ISPRS Journal of Photogrammetry and Remote Sensing 139 (2018) 119–132

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Integration of aerial oblique imagery and terrestrial imagery for optimized 3D modeling in urban areas

Bo Wu [a,*], Linfu Xie [a,b], Han Hu [a,c], Qing Zhu [c], Eric Yau [d]

[a] Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong
[b] State Key Laboratory of Information Engineering in Surveying Mapping and Remote Sensing, Wuhan University, 129 Luoyu Road, Wuhan, Hubei, PR China
[c] Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University, Chengdu, Sichuan, PR China
[d] Ambit Geospatial Solution Limited, Sai Ying Pun, Hong Kong

ARTICLE INFO

Article history:
Received 15 November 2017
Received in revised form 28 February 2018
Accepted 1 March 2018
Available online 13 March 2018

Keywords:
Aerial oblique imagery
Terrestrial imagery
Photogrammetry
3D modeling

ABSTRACT

Photorealistic three-dimensional (3D) models are fundamental to the spatial data infrastructure of a digital city, and have numerous potential applications in areas such as urban planning, urban management, urban monitoring, and urban environmental studies. Recent developments in aerial oblique photogrammetry based on aircraft or unmanned aerial vehicles (UAVs) offer promising techniques for 3D modeling. However, 3D models generated from aerial oblique imagery in urban areas with densely distributed high-rise buildings may show geometric defects and blurred textures, especially on building façades, due to problems such as occlusion and large camera tilt angles. Meanwhile, mobile mapping systems (MMSs) can capture terrestrial images of close-range objects from a complementary view on the ground at a high level of detail, but do not offer full coverage. The integration of aerial oblique imagery with terrestrial imagery offers promising opportunities to optimize 3D modeling in urban areas. This paper presents a novel method of integrating these two image types through automatic feature matching and combined bundle adjustment between them, and based on the integrated results to optimize the geometry and texture of the 3D models generated from aerial oblique imagery. Experimental analyses were conducted on two datasets of aerial and terrestrial images collected in Dortmund, Germany and in Hong Kong. The results indicate that the proposed approach effectively integrates images from the two platforms and thereby improves 3D modeling in urban areas.

1. Introduction

Photorealistic three-dimensional (3D) models are fundamental to the spatial data infrastructure of a digital city, and have numerous potential applications in areas such as urban planning, urban management, urban monitoring, and urban environment studies (Haala and Kada, 2010; Qiao et al., 2010; Spagnuolo, 2014). The generation of 3D city models usually requires significant human intervention. Recent developments in aerial oblique photogrammetry have enabled the automatic generation of 3D mesh models in urban areas using aerial oblique images collected by cameras on aircraft or unmanned aerial vehicles (UAVs). Aerial oblique imagery is currently one of the commonly-used datasets for city-scale 3D reconstruction (Moe et al., 2016). However, in urban areas with densely distributed high-rise buildings, 3D mesh models generated from aerial oblique images may show geometric defects, e.g., inaccurate shapes, holes, merging of objects, and blurred textures, especially when representing building façades (see the example in Fig. 1(a)), due to problems such as occlusion and large camera tilt angles.

Meanwhile, terrestrial mobile mapping systems (MMSs) have been widely used in recent years to collect street-view data including images and/or laser scanning measurements. However, terrestrial MMS data do not provide full coverage (e.g., roof information is lacking). And in urban areas, Global Navigation Satellite System (GNSS) positioning may be inaccurate due to the blocking of satellite signals by tall buildings. The direct geo-referencing error of MMSs may become as large as 5 m or even worse for GNSS outrage (Chu and Chiang, 2016), which is below the required level of accuracy (better than 1 m) for city modelling (Gruen et al., 2013; Jende et al., 2016a,b). As aerial oblique imagery and terrestrial MMS data are now widely available for urban areas (see Fig. 1), the

* Corresponding author.
E-mail address: bo.wu@polyu.edu.hk (B. Wu).

https://doi.org/10.1016/j.isprsjprs.2018.03.004

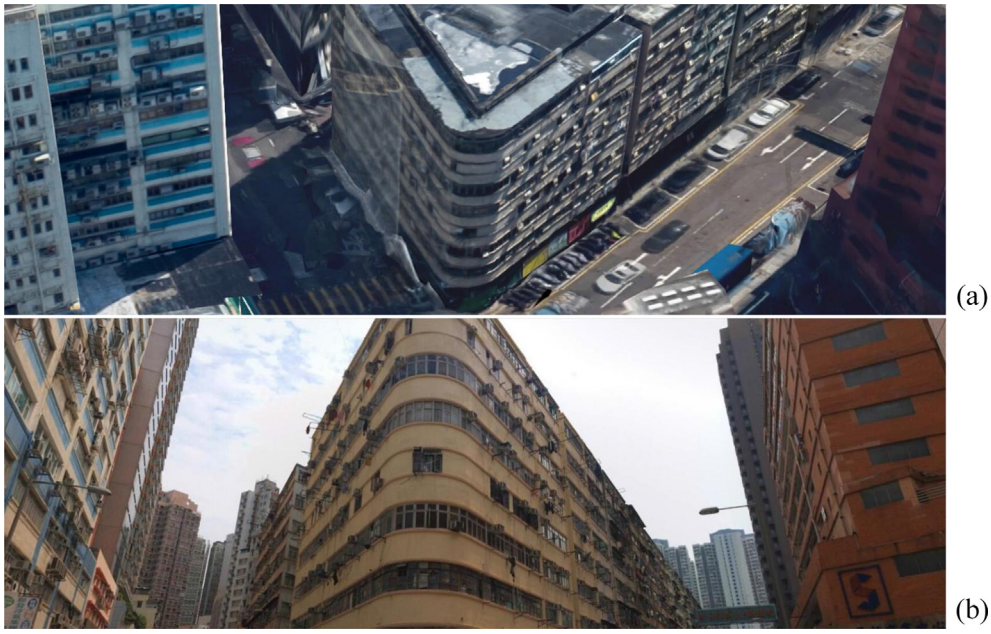

Fig. 1. Aerial oblique imagery and terrestrial MMS imagery in Cheung Sha Wang, Hong Kong. (a) 3D mesh model generated from aerial oblique imagery showing geometric defects and blurred textures, and (b) terrestrial MMS imagery of the same area.

integration of aerial oblique imagery with terrestrial imagery offers a promising means of optimizing 3D city modeling.

In this paper, we present a novel approach integrating the processing of aerial oblique imagery with terrestrial imagery to optimize 3D modeling in urban areas. The inputs comprise aerial and terrestrial image datasets and their respective initial image orientation parameters. By matching feature points between aerial oblique images and terrestrial images, combined bundle adjustment (BA) is performed for these two image datasets, yielding optimal image orientation parameters that better co-register the aerial and terrestrial datasets. These improved image orientation parameters and the sparse point clouds of the combined image block retrieved using our BA approach can then be imported to generate dense point clouds and surface models with better geometric accuracy using existing scene reconstruction software. Finally, the textures of building façades are optimized using image patches from terrestrial views, resulting in higher-quality 3D models.

The remainder of the paper is organized as follows. Section 2 briefly reviews existing works on 3D modeling based on aerial and terrestrial imagery. In Section 3, the proposed approach and its key steps are presented in detail. The performance of the proposed approach is evaluated in Section 4 using paired sets of aerial and terrestrial images. Conclusions are drawn and discussed in Section 5.

2. Related work

As the characteristics of aerial and terrestrial datasets are complementary, researchers have explored joint data processing methods to compensate for the failings of methods based on aerial or terrestrial data alone. Early works focused on methods of integrating airborne and ground-based laser scanning data, such as data registration (Von Hansen et al., 2008; Wu et al., 2013), absolute geo-referencing for visibility modeling of vegetation obstructions (Murgoitio et al., 2014), object detection and 3D city modeling (Böhm and Haala, 2005; Kedzierski and Fryskowska, 2014). Other researchers explored the potential of range data and images across these two platforms. Aerial images were jointly processed with terrestrial laser scanning data for digital elevation model generation (Ouédraogo et al., 2014), geological structure analysis (Chen et al., 2015), the seamless mapping of river channels (Flener et al., 2013), the 3D mapping and monitoring of open-pit mine areas (Tong et al., 2015), archaeological documentation (Balletti et al., 2015), and 3D urban reconstruction (Toschi et al., 2017). Recently, various applications of the integration of aerial and terrestrial imagery have been investigated, e.g. cross-platform registration (Jende et al., 2016a,b), the adjustment of mobile platform locations in Global Navigation Satellite System (GNSS) denied environments (Jende et al., 2017), and the geo-referencing of panoramic image sequences in complex urban scenes (Ji et al., 2015).

The degree of automation of image-based modeling pipelines has been significantly boosted by recent advances in photogrammetry and computer vision communities, which enable high-quality 3D point clouds to be generated from calibrated camera images (Nocerino et al., 2013). Meanwhile, novel commercial software and open-source packages allow even non-professional mapping practitioners to produce 3D models with high-resolution textures. Capitalizing on the newfound ability to acquire comprehensive images from both ground and non-ground views, aerial photogrammetry has been combined with terrestrial photogrammetry for accurate 3D reconstruction, with applications in areas such as cultural heritage protection (Bolognesi et al., 2014; Balletti et al., 2015), archaeology (Balletti et al., 2015), and urban environment studies (Rumpler et al., 2017). In addition, novel scientific benchmarking datasets facilitate the assessment of various algorithms and methodologies for image orientation and dense matching using ground-based and airborne images (Nex et al., 2015).

Off-the-shelf solutions for automatic 3D city modeling based on multi-view images involve two key sequential steps: structure from motion (SfM) (Hu et al., 2015; Gerke et al., 2016a,b; Xie et al., 2016) and multi-view stereo (MVS) (Hirschmüller, 2008; Zhu et al., 2010; Galliani et al., 2015; Hu et al., 2016). First, in the processing pipeline, feature point matching (Lowe, 2004) is used to automatically obtain tie points between views of images. Then, the orientation parameters of the images are then refined by bundle adjustment at the SfM stage (Agarwal et al., 2010). Next,

MVS is used to generate dense image matching point clouds (Furukawa and Ponce, 2010), which are triangulated (Kazhdan and Hoppe, 2013) and textured (Waechter et al., 2014) to give 3D mesh models.

