community is the patchbased multi-view stereo (PMVS) invented by Furukawa and Ponce (2010). In this method, corresponding points in multiple images are used to construct an initial set of patches to represent the scene, and the patches are repeatedly expanded to improve their density through enforcing photometric consistency and global visibility constraints to improve reconstruction accuracy. Based on the oriented images and the corresponding dense point clouds, a 3D mesh model of the surface can be reconstructed and textured using algorithms such as the Poisson reconstruction algorithm (Waechter et al. 2014), which produces watertight surfaces from oriented point clouds. Figure 23.2 is an example of automatically generated 3D models in Central Hong Kong using aerial oblique images based on SfM and MVS.

23.3.2 Integrated 3D Mapping from Multiple-Source Data

Apart from the above advances in oblique photogrammetry, there is a trend of integrating multiple-source images and laser-scanning data collected from different remote sensing platforms—for example, satellite, aircraft, unmanned aerial vehicle (UAV), and mobile mapping systems (MMS)—for better 3D mapping and modeling in urban areas (Wu et al. 2015, 2018).

Images and laser-scanning point clouds collected by different types of remote sensing platforms are widely used for 3D mapping and modeling. However, the 3D mapping results derived from different sensors and platforms usually show inconsistencies in the same area. Wu et al. (2015) presented an integrated 3D mapping model for the integrated processing of satellite imagery and airborne LiDAR data. In this model, the EO parameters of images, tie points matched in the overlapping images, and selected LiDAR points are used as inputs for a combined adjustment, and local constraints, including a vertical constraint and a horizontal constraint, are applied to ensure the consistencies between these two types of data are reduced and the geometric accuracies of the mapping results are improved.

The integrated 3D mapping model was further extended for integrated processing of images and laser scanning point clouds collected from UAV and MMS platforms (Wu et al. 2018). Aerial oblique photogrammetry offers promising solutions for 3D mapping and modeling in urban areas. However, in metropolitan areas such as Hong Kong, where high-rise buildings are densely distributed, there are usually



(a) Aerial oblique images collected in Central, Hong Kong



(b) Automatically generated 3D models from the aerial oblique images

Fig. 23.2 SfM and MVS for automatic 3D modeling from aerial oblique images

geometric defects in the 3D models generated from aerial oblique imagery, and the textures on building façades are usually blurred. These problems are related to the common occlusion situations and large camera tilt angles of aerial oblique imagery. Meanwhile, MMS can collect ground images and laser scanning point clouds on the ground, which provides a dataset complementary to the aerial data. The integrated processing of images and laser scanning data collected from UAV and MMS platforms offers promising opportunities to optimize 3D modeling in urban areas. The integrated 3D mapping of aerial and ground datasets includes three main steps: (1) automatic feature matching between the aerial and ground data to remove

their geometric inconsistencies; and (3) optimal selection of aerial and ground data for the best textural quality and minimum occlusions. Figure 23.3 shows an example of the integrated 3D mapping from UAV and MMS images collected in Kowloon Bay, Hong Kong. Figure 23.3 indicates that the integration of aerial and ground data



(a) 3D models from UAV images



(b) 3D models from integrated processing of UAV and MMS images

Fig. 23.3 Integrated 3D mapping of UAV and MMS images in Kowloon Bay, Hong Kong

shows a promising solution for generating 3D city models of the best geometry and quality. With the MMS data, the geometry and quality of the 3D mesh models at the street level are significantly improved, compared with those from aerial images only.

23.4 Summary

Photogrammetry is the most robust, efficient, economical, and flexible method for 3D mapping and modeling, regardless of the challenges ahead. Photogrammetry has been and will continue to be the representative and influential technology for obtaining 3D information. The latest advances in photogrammetry such as SfM, MVS, and integrated 3D mapping, offer great potential for optimized and enhanced 3D mapping and modeling in urban areas at both city scale and street level. Photogrammetry can be used as the primary technology to create the 3D spatial-data infrastructure for a digital city, which can be widely used to support applications in, for example, urban planning and design, urban management, urban environmental studies, and the development of smart cities.

