*Article*

# AROSICS: An Automated and Robust Open-Source Image Co-Registration Software for Multi-Sensor Satellite Data

**Daniel Scheffler** [1,2,*] (ID)**, André Hollstein** [1] (ID)**, Hannes Diedrich** [1]**, Karl Segl** [1] **and Patrick Hostert** [2,3]

1. Helmholtz Centre Potsdam GFZ German Research Centre for Geosciences, Section Remote Sensing, Telegrafenberg, 14473 Potsdam, Germany; andre.hollstein@gfz-potsdam.de (A.H.); hannes.diedrich@gfz-potsdam.de (H.D.); segl@gfz-potsdam.de (K.S.)
2. Geography Department, Humboldt University Berlin, Unter den Linden 6, 10099 Berlin, Germany; patrick.hostert@geo.hu-berlin.de
3. Integrated Research Institute on Transformations of Human-Environment Systems (IRI THESys), Humboldt University Berlin, Unter den Linden 6, 10099 Berlin, Germany
* Correspondence: daniel.scheffler@gfz-potsdam.de; Tel.: +49-331-288-1198; Fax: +49-331-288-1192

**Abstract:** Geospatial co-registration is a mandatory prerequisite when dealing with remote sensing data. Inter- or intra-sensoral misregistration will negatively affect any subsequent image analysis, specifically when processing multi-sensoral or multi-temporal data. In recent decades, many algorithms have been developed to enable manual, semi- or fully automatic displacement correction. Especially in the context of big data processing and the development of automated processing chains that aim to be applicable to different remote sensing systems, there is a strong need for efficient, accurate and generally usable co-registration. Here, we present AROSICS (Automated and Robust Open-Source Image Co-Registration Software), a Python-based open-source software including an easy-to-use user interface for automatic detection and correction of sub-pixel misalignments between various remote sensing datasets. It is independent of spatial or spectral characteristics and robust against high degrees of cloud coverage and spectral and temporal land cover dynamics. The co-registration is based on phase correlation for sub-pixel shift estimation in the frequency domain utilizing the Fourier shift theorem in a moving-window manner. A dense grid of spatial shift vectors can be created and automatically filtered by combining various validation and quality estimation metrics. Additionally, the software supports the masking of, e.g., clouds and cloud shadows to exclude such areas from spatial shift detection. The software has been tested on more than 9000 satellite images acquired by different sensors. The results are evaluated exemplarily for two inter-sensoral and two intra-sensoral use cases and show registration results in the sub-pixel range with root mean square error fits around 0.3 pixels and better.

**Keywords:** image co-registration; Python; sub-pixel; Fourier shift theorem; optical; radar; phase correlation; geometric pre-processing; inter-sensor; intra-sensor

## 1. Introduction

Remote sensing datasets are usually provided as georeferenced image datasets. Nevertheless, often, slight displacements between images of different sensors or even within the same sensor exist. Precise co-registration is a fundamental requirement for many remote sensing applications, such as environmental mapping, change detection, image fusion or mosaicking [1–5]. Even misregistrations in the sub-pixel range can affect the reliability of change detection maps [4,6]. In recent years, various

image registration algorithms have been proposed. However, accurate detection and correction of geospatial misalignments between images still presents challenges [7,8]. There is therefore a need for fast and efficient image co-registration algorithms, especially considering the rapidly growing remote sensing data volumes available during recent years and the need for fully automated pre-processing and analysis pipelines that make use of multi-sensoral and multi-temporal data sources. The overall challenge is to provide a fully operational algorithm that can handle the heterogeneity of satellite data acquired all over the world at different times by different sensors with varying atmospheric conditions and differing acquisition and illumination geometries.

In general, two groups of image registration techniques exist: intensity based and feature based [9]. Intensity-based techniques rely on the recognition of similar grey value patterns that appear in both the reference and the target image. Proposed algorithms differ in terms of how this similarity is detected and how it is quantified. The general approach of feature-based techniques is to detect the image positions of ground objects based on distinct recognition features, such as crossroads, buildings, borders of agricultural parcels, well-defined mountain ridges or other sharply delineated objects that are not subject to change in time and space. Both groups have advantages and disadvantages with regard to different application fields. Feature-based techniques are often computationally inexpensive and have the potential to precisely quantify displacement vectors in the sub-pixel range while requiring distinctive objects and often suffering from unevenly distributed positive matches [10]. Consequently, the success rate is strongly dependent on the appearance and composition of the Earth's surface, prohibiting operational applicability. In contrast, intensity-based techniques also allow sub-pixel accuracies to be achieved, and they are not dependent on the presence of significant image features. However, they are limited to non-excessive image distortions, and especially correlation-based techniques have a tendency to require more computational power [9,11].

