CO-REGISTRATION OF TERRESTRIAL AND UAV-BASED IMAGES – EXPERIMENTAL RESULTS

M. Gerke, F. Nex, P. Jende

University of Twente, Faculty of Geo-Information Science and Earth Observation (ITC),
Department of Earth Observation Science, The Netherlands {m.gerke,f.nex, p.l.h.jende}@utwente.nl

KEY WORDS: Photogrammetry, Orientation, Data, Integration, Multisensor, Accuracy, Estimation

ABSTRACT:
For many applications within urban environments the combined use of images taken from the ground and from unmanned aerial platforms seems interesting: while from the airborne perspective the upper parts of objects including roofs can be observed, the ground images can complement the data from lateral views to retrieve a complete visualisation or 3D reconstruction of interesting areas. The automatic co-registration of air- and ground-based images is still a challenge and cannot be considered solved. The main obstacle is originating from the fact that objects are photographed from quite different angles, and hence state-of-the-art tie point measurement approaches cannot cope with the induced perspective transformation. One first important step towards a solution is to use airborne images taken under slant directions. Those oblique views not only help to connect vertical images and horizontal views but also provide image information from 3D-structures not visible from the other two directions. According to our experience, however, still a good planning and many images taken under different viewing angles are needed to support an automatic matching across all images and complete bundle block adjustment. Nevertheless, the entire process is still quite sensible – the removal of a single image might lead to a completely different or wrong solution, or separation of image blocks.
In this paper we analyse the impact different parameters and strategies have on the solution. Those are a) the used tie point matcher, b) the used software for bundle adjustment. Using the data provided in the context of the ISPRS benchmark on multi-platform photogrammetry, we systematically address the mentioned influences. Concerning the tie-point matching we test the standard SIFT point extractor and descriptor, but also the SURF and ASIFT-approaches, the ORB technique, as well as (A)KAZE, which are based on a nonlinear scale space. In terms of pre-processing we analyse the Wallis-filter. Results show that in more challenging situations, in this case for data captured from different platforms at different days most approaches do not perform well. Wallis-filtering emerged to be most helpful especially for the SIFT approach. The commercial software pix4dmapper succeeds in overall bundle adjustment only for some configurations, and especially not for the entire image block provided.

1. INTRODUCTION

Multiplatform image data is very interesting for many applications. Unmanned Aerial Vehicles (UAV) are getting more mature and fully automatic processing workflows are in place which help turning the image set into point clouds or more advanced products. At the same time and due to the availability of easy-to-use end-user software also hand-held cameras are used by researchers from a variety of disciplines to model objects. Examples are as-is-modelling of buildings, archaeology/cultural heritage, cadastre/city modelling. In order to model the outer faces of buildings entirely, with great detail, and with a minimum amount of occlusions, the object should be photographed from many different viewpoints. Those should be at different heights and enclosing a variety of angles with the object. In case of complex architectures such as intrusions, extrusions (like balconies), a UAV can offer favourable viewpoints to avoid or minimize occlusions. In addition the roof should be captured from conventional nadir-looking views.
The processing pipelines proposed and implemented in research and commercial products work well, especially when the following conditions are met:

- Approximate position and viewing direction of cameras are known: to more reliably find matching mates for each image and exclude unlikely matches.
- Sequence of image acquisition resembles overlapping configuration (adjacent images also have similar time stamp)
- Viewing direction change between overlapping images is small (i.e. perspective distortion is small): because most key point descriptors, or image matchers, are not invariant with respect to large perspective distortions.
- Lighting conditions for overlapping images are similar: again, some key point descriptors are sensitive to global grey value distribution changes
- Object does not show repetitive patterns: similarity between areas in the object will lead to wrong matches. Especially in buildings with symmetrical façade-object arrangements (windows/doors) this is a problem.

In practice, however, an image block configuration might not be perfect: GPS which is used to estimate approximate camera location might compute largely wrong positions because of signal obstruction and multipath effects in urban canyons, or close to buildings in general. In addition, when several image blocks, taken from different platforms are merged, the valuable adjacency information from the image capture time gets lost, at least for matches between the blocks.
When object planes enclose large angles – like right-angled building facades – special attention has to be taken when images are captured to ensure that the image matcher still finds enough valid tie points in adjacent images. The lighting might become problematic, as well. First, when we do not have diffuse light, parts of the building may be covered by strong shadow, e.g. casted by extrusions of the building itself. Second, depending on solar elevation angle, building height and camera height, glare might lead to underexposure of building details, and third, when images from different sensors are matched. Repetitive patterns which are standard in many architectural designs are a problem especially when images at very high resolution are used: single shots then only cover parts of the façade and the context might get lost completely: when a single window is photographed on two images it might not be possible to decide whether this is actually the same or just another window of the same type.

