ADVANCED TIE FEATURE MATCHING FOR THE REGISTRATION OF MOBILE MAPPING IMAGING DATA AND AERIAL IMAGERY

P. Jende [a, *], M. Peter [a], M. Gerke [a], G. Vosselman [a]

[a] Dept. of Earth Observation Science, Faculty ITC, University of Twente, 7514 AE Enschede, The Netherlands
(p.l.h.jende, m.s.peter, m.gerke, george.vosselman)@utwente.nl

Commission V, WG V/3

KEY WORDS: Mobile Mapping, Orientation, Accuracy, Estimation, Feature Matching, Template Matching

ABSTRACT:

Mobile Mapping's ability to acquire high-resolution ground data is opposing unreliable localisation capabilities of satellite-based positioning systems in urban areas. Buildings shape canyons impeding a direct line-of-sight to navigation satellites resulting in a deficiency to accurately estimate the mobile platform's position. Consequently, acquired data products' positioning quality is considerably diminished. This issue has been widely addressed in the literature and research projects. However, a consistent compliance of sub-decimetre accuracy as well as a correction of errors in height remain unsolved.
We propose a novel approach to enhance Mobile Mapping (MM) image orientation based on the utilisation of highly accurate orientation parameters derived from aerial imagery. In addition to that, the diminished exterior orientation parameters of the MM platform will be utilised as they enable the application of accurate matching techniques needed to derive reliable tie information. This tie information will then be used within an adjustment solution to correct affected MM data.
This paper presents an advanced feature matching procedure as a prerequisite to the aforementioned orientation update. MM data is ortho-projected to gain a higher resemblance to aerial nadir data simplifying the images' geometry for matching. By utilising MM exterior orientation parameters, search windows may be used in conjunction with a selective keypoint detection and template matching. Originating from different sensor systems, however, difficulties arise with respect to changes in illumination, radiometry and a different original perspective. To respond to these challenges for feature detection, the procedure relies on detecting keypoints in only one image.
Initial tests indicate a considerable improvement in comparison to classic detector/descriptor approaches in this particular matching scenario. This method leads to a significant reduction of outliers due to the limited availability of putative matches and the utilisation of templates instead of feature descriptors. In our experiments discussed in this paper, typical urban scenes have been used for evaluating the proposed method. Even though no additional outlier removal techniques have been used, our method yields almost 90% of correct correspondences. However, repetitive image patterns may still induce ambiguities which cannot be fully averted by this technique. Hence and besides, possible advancements will be briefly presented.

1. INTRODUCTION

Mobile Mapping data products are a valuable, additional source of geo-information especially to extend coverage and enhance detail in urban areas. However, these areas in particular cause difficulties for satellite-based direct georeferencing techniques, if buildings or other tall structures obstruct the necessary line-of-sight between the MM platform and the respective navigation satellites. Hence, MM data products' absolute accuracy may be impaired.

To tackle that problem, many authors rely on other sources of exterior orientation information, such as digital maps, aerial imagery or ground control points (Ji, Shi et al. (2015); Jaud, Rouveure et al. (2013); Kümmerle, Steder et al. (2011); Levinson and Thrun (2007)). A similarity of available approaches is a registration between the MM data and the reference data to yield enough correspondences which are utilised as constraints in an adjustment or filter solution, respectively. The majority of these methods utilises mobile laser scanning data, and approaches relying on MM images, such as Ji, Shi et al. (2015) compensate for matching errors within the filtering stage rendering a reliable registration unnecessary by accepting correct but mediocre correspondences. Moreover, these methods do not compensate for vertical errors, and cannot comply with a consistent decimetre accuracy.
In our research project, high-resolution aerial nadir and oblique images will be used to provide the required exterior orientation parameters. Although this paper will focus solely on the registration between MM and ortho-images computed from aerial nadir images, it constitutes the basis for further developments with respect to the registration of MM and oblique images.

A registration of images with a non-standard geometry and different sensor setup is not a trivial task. Either these differences have to be accounted for by feature detectors and descriptors, or by pre-processing to converge the images. State-of-the-art algorithms for the extraction of feature keypoints can account for differences in scale, rotation, illumination and perspective all to a certain degree (Alcanterilla, Nuevo et al. (2013); Rublee, Rabaud et al. (2011); Levi and Hassner (2015)), but bridging great overall variations e.g. between MM and aerial images has not been achieved yet.

