

IMPROVING HYSPEX SENSOR CO-REGISTRATION ACCURACY USING BRISK AND SENSOR-MODEL BASED RANSAC

P. Schwind^{1a,*}, M. Schneider^{1a}, R. Müller^{1a}

^{1a}German Aerospace Center (DLR), EOC, 82234 Oberpfaffenhofen, Germany -
(peter.schwind, mathias.schneider, rupert.mueller)@dlr.de

Commission I, WG I/4

KEY WORDS: Matching, Estimation, Orthorectification, Photogrammetry, Hyper spectral

ABSTRACT:

In this paper a method to improve the co-registration accuracy of two separate HySpex SWIR and VNIR cameras is proposed. The first step of the presented approach deals with the detection of point features from both scenes using the BRISK feature detector. After matching these features, the match coordinates in the VNIR scene are orthorectified and the resulting ground control points in the SWIR scene are filtered using a sensor-model based RANSAC. This implementation of RANSAC estimates the boresight angles of a scene by iteratively fitting the sensor-model to a subset of the matches. The boresight angles which can be applied to most of the remaining matches are then used to orthorectify the scene. Compared to previously used methods, the main advantages of this approach are the high robustness against outliers and the reduced runtime. The proposed methodology was evaluated using a test data set and it is shown in this work that the use of BRISK for feature detection followed by sensor-model based RANSAC significantly improves the co-registration accuracy of the imagery produced by the two HySpex sensors.

1. INTRODUCTION

To prepare for the hyperspectral satellite mission EnMAP (Environmental Mapping and Analysis Program) (Storch et al., 2010), the German Aerospace Center (DLR) acquired two HySpex cameras from Norwegian company NEO in 2011 (Baumgartner et al., 2012). For now these cameras are mainly used to develop new and optimize existing processing methods for hyperspectral data, but in the future they will also serve to validate EnMAP data.

Similar to EnMAP, the used HySpex configuration uses two separate sensors to acquire images in the SWIR and VNIR range of the electromagnetic spectrum. The two cameras have slightly differing view angles, meaning they don't acquire the same area on the ground at exactly the same time. As the cameras are mounted onto an airborne platform, the attitude during this fraction of time can deviate significantly. The extent of these attitude deviations can be recognized visually at the wavy boundaries of the scene shown in Figure 1. In combination with the view angle difference, these deviations lead to geometric co-registration inaccuracies between the SWIR and VNIR imagery. To correct these inaccuracies, the boresight angles between the two sensors have to be estimated precisely and the two scenes need to be orthorectified with these angles taken into account. To estimate the boresight angles, first of all the two scenes need to be matched using a feature-detector. As the two sensor have spectrally overlapping channels, depending on the scene content, this step should usually result in a considerable amount of matches. After removing matching outliers and by applying the known sensor model, the boresight angles between the sensors can be estimated and the scenes can finally be orthorectified (Müller et al., 2005).

There are several well-known matching algorithms which could possibly be used for the extraction of feature matches between the two scenes, e.g. Förstner (Förstner and Gülch, 1987), SIFT (Lowe, 2004), SURF (Bay et al., 2006). For this application however, we decided to test a comparably new method called BRISK

(Binary Robust Invariant Scalable Keypoints) (Leutenegger et al., 2011). While the detection stage of BRISK is essentially a FAST detector (Rosten and Drummond, 2006) applied to a scale-space, the novelty of BRISK lies mostly in its binary descriptor which is based on gradients computed within a fixed sampling pattern. As the BRISK algorithm was designed with a focus on limited complexity, its developers promise similar or better matching performance than state-of-the-art algorithms such as SIFT or SURF at a significantly reduced runtime. One of the goals of this work was to investigate if these expectations can be met when applying the algorithm to HySpex data.