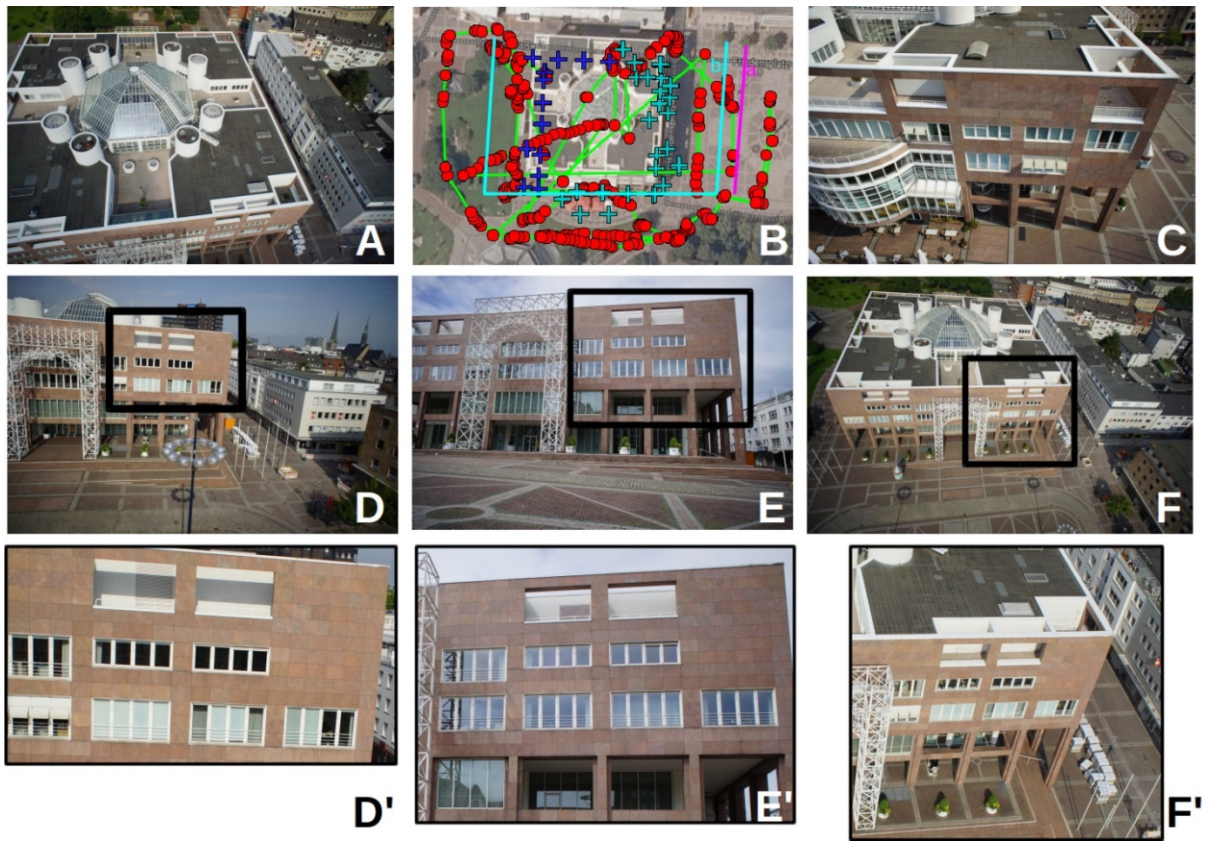

Figure 1: Overview on benchmark dataset. A,C,D,F: UAV images from different locations and under different angles, B: GPS-based image position approximations overlaid on google earth view (red dots), GCP/CPs (crosses), magenta and turquoise lines: used facades for software test, section 4, subset a and b, respectively, E: terrestrial image. D', E', F': cut outs of the respective image.

A challenging dataset, composed of terrestrial and UAV-based images is provided to the research community in the framework of the multi-platform photogrammetry benchmark (Nex et al., 2015), which is supported by the ISPRS scientific initiative 2014-2015[1].

One of the released datasets is a combined UAV/terrestrial image block, and the task for participants is to co-register the images and fine adjust the bundle block using ground control points provided. In Fig. 1 some overview is given including sample images. Fig. 1 B) shows the approximate positions of images around the centre building (municipality hall in Dortmund, Germany). In the northern part, where images were captured in a narrow street, the positions are largely off the right location. Images A,D,E,F show the same part of the building from different perspectives. Although the same camera (Sony Nex 7) was used for both – terrestrial image captures and on board the UAV – the colour and grey value distribution is quite different, in particular between terrestrial shots (E) and the UAV-based images. In addition, the sky/clouds are reflected in the windows in the terrestrial shots. During the terrestrial acquisition the weather was very bad – the sky was cloudy and the campaign was interrupted frequently by heavy rain. During that day it was not possible to operate the UAV. Therefore, the acquisition has been conducted about 4 weeks later. Then the weather was very good, led to no diffuse light and clear sky (compare D and E). As far as the building architecture is concerned the window and other elements do not show large variations, but large repetitions. Note also that image C) shows another façade of the building, but actually the window elements and their arrangement are similar to the façade shown in A,D,E,F.

In this paper we present experiments and their results focussing on the issues: how do current state-of-the-art tie point matching algorithms perform on this dataset and which influence does image pre-processing have? To this end, several image combinations (same platform/across platform) are matched. We perform outlier removal based on a RANSAC approach exploiting the epipolar constraint in stereo matches, using the essential matrix. The Wallis filter (Wallis, 1976) was applied in a separate experiment. Apart from testing different stereo matching techniques we analyse the performance of one software package. The entire structure-from-motion workflow including image matching and bundle adjustment is tested. The solution offered by pix4d (pix4dmapper) is used for those tests.

2. RELATED WORK

In literature we find a multitude of approaches to tie point extraction, description and matching. The aim of this paragraph is not to give a comprehensive overview, the interested reader might want to refer to (Dahl et al., 2011, Levi and Hassner, 2015) for some general overview. Urban and Weinmann (2015) tested state-of-the-art key point extractors and detectors in the context of co-registration of terrestrial laser scans. To this end the authors compared the keypoint extractors SIFT, SURF, ORB and A-KAZE on depth-images derived from the single scans. Used approaches are of a different type in the sense that ORB is based on corner detection, and the others extract blobs in the scale space. As far as keypoint descriptors are concerned

[1] http://www2.isprs.org/commissions/comm1/icwg15b/benchmark/data-description-Image-orientation.html

we can distinguish between gradient-based descriptors, finally encoded in floating numbers, and descriptors computed from intensity differences and encoded as binary strings. The latter one is supposed to be more computational efficient, but many of those are known to be more sensitive to noise. In addition to the aforementioned key point extractors/detectors we add the SIFT-based ASIFT and the KAZE approaches when we evaluate the performance of the methods for the benchmark dataset. One pre-requisite for all used descriptors is that they are both, scale and rotation invariant.