When matching feature points, traditional aerial images with nadir views present only small differences in scale, rotation, and perspective distortion, because the roll and pitch angles of flight are relatively small. Traditional software packages rely on classical image matching strategies based on corner detectors such as Förstner (Förstner, 1986) and Harris (Harris and Stephens, 1988), with the normalized correlation coefficient as the matching metric. However, these methods are unsuitable for oblique image matching because such corner features are sensible to geometric differences. Since the groundbreaking introduction of the scale invariant feature transform algorithm (SIFT) (Lowe, 2004), the use of invariant features has offered another paradigm for image matching. Follow-up studies have been conducted to improve various aspects of SIFT, such as its dimensionality (Mikolajczyk et al., 2005), calculation speed (Bay et al., 2008), and distinctiveness (Tola et al., 2010). However, as SIFT-like features are not essentially affine invariant, the performance of this method decreases dramatically when translation tilt exceeds 25° (Mikolajczyk et al., 2005).

SfM can be used to reconstruct a 3D scene from given images by bundle adjustment (Triggs et al., 2000). Bundle adjustment optimizes camera poses and orientations, as well as the triangulated 3D points obtained from correspondences during feature matching. Bundle adjustment approaches are divided into two categories: sequential and global (Schonberger and Frahm, 2016). Sequential methods perform reconstruction from a minimal robust cluster, such as a pairwise model or a reconstructed triplet, incrementally adding new images to existing clusters. This approach performs well when the initial orientation parameters of images are inaccurate or even missing, but computation cost increases with each increment in reconstruction; a divide-and-conquer strategy can be adopted to reduce computation cost (Snavely et al., 2008). In contrast, global methods generally estimate relative orientations of all the images in one go and estimate global rotation and translation separately. However, ensuring the convergence of the global optimization algorithms may be difficult, demanding robust initial estimations and reliable outlier detection (Moulon et al., 2013; Toldo et al., 2015; Schonberger and Frahm, 2016).

Once the images are oriented, dense image matching (DIM) can be used to turn 2D images into 3D point clouds (Nex and Gerke, 2014; Forlani et al., 2015). Existing commercial solutions, e.g., ContextCapture (Bentley, 2018), PixelFactory (Airbus Intelligence, 2018), and PhotoScan (Agisoft, 2018), efficiently generate 3D meshes with textures from point clouds and oriented images. Texturing is also an important stage in the process of obtaining final 3D mesh models. However, this process is very challenging, due to problems such as changes in illumination and exposure (Tan et al., 2008), non-rigid scene parts, unreconstructed occluding objects, and image scales that may vary by several orders of magnitude between close-up views and distant overview images (Waechter et al., 2014). In addition, as textures are sampled through each triangle, texture maps should be packed into a single image using approximate bin packing algorithms (Zhou et al., 2004).

As previously mentioned, the integration of aerial oblique imagery and terrestrial imagery offers promising opportunities for optimized 3D city modeling, as the corresponding types of data processing are complementary and possible occlusions can be alleviated by combining aerial and terrestrial views. However, differences in view perspective between aerial oblique imagery and terrestrial imagery may be as large as 90°, and differences in scale between the two types of image are severe. In practice, even methods reported to be fully affine invariant (Morel and Yu, 2009) cannot find valid matches between them. Although a few researchers have investigated aerial and terrestrial image matching (Morel and Yu, 2009) and offered some preliminary results (Gerke et al., 2016a,b), little substantial progress has been made in this area. In addition, the orientation parameters of terrestrial images obtained through the GNSS and inertial measurement unit (IMU) may show systematic deviation, and unstable positioning accuracies between terrestrial image blocks may arise due to various signal environments in urban areas. To address these problems, this paper presents an approach integrating the processing of aerial oblique imagery and terrestrial imagery to optimize 3D modeling in urban areas. This approach has the following novel aspects: (1) automatic feature matching between aerial oblique imagery and terrestrial imagery by base-plane fitting and image rectification; (2) combined bundle adjustment of aerial and terrestrial images via a two-step optimization strategy that ensures both inter-platform stability and cross-platform geometrical consistency; and (3) geometric refinement and textural optimization of the aerial 3D mesh model using co-registered terrestrial images.

3. Integration of aerial oblique imagery and terrestrial imagery

3.1. Overview of proposed approach

A fundamental step in the integrated processing of aerial oblique imagery and terrestrial imagery for 3D modeling is to ensure that they are georeferenced or, at least, have a unique spatial reference. As aerial oblique images and terrestrial images are captured separately by sensors onboard different platforms, inconsistencies between the two datasets are inevitable. To integrate the datasets for synergistic uses, the inconsistencies must be removed by feature matching and combined bundle adjustment.

Fig. 2 shows the overall workflow of the proposed approach. First, initial exterior orientation (EO) parameters of the aerial and terrestrial imagery are obtained from off-the-shelf SfM solutions through self-calibrating bundle adjustment with available GNSS data. DIM point clouds and initial 3D mesh models are derived from the aerial and terrestrial imagery, respectively. This ensures the internal geometrical consistency of the two image sets, respectively. Next, the DIM point clouds derived from the aerial images are used to produce a footprint of the target buildings, and roof heights are estimated accordingly. Vertical planes serving as candidate base planes are fitted from DIM point clouds near the façade area. Accordingly, both the aerial and the terrestrial images are rectified onto the base planes obtained from visibility analysis. The visibility analysis is based on the image EO parameters and the fitted vertical planes. Feature matching is then conducted with the rectified aerial and terrestrial images, and pixel coordinates are transformed to their original images. The matched feature points are input into a combined bundle adjustment process to co-register the aerial and terrestrial images. Finally, refined EO parameters for both image sets and the sparse 3D point clouds obtained from the combined bundle adjustment are input into an existing MVS procedure (i.e., ContextCapture (Bentley, 2018)) to improve geometric reconstruction, and texture mapping is optimized using patches from the high-resolution terrestrial images to substitute for the blurred textures of building façades in the aerial images. This yields optimized photorealistic 3D mesh models with refined geometry and texture.

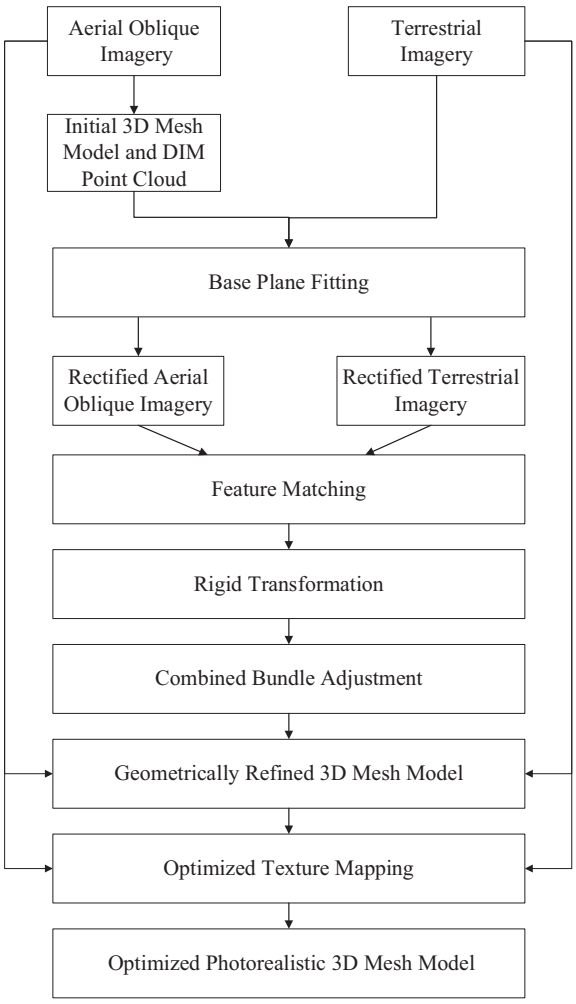

Fig. 2. Overall workflow of the proposed approach.

3.2. Feature matching between aerial oblique imagery and terrestrial imagery

Aerial oblique imagery and terrestrial imagery show large geometric differences, such as perspective distortion and scale differences, as illustrated in Fig. 3. This creates a significant obstacle to feature matching. As the major scenes in terrestrial images are building façades, and the building roofs observed in aerial images are invisible in terrestrial images, feature matching is assumed to involve mainly on building façades. Accordingly, we first project both the aerial and terrestrial images onto base planes based on building façades, and then conduct feature matching with the rectified images on the base planes, which have similar perspective and scale characteristics.

(1) Determination of the base plane

The first step is to determine a base plane (or base planes in complex cases) for the terrestrial images. Initial 3D point clouds can be derived from the aerial oblique images from an off-the-shelf pipeline. The EO parameters for both the aerial and the terrestrial images can be obtained from their separated data processing. Although the two datasets may show systematic inconsistencies in orientation parameters, e.g., global shifts or scale differences, they can be treated as roughly co-registered. Meanwhile, DIM point clouds from the aerial oblique images can be used to fit finite vertical planes. These planes provide candidate base planes for further image rectification. The orientation parameters for each type of image are used to determine the visibility of each candidate plane; only visible planes on both types of images are selected as base planes.

(2) Rectification of aerial and terrestrial images

After determining the base plane and its boundaries, the rectification process is similar to the classical method of orthorectification in photogrammetry (Hu et al., 2016). The ground sampling distance of the aerial images is used as the target resolution of the rectified images. The plane is then sampled using grids at the above resolution. Each grid is projected onto the original image and interpolated to obtain a pixel on the rectified image. In this way, the scale differences and perspective distortions are alleviated or even removed. As illustrated in Fig. 3, the building façades in the rectified aerial images resemble those in the rectified terrestrial images, which enhances the accuracy of subsequent feature matching.

(3) Feature matching with the rectified images

After rectification, the building façades in images with the same base plane should have similar geometric properties. Feature points are then extracted from the images using the SIFT detector. Descriptors are constructed for the feature points. The two nearest neighbors in the descriptor space are identified using an approximate nearest neighbors search. A cross-check is conducted to ensure that the correspondences are all unique. As the epipolar geometry of the rectified images is not determined, a more general homograph model is used as the geometric kernel for random sample consensus outlier detection (RANSAC) (Fischler and Bolles, 1981). A recent variant of RANSAC, A Contrario-RANSAC (Moisan et al., 2012), which features automatic threshold determination and better inlier retrieval, is implemented. After outliers have been removed, pairwise matches are connected to form tracks using a connected component algorithm. A grid mask is used to ensure that the selected matches are evenly distributed. Finally, the image coordinates of the matches are inversely transformed to the original image space for the subsequent combined bundle adjustment.