Extensive research has been conducted; several review papers and comparison studies detail improvements in accuracy and reliability and how to mitigate limitations on both groups of co-registration techniques [7,9,12,13]. Moreover, many algorithms have been proposed to automate the time-consuming process of tie point (TP) localization (e.g., [3,5,13–18]). However, remote sensing applications place high demands on the robustness and accuracy of the co-registration approach used, particularly if different sensors are involved and images cover long time periods and varying surface conditions of the Earth. Different sensor noise levels, variable cloud coverages, acquisition and illumination angles, and various ground sampling distances, spectral band positions and geographic projections must all be considered in the design of a robust co-registration approach. Only a few algorithms have been proposed that are explicitly designed to handle the challenges arising from multi-sensoral and multi-temporal remote sensing image co-registration. Wong and Clausi (2007) presented ARRSI, a MATLAB-based, automated approach that was based on principle moments of phase-congruency in connection with a variation of RANdom SAmple Consensus (RANSAC) as outlier detection, called Maximum Distance Sample Consensus (MDSAC) [15]. It is applicable to inter- and intra-sensoral use cases. The COSI-Corr package was developed in 2007 by Leprince et al. and provides automatic co-registration and orthorectification of optical satellite images acquired by pushbroom sensor systems. It uses phase correlation and has been implemented as an IDL algorithm [16]. Gao et al. (2009) proposed a cross-correlation based approach called AROP for Landsat and Landsat-like data that can automatically register satellite images acquired by multiple sensors and robustly handle different spatial resolutions and projections [17]. Behling et al. (2014) also utilized a cross-correlation-based approach for detecting and correcting global X/Y translations between satellite image time series of multiple sensors in a robust and automated way [5]. Yan et al. (2016) used a hierarchical image matching approach and combined intensity- and feature-based techniques to automatically co-register Landsat-8 and Sentinel-2 data globally [18]. However, all these algorithms either require many input parameters or preprocessing steps and are thus difficult to use, are embedded in proprietary (and often expensive) software, or are not designed to correct local displacements.

The goal of this paper is thus to provide a generic and automated co-registration tool for georeferenced but only inadequately matching satellite images to achieve sub-pixel accuracies. The algorithm should further depend only on a minimum of input parameters, while allowing users to fine-tune the algorithm to match their individual needs. The following requirements have accordingly been identified and taken into account:

- generic applicability to multi-sensoral and multi-temporal use cases, i.e., robustness against:

  ○ varying acquisition and illumination geometries
  ○ varying atmospheric conditions and land cover dynamics
  ○ different sensor noise levels, ground sampling distances, spectral band positions
  ○ different geographic projections, uneven coordinate grid alignments

- minimal impairment of data quality from resampling
- computational efficiency and being deployable on powerful server environments as well as on customary desktop computers, while optimally utilizing available resources
- simple integration into existing remote sensing processing pipelines
- simple usability though an automated software installation routine following customary Python standards, an easy-to-use software interface, a minimum of obligatory input parameters and support for a wide range of raster image formats
- public availability, using an open-source licence-agreement