Due to the similarity between spectrally overlapping channels, it is often possible to find thousands of matches for one HySpex scene. In previous comparable processing chains, we removed outliers within these matches iteratively using an algorithm denoted ESTIMATE hereafter. In each iteration, the boresight angles were estimated by fitting the known sensor model to the matches and the match with the maximum deviation from this fit was removed. This process was repeated until the deviation of the worst match remained below a given threshold. While this method works very reliably if the number of outliers is not too big, due to the iterative approach, the turnaround time (TAT) for thousands of matches can be very long. To reduce the turnaround time of the entire processing chain, this step was replaced by a RANSAC (RANDOM SAMPLE CONSENSUS) (Fischler and Bolles, 1981) based approach. Instead of fitting the sensor model to all matches, it is fitted to a randomly selected subset and the resulting boresight angles are then tested against all matches. After a limited number of iterations, the estimated boresight angles which can be applied to most matches are accepted as the correct parameters. The two main advantages of this approach compared to ESTIMATE are a reduction in runtime and a higher robustness against outliers in the match set.

Once the boresight angles have been estimated using the proposed approach, the images of both sensors can be orthorectified independently and, assuming the matching worked, should be well registered to each other. To evaluate if this is the case, the

*Corresponding author



Figure 1: Orthorectified SWIR channel

entire processing chain was tested rigorously.

In the following sections, first the main objectives of this work are defined (Section 2.). In Section 3., the technical details of the used sensors as well as the applied methods are laid out, followed by Section 4. where it is explained how these methods are combined and implemented in practice. Finally the proposed algorithm is evaluated in Section 5. and the obtained results are discussed in Section 6..

2. OBJECTIVES

Since the co-registration accuracy between HySpex SWIR and VNIR data is not sufficient when using laboratory and in-flight measurements alone, the goal of this work is to find a way to improve this accuracy substantially. To achieve this, in this paper we propose matching the two scenes using BRISK and estimating the boresight angle difference between them by using RANSAC in combination with the sensor-model of HySpex. Subsequently, a detailed evaluation is performed to measure if and how much the co-registration accuracy between the HySpex SWIR and VNIR sensors can be improved using the proposed methodology.

3. TECHNICAL BACKGROUND

In this section, first the special characteristics of the HySpex are specified, to clarify why this comprehensive processing approach is necessary. Then a short overview of the BRISK detector is provided, followed by an outline of the developed sensor-model based RANSAC.

3.1 HySpex

The company NEO offers five different cameras under the HySpex brand. To prepare for the upcoming hyperspectral satellite mission EnMAP, DLR acquired the VNIR-1600 and SWIR-320m-e models with a FOV expander in 2011. For the specifications of the cameras refer to Table 1. The cameras are installed on an airplane with slightly differing viewing angles, meaning they acquire the same areas on the ground with a small time difference. As the airborne platform is not perfectly stable, the attitude during this fraction of time can deviate significantly. The wavy boundaries of the orthorectified scene in Figure 1 illustrate this instability. In combination with the view angle difference, this time difference leads to geometric co-registration inaccuracies between the SWIR and VNIR imagery. An example showing this co-registration inaccuracy for an orthorectified SWIR and VNIR scene is depicted in Figure 2.

To improve the accuracy, the precise boresight angles between the two scenes need to be considered during orthorectification. It is possible to estimate the boresight angles of a sensor by fitting the known sensor model of the camera to ground control points located in the scene. In this work, by orthorectifying one of the two HySpex scenes, it is used as ground control for the other



Figure 2: Subsection of uncalibrated SWIR and VNIR scenes

Table 1: HySpex specifications

	VNIR-1600	SWIR-320m-e
Spectral range (nm)	415 – 991	967 – 2496
Bands	160	256
Spatial pixels	1600	320
FOV across track /with expander (°)	17/34	13.5/27
Pixel FOV (across/along track; mrad)	0.18/0.36	0.75/0.75

scene. Two channels from both sensors with similar wavelength are matched using BRISK, and the boresight angles are estimated by applying the sensor-model based RANSAC algorithm to the matches.