2.1 SIFT/ASIFT

The Scale Invariant Feature Transform SIFT (Lowe, 2004) became a standard in computer vision and photogrammetry. It works in scale space which is derived by image convolution with a Gaussian kernel. Extrema in the DoG (Difference-of-Gaussian) constitute keypoint candidates. After removing candidates along edges or in low contrast regions, a higher order function is fitted to derive sub-pixel accuracy. Scale invariance is achieved implicitly through the realisation of a scale pyramid. The keypoint descriptor is derived by computing gradient histograms in all directions. The area around the point is subdivided into 4x4 regions and in each region the orientation histogram is computed in 8 angular bins. Those 4x4x8 bins are concatenated and stored as a 128-dimensional descriptor, along with the dominant scale. Since the main gradient direction is derived as well and rotations are normalised accordingly, the descriptor is rotation invariant.
Morel and Yu (2009) extended the SIFT approach in order to achieve invariance under affine image transformations (ASIFT-affine SIFT). To this end, the images are stepwise rotated around both axis. For each of such re-projections SIFT points are computed and described.

2.2 SURF

In contrast to SIFT, the Speeded-Up Robust Features SURF detector (Bay et al., 2008) does not work with the DoG, but with the Hessian matrix in the scale space pyramid. Local maxima of the matrix determinant in image- and scale space constitute candidates for keypoints. Similar to SIFT, uncertain hits are removed and accuracy is increased through sub-pixel interpolation. Rotation invariance of the descriptor is also derived by first computing the dominant gradient direction. Image subdivision in 4x4 regions is also done similar to SIFT, but in this case Haar wavelets are computed to describe the local gradients in the frequency domain. Four descriptors per sub region are computed, leading to 4x4x4=64 entries per point.

2.3 ORB

The full name of the acronym ORB (Rublee et al, 2011), namely Oriented FAST and rotated BRIEF, already tells that this is a combination of an improved version of the FAST feature detector (Rosten and Drummond, 2005) and the rotational invariant BRIEF (Calonder et al., 2010) descriptor.
The FAST (Features from Accelerated Segment Test) descriptor finds keypoints basically by comparing intensity differences around each pixel in the image of interest, where the pattern of tests has a circular shape. The two main extensions within ORB are that those detections are done in scale space, i.e. adding scale invariance, and that rotation angle information is added.
As far as keypoint description is concerned, the authors of ORB added rotation invariance and unsupervised learning to select an ideal pairing of pixel samples to the BRIEF descriptor.

2.4 KAZE/A-KAZE

One basic idea behind the KAZE (Alcantarilla et al., 2012) point extractor is to use a non-linear scale space to enable scale invariance. Compared to the Gaussian scale space derivation as done by other approaches, a non-linear scale space, in this case realized by nonlinear diffusion filtering, preserves edges while reducing noise at the same time. In terms of computational time, however, it is reported that the employed additive operator splitting (AOS) is quite inefficient. Therefore, in Alcantarilla et al. (2013) the authors propose A-KAZE, where the A stands for accelerated. Here, the fast explicit diffusion (FED) is used to compute the scale representation.
Similar to SURF, in both cases, the Hessian matrix is employed to find salient points. For KAZE, point description is undertaken by a modified variant of SURF (M-SURF, Agrawal et al., 2008) adapted to the non-linear scale space. AKAZE, however, utilizes a binary description based on a modified version of the Local Difference Binary method proposed by Yang and Cheng (2012).

2.5 Matching and inlier filtering

In order to match keypoints for each candidate in an image the closest mate in terms of descriptor-space distance is searched for. This descriptor-based matching is then followed by the so-called ratio test. The distance in descriptor space between the best and second best match is computed, and the ratio should not exceed a certain threshold. In this way outliers can be removed since the assumption is that also in the descriptor space a valid match should not be isolated.
In order to enhance filtering inliers, a two-step approach is pursued, exploiting the perspective camera model: since the camera models are known for the benchmark description, a RANSAC-based filtering is done using an estimation of the essential matrix E. To this end, the image coordinates are first normalized employing the camera calibration matrix K (Hartley and Zisserman, 2008). The essential matrix can be computed using the 5-point algorithm by Nistér (2004) or Li and Hartley (2006). We used the MATLAB implementation provided by the latter. In a second step the filtered points are again processed by a RANSAC-based filtering using the fundamental (F)-matrix estimation. We found out that using only F-matrix-based filtering in cases with a large number of outliers might not converge. In those cases the geometric constraints imposed by the essential matrix help. However, since the camera calibration might not be known too well, we also observed that still a significant number of outliers are present afterwards. Therefore, an F-matrix-based filtering applied to the remaining matches reduces the number of outliers considerably.

2.6 Image enhancing prior to point extraction

As far as image pre-processing is concerned Jazayeri and Fraser (2008) reported that an image enhancement with the Wallis filter (Wallis, 1976) helped to significantly improve corner point detection. Therefore, after testing the tie point matching with the original images we will perform the same with Wallis-filtered images on selected pairs. This filter is an adaptive contrast filter, working in local windows. In contrast to many other global filters, image details will remain and contrast and brightness is balanced over the entire image.

3. TIE POINT MATCHING

In order to test the 6 point extractor/descriptor combinations for the benchmark test data, a setup has been defined as follows.

We selected 10 different image combinations which reflect all the challenges we are facing in this dataset as mentioned in section 1. In Figures 2 and 3, the four selected terrestrial and UAV images, respectively, are shown.