* Corresponding author

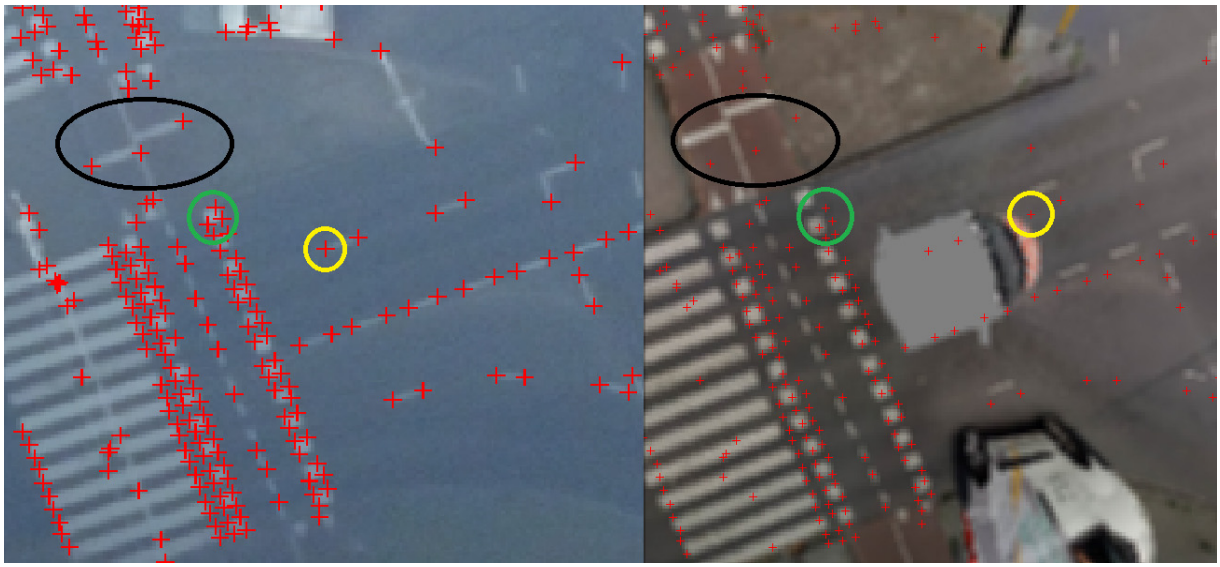

Figure 1. Left: aerial image with extracted Förstner keypoints; Right: aerial image's keypoints back-projected into MM image. Coloured circles illustrate the horizontal error. Moreover, the ambiguity problem with repetitive road markings becomes apparent using the example of the green circle; three corners have been identified in the aerial image. The back-projected keypoints, however, are now closer to the adjacent square-shaped road marking which may lead to a wrong correspondence.

Thus, this paper will discuss a matching strategy based on a pre-processing step to approximate the MM data to the aerial nadir image by performing an ortho-projection of the panoramic MM image. Moreover, the MM image's resolution and sharpness is also adjusted to increase the resemblance to the aerial reference data. Although the exterior orientation of the MM data may be imprecise, it is used to confine the matching procedure. However, the similarity between aerial nadir and MM images is still limited due to variations of contrast, illumination, image content, original perspective etc., reducing the efficiency of feature-based matching considerably. Road markings, however, remain a common element for image registration in urban areas with the intrinsic disadvantage to be repetitive. As a consequence, a feature-based matching may fail to find the correct correspondences due to the ambiguity of the computed feature description. Our experiments in the past have shown (Jende, Hussnain et al. 2016), that repetitive road markings can prevent a correct registration regardless of the employed detector/descriptor combination or the exclusion of various invariances. In conclusion, a feature matching relying on a separate detection and description phase in both input images is immanently sensitive to the overall image differences in this non-standard scenario.

Therefore, the strategy presented in this paper comprises a feature detection only in the aerial image, a kd-tree neighbourhood search for areas with sparse keypoints and a cross-correlation based template matching. Using an area-based approach is highly facilitated by the images' similar perspective as the integration of invariances, such as change in perspective, large scale differences or rotation is nonessential. In addition to that, the procedure can be designed independently from a feature detection in two images and enables the definition of the size of the moving window. Moreover, a template may alleviate differences in image details to a better extent than descriptor-based approaches, and consequently allows for a bridging of sensor differences. Even though this method yields only a few, but reliable correspondences, it allows for the determination of initial transformation parameters. Depending on the distribution and the number of correspondences, they may either serve as an outlier mask for a subsequent feature matching or already as an input for an adjustment to rectify the MM data set.

2. METHOD

2.1 Pre-processing

2.1.1 Pre-processing of aerial images: Our test data set comprises 15 aerial nadir images of Rotterdam (NL) with a maximum overlap of 6 images. These images were all combined to an ortho-mosaic with a ground sampling distance of 12 centimetres.

