IMAGE MATCHING ERROR DETECTION WITH FOCUS ON MATCHING OF SAR AND OPTICAL IMAGES

Soukal Peter, Baltsavias Emmanuel *

Institute of Geodesy and Photogrammetry, ETH Zurich, Wolfgang-Pauli-Str. 15, 8093 Zurich, Switzerland
(psoukal@student.ethz.ch , manos@geod.baug.ethz.ch)

KEY WORDS: Multisensor Data, Optical, SAR, Matching, Least Squares Matching, Matching Quality Characterisation, Matching Error Detection

ABSTRACT: In this paper, we present matching of multitemporal optical and SAR (Synthetic Aperture Radar) images of various spatial resolutions and naturally different bands. The aims of such matching could be various, like image co-registration, needed for various applications, where a combination of SAR and optical images is needed, as. e.g. in disaster and hazard monitoring and mapping. Here, the application idea is to transfer GCPs (Ground Control Points) from SAR to optical images. SAR images, at least with the newest satellite platforms, can be georeferenced without the use of GCPs down to an accuracy of few dm, while optical sensors, especially high-resolution ones, often need GCPs to achieve the highest possible accuracy. The SAR images are assumed to be orthorectified and a related DSM (Digital Surface Model) exists. Thus, through matching, we can automatically determine planimetric coordinates of GCPs for optical images, which are needed for an accurate optical image georeferencing, while the height is interpolated from the given DSM. However, matching SAR and optical images is so difficult that the great majority of matching results are wrong (only 1 out of about 1,000 match points is correct). In these investigations, we used LSM (Least Squares Matching), as it is very accurate, but most importantly it can provide different statistics, which can be used for matching error detection and elimination. Thus, we developed novel methods for an LSM matching quality characterization and error detection. The same methods, with some minor modifications, can be used for quality characterization for various other matching methods, like cross-correlation, or even feature-based matching. For matching of SAR and optical images, very variable test datasets were used and the results are based on visual inspection of the match-points, very promising, with poorer performance for low spatial resolution images of about 25 m GSD (Ground Sampling Distance). The same matching error detection method has been used for a much more important application, the automatic DSM generation via image matching. Preliminary results show that performing such tests after matching lead to a much more accurate DSM (error standard deviation decreased from 9.3 m to 1.3m from no error detection to the method proposed below). They also provide an accuracy measure for each DSM point, which could be used for various purposes, among them DSM fusion.

1. INTRODUCTION AND AIMS

The primary aim of this work was initially to do automatic point transfer via image matching between SAR and optical images. However, it was shown that our methods had a much wider significance, namely the automatic error detection in matching of optical images for automated DSM generation, focussing here on use of Least Squares Matching. In our concrete application, the SAR images are orthorectified and have an underlying DSM. SAR orthorectification can be performed without GCPs, with few exceptions (especially for older sensors). The point transfer then provides GCPs for optical image georeferencing, avoiding the costly and difficult GCP acquisition by other means. About 20-40 (usually well-distributed) points are needed, half of them as GCPs, the rest as check points. We aim at matching (point transfer) accuracy of about 0.5-2 pixels. The same procedure is needed for other purposes, e.g. co-registration of SAR and optical images (e.g. in fast response applications, as post-hazard mapping), where optical and SAR images need to be fused, whereby optical images provide better interpretability and usually resolution, while SAR weather and day/night independence. SAR and optical images are spectrally and geometrically very different and have multitemporal and other (e.g. noise characteristics) differences, making matching extremely challenging. Only a very small % of the matches (0.01 to 1%) is correct. Thus, we need to derive quality values for each match-point and based on that do error detection and keep only the N best points needed. This finally turned out to be the most important and challenging part of the work, and not the matching itself. It also has a much larger significance, as the same quality characterization and error detection methods can be used in matching of optical images for automatic DSM generation, or matching of SAR images for other purposes, using the same matching method as here, and with adaptations also for other matching methods, e.g. using cross-correlation. This can solve the problem of deriving quality values for each match-point, which is essential not only for error detection and reliable automated processing, but also other applications like data (e.g. DSM) fusion.

* Corresponding author.

Previous research work performed on the paper topic is quite limited with major work given by Christophe et al. (2008), Dare and Dowman (2001), Goncalves and Dowman (2002) and Reinartz et al. (2001), while we do not know of any commercial software providing this functionality, at least reliably.

2. INPUT DATA

The 15 x 15 km^2 test area, close to the city of Thun, Switzerland, had substantial elevation differences and many land-cover classes (see Figure 1).

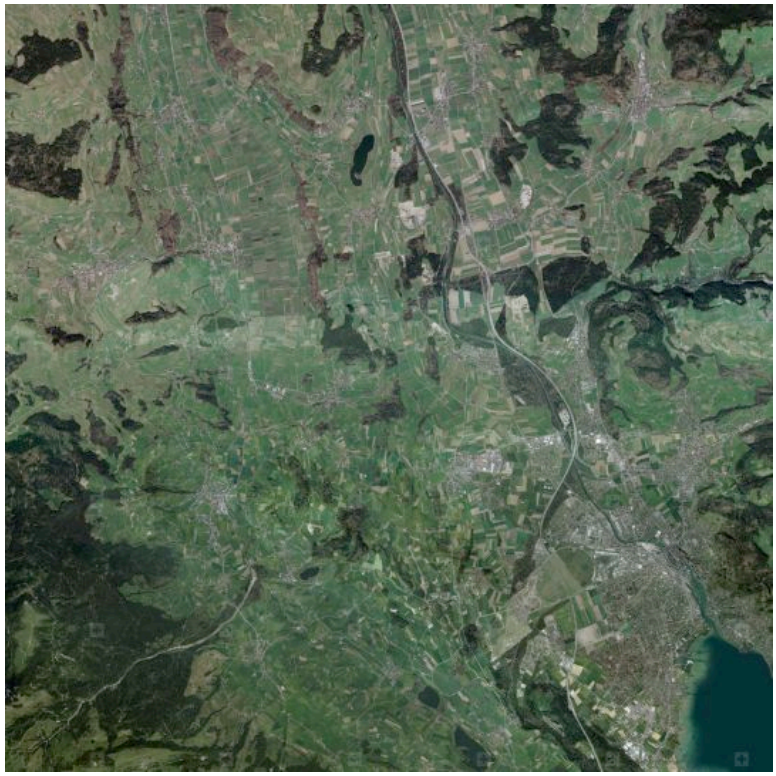

Figure 1: Common test area of SAR and optical images.

The area was imaged by five SAR and optical datasets each. These were combined in pairs, such that the GSD of SAR-Optical image pairs is similar (matching and GCP transfer makes no sense, if the GSD between SAR and optical images is too different). The major image parameters are listed in Table 1. All optical data was delivered as 8-bit and all SAR images had HH polarisation, except ERS-1, which had VV. A part of the overlapping area of the images is shown in Figure 2 for three SAR and three optical images.

Table 1: Major image parameters. For each platform/sensor in brackets, the band and GSD.

SAR Images	SAR Acquisition Date	Optical Images	Optical Acquisition Date
TerraSAR-X (X, 3m)	10.2008	ALOS/PRISM (PAN, 2.5m)	1.2006
TerraSAR-X (X, 3m)	10.2008	SPOT-5/HRG (PAN, 2.5m)	8.2005
RADARSAT-1 (C, 7m)	2.2008	LANDSAT-7/ETM+ (PAN, 15m)	3.2000
ALOS/PALSAR (L, 8m)	11.2008	LANDSAT-7/ETM+ (PAN, 15m)	3.2000
ENVISAT/ASAR (C, 25m)	4.2008	IRS-P6/LISS-3 (NIR, 23.5m)	7.2006
ENVISAT/ASAR (C, 25m)	4.2008	LANDSAT-7/ETM+ (R, 30m)	6.2001
ERS-1/AMI (C, 25m)	-	IRS-P6/LISS-3 (NIR, 23.5m)	7.2006
ERS-1/AMI (C, 25m)	-	LANDSAT-7/ETM+ (R, 30m)	6.2001

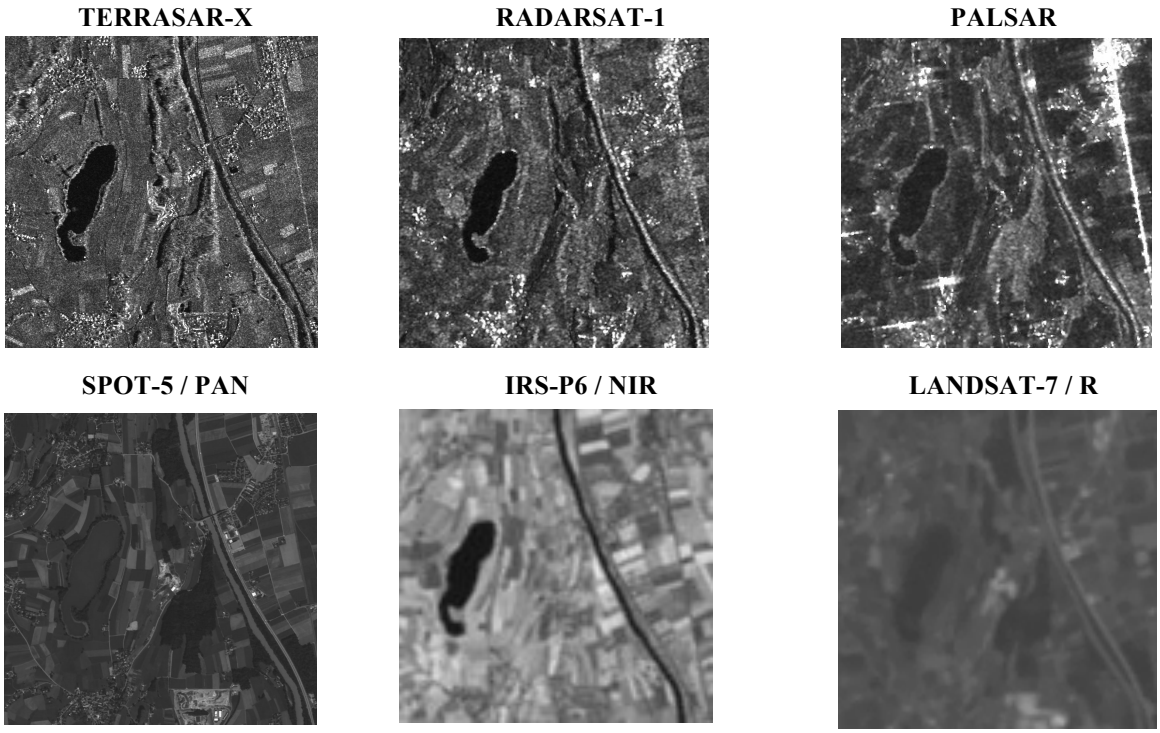

Figure 2: Examples of image parts from the different sensor data to be matched. Notice the huge differences not only between SAR and optical images (top to bottom), but also between SAR and between optical images.

3. METHODOLOGY

The above SAR-Optical pairs were processed using the steps as described below and in Figure 3.

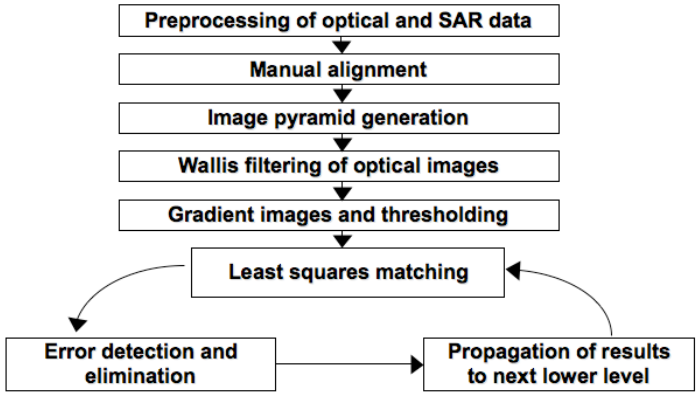

Figure 3: Overview of processing steps.

- Preprocessing to reduce noise (for SAR, especially speckle) ;
- Coarse co-registration of optical to SAR images (performed manually with a polynomial transformation (1-2 degree) and remaining maximum parallaxes in this case of ca. 38 pixels) ;
- Generation of image pyramids (using a 3x3 Gaussian filter, 5 pyramid levels and a resolution reduction factor 2, so that maximum parallaxes on the top pyramid level are 2 pixels, the latter required as approximations for the specific matching method used (LSM)) ;
- At each pyramid level, Wallis filtering (Wallis, 1976) for optical images to enhance contrast, especially in shadow areas, and get more edges (not implemented for SAR, as remaining speckle noise is large and non-salient edges would be much enhanced) ;
- At each pyramid level, computation of edge gradients (to increase similarity between SAR and optical images compared to original grey values) and adaptive thresholding of edge gradients (to reduce non-salient edges due to remaining noise), see Figure 4 ;

- At each pyramid level, before matching elimination of match-points with no or very weak texture (detecting poor x- or y- grey-level gradients, as there matching will fail) ;
- At each pyramid level, LSM (Gruen and Baltsavias 1988) and computation of quality measures, match-point quality characterisation, error detection and elimination, and except for the last pyramid level 0 (original images) derivation of approximations for the lower level ;
- At level 0, match-point quality characterisation, error detection and elimination, and final decision.

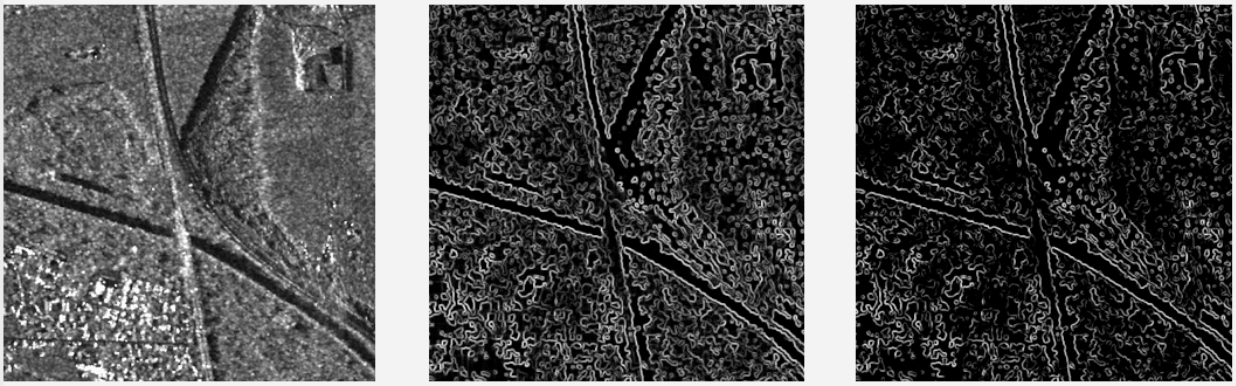

Figure 4: Left: original image (SAR) ; Middle: gradients ; Right: thresholded gradients.

Details (limited due to lack of space) on some of the above steps are listed below, others in Braun (2010) and Soukal (2012). Noise reduction includes an adaptive local method for optical images and anisotropic diffusion for SAR ones. Elimination of poor texture points before matching leads to 2-13% less match-points, increasing with lower level. LSM was used, as it is very accurate and provides many quality criteria, even if slow and requiring good approximations (a few pixels for the parallaxes, i.e. shifts of the affine transformation used in LSM). Patches of 13x13 pixels from the SAR image were used as reference and the optical image patches (search) were affine transformed to them in an iterative fashion, minimizing the squared sum of residuals (here edge gradient differences). All image points, apart a small image border, were matched. Although LSM is an area-based matching (ABM) method, using gradients for matching, makes it a hybrid approach, i.e. something between ABM and feature-based matching (FBM), where here features are edges. Preliminary investigations extracting points or edges with interest operators or edge detectors, showed that a classical FBM would be more difficult than our approach, as points are too many, not distinct and very different between SAR and optical images, while one-pixel-wide edges are not only different and less than our edge gradients but also introduce false parallaxes.

The 11 quality measures provided by LSM are: number of iterations, normalized cross-correlation computed separately after iteration convergence, sigma a posteriori, estimated values of 6 affine parameters and the standard deviations of the shifts (divided by sigma a posteriori). They are checked whether they follow a normal distribution, needed for deriving quality measure thresholds, and this was indeed the case, with the exception of the number of iterations, and much less some skewness in the normal distribution, caused by much more poor than good match-points in this matching application. Regarding these 11 quality measures, poor match-points are on one side of the distribution, except for the 6 affine parameters, which are two-sided, and the number of iterations, which for one iteration are also rather suspicious. For error detection after matching, we optionally use at each level the following tests:

1. Bi-directional matching consistency (matching search to reference patch and opposite, with a difference of more than 1 pixel between the two matchings for the reference point leading to rejection), points rejected are e.g. 39 - 72% for the Landsat/Radarsat pair, increasing as pyramid levels decrease.
2. Combination of quality measures. Each measure should "pass" in an OR fashion, whereby thresholds for each measure are derived using robust statistics, median and the robust approximation of the standard deviation, s(MAD), with MAD = median absolute deviation from the median and s(MAD) = 1.484 * MAD, and a threshold for each measure derived as (median +/- n * s(MAD)), with n derived based on long a priori knowledge and empirical tests. Points rejected after this test are e.g. 99.3- 93.4% for the Landsat/Radarsat pair, decreasing as pyramid levels decrease.
3. The measures of test 2 above are each given a normalised score (0-1) using a separate fuzzy function. Then, a single quality value is computed using a weighted averaging of each quality measure score, whereby weights for each criterion are determined empirically. This allows selecting the N best match-points, in this work the best 100.

4. At the end of the tests, the local consistency of the x- and y-parallax of a point is tested and if enough neighbours exist, the point is deleted or its parallaxes are corrected to the median parallaxes of the neighbourhood. The % of deleted or corrected parallaxes is small (< than 5%). Thus, this test could be generally omitted, especially if the previous ones perform well.

We evaluated two test combinations (1 and 2):

- Combination 1: tests 1, 2, 4 and ;
- Combination 2: test 3 (we tried also test 1 after 3 but too many points, many correct, were rejected, indicating the ad-hoc nature of this test).

The above tests can be performed in the whole image or in tiles to ensure good point distribution. Due to lack of space, we do not treat details of derivation of approximations for parallaxes (shifts of affine transformation), which are very crucial, especially for LSM, and influence how many pyramid levels are needed, how matching results are propagated if more than one level needed, and the number of match points at 0 pyramid level. As a short statement, we use for approximations' propagation only the good match-points (passing the error detection tests at each level), with a nearest neighbour interpolation of the parallaxes for the lower level points. For the second top level, we derive approximations and match all points, for lower levels only points within a neighbourhood (here 9x9 pixels) of each good match-point of the upper level.

4. RESULTS

Of the above 2 test combination methods, the 2nd is much more preferable and discussed below. It is faster (does not need the time-consuming bi-directional matching), provides more reliable results, a single quality value per point and gives control over the N best points kept. To ensure a good point distribution, which is often necessary for GCPS and georeferencing, for combination method 2, we select the best 100 points on pyramid level 0, plus the best point in image tiles (here 3x3) that are empty, i.e. have none of the N best points (see Figure 5).

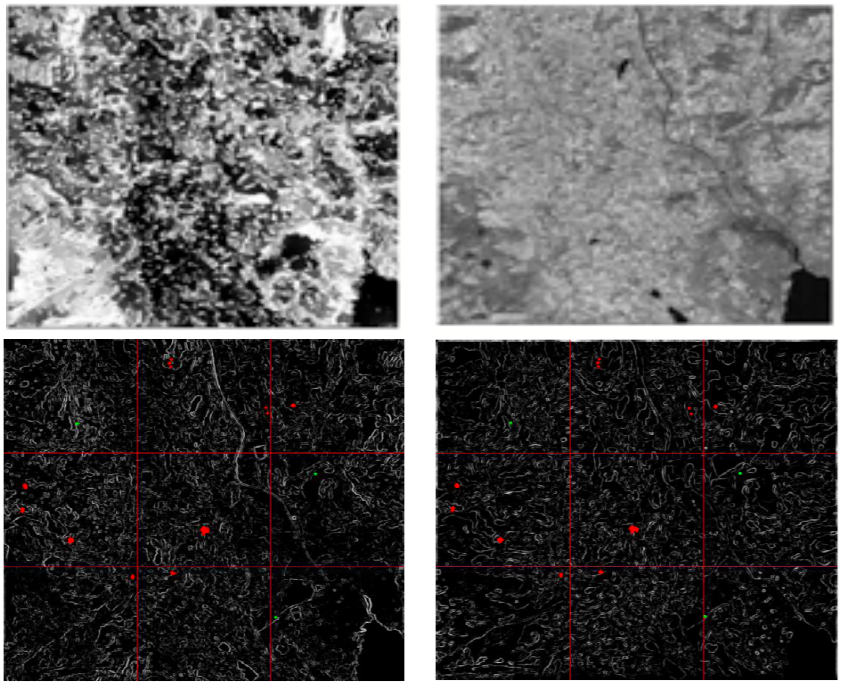

Figure 5: Radarsat-1 (left), Landsat-7/ETM+ (right), top grey level images, bottom thresholded gradient images with 100 best points (in red) and best point in empty 3x3 tiles (in green).

The evaluation of the results is very difficult as there are no reference data (can hardly be on image level, contrary to reference DSMs) and manual measurements are mostly even worse than matching results due the large differences between the matched images. Thus, we use a visual inspection of match-points using the original grey level images (not the image gradients) and a classification in % of correct, unsure and false matchings. The results are listed in Table 2. It is obvious that using low resolution images like ERS-1/Envisat with Landsat-7 (30m GSD)/IRS-P6 leads to worse quality. Regarding ERS-1 this is also due to the poor image quality (especially unsharpness) of the image. However, this visual inspection is subjective and very difficult, especially for the low

resolution images. Thus, to find the real quality of the matching results we plan in the near future to georeference the optical images with these 100+ and the 20 best points and analyse the residuals of GCPs and check points (in the latter case, when only 20 GCPs are used).

Table 2: Correctness of results on pyramid level 0 for all image pairs. With yellow and red, results are marked that are of inferior quality.

SAR - Optical Image Pair	No. of Best Match Points (Level 0)	Correct (%)	Unsure (%)	False (%)
TerraSAR-X (X, 3m) - ALOS/PRISM (PAN, 2.5m)	106	98.1	0.0	1.9
TerraSAR-X (X, 3m) - SPOT-5/HRG (PAN, 2.5m)	103	89.1	8.9	2.0
RADARSAT-1 (C, 7m) - LANDSAT-7/ETM+ (PAN, 15m)	102	98.0	2.0	0.0
ALOS/PALSAR (L, 8m) - LANDSAT-7/ETM+ (PAN, 15m)	104	100.0	0.0	0.0
ENVISAT/ASAR (C, 25m) - IRS-P6/LISS-3 (NIR, 23.5m)	103	57.3	35.0	7.7
ENVISAT/ASAR (C, 25m) - LANDSAT-7/ETM+ (R, 30m)	105	62.9	29.5	7.6
ERS-1/AMI (C, 25m) - IRS-P6/LISS-3 (NIR, 23.5m)	106	54.7	39.6	5.7
ERS-1/AMI (C, 25m) - LANDSAT-7/ETM+ (R, 30m)	104	0.0	52.0	48.0

5. EXTENSION OF THE MATCHING ERROR DETECTION METHOD TO DSM GENERATION

The above method was used for a much more important application, namely automatic DSM generation with matching of optical images. Unfortunately, most commercial, but also research, matching methods and respective software do not provide quality criteria for each match point (or even if they do, the criteria are not reliable), and thus they also do not provide reliable methods for matching error detection. In the test below, we used data from a thorough, well-organised, international test of the German Society of Photogrammetry and Remote Sensing (DGPF), where also accurate reference DSM data from airborne laser scanning (ALS) existed (Cramer, 2010). Among the available data, we selected some image parts (in this case DMC images) and performed LSM with the same program as above. Then, we (a) accepted all match-points ; (b) rejected points with a cross-correlation coefficient (a commonly-used criterion) of less than 0.6 and (c) accepted points only that passed the test 2 mentioned in Section 3 above. Naturally, the number of points accepted from (a) to (c) was decreasing from 220,000 to 151,000 points. The matched pixel coordinates were transformed via the given image orientation to 3D coordinates and these were compared to the 3D ALS reference data. Statistical results of the differences are listed in Table 3, and a colour-coded visualisation of the differences in Figure 6. As it can be seen from the Table 3 and Figure 6, the matching DSM after error detection by our method delivers much more accurate results with less blunders.

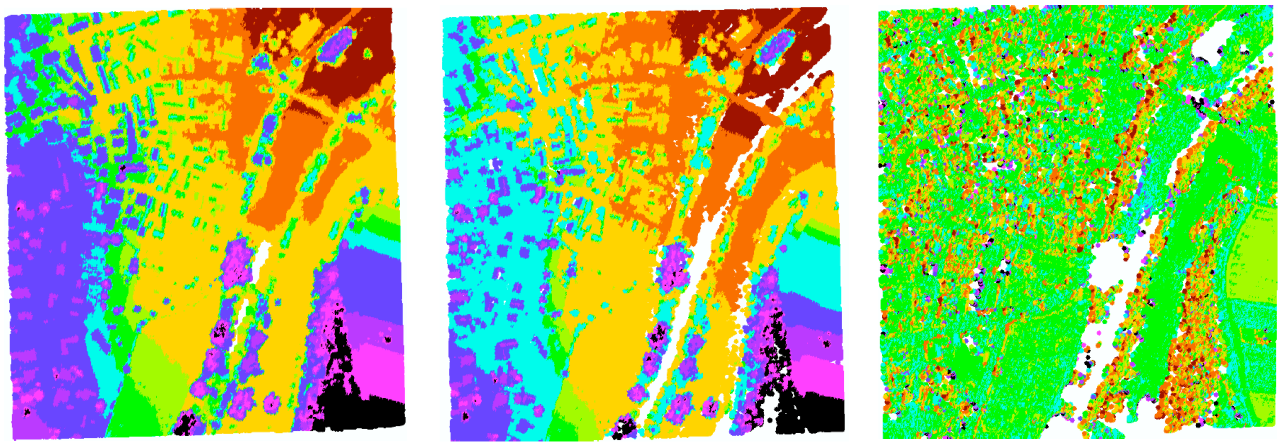

Figure 6: Colour-coded differences of matching DSM to reference ALS DSM. Left: No error detection ; Middle: Cross-correlation < 0.6 ; Right: Test no. 2 above. Greenish colours represent the smallest errors, white non-matched/rejected match-points. Note: although in the 3 figures the same numbers of intervals are used, their range varies, especially for the right figure, where the range of each colour is smaller.

Table 3: Image matching with three error detection methods, as compared to reference ALS data.

Error Detection	Mean (m)	Median (m)	Standard Deviation (m)
None	0.71	1.65	9.32
Cross-correlation < 0.6	0.30	1.08	9.30
Test no. 2 above	0.41	0.12	1.34

6. DISCUSSION AND CONCLUSIONS

The above results show that our approach delivers very good results, although the test data was very variable and difficult. The quality characterization and error detection for matching are innovative. Further research may include refinement of the above methods for error detection (use DSM slope and roughness, land-cover and a subsequent sensor orientation as mentioned above), estimation of absolute (instead of relative) single quality values (should be expressed for this case in pixels), avoidance of manual coarse co-registration (using approximate sensor position and attitude data, especially RPCs (Rational Polynomial Coefficients)) and tests of these methods in matching of only optical or only SAR images and for other matching methods and associated quality criteria. Furthermore, the results of error detection of matching of optical images are a strong indication that the proposed method can be very helpful and be used for far more important applications, like DSM generation.

ACKNOWLEDGEMENTS

This work has been supported by Joel Braun (ETHZ) within his M.Sc. thesis, and Nusret Demir (ETHZ) within projects that have been financed by the Swiss Space Office (Switzerland) and CTI (Switzerland), that we greatly acknowledge. Furthermore, we thank our partner in these projects Sarmap SA (www.sarmap.ch) for fruitful input and data, and DGPF (www.dgpf.de) for provision of data.

REFERENCES

Braun, J., 2010. Co-registration and Ground Control Point transfer between optical and SAR satellite images. M.Sc. Thesis, Institute of Geodesy and Photogrammetry, ETH Zurich.

Christophe, E., Inglada, J., Giros A., 2008. Orfeo toolbox: a complete solution for mapping from high resolution satellite images. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. 37, Part B4, pp. 1263-1268.

Cramer, M., 2010. The DGPF-Test on Digital Airborne Camera Evaluation - Overview and Test Design. Photogrammetrie Fernerkundung Geoinformation (2): 75-84.

Dare, P., Dowman, I., 2001. An improved model for automatic feature-based registration of SAR and SPOT images. ISPRS Journal of Photogrammetry and Remote Sensing 56 (1): 13-28.

Goncalves, J., Dowman, I., 2002. Precise orientation of SPOT panchromatic images with tie points to a SAR image. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. 34, Part 3A, pp. 125-130.

Gruen, A., Baltsavias, E.P., 1988. Geometrically Constrained Multiphoto Matching. Photogrammetric Engineering & Remote Sensing 54(5): 633 - 641.

Reinartz, P., Müller, R., Schwind, P., Suri, S., Bamler, R., 2011. Orthorectification of VHR optical satellite data exploiting the geometric accuracy of TerraSAR-X data. ISPRS Journal of Photogrammetry and Remote Sensing 66 (1): 124-132.

Soukal, P., 2012. Error Detection in Automatic Point Matching between SAR and Optical Satellite Images. M.Sc. Thesis, Institute of Geodesy and Photogrammetry, ETH Zurich.

Wallis, R., 1976. An approach to the space variant restoration and enhancement of images. Proc. Symposium on Current Mathematical Problems in Image Science, Naval Postgraduate School, Monterey, CA.