EVALUATION OF MATCHING STRATEGIES FOR IMAGE-BASED MOBILE MAPPING

S. Cavegn [a, b], N. Haala [b], S. Nebiker [a], M. Rothermel [c], T. Zwölfer [c]

[a] Institute of Geomatics Engineering, FHNW University of Applied Sciences and Arts Northwestern Switzerland, Muttenz, Switzerland - (stefan.cavegn, stephan.nebiker)@fhnw.ch
[b] Institute for Photogrammetry, University of Stuttgart, Germany - norbert.haala@ifp.uni-stuttgart.de
[c] nFrames GmbH, Stuttgart, Germany - (mathias.rothermel, thomas.zwoelfer)@nframes.com

Commission III, WG III/4

KEY WORDS: Image, Matching, Sequence, Mobile Mapping, Stereo, Rectification

ABSTRACT:

The paper presents the implementation of a dense multi-view stereo matching pipeline for the evaluation of image sequences from a camera-based mobile mapping system. For this purpose the software system SURE is taken as a basis. Originally this system was developed to provide 3D point clouds or DEM from standard airborne and terrestrial image blocks. Since mobile mapping scenarios typically include stereo configurations with camera motion predominantly in viewing direction, processing steps like image rectification and structure computation of the existing processing pipeline had to be adapted. The presented investigations are based on imagery captured by the mobile mapping system of the Institute of Geomatics Engineering in the city center of Basel, Switzerland. For evaluation, reference point clouds from terrestrial laser scanning are used. Our first results already demonstrate a considerable increase in reliability and completeness of both depth maps and point clouds as result of the matching process.

1. INTRODUCTION

Street-level mobile mapping has evolved into a highly efficient mapping technology. Frequently these systems use LiDAR sensors to provide area-covering 3D point clouds. However, advanced algorithms for dense stereo image matching alternatively allow for efficient and accurate 3D data capture using camera-based systems. The highly redundant metric imagery data can be used for a high-fidelity scene representation with an unparalleled level of detail and for the extraction of dense 3D information at the resolution of the available imagery. Furthermore, the imagery can also be used for improving the direct georeferencing solution using integrated sensor orientation approaches (Eugster et al., 2012; Nebiker et al., 2012).

Dense stereo image matching, as for example realised in our stereo matching software SURE (Rothermel et al., 2012), has proven to perform well in aerial imaging, both for standard nadir (Haala, 2014) and oblique image configurations (Cavegn et al., 2014), as well as in close range terrestrial scenarios (Wenzel et al., 2014). This was our motivation to apply this approach for dense matching of mobile mapping imagery. While other approaches aiming at 3D scene reconstruction from street-level imagery (Pollefeys et al., 2008; Gallup, 2011) are based on plane-sweeping, the pipeline implemented in SURE is based on multiple matching of stereo image pairs. This presumes an image rectification process as an important pre-processing step, which had to be adapted for imagery captured with motion predominantly in viewing direction. Standard rectification approaches fail for epipoles located inside the stereo image pairs. This situation is unlikely for airborne imagery, where the viewing direction is roughly perpendicular to the platform motion, but typical in case of mobile mapping systems with forward or backward looking cameras. To overcome this problem, the polar rectification technique was additionally implemented in the SURE framework. By these means, stereo matching is no longer restricted to stereo images at common timeframes but can easily be extended to an in-sequence matching of images from identical cameras. This considerably extends potential configurations for multiple image pairs, so that an increase in reliability, completeness and accuracy of both depth maps and point clouds as result of the matching process can be expected.

In this paper we introduce the new multi-image matching capabilities for in-sequence and stereo configurations and subsequently evaluate them with imagery captured by a mobile mapping system described in section 2. Section 3 presents the complete processing pipeline including accurate georeferencing, image rectification and point filtering incorporating the dense stereo matching process as the main contribution of the paper. Furthermore, the selection of different image configurations for multi-view stereo is described and the investigated test scenarios in image and object space are presented in section 4.

2. MOBILE MAPPING PLATFORM AND TEST SCENARIO

For our investigations, the mobile mapping system of the Institute of Geomatics Engineering (IVGI), University of Applied Sciences and Arts Northwestern Switzerland (FHNW) was used. Test data was collected at a busy junction in the city center of Basel, Switzerland.

2.1 Mobile mapping system

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 (see Figure 1). The forward looking stereo cameras have a resolution of 4008 x 2672 pixels at a pixel size of 9 µm, a focal length of 21 mm and resulting fields-of-view of approx. 80° in horizontal and 60° in vertical direction (Burkhard et al., 2012). While the forward stereo base is 0.905 m in this case, stereo cameras pointing back-right are separated by 0.779 m and stereo cameras looking left by 0.949 m (see Figure 2). Both stereo systems incorporate HD cameras with a resolution of 1920 x 1080 pixels, a pixel size of 7.4 µm and a focal length 8 mm.

To enable direct georeferencing of the imagery captured at typically 5 fps, a NovAtel SPAN inertial navigation system is used. The navigation system consists of a tactical grade inertial measurement unit with fiber-optics gyros of the type UIMU-LCI and a L1/L2 GNSS kinematic antenna. In the case of good GNSS coverage and post-processing, these sensors enable an accuracy of horizontally 10 mm and vertically 15 mm (NovAtel, 2015). The accuracies of the attitude angles roll and pitch are specified with 0.005° and heading with 0.008°. A GNSS outage of 60 seconds lowers the horizontal accuracy to 110 mm and the vertical to 30 mm.

