Extracting 3D Urban Models from Oblique Aerial Images

Norbert Haala, Mathias Rothermel, Stefan Cavegn
Institute for Photogrammetry
Stuttgart University
Stuttgart, Germany
Forname.Lastname@ifp.uni-stuttgart.de

***Abstract*—Oblique airborne cameras are increasingly used for area covering image collection of building facades in urban environments. Until recently, these images were mainly used to improve the visual appearance of relatively simple 3D building models by providing suitable facade texture. Meanwhile, oblique airborne images are also used to generate dense 3D point clouds using state-of-the-art pixel-based multi-stereo image matching. Subsequently, these point clouds can then enhance the amount of geometric and semantic detail of 3D urban models especially for the depicted building facades. However, occlusions and large view-point changes are especially demanding during dense image matching can be very challenging for oblique imagery in complex urban environment. As described in the paper, our matching pipeline tackles these problems by a coarse-to-fine modification of the SGM method. Furthermore the raw 3D point clouds are efficiently analyzed and filtered in subsequent steps, thus 3D data capture from oblique airborne imagery while aiming at the extraction of 3D urban models becomes feasible.**

I. Introduction

The automatic reconstruction of urban 3D models from remote sensing data has been an important research topic for almost two decades [1]. Airborne or spaceborne images in principle provide all information required to solve this task. However, automatic image understanding is still hard to realize. Thus, existing reconstruction systems frequently use 3D point clouds or 2.5D Digital Elevation Models (DEM) as alternative input. By these means, the interpretation task can be restricted to explicit geometric information, which helps to facilitate the development of automatic tools for 3D building reconstruction [2]. Until recently, the required 3D point clouds were typically collected by airborne LiDAR. Meanwhile, state-of the-art multi-stereo image matching can alternatively provide 3D point clouds and DEM raster representations at a remarkable accuracy, density and reliability. If airborne imagery is captured at a sufficient overlap, the available redundancy enables recent matching software to generate 3D points at a density, which corresponds to the ground sampling distance (GSD) of the original images [3].

Automatic interpretation of these airborne point clouds usually results in 3D city models, which consist of a digital elevation model and relatively coarse building representation. Such geometry models are termed as LoD 2 (Level of Detail) as defined in the standard format CityGML [4]. This relates to building representations with detailed roof structures and planar facades. Usually, this amount of detail is sufficient for simulations or visualizations at small or medium scale. However, a number of applications require more complex so-called LoD 3 models with an explicit geometric representation of building facades. This for example includes elements like doors and windows. Since such features are difficult to extract from standard airborne data due to viewpoint restrictions, terrestrial platforms like mobile mapping systems are frequently used. Terrestrial LiDAR points as provided from such systems can then be used for the extraction of detailed building facades [5]. Complete city coverage by mobile mapping systems are already offered by a number of service providers. However, if compared to airborne platforms, data collection is still relatively expensive. Due to this reason, oblique aerial images are also used as alternative data source in urban environments. As an example, they are integrated in global map services, such as Google Maps, not only for visualization purposes but also for information extraction. An overview of the considerable number and variety of current oblique aerial cameras is given in [6]. Oblique images are taken with the optical axis deviating from the vertical, i.e. nadir direction. Since building facades and other vertical objects are well visible in these images, they are in principle very suitable for 3D data collection at such structures. However, dense matching in such scenarios is challenging with potential difficulties arising from large scale variations due to a higher depth of field, greater illumination changes and multiple occlusions. Thus, suitable tools for reliable data processing are required.

Within paper we discuss the use of our dense matching pipeline SURE for data capture at building facades from oblique aerial images. As presented in section II, it applies a variant of Semi-Global-Matching for the generation of 2D elevation data from highly overlapping, redundant imagery. Furthermore, a coarse to-fine modification of SGM is applied, which dynamically reduces search ranges for pixel correspondences. Exemplary results for oblique airborne images are then demonstrated in section III of the paper.