Here, we use phase correlation, an intensity-based registration technique, as the core technology to detect sub-pixel displacements. This way, we remain independent of distinct image features that are possibly unevenly distributed or completely missing, especially considering the background of highly dynamic surface coverages within multi-temporal satellite data. Phase correlation relies on deriving image displacements in the frequency (Fourier) domain and enables quick calculation of relative translational offsets between two images in the same dimension. Compared with other intensity-based algorithms such as classical cross-correlation, phase correlation delivers significantly higher accuracy due to a distinct peak in the cross-power spectrum indicating the point of registration [9,19–21]. It delivers excellent co-registration results, even in the case of poor signal-to-noise ratios and substantial ground cover changes between different images, e.g., due to seasonal vegetation dynamics [19,22]. Adding to its robustness against albedo differences and its computational efficiency [23,24], an intensity-based co-registration approach, such as phase correlation, is highly suited for a generic application to multi-sensor remote sensing datasets—provided that the geometric displacements follow a more or less affine or polynomial pattern, which would limit a phase correlation approach [7,9,20]. However, since remote sensing data are usually distributed as georeferenced datasets (even though not always precisely matching), it can be reasonably assumed that the input datasets are already roughly matching and do not show any severe geometric artefacts. This is certainly the case considering state-of-the-art Earth observation data, such as Sentinel-2, TerraSAR-X or the novel Landsat collection data [25–27].
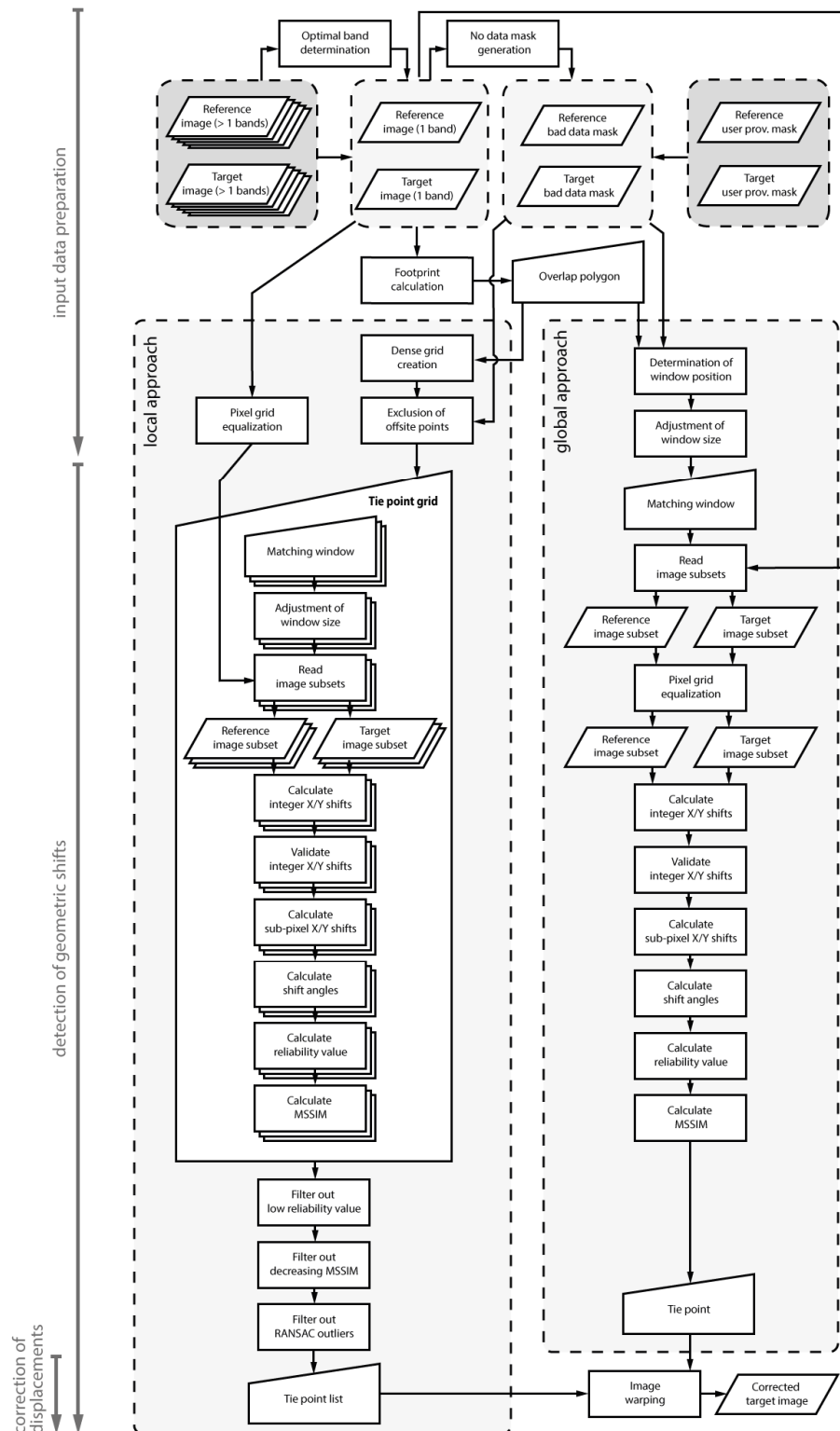
## 2. Processing Chain