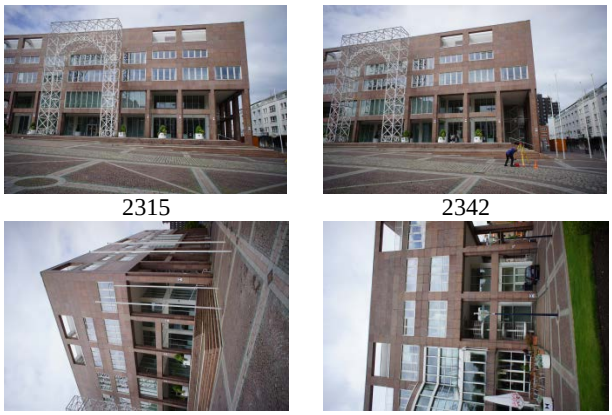

Figure 2: Terrestrial images used for the tests

3.1 Stereo pair matching in original images@25%

The images are resampled to 25% of the original size. This is simply done for practical reasons: the computation time we need to perform all the experiments can be reduced. We may assume that the number of matches is smaller compared to a higher resolution, so the visualization of matches is better legible. Anyhow, we believe that this reduction of resolution is valid since we are only interested in the relative performance between the approaches, given several typical stereo image combinations.
However, in a later step we will show experiments with full resolution images for selected stereo pairs. We also add results from matching of pre-processed images, in particular after Wallis filtering.
All experiments have been conducted with the Matlab/OpenCV[2] implementation of the approaches used. An exception is ASIFT for which we use the C-implementation provided by the authors (Morel and Yu, 2009).

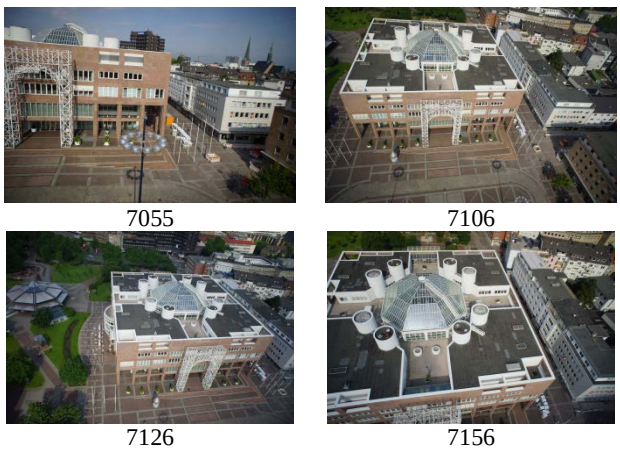

Figure 3: UAV images used for the tests

Pair 1 – within terr-1: 2315- 2342 (standard)
In Fig 4., the upper part the matches of SURF are displayed while in the lower part for each approach the number of inlier matches after E-/and F-based filtering (blue) is shown and compared to the real number of inliers (from visual inspection). Compared to all other pairs, this is the simplest case for matchers: a similar scale, similar viewing direction, just small baseline, same time of capture, i.e. no illumination changes. In terms of the number of reliable matches, the SURF approach outperforms all the others; however, all methods would give enough reliable matches for practical applications.

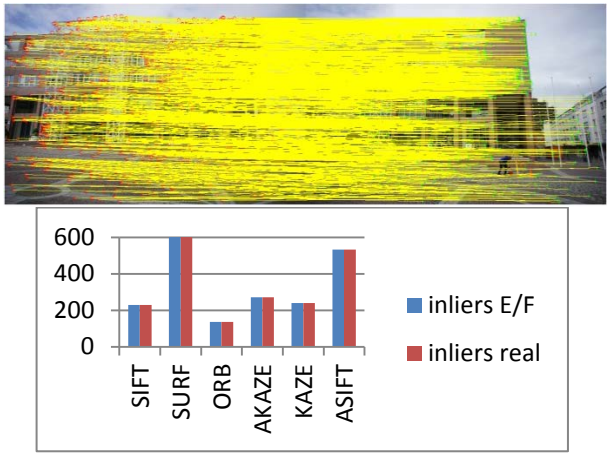

Figure 4: pair-1 results. Upper: SURF matches, lower: for each number of matches after E-/and F-based filtering (blue) and real inliers (red)[3]

Pair 2 – within terr-2: 2392-2342 (perspective/rotation)
This pair is a bit more challenging than the first one in two respects: the viewpoint and image rotation changes in a way to include the 2nd facade. The rotation is quite small, though. In addition the orientation of one camera changes from landscape to portrait mode. This is done in close range projects for two reasons. Sometimes the opening angle of the camera in landscape mode does not allow for acquiring the façade in the entire vertical direction. Additionally, by rotating the camera by 90° the self-calibration of interior parameters is supported, in particular the estimation of the principal point.

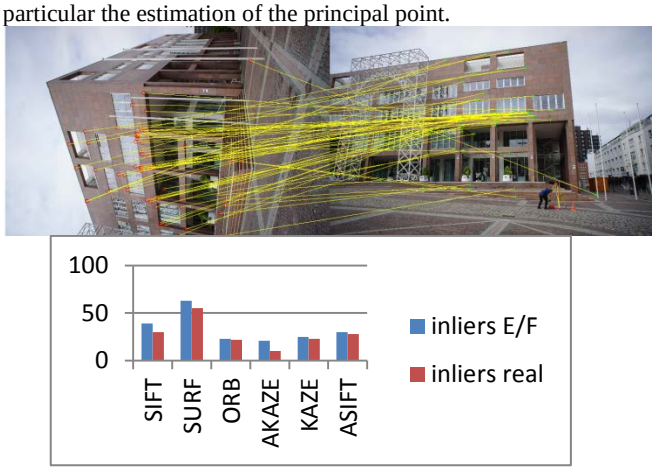

Figure 5: pair-2 results. Upper: SURF matches, lower: for each number of matches after E-/and F-based filtering (blue) and real inliers (red)

The result (see Figure 5) show that all matchers yield much fewer valid tie point connections compared to pair 1. Again

[2] OpenCV 3.0, including the API MEXOpenCV to use OpcnCV functions in Matlab: https://github.com/kyamagu/mexopencv

[3] A visualization of matches from all six approaches for all 10 matching pairs is available as additional material on research gate, see http://www.researchgate.net/profile/Markus_Gerke

SURF delivers most inliers, but especially the A-KAZE inliers might be too few in a practical setup.