3.2 BRISK

The BRISK algorithm used for the matching of the SWIR and VNIR scenes is a feature detector with a working principle similar to better known algorithms such as SIFT or SURF. Just like these algorithms, BRISK can be subdivided into 4 main processing steps:

1. Scale space Keypoint Detection
2. Keypoint filtering and sub-pixel localization
3. Orientation assignment
4. Descriptor generation

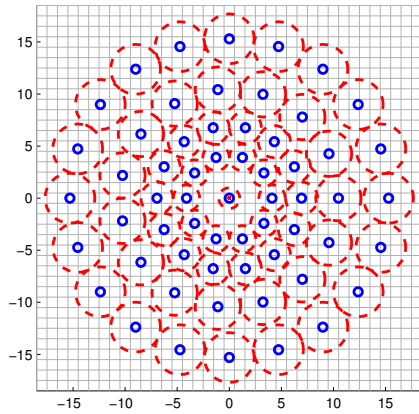


Figure 3: The sampling pattern used to estimate the orientation of a keypoint and to generate the descriptor. The blue circles are the centers of the sampling locations, the red circles represent areas filtered using Gaussian smoothing (Leutenegger et al., 2011).

First a scale space is created by repeatedly down-sampling the input image. Keypoint candidates in this scale space are selected using the FAST 9-16 detector (Rosten and Drummond, 2006). Next the keypoints are filtered by performing a 3D non-maxima suppression within the scale space pyramid. The positions of the remaining keypoints are then interpolated with sub-pixel accuracy within the pyramid. For this purpose a 2D quadratic function is fitted separately to the 3×3 region surrounding a keypoint as well as the corresponding regions in the neighbouring layers of the scale space. The resulting 3D maxima are then used for a 1D parabola fit to determine the 3D sub-pixel location in the scale space. For the orientation assignment and the descriptor generation a fixed sampling pattern is used (see Figure 3). The orientation is calculated by computing the average gradient of all long distance pairs within this pattern. *Long distance pair* in this case refers to all pairs with a distance bigger than $13.67t$ (t being the scale of the keypoint). Finally the descriptor is generated by rotating the sampling pattern by the estimated orientation and comparing all 512 pair combinations of this pattern with a distance smaller than $9.75t$. This finally results in a binary descriptor with 512 entries.

This brief description of the BRISK detector omits some of the technical details of the full algorithm. For a detailed description, please refer to (Leutenegger et al., 2011).

In practice the BRISK detector has 3 parameters which should be specified for each run. A threshold for the FAST detector, the number of octaves that should be created within the scale space and the matching ratio used to specify how different the two closest matching candidates for any keypoint have to be to accept the first candidate as a match. As the presented method is meant to be used within an operational processing chain, the selected parameters were chosen with a focus on automation. For the FAST threshold a very low value of 2 was used. As this would always result in a very big number of keypoints only the 400 000 matches with the best FAST response were finally used for the matching. The advantage of this approach is, that the FAST threshold does not have to be fine-tuned manually for every image pair, meaning the 400 000 best keypoints will always be selected for each scene automatically. The number of pyramid octaves is set to the empirically determined value of 4 and to obtain as many matches as possible the matching ratio is set to 0.7. A small drawback of this relatively high ratio is that even though it usually increases the total number of correct matches it also increases the number

and ratio of mismatches. However, as will be shown in Section 5., RANSAC is robust enough to deal with these outliers.

3.3 SENSOR-MODEL BASED RANSAC

RANSAC is an algorithm frequently used in computer vision and remote sensing applications to remove outliers from a set of matches. Even though most implementations probably use an affine or polynomial transformation model for RANSAC, depending on the application, any kind of mathematical model can be used to transform the matches from one image to the other. The most accurate model available for the used data in this case is the geometric sensor model of HySpex (see Figure 4). Using this model the geometric mapping in Earth-Centered Earth-Fixed (ECEF) coordinates r_{Object}^{ECEF} of each pixel of an image can be derived from the position of the sensor r_{Sensor}^{ECEF} and the Line of Sight (LoS) vector r_{LoS}^{ECEF} :

$$r_{Object}^{ECEF} = r_{Sensor}^{ECEF} + \lambda \cdot r_{LoS}^{ECEF} \quad (1)$$