3.3. Combined bundle adjustment of aerial oblique imagery and terrestrial imagery

After acquiring tie points from the feature matching of the aerial and terrestrial images, combined bundle adjustment is conducted to optimize the image EO parameters and thereby remove inconsistencies between the two datasets. This process of adjustment should not be conducted freely, for the following two reasons. (1) The aerial and terrestrial images should have relatively good inter-platform consistency, as can be reflected by the EO parameters of the images within each platform. Cross-platform geometrical consistency, as can be reflected by the EO parameters of the images cross the aerial and terrestrial platforms, is the main problem. (2) Although the tie points may cover a large area of the terrestrial images, they may only cover a small area of the aerial images, and traditional approach of free optimization of orientation parameters for all images in the bundle block will lead to unstable results. To conduct combined bundle adjustment, we propose the two-step optimization strategy illustrated in Fig. 4, which ensures both inter-platform stability and cross-platform consistency. In the first step, inter-platform image orientations are maintained and the problem of cross-platform non-conformity is solved. In the second step, the inter-platform consistency is refined

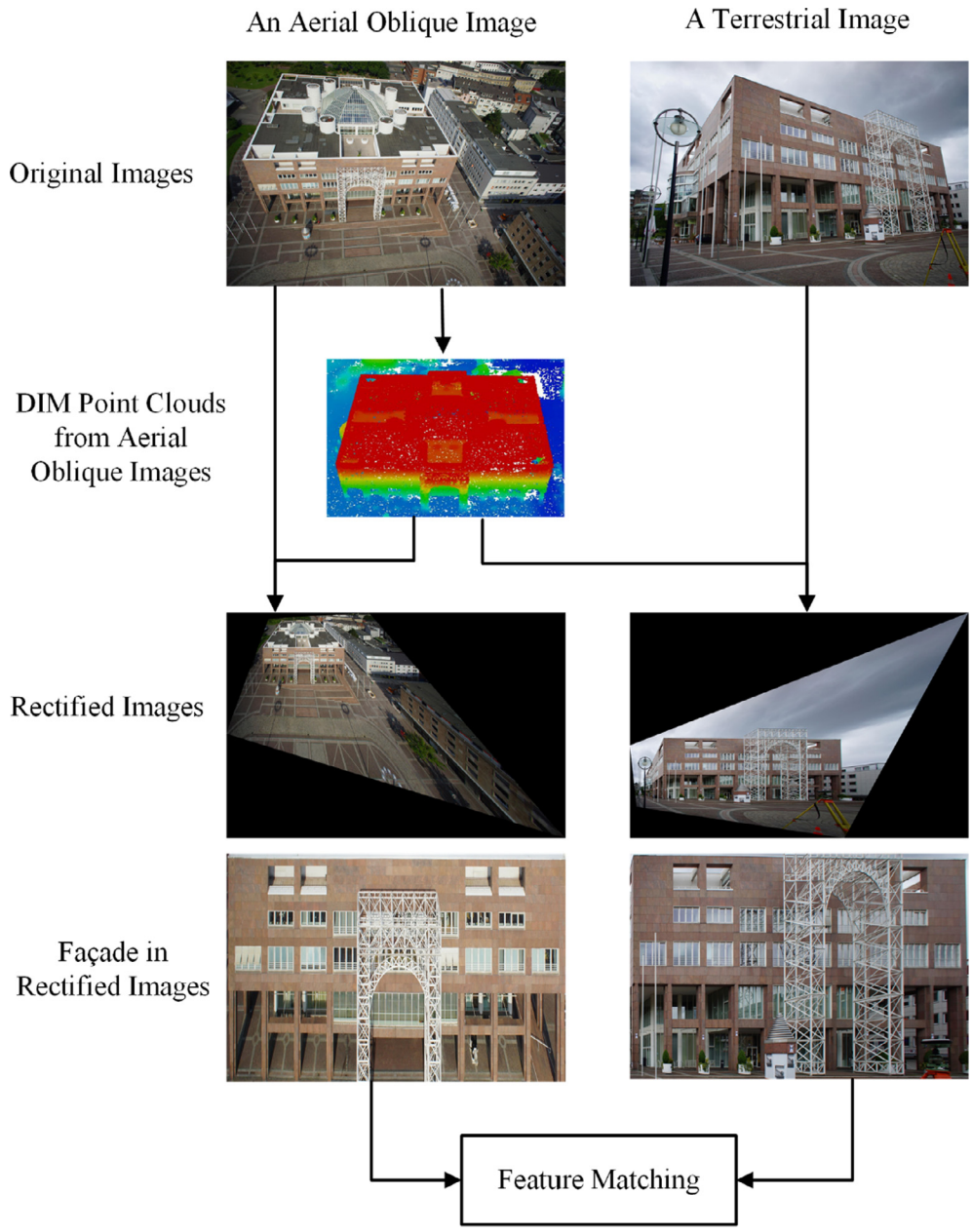

Fig. 3. Rectification on base planes for the feature matching of aerial and terrestrial images.

and the cross-platform consistency obtained in the first step are used as rigid constraints.

(1) Cross-platform rigid transformation

As the potential mated image pairs between aerial oblique images and terrestrial images cannot be evenly distributed, free bundle adjustment which optimizes the orientation parameters for all images based on the tie point observations is likely to destroy the already well-recovered cross-platform consistency achieved by the off-the-shelf software system. Thus, in the first step, we treat images from aerial and terrestrial platforms as two rigid groups, and adjust only the global rotation, translation, and scale parameters, using the following equation:

$$\begin{aligned} R_1 &= RR_0 \\ P_1 &= sRP_0 + T \end{aligned} \tag{1}$$

where (R_0, P_0) and (R_1, P_1) represent the original and transformed rotational and positional EO parameters of the images, respectively; s is a global scale factor; R represents a global rotation matrix; and T indicates a 3D translation vector.

Separate sets of rigid transformation parameters are used for the aerial and terrestrial image groups. The observation equations for the rigid transformation can be represented in matrix form using the following equation:

$$V = AX - L, P \tag{2}$$

where X is the unknown vector to be solved, containing the transformation parameters (s, R, T); L is the observation vector; A is the coefficient matrix containing the partial derivatives from each observation; and P is the a priori weight matrix of the observations, which reflects the measurement quality and the contribution of the observation to the final result. Normally, aerial oblique images provide good overall accuracy from the off-the-shelf

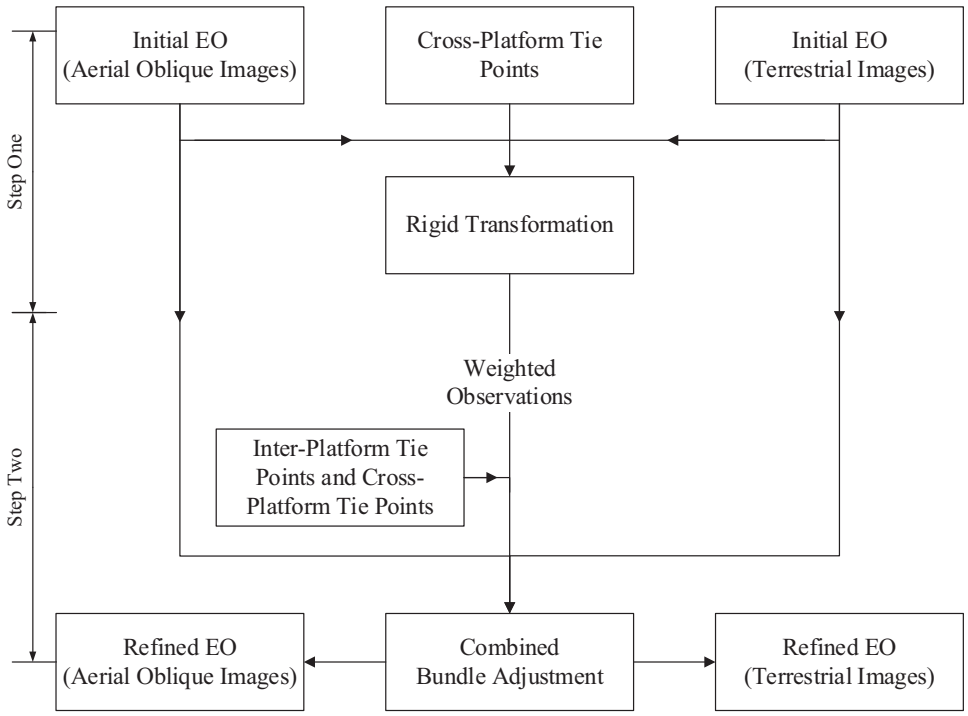

Fig. 4. Proposed two-step combined bundle adjustment strategy.

pipeline and cover a large area; in contrast, terrestrial images tend to cover only segmented regions within the area, and their geometric accuracy may be reduced by GNSS tracking loss in urban areas. In Equation (2), therefore, we award larger weights to the EO parameters of the aerial oblique images than to those of the terrestrial images. The EO parameters of the aerial oblique images are thus adjusted less and those of the terrestrial images are adjusted more. If the off-the-shelf pipeline already provides good overall accuracy for the aerial oblique images, e.g., favorable statistics of residuals with checkpoints in object space, infinite weights can be assigned to their EO parameters to ensure that they are not changed in the process. In this case, only the set of transformation parameters to transform the terrestrial images to the aerial oblique images are obtained.

The unknown parameters in Equation (2) can be clustered into three groups: the EO parameters of the images in each dataset; the rigid transformation parameters of the two datasets; and the 3D coordinates of the tie points. Their initial values can be estimated as follows. For the first group of parameters, initial EO parameters obtained from the off-the-shelf pipeline (for the aerial images) or the MMS (for the terrestrial images) can be used. For the second group of parameters, identity values are adopted as initial values. For the third group, the 3D coordinates obtained through photogrammetric space intersection using the initial EO parameters can be used. Based on these initial values, incremental corrections to the unknowns can be made iteratively, and the unknowns can finally be solved in a least-squares manner.

(2) Combined bundle adjustment with rigid constraints

Although the above step optimizes cross-platform geometric consistency, the already-achieved inter-platform accuracy may have degenerated to some degree due to scale variation. Therefore, in the second step, we seek to obtain optimal image orientation parameters via both inter-platform and cross-platform methods. In this step, all tie points (whether inter-platform or cross-platform) serve as image observations, optimizing orientation parameters for all images within the bundle adjustment process based on the co-linearity equation. Meanwhile, the rigid transformation parameters obtained in the first step are treated as weighted observations to constrain that the previously achieved cross-platform consistency is retained to some extent. Thus, the combined bundle adjustment process in the second step includes the following two types of observation:

$$V_1 = A_1X_1 - L_1, P_1$$
$$V_2 = A_2X_2 - L_2, P_2$$

where V_1 represents the residual vectors of the image observations in the co-linearity equation and V_2 stands for the deviations of the EO parameters from those computed through the previous rigid transformation.