Pair 4 – within UAV-1: 7106 -7126 (standard)
In this case, we test a pair similar to pair 1 in the sense that it resembles a simple situation: Similar viewing direction of images, with another viewpoint shifted in façade direction. Similar to pair 1, SURF and ASIFT provide the most matches, but SURF only achieved a comparable result to SIFT (Fig. 6).

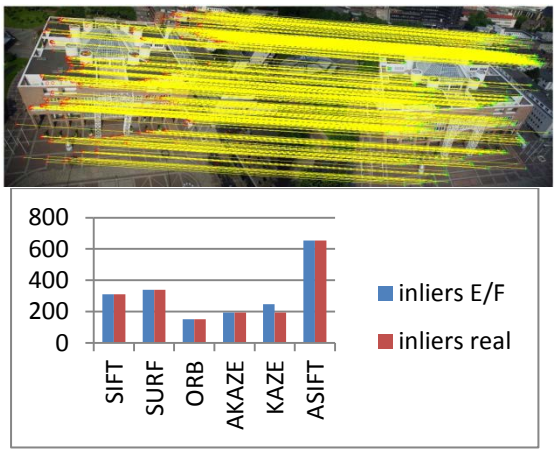

Figure 6: pair-4 results. Upper: ASIFT matches, lower: for each number of matches after E-/and F-based filtering (blue) and real inliers (red)

All other matchers provide a similar number of valid matches as in the example of pair 1. Another remarkable observation is that the ORB and A-KAZE matches are not as well distributed as the matches from the other methods

Pair 5 – within UAV-2: 7106-7055 (oblique- horizontal)
This pair is typical for UAV image blocks in the context of 3D building modelling: the camera is tilted to include different angles with the nadir direction. In this way, horizontal views (for façades) can be connected to vertical views (for the roof). In addition more complex object structures like in- or extrusions can be modelled since occlusions get minimized.

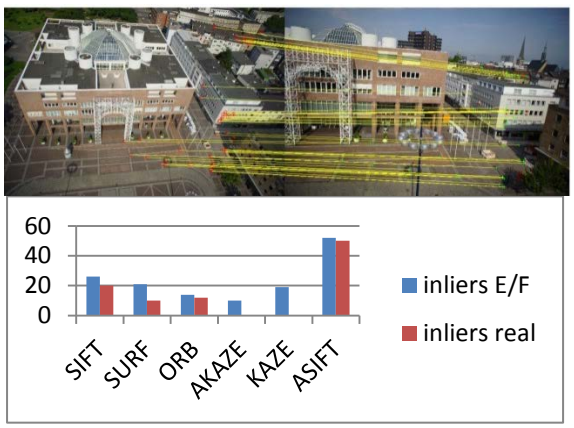

Figure 7: pair-5 results. Upper: ASIFT matches, lower: for each number of matches after E-/and F-based filtering (blue) and real inliers (red)

Again, the ASIFT result is significantly better than the others (Fig. 7). While standard SIFT still yields around 50% of the number of ASIFT matches, the number of real inliers from the others is much less. Especially (A)KAZE do not deliver valid matches at all.

Pair 6 – within UAV-3: 7106-7156 (oblique-vertical)
Again, this dataset is typical for 3D building modelling, like pair 5, but this time a nadir view is connected to the slanted one.

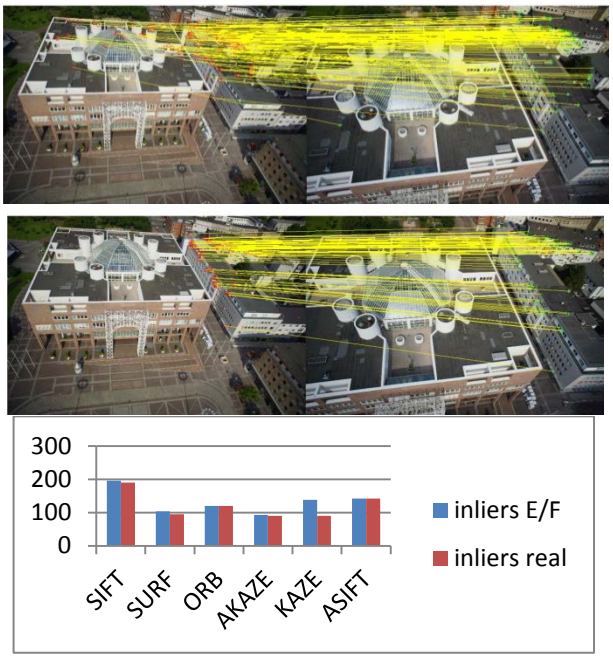

Figure 8: pair-6 results. Upper: SIFT, middle: ORB, lower: for each number of matches after E-/and F-based filtering (blue) and real inliers (red)

Although the nature of perspective transformation imposed by the different camera nick is similar as in the previous case, here many more matches in general can be found – all methods yield around 100 valid matches, around double compared to SIFT. Remarkably SIFT delivers more matches than ASIFT. However, it is also visible from the match visualization (Fig. 8) that the distribution of matches from (A)SIFT is much better than that one from the other approaches – ORB, AKAZE and KAZE just return matches in a very small area of the scene.

Pairs 3 and 7 – within terr-3: 2315 -2793 and within UAV-4: 7055-7336 (wrong pair)
The object arrangement is quite similar to pair 2 and pair 5, respectively. However, the images of the pair show different parts of the building, so *each match is a false* match. Therefore in the bar chart (Fig. 9) the "real" inliers are omitted since there cannot be any real inliers.