### 2.1. Algorithm Overview

The presented algorithm is called "AROSICS" (Automated and Robust Open-Source Image Co-Registration Software). It is based on a phase correlation approach as proposed by Foroosh et al., 2002 [19] and makes use of Fourier shift theorem [28], enabling the determination of precise X/Y offsets at a given geographical position. However, phase correlation can only be used for two monochromatic (single-band) input images with exactly the same pixel dimensions, representing roughly the same (or slightly shifted) geographical position on the Earth's surface, ideally also with similar pixel intensity values. This is usually not fulfilled when dealing with multi-temporal and multi-sensoral image data

(see Section 1). Therefore, it was necessary to combine the pure phase correlation approach with additional processing and evaluation modules that are all integrated within the AROSICS framework, representing the intrinsic innovation of the presented work.

The workflow is divided into three major processing steps (Figure 1).



**Figure 1.** Flow chart of AROSICS (Automated and Robust Open-Source Image Co-Registration Software).

- *Input data preparation (Section 2.2)* including: (1) automatic selection of the spectral bands from target and reference image to be used for co-registration; (2) no-data mask generation and combination with user-provided masks (e.g., containing clouds and cloud shadows) to separate bad-data masks for reference and target image; (3) calculation of the respective footprint polygons and corresponding overlap area; (4) pixel coordinate grid equalization; and (5) adjustment of matching window positions and sizes
- *Detection of geometric shifts (Section 2.3)*, either in a moving-window manner for each point of a dense grid (local co-registration approach) or for only a single coordinate position (global co-registration approach), including validation
- *Correction of displacements (Section 2.4)*, by warping the target image under the use of a list of tie points (local co-registration approach) or a single X/Y shift vector (global co-registration approach)

The presented workflows for correcting local or global displacements are separate algorithms and both part of AROSICS. The **local co-registration approach** applies phase correlation in a moving-window manner to a (by default) regular grid of coordinate points and estimates X/Y translations for each point within the overlap area of the input images. We validate the resulting shift vector grid based on a set of independent quality metrics (see Section 2.3.2). The remaining points that passed all validation steps are then used as a list of tie points to fit an affine transformation model for warping the target image and thus for correction of detected local geometric misregistrations. In contrast, the **global co-registration approach** computes the displacements at only one coordinate position within a small image subset (from here: matching window). The target image is then shifted on the basis of a single displacement vector. This procedure is computationally inexpensive and thus very fast, which is not the case for the local approach. Moreover, it allows us to optionally disable image warping to correct sub-pixel X/Y shifts without any resampling—only by adjusting the geocoding metadata. This preserves initial pixel values, while at the same time not matching the coordinate grid of the reference image. However, global co-registration assumes that the whole misregistration between reference and target image can be described by a constant (global) translation in the X and Y directions. It might be useful if the displacement patterns are sufficiently known to confirm this a priori or if any modification of original pixel values by resampling is an issue and needs to be strictly avoided. Against a backdrop of generic and operational applicability to multi-sensoral and multi-temporal image data, we focus on the local co-registration approach here.

### 2.2. Input Data Preparation

#### 2.2.1. Obligatory and Optional Input Data and Parameters

Obligatory input data and parameters have been intentionally omitted as far as possible. Instead, we incorporated a set of default and fall-back values along with automated algorithms for parameter derivation, which facilitates the use of the AROSICS framework but also enables fine-tuning of results by providing user-defined data. Consequently, only a reference and a target image and a grid resolution (in case of local co-registration) defining the distance of tie points within the generated dense grid have to be provided. An appropriate grid resolution depends on the complexity of the expected misregistration patterns. Small point distances enable the correction of more complicated distortions but increase computational load. The input images must have valid geocoding information corresponding to any geographic or UTM map projection (equal geographical datum). Various raster image formats are supported [29]. In addition, it is always reasonable to provide a mask of areas with non-opaque clouds in the input images to decrease error probabilities in co-registration and to speed up processing. No-data masks indicating the image areas filled by the no-data value are automatically generated. The no-data value is read from the input image metadata if specified there or alternatively automatically derived from a $3 \times 3$ window at each image corner. Generated and user-provided masks are combined to separate (1-bit) bad-data masks for reference and target image.