As the number of tie points matched between the aerial oblique imagery and the terrestrial imagery may be significantly lower than that of tie points matched between images belonging to the same platform, the weights given to the image observations should not be distributed identically. Although fewer cross-platform tie points may be obtained, they are more important than inter-platform tie points. Therefore, we award larger weights to the cross-platform tie points. Note that, in this step, a self-calibrating combined bundle adjustment which re-compute the interior orientation parameters of images may help improve the performance.

Via the two-step combined bundle adjustment of the aerial and terrestrial images, the terrestrial images are co-registered to the aerial images and their inter-platform geometrical consistency is maintained.

3.4. Geometric refinement and textural optimization of aerial 3D mesh models

After the combined bundle adjustment, the aerial and terrestrial images, their improved orientation parameters, and the sparse 3D point clouds generated from the inter-platform and cross-platform tie points from the combined bundle adjustment are imported to

an existing MVS procedure (i.e., ContextCapture (Bentley, 2018)), allowing dense point clouds and 3D mesh models with refined geometric accuracy to be generated.

Having obtained 3D mesh models with enhanced geometric accuracy, we remap the blurred textures of the building façades using the high-resolution terrestrial images. As the EO parameters of the terrestrial images have already been adjusted and co-registered to ensure consistency with the aerial images, the areas of texture in the terrestrial images should be consistent with those in the 3D mesh models. This assumption is used to automatically update the textures of the building façades, which may be blurred or have an undesirably low resolution.

The vertexes of each triangle in the 3D mesh model in the façade area are projected to the oriented terrestrial images to receive new texture coordinates. To select the optimal terrestrial image, with desirable texture and resolution characteristics, we first calculate the normal vectors of the detected façades. These normal vectors are intersected with the orientation angles of the images, and those with smaller intersection angles are selected for texture mapping, as they result in better viewing conditions. In addition, adjacent triangles are used to select textures from the same image to minimize the inevitable seams. Furthermore, as the illumination conditions when collecting the aerial and terrestrial images may not have been consistent, the photometric quality of the images may differ. The radiometric property of the terrestrial images can be corrected to match that of the aerial images. We perform histogram matching between the texture images of the same area. First, façade image patches from the terrestrial images and aerial oblique images are extracted; next, histograms for the aerial oblique images are calculated and set as a reference. Finally, histograms for the terrestrial image patches are matched with the reference histograms.

4. Experimental analysis

4.1. Experimental data description

In order to evaluate the performance of the proposed approach, a benchmark dataset collected at Dortmund, Germany is used for systematic experimental analysis and comparing with existing works. While a more challenging dataset collected in Hong Kong is employed to further verify the usability of the proposed approach.

The first dataset comprises a UAV image block and a terrestrial image block provided by the International Society for Photogrammetry and Remote Sensing (ISPRS) and EuroSDR (Nex et al., 2015). The images were taken in the center of Dortmund, Germany. The UAV images were captured by a Sony Nex-7 mounted on a multi-rotor DIJ S800, and the terrestrial images were acquired by the same camera on the ground (see Fig. 3). The GSDs for UAV and terrestrial images are ranging from 1 cm to 3 cm. Together with the images and their initial positions, 13 ground control points (GCPs), 27 checkpoints, and the terrestrial laser scanning point clouds are provided. Note that, all the GCPs and checkpoints are measured on terrestrial images only. A detailed description of this dataset is available in (Nex et al., 2015).

The second dataset was collected in Cheung Sha Wan, Hong Kong. Both the nadir and the oblique images were obtained using a UAV-borne Sony Ilce-qx1 camera, and the terrestrial images were collected by an UltraCam Mustang MMS. The GSDs for aerial images are ranging from 6 cm to 8 cm, and the minimal GSD for the terrestrial imagery is about 2 cm. The target building and representative aerial and terrestrial MMS images are shown in Fig. 1. Compared with the Dortmund dataset, the Hong Kong dataset is more challenging due to the larger differences in spatial resolutions and viewing directions between the aerial and terrestrial images.

Since the benchmark dataset enables comparing with existing works, the performance evaluation of the proposed feature matching and bundle adjustment approaches are conducted for this dataset only. For both datasets, the 3D mesh models generated by integrating aerial and terrestrial imagery are evaluated both qualitatively and quantitatively.

4.2. Evaluation of the feature matching approach

In order to evaluate the performance of the proposed feature matching strategy, four challenging image pairs from the first dataset are selected and compared with existing works. In the experiments, the GPU implementation of SIFT (GPU-SIFT) algorithm (Wu, 2007) is selected as a standard method for feature matching on rectified images. Both pairwise feature matching results and matching for the entire image block are compared. For pairwise feature matching, comparison of numbers of matches and inliers are carried out among the proposed approach, the GPU-SIFT method, and other feature matching approaches including the SIFT (Lowe, 2004), SURF (Bay et al., 2008), ORB (Rublee et al., 2011), AKAZE (Alcantarilla et al., 2013), KAZE (Alcantarilla et al., 2012), and ASIFT (Morel and Yu, 2009) as reported in Gerke et al. (2016a). Meanwhile, for the matching results for an entire image block, comparisons are conducted between the proposed approach and the standard GPU-SIFT method, and combined bundle adjustment from the matching results is involved to evaluate the matching performances.

The pairwise feature matching results are shown in Figs. 5–8. In pair-1 (see Fig. 5) which includes two terrestrial images with orientation changes of nearly 90°, the proposed method outperformed all the other methods. Traditional GPU-SIFT method obtained 127 correct matches while the proposed methods obtained 558 correct matches. Meanwhile, for the other six methods, only SURF got more than 50 correct matches.

A more challenging image pair is shown in Fig. 6. The images were captured on different platforms and scales. Besides, since the images were collected on different dates, lighting conditions were different (see the windows on façade). In this case, the proposed method successfully obtained 148 correct matches, while the GPU-SIFT and SURF obtained 14 and 20 correct matches, respectively. Other methods failed to obtain any correct matches.

As shown in Fig. 7, pair-3 is also a cross-platform image pair but with larger differences in image scales and viewing angles. Besides, radiometric differences are distinct because of the different lighting conditions. Apparently, the proposed method is the only one that successfully found a sufficient number of correct matches for further integration processing. Traditional approaches failed in this case possibly due to severe scale changes and perspective distortions.

Among the four test image pairs, pair-4 is the most challenging one. As illustrated in Fig. 8, the building façade is most visible in the terrestrial image and serves as the foreground of the image, while the same building façade is fully visible in the UAV image but from a far and tilt view. Under such circumstance, although the proposed method found 22 correct matches, the successful matching rate is decreased. At the same time, all other approaches failed in this case.

In view of the results of pairwise image matching, the performance of the proposed aerial-ground image matching strategy is affirmed to be effective. Compared with other existing methods, the proposed approach involves image rectification that alleviates the problems of perspective distortion and scale difference, which substantially increases the number of correct matches.

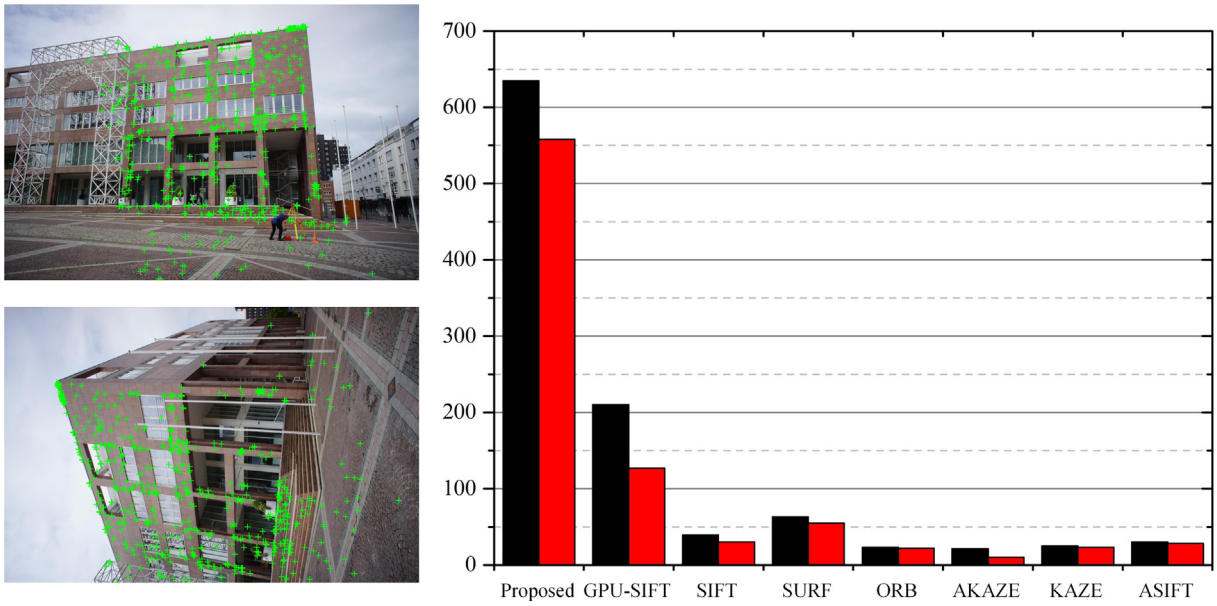

Fig. 5. Pairwise feature matching comparison on pair-1 (2342–2392). Left: matches obtained by the proposed method; Right: number of matches obtained by different methods, black bars indicate match numbers while red bars indicate inliers. Results for SIFT, SURF, ORB, AKAZE, KAZE, ASIFT are referred from Gerke et al. (2016a).