The scale factor λ can be computed by intersecting r_{LoS}^{ECEF} with the Earth ellipsoid or a DEM. The Line of Sight (LoS) vector r_{LoS}^{ECEF} itself can be calculated using

$$r_{LoS}^{ECEF} = R_{Body}^{ECEF} \cdot R_{Sensor}^{Body} \cdot r_{LoS}^{Sensor} \quad (2)$$

where R_{Body}^{ECEF} is the rotation matrix used to change from the body to the ECEF coordinate system, R_{Sensor}^{Body} is the rotation matrix from the sensor to the body coordinate system and r_{LoS}^{Sensor} is the interior orientation. The position r_{Sensor}^{ECEF} is known for each scene from in-flight measurements, while the interior orientation r_{LoS}^{Sensor} is known for each pixel of the scan line from laboratory measurements. The rotation matrix R_{Sensor}^{Body} , which includes the boresight angles, is also known to some degree from pre-flight measurements, but to achieve an accurate co-registration, the boresight angles have to be estimated more precisely after image acquisition.

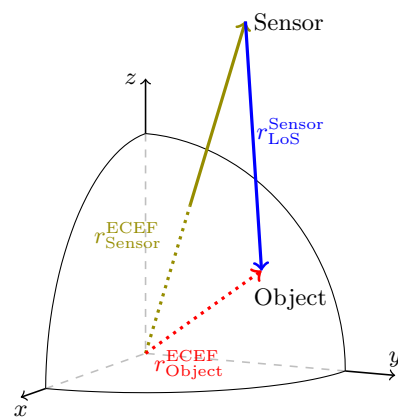


Figure 4: The used sensor model

To obtain the boresight angle difference between SWIR and VNIR using sensor-model based RANSAC, initially it is necessary to have geo-coordinates in one scenes (VNIR) and their corresponding pixel-coordinates in the other scene (SWIR). Next, using the sensor model of the SWIR sensor, two randomly selected matches of the SWIR scene are orthorectified. The boresight angles used for these two matches are estimated by minimizing the sum of squares to the corresponding geo-coordinates of the matches in the VNIR scene. If the least squares optimization succeeds, the sensor model using the estimated boresight angles is applied to

all remaining matches in the SWIR scene and the resulting geo-locations are compared to the positions of the matches in the VNIR scene. Then, all matches with a distance bigger than a set threshold (measured in meters) are discarded. This entire process is repeated for a limited number of iterations and the boresight angles candidates which can be applied to the most matches are kept as the final parameters.

As the focus of this work is to improve only the relative co-registration accuracy between the two HySpex instruments, only the boresight angle difference between SWIR and VNIR has to be estimated. Even though this was not done for this work, it should be noted that the absolute boresight angles could be estimated using the same methodology using a reference image.

4. METHODOLOGY

The algorithms described above are combined in the processing chain depicted in Figure 5 to process the SWIR and VNIR images. First, point matches between two acquired scenes are computed using the BRISK feature detector. Even though the original scenes show significant distortions because of attitude deviations during the acquisition (see Figure 1), the matching process usually results in a considerable amount of matches, as the view angle difference between SWIR and VNIR is not very big, meaning both scenes feature very similar distortions. To further improve the matching performance, two channels acquired at a similar wavelength are matched with each other and the size of the SWIR scene is increased to match the dimensions of the VNIR scene. Although the BRISK detector is actually scale invariant, bringing the two scenes to the same geometry allows matching without rotation and scale invariance which improves the number and ratio of correct matches somewhat. After matching, the VNIR scene is orthorectified using the sensor-model and the attitude and position values measured in-flight. The coordinates of the matches in the VNIR scene are also orthorectified during this processing step. The VNIR geo-coordinates and the corresponding SWIR pixel coordinates are then used as input for the RANSAC processing, where outliers are removed and the boresight angle difference between the SWIR/VNIR sensors is estimated. Finally the estimated boresight angles are used to orthorectify the SWIR scene. The end result of this entire process are two SWIR and VNIR ortho images in the same coordinate system.