References

- ASPRS (1998) Definitions of photogrammetry, remote sensing, and GIS. American Society for Photogrammetry and Remote Sensing. https://www.asprs.org/a/society/about.html. Accessed 20 Oct 2019
- Furukawa Y, Ponce J (2010) Accurate, dense, and robust multiview stereopsis. IEEE Trans Pattern Anal Mach Intell 32(8):1362–1376
- Galliani S, Lasinger K, Schindler K (2015) Massively parallel multiview stereopsis by surface normal diffusion. In: Proceedings of the IEEE international conference on computer vision, pp 873–881
- Gálvez-López D, Tardos JD (2012) Bags of binary words for fast place recognition in image sequences. IEEE Trans Rob 28(5):1188–1197
- Gerke M, Nex F, Remondino F, Jacobsen K, Kremer J, Karel W, Huf H, Ostrowski W (2016) Orientation of oblique airborne image sets: experiences from the ISPRS/EUROSDR benchmark on multi-platform photogrammetry. Int Arch Photogramm Remote Sens Spat Inf Sci 41:185–191
- Hirschmuller H (2008) Stereo processing by semiglobal matching and mutual information. IEEE Trans Pattern Anal Mach Intell 30(2):328–341
- Konecny G (1985) The international society for photogrammetry and remote sensing—75 years old, or 75 years young. Photogramm Eng Remote Sens 51(7):919–933
- Lhuillier M, Quan L (2002) Match propagation for image-based modeling and rendering. IEEE Trans Pattern Anal Mach Intell 24(8):1140–1146
- Lowe DG (2004) Distinctive image features from scale-invariant key points. Int J Comput Vis 60(2):91–110

- Musialski P, Wonka P, Aliaga DG, Wimmer M, Gool LV, Purgathofer W (2013) A survey of urban reconstruction. In: Proceedings of computer graphics forum. Wiley online library, pp 146–177
- Qiao G, Wang W, Wu B, Liu C, Li R (2010) Assessment of geo-positioning capability of high-resolution satellite imagery for densely populated high buildings in metropolitan areas. Photogramm Eng Remote Sens 76(8):923–934
- Rupnik E, Nex F, Toschi I, Remondino F (2015) Aerial multi-camera systems: accuracy and block triangulation issues. ISPRS J Photogramm Remote Sens 101:233–246
- Schonberger JL, Frahm J (2016) Structure-from-motion revisited. In: Proceedings of 2016 IEEE conference on computer vision and pattern recognition (CVPR), Las Vegas, 26 June–1 July 2016
- Snavely N, Seitz SM, Szeliski R (2008) Skeletal graphs for efficient structure from motion. In: Proceedings of 2008 IEEE conference on computer vision and pattern recognition (CVPR), Anchorage, 24–26 June 2008
- Toldo R, Gherardi R, Farenzena M, Fusiello A (2015) Hierarchical structure-and-motion recovery from uncalibrated images. Comput Vis Image Underst 140:127–143
- Waechter M, Moehrle N, Goesele M (2014) Let there be color! Large-scale texturing of 3D reconstructions. In: Proceedings of 2014 European conference on computer vision, pp 836–850
- Wang Z (1998) The research program of the full digitized and automatic mapping system. J Wuhan Tech Univ Surv Map 23(4)
- Wu B (2017) Photogrammetry—3D from imagery. In: Richardson D, Castree N, Goodchild MF, Kobayashi A, Liu W, Marston RA (eds) The international encyclopedia of geography. Wiley, New York, pp 1–13. https://doi.org/10.1002/9781118786352.wbieg0942
- Wu B, Zhang Y, Zhu Q (2011) A triangulation-based hierarchical image matching method for wide-baseline images. Photogramm Eng Remote Sens 77(7):695–708
- Wu B, Zhang Y, Zhu Q (2012) Integrated point and edge matching on poor textural images constrained by self-adaptive triangulations. ISPRS J Photogramm Remote Sens 68:40–55. https:// doi.org/10.1016/j.isprsjprs.2011.12.005
- Wu B, Tang S, Zhu Q, Tong K, Hu H, Li G (2015) Geometric integration of high-resolution satellite imagery and airborne LiDAR data for improved geopositioning accuracy in metropolitan areas. ISPRS J Photogramm Remote Sens 109:139–151
- Wu B, Xie L, Hu H, Zhu Q, Yau E (2018) Integration of aerial oblique imagery and terrestrial imagery for optimized 3D modeling in urban areas. ISPRS J Photogramm Remote Sens 139:119–132
- Ye L, Wu B (2018) Integrated image matching and segmentation for 3D surface reconstruction in urban areas. Photogramm Eng Remote Sens 84(3):35–48



Bo Wu is with the Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University. His research interests are mainly in the areas of Photogrammetry and Planetary Remote Sensing. He has worked on 3D city modeling projects, and the Mars and Moon exploration missions funded by NASA and China. **Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