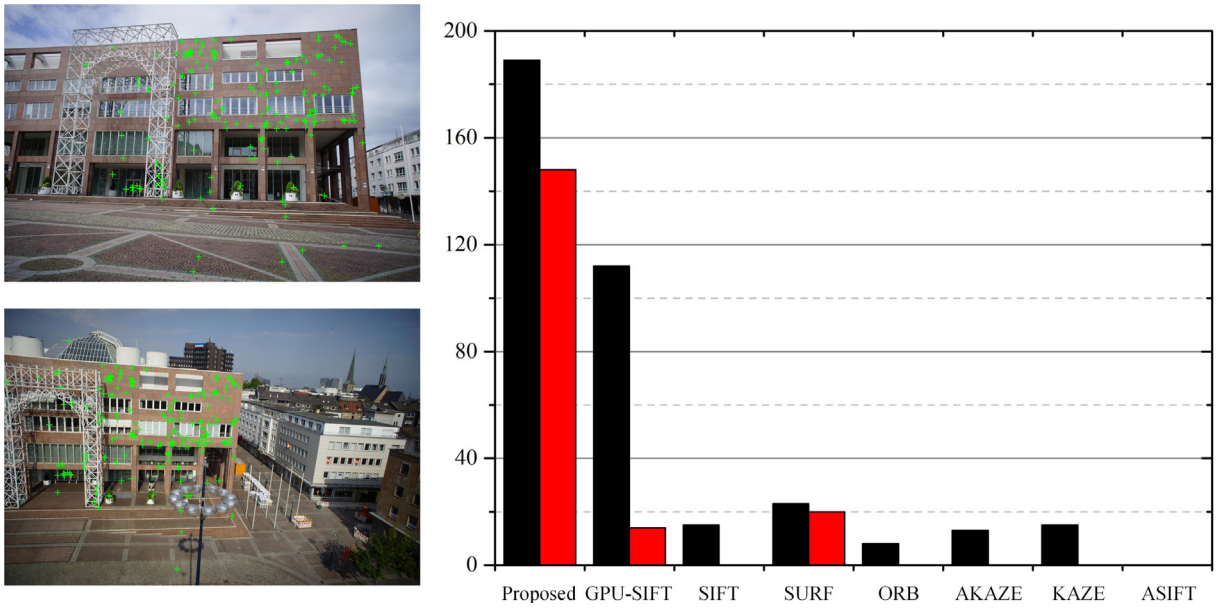

Fig. 6. Pairwise feature matching comparison on pair-2 (2315–7055). Left: matches obtained by the proposed method; Right: number of matches obtained by different methods, black bars indicate match numbers while red bars indicate inliers. Results for SIFT, SURF, ORB, AKAZE, KAZE, ASIFT are referred from Gerke et al. (2016a).

Apart from the above pairwise image matching, comparison of matching results in image blocks are also conducted between the proposed feature matching method and the standard GPU-SIFT method. The numbers of the matched points from each method are counted and the performances of using the matched points for combined BA are analyzed. The results are shown in Table 1. Using the proposed method, in the entire image block, 1,031,116 tie points are automatically generated, but only 7,189 points connect the aerial and terrestrial imagery. In contrast, the total number of tie points obtained by the GPU-SIFT is evidently less than that from the proposed method, but interestingly the number of cross-platform tie points is decuple of the former. Then, combined BA is conducted using the two sets of tie points, respectively. The results show that using the tie points provided by the proposed method succeeded in the combined BA. Meanwhile, the combined BA using the tie points from GPU-SIFT failed, indicating that a large number of cross-platform tie points may be overwhelmed by mismatches.

4.3. Evaluation of the combined bundle adjustment approach

In order to evaluate the performance of the proposed two-step combined BA approach, comparisons with the standard BA method are conducted by analyzing residuals of checkpoints in object space. Besides, for the purpose of evaluating the achieved geometrical accuracy in object space by integrating aerial and terrestrial imagery, comparisons with those from a single source of images are also carried out. For the benchmark dataset, since all the GCPs

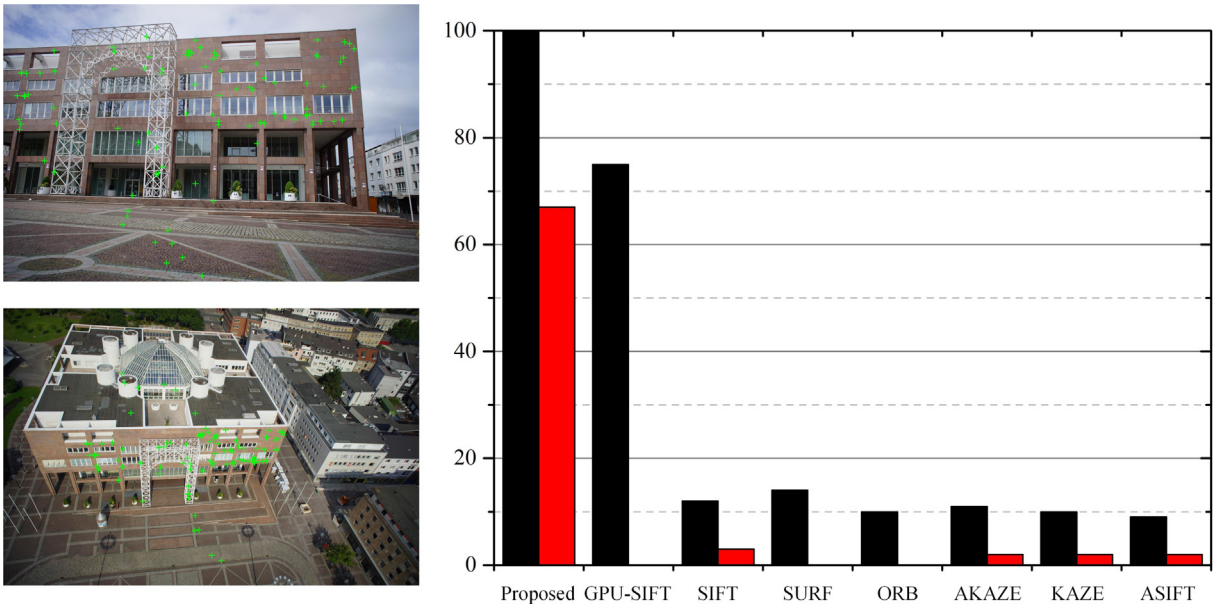

Fig. 7. Pairwise feature matching comparison on pair-3 (2315–7106). Left: matches obtained by the proposed method; Right: number of matches obtained by different methods, black bars indicate match numbers while red bars indicate inliers. Results for SIFT, SURF, ORB, AKAZE, KAZE, ASIFT are referred from Gerke et al. (2016a).

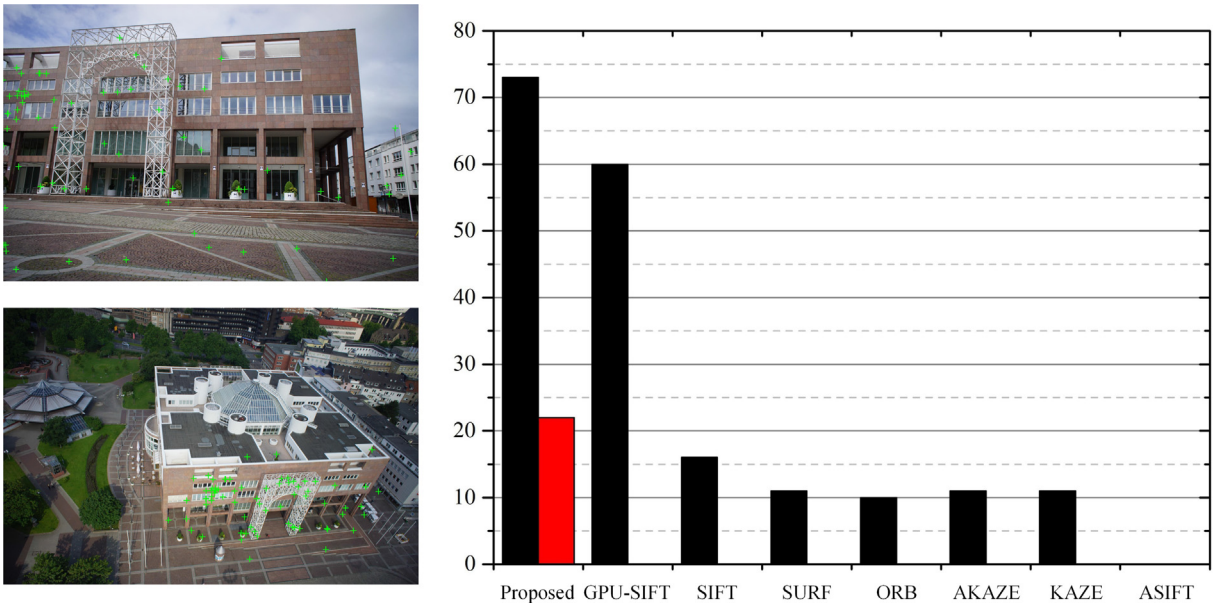

Fig. 8. Pairwise feature matching comparison on pair-4 (2315–7126). Left: matches obtained by the proposed method; Right: number of matches obtained by different methods, black bars indicate match numbers while red bars indicate inliers. Results for SIFT, SURF, ORB, AKAZE, KAZE, ASIFT are referred from Gerke et al. (2016a).

Table 1
Comparison of feature matching results in image blocks.

Method	Total number of tie points	Inner-platform tie points	Cross-platform tie points	Combined bundle adjustment
Proposed	1,031,116	1,023,927	7189	Succeeded
GPU-SIFT	644,550	570,452	74,098	Failed

and checkpoints are measured on terrestrial images only, as mentioned previously, only the residuals of checkpoints for terrestrial images are analyzed and presented here.

The residuals of checkpoints in object space from different methods are shown in Table 2. The root-mean-square error (RMSE) values in the X, Y, and Z directions for the proposed BA method are 0.028 m, 0.069 m, and 0.036 m, respectively, indicating that the approach achieves favorable planar and vertical accuracies. The RMSE in XYZ is about 0.083 m, indicating that desired absolute geometrical accuracy has been achieved by the proposed approach. As for the standard BA method, statistics in the Y direction are rather competitive with the proposed method with an RMSE of

Table 2
Statistics of residuals of checkpoints after combined bundle adjustment. Proposed: results for the proposed two-step BA method (bold values); Standard: results for the standard BA method; Terrestrial Only: results for terrestrial image block using the standard BA method.