II. Dense Matching in Urban Aeras

Tools for DSM generation by automatic stereo image matching are available for more than two decades. However,

978-1-4799-6652-3/15/$31.00~ © 2015 IEEE

considerable software improvements still were achieved only recently. One example is the success of algorithms like the Semi-Global Matching (SGM) stereo method [7]. Our stereo algorithm is a coarse-to-fine modification of this SGM approach. There the problem of dense matching is stated as densely estimating the correspondences across a reference view as base image and a second frame as match image. In the first stage, a photo-consistency measure is computed and stored for each potential correspondence pair. The search range for correspondences is given by the 1D horizontal epipolar line and a pre-defined, constant disparity search range. Secondly, costs of photo-consistency are accumulated along eight image paths in order to force piecewise smoothness of the underlying surface. The set of minimal accumulated costs then yields a set of disparities corresponding to a strong local minimum of an energy function. In contrast to the original approach, we evaluate potential correspondences only in a reduced search range as visualized in Fig. 1. Within our approach the search range is limited for each pixel individually based on disparities available from disparity maps already computed on the next higher pyramid level at a lower resolution.

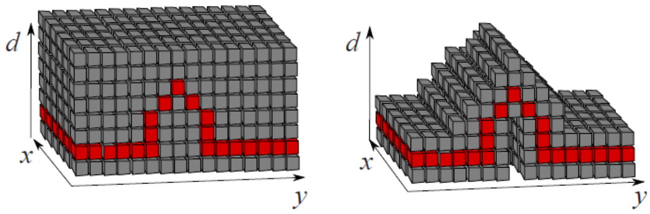

Fig. 1. Visualization of cost structures of classic SGM (left) and the dynamic solution (right). Red cubes represent cost for the true correspondences. Gray cubes mark the costs of potential correspondences.

The choice of a suitable range is crucial to assure that small details are still reconstructed completely even at large depth changes. In our algorithm, ranges are derived by analysis of the local surface structure [8]. On the coarsest pyramid level, correspondences are searched along the full range of epipolar lines, which is then adapted during the following levels. The pixel-wise SGM approach provides a dense point distribution, while the global approximation on paths enables a reasonable runtime even for large imagery. By these means, our matching generates a corresponding 3D point for each pixel of the central base image, which is matched against all overlapping neighbor images.

In order to combine the matching results from all captured images, a gridding is optionally realized. This results in a 2.5D Digital Surface Model (DSM). Within this process the median values of all points within a raster cell are used to compute the respective DSM elevation. Typically, a DSM grid width is selected, which corresponds to the GSD of the original imagery. Fig. 2 shows a textured DSM as output of this matching pipeline using standard airborne imagery. The example covers a densely built-up urban area from the city of München. The imagery is part of a recent benchmark on high density image matching for DSM computation [3]. It was captured by the large frame airborne camera DMC II 230 at a GSD of 10cm. To support the matching process in such an urban scenario, sufficient redundancy is provided by suitable image overlap. Thus, images were captured 80% in flight and 80% cross flight overlap. This provides up to fifteen images per object point.

Fig. 3 shows gives a more detailed view on our matching results for this test area. There the available imagery is additionally used to match texture against the almost vertical structure of the DSM raster at building facades. The DMC II 230 camera features a Field of View (FoV) of 49.9° cross track and 47.3° along track. If imagery from such a wide angle camera is available at sufficient overlap even standard nadir configurations can provide facade texture at resolution sufficient for a number of applications. However, the extraction of facade geometry by image matching presumes the availability of multiple views also for building facades. This is usually only available from oblique camera systems.

Fig. 2. Textured DSM using dense image matching for highly overlapping standard aerial imagery.

Fig. 3. Textured DSM using dense image matching for highly overlapping standard aerial imagery.