2.1.2 Ortho-projection and blurring of MM image: As a necessary pre-processing step, the MM data is ortho-projected in order to decrease the perspective differences between the input images. A horizontal plane representing the actual ground is computed based on the location of the MM sensor and its fixed height above the ground. A grid spacing which equals the aerial ortho-image's resolution is defined, and the corresponding MM panoramic image's pixels are projected onto the ground plane. Due to the definite description of the plane, all pixels contain approximated world coordinates. However, if the surrounding surface around the MM recording location is non-planar, this projection may lead to distortions. Although the deviation of the computed plane from the actual surface is minor especially in the immediate vicinity of the MM platform, a strategy has been developed to account for this issue. Belonging to the task of adjustment, however, this method will be presented in the future. After the ortho-projection, the MM image is blurred using a Gaussian filter to increase the resemblance to the aerial image for the step of template matching.

2.2 Matching and estimation of transformation

2.2.1 Feature detection: Feature keypoints are only detected in the aerial ortho-image using a modified version of the Förstner operator (Förstner and Gülch 1987) proposed by Köthe (2003). The major advancement is a revision of the structure tensor to extract corners at twice the resolution of the original image. This results in a higher accuracy and quality of the identified corners.

Corner features are suitable to represent road markings and other geometrical structures on the ground. To circumvent individual parametrisations as well as the risk of overfitting feature detection for a specific data set and first of foremost as a consistent and reliable keypoint detection in two disparate images cannot be guaranteed, feature keypoints are only detected in the aerial image. Back-projecting the identified keypoints into the MM image in a later step and using a template for matching, renders feature detection in the second image dispensable.

2.2.2 Avoiding repetitive patterns: Subsequently to feature detection, a kd-tree is computed to organise the detected keypoints and to allow for neighbourhood search. Repetitive road markings (see e.g. Figure 1) consist of many corners, thus returning many keypoints. Therefore, a kd-tree is used to identify isolated keypoints with fewer neighbours (i.e. 7 keypoints) in its extended surrounding of 50 by 50 pixels or 6 by 6 metres. This strategy simply avoids employing keypoints detected at corners of repetitive road markings for matching. In the future, however, these difficult features should be also utilised for a registration. A possible approach will be discussed in the last section of this paper.

2.2.3 Template Matching: By back-projecting the aerial image's keypoints into the MM image, the horizontal offset between the data sets becomes visible (Figure 1). Subsequently, for every keypoint in the aerial image, a template with 16 by 16 pixels is defined. To constrain the search space by still taking enough error margin into account, a 40 by 40 pixel neighbourhood around the back-projected keypoint in the MM ortho-image is used. Future work will, however, allow for an automatic and flexible definition of this window to account for different magnitudes of error in the MM image. Moreover, another feasible solution for the future would involve using parts of the MM image as a template, thus designing the process vice versa. Corresponding parts of the image are now cross-correlated. A threshold determines whether the peak of the normalised values in the cross-correlation matrix indicates a valid match. Moreover, the location of the peak is assumed to correspond to the extracted keypoint in the aerial image, thus the back-projected keypoint is moved to the peak's position allowing for a pixel-to-pixel accurate correspondence between the two images.

2.2.4 Estimation of transformation: The aerial and the MM image share approximately the same projection reducing the degrees of freedom to translation, uniform scaling and rotation. Scale is induced by an estimated ground sampling distance of the aerial ortho-image and rotation by a little deviation of the yaw axis of the MM platform from the north orientation. Thus, translation is the most crucial parameter.
As a similarity transform with only four degrees of freedom, two point pairs are needed to compute the parameters. The resulting transformation can be then employed as an outlier mask for a subsequent feature-based matching approach using e.g. state-of-the-art detector/descriptor combinations, such as those mentioned in the introduction. Since this approach finds an optimal solution with a least squares estimation, more than two non-collinear point correspondences can be used with the disadvantage of being not very robust against outliers. However, further experiments will show if the computed transform is reliable enough to directly serve as an input for the orientation update to rectify the MM data.

3. EXPERIMENTAL STUDY

3.1 Experimental setup

For this research project, the entire city of Rotterdam (NL) is being used as a test area as it offers a typical urban canyon scenario with a great number of high-rise buildings. For this specific test case, however, 14 tiles in Rotterdam's city centre cropped from the aerial ortho-image have been selected (Figure 2).