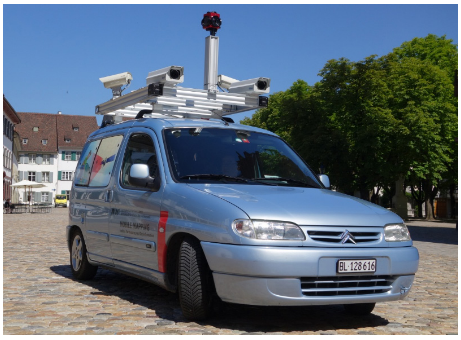


Figure 1. IVGI mobile mapping system

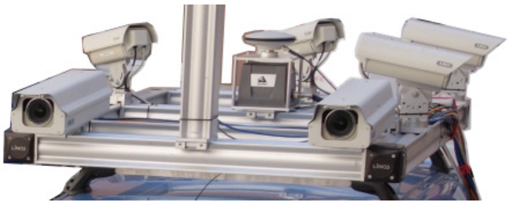


Figure 2. Sensor configuration of the IVGI MMS

2.2 Test area and test data

The test site depicted in Figure 4 is located at a very busy junction of five roads in the city center of Basel, Switzerland. It includes three tramway stops resulting in many overhead wires and is surrounded by rather large commercial properties (see Figure 3). This creates a very challenging environment for GNSS positioning. Furthermore, construction work as well as a large number of moving objects in the form of pedestrians, cars and tramways is in existence.

Figure 3. Mobile mapping imagery of image sequence 1 (left) and image sequence 3 (right) showing the test area

Mobile mapping data was acquired in July 2014 as part of a complete survey of the city-state of Basel by the FHNW spin-off company iNovitas AG. While eight image sequences were available in the test area, forward imagery of the three sequences depicted in Figure 4 was selected for further investigations. Two roads were mapped from trajectories in both directions and one road in one direction only. The selected test data was acquired at different dates and daytimes, which reflects typical real-world situations (see Table 1). An along-track distance between successive image exposures of 1 m was targeted, but larger distances occurred at velocities higher than 18 km/h since the maximum frame rate was 5 fps.

		1 (red)	2 (green)	3 (blue)
Date		24.7.2014	27.7.2014	27.7.2014
Time		10:20	11:53	11:57
Condition		sunny	cloudy	cloudy
Number of images		246	314	170
Length [m]		164	173	108
Along-track spacing	mean [m]	1.34	1.11	1.29
	min [m]	0.93	0.88	1.03
	max [m]	1.97	1.60	1.49

Table 1. Characteristics of the three selected image sequences

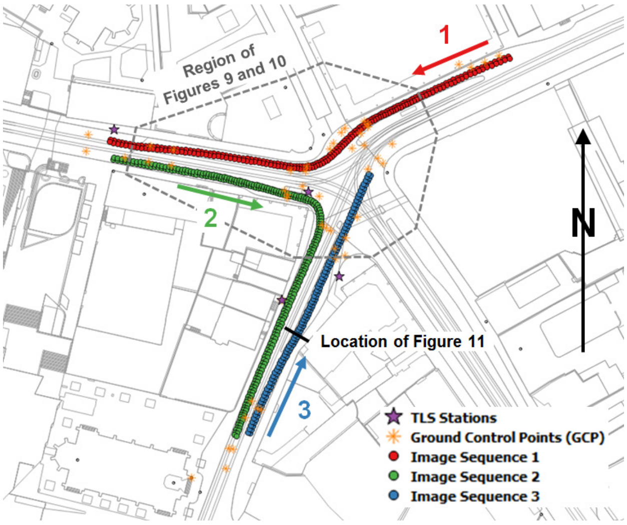


Figure 4. Base map of the test area with overlaid projection centers of the selected image sequences, GCP and TLS stations (Source: Geodaten Kanton Basel-Stadt)

2.3 Reference data

In order to assess the performance and quality of different matching strategies, there was a demand for independent and highly accurate reference data. Hence, four 360° terrestrial laser scans (TLS) recording XYZ point geometry and intensity were performed using a Leica ScanStation P20 on 19.3.2015. By means of several cadastral reference points and by registration using Leica Cyclone, an absolute 3D TLS accuracy of 1-2 cm was obtained. In addition, coordinates of 70 ground control points (GCP) were determined by a Leica Nova MS50 total station with a 3D accuracy of better than 1 cm on 18.3.2015. All measurements were performed in the Swiss reference frame LV95 and orthometric heights LN02.

3. PROCESSING PIPELINE FOR STEREO IMAGE SEQUENCES

The results from direct georeferencing of the captured stereo image sequences can already be used as input for dense image matching. However, there might still be some remaining systematic effects which were eliminated in a refined georeferencing step based on an additional bundle block adjustment as described in section 3.1. Especially if imagery has been captured at high redundancy, the selection of suitable combinations is an important step during multi-view stereo. Thus, as described in section 3.2, different matching configurations were selected from the mobile mapping sequence. Potential configurations from mobile mapping image sequences also consist of stereo pairs captured from cameras oriented in moving direction at different time frames. As already discussed, this presumed a newly implemented polar rectification of stereo imagery to allow for in-sequence matching of backward and forward looking cameras of our mobile mapping system (section 3.3). As a consequence, the following structure computation as implemented in SURE had to be modified additionally (section 3.4). Finally, a filtering and fusion step of the respective point clouds was applied (section 3.5).

3.1 Refined georeferencing by bundle block adjustment

In order to be able to compare the results of terrestrial laserscanning and dense image matching, direct georeferencing by post-processing using NovAtel Inertial Explorer had to be improved. Since not only position updates were required (Eugster et al., 2012), but also the attitude angles needed to be updated, bundle adjustment using PhotoScan was performed and the projection center coordinates of direct georeferencing served as initial values. As can be seen in Table 2, deviations of direct georeferencing and bundle adjustment lie in the range of a few decimetres, mainly caused by GNSS signal outages. 3D georeferencing accuracy on the totally used 50 ground control points (GCP) amounts to around 2-3 cm or circa 1/3 pixel (see Table 3). An overall projections error of 0.67 pixel was computed for image sequence 1, 0.65 pixel for image sequence 2 and 0.76 pixel for image sequence 3.

