

# Image-based 3D Data Capture in Urban Scenarios

Norbert Haala, Mathias Rothermel, Stuttgart

## ABSTRACT

Presuming that airborne imagery is available at a sufficient overlap, state-of-the-art multi-stereo matching can generate DSM raster representations at an accuracy and resolution which corresponds to the ground sampling distance (GSD) of the original images. For such scenarios recent matching software exploits the resulting redundancy and derives surface representations at a remarkable accuracy and reliability. Typically, DSM rasters are generated as a standard result at a grid size corresponding to the average pixel footprint by a rather simple fusion of the 3D point clouds from multi-view matching. While such 2.5D models are suitable for a number of applications, high resolution data capture in complex urban environments requires the reconstruction and representation of 3D representations. This is especially true while aiming at the geometric reconstruction of objects with distinct structure like urban furniture or building façades. After a brief introduction in the state-of-the-art on high density image matching for DSM computation, this generation of filtered point clouds and 3D meshes within our multi-view reconstruction pipeline is discussed for both imagery aerial cameras and camera based mobile mapping systems. The results are especially beneficial while aiming at high quality visualisations and geometric data capture in urban scenarios.

## 1. INTRODUCTION

Research on automatic interpretation of remote sensing data to provide 3D representations of complex urban environments remains an ongoing research topic. Overviews on automatic 3D building reconstruction based on 3D point clouds or 2.5D Digital Surface Models (DSM) are e.g. given by (Haala & Kada, 2010) and (Brenner, 2010). Originally, the basic 3D information for such applications was provided by airborne LiDAR. Meanwhile, state-of-the-art multi-stereo matching is mainly used to generate 3D point clouds and DSM raster representations to provide the required geometric information. Especially if airborne imagery is available at a sufficient overlap, recent matching software exploits this redundancy to derive 3D points at a remarkable accuracy and reliability. The potential of such pixel-wise multi-stereo image matching for automatic photogrammetric 3D data capture is for example demonstrated by the joint ISPRS/EuroSDR project “Benchmark on High Density Aerial Image Matching” (Haala, 2014). State-of-the-art commercial matching tools generate DSM raster data sets at an accuracy and resolution which corresponds to the ground sampling distance (GSD) of the original images. As briefly summarized in section 2, this also holds true for rather complex urban environments. If these results are used for the generation of 3D city models as e.g. described by (Lafarge & Mallet, 2012), the resulting building representations provide detailed roof structures but are restricted to planar façades. Such representations at the Level of Detail (LoD) 2 are sufficient for simulations or visualizations at small or medium scale. In contrast, a number of applications require explicit information for the building façades. This presumes the LoD3 including elements like doors and windows. However, such features are difficult to extract from standard airborne data due to viewpoint restrictions of nadir imagery.

This is one reason why oblique aerial cameras are becoming an important alternative for image collection and information extraction in urban environments. Since building façades and other vertical objects are well visible, dense matching of oblique images provides 3D point clouds also for such structures. However, dense matching in such scenarios is frequently aggravated. Potential difficulties arise from large scale variations due to a higher depth of field, greater illumination changes and multiple occlusions. Even more important, compared to standard 2.5D processing, which is sufficient for DSM raster generation from nadir imagery, geometric processing is now required in true 3D space. As discussed in section 3, this includes matching steps like filtering and meshing of the generated 3D

points. Furthermore, surface representation has to be adapted by adding the capability to encode 3D structure.

While oblique airborne images are very suitable for area covering 3D data capture in urban scenarios, highest resolution and further increase in the available amount of detail presumes the use of terrestrial platforms as provided by street-level mobile mapping. As discussed in section 4, advanced algorithms for dense stereo image matching alternatively allow for point cloud generation both from oblique aerial imagery as well as from camera-based mobile mapping systems. These results can then be used for automatic reconstruction of detailed building models (Tutzauer & Haala, 2015).

## 2. URBAN DSM GENERATION FROM AIRBORNE NADIR IMAGERY

Recent innovations in matching algorithms considerably improved the quality of elevation data from aerial images. Pixel-wise multi-stereo matching can generate 3D point clouds and Digital Surface Models (DSM) at a resolution, which corresponds to the ground sampling distance GSD of the original images. As an example, semi-global matching introduced by (Hirschmüller, 2008) provides a well performing solution, which is currently implemented in a number of software tools for image based DSM generation. The state-of-the-art of dense image matching (DIM) solutions provided from research institutes and photogrammetric software vendors is e.g. documented by the benchmark on high density image matching for DSM computation (Haala, 2014) which compares results from 10+ DIM solutions.