As mentioned before (see Section 3.3) either a DEM or the Earth ellipsoid can be used for the calculation of the sensor model. Using a detailed DEM for the processing is of course preferable as this will lead to more accurate results in the estimation process. Unfortunately a DEM is not always available. That is why the implemented method allows processing with or without a DEM. In fact if no DEM is provided it is even possible to provide no height value at all for the processing. In this case the average height difference from the reference ellipsoid for the entire scene is estimated in addition to the boresight angles. As will be shown in Section 5., if a large number of matches is provided and the processed scenes feature mostly flat terrain, this *makeshift solution* works quite accurately, even when the incidence angle difference between the two scenes is very small.

4.1 Implementation

The BRISK detector used here was implemented in C++ as a module of the DLR in-house image processing software XDibias. For the matching of the BRISK descriptors the LSH (locality-sensitive hashing) algorithm in the publicly available FLANN (Fast Library for Approximate Nearest Neighbors (Muja and Lowe,

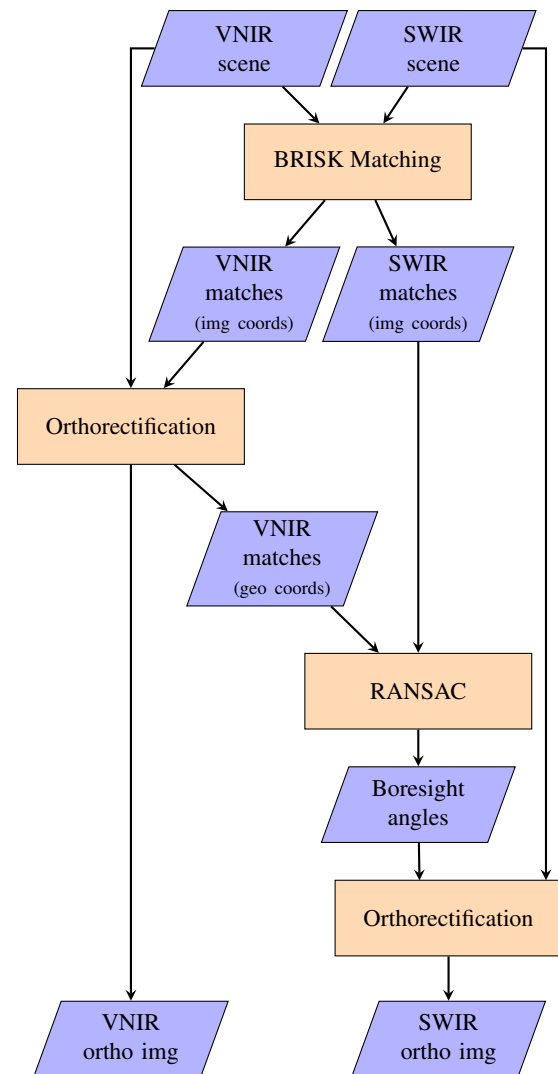


Figure 5: Methodology overview (without DEM)

2009)) library was used. Orthorectification was performed using the existing XDibias module ORTHO (Müller et al., 2002) and the sensor-model based RANSAC approach was implemented as a new XDibias module, reusing the same sensor-model functions already implemented for ORTHO.

5. EVALUATION

5.1 Methodology

The used evaluation approach is depicted in Figure 6. After processing a HySpex VNIR and SWIR scene using the methodology presented in Section 4. the resulting ortho-images are matched again using SIFT. The SIFT matches are used as input for an elaborate local least squares (LLSQ) fine matching which brings the matches to an accuracy of ~ 0.1 pixels and eliminates mismatches. For the remaining matches, the root mean squared error (RMSE) is calculated to determine how accurately the two scenes are co-registered.

For comparison this entire procedure was performed with (RANSAC+DEM) and without DEM (RANSAC) and using the program ESTIMATE (see Section 1.) for the derivation of the boresight angles. For all three methods a threshold of 2 m was used to remove outliers from the matches.