	X [mm]	Y [mm]	Z [mm]	3D [mm]
1	301	38	423	520
2	42	36	75	93
3	27	13	76	81

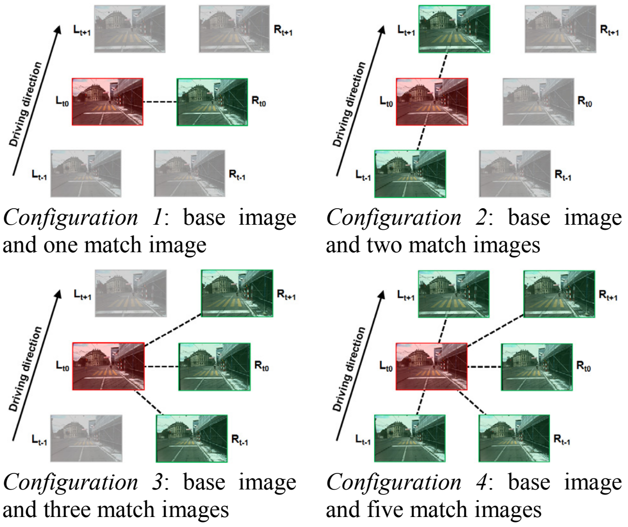
Table 2. Deviations of direct georeferencing and bundle adjustment

	No. GCP	X [mm]	Y [mm]	Z [mm]	3D [mm]	3D [px]
1	27	14	11	4	18	0.34
2	23	19	25	10	33	0.21
3	16	8	15	12	21	0.31

Table 3. Overall GCP errors

3.2 Selection of matching configurations

For processing and investigations of mobile mapping image sequences, the four matching configurations depicted in Figure 5 were selected. Configuration 1 represents standard stereo matching with one base and one match image for which a great number of algorithms exist (Scharstein & Szeliski, 2002). Configuration 2 stands for the case in which only mono imagery would be available thus matching the base image with the previous and the following image. In configuration 4, the base image is matched with all five neighbouring images. Six images are also used by Vogel et al. (2014) for evaluating their powerful dense 3D scene flow method which estimates both the depth and the 3D motion field of dynamic scenes. Omitting the match images of configuration 2 from configuration 4 leads to configuration 3.



Configuration 1: base image and one match image

Configuration 2: base image and two match images

Configuration 3: base image and three match images

Configuration 4: base image and five match images

Figure 5. Selected matching configurations

3.3 Polar rectification for in-sequence matching

Image rectification is the process of warping a pair of images in a way such that epipolar lines across two views are horizontal. In other words an arbitrary object point is projected to the rectified views to identical rows. This is a common pre-processing step in dense image matching algorithms to reduce computational complexity by restricting the search for corresponding pixels across two views to one dimensional epipolar lines. Despite of the straight forward implementation, the approaches proposed by Fusiello et al. (2000) and Loop & Zhang (1999) do not work for arbitrary geometric configurations of stereo pairs. Both algorithms construct virtual image planes parallel to the baseline connecting the two camera centers. Then homographic mapping is used to project the original images onto the virtual images planes. This is in particular problematic for motion in viewing direction, since virtual and original image planes are close to perpendicular which results in huge image dimensions and large distortions of the rectified imagery, as it is e.g. visible for the investigated configuration 4 in Figure 6. On the one hand, this increases processing times – and even more important – makes dense matching challenging because heavily distorted images cause problems in the computation of similarity measures.

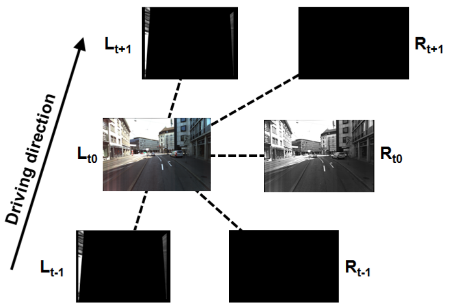


Figure 6. Rectification results by Fusiello et al. (2000)

A rectification method for general motion, also handling pure forward motion, as present in our mobile mapping application is proposed by Pollefeys et al. (1999). Our implementation is based on their approach, where rectification is performed by sampling corresponding half epipolar lines across two views. Thereby half epipolar lines are the line segments defined by subdividing the epipolar line at the epipole. Half epipolar lines are subsequently processed in circular scheme. Corresponding lines in the images are then arranged in parallel image rows in the rectified images $\mathbf{I}_{r,\theta}$, as shown in Figure 7.

The dimensions of the resulting images are limited by enforcing the distances of subsequent epipolar line Δ_e such that each pixel on the image border opposing the epipole is sampled exactly once. Additionally this adaptive sampling implicitly avoids the occurrence of pixel compression in $\mathbf{I}_{r,\theta}$, but necessitates a lookup table to invert the transformation. Epipoles close to, or inside the images entail heavily distorted image regions, where depth determination is inherently not accurate (Pollefeys et al., 1999). To avoid the influence of this singularity, we detect the affected image region automatically and discard it from further processing.

Following the rectification step, dense matching is carried out on $\mathbf{I}_{r,\theta}$. Interpolation in a lookup table, which is established during the rectification process, is used to determine the corresponding Cartesian coordinates u, v from polar coordinates r, θ.