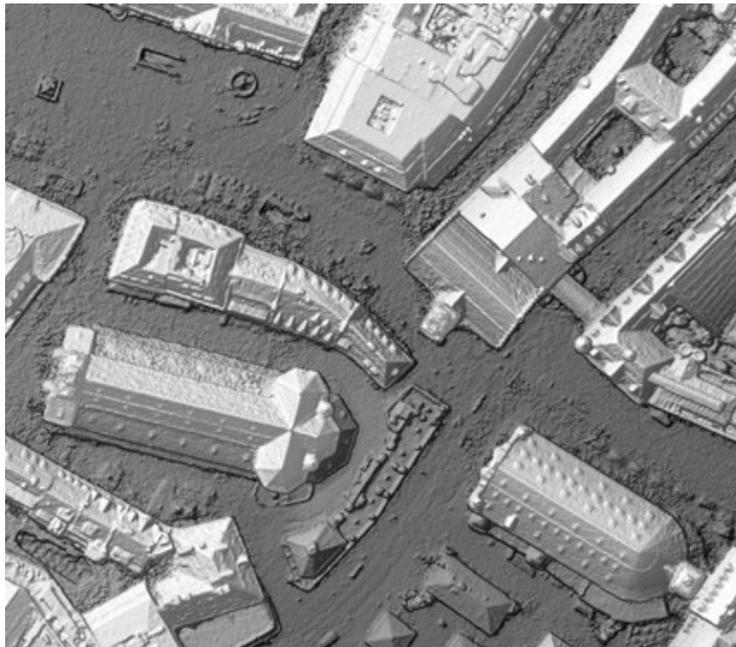


Fig. 1: Shaded DSM from nadir aerial imagery for test area München.

An exemplary result from this benchmark for a complex built-up area is depicted in Fig. 1. It shows a small subset of the raster DSM to be generated for the city of München. For this purpose imagery captured by the large frame airborne camera DMC II 230 at a GSD of 10cm had to be used. The matching process for this urban scenario is supported by a rather large image overlap to guarantee sufficient redundancy. The available 80% in flight and 80% cross flight lap results in up to fifteen images per object point. This helps to detect erroneous matches and supports the generation of DSM at vertical accuracies close to the sub-pixel level.

## 2.1. Pixel-wise multi-view stereo matching

As an example, the SGM-based approach implemented in our software SURE (Rothermel et al. 2012) benefits from this redundancy by solving the correspondence problem for multiple image pairs. Ideally, this leads to one 3D coordinate or depth per pixel for each image, which is then represented by a so-called depth map generated for each image frame. High redundant image sets with large overlap usually enable a repeated reconstruction of the same surface point within the image sequence. This leads to a significant amount of redundant observations of varying precision and reliability. The reason for these variances in reconstruction quality of depths and thus of 3D point coordinates across multiple stereo models are manifold. They result from differences in image scale as well as differences in the number of observations for depth reconstruction available during image matching and forward intersection. Further reasons are variances of ray intersection angles and of course matching errors across multiple views. Such matching errors can for example occur due to image blur, fronto parallel effects, or pixel locking effects.

## 2.2. Redundant 3D point computation from multiple-matches

In order to reduce the generated data while removing erroneous depths by exploiting the available redundancy, an adequate fusion strategy is required. This fusion is exemplarily demonstrated in Fig. 2 which depicts results from the evaluation of a UAV data set (Rothermel et al., 2014a).