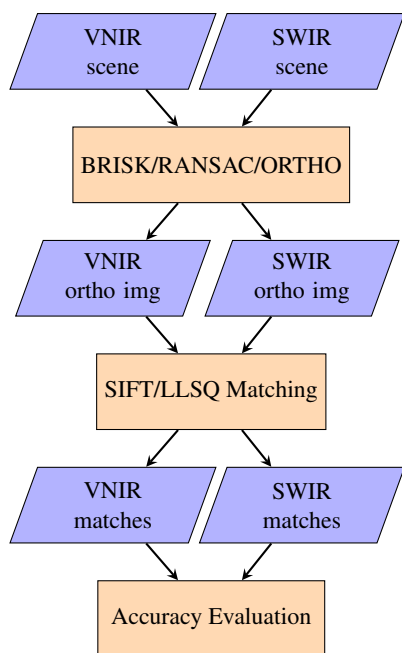


Figure 6: Evaluation approach

5.2 Dataset

The used HySpex test scene was acquired on 18th of June, 2014 over the city of Kaufbeuren, Germany and features both urban and rural areas. It was recorded from a height of ~ 1 km resulting in a ground sampling distance of ~ 0.5 m for VNIR and ~ 1.25 m and ~ 2.5 m for SWIR in along and across-direction respectively. The VNIR scene covers 1600×7800 pixels (width \times height) and the SWIR scene 320×3900 pixels.

The DEM used for the orthorectification and RANSAC is based on stereo imagery acquired using the 3K camera system (Kurz et al., 2012). It was generated using the semi-global matching based processor presented in (d'Angelo and Reinartz, 2011) and has a GSD of 25 cm.

5.3 Results

For the tested scenes 37 936 matches are found in 141 s using BRISK (see Figure 7). As shown in Table 2, 34 842, 28 080 or 15 505 of these matches remain when using RANSAC+DEM, RANSAC and ESTIMATE respectively. Using RANSAC without a DEM performs the most quickly with only 59 s. Using a DEM for RANSAC slows down the processing by a factor of ~ 25 to 1583 s. However even that is still almost twice as fast as the iterative ESTIMATE approach which takes 2725 s. The calculated boresight angles are very similar for all three methods. Only for the RANSAC case without DEM, the yaw angle deviates somewhat. On a side note, as mentioned in Section 4, RANSAC without a DEM estimated the height of the scene. In this case, after using a start value of 0 m for the least squares estimation, the average height estimated for all the matches was 754.9 m. According to the 3K DEM, the real average height of the area is 762.2 m. Considering the small view angle difference between the SWIR and VNIR cameras, the accuracy of the height estimation in this case is very good.

Using the estimated boresight angles shown in Table 2, both HySpex scenes were orthorectified and the resulting scenes were matched using SIFT and LLSQ. In Table 3 the number of matches and the root-mean-square error (RMSE) for each VNIR/SWIR

Table 2: Processing Results

	Matches (#)	TAT (s)	Boresight angles (°) (Roll/Pitch/Yaw)
RANSAC+DEM	34 842	1483	-1.52/-0.50/-0.45
RANSAC	28 080	59	-1.53/-0.48/-0.33
ESTIMATE	15 505	2725	-1.51/-0.48/-0.40

Table 3: Accuracy Evaluation

	Matches (#)	RMSE (m) (across/along)
Without improvement	3398	26.91/11.81
RANSAC+DEM	4450	0.79/0.66
RANSAC	4207	0.94/0.97
ESTIMATE	4264	0.64/0.88

pair is listed. Without any kind of calibration, the scenes have an RMSE of 26.91 m and 11.81 m in across and along track direction respectively. This error can clearly be seen in Figure 2. The RMSE values for the 3 corrected SWIR/VNIR pairs are all very similar. Only for RANSAC without a DEM the error is somewhat higher compared to the other two cases. Nevertheless in all cases the errors lie below the subpixel level. In Figure 8, the accuracy of the scenes corrected using the RANSAC+DEM boresight angles is illustrated.