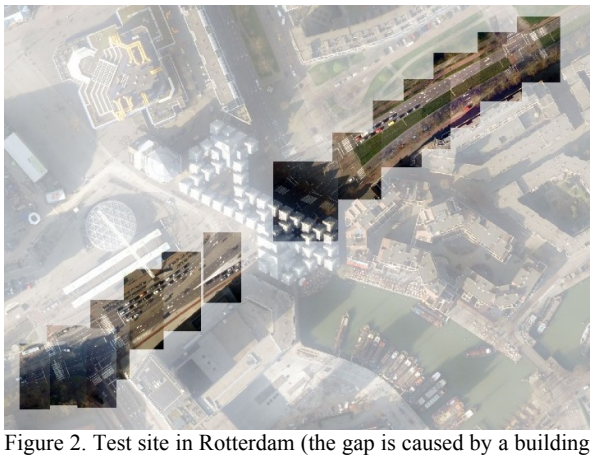

Figure 2. Test site in Rotterdam (the gap is caused by a building spanning the road)

These 14 tiles include difficult lighting and contrast conditions, repetitive road markings as well as great differences in the image content due to bustling traffic and vegetation. For every aerial image tile its centre coordinate is used to retrieve the corresponding MM ortho-image with an extent of approximately 21 metres side length. The presented approach has been applied to every tile of the test case, and no additional outlier removal technique has been conducted nor has the parametrisation been changed.

In order to evaluate the performance of the proposed method, visual checks as well as a comparison with feature-based approaches were conducted. For that purpose, a combination of the AGAST corner detector (Mair, Hager et al. 2010) and SURF will be used (Bay, Ess et al. 2008). AGAST is a corner detector which is not scale nor rotational invariant. This is intended as the rotation and scale of keypoints at a local level in this setup is negligible, and would introduce unnecessary ambiguities to the feature description. SURF is a widely-used float descriptor, which is – as default – scale and rotational invariant, but both parameters are estimated in the detector phase. Thus, by having no such information available from AGAST, SURF cannot account for these invariances. Alternatively, rotational invariance could be achieved by using Upright-SIFT (ibid). Moreover, coarse orientation information is also provided for the matching step by defining a search window around a back-projected aerial image keypoint, reducing the number of possible matches accordingly.

Additionally, a second feature-based matching approach with Förstner keypoint detection (Köthe 2003) and SURF description will be used for the comparison. For both methods, their default detector and descriptor parameter sets and a RANSAC-based outlier removal have been used. The inliers are then compared against the matching result of the proposed method. It is evident, that the remaining matches might be wrong, if RANSAC converged to a wrong solution. Involving all matches, however, would make a comparison difficult and insignificant as these two methods return far more matches than the proposed method.

To interpret the results correctly, it has to be mentioned, that the aim of the proposed method is to be as reliable as possible to find an initial estimate of the transformation between the aerial nadir and MM ortho-image. Thus, if a good ratio of matches versus inliers is available, the estimated transformation is more likely to be correct. Consequently, a few good matches are favoured over a high number of mediocre matches. To this end, the visual check has been quite strict, only labelling correct corner-to-corner or rather point-to-point correspondences as valid. However, a registration between MM and aerial data might not always be possible due to the lack of common features or different image content etc. Since the entire workflow includes a registration among adjacent MM images, this issue is not very crucial as a MM tile missing a direct correspondence to the aerial reference can be bridged accordingly.

3.2 Overview of experimental results

The three charts below (Figure 3, Figure 4 & Figure 5) show the overall results of the conducted experiment. In general, the number of matches varies among the three methods. But more importantly, the number of correct correspondences diverges significantly (see Table 1).

Table 1 Summary of matches, inliers and averages of all test tiles

Method	Number of matches	Number of inliers	Avg matches/tile	Avg inliers/tile	Ratio
Förstner /CC	86	75	6,1	5,4	87,2
AGAST /SURF	243	58	17,4	4,1	23,9
Förstner /SURF	172	9	12,3	0,6	5,2

With almost 90% of correct correspondences of the total number of identified matches, the presented method clearly outperforms the other methods which rely on a classic feature detector/descriptor combination. Matching images with a non-standard geometry is indeed quite difficult, even though coarse orientation parameters could be utilised. The aim, however, is to develop a method to be very reliable with respect to deriving an initial transformation between the aerial nadir and MM ortho-image.

In case of the presented method, a transformation estimate could be derived in 9 out of 14 tiles, resulting in a success rate of 64%. There were, admittedly, two tiles (5 & 13) without any correspondences making an estimation impossible.