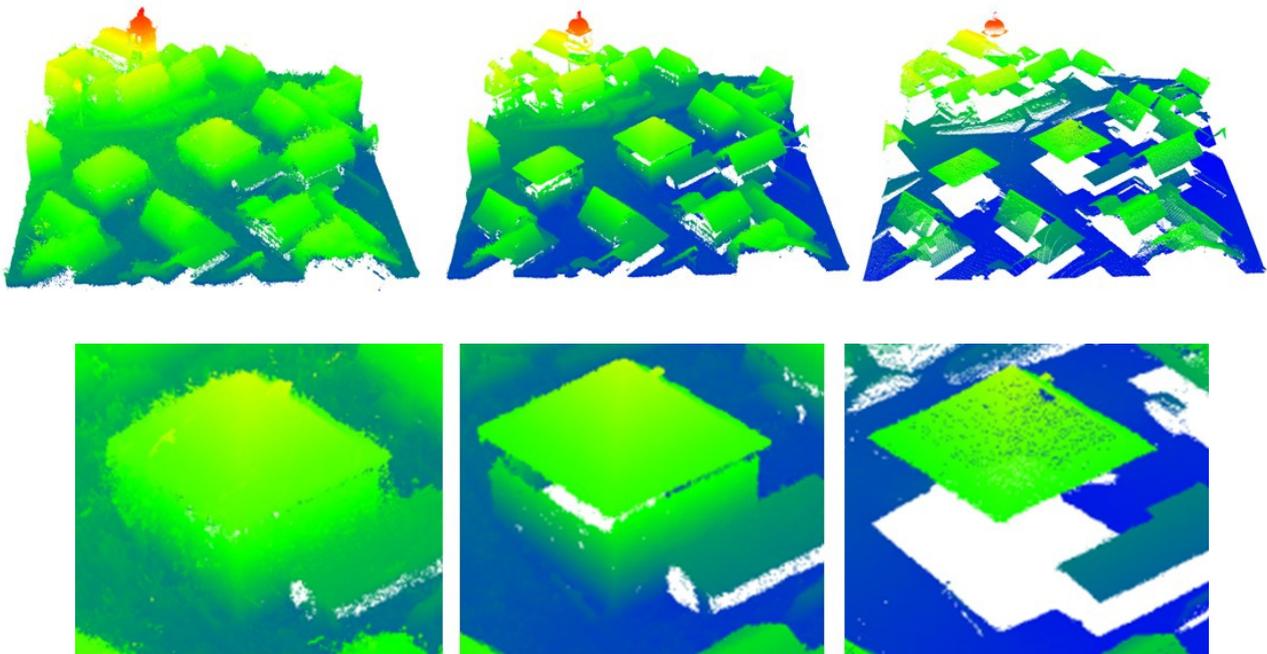


Fig. 2: Filtering of 3D points generated from stereo matching for larger scene (top) and detailed view (bottom) with point cloud from stereo matches (left), reduced points from multi-view constraints (middle) and gridded points (right).

The implemented pixel-wise SGM approach matches each pixel of the central base image against all overlapping neighbor images. Thus, multiple disparity maps are generated by the matching process. For each pixel of the base image they provide a parallax, which links it to the corresponding pixels of the respective match image. Each stereo match can thus be used to generate one corresponding 3D point by spatial intersection. Due to the availability of multiple stereo matches, each pixel of the base image can result in multiple 3D points. This results in a very dense point cloud depicted in the left images of Fig. 2. While the top row depicts an overview of the urban scene, the bottom row gives a more detailed view.

Point clouds derived from single stereo pairs are noisy due to matching errors and contain outliers as it is e.g. visible on the left images in Fig. 2. To eliminate this noise in the following step, the redundant per-pixel depth estimations for the respective match images are exploited. As mentioned, these multiple estimations result from the fact, that the implemented MVS approach matches a base or master image  $\mathbf{I}_b$  against  $N$  match or slave images  $\mathbf{I}_m$ . In order to fuse these multiple matches, redundant depth estimates as available for each pixel in the base image are combined while aiming at geometric consistency (Rothermel et al., 2014b). This consistency check helps to improve precision and to reduce outliers. Within this step, one 3D point is computed for each pixel of the central base image, while in principle each additional match against the multiple search images adds one image ray per point during the computed spatial intersection. Let  $\mathbf{x}_b$  be the normalized coordinates in the base image  $\mathbf{I}_b$  and  $\mathbf{x}_{m,i}$  the set of corresponding coordinates in the match images  $\mathbf{I}_m$  for  $i = 1 \dots N$ . First we use forward intersection between base image  $\mathbf{I}_b$  and match image  $\mathbf{I}_m$  as well as the respective normalized correspondences  $\mathbf{x}_b = (x_b, y_b)$  and  $\mathbf{x}_m = (x_m, y_m)$  to compute the depths  $d_i$  along  $\mathbf{x}_b$  by spatial intersection. Moreover, for each depth  $d_i$  an uncertainty interval  $\sigma_{l,i}$  in image space is propagated to an interval  $\sigma_{o,i}$  on  $\mathbf{x}_b$ . If the propagated uncertainty ranges overlap, the depths on the image ray defined on  $\mathbf{x}_b$  are considered consistent and used within the final depth computation. The final point coordinate is derived by minimizing the reprojection error across the incorporated views. For rectified imagery based on homographies (i.e. the stereo normal case) a direct solution to this problem exists, which is used to speed up computations. The result of this step is shown in the middle of Fig. 2. As it is visible, this process of correspondence linking considerably reduces the noise of the respective point cloud and thus improves the geometric quality of the object points. During processing typically each image of a block is also used as a base image. This is the reason why the point clouds shown in the middle of Fig. 2. still are very dense for that highly overlapping image block. Thus, while finally aiming at suitable point densities, which correspond to the GSD of the available imagery, further filter steps are necessary.