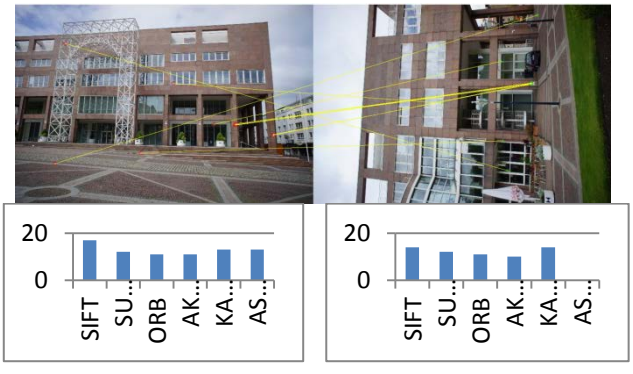

Figure 9: pair-3 results. Upper: SIFT matches for terrestrial wrong matches, lower: for each number of matches after E-/and F-based filtering (blue) (left: terrestrial, pair-3, right: UAV, pair-7)

Although all extractor/descriptor combinations deliver only a few inlier matches after E-/F-filtering, it is remarkable that for instance AKAZE does not produce fewer matches compared to pair-2 where a valid image combination was used. When the matches are visualized (Fig. 6), one can see that most of the matches are at window frame corners, and columns. For the airborne case, however, ASIFT-based matching did not result in inliers.

Pair 8 – across terr/UAV-1: 2315-7055 (similar viewing direction)
From a geometrical point of view this pair is standard and thus comparable to pair 1. The building façade is photographed with the same camera, from the same direction, only the UAV took a higher altitude as the terrestrial shot. However, this pair is special in the sense that the illumination of the scene is quite different, see the description of the acquisition campaign in section 1.

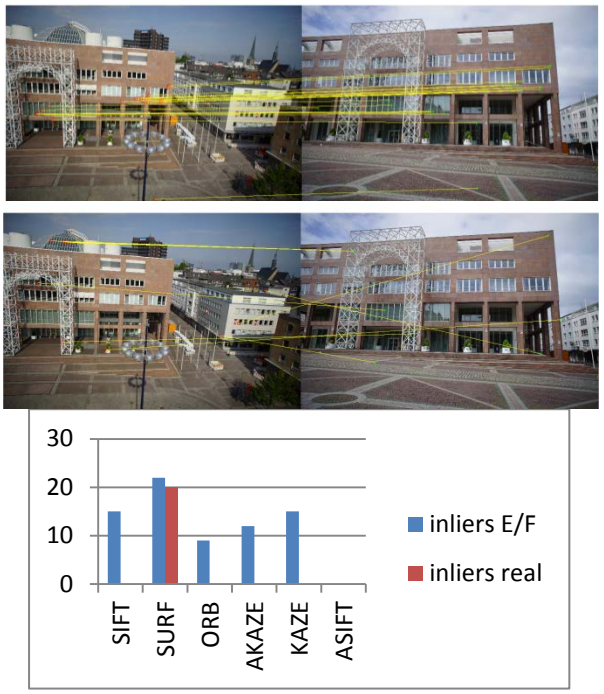

Figure 10: pair-8 results. Upper: SURF, middle: SIFT, lower: for each number of matches after E-/and F-based filtering (blue) and real inliers (red)

Achieved results are poor, cf. Fig. 10: although in all cases inliers remain after RANSAC-based filtering, all of them are invalid, except for the SURF results which are still acceptable. However, with about only 20 inliers, this result is worse compared to pair 1 (more than 600).

Pair 9 – across terr/UAV-2: 2315-7106 (horizontal- oblique)
This combination is similar to pair 5 (within UAV-2): a horizontal view is combined with a slanted view from the air. Here, basically all matches fail. Almost all remaining inliers after RANSAC filtering (10 to 15) are actually wrong matches. When comparing to the previous pair, this result is reasonable since the geometric transformation imposed by the camera tilt makes the task not easier, and in addition the images show the same unfavourable radiometric differences as in pair 8.

Pair 10 – across terr/UAV-3: 2315-7126 (horizontal-oblique)
The idea for this combination is the same as for pair 9: a classical terrestrial view is combined with a slanted oblique view. This pair, however, is even more challenging than the previous one since the common area is more towards the background of the UAV image. Hence, the scale difference between both images is quite large. As one could expect, no approach produced usable results.

3.2 Alternative setups

In order to experimentally analyze the impact of image rescaling and Wallis filtering, we performed the same experiments with a selected pair (pair 8) again. We selected pair 8 because on the one hand it seems to be challenging as the two images were captured by the two different platforms at different days, but on the other hand - from a geometrical point of view - it should not be too difficult (see description above).

Full resolution matching
In practical projects it is important to reach the highest geometric accuracy, therefore it is advised to match full resolution imagery. In our experiments, however, we found out that – at least in the used setups and implementation – there is no significant improvement in terms of real inliers. Although the number of keypoints grows almost linearly with the image resolution, the absolute number of inliers, i.e. not the inlier ratio, remains somewhat stable for all approaches.

Wallis filtering
After Wallis filtering, the matching result for pair 8 improved significantly, but only for SIFT matching. While SIFT does not produce a single real inlier in the original images (cf. Fig. 10, middle), it results in more than 100 real inliers in the Wallis-filtered images, see Fig. 11.

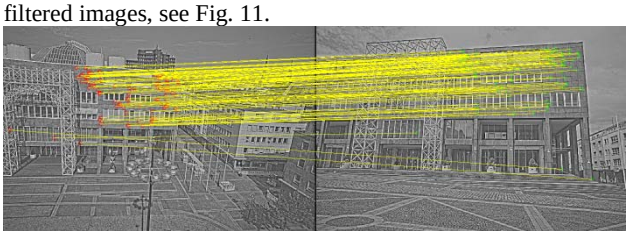

Figure 11: pair-8 results: SIFT matching in Wallis-filtered images.