III. Meshed Surface Models from Oblique Aerial Images

On example of such a camera system is the medium format camera Leica RCD30 Oblique Penta. As shown in Fig. 4 the five camera heads are mounted with tilt angles of 35°. All camera heads were equipped with Leica NAG-D 50 mm lenses. The sensor has a size of 60 MP, a pixel size of 6 µm, a radiometric resolution of 14 bit, multi-directional motion compensation and a maximum rate of 1.8 seconds per image. For the investigated data set a flying height of around 520 m above ground and the calibrated focal length of 53 mm resulted in a GSD of 6 cm and an image scale of around 9800 in nadir view as well as a GSD of 6-13 cm for all four oblique views. Fig. 5 shows exemplary image sections from all camera heads. There the oblique views guarantee a complete visibility of building facades and other vertical objects as well as building footprints. The data set is also made available in a current benchmark, which is described in more detail in [9].

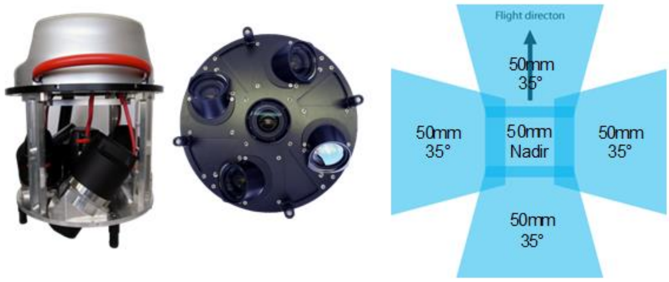

Fig. 4. Leica RCD30 Oblique Penta camera configuration

Fig. 5. Exemplary image section captured from all four oblique directions and the nadir view.

Surface geometry is commonly represented by a 2.5D DSM or 2.5D and 3D triangle meshes. If dense multi-image matching is applied, these structures have to be extracted from the densely matched 3D point clouds. In many applications, extensive simplification of this geometry is required to improve the interoperability. During processing of airborne nadir imagery typically 2.5D structure of reconstructed surfaces are assumed. As already discussed in section II, depth estimations and dense 3D point clouds can be fused using orthographic projection onto a plane. This DSM rasterization and cell-wise filtering provides one height value per cell. Thereby the number of matched 3D points is reduced significantly whereas redundancy is exploited to eliminate outliers and increase accuracy. While such a 2.5D assumption is sufficient for standard airborne nadir scenarios, the processing of oblique aerial images depicting urban scenes requires a more general approach. One key idea of our algorithm is to generate a 2D triangulation for each available depth map in the image sequence using a restricted quadtree (RQT). On the one hand this guarantees matching triangulations, on the other hand this creates the possibility to reduce points in the noise range not contributing to the geometry. Apart from the computation of reliable point coordinates the approach also generates robust surface normals. This is also of interest, since the latter can serve as input for subsequent meshing techniques. In our pipeline, the concept of RQT is applied to construct meshes from depth maps as generated previously by the multi-view matching pipeline. These meshes are subsequently fused based on the available normals and corresponding accuracy information [8].

Fig. 6. Facade image in oblique view and corresponding 3D point from dense image matching.

An exemplary result is given in Fig. 6. The top view shows a building in one of the available oblique aerial images, while the bottom view depicts the mesh patches extracted by our dense matching approach. For this data set the available overlap is 70% and 50% in flight and across flight, respectively for the nadir view. Despite the fact that the resulting redundancy is limited particularly for the facades, the amount of geometric detail as provided from the 3D point is still significant.

Fig. 7. 3D point cloud at building facades for complex urban area.

Fig. 7 gives an example for a 3D point cloud from dense multi-stereo image matching using oblique aerial imagery. FIG.As it is visible for this larger area, the reconstruction is complete while 3D structures at building facades such as balconies are successfully reconstructed.