### 2.3. Filtering by gridding

Typically, a DSM raster is generated as the final outcome from geometric evaluation of standard airborne imagery. During structure computation from multiple-matches described in the preceding section, in principle one 3D point is generated for each pixel of the respective base images. For typical overlaps of aerial image blocks, which are frequently captured at an 80% in-flight lap, this still results in a considerably point density. To generate the aspired 2.5D DSM raster structure, a rather simple fusion of the 3D point clouds from multi-view matching can be realized. For this purpose, all 3D points from the preceding redundant computation from multiple-matches are assigned to a  $n \times m$  grid, which is defined parallel to the ground. Typically, the dimension of a single grid cell corresponds to the average pixel footprint. Within a first processing step, the number of average observations per cell  $n_a$  is computed. Then the final height value of each grid cell is derived as the median of the  $n_a$  largest z-components of the points assigned to each single cell. As shown, in the right picture of Fig. 2, the computation of one height value per cell significantly reduces the number of points while efficiently exploiting the available redundancy to eliminate outliers and increase precision. Moreover, this filter strategy efficiently removes façade points as desired for DSM and true-ortho production.

## 3. VISUALISATION OF TRIANGULATED DSM RASTER

Since the data set depicted in Fig. 2 was generated from UAV imagery with high overlap and varying look angles, a rather large number of 3D points were also reconstructed at the building façades. Typically the amount of captured 3D structure is limited if nadir views from standard aerial image

flights are used. In such scenarios the restricted look angles result in rather sparse reconstructions of 3D façade geometry. However, for visualization purposes rather appealing results can be obtained by mapping aerial image texture against the available 2.5D DSM data.

Since visualization pipelines are based on triangle meshes, these primitives have to be extracted from the reconstructed DSM raster as a first processing step. For this purpose multi-resolution triangulations are available in computer graphics since decades (Lindstrom et.al. 1996). Originally, these approaches aimed on the visualization of DTM data. However, if for visualization of so-called height fields in built-up areas high resolution DSM data are used, some modifications of the triangulation process are required. Since computational complexity of both, visualization and texture mapping directly depend on the number of triangles, it is desirable to construct meshes which solely consider elevation data contributing to the actual geometry and neglect data possessing elevation variances close to the noise level. An ideal framework for this task is given by restricted quad-trees (RQT) as proposed by Pajarola (1998), which are therefore integrated in our pipeline (Rothermel et al., 2014b). A nice property of RQTs is that generated meshes are matching (crack-free) triangulations. Furthermore, incorporated vertices are guaranteed to satisfy a certain error criteria. In our case this criteria assures that spatially neighboring elevations are only added to the mesh if their elevation variance is larger than the local noise level. Since it operates in a 2D space, the triangulation process is efficient compared to alternative 3D approaches.

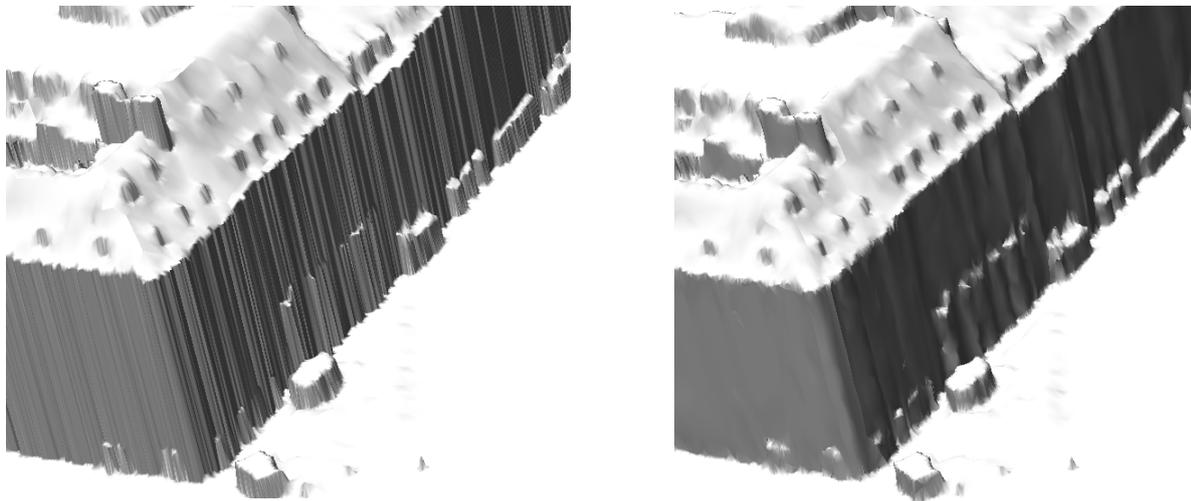


Fig. 3: Shaded visualisation of triangulated DSM raster before (left) and after (right) smoothing.