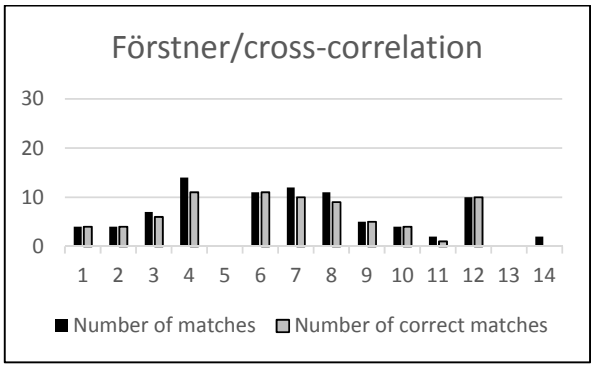

Figure 3. Matching results of proposed method across all 14 tiles

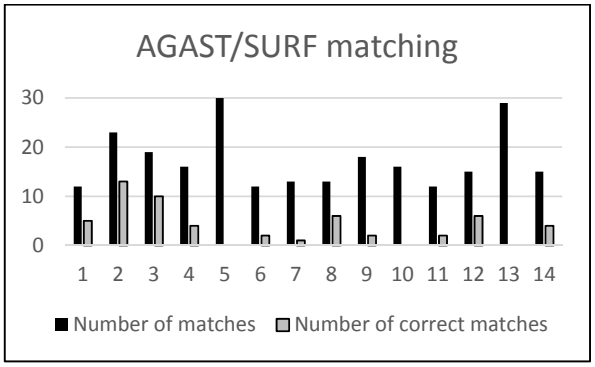

Figure 4. Matching results of AGAST detection and SURF description across all 14 tiles

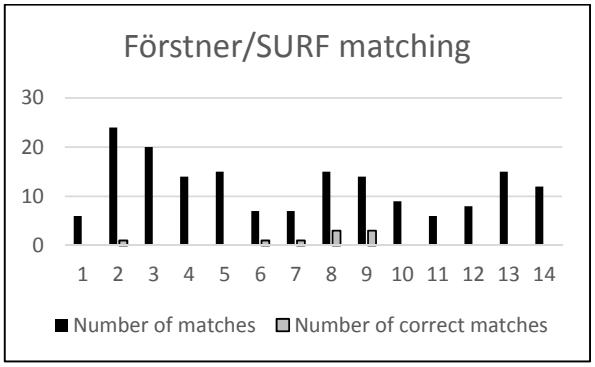

Figure 5. Matching results of Förstner detection and SURF description across all 14 tiles

3.3 Discussion of experimental results

In this section, some examples will be shown and discussed with regard to the individual performance of the respective method. In the figures shown below, red indicates a wrong and green a correct match.

3.3.1 Tile 2:

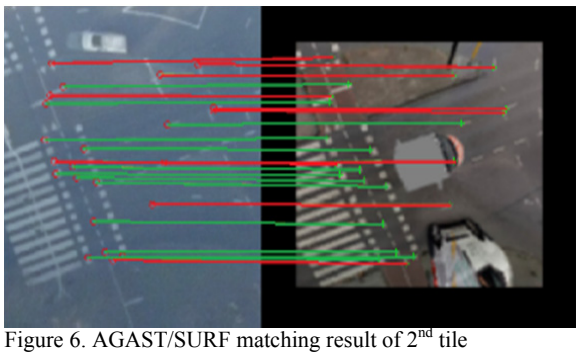

Figure 6. AGAST/SURF matching result of 2nd tile

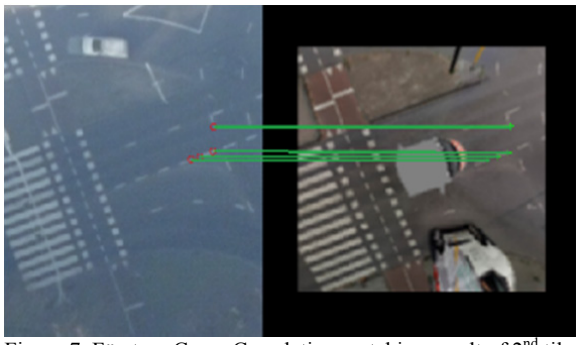

Figure 7. Förstner Cross-Correlation matching result of 2nd tile

Tile 2 shows a typical urban scenario with a zebra crossing. These repetitive road markings hamper the computation of distinct keypoint descriptors considerably. Figure 6 shows the matching result achieved with an AGAST detection and a SURF description. The result is good with only 10 mismatches out of 23 correspondences found in total. On the other hand, the result obtained by the method presented in this paper shown in Figure 7 only found 4, but therefore only correct correspondences.