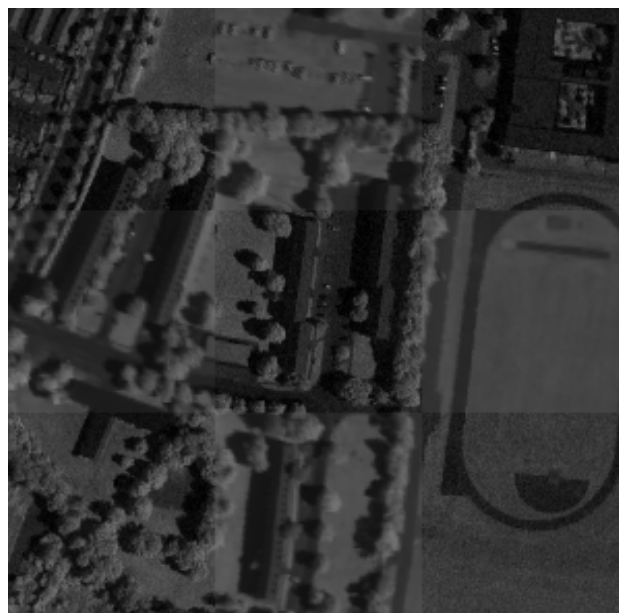


Figure 8: Subsection of co-registered SWIR and VNIR scenes

6. DISCUSSION AND CONCLUSIONS

As the boresight angles derived by all three procedures are very similar, the resulting accuracies of the corrected scenes do not vary considerably. The remaining small error is probably a result of sensor model, matching as well as DEM inaccuracies. In addition to that the LLSQ matching used for evaluation also has an inaccuracy of ~ 0.1 pixels. Nevertheless, compared to the RMSE of 26.91 m and 11.81 m for the uncalibrated scenes, the remaining errors are negligible.

While the accuracy relative to the previously used ESTIMATE

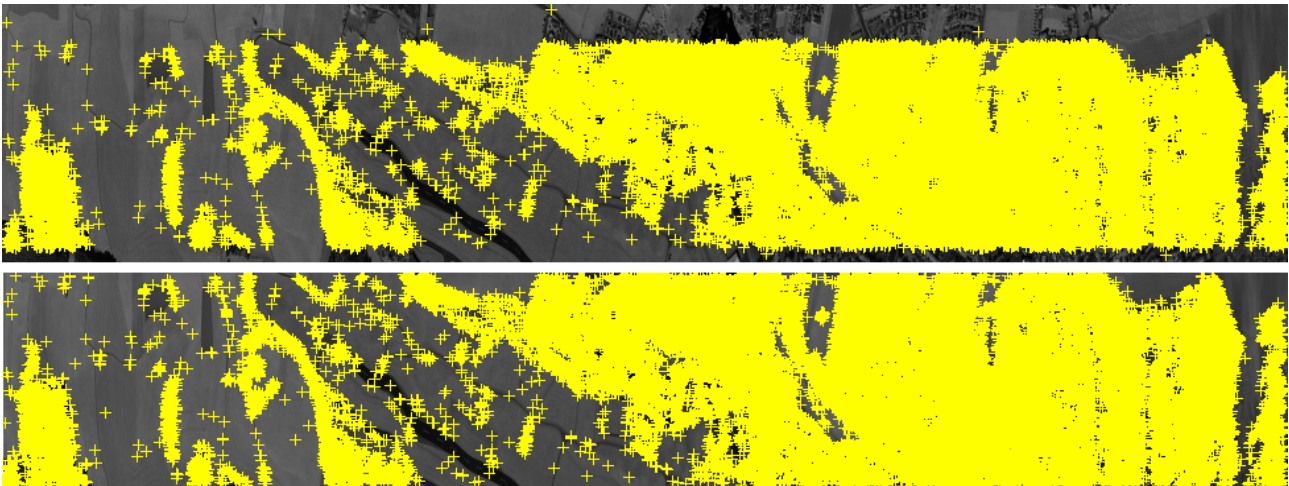


Figure 7: BRISK matches found in VNIR(top) and SWIR(bottom; scaled up) scenes

program does not really change, the reduction in the turnaround time is significant. Especially when using RANSAC without a DEM, a 46 times shorter turnaround time is achieved with a minor reduction in accuracy. Another advantage of RANSAC, which does not really take effect in this use case, is the higher robustness against outliers. As ESTIMATE uses all points to build a model each iteration before removing the point with the highest deviation, a high proportion of outliers might result in correct matches being eliminated. For RANSAC, in theory the correct matches can always be found with a high probability even for a high proportion of outliers. In practice of course this becomes impractical at some point as a higher proportion of outliers means that more iterations have to be performed by RANSAC.