In areas of large depth variations, in particular at façades, triangles possessing very large side ratios are produced. Moreover, the orientations of these triangles are subject to large variances. This results from the fact that reconstructed roof edges typically are not represented by exactly straight lines in the reconstructed DSM. Since the normals of triangulated faces are required during shading, this directly limits the visual quality especially at building façades. This is demonstrated exemplarily in the left image of Fig. 3. Therefore, we correct suspicious faces by a mesh smoothing process. For this purpose we first identify badly shaped triangles analysing the triangles side ratio. These faces are then subdivided until a minimal side length of each triangle is equal or smaller than the GSD of the available imagery. The resulting triangles are then subject to a Laplacian smoothing and are fused again using a quadric error metrics to preserve geometry. The RQT framework offers the possibility to insert additional vertices. This enables tiled processing schemes with tiles possessing well defined geometry at border vertices. As a result, the approach scales well to large datasets and enables the generation of matching triangulations for country-wide projects. Fig. 3 demonstrates the results of this smoothing process based on a shaded visualization of the triangulated DSM raster before (left)

and after (right) smoothing. This example was created based on the test data set München already used in the example of Fig. 1.



Fig. 4. Texture mapping for triangulated DSM raster, visualisation of vertices (top) and faces (bottom).

A larger area of this München data set was also used for the example given in Fig. 4. It again depicts a meshed representation of the original raster DSM after the implemented smoothing process. While the top of Fig. 4 depicts the wireframe of the model to visualize the structure of the generated meshes, the bottom image shows the respective triangles filled with their corresponding texture as provided from the aerial images.

The data set München features imagery captured by the DMC II 230 camera, which features a Field of View (FoV) of  $49.9^\circ$  cross track and  $47.3^\circ$  along track. If imagery from such a wide angle camera is available at a sufficient overlap (e.g. 80%/80% for this example) even standard nadir configurations can provide façade texture at resolution sufficient for a number of applications. In general a triangle of the mesh is seen in multiple views arising the question of how to texture the face. In our pipeline we follow the approach proposed by Waechter et al. (2014) to pick an adequate, consistent texture. It has to be note that during our RQT-based triangulation, the geometry of the original DSM as generated from our dense multi-image-matching pipeline is not changed. Since only virtual façade points are modified, the meshed original elevation data is still valid. However, façade vertices are

only approximated and not a result of the reconstruction process itself, thus they remain planar as it is visible in the examples in Fig. 3 and Fig. 4. Since this can be acceptable for applications like visualisations, structural information on building façades is extremely beneficial for further interpretation. Even more important, representations based on a raster DSM result in artifacts for object parts violating the implicit 2.5D assumption. This is especially true for undercuts like they are present for the tree areas in Fig. 4.

#### 4. RECONSTRUCTION OF FAÇADE GEOMETRY

To overcome this problem, geometric reconstruction in 3D space is required. However, this presumes the availability of suitable image data, which can be provided either from oblique aerial cameras or terrestrial viewpoints.

##### 4.1. Evaluation of oblique aerial imagery

Especially in complex urban environments, oblique airborne imagery are more and more used as auxiliary source of information, which depict building façades and footprints. Since they are easy to interpret also for non-expert users they are frequently integrated for visualization in global map services, such as Google Maps. Furthermore, they can be used for 2.5D or 3D information extraction in applications like monitoring, urban area classification or administration services. In principle, oblique images are also very suitable for image matching while aiming at the generation of dense 3D point clouds in the context of 3D city modelling. However, applying DIM algorithms to oblique imagery introduces some major new challenges to the processing pipeline. In addition to greater illumination changes these include multiple occlusions as well as large scale variations due to a higher depth of field. We overcome the significant increase in disparity search space while avoiding higher requirements concerning processing time and memory by employing a modified SGM method called tSGM (Rothermel et al., 2012). This method determines the search space for every pixel individually by a pyramid based multi-resolution approach. This is especially beneficial for point cloud generation from oblique imagery. However, while aiming at the generation of 3D surface representations at building façades an adaption of the original filtering process already discussed in section 2.2 is required.