		X (m)	Y (m)	Z (m)	XYZ (m)
Mean	Proposed BA	***−0.006***	***−0.051***	***0.017***	***0.054***
	Standard BA	−0.004	−0.041	0.079	0.089
	Terrestrial only	0.169	−0.076	−0.077	0.201
Std. dev.	Proposed BA	***0.015***	***0.008***	***0.002***	***0.017***
	Standard BA	0.012	0.007	0.007	0.016
	Terrestrial only	0.051	0.002	0.021	0.055
RMSE	Proposed BA	***0.028***	***0.069***	***0.036***	***0.083***
	Standard BA	0.063	0.061	0.086	0.123
	Terrestrial only	0.217	0.110	0.115	0.269

0.061 m. Besides, the RMSE in the X and Z directions are 0.063 m and 0.086 m, respectively; both are double or even greater than those achieved by the proposed method. The achieved accuracies in the X and Z directions are more noticeable than that in the Y direction; this may relate to the uneven distribution of cross-platform tie points, which were used to evaluate the rigid transformation parameters in the first step of the proposed BA strategy. Furthermore, in terms of the RMSE in 3D, the standard BA is about 0.123 m. Overall, the proposed BA method outperforms the standard BA method for about 48% in terms of geometrical accuracy.

Notice that, for the image block comprising of terrestrial images solely, the achieved accuracies are obviously weaker than those from the integration of aerial and terrestrial images. The RMSE in all three directions is greater than 0.1 m and the RMSE in 3D is 0.269 m. The results indicate that by integrating with aerial oblique images, the connectivity of the terrestrial image block is strengthened and the bundle block becomes more robust as more connected image observations are included.

4.4. Evaluation of the generated 3D mesh models

Both the Dortmund dataset and the Hong Kong dataset are adopted to evaluate the qualities of the 3D mesh models generated by integration of the aerial and terrestrial images. Both qualitative and quantitative comparisons are conducted. A qualitative comparison is carried out by visualizing the 3D mesh models generated using the proposed approach versus those generated by off-the-shelf software from aerial oblique imagery only. To quantitatively measure the geometrical accuracy of the generated 3D mesh models, we use geo-referenced terrestrial laser scanning point clouds as ground truth and calculate the unsinged and signed cloud-mesh distance (CMD) for the 3D mesh models generated solely using aerial oblique images and those generated from the proposed integration approach of aerial and terrestrial images. The computation of CMD is performed using the CloudCompare method (Girardeau-Montaut, 2015), of which the cloud-mesh corresponding relationships are defined as the nearest mesh for each point. The unsigned CMD is used to characterize the absolute geometric errors, while the signed CMD helps to further evaluate the achieved accuracies of 3D mesh models.

4.4.1. Results for the Dortmund dataset

Fig. 9 provides a side-by-side comparison of the 3D mesh models generated using the two methods. The first column in Fig. 9 visualizes the 3D mesh models as viewed from southwest to northeast. Both models look acceptable. However, significant differences in geometrical and textural quality become evident in the second column, which visualizes the 3D mesh models as viewed from northeast to southwest. This improvement is even clearer in the enlarged views in the third column. The aerial oblique images alone provide unsatisfactory viewing angles of the north wall of the building, leading to unfavorable results for this part. The proposed approach significantly improves the geometrical and textural quality by incorporating terrestrial images for integrated processing.

The CMD of the four façades is calculated separately, as terrestrial laser scanning data are available for each of the four façades. The CMD distribution and statistics are shown in Fig. 10 and Table 3, respectively. Fig. 10 clearly shows that in the higher part of the building, unsigned CMD is low for both 3D mesh models; and that in the lower part of the building, unsigned CMD is significantly higher for the model generated from aerial oblique images solely than the one from integrated processing of aerial and terrestrial images, due to possible occlusion and imperfect viewing angles caused by the complexity of city scene. As illustrated in Table 3, the mean unsigned CMD obtained for the 3D mesh models generated from aerial images alone ranges from 0.127 m to 0.212 m, with standard deviations all greater than 0.24 m, while the signed CMD varies from 0.019 m to 0.098 m with even larger standard deviations. In contrast, the proposed approach largely improves the geometrical accuracy of the resulting 3D mesh models. The mean unsigned CMD for the model generated using the proposed approach fluctuates between 0.051 m and 0.071 m, and the maximum standard deviation is 0.161 m. The mean of signed CMD is even at millimeter level. Note that the mean signed/unsigned CMD for the north façade, which was captured via unsatisfactory viewing angles in the aerial oblique images, decreases largely as a result of incorporating the terrestrial view images, again indicating a remarkable improvement for this part (see also Fig. 9).

4.4.2. Results for the Hong Kong dataset

Fig. 11 provides a side-by-side comparison of the 3D mesh models generated for the Hong Kong dataset. The first column in Fig. 11 visualizes the 3D mesh models as viewed from top to bottom. The rooftops generated using the two methods look similar, as this portion of the scene is invisible in the terrestrial images. In the second column, the visualization of the 3D mesh models as viewed from west to east reveals a significant difference in the geometrical and textural quality of the building façades, as further demonstrated in the enlarged views in the third column. In the first row in Fig. 11, some geometrical details are obscured or even absent, and the texture is seriously blurred and distorted. The aerial oblique images have unsatisfactory viewing angles and low image resolutions, leading to unfavorable 3D modeling results for this part. Integrating aerial oblique images with terrestrial images using the proposed approach provides much better geometrical and textural accuracy for the building façade.

As the available laser scanning data for this dataset covers only the bottom part of the building, the quantitative evaluation is limited to this area. The CMD distribution and statistics for the derived

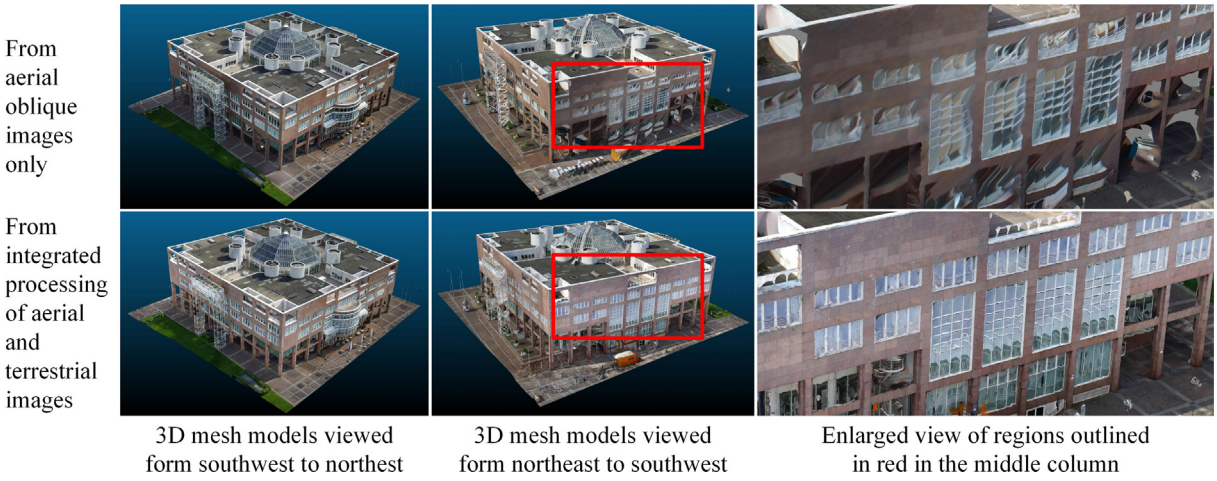

Fig. 9. 3D mesh models generated from aerial oblique images only (first row) and from the integration of aerial oblique images and terrestrial images (second row) for the Dortmund dataset.

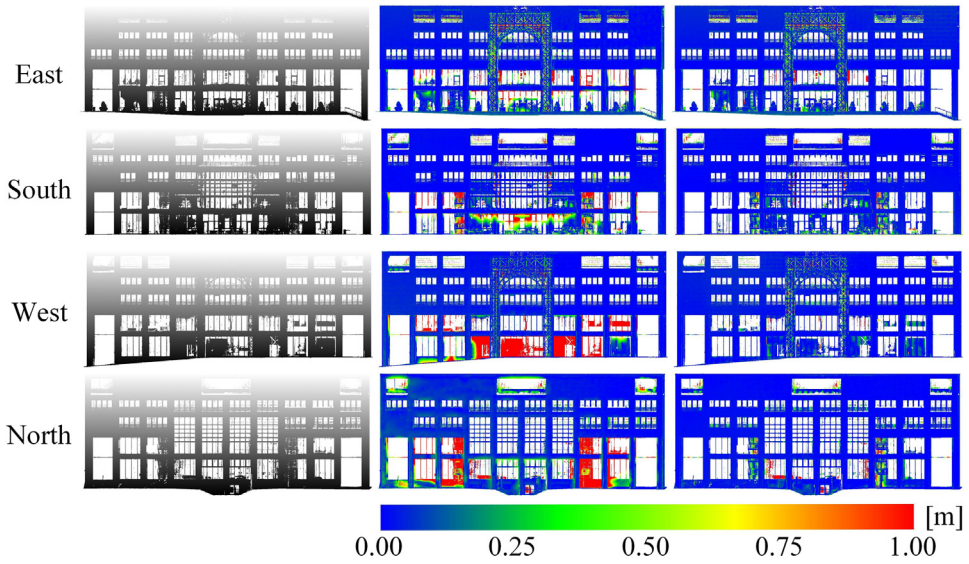

Fig. 10. CMD figures for the Dortmund dataset. The left-hand, middle, and right-hand columns illustrate laser scanning point clouds shaded by height; unsigned CMDs for the 3D mesh model from aerial oblique images only; and unsigned CMDs generated from the integrated processing of aerial and terrestrial images, respectively.

Table 3
CMD statistics for the Dortmund dataset.

	From aerial oblique images only				From integrated processing of aerial and terrestrial images			
	Mean (m)		Std. dev. (m)		Mean (m)		Std. dev. (m)	
	Signed	Unsigned	Signed	Unsigned	Signed	Unsigned	Signed	Unsigned
East	0.032	0.135	0.281	0.248	0.009	0.071	0.158	0.141
South	0.019	0.127	0.296	0.267	−0.009	0.054	0.139	0.128
West	0.078	0.179	0.355	0.317	−0.003	0.051	0.120	0.109
North	0.098	0.212	0.378	0.327	−0.002	0.068	0.174	0.161

3D mesh models are given in Fig. 12 and Table 4. As shown in Fig. 12, the geometrical accuracy of the 3D mesh model generated from the aerial oblique images alone is relatively low for the vertexes surrounding the extruded part of the façade, mainly due to inferior image detail, which limits the accuracy of depth reconstruction. After integrated processing with terrestrial images, the mean unsigned CMD falls from 0.158 m to 0.096 m and the standard deviation decreases from 0.109 m to 0.075 m. Meanwhile, the mean signed CMD also changes from -0.117 m to -0.045 m. This confirms the higher geometrical accuracy of the 3D mesh models acquired by integrating aerial and terrestrial images using the proposed approach.