3.3 Summary of tie point matching results

The results give some interesting insights to the performance of the selected tie point extractors and matchers. Since in terms of outlier removal and image scale we used the same settings for all approaches, we can at least compare the results relatively to each other. For the standard case, namely to match images taken from the same sensor and platform, and showing similar illumination conditions, all matchers perform satisfactorily, see pairs 1 and 4. Under certain image transformations other than similarity, most approaches show difficulties. In those cases ASIFT and SURF performed best, but also SIFT works well, especially when nadir UAV images are combined with slanted UAV views. When it comes to image pairs across platforms, and in this case being taken under quite different illumination conditions, the number of inliers drops drastically. Only in the simplest combination, when the geometry of image acquisition is similar (pair 8), a fairly decent result was obtained from

SURF only. We might suspect that the employed Hessian matrix is less sensitive to large illumination changes.
After Wallis-filtering, only the SIFT result improved significantly – from no real inliers in the original image to around 100 in the Wallis-case. To use the full resolution of the original image, in turn, does not lead to an improvement of descriptor-based matching.

4. STRUCTURE-FROM-MOTION TESTS

When end-users are working with software packages they often have no influence on the entire workflow. Especially the keypoint extraction and description is normally hard coded and the algorithm behind is not disclosed. We tested the initial performance of such software, in particular with the challenging data of this benchmark. To this end, the pix4dmapper by Pix4d[4] has been analysed. For all experiments we did not use the ground control point information provided in the benchmark dataset in order to be independent from external tie information.

We undertook the following experiments with this software. First, we defined three sub-datasets, considering only terrestrial or UAV images as well as both datasets:
a) all images showing only one façade of the building (East façade, the same side as used for the tie point matching experiments);
b) three façades: all terrestrial images except for North side where the largest problem with initial GPS location is observed. This set, however, includes already all UAV images;
c) entire dataset.
The location of subsets for a) and b) are also indicated in Figure 1, B).

By analyzing subset a), we can focus only on the across-platform matching and bundle adjustment performance since the approximate GPS location is reasonable for all images. Also for subset b), the GPS locations are good, but the geometry of buildings is more challenging since multiple oriented façades plus roof and ground planes are involved. The full dataset (subset c)), finally, is the most challenging since the poor GPS locations from the terrestrial images showing the North façade are included. For all subsets we performed several software runs: only terrestrial, only UAV, combination terrestrial and UAV. This was done twice: on original and Wallis-filtered images, and on the first pyramid (half image resolution) only.

Configuration	original	Wallis
UAV only		
UAV_set a) – East Facade	☑	☑
UAV_set c) – all	☑	☑
Terrestrial only		
Terrestrial_set a) – East Facade	☑	☑
Terrestrial_set b) – East, South, West	☑	☒
Terrestrial_set c) – all	☒	☒
Combinations		
UAV_set a) and Terrestrial_set a)	☒	(☑)
UAV_set c) and Terrestrial_set b)	☒	☒
UAV_set c) and Terrestrial_set c)	☒	☒

Table 1: Experimental results with pix4dmapper on several image block subsets.

[4] http://www.pix4d.com

Table 1 gives an overview on the results obtained. The tick mark indicates that all images are adjusted in one connected bundle block. The cross mark is used when the block got divided into sub blocks. There is one case (first combination with Wallis-filtered images), where pix4dmapper resulted in a stable block, but 5 UAV images got excluded from the solution.

The software had no problems when only the UAV images were used. The quality report reveals that images are well connected, see Figure 12. The results shown are from set c (all UAV images), using the original images. Using the Wallis-filtered images yields a similar result.

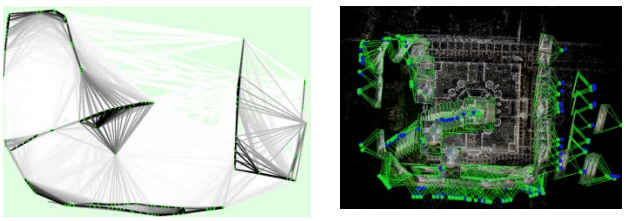

Figure 12: Image connectivity (left), tie point cloud and camera locations (right), all UAV images

The terrestrial image set was solved without bigger problems as expected in the one-façade-only case (set a). The next challenging set b was only solved as one image block when the original images got employed. The entire terrestrial image block, however, (set c) was not adjusted successfully at all. This observation might support the assumption that good GPS location observations are necessary to support the entire matching and adjustment process.

A typical block configuration is shown in Figure 13: the terrestrial block got separated; in particular the images from the North façade are not connected to the others.

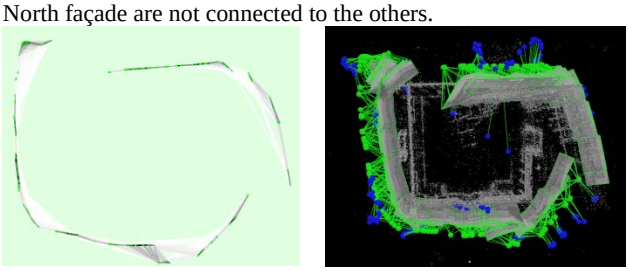

Figure 13: non-connected terrestrial image-block, Wallis-filtered images

For the combined UAV- and terrestrial image configurations, only the most simple one (set a) got solved, but only for the Wallis-filtered images and even in that case some cameras are not included in the block, cf. Fig. 14. This observation somehow backs up our observation from section 3, namely that the matching across the multi-platform dataset, especially the fact that the illumination is quite different, challenges commercial state-of-the-art software, as well.

5. CONCLUSIONS

It turns out that the ISPRS benchmark dataset is a challenging, but at the same time also realistic example for close range/UAV image blocks. State-of-the-art tie point matching approaches show good results in some published work. However, in this case still the traditional SIFT, in combination with Wallis filtering outperforms all other approaches for a mixed-platform and illumination image pair. The selected images are all from the released dataset, i.e. all the experiments can also be conducted by interested researchers.