Our multi-stereo-matching produces a depth image for the respective base image by linking correspondences across views. Since all (base) images overlap also resultant depth maps do. Such depth map based DIM algorithms scale well by design and thus allow for the reconstruction of large data sets. Nevertheless, the non-trivial problem of an adequate fusion algorithm to extract a consistent model arises. When fusing redundant depths from different maps, a key concern is to pick the most precise and robust observation to represent the surface. Precision varies due to variances in ray intersection angles and image scale, the number of redundant observations, as well as the quality from stereo matching itself. The approach for 2.5D fusion based on median filtering described in section 2.3 yields acceptable results in terms of precision and blunders. However, it benefits from the implicit assumption for potential surface normals by the DSM grid plane. This plane defines the spatial direction for which filtering is carried out. While this is reasonable for 2.5D DSM processing, normals of the surfaces largely vary for 3D data fusion scenarios. Thus, our approach first estimates a surface representation based on triangles comprising robust normals. Based on this initial surface, median filtering is applied. More precisely, we project the current model into a depth map to be added. The depth map is triangulated using a RQT, which adapts triangle sizes to the local noise levels. If depth map and model faces are redundant, the more consistent triangle is chosen while the remaining one is discarded. The consistency or goodness of a face is determined by combining precision and reliability. Precision is approximated by the pixel footprint while reliability is approximated by the

density of the reconstruction in the local neighborhood of the respective face. By sequentially adding depth maps we construct a first representation of the surface based on face patches. In a second step, redundant depth estimates for each oriented point are collected by projection of the point coordinates to the depth images. These so obtained depths imply scalar offsets along the point's normal. These offsets are median filtered and the result defines the final point coordinates. Furthermore, visibility checks are used for blunder detection. Based on the refined vertex positions and normals, a consistent mesh can be extracted. For this purpose Poisson reconstructions as e.g. described by Kazhdan & Hoppe (2013) is applied.



Fig. 5: Façade image in oblique view (left) and corresponding 3D point cloud (right) from dense image matching.

A result of this 3D point generation process from oblique aerial imagery at a building façade is given in Fig. 5. Fig. 5 (left) shows an exemplary section of the used imagery captured by the oblique camera system Leica RCD30 Oblique Penta. This medium format camera features five camera heads, which are mounted with tilt angles of  $35^\circ$ . The data set is part of the current ISPRS/EuroSDR Benchmark on High Density Aerial Image Matching described in more detail in (Cavegn et al. 2014). The imagery was captured for an area in the city of Zürich at an approximate image overlap of 70% in flight and 50% across flight direction in the nadir view. The flying height of around 520 m above ground and

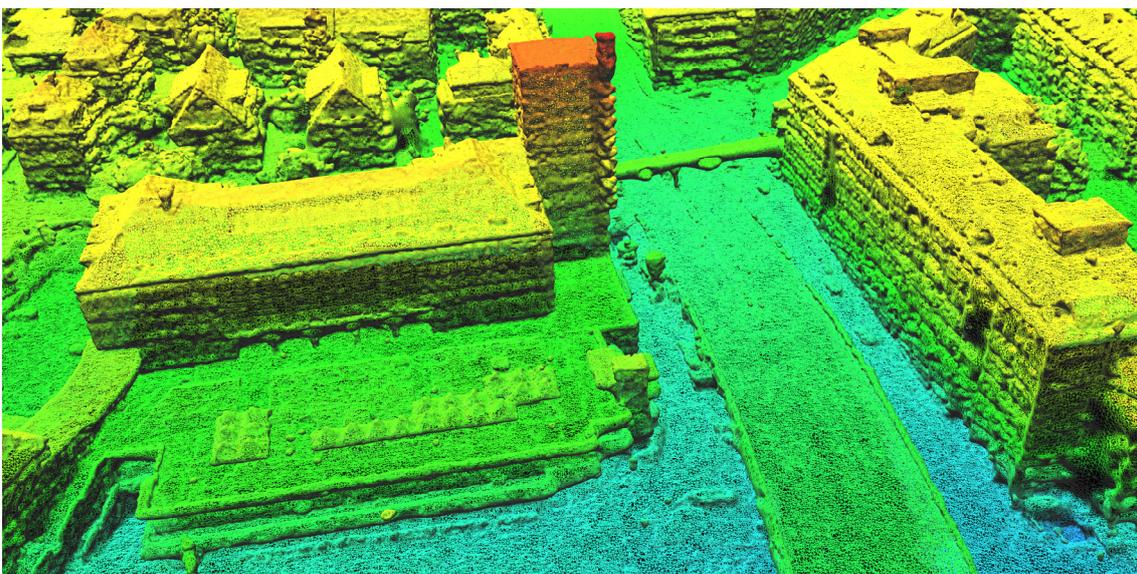


Fig. 6: Complex urban area represented by 3D meshes from multi-stereo matching of oblique aerial images.