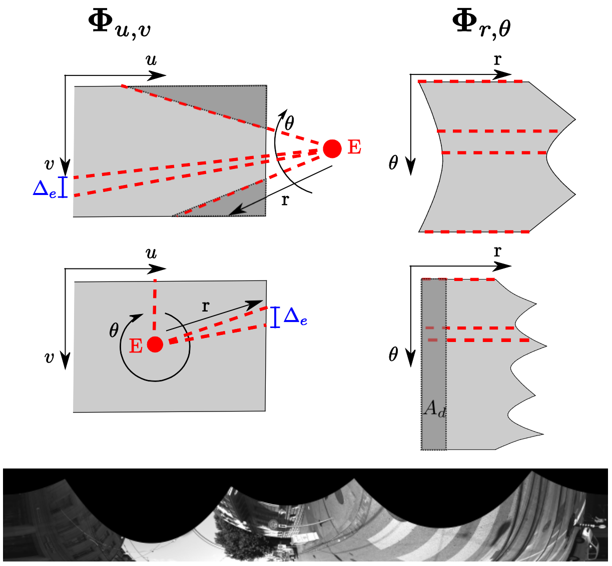


Figure 7. Polar rectification process and rotated polar rectified image (bottom row). Transformation of $\mathbf{I}_{u,v}$ to $\mathbf{I}_{r,\theta}$ in the case the epipole is located outside the image (top row) and in the case the epipole is located inside the image (middle row). A_d is the distorted image region which will be removed for depth estimation.

3.4 Structure computation

Once image pairs are rectified, disparity maps are computed. In the implemented Multi-View-Stereo (MVS) approach, spatially proximate disparity maps are fused by linking redundant depth estimates claiming geometric consistency. Eventually 3D points are triangulated from the sets of consistent disparity values. The rectification methods described in Fusiello et al. (2000) and Loop & Zhang (1999) produce configurations of rectified image pairs known as the stereo normal case (Kraus, 1994). Therefore, the 3D coordinates minimizing the reprojection error can be directly computed (Rothermel et al., 2014). In contrast, for arbitrary geometric configurations of image pairs, including polar rectified imagery, computing the depth minimizing the reprojection error is only possible by iterative approaches. To maintain fast processing speed we minimize the error in object space by computing the depth for all redundant disparities per stereo pair and then simply perform weighted averaging taking into account the uncertainties of single observations.

For structure computation, given a pair of calibrated images $\mathbf{I}_b$ and $\mathbf{I}_m$ as well as the respective normalized correspondences $\mathbf{x}_b = (x_b, y_b)$ and $\mathbf{x}_m = (x_m, y_m)$ both depth and 3D coordinates can be derived by forward intersection. First the intersection problem is reduced to 2 dimensions by reformulation using the epipolar plane. Therefore, a coordinate system is defined with the origin located in the optical center $\mathbf{C}_m$. The x-axis of the new coordinate system $\mathbf{\Phi}_m$ is defined by the epipolar line $\mathbf{l}_m$ in $\mathbf{I}_m$, the z-axis is the normal vector $\mathbf{n}_\pi$ of the epipolar plane π. Both entities are defined with respect to the camera coordinate system $\mathbf{\Phi}_m$ of the camera m. The y-axis is then constructed perpendicular to $\mathbf{n}_\pi$ and $\mathbf{l}_m$.

To form the triangle equation the vector of the baseline $\mathbf{B} = (B_x, B_y)$ and $\mathbf{x}_b$ are expressed with respect to $\mathbf{\Phi}_m$. The depth d corresponding to the 3D point can then be computed by requiring the two dimensional vector equation of triangles to be zero.

$$d\mathbf{x}_b - \alpha\mathbf{x}_m + \mathbf{B} = \mathbf{0} \quad (1)$$

Solving for d leads to

$$d = \frac{B_y x_m - B_x y_m}{y_m x_b - y_b x_m} \quad (2)$$

Eventually the final object coordinates can be computed from d. The implemented MVS approach computes redundant depth estimations by selecting a base or master image which is then matched against N match or slave images. As a result for each pixel in the base image several redundant depth estimations are available and further used to reduce outliers by checking for geometric consistency and to improve precision. Let $\mathbf{x}_b$ be the normalized coordinates in $\mathbf{I}_b$ and $\mathbf{x}_{m,i}$ the set of corresponding coordinates in the match images for $i = 1 \ldots N$. First we use equation 2 to compute the depths d_i along $\mathbf{x}_b$. Moreover, for each depth d_i an uncertainty interval $\sigma_{I,i}$ in image space is propagated to an interval $\sigma_{O,i}$ on $\mathbf{x}_b$. If the propagated uncertainty ranges overlap, the depths are considered consistent and used within the final depth computation. Taking into consideration the relative geometric precision implied by the propagated uncertainties $\sigma_{O,n}$ of the single observations, we compute the final depth using weighted averaging.

$$d = \frac{\sum_i \frac{d_i}{\sigma_{O,i}}}{\sum_i \frac{1}{\sigma_{O,i}}} \quad (3)$$

Note that since no further assumptions of rectification geometries are made, it is possible to incorporate correspondences of all three mentioned rectification methods into the same triangulation process.