### 2.2.2. Determination of Optimal Spectral Bands Used for Matching

In the case of a multi- or hyperspectral input image to be co-registered, automatic determination of the specific spectral band (wavelength position) to be used as input image for co-registration is essential to ensure that the intensity information in the reference and target image is as similar as possible. While phase correlation seems robust against intensity differences (regardless of the origin), as long as the actual image content, and, thus, its contained frequency information, does not vary too much, it has been reported that the robustness of intensity-based co-registration algorithms correlates with the similarity of the pixel value intensity of the input images [30]. Therefore, the implemented algorithm respects the wavelength position of the target band. It automatically chooses the optimal band of the reference image by determining the band with the minimal spectral distance, given that the respective image metadata is available. Otherwise, the first band is chosen.

If co-registration targets a multi-band data cube, the algorithm can be either applied band-by-band in a loop, or the co-registration results of a single band can be applied to the remaining bands. The computational expense of the former will be higher, while a better co-registration accuracy of the target image bands to the corresponding reference band may be achieved.

### 2.2.3. Calculation of Overlap Area between the Input Datasets

Knowing the exact spatial overlap area of the two input images, excluding any no-data areas, is the prerequisite for automatically determining coordinate positions for each matching window used for calculating spatial shifts. Furthermore, it efficiently helps to reduce the computational workload needed when performing local co-registration based on a large number of tie points, because all points of the dense grid where one of the input images only contains no-data values can be excluded from processing. Thus, our algorithm first calculates footprint polygons of reference and target images by applying raster-to-vector conversion to the previously generated no-data masks and then identifies their common extent to be used as overlap polygon.

### 2.2.4. Pixel Grid Equalization

Remote sensing images are often delivered with uneven pixel coordinate grids, especially in multi-sensoral cases where spacing and grid origins usually differ and images may be provided with different map projections. We therefore resample one image to the exact pixel spacing, origin and map projection of the other (based on the respective geocoding information and projections) before applying the phase correlation algorithm for detecting geometric shifts. Since having the same number of rows and columns is one of the prerequisites for applying phase correlation, an equalization of the pixel grids is indispensable. The algorithm always performs a down-sampling of the image with the higher spatial resolution to the lower resolved one. Performing an up-sampling of the lower resolved image would not increase co-registration accuracy but would slow down processing. By default, we use a cubic resampling technique here to avoid positional changes of ground objects within the images that are solely caused by the resampling algorithm (e.g., in case of nearest neighbour resampling).

This processing step is performed before applying phase correlation to the dense grid in case of local co-registration, whereas in the global case, we embedded it into the displacement calculation (Figure 1). The intention is to avoid computational overhead by only resampling the actually used image subset (within the matching window) in the global case while preventing repeated resampling of the same position caused by overlapping neighboured matching windows in local co-registration.

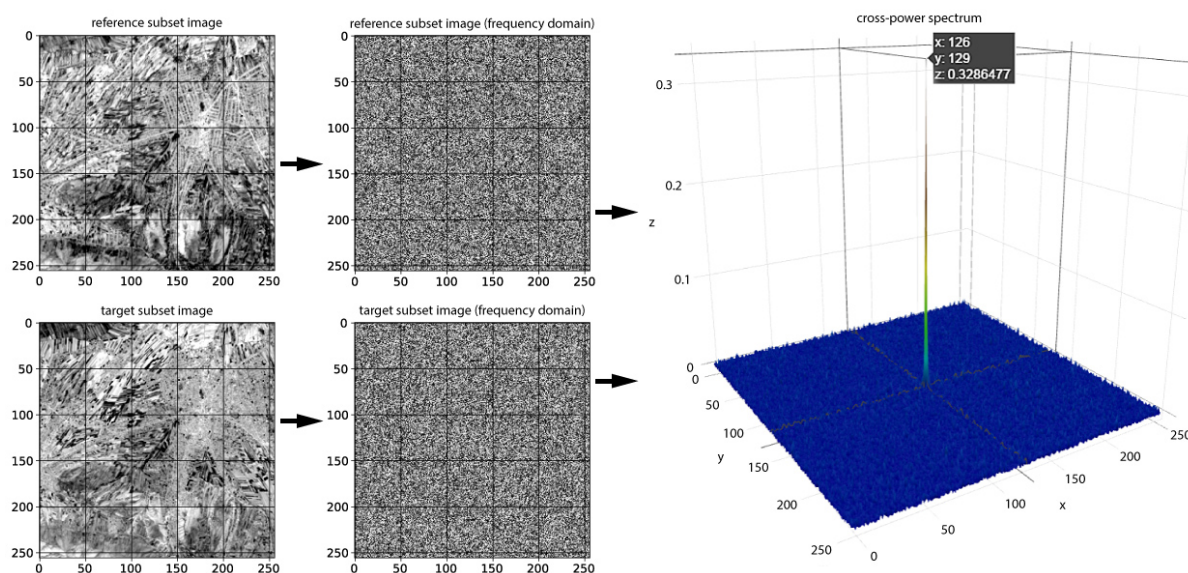### 2.2.5. Adjustment of Matching Window Positions and Sizes

The positions of the matching windows are given by the previously created dense grid of coordinates in case of local co-registration. For the global approach, the coordinate position defaults to the centre of the overlap polygon, thereby avoiding those image areas that are flagged as "bad position" according to respective bad-data masks.

The size of the matching window follows a default (256 × 256 pixels) or a user-provided setting and does not change within the overlap area used for co-registration. This is essential to keep the results of neighbouring displacement vectors comparable, since the precision of the calculated displacement changes with the size of the matching window [11,22]. However, at the edge of the overlap area, it is needed to ensure that the matching window does not protrude into the no-data area of reference or target image. Otherwise, the edge between the data and no-data areas of the image may affect the result of geometric shift detection via phase correlation due to its own frequency response in the Fourier domain. Consequently, we adapt the window size at the edges, whereby the reduced size influences the result of a later reliability check (Section 2.3.2) and is thus respected during validation of the shift vector grid.

## 2.3. Detection of Geometric Shifts

### 2.3.1. Calculation of Geometric Shifts within the Matching Window

For calculating geometric shifts between the input images within a single matching window, both subset images are transformed into the frequency domain – in this study using FFTW (Fastest Fourier Transform in the West), the fastest freely available implementation of discrete Fourier transform (DFT) that is described in detail by Frigo and Johnson in 2005 [31] and 1997 [32]. The two resulting images in the frequency domain are phase-correlated to generate their cross-power spectrum, which is then transformed back into spatial domain using inverse FFTW [9,13,22,33,34]. The normalized form of the cross-power spectrum in the spatial domain demonstrates a distinct sharp peak at the point of registration of the input images (Figure 2), which can be used for quantification of image displacements [7,9,16,19,24,33]. Integer shifts can be derived from the distance between the position of the maximum peak and the centre position of the spectrum in the X and Y directions [33].



**Figure 2.** Subset images within the matching window in spatial and frequency domains and the corresponding cross-power spectrum (spatial domain).

The subset image of the target dataset, i.e., the image content within the matching window, is then temporarily moved according to the calculated integer shifts. This is realized by adjusting the corner coordinates of the matching window and recreating the subset image for the target dataset at the new position. To derive sub-pixel shifts, both subset images (reference and integer corrected target) are again transformed into the frequency domain, and a refreshed cross-power spectrum is created. Sub-pixel shifts can then be estimated based on formulas 1 and 2, where $v_{(0,0)}$ represents the value

of the refreshed cross power spectrum at the peak position and $v_{(1,0)}$ and $v_{(0,1)}$ represent the direct neighbours in X- and Y-direction. Corresponding mathematic derivations can be found in [19].

$$\Delta x = \frac{v_{(1,0)}}{v_{(1,0)} \pm v_{(0,0)}} \qquad (1)$$
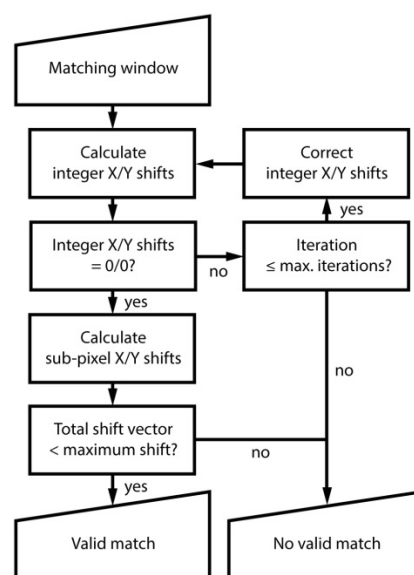
$$\Delta y = \frac{v_{(0,1)}}{v_{(0,1)} \pm v_{(0,0)}} \qquad (2)$$

The calculated sub-pixel shifts are added, and the total shifts in pixel units converted into map units, with respect to the ground sampling distance of the reference image. In the case of the local shift correction approach, a tie point grid is generated by combining total map shifts with the absolute map coordinate offsets, as provided by the geocoding information of the input images.

### 2.3.2. Validation of Calculated Spatial Shifts

The reliability of the X/Y displacements is connected to the pixel value similarity of the input images induced by, e.g., varying illumination or land cover changes or by different sensor noise levels. This is also supported by [9,22], and potential effects are quantified by implementing a total of five complementary validation techniques. Their individual performance depends on the image content of the input images and the pattern of misregistration. They build upon each other and aim to avoid different sources of erroneous shift detection but may also be deactivated on demand. However, optimal validation results have been observed by combining all the validation techniques presented below.

First, a validity check of the calculated integer shifts has been implemented (Figure 3). It is assumed that correcting the integer shifts within the respective matching window by moving the target subset image must result in zero displacements. If a subsequent calculation still yields non-zero X/Y shifts, the found match is understood to be false-positive and rejected. In this case, we apply the newly calculated integer shifts and then repeat the shift calculation algorithm to assess whether remaining integer shifts are now zero. This continues until a maximum number of iterations (5, by default) is reached. If shifts remain, the matching fails for the respective geographic position. Otherwise, a valid match (in terms of integer displacements) is found, and the sub-pixel displacements are finally derived, as explained above.



**Figure 3.** Flow chart of integer shift validation during calculation of sub-pixel X/Y shifts.

After successful calculation of sub-pixel X/Y shifts, a threshold check (second validation method) is applied to eliminate unrealistically large displacements. This enables incorporation of the knowledge of the user by simply rejecting all tie points that exceed a certain vector length (by default 5 reference image pixels). However, the usage of this possibility has been left optional to maintain wide applicability of the proposed algorithm.

It is still possible that some registration points do not match correctly, e.g., due to repetitive patterns in one of the input images that introduce pseudo-matches. To address that, we combined another three validation techniques (3–5) to effectively filter outlier tie points in the case of local displacement correction. Two of those validation metrics are separately calculated for each tie point of the dense grid, directly after calculating sub-pixel shifts.

The first (validation method 3) analyses the three-dimensional shape of the cross-power spectrum and aims to quantify the sharpness of the peak (Figure 2) as a measure for assessing the reliability of the respective tie point. It returns a reliability percentage *R* that is calculated using the following formulas:

$$\mu_{peak} = \frac{1}{N} \sum_{\substack{-1 \leq r \leq 1 \\ -1 \leq c \leq 1}} v_{(r,c)} \tag{3}$$

$$\mu_{remain} = \frac{1}{N} \sum_{\substack{-1 > r > 1 \\ -1 > c > 1}} v_{(r,c)} \tag{4}$$

$$\sigma_{remain} = \sqrt{\frac{1}{N} \sum_{\substack{-1 > r > 1 \\ -1 > c > 1}} \left( v_{(r,c)} - \mu_{remain} \right)^2} \tag{5}$$

$$R = 100 - 100 \frac{(\mu_{remain} + 3\,\sigma_{remain})}{\mu_{peak}} \tag{6}$$

Because $v_{(r,c)}$ represents the value of the cross-power spectrum at a specific position and N is the number of involved pixels, the mean power within a $3 \times 3$ window centred at the position of the peak of the spectrum is related to the mean power of the remaining spectrum plus its three-fold standard deviation. This results in a reliability percentage for each tie point and enables the exclusion of those points with a value below a certain threshold from later calculation of geometric transformation parameters. Empirical tests (not shown) indicated that a threshold of 30% can be recommended, below which tie points are rejected.

The fourth validation method that is also performed for each point of the dense grid evaluates the similarity of the subset image content within the matching window before and after shift correction. It relies on the assumption that successful correction of an erroneous co-registration leads to increased similarity between the reference and target image. As a measure for image similarity, the Mean Structural Similarity Index (MSSIM) has been chosen, as it is reported to be highly sensitive to even marginal image displacements [35,36].

MSSIM returns a value between 0 and 1, whereas 0 represents no match and 1 represents a perfect match. To calculate MSSIM before and after shift correction, it is first applied to the input data of phase correlation, and then the subset target image is globally shifted, according to calculated sub-pixel shifts for the corresponding position; then, MSSIM is calculated again. All points of the dense tie point grid where MSSIM decreases are flagged and later excluded from estimating transformation parameters.

The fifth and final validation technique is not applied point-by-point, but it incorporates all calculated shift vectors at once. This is essential for avoiding artefacts in the co-registration result due to outliers within the dense grid point data. For that purpose, the widely used state-of-the art

algorithm RANSAC [37] has been chosen, coupled with the assumption that in case of remotely sensed satellite data, an erroneous co-registration can be roughly corrected using an affine transformation. In the AROSICS framework, RANSAC has been incorporated to automatically estimate the parameters of an assumed affine transformation between reference and target image and thus to identify outliers among the previously calculated shift vector grid. Parameter estimation is automatically performed by RANSAC in an iterative process, but it has some free input parameters to be set. The most important one (according to our experience) is the threshold used to separate inliers and outliers, which is also supported by [38–40]. In this context, the optimal threshold value highly depends on the heterogeneity of the input sample and, hence, on the variance and reliability of the calculated displacement vectors. This variance is, in turn, connected to the input images itself, meaning that, for example, extensive cloud coverage may strongly influence the calculated tie points, as it often causes repetitive image patterns leading to a greater proportion of false-positives [9]. In addition, RANSAC suffers from high outlier proportions within the input sample [38]. To overcome that and avoid an erroneous estimation of affine parameters, multiple strategies have been implemented in this validation step.

First, the opportunity to consider bad-data masks for the input images helps to effectively exclude such critical image areas from the matching process (Section 2.2). Thus, the inlier-outlier variance within RANSAC input data is reduced, and the probability of an outlier dominated input sample is significantly decreased. However, it must be noted that, on the one hand, cloud masks are usually generated automatically and possibly contain misclassifications. On the other hand, the provision of such masks has been intentionally kept optional, because they are not always available.

As a further way to increase the robustness of outlier determination, RANSAC has been implemented as the last level of validation. Consequently, the input sample of RANSAC is already pre-filtered, in the sense that tie points with low reliability value or decreasing MSSIM are no longer passed to RANSAC.

Finally, an iterative approach is used to automatically optimize the abovementioned RANSAC threshold separating inliers from outliers. RANSAC is run multiple times while observing the proportion of inliers and outliers in the result to identify a threshold that flags a realistic number of outliers within the input data. The first iteration starts with an initial guess of the threshold, which is then iteratively increased or decreased, depending on whether the flagged outlier proportion is too low or too high. Empirical sensitivity tests (not shown) indicated that a 10% proportion of outliers seemed to be suitable for most use cases. Thus, the algorithm terminates as soon as $10\% \pm 2\%$ of the input tie points have been marked as outliers.

### 2.4. Correction of Displacements

In the case of co-registration, respecting local distortions is pursued; correction of geometric misregistrations is based on a list of tie points that fulfil our set of validation criteria and quality indicators. Otherwise, in case of global co-registration, only one tie point specifies an X/Y translation to be applied to the target dataset. If the input images have been provided with unequal map projections, the map projection of the calculated displacements is respected during image warping. The warping itself is based on a cubic resampling technique with the ability to align the pixel grid of the target image to the one of the reference image or to change target spatial resolution. Resampling is only performed once during the whole co-registration workflow to avoid undesired degradation of geometric and spectral image quality. Global co-registration allows us to completely avoid resampling so that pixel values of the target image are kept unchanged. In that case, only geocoding image metadata is adjusted—pixel grids are not aligned.

## 3. Sample Data

The proposed algorithm was tested and evaluated based on various satellite datasets acquired by different sensors at different times and geographic positions (Table 1.). The data include inter-sensoral use cases to show registration results with typical registration difficulties of multi-sensoral and

multi-temporal imagery (see Section 1) as well as intra-sensoral use cases that show both band-to-band registration performance (within the same multi-spectral dataset) and registration between multiple acquisitions of the same sensor. The selected datasets intentionally exhibit high cloud coverages, no-data areas, image edges diagonally crossing the dataset, strong relief and large differences in ground sampling distances to demonstrate the functionality and effectiveness of the AROSICS workflow.