Fig. 7: Texture mapped visualisation of 3D reconstruction depicted in Fig. 6.

the calibrated focal length of 53 mm resulted in a GSD of 6 cm in nadir view as well as a GSD of 6-13 cm for all four oblique views. The 3D point cloud from image matching in Fig. 5 (right) demonstrates the amount of structure, which can be extracted at a building façade for this data set. Fig. 6 shows the result from 3D meshing for a larger part of that test area. Please note that in contrast to the meshing results based on a DSM as depicted in Fig. 4, the 3D meshing in Fig. 6 actually represents 3D structure e.g. at building façades from balconies or other protrusion. For better interpretation of the results Fig. 7 shows the texture mapped visualisation of the reconstruction already given in Fig. 6. Despite the successful reconstruction of true 3D geometry, the amount of detail for this test is of course limited by the resolution of the oblique imagery and the available image overlap. Thus, despite the considerable progress in 3D data capture from airborne imagery, for highest resolution and accuracy of 3D object reconstruction in complex urban environments, still data capture from terrestrial viewpoints is required.

#### 4.2. Terrestrial point clouds from mobile mapping imagery

Area covering terrestrial data collection can be realized efficiently by street-level mobile mapping. Frequently these systems use LiDAR sensors to provide 3D point clouds for applications like road management or rural and urban structuring. Meanwhile, the progress in dense multi-stereo image matching enables 3D data capture using camera-based systems as a suitable alternative. In order to investigate the feasibility of our matching pipeline for evaluation of such imagery, test data was captured using the mobile mapping system of the Institute of Geomatics Engineering (IVGI), University of Applied Sciences and Arts Northwestern Switzerland (FHNW). Fig. 8 shows the platform and sensor configuration. The system features several industrial stereo cameras with CCD sensors and a radiometric resolution of 12 bit as well as a Ladybug5 panorama camera which are all mounted on a rigid platform. The forward looking stereo cameras used in the test have a resolution of 4008 x 2672 pixels at a pixel size of 9  $\mu\text{m}$ , a focal length of 21 mm and resulting fields-of-view of approximately 80° in horizontal and 60° in vertical direction. The test data were captured by the iNovitas AG during a complete survey of the city-state of Basel.

An exemplary image of the evaluated sequence is shown in Fig. 9 (left). It was captured at a busy junction in the city center of Basel. Due to the presence of complex buildings with large windows and smaller objects like the overhead wires of the tramway, the scene is rather challenging for the matching process. For the first investigations described in more detail by Cavegn et al. (2015) stereo



Fig. 8: Platform and sensor configuration of the IVGI mobile mapping system.



Fig. 9: Example image of investigated sequence (left) and point cloud from matching (right).

imagery captured by the two forward looking cameras was used. The base length for this stereo configuration also depicted in Fig. 8 is 0.905m. The image sequence used during multi-view stereo matching was captured at a frame rate of 5 fps.

While in principle, a number of image combinations can be used during matching, the exemplary point cloud in Fig. 9 (right) was generated by matching a base image against five neighbors. If the left camera provides the base image at  $t_1$ , two match images are captured from the same camera at the previous and following viewpoint,  $t_0$  and  $t_2$ , respectively. The remaining three match images are provided from the right cameras and are captured at  $t_0$ ,  $t_1$  and  $t_2$ . For each frame, a new base image is used, which results in a new depth image every 0.2 seconds. For the forward looking stereo combination the camera of course moves in the look direction. Together with the high frame rate, this results in a considerable overlap between the 3D point clouds generated for the respective depth images. To exploit this redundancy for stereo model selection and filtering an octree based approach is used (Wenzel et al., 2014). Despite the fact that only a part of the available imagery is integrated in the processing pipeline, the first results depicted in Fig. 9 (right) already demonstrate the potential of 3D data capture in urban areas from camera based mobile mapping systems.

## 5. CONCLUSION

Within the paper we presented the current state of 3D data capture in urban areas by dense multi-view-stereo for different scenarios using different camera configurations. DSM raster grids at the resolution of the captured imagery are already generated by a number of commercial software solutions. In contrast to such 2.5D raster representations typically generated from standard airborne

nadir imagery, high resolution reconstruction of complex 3D structures in urban environments presumes an object representation by 3D meshes. This was demonstrated exemplarily for the evaluation of oblique aerial imagery. Challenges for the matching process due to large scale variations and multiple occlusions could be overcome by the use of restricted quad-trees, which proved to be especially beneficial for filtering and meshing of the generated 3D point clouds. As demonstrated, façade geometry of the depicted building can be reconstructed successfully from oblique images. The amount of detail can be further increased using image sequences from mobile mapping systems. Since such systems frequently apply multiple cameras, a lot of stereo configurations are possible. In combination with the complexity of urban scenes as captured from terrestrial viewpoints, this further increases the demands on robustness and flexibility of the multi-view matching pipeline. Nevertheless, dense image matching still provides high quality point clouds with depth information for almost every pixel.

## 6. REFERENCES

- Brenner, C. (2010) Building Extraction. Book chapter. In: Airborne and Terrestrial Laser Scanning. George Vosselman, Hans-Gerd Maas (Eds.), Whittles Publishing.
- Cavegn, S., Haala, N., Nebiker, S., Rothermel M. & Zwölfer, T. (2015) Evaluation of Matching Strategies for Image Based Mobile Mapping, to appear in CMRT.
- Cavegn, S., Haala, N., Nebiker, S., Rothermel, M. & Tutzauer, P. (2014) Benchmarking high density image matching for oblique airborne imagery. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, Vol. XL-3.
- Haala, N. & Kada, M. (2010) An update on automatic 3D building reconstruction. *ISPRS Journal of Photogrammetry and Remote Sensing* (65), 2010, pp. 570-580.
- Haala, N. (2014) Dense Image Matching Final Report, EuroSDR Official Publication N°64, pp. 115-145, 2014.
- Haala, N. & Rothermel, M. (2012) Dense Multi-Stereo Matching for High Quality Digital Elevation Models. *Photogrammetrie – Fernerkundung – Geoinformation (PFG)*, 4, pp. 331-343.
- Hirschmüller, H. (2008) Stereo Processing by Semi-Global Matching and Mutual Information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30 (2), pp. 328-341.
- Kazhdan, M. & Hoppe, H. (2013) Screened Poisson surface reconstruction. *ACM Trans. Graph.* 32, 3, Article 29 (June 2013).
- Lafarge, F. & Mallet, C. (2012) Creating large-scale city models from 3D-point clouds: a robust approach with hybrid representation. *International Journal of Computer Vision*, 99 (1), pp. 69-85.
- Lindstrom, P., Koller, D., Ribarsky, W., Hodges, L., Faust, N. & Turner, G. (1996) Real-time, continuous level of detail rendering of height fields. In: *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques (SIGGRAPH '96)*.
- Pajarola, R. (1998) Large scale terrain visualization using the restricted quadtree triangulation. In: *Proc. IEEE Visualization'98*, pp. 19-26, 1998.

- Rothermel, M., Bulatov, D., Haala, N & Wenzel, K. (2014a) Fast and Robust Generation of Semantic Urban Terrain Models from UAV Video Streams. 22nd International Conference on Pattern Recognition DOI: 10.1109/ICPR.2014.112.
- Rothermel, M., Haala, N., Fritsch, D. (2014b) Generating oriented pointsets from redundant depth maps using restricted quadtrees. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, Vol. XL-3, 2014.
- Rothermel, M., Wenzel, K., Fritsch, D. & Haala, N. (2012) SURE: Photogrammetric Surface Reconstruction from Imagery. In: *Proceedings LC3D Workshop*, Berlin, Germany.
- Tutzauer, P. & Haala, N. (2015) Façade Reconstruction Using Geometric and Radiometric Point Cloud Information *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, Vol. XL-3/W2, pp. 247-252.
- Waechter, M., Moehrle, N. & Goesele, M. (2014) Let There Be Color! – Large-Scale Texturing of 3D Reconstructions. *ECCV 2014, Lecture Notes in Computer Science*, Vol. 8693, 2014, pp. 836-850.
- Wenzel, K., Haala, N. & Fritsch, D. (2014) Stereo model selection and point cloud filtering using an out-of-core octree. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, Vol. XL-3, 2014, *ISPRS Technical Commission III Symposium*, 5-7 September 2014, Zurich, Switzerland, pp. 373-380.