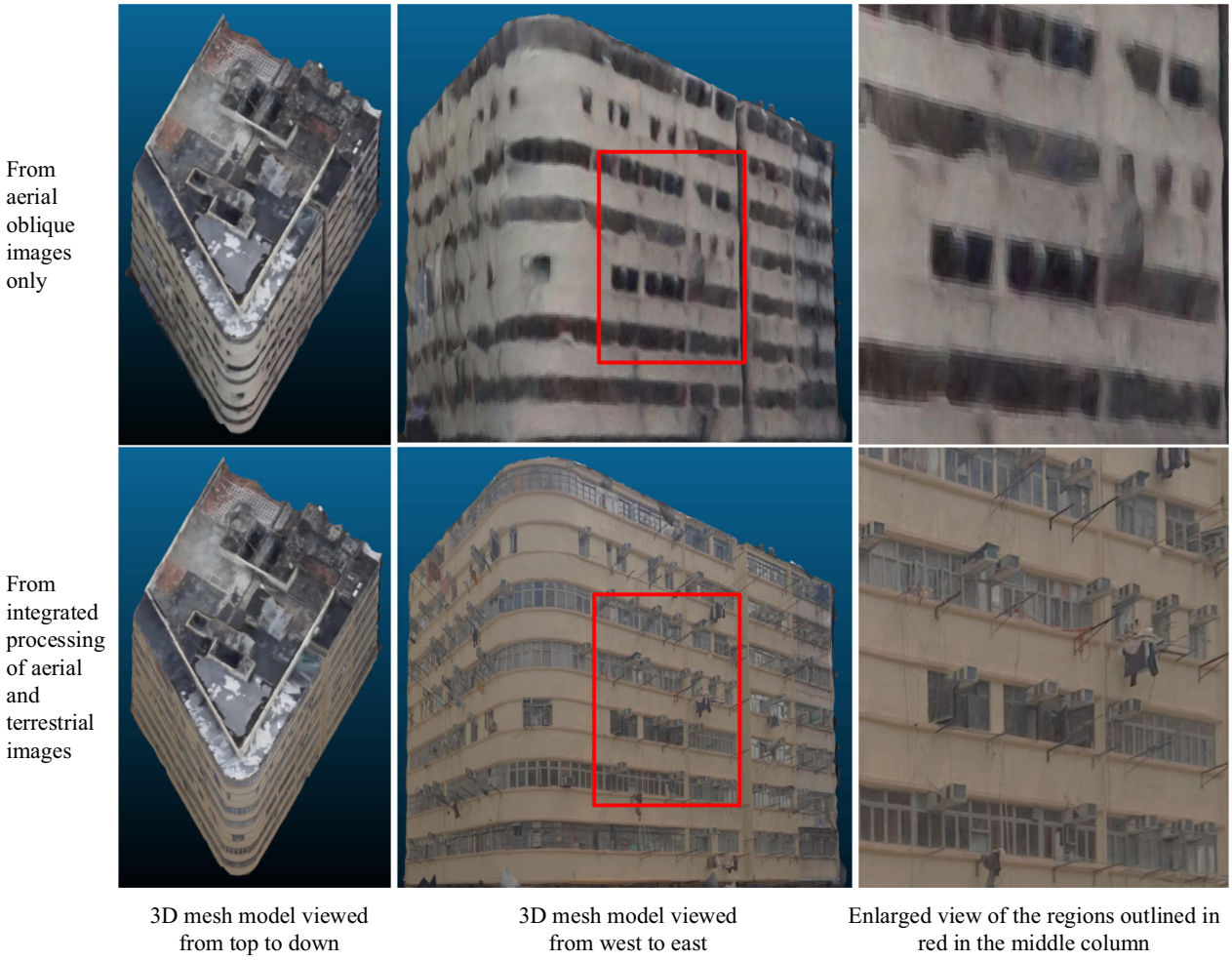

Fig. 11. 3D mesh models generated from the aerial oblique images only (first row) and from the integration of aerial oblique and terrestrial images (second row) for the Hong Kong dataset.

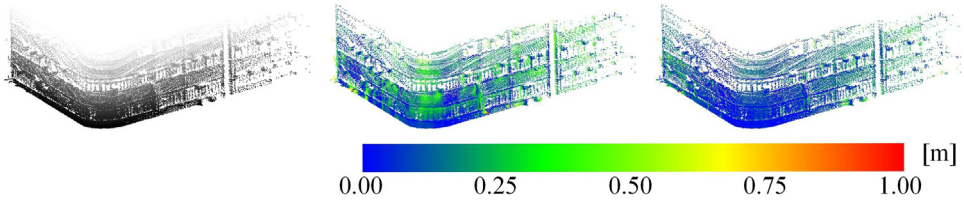

Fig. 12. CMD figures for the Hong Kong dataset. The left-hand, middle, and right-hand columns illustrate laser scanning point clouds shaded by height; unsigned CMDs for the 3D mesh model from aerial oblique images only; and unsigned CMDs for the integrated processing of aerial and terrestrial images, respectively.

Table 4
CMD statistics for the Hong Kong dataset.

From aerial oblique images only				From integrated processing of aerial and terrestrial images			
Mean (m)		Std. dev. (m)		Mean (m)		Std. dev. (m)	
Signed	Unsigned	Signed	Unsigned	Signed	Unsigned	Signed	Unsigned
−0.117	0.158	0.152	0.109	−0.045	0.096	0.113	0.075

5. Conclusions

In this paper, we propose an approach integrating aerial oblique imagery with terrestrial imagery to optimize 3D modeling in urban areas. We present a novel method of feature matching between images from the two platforms, a robust two-step combined bundle adjustment strategy, and effective techniques for geometric refinement and textural optimization. Two datasets are subjected to quantitative and qualitative analysis to evaluate the proposed approach. The results indicate that the proposed approach

effectively integrates aerial oblique with terrestrial imagery to yield 3D mesh models with greater geometric accuracy and superior texture.

The significance of the proposed approach lies in its potential to advance urban 3D mapping and modeling. Integrating aerial oblique imagery and terrestrial MMS data not only improves the performance of integrated 3D modeling but may also solve the problem of GNSS tracking loss in MMS in urban areas. The integrated processing of aerial oblique imagery with terrestrial MMS data generates an integrated aerial–ground environment. It offers a promising platform for optimizing 3D mapping and modeling at both city scale and street level, as the aerial and terrestrial datasets supplement each other. The integrated aerial–ground environment can be used for numerous 3D mapping and modelling tasks with improved performances.

Acknowledgements

This work was supported by a grant from the Research Grants Council of Hong Kong (Project No. PolyU 152211/18E) and grants from the Hong Kong Polytechnic University (Project No. 1-ZE24, Project No. 4-BCCQ). The authors gratefully acknowledge the provision of the datasets by ISPRS and EuroSDR, released in conjunction with the ISPRS Scientific Initiatives 2014 and 2015, led by ISPRS ICWG I/Vb.

References

Agarwal, S., Snavely, N., Seitz, S.M., Szeliski, R., 2010. Bundle adjustment in the large. In: Science, L.N.I.C. Berlin, Heidelberg: Springer, pp. 29–42.

Agisoft, 2018. Agisoft PhotoScan, <http://www.agisoft.com> (accessed 17 January, 2018).

Airbus Intelligence, 2018. Pixel Factory The power of an industrial solution in your hands <http://www.intelligence-airbusds.com/en/161-pixel-factory> (accessed 17 January, 2018).

Alcantarilla, P., Nuevo, J., Bartoli, A., 2013. Fast Explicit Diffusion for Accelerated Features in Nonlinear Scale Spaces. In: Proc. British Machine Vision Conference, pp. 13.1-13.11.

Alcantarilla, P.F., Bartoli, A., Davison, A.J., 2012. KAZE features. In: Proc. European Conference on Computer Vision, Springer, pp. 214–227.

Balletti, C., Guerra, F., Scocca, V., Gottardi, C., 2015. 3D integrated methodologies for the documentation and the virtual reconstruction of an archaeological site. ISPRS—International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XL-5/W4, pp. 215–222.

Bay, H., Ess, A., Tuytelaars, T., Van Gool, L., 2008. Speeded-up robust features (SURF). Comput. Vision Image Understand. 110 (3), 346–359.

Bentley, 2018. ContextCapture Create 3D models from simple photographs, <https://www.bentley.com/en/products/brands/contextcapture> (accessed 17 January, 2018).

Böhm, J., Haala, N., 2005. Efficient integration of aerial and terrestrial laser data for virtual city modeling using lasermaps. Proceedings of the ISPRS Workshop "Laser scanning 2005", Enschede, the Netherlands.

Bolognesi, M., Furini, A., Russo, V., Pellegrinelli, A., Russo, P., 2014. Accuracy of cultural heritage 3D models by RPAS and terrestrial photogrammetry. Int. Arch. Photogram., Remote Sens. Spatial Inform. Sci. 40 (5), 113.

Chen, N., Ni, N., Kapp, P., Chen, J., Xiao, A., Li, H., 2015. Structural analysis of the Hero Range in the Qaidam Basin, Northwestern China, using integrated UAV, terrestrial LiDAR, Landsat 8, and 3-D seismic data. IEEE J. Select. Top. Appl. Earth Observ. Remote Sens. 8 (9), 4581–4591.

Chu, C., Chiang, K., 2016. The performance of a tight INS/GNSS/Photogrammetric integration scheme for land based MMS applications in GNSS denied environments. The International Archives of Photogrammetry, Remote Sensing and Spatial Information sciences XLI-B1, pp. 551–557.

Fischler, M.A., Bolles, R.C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM 24 (6), 381–395.

Flener, C., Vaaja, M., Jaakkola, A., Krooks, A., Kaartinen, H., Kukko, A., Kasvi, E., Hyyppä, H., Hyyppä, J., Alho, P., 2013. Seamless mapping of river channels at high resolution using mobile LiDAR and UAV-photography. Remote Sens. 5 (12), 6382–6407.

Forlani, G., Roncella, R., Nardinocchi, C., 2015. Where is photogrammetry heading to? State of the art and trends. Rendiconti Lincei 26 (1), 85–96.

Förstner, W., 1986. A feature based correspondence algorithm for image matching. Int. Arch. Photogram., Remote Sens. Spat. Inform. Sci. 26 (3), 150–166.