Concerning BRISK, which was not evaluated in detail in this work, the fact that less than 10% of the matches are mismatches shows that the matching algorithm provides highly reliable matches. This is especially significant as all the BRISK parameters were selected with a focus on automation of the processing chain instead of fine-tuning them for this application. For the future, a more detailed analysis of the applicability of the BRISK detector to remote sensing data is planned.

REFERENCES

- Baumgartner, A., Gege, P., Köhler, C., Lenhard, K. and Schwarzmaier, T., 2012. Characterisation methods for the hyperspectral sensor HySpex at DLR's calibration home base. In: *SPIE Remote Sensing 2012*, pp. 1–8.
- Bay, H., Tuytelaars, T. and Van Gool, L., 2006. SURF: Speeded Up Robust Features. In: *Computer Vision—ECCV 2006*, Springer, pp. 404–417.
- d'Angelo, P. and Reinartz, P., 2011. Semiglobal Matching Results on the ISPRS Stereo Matching Benchmark. In: C. Heipke, K. Jacobsen, F. Rottensteiner, S. Müller and U. Sörgel (eds), *High-Resolution Earth Imaging for Geospatial Information*, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Institute of Photogrammetry and GeoInformation, Leibniz Universität Hannover, pp. 1–6.
- Fischler, M. A. and Bolles, R. C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM* 24(6), pp. 381–395.
- Förstner, W. and Gülch, E., 1987. A fast operator for detection and precise location of distinct points, corners and centres of circular features. In: *Proc. ISPRS intercommission conference on fast processing of photogrammetric data*, pp. 281–305.
- Kurz, F., Türmer, S., Meynberg, O., Rosenbaum, D., Runge, H., Reinartz, P. and Leitloff, J., 2012. Low-cost optical camera systems for real-time mapping applications. *Photogrammetrie-Fernerkundung-Geoinformation* 2012(2), pp. 159–176.
- Leutenegger, S., Chli, M. and Siegwart, R. Y., 2011. BRISK: Binary Robust Invariant Scalable Keypoints. In: *Computer Vision (ICCV), 2011 IEEE International Conference on*, IEEE, pp. 2548–2555.
- Lowe, D. G., 2004. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision* 60(2), pp. 91–110.
- Muja, M. and Lowe, D. G., 2009. Fast approximate nearest neighbors with automatic algorithm configuration. In: *International Conference on Computer Vision Theory and Application VISS-APP'09*, INSTICC Press, pp. 331–340.
- Müller, R., Lehner, M., Müller, R., Reinartz, P., Schroeder, M. and Vollmer, B., 2002. A program for direct georeferencing of airborne and spaceborne line scanner images. In: *Integrating Remote Sensing at the Global, Regional and Local Scale*, Denver CO, Nov. 10-15 2002, Vol. XXXIV, Part 1; Commission I, Vol. 1, pp. 148–153.
- Müller, R., Lehner, M., Reinartz, P. and Schroeder, M., 2005. Evaluation of spaceborne and airborne line scanner images using a generic ortho image processor. In: *Proc. of High Resolution Earth Imaging for Geospatial Information*, ISPRS Hannover Workshop, Commission I WG, Vol. 5, p. 2005.
- Rosten, E. and Drummond, T., 2006. Machine learning for high-speed corner detection. In: *Computer Vision—ECCV 2006*, Springer, pp. 430–443.
- Storch, T., Eberle, S., Makasy, C., Maslin, S., de, A. M., Mißling, K.-D., Mühle, H., Müller, R., Engelbrecht, S., Gredel, J. and Müller, A., 2010. On the design of the ground segment for the future hyperspectral satellite mission EnMAP. In: *IEEE Aerospace Conference Proceedings*, Big Sky, MT, USA, 6-13 March, pp. 1–11.