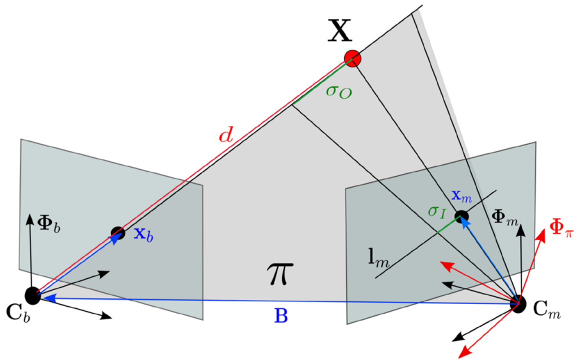


Figure 8. Structure computation in the epipolar plane

3.5 Filtering and fusion of point clouds

Forward looking imagery of the three selected image sequences (see Figure 4) was processed exploiting configurations 1 and 4 (see Figure 5). The overall objective in this case was to achieve an optimal scene in object space with not too dense but accurate points. It turned out that processing with standard parameters already leads to good results. Potential fine-tuning does not cause a significant improvement, while filtering in object space is crucial for our aim. Since all available forward looking imagery was incorporated, high similarity of neighbouring images and low scale differences of specific regions could be ensured which is essential for the dense image matching step. Filtering was carried out using the tool described in Wenzel et al. (2014). The three point clouds were subsequently fused which resulted in the point clouds depicted in Figure 9 and Figure 10.

While almost all moving objects are eliminated when five match images are incorporated (see Figure 10), several parts of moving objects remain if just one stereo pair is matched (see Figure 9). The scene depicted in Figure 9 also contains more clutter and a considerable number of sky points around overhead wires. The fact that more façades are mapped in Figure 10 is due to the minimum forward intersection angle of 2° for both configurations.

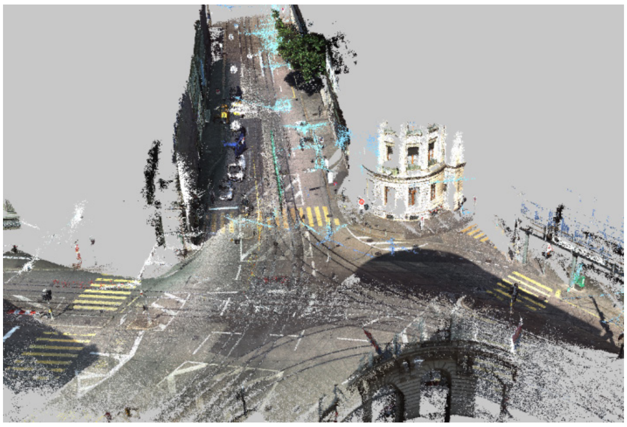


Figure 9. Filtered and fused point cloud exploiting one match image per base image (configuration 1)

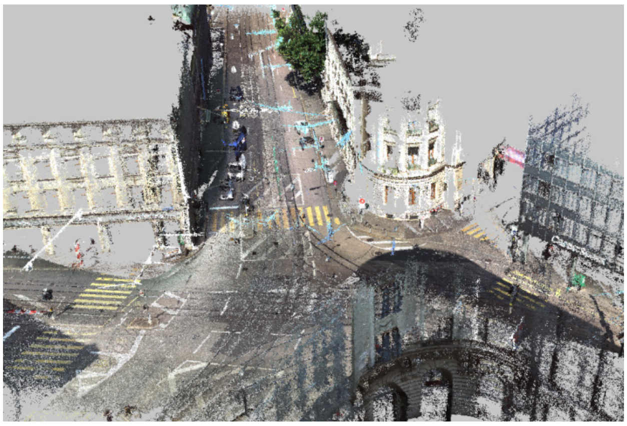


Figure 10. Filtered and fused point cloud exploiting five match images per base image (configuration 4)

4. ACCURACY INVESTIGATIONS

Visual comparisons enable a general impression, but they cannot disclose all effects. Therefore, numerical accuracy values needed to be computed for deviations in image space as well as for deviations of several road and façade patches in object space.

4.1 Investigations in image space

Accurate depth maps are fundamental for mono-plotting applications. Thus, comparisons of depth maps, which were either generated directly by the SURE triangulation module or by back-projecting point clouds to the viewing geometry of a base image, were performed. Same as Geiger et al. (2012) who did not interpolate ground truth disparity maps in order to avoid artificial errors, there was no interpolation in our case as large parts in several depth maps are missing and since the quality of single raw depth values needed to be evaluated. Depth deviations were only computed for pixels holding values for both depth maps and only deviations smaller than 50 cm were considered for RMSE and mean.

All investigations in image space which are discussed in the following were carried out using the image depicted in Figure 11. Firstly, relative comparisons were conducted whereby the depth map generated by configuration 4 served as reference. RMSE values for configurations 1-4 and 3-4 are close to 60 mm, for configuration 2-4 approx. 80 mm (see Table 4). Configuration 2-4 shows an opposite behaviour compared to the two others and the mean is 18 mm. Configuration 3-4 delivers the most façade points. In 2-4 and 3-4 the region close to the epipole where depth estimations are not accurate and thus eliminated (see section 3.3) is well indicated. This effect causes that a considerable amount of road surface points are not mapped.

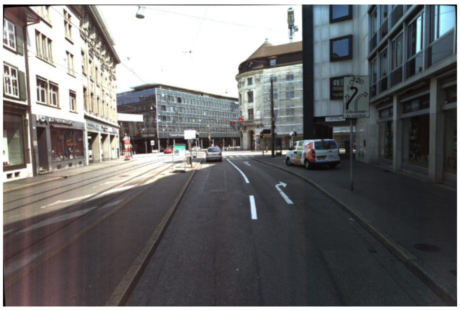
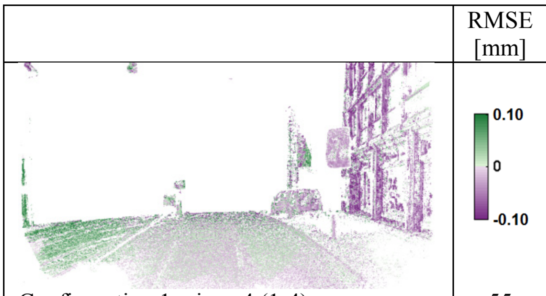


Figure 11. Base image from image sequence 3 for investigations in image space

	RMSE [mm]	Mean [mm]
		
Configuration 1 minus 4 (1-4)	55	-11

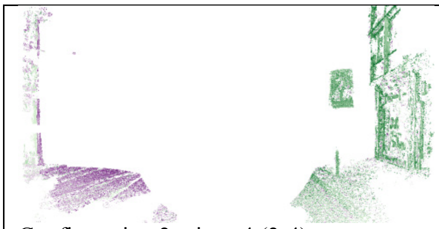
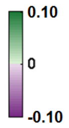
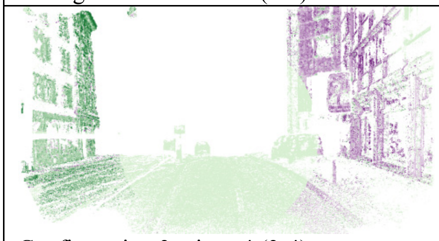
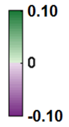
 Configuration 2 minus 4 (2-4)	 79	18
 Configuration 3 minus 4 (3-4)	 57	-3

Table 4. Deviations of depth maps generated by the SURE triangulation module

Secondly, terrestrial laser scanning points projected to image space were provided as reference and were compared with dense image matching point clouds generated by the SURE triangulation module for all configurations. The trend of positive and negative deviations remains identical for all configurations (see Table 5). RMSE values between 145 and 150 mm were computed for configurations 1, 3 and 4, which is significantly lower than 174 mm for configuration 2. There are almost no differences between 3-TLS and 4-TLS.

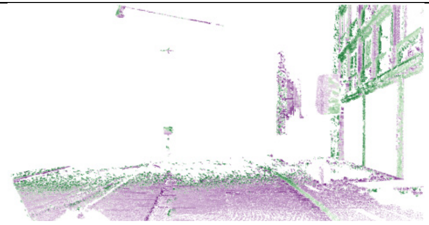
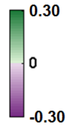
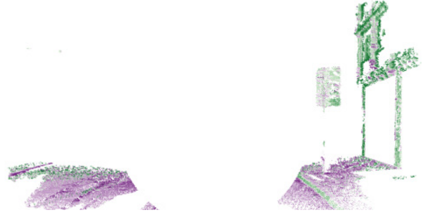
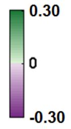
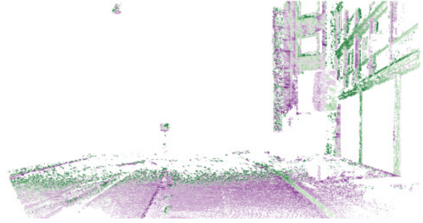
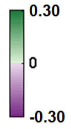
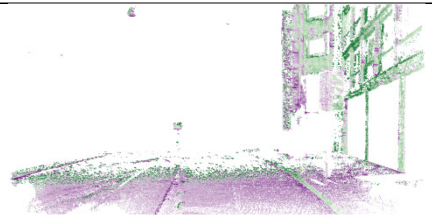
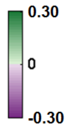
	RMSE [mm]	Mean [mm]
 Configuration 1 minus TLS (1-TLS)	 145	-17
 Configuration 2 minus TLS (2-TLS)	 174	-8
 Configuration 3 minus TLS (3-TLS)	 146	-10
 Configuration 4 minus TLS (4-TLS)	 150	-8

Table 5. Point clouds generated by the SURE triangulation module (1 base image) minus TLS

Lastly, deviations between the filtered and fused point clouds from dense image matching (see Figure 9 and Figure 10) and terrestrial laserscanning were computed. Mean values are similar, the RMSE value for configuration 1 is lower, but there are also fewer façade points which are much more inaccurate than road surface points (see Table 6).

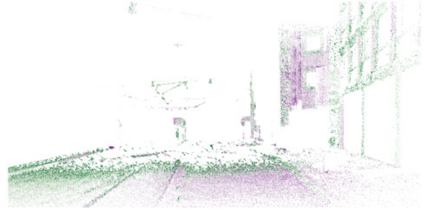
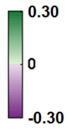
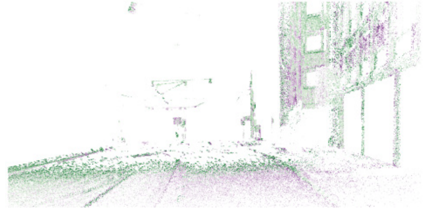
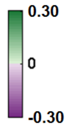
	RMSE [mm]	Mean [mm]
 Fused configuration 1 minus TLS (f1-TLS)	 147	23
 Fused configuration 4 minus TLS (f4-TLS)	 165	27

Table 6. Filtered and fused point clouds (several base images, Figure 9 and Figure 10) minus TLS

In summary, configuration 1 provides high accuracy due to a large base and thus a great number of sophisticated stereo algorithms is in existence. Configuration 2 loses many points around the epipole and suffers from limited accuracy since there is almost no base. The differences between configuration 3 and 4 are not significant. However, especially compared to the standard stereo configuration 1, the increasing number of match images as available in configuration 4 delivers significantly more points especially on the boundaries.

4.2 Investigations in object space

Accuracy of 3D reconstruction can be well assessed by investigations in object space. Hence, the evaluation procedure described by Cavegn et al. (2014) was used for 11 selected patches which are predominantly planar and cover road surface parts with shadows, rails and road marking as well as façade parts of different structure and material (see Table 7). All patches were arbitrarily defined by four points and then, both TLS and DIM point clouds were extracted using Leica Cyclone (see Figure 12). DIM patches were extracted from one of three filtered and fused point clouds, i.e. one point cloud for each image sequence. All reference TLS point cloud patches were subsampled to a distance of 1 cm. TLS and DIM grids of 5 cm spacing were used for the computation of deviations.

Figure 13 exemplarily shows the deviations of road patch 2 for configurations 1 and 4. Differences are not significant and rails are clearly indicated by positive deviations while road surface deviations are mainly negative. Cross- and along-track profiles reveal an offset of about 1 cm for both configurations compared to TLS and there is a little less noise for configuration 4 (see Figure 14).

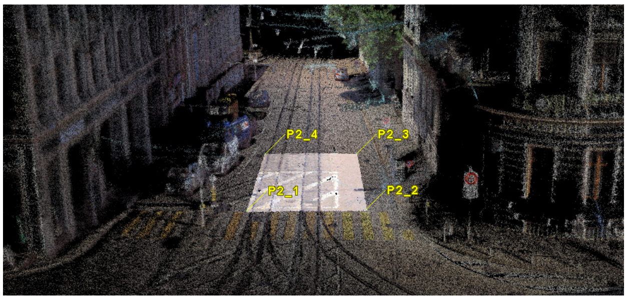


Figure 12. Extraction of road patch 2 using Leica Cyclone

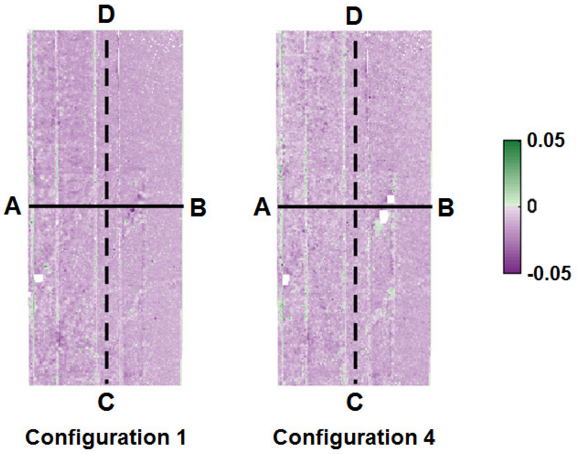


Figure 13. Deviations DIM-TLS of road patch 2

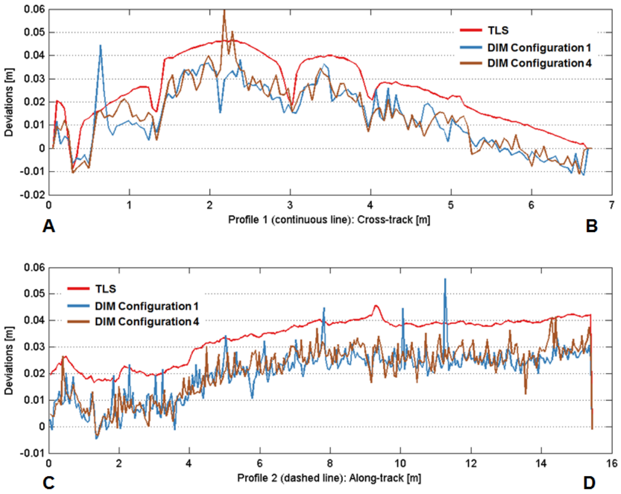


Figure 14. Cross-track and along-track profile of road patch 2

In Table 7 numerical values for the road patches P1-P5 and the façade patches P6-P11 are given. While the size of road patches ranges from 80 to 140 m^2, the size of façade patches lies between 20 and 120 m^2. Configuration 1 delivers more points than configuration 4 for road patches which is due to the fold parameter for the SURE triangulation module (fold 1 for configuration 1 and fold 2 for configuration 4). For façade patches all but patch 7 have a higher density for configuration 4 than configuration 1. The reason for this is a minimum angle value of 2° for the SURE triangulation module for both configurations. Low density values are caused by difficult matching conditions which is a large shadow area for patch 3 and a high number of road marking points whose pixel information is partly overexposed for patch 5. RMSE, mean and standard deviation values are almost identical for configurations 1 and 4 in case of road patches. Patch 11 features significantly more points as this is the only investigated façade which is almost perpendicular to the driving direction. Mean and thus RMSE values are larger for configuration 4 than for configuration 1, but there is the same standard deviation value. The larger values of patch 10 are due to different façade levels and the non-planarity. Standard deviation values for road patches are approx. 1 cm and around 5 times larger for façade patches.

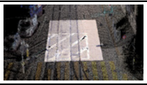
	Patch size [m^2]	Density [Points / m^2]	RMSE DIM-TLS [mm]	Mean DIM-TLS [mm]	SD DIM-TLS [mm]	
P1 c1	103	1742	7	-6	5	
P1 c4	103	900	7	-5	5	
P2 c1	105	1466	13	-12	6	
P2 c4	105	909	12	-11	6	
P3 c1	82	692	22	-20	9	
P3 c4	82	318	20	-18	9	
P4 c1	90	1729	14	-5	13	
P4 c4	90	1006	14	-3	14	
P5 c1	138	1063	12	10	7	
P5 c4	139	624	13	10	7	
P6 c1	81	1166	47	32	34	
P6 c4	81	1340	65	51	40	
P7 c1	61	729	57	46	33	
P7 c4	61	701	74	66	33	
P8 c1	119	908	72	-6	72	
P8 c4	119	1152	75	20	72	
P9 c1	38	1006	74	-68	30	
P9 c4	38	1132	88	-81	34	
P10 c1	85	606	89	-63	63	
P10 c4	85	1113	121	-109	54	
P11 c1	23	1580	67	-12	65	
P11 c4	23	2106	79	45	65	

Table 7. Density and deviation values for all road and façade patches by SURE 1 match image (configuration 1, c1) and SURE 5 match images (configuration 4, c4)

To sum up, configuration 1 enables high accuracy and the lack of façade points could be overcome by adapting the minimum angle parameter for the SURE triangulation module. Since the filter which was used has been implemented for terrestrial and airborne scenarios, an adaptation to the problem of motion in viewing direction could even result in better results. Nonetheless, filtering is a trade-off between point density and clutter or outliers.

5. CONCLUSIONS AND OUTLOOK

Within this paper the extension of our dense matching framework SURE to data of the image-based IVGI mobile mapping system was presented. First results for different stereo image configurations obtained in image and object space are promising. However, full benefit of the improved redundancy made available from additional stereo pairs in viewing direction will need further adaption of the 3D point filter algorithms. Standard stereo matching based on imagery captured at the same point of time from the two forward looking stereo cameras (configuration 1) already provided good accuracy, which did not significantly increase by additional use of images captured at different time frames (configuration 4). Nevertheless, the increased redundancy already improved reliability and completeness for 3D reconstruction. A further improvement of the results is to be expected especially if imagery from the back-right, left and panorama camera will be integrated. Future work will also include the improvement of automated processes for sub-pixel accurate co-registration/georeferencing of large image sequences for urban mapping.

6. ACKNOWLEDGEMENTS

Thanks are due to iNovitas AG (Baden-Dättwil, Switzerland) and to the City of Basel (Bau- und Verkehrsdepartement Kanton Basel-Stadt) for providing the mobile mapping data. Some of this work was co-financed by the Swiss Commission for Technology and Information CTI as part of the infraVIS project.

7. REFERENCES

Burkhard, J., Cavegn, S., Barmettler, A. & Nebiker, S., 2012. Stereovision Mobile Mapping: System Design and Performance Evaluation. In: *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, Melbourne, Australia, Vol. XXXIX, Part B5, pp. 453-458.

Cavegn, S., Haala, N., Nebiker, S., Rothermel, M. & Tutzauer, P., 2014. Benchmarking High Density Image Matching for Oblique Airborne Imagery. In: *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, Zurich, Switzerland, Vol. XL-3, pp. 45-52.

Eugster, H., Huber, F., Nebiker, S. & Gisi, A., 2012. Integrated Georeferencing of Stereo Image Sequences Captured with a Stereovision Mobile Mapping System – Approaches and Practical Results. In: *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, Melbourne, Australia, Vol. XXXIX, Part B1, pp. 309-314.

Fusiello, A., Trucco, E. & Verri, A., 2000. A compact algorithm for rectification of stereo pairs. *Machine Vision and Applications*, 12(1), pp. 16-22.

Gallup, D., 2011. Efficient 3D Reconstruction of Large-Scale Urban Environments from Street-Level Video. PhD thesis, University of North Carolina, Chapel Hill, USA.

Geiger, A., Lenz, P. & Urtasun, R., 2012. Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite. In: *IEEE Conference on Computer Vision and Pattern Recognition*, Providence, USA, pp. 3354-3361.

Haala, N., 2014. Dense Image Matching Final Report. EuroSDR Publication Series, Official Publication No. 64, pp. 115-145.

Kraus, K., 1994. *Photogrammetrie - Band 1*. Ferd. Dümmlers Verlag, ISBN 3-427-78645-5.

Loop, C. & Zhang, Z., 1999. Computing Rectifying Homographies for Stereo Vision. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Fort Collins, USA, pp. 125-131.

Nebiker, S., Cavegn, S., Eugster, H., Laemmer, K., Markram, J. & Wagner, R., 2012. Fusion of Airborne and Terrestrial Image-based 3D Modelling for Road Infrastructure Management - Vision and First Experiments. In: *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, Melbourne, Australia, Vol. XXXIX, Part B4, pp. 79-84.

NovAtel, 2015. UIMU-LCI Tactical Grade Fiber-Optic Gyros (FOG) Inertial Measurement Unit (IMU). http://novatel.com/products/span-gnss-inertial-systems/span-imus/uimu-lci/ (15.4.2015).

Pollefeys, M., Koch, R. & Van Gool, L., 1999. A simple and efficient rectification method for general motion. In: *International Conference on Computer Vision*, Kerkyra, pp. 496-501.

Pollefeys, M. et al., 2008. Detailed Real-Time Urban 3D Reconstruction from Video. *International Journal of Computer Vision*, 78(2-3), pp. 143-167.

Rothermel, M., Wenzel, K., Fritsch, D. & Haala, N., 2012. SURE: Photogrammetric Surface Reconstruction from Imagery. In: *Proceedings LC3D Workshop*, Berlin, Germany.

Rothermel, M., Bulatov, D., Haala, N. & Wenzel, K., 2014. Fast and Robust Generation of Semantic Urban Terrain Models from UAV Video Streams. In: *22nd International Conference on Pattern Recognition (ICPR)*, Stockholm, Sweden, pp. 592-597.

Scharstein, D. & Szeliski, R., 2002. A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. *International Journal of Computer Vision*, 47(1-3), pp. 7-42.

Vogel, C., Roth, S. & Schindler, K., 2014. View-Consistent 3D Scene Flow Estimation over Multiple Frames. In: *13th European Conference on Computer Vision (ECCV)*, Zurich, Switzerland, Part IV, LNCS 8692, pp. 263-278.

Wenzel, K., Rothermel, M., Fritsch, D. & Haala, N., 2014. Filtering of Point Clouds from Photogrammetric Surface Reconstruction. In: *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, Riva del Garda, Italy, Vol. XL-5, pp. 615-620.