Furukawa, Y., Ponce, J., 2010. Accurate, dense, and robust multiview stereopsis. IEEE Transact. Pattern Anal. Mach. Intell. 32 (8), 1362–1376.

Galliani, S., Lasinger, K., Schindler, K., 2015. Massively parallel multiview stereopsis by surface normal diffusion. Proceed. IEEE Int. Conf. Comput. Vision, 873–881.

Gerke, M., Nex, F., Jende, P., 2016a. Co-registration of terrestrial and UAV-based images – experimental results. Int. Arch. Photogram., Remote Sens. Spat. Inform. Sci., 11–18

Gerke, M., Nex, F., Remondino, F., Jacobsen, K., Kremer, J., Karel, W., Huf, H., Ostrowski, W., 2016b. Orientation of oblique airborne image sets-experiences from the ISPRS/EUROSDR benchmark on multi-platform photogrammetry. Int. Arch. Photogram., Remote Sens. Spat. Inform. Sci. 41, 185–191.

Girardeau-Montaut, D., 2015. Cloud Compare—3D Point Cloud and Mesh Processing Software. Open Source Project, <http://www.danielgm.net/cc> (accessed 17 January, 2018).

Gruen, A., Huang, X., Qin, R., Du, T., Fang, W., Boavida, J., Oliveira, A., 2013. Joint processing of UAV imagery and terrestrial mobile mapping system data for very high resolution city modeling XL-1/W2 Int. Arch. Photogram. Remote Sens. Spat. Inform. Sci., 175–182

Haala, N., Kada, M., 2010. An update on automatic 3D building reconstruction. ISPRS J. Photogram. Remote Sens. 65 (6), 570–580.

Harris, C., Stephens, M., 1988. A combined corner and edge detector. In: Proceedings of the Alvey Vision Conference, Manchester, UK, pp. 147–151.

Hirschmüller, H., 2008. Stereo processing by semiglobal matching and mutual information. IEEE Transact. Pattern Anal. Mach. Intell. 30 (2), 328–341.

Hu, H., Chen, C., Wu, B., Yang, X., Zhu, Q., Ding, Y., 2016. Texture-aware dense image matching using ternary census transform. ISPRS annals of photogrammetry. Remote Sens. Spat. Inform. Sci., 59–66

Hu, H., Zhu, Q., Du, Z., Zhang, Y., Ding, Y., 2015. Reliable spatial relationship constrained feature point matching of oblique aerial images. Photogram. Eng. Remote Sens. 81 (1), 49–58.

Jende, P., Hussnain, Z., Peter, M., Oude Elberink, S., Gerke, M., Vosselman, G., 2016a. Low-level tie feature extraction of mobile mapping data (MLS/images) and aerial imagery. Int. Arch. Photogram. Remote Sens. Spat. Inform. Sci., 19–26

Jende, P., Nex, F., Gerke, M., Vosselman, G., 2017. Fully automatic feature-based registration of mobile mapping and aerial nadir images for enabling the adjustment of mobile platform locations in GNSS-denied urban environments. ISPRS—Int. Arch. Photogram. Remote Sens. Spat. Inform. Sci., 317–323

Jende, P., Peter, M., Gerke, M., Vosselman, G., 2016b. Advanced tie feature matching for the registration of mobile mapping imaging data and aerial imagery. Int. Arch. Photogram. Remote Sens. Spat. Inform. Sci., 617–623

Ji, S., Shi, Y., Shan, J., Shao, X., Shi, Z., Yuan, X., Yang, P., Wu, W., Tang, H., Shibasaki, R., 2015. Particle filtering methods for georeferencing panoramic image sequence in complex urban scenes. ISPRS J. Photogram. Remote Sens. 105, 1–12.

Kazhdan, M., Hoppe, H., 2013. Screened poisson surface reconstruction. ACM Transact. Graph. 32 (3), 29.

Kedzierski, M., Fryskowska, A., 2014. Terrestrial and aerial laser scanning data integration using wavelet analysis for the purpose of 3D building modeling. Sensors 14 (7), 12070–12092.

Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vision 60 (2), 91–110.

Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., Van Gool, L., 2005. A comparison of affine region detectors. Int. J. Comput. Vision 65 (1–2), 43–72.

Moe, K., Toschi, I., Poli, D., Lago, F., Schreiner, C., Legat, K., Remondino, F., 2016. Changing the production pipeline-use of oblique aerial cameras for mapping purposes. Int. Arch. Photogram. Remote Sens. Spat. Inform. Sci. 41.

Moisan, L., Moulon, P., Monasse, P., 2012. Automatic homographic registration of a pair of images, with a contrario elimination of outliers. Image Process. On Line 2, 56–73.

Morel, J., Yu, G., 2009. ASIFT: A new framework for fully affine invariant image comparison. SIAM J. Imaging Sci. 2 (2), 438–469.

Moulon, P., Monasse, P., Marlet, R., 2013. Global fusion of relative motions for robust, accurate and scalable structure from motion. Proceed. IEEE Int. Conf. Comput. Vision, 3248–3255.

Murgoitio, J., Shrestha, R., Glenn, N., Spaete, L., 2014. Airborne LiDAR and terrestrial laser scanning derived vegetation obstruction factors for visibility models. Transact. GIS 18 (1), 147–160.

Nex, F., Gerke, M., 2014. Photogrammetric DSM denoising. Int. Arch. Photogramm., Remote Sens. Spat. Inform. Sci. 40 (3), 231.

Nex, F., Gerke, M., Remondino, F., Przybilla, H.J., Bäumker, M., Zurhorst, A., 2015. ISPRS benchmark for multi-platform photogrammetry. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences II-3/W4.

Nocerino, E., Menna, F., Remondino, F., Saleri, R., 2013. Accuracy and block deformation analysis in automatic UAV and terrestrial photogrammetry—Lesson learnt. ISPRS Annals Photogram., Remote Sens. Spat. Inform. Sci. 2, 5.

Ouédraogo, M.M., Degré, A., Debouche, C., Lisein, J., 2014. The evaluation of unmanned aerial system-based photogrammetry and terrestrial laser scanning to generate DEMs of agricultural watersheds. Geomorphology 214, 339–355.

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. Photogram. Eng. Remote Sens. 76 (8), 923–934.

Rublee, E., Rabaud, V., Konolige, K., Bradski, G., 2011. ORB: An efficient alternative to SIFT or SURF. In: Proc. Computer Vision (ICCV), 2011 IEEE international conference on, IEEE, pp. 2564-2571.

Rumpler, M., Tscharf, A., Mostegel, C., Daftry, S., Hoppe, C., Prettenthaler, R., Fraundorfer, F., Mayer, G., Bischof, H., 2017. Evaluations on multi-scale camera networks for precise and geo-accurate reconstructions from aerial and terrestrial images with user guidance. Comput. Vision Image Understand. 157, 255–273.

Schonberger, J.L., Frahm, J., 2016. Structure-from-motion revisited. Proceed. IEEE Conf. Comput. Vision Pattern Recogn., 4104–4113

Snavely, N., Seitz, S.M., Szeliski, R., 2008. Skeletal graphs for efficient structure from motion. Proceedings of the Conference on Computer Vision and Pattern Recognition, p. 2.

Spagnuolo, L.M., 2014. State of the art in surface reconstruction from point clouds. Eurographics Star Rep. 1 (1).

Tan, Y., Kwoh, L.K., Ong, S.H., 2008. Large scale texture mapping of building facades. Int. Arch. Photogram. Remote Sens. Spat. Inform. Sci. 37 (2).

Tola, E., Lepetit, V., Fua, P., 2010. Daisy: an efficient dense descriptor applied to wide-baseline stereo. IEEE Transact. Pattern Anal. Mach. Intell. 32 (5), 815–830.

Toldo, R., Gherardi, R., Farenzena, M., Fusiello, A., 2015. Hierarchical structure-and-motion recovery from uncalibrated images. Comput. Vision Image Understand. 140, 127–143.

Tong, X., Liu, X., Chen, P., Liu, S., Luan, K., Li, L., Liu, S., Liu, X., Xie, H., Jin, Y., 2015. Integration of UAV-based photogrammetry and terrestrial laser scanning for the three-dimensional mapping and monitoring of open-pit mine areas. Remote Sens. 7 (6), 6635–6662.

Toschi, I., Ramos, M.M., Nocerino, E., Menna, F., Remondino, F., Moe, K., Poli, D., Legat, K., Fassi, F., 2017. Oblique photogrammetry supporting 3d urban reconstruction of complex scenarios. ISPRS: Int. Arch. Photogram. Remote Sens. Spat. Inform. Sci., XLII-1/W1, pp. 519–526.

Triggs, B., McLauchlan, P.F., Hartley, R.I., Fitzgibbon, A.W., 2000. Bundle Adjustment: A modern Synthesis. Springer, Berlin, Heidelberg, pp. 153–177.

Von Hansen, W., Gross, H., Thoennessen, U., 2008. Line-based registration of terrestrial and airborne LIDAR data. Int. Arch. Photogram., Remote Sens. Spat. Inform. Sci. 37, 161–166.

Waechter, M., Moehrle, N., Goesele, M., 2014. Let there be color! Large-scale texturing of 3D reconstructions. Proceedings of European Conference on Computer Vision, Springer, pp. 836–850.

Wu, B., Guo, J., Hu, H., Li, Z., Chen, Y., 2013. Co-registration of lunar topographic models Derived from Chang'E-1, SELENE, and LRO laser altimeter data based on a novel surface matching method. Earth Planet. Sci. Lett. 364, 68–84.

Wu, C., 2007. SiftGPU: A GPU implementation of scale invariant feature transform (SIFT). URL <https://github.com/pitzer/SiftGPU> (accessed 15th September, 2017).

Xie, L., Hu, H., Wang, J., Zhu, Q., Chen, M., 2016. An asymmetric re-weighting method for the precision combined bundle adjustment of aerial oblique images. ISPRS J. Photogram. Remote Sens. 117, 92–107.

Zhou, K., Synder, J., Guo, B., Shum, H., 2004. Iso-charts: stretch-driven mesh parameterization using spectral analysis. In: Proceedings of the 2004 Eurographics/ACM SIGGRAPH symposium on geometry processing, ACM, pp. 45–54.

Zhu, Q., Zhang, Y., Wu, B., Zhang, Y., 2010. Multiple close-range image matching based on a self-adaptive triangle constraint. Photogram. Rec. 25 (132), 437–453.