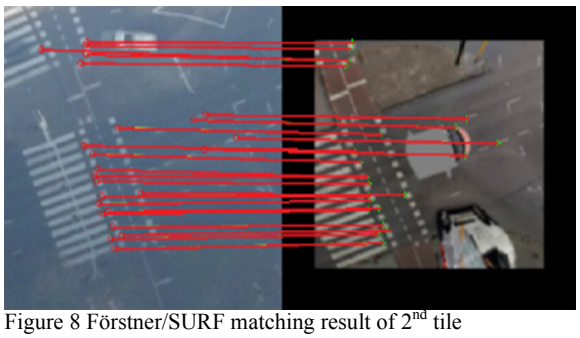

Figure 8 Förstner/SURF matching result of 2nd tile

Figure 8 depicts the results achieved with a Förstner corner detection and a SURF description. It is clear, that the repetitive appearance of the road markings hinders a successful registration. Among 24 matches, there is not a single correct correspondence.

3.3.2 Tile 6:

Figure 9 AGAST/SURF matching result of 6th tile

The result shown in Figure 9 is quite interesting as it ideally illustrates the problem with feature detection in two images from different sensors. Due to the overall differences, but the same parametrisation of the keypoint detection method, different keypoints are returned (see Figure 10) resulting in wrong correspondences. In this case, only 2 out of 12 matches were correct. Resolving this issue by adapting the feature detector's parameters to each image data set, however, may involve the risk of overfitting.

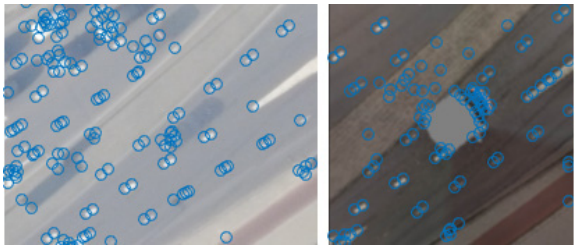

Figure 10 AGAST keypoints in aerial nadir and MM ortho-image

The presented method can compensate for these kind of problems by having a keypoint detection in only one image leading to a substantially better result of 11 correct correspondences (Figure 11).

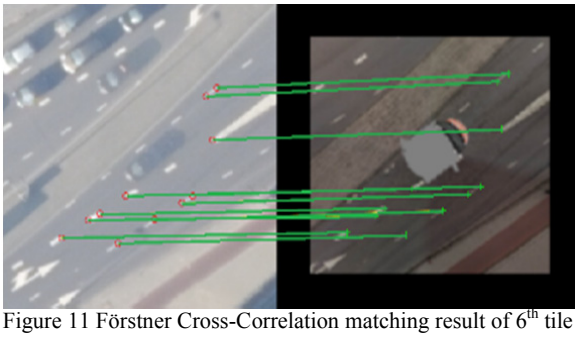

Figure 11 Förstner Cross-Correlation matching result of 6th tile

3.3.3 Tile 8: Not only different sensor systems lead to difficulties in matching, but also changing image content. A match is either valid, if the correlation value is above a certain threshold or below a defined descriptor distance. Since all three methods utilise some sort of search window, the number of matching candidates is already reduced to a high degree. If however, the image content changes – in that case – the MM car covering the zebra crossing, the closest and highest correlation value are the neighbouring strips of the zebra crossing (Figure 12). This can be tackled either by adjusting the correlation threshold or introducing a constraint preventing the assignment of a match which is below a certain clearance.

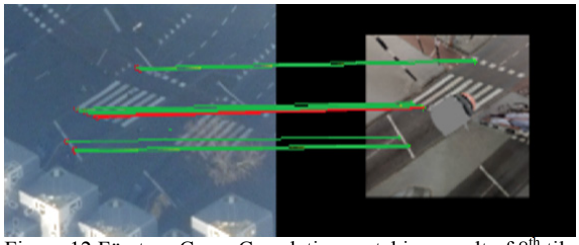

Figure 12 Förstner Cross-Correlation matching result of 8th tile

3.3.4 Tile 9: Tile 9 comprises of many elements impeding a successful matching. There is vegetation overgrowing parts in the aerial nadir image, repetitive road markings and a great difference in contrast. Figure 13 shows the result of the Förstner detection in combination with the SURF description. Out of 14 matches in total, just 3 are valid correspondences.

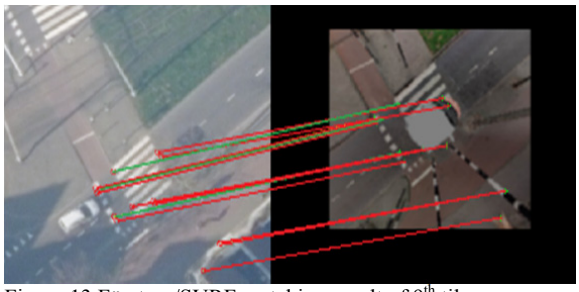

Figure 13 Förstner/SURF matching result of 9th tile

With the introduction of a keypoint neighbourhood constraint preventing a template matching in an area with a high keypoint density, the proposed method is able to identify only correct correspondences between these two images (Figure 14).