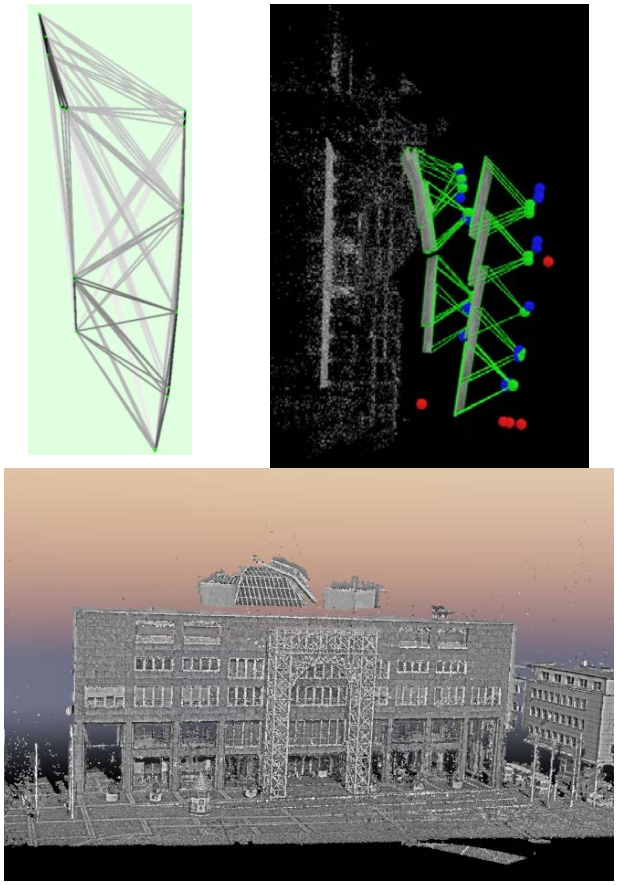

Figure 14: UAV and terrestrial, set a, Wallis-filtered image. Upper left: image connections, upper right: tie point cloud and camera locations (red: not adjusted), lower: densified point cloud.

Using pix4dmapper we found out that the entire block of images gets split-up into several sub-blocks. Our assumption is that one reason for this failure is a combination of a low number of inter-platform image matches, with a bad approximate geo-location for images, especially in the terrestrial part.

ACKNOWLEDGEMENTS

Data acquisition and pre-processing was made feasible through the funds provided by ISPRS (Scientific Initiative) and EuroSDR. We thank Pix4D for providing us a research license of pix4dmapper.

REFERENCES

Agrawal, M., Konolige, K. and Blas, M. R., 2008. CenSurE: center surround extremas for realtime feature detection and matching. *Proceedings of the European Conference on Computer Vision*, Vol. IV, pp. 102–115.

Alcantarilla, P. F., Bartoli, A. and Davison, A. J., 2012. KAZE features. *Proceedings of the European Conference on Computer Vision*, Vol. VI, pp. 214–227.

Alcantarilla, P. F., Nuevo, J. and Bartoli, A., 2013. Fast explicit diffusion for accelerated features in nonlinear scale spaces. *Proceedings of the British Machine Vision Conference*, pp. 13.1–13.11.

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

Calonder, M., Lepetit, V., Strecha, C. and Fua, P., 2010. BRIEF: binary robust independent elementary features. *Proceedings of the European Conference on Computer Vision*, Vol. IV, pp. 778–792.

Dahl, A. L., Aanæs, H. and Pedersen, K. S., 2011. Finding the best feature detector-descriptor combination. *Proceedings of the International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission*, pp. 318–325.

Hartley, R. and Zisserman, A., 2008. Multiple view geometry in computer vision. University Press, Cambridge, UK.

Jazayeri, I. and Fraser, C.S., 2008. Interest operators in close-range object reconstruction, *ISPRS Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXVII-B5*, pp. 69-74.

Levi, G. and Hassner, T., 2015. LATCH: Learned Arrangements of Three Patch Codes, *arXiv preprint* arXiv:1501.03719.

Li, H and Hartley, R., 2006. Five-Point Motion Estimation Made Easy. *Proceedings of the 18th International Conference on Pattern Recognition, ICPR*, pp.630-633.

Lowe, D. G., 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60(2), pp. 91–110.

Morel, J-M. and Yu, G., 2009. ASIFT: A New Framework for Fully Affine Invariant Image Comparison. *SIAM Journal on Imaging Sciences* 2(2): 438–469.

Nex, F., Gerke, M, Remondino, F., Przybilla, H.-J., Bäumker, M. and Zurhorst, A., 2015. ISPRS Benchmark for multi-platform photogrammetry. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. II-3/W4*, pp. 135-142.

Nistér, D., 2004. An efficient solution to the five-point relative pose problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26(6), pp. 756–770.

Rosten, E. and Drummond, T., 2005. Fusing points and lines for high performance tracking. *Proceedings of the International Conference on Computer Vision*, Vol. 2, pp. 1508–1515.

Rublee, E., Rabaud, V., Konolige, K. and Bradski, G., 2011. ORB: an efficient alternative to SIFT or SURF. *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2564–2571.

Urban, S. and Weinmann, M., 2015. Finding a good feature detector-descriptor combination for the 2D keypoint-based registration of TLS point clouds. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. II-3/W5*, pp. 121–128.

Wallis, K.F. (1976) Seasonal adjustment and relations between variables *Journal of the American Statistical Association*, 69(345) pp. 18-31.

Yang, X. and Cheng. K-T., 2012. LDB: An ultra-fast feature for scalable augmented reality on mobile devices. *IEEE Int. Symposium on Mixed and Augmented Reality*, pp. 49-57.