IV. Conclusions and Outlook

Compared to standard nadir imagery, the use of oblique aerial imagery within a dense image matching pipeline significantly increases the range of disparities in stereo image pairs to be considered. This larger search space is potentiall demanding with respect to processing time and memory requirements. These problems could be solved successfully by employing a modified SGM method which determines the search space for every pixel individually using a pyramid based multi-resolution approach. We show that by this modification, ambiguities in the correspondence problem are reduced and completeness of the reconstruction can be increased. Typically image sets for the purpose of 3D reconstruction are collected with high image overlap. By exploiting the available high redundancy precision is improved, gross errors are eliminated and completeness is guaranteed. Dense image matching can provide high quality point clouds with depth information for almost every pixel at accuracies up to ground sampling distance.

In addition to the further improvement of our dense image matching pipeline SURE, current work mainly aims at the integration of 3D point clouds for geometric enhancement of existing building models. An exemplary scenario is given in Fig. 8. It shows a typical 3D building model as it is already available for a number of municipalities. This so-called LoD2 building model represents facades by planar polygons. In figure 8, point clouds from image matching are additionally overlaid. As described in [5] such point clouds can be used to further increase the amount of detail at building facades by extracting features like doors and windows.

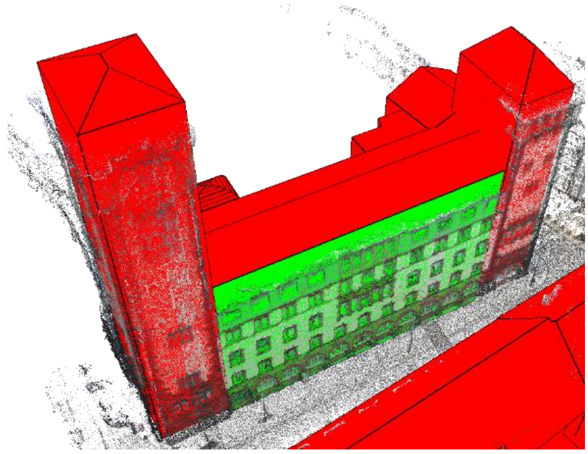

Fig. 8. Existing LoD2 building model with point cloud overlaid.

The oblique aerial imagery used within this paper are provided from the ISPRS/EuroSDR benchmark on High Density Image Matching for DSM Computation (Fig. 5 and 6) and the ISPRS/EuroSDR benchmark data set for multi-platform very high resolution photogrammetry (Fig. 7). This is gratefully acknowledged.

References

[1] N. Haala and M. Kada, An update on automatic 3D building reconstruction, ISPRS Journal of Photogrammetry and Remote Sensing (65), 2010, pp. 570–580.

[2] C. Brenner Building Extraction. Book chapter. In: Airborne and Terrestrial Laser Scanning. George Vosselman, Hans-Gerd Maas (eds.), Whittles Publishing, 2010

[3] N. Haala, The Landscape of Dense Image Matching Algorithms, Photogrammetric Week 2013, Wichmann Verlag, Berlin/Offenbach, 2013, pp. 271-284.

[4] G. Gröger an dL. Plümer, L., CityGML – Interoperable semantic 3D city models. In ISPRS Journal of Photogrammetry and Remote Sensing (71), 2012

[5] S. Becker, Generation and application of rules for quality dependent façade reconstruction. ISPRS Journal of Photogrammetry and Remote Sensing 64 (6), 2009, pp. 640–653

[6] M. Lemmens, Digital Oblique Aerial Cameras (1). A Survey of Features and Systems. GIM International, 28(4), 2014, pp. 20-25

[7] H. Hirschmüller, Stereo Processing by Semi-Global Matching and Mutual Information. IEEE Transactions on Pattern Analysis and Machine Intelligence, 30 (2), 2008, pp. 328 - 341.

[8] M. Rothermel, N. Haala, D. Fritsch, Generating oriented pointsets from redundant depth maps using restricted quadtrees, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., Volume XL-3, 2014

[9] S. Cavegn, N. Haala, S. Nebiker, M. Rothermel, P. Tutzauer, Benchmarking high density image matching for oblique airborne imagery, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., Volume XL-3, 2014