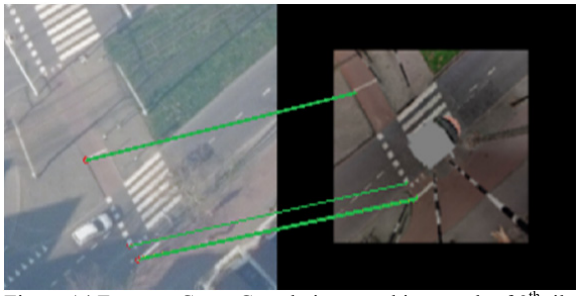

Figure 14 Förstner Cross-Correlation matching result of 9th tile

3.3.5 Tile 12: This example is quite similar to tile 6 where a successful registration has to rely mainly on sparse road markings. Classic detector/descriptor approaches do most likely fail in such a scenario since computed descriptors are not unique, and keypoints detected across two images originating from different sensors may not be the identical.

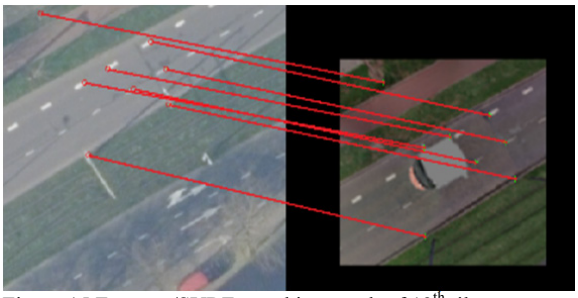

Figure 15 Förstner/SURF matching result of 12th tile

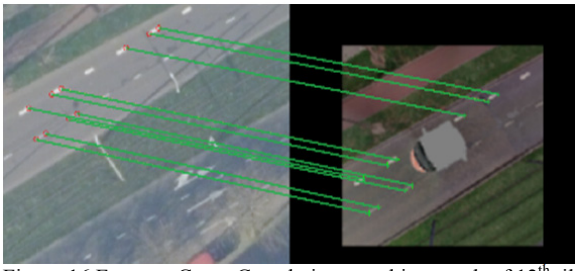

Figure 16 Förstner Cross-Correlation matching result of 12th tile

Comparing Figure 15 and Figure 16, it becomes evident that a cross-correlation based matching in combination with a feature detection in only one image is more reliable and robust than a descriptor-based approach.

3.3.6 Tile 13: The method proposed in this paper did not return any matches with the data from tile 13, but the other two methods did not succeed either, even though a lot of matches but no correct correspondences have been identified. Again repetitive road markings, overgrowing vegetation and different image content prevent a successful registration (Figure 17).

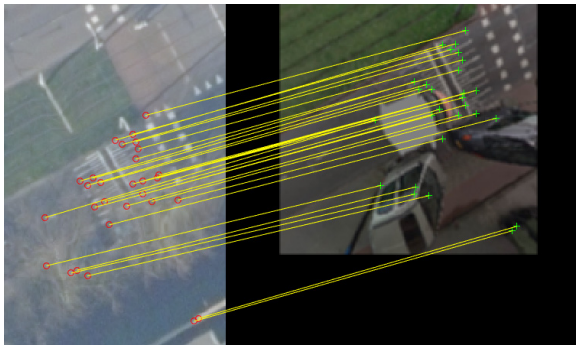

Figure 17 AGAST/SURF matching result of 13th tile with not a single correct correspondence

4. CONCLUSION

A substantial part within this research project is a reliable registration between MM images and the reference data set. In order to solve a matching problem involving this non-standard geometry, standard approaches are likely to fail or do not yield reliable nor consistent results. Converging the images by ortho-projecting the panoramic MM image was a first important step to simplify the images' geometry for matching. The second important strategy was to exploit the images' orientation parameters which enable the use of search windows to reduce the number of matching candidates. In this paper, the emphasis has been laid on the third step to detect features in only one image, on the one hand, and to use a template matching approach instead of a feature description on the other hand. The presented results and their comparison with descriptor-based methods could show, that not just the quality of individual correspondences, but also the reliability and robustness could be increased. Moreover, in 9 out of 12 cases where matches have been identified, a transformation could be estimated.

Currently, a kd-tree is being used to identify areas with a low keypoint density. This is indeed a fast solution to count keypoints in a window, but will actually serve as the basis for further developments. A problem which has not been tackled yet is how to cope efficiently with repetitive road markings. In most of the cases they can be simply avoided by this very neighbourhood search, but future developments should also exploit their geometric properties for registering the images. In the past, there were certain endeavours to determine which parts of an image belong together, such as perceptual grouping (Lowe 1985). Even though the method should be kept as simple as possible, and high-level feature approaches are likely to be immoderate also regarding their computational costs (Tournaire, Soheilian et al. 2006), the geometric interrelation between neighbouring keypoints will be utilised to determine whether certain road markings are repetitive. To this end, a check may be introduced, if a tuple of keypoints is collinear and complies to a set of rules.

Additionally, a network of correspondences between aerial to MM image matches and matches among MM images will be designed. Firstly, correspondences between MM images could contribute to the adjustment as they allow for the correction of relative errors along the MM platform's trajectory, and secondly, they may recover the orientation of individual MM tiles which do not have a direct correspondence to the aerial image.

ACKNOWLEDGEMENTS

This project is funded and supported by the Dutch Technology Foundation STW, which is part of Netherlands Organisation for Scientific Research (NWO), and which is partly funded by the Ministry of Economic Affairs.

REFERENCES

Alcanterilla, P. F., J. Nuevo and A. Bartoli, 2013. Fast Explicit Diffusion for Accelerated Features in Nonlinear Scale Spaces. *IEEE Trans. Patt. Anal. Mach. Intell* 34.7 (2011): 1281-1298.

Bay, H., A. Ess, T. Tuytelaars and L. Van Gool, 2008. Speeded-up robust features (SURF). *Computer vision and image understanding* 110(3), pp. 346-359.

Förstner, W. and E. Gülch, 1987. "A fast operator for detection and precise location of distinct points, corners and circular features." *Proc. ISPRS intercommission conference on fast processing of photogrammetric data*. 1987.

Jaud, M., R. Rouveure, P. Faure and M.-O. Monod, 2013. Methods for FMCW radar map georeferencing. *ISPRS Journal of Photogrammetry and Remote Sensing* 84(0), pp. 33-42.

Jende, P., Z. Hussnain, M. Peter, S. Oude Elberink, M. Gerke and G. Vosselman, 2016. Low-Level Tie Feature Extraction of Mobile Mapping Data (MLS/Images) and Aerial Imagery. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Lausanne, Switzerland, XL-3/W4, pp. 19-26.

Ji, S., Y. Shi, J. Shan, X. Shao, Z. Shi, X. Yuan, P. Yang, W. Wu, H. Tang and R. Shibasaki, 2015. Particle filtering methods for georeferencing panoramic image sequence in complex urban scenes. *ISPRS Journal of Photogrammetry and Remote Sensing* 105(0), pp. 1-12.

Köthe, U., 2003. Edge and junction detection with an improved structure tensor. *Pattern Recognition, Proceedings* 2781, pp. 25-32.

Kümmerle, R., B. Steder, C. Dornhege, A. Kleiner, G. Grisetti and W. Burgard, 2011. Large scale graph-based SLAM using aerial images as prior information. *Auton. Robots* 30(1), pp. 25-39.

Levi, G. and T. Hassner, 2015. LATCH: Learned Arrangments of Three Patch Codes. *arXiv* 1501 (2015)

Levinson, J. and S. Thrun, 2007. Map-Based Precision Vehicle Localization in Urban Environments. *Robotics: Science and Systems*, Vol. 4, p. 1.

Lowe, D. G., 1985. Perceptual Organization and Visual Recognition, *Kluwer international series in engineering and computer science. Robotics: vision, manipulation and sensors*, 1985

Mair, E., G. D. Hager, D. Burschka, M. Suppa and G. Hirzinger, 2010. Adaptive and generic corner detection based on the accelerated segment test. *Computer Vision–ECCV 2010*, Springer, pp. 183-196.

Rublee, E., V. Rabaud, K. Konolige and G. Bradski, 2011. ORB: an efficient alternative to SIFT or SURF. *Computer Vision (ICCV), 2011 IEEE International Conference on Computer Vision*, Barcelona, 2011, pp. 2564-2571.

Tournaire, O., B. Soheilian and N. Paparoditis, 2006. Towards a Sub-Decimetric Georeferencing of Ground-Based Mobile Mapping Systems in Urban Areas: Matching Ground-Based and Aerial-based Imagery Using Roadmarks. *ISPRS Commission I Symposium - From sensors to imagery*, Paris, France.