A COMPARISON AND COMBINATION OF METHODS FOR CO-REGISTRATION OF MULTI-MODAL IMAGES

Charis Lanaras, Emmanuel Baltsavias, Konrad Schindler

ETH Zurich, Institute of Geodesy and Photogrammetry, Stefano-Franscini-Platz 5, 8093 Zurich, Switzerland
{lanaras, manos, schindler}@geod.baug.ethz.ch

KEY WORDS: Disasters, rapid mapping, data integration, image co-registration, image matching, error detection

ABSTRACT: Combined usage and analysis of images from different sensors for various applications, including disaster monitoring, often needs first an image co-registration. Co-registration is based on automated matching of corresponding image features (e.g. just 10-40) and becomes very difficult when the images differ a lot. Perhaps the most difficult case is that of co-registering optical and SAR images, while also multispectral images can vary a lot. In difficult co-registration cases, the point correspondences are mostly wrong. Thus, an additional problem is to find and automatically eliminate matching errors. In this paper, we present and compare various matching methods and show some benefits when combining them. The two main methods used include Mutual Information (MI) and a Discrete Fourier Transform method (FTCC). We extend the methods to also estimate the matching quality and exclude blunders, showing also results on that. Additionally, in one test a previously used Least Squares Matching (LSM) method was employed. Our test data include SAR and optical images, from the Tohoku Earthquake and Tsunami, 2011 in Japan and a region in Thun, Switzerland. Since there was no ground truth, the results for Tohoku were checked visually. As an extra test dataset we use very different AVHRR multispectral images, which are already co-registered. By introducing known geometric distortions we can perform quantitative evaluations with this data. Both MI and FTCC show a quite robust performance, with FTCC showing generally more blunders, but, when points are correct, slightly better accuracy. Matching quality evaluation and less matching method combination reduce the number of blunders very significantly to almost complete elimination.

1. INTRODUCTION AND INPUT DATA

The RAPIDMAP project (Cho et al., 2014) aims at integrating Remote Sensing (RS) and Geographic Information Systems (GIS) for resilience against disasters. In case of disasters and hazard monitoring, rapid mapping of the affected areas using RS and GIS is of great importance. Moreover, the continuous increase of imaging satellite sensors has necessitated the combined usage and analysis of such sensor data, often after an image co-registration. Co-registration has been an open task in the past and there is still some on-going research (see some references in Section 2.1). Here, we present a comparison (and partly combination) of three methods matching very different images and methods for matching blunder detection.

The datasets used include three cases. The first case is the Tohoku area in Japan, which was hit by the 2011 Tsunami on March 11th. The images acquired consist of pre- and post- disaster images. From the whole image area, here only a part was used, covering approx. 11.5 x 6 km and located close to the Sendai airport (Figure 1). The images were resampled (where needed) to a common GSD of 5m. A second dataset, located close to the city of Thun (Switzerland), with substantial elevation differences and many land-cover classes, is further used to evaluate the methodology. An overview of the SAR/Optical images for the first two datasets is given in Table 1.

Table 1. Overview of the SAR and optical data used.

		Satellite / Sensor	**Date**	**GSD**	**Wavelength**
Tohoku	SAR	TerraSAR-X	21.09.2008	5 m	X-band, 3 cm
			20.10.2010	5 m	
			12.03.2011	5 m	
			23.03.2011	5 m	
	Optical	FORMOSAT-2	11.03.2011	2 m	0.45 - 0.90 µm (PAN+NIR)
			19.03.2011	2 m	
Thun	SAR	TerraSAR-X	10.2008	3 m	X-band, 3 cm
	Optical	ALOS/PRISM	1.2006	2.5 m	0.52 - 0.77 µm (PAN)

The third dataset used consists of images from the Advanced Very High Resolution Radiometer (AVHRR) sensor on board of satellite NOAA-17. The images were taken on 08.02.2008 and contain 5 channels (see Table 2) with many radiometric differences. The images were orthorectified with the same DEM and orientation and the GSD is 1 km.

Here, 4 channels were used for co-registration, with channel number 2 as matching reference. Since channel 1 is very similar to channel 2 it was left out of the evaluation.

Table 2. The AVHRR channels' wavelengths.

Channel	1	2	3A	4	5
Wavelength [μm]	0.58-0.68	0.725-1.0	1.58-1.64	10.3-11.3	11.5-12.5

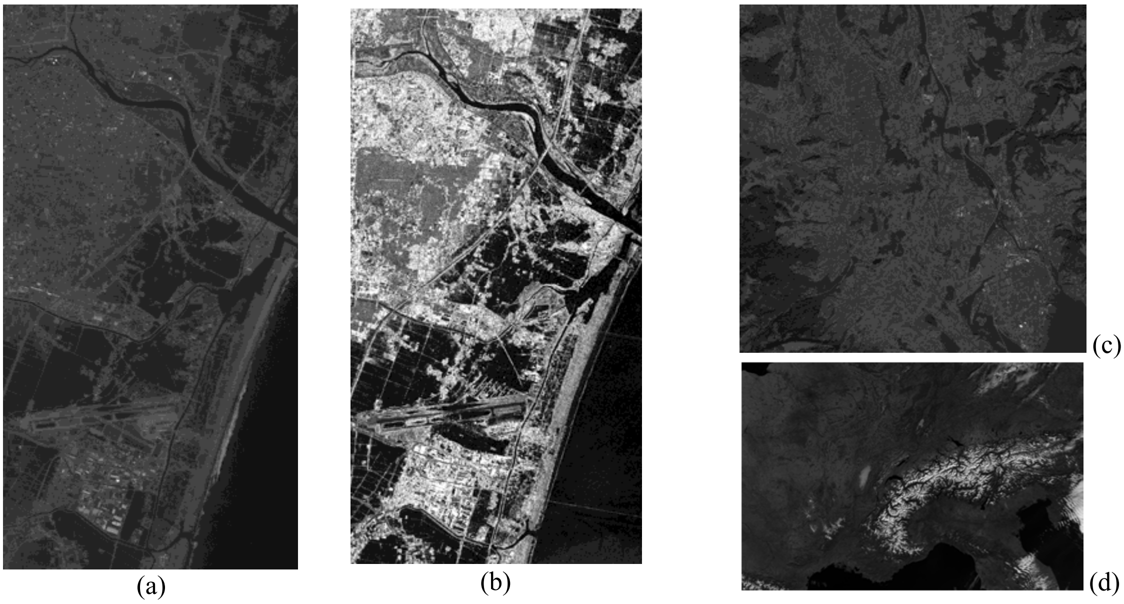

Figure 1. Region close to the Sendai airport, which was severely hit by the 2011 Tsunami: (a) is the FORMOSAT-2 (© NSPO) image and (b) the TerraSAR-X image, taken just after the tsunami. In (c) the ALOS/PRISM image of the region close to Thun and (d) channel 2 of the AVHRR sensor.

2. METHODOLOGY

2.1 Matching Methods and Metrics Used

Matching multi-modal images is a task, which cannot be solved with traditional matching methods. Thus, methods are being aimed at that best suit images with different intensities. Three approaches were used for this paper. The matching procedure in all following methods can be described as following: small samples (patches) of the images are selected and then their relative transformation is searched for in order to have a match. This approach belongs to the intensity-based methods, which compare intensity patterns. Based on these correspondences between the centres of the patches a global geometrical transform can be computed. One basic parameter that has to be set is the size of the patches, which differs in each method and dataset.

2.1.1 Mutual Information (MI). Mutual Information (MI) methods for registering multi-modal images are mentioned in Inglada and Giros (2004) and Reinartz et al. (2011). MI is an entropy-based metric that is connected to information theory and indicates the statistical dependence between two random variables. This metric is used to maximize the relative intensity values of coincident pixels, so in this case it is suited for images with different intensities. The implementation of Mattes MI (Mattes et al., 2001) included in the ITK Toolkit (http://www.itk.org) is used in this work. The metric is further used with a one-plus-one revolutionary optimization module (Styner et al., 2000). This method uses a shift transformation model. Matching is with subpixel accuracy.

2.1.2 Fourier Transform Cross-Correlation (FTCC). This method uses a single-step discrete Fourier transform (DFT) algorithm, which computes the cross-correlation in the frequency domain. Using DFTs leads to gains in computational speed and memory requirements in comparison to the usual FFT approach, while in this implementation one can use large oversampling factors in order to achieve sub-pixel accuracy,. By increasing the resolution of the patches to match by a factor of k (here k=100 was used), the maximum correlation can be located at a higher resolution, with a precision of 1/k pixels (Guizar-Sicairos et al., 2008). This method uses a shift transformation model.

2.1.3 Least Squares Matching (LSM). LSM based on the Adaptive Least Squares Correlation (Gruen and Baltsavias, 1988) is a technique optimized for image matching by computing local geometrical image shaping (in this case an affine transformation) during least squares iterations. LSM can be very accurate, but is slow and requires good starting approximations (in the order of a few pixels). Because of the big radiometric differences of the images to match with this metric, the thresholded image gradients were computed and used for matching. In previous work (Soukal and Baltsavias, 2012), quality criteria for every match point were determined and statistical measures were used to evaluate them in order to identify the wrong matches. LSM is used only with the AVHRR dataset.

2.2 Filtering of the Matches

During matching, it is expected that in some cases the matching will fail, especially when the images differ a lot. Thus, it is necessary to identify and remove the blunders. MI and FTCC have been extended in order to extract some quality criteria for each match point and eliminate blunders.

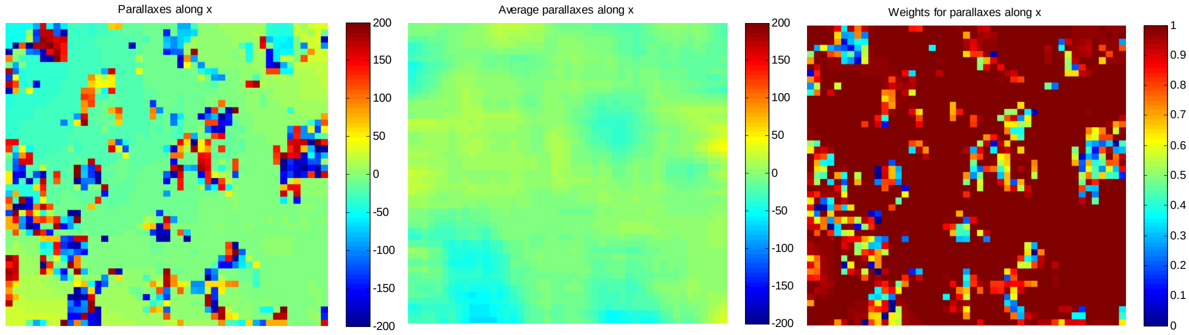

Figure 2. X-shifts of the matching regular grid (left), the average x-shifts computed for every point in its 9x9 neighbourhood (centre) and the final weight for local x-shift changes computed for every point (right), based on the difference of its x-shift from its average local x-shift.

Here, the following procedure has been used to determine quality values for MI and FTCC:

1. Local shifts are taken into account. Based on the shifts along x- and y-directions (Figure 2, left), an average x- and y-shift is computed around a 9x9 neighbourhood of each point (Figure 2, centre) and subtracted from its x- and y-shift. Thus, we can detect points, with different shifts than their neighbourhood. The final differences are normalized between 0 and 1, by dividing the square of the differences with the maximum difference squared (Figure 2, right). By averaging both weights for x- and y-shifts, the final weight is computed.
2. Secondly, the normalized cross-correlation (NCC) in the gradient domain is computed and used as a quality value. By taking a window and centring it at the matched points in the gradient images and sliding it in both x- and y-directions the NCC is computed for every pixel in a window. The size of the window depends on the matching window (here half its size). First, the value of NCC at the match point is divided by the maximum NCC value within the window and this ratio is stored. We also keep the location (in x and y) of the maximum NCC value with respect to the match point.
3. Another criterion for evaluating the matches is the standard deviation of the intensities within the patches. The assumption behind this is that windows that have a large standard deviation will have greater contrast and thus matching would be more successful, compared to images with low standard deviation (little to no texture). These values are also normalized from 0 to 1, by dividing all with the maximum value of standard deviation.
4. As a final quality criterion the matching criterion of every method can be considered. Here, only the maximum value of MI of the matched patches, normalized by dividing with the highest value, is used.

Finally, the criteria that are used for filtering are:

1. x-shift (from approx. location – not normalized)
2. y-shift (from approx. location – not normalized)
3. Score from shifts (normalized from 0 to 1)
4. Ratio of centre point NCC to maximum NCC within the window (typical values in the range from 0 to 1)
5. x-coordinate of maximum NCC (difference from patch centre – not normalized)
6. y-coordinate of maximum NCC (difference from patch centre – not normalized)
7. Standard deviation of optical patch (before matching – normalized from 0 to 1)
8. Standard deviation of SAR patch (before matching – normalized from 0 to 1)
9. (Optional) Mutual Information value (normalized from 0 to 1)

The filtering procedure uses the quality criteria described in this section and is applied in the following way. For each matched point we append into a vector all the quality criteria. Under the assumption that there are some correct matches, their quality criteria should share similar values, so there will be an area in the vector space that is dense, as these points would create a cluster (Figure 3). This cluster should be the largest in the vector space, but it is not necessary that it contains more than half of the points. The space, in which the vector endpoints are, gets sparser as the dimensionality of the vectors grows (when using more quality criteria) and it is easier to find a cluster. If the wrong points have random quality criteria it is more likely that they will be lying far away from other points. So in order to find the good matches we compute from every point in this space, its Euclidean distance to every other point and sum up all the distances. Points with the lowest sum of distances will indicate that they are close to many other points, while a larger sum of distances will indicate probable wrong matches. For the criteria with normalised values (0,1), we subtract from the values their mean (to centre them) and multiply by the mean standard deviation of the first two criteria. Thus, all criteria cover approximately the same value range in the vector space, with pixels units.

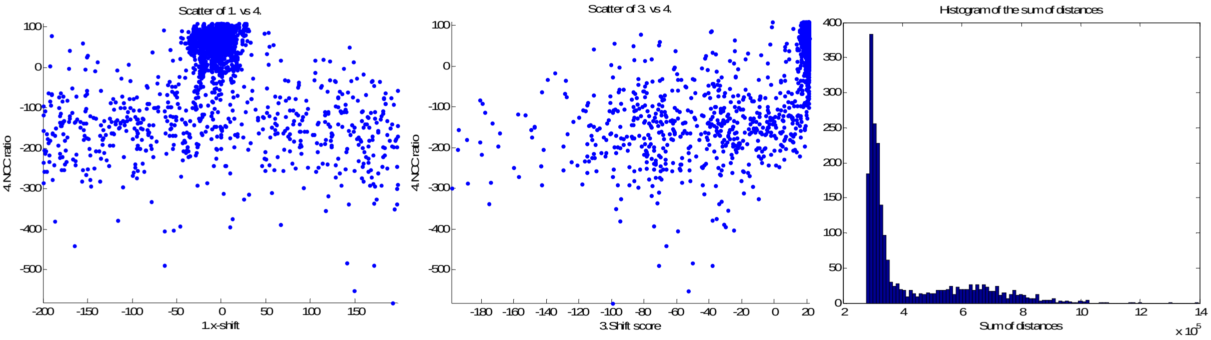

Figure 3. Scatter plots with only two dimensions each time in the “quality value” space, which show the locations with high point density (left and centre). On the right, the histogram of the sums of the distances.

The sums of the distances are sorted. Figure 3 (right) shows their histogram with the best points on the histogram left. The selection of good points is done manually by giving the number or percentage of the best points to keep. This threshold is application and image dependent. In the future, we plan to automate its selection.

3. RESULTS

3.1 Japan Dataset

For this dataset, the MI and FTCC methods were used. For this experiment a 301x301 sized window was centred at the points to be searched and for covering the whole image a regular grid was used with an interval of 75 pixels. The input data given in Table 1 are sorted into 4 pairs. The first two pairs consist of the FORMOSAT-2 image before the disaster combined with the SAR pre-disaster images (i.e.21.09.2008 and 20.10.2010) respectively. The third and fourth pairs consist of the FORMOSAT-2 image after the tsunami and the two post-disaster TerraSAR-X images (i.e. 12.03.2011 and 23.03.2011).

3.1.1 Comparison of methods before and after filtering. First of all, before performing any filtering of the matches and to be able to make comparisons and evaluate the results, the matches were manually checked by visual interpretation (with an accuracy of 1-3 pixels) and classified to correct and false matches. The summary of the results is given in Table 3 and a visual comparison of the computed shifts for pair 4 in Figure 4. Pair 3 has the biggest number of correct matches. This is due to the larger flooded area (see dark areas in Figure 1 (a) and (b)), which makes the contrast of features protruding from the flooded area easier to match. Due to some problems at the borders of the images, some numerical problems occurred and the matches were discarded. Points, which were matched correctly in both methods, were further used to check the consistency of both methods. In Table 4 the filtering performances are given, including the manually defined threshold, the number of points left after filtering and how many of the points are correct, based on the visual check. The filtering results of pair 4 are shown in Figure 4 (right part). The performance of this filtering leads to 90% or higher correct points.

Table 3. Summary of manual visual evaluation of the matches for all four Japan pairs.

Pair	Initial points	MI matches without numerical problems	FTCC matches without numerical problems	MI correct matches	FTCC correct matches	Common correct matches
1	324	323	298	265	172	149
2	324	321	296	254	164	138
3	324	312	323	274	297	270
4	324	312	315	275	183	166

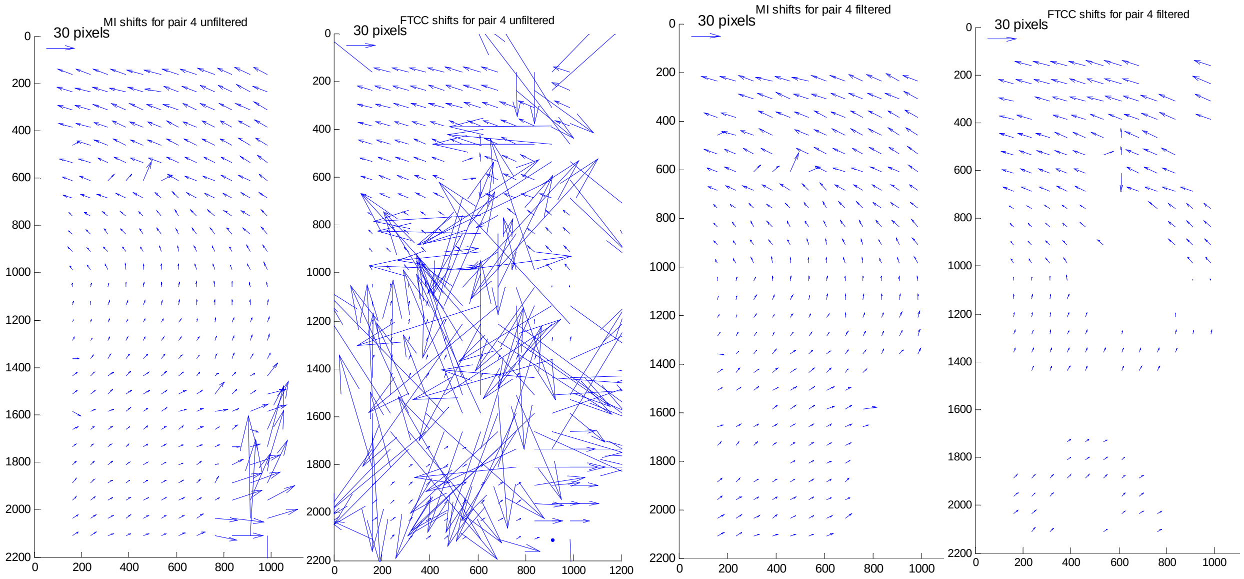

Figure 4. The shifts computed for pair 4 in the area given in Figure 1. From left to right: MI and FTCC shifts before filtering, MI and FTCC shifts after filtering. Scale of vectors in the upper left corner.

Table 4. Numerical evaluation of the points left after the filtering and the parameter (% of points to keep), which was manually set. Correct points are estimated based on visual check.

Pair	MI points after filtering	% of points to keep	Correct filtered points (no. / %)	FTCC points after filtering	% of points to keep	Correct filtered points (no. / %)
1	242	75%	230 / 95%	167	56%	163 / 98%
2	241	75%	227 / 94%	160	54%	151 / 94%
3	265	85%	259 / 98%	284	88%	280 / 99%
4	262	84%	254 / 97%	176	56%	168 / 95%

3.1.2. Combination of Methods. Another way to evaluate the relative (and partly absolute) correctness of the two methods is to compare the results of the methods at the same matching points and keep as good points those with small differences. This second filtering can be performed even to the original matching results, though it makes more sense to apply it after the filtering using the quality criteria. A comparison of MI and FTCC after the first filtering is shown in Table 5. Since after the first filtering, a few blunders still exist we give in Table 5 the median values of the differences and their median absolute deviation from the median (MAD) (multiplied by 1.4826 for normal distribution to obtain a consistent estimate of the standard deviation). We discard a point as blunder (in the 2nd filtering), if abs(point(x)-median(x)) > 3*1.4826*MAD(x) (and similarly for y). The remaining points (Table 5) were all classified as correct based on the visual check. The difference of these points can vary up to 2-3 pixels.

Table 5. Median difference and MAD (multiplied by 1.4826) between the FTCC and MI methods derived from the filtered matches.

Pairs	Median difference		1.4826*MAD		No. of points after / before 2nd filtering
	x	y	x	y	
1	-0.15	0.46	1.00	0.79	117 / 123
2	-0.51	0.56	0.92	0.97	115 / 128
3	-0.58	0.08	0.59	0.43	251 / 263
4	-0.93	-0.12	1.05	0.96	149 / 161

3.2 Thun Dataset

The Thun dataset is not related to any disaster, but rather includes a larger area than for the Sendai region. The matching window used in this case is 401x401 pixels and the density of the regular grid is 100 pixels. Out of the total number of 2209 points to match, we use the filtering procedure to select the best matches (Figure 5). No visual check was performed for this test, due to the high number of points. For the MI method a threshold of 70% was used and for

FTCC 60%, which results to 1519 and 1289 matched points respectively. Out of these points, 956 were matched by both methods, and among them 651 points (68%) differ less than 2 pixels and 793 points (83%) differ less than 4 pixels. Nevertheless, after the first filtering some wrong matches still exist, mostly in the case of FTCC. By combining these two methods, we can increase the certainty of matching correctness, and also perform a second filtering excluding points where results differ a lot (see FTCC blunders with very different MI results in Figure 5).

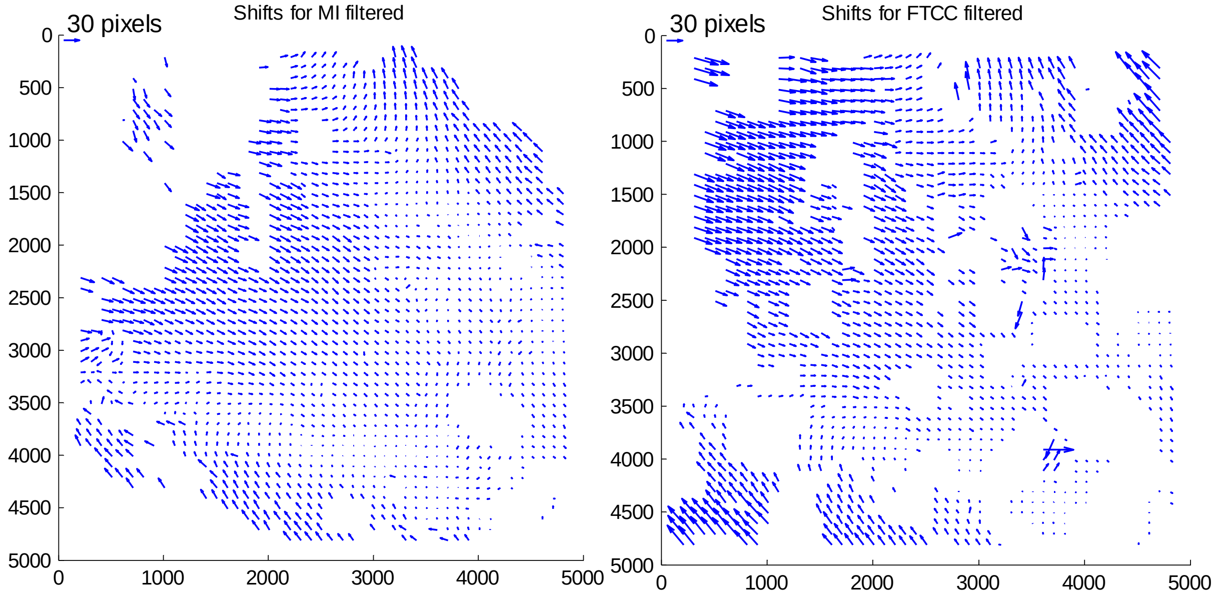

Figure 5. Filtered matches of the two methods, MI (left) and FTCC (right). Scale of vectors in the upper left corner.

3.3 AVHRR dataset

The AVHRR dataset is used to make some quantitative numerical evaluations, since its channels are almost co-registered, thus can be used quasi as reference. In some cases, between some channels small shifts are observed with all three evaluated methods. These shifts are stored and used below, as initial shifts of the “co-registered” images. The search window used is 149x149 pixels for MI and FTCC and 41x41 for LSM. The grid interval is 50 pixels and is common for all three methods. For matching, thresholded image gradients were used, after enhancing the contrast, since they led to more correct matches. Matching these images is a task that differs from the previous tests because the image scale is very coarse and there are no linear features, such as rivers and roads. Up to now, point correspondences were searched for exactly at the same location in both images to be matched. For the AVHRR dataset, we introduce a known similarity transformation of the regular matching grid (with origin the upper left image corner) and search for correspondences on this distorted grid (Figure 6), i.e. we change the starting positions of the matching. If the matching is correct, then the estimated shifts should eliminate the introduced error and permit the estimation (recovery) of the similarity transformation parameters (Figure 6). Before computing the transformation, we subtract from the matching results the initial shifts of the co-registered images mentioned above. This procedure can check the matching accuracy, also in dependence of the quality of the starting match positions. After matching, the first filtering was performed forMI and FTCC, however, only a few points were rejected. For LSM, points were excluded, if they needed more than

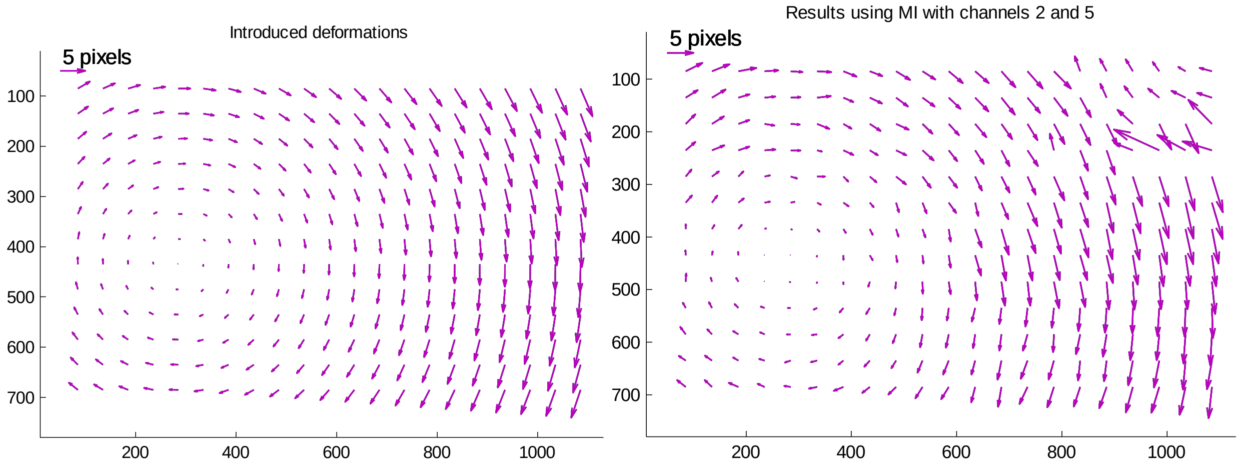

Figure 6. The inserted distortions of the grid in the AVHRR dataset (left) and the recovered distortions by MI after matching channels 2 and 5 (here without filtering).

25 iterations in the least-squares adjustment. Finally, based on the filtered results, the inverse transformation is estimated with RANSAC, which reduces the effect of blunders that remained after filtering. The numerical results of the transformation estimated can be found in Table 6. LSM has clearly worse performance than MI and FTCC, while the latter is better than MI regarding the estimation of the shifts. With both methods, the estimated shifts of the similarity transformation were never wrong more than 0.2 pixels.

Table 6. True and recovered transformation parameter values for three image pairs after the first filtering and RANSAC.

Channel Pairs		2-3A			2-4			2-5		
Transformation parameters	**True values**	Recov. MI	Recov. FTCC	Recov. LSM	Recov. MI	Recov. FTCC	Recov. LSM	Recov. MI	Recov. FTCC	Recov. LSM
x-shift [px]	**3.00**	2.80	2.92	2.81	3.20	2.98	2.27	3.14	2.99	2.39
y-shift [px]	**-2.00**	-2.02	-1.85	-1.81	-1.92	-1.97	-1.69	-1.92	-1.97	-1.62
rotation [deg]	**0.40**	0.40	0.39	0.37	0.40	0.40	0.32	0.40	0.40	0.31
scale	**1.000**	1.000	1.001	1.000	1.001	1.000	1.000	1.000	1.000	0.999

4. DISCUSSION AND CONCLUSIONS

In this paper we compared three methods for image co-registration and also produced some results with a combination of MI and FTCC. We introduce a new approach in performing a filtering of the matches to exclude blunders, which delivers good results, when compared to manually checked points. We have performed three tests involving very different satellite images, having various GSDs (from 2m to 1km). Test results showed that both MI and FTCC, using large matching patches, are quite insensitive to image differences and do not require good starting approximations. LSM, using smaller patches and without filtering, performs clearly worse. The computation time of MI and FTCC is similar, typically 1min for 300 points (using large matching patches and MatLab code). FTCC usually produces more and larger blunders, however due to their large size, they can be easily detected and excluded. FTCC also seems to produce slightly more accurate results than MI, although a possible subpixel accuracy could not be visually checked. In the Tohoku dataset, filtering produced results that were 94% to 99% correct, after visual control. This can be further improved by comparing the results of the two methods and keeping only points with small differences. A quantitative evaluation was made with the AVHRR images, by estimating after matching an a priori introduced transformation, showing a good estimation of the transformation parameters for both MI and FTCC. The procedures are automated, with the exception of a manually set threshold. The methods can be favourably used in co-registration of very different images for various applications, including rapid disaster mapping. Future work will concentrate on the improvement of filtering.

ACKNOWLEDGEMENTS

We gratefully acknowledge the financial support of RAPIDMAP by Concert-Japan, sarmap sa for providing the Thun SAR images and the University of Bern for providing the AVHRR dataset.

REFERENCES

Cho, K., Wakabayashi, H., Yang, C.H., Soergel, U., Lanaras, Ch., Baltsavias, E., Rupnik, E., Nex, F., Remondino, F., 2014. Rapidmap project for disaster monitoring. Proc. of 35th Asian Conference on Remote Sensing, 27-31 Oct., Nay Pyi Taw, Myanmar.

Gruen, A., Baltsavias, E., 1988. Geometrically constrained multiphoto matching. Photogrammetric Engineering and Remote Sensing, 54(5), pp. 633-641.

Guizar-Sicairos, M., Thurman, S. T., Fienup, J. R., 2008. Efficient subpixel image registration algorithms. Optics letters, 33 (2), pp. 156-158.

Inglada, J., Giros, A., 2004. On the possibility of automatic multisensor image registration. IEEE Transactions on Geoscience and Remote Sensing, 42(10), pp. 2104-2120.

Mattes, D., Haynor, D.R., Vesselle, H., Lewellen, T., Eubank, W., 2001. Nonrigid multimodality image registration. Medical Imaging 2001: Image Processing. SPIE Publications, 3 July, pp. 1609–1620.

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), pp. 124-132.

Soukal, P., Baltsavias, E., 2012. Image matching error detection with focus on matching of SAR and optical images. Proc. of 33rd Asian Conference on Remote Sensing, 26-30 Nov., Pattaya, Thailand, pp. 1436-1442.

Styner, M., Brechbuhler, C., Szckely, G., Gerig, G., 2000. Parametric estimate of intensity inhomogeneities applied to MRI. IEEE Transactions on Medical Imaging, 19(3), pp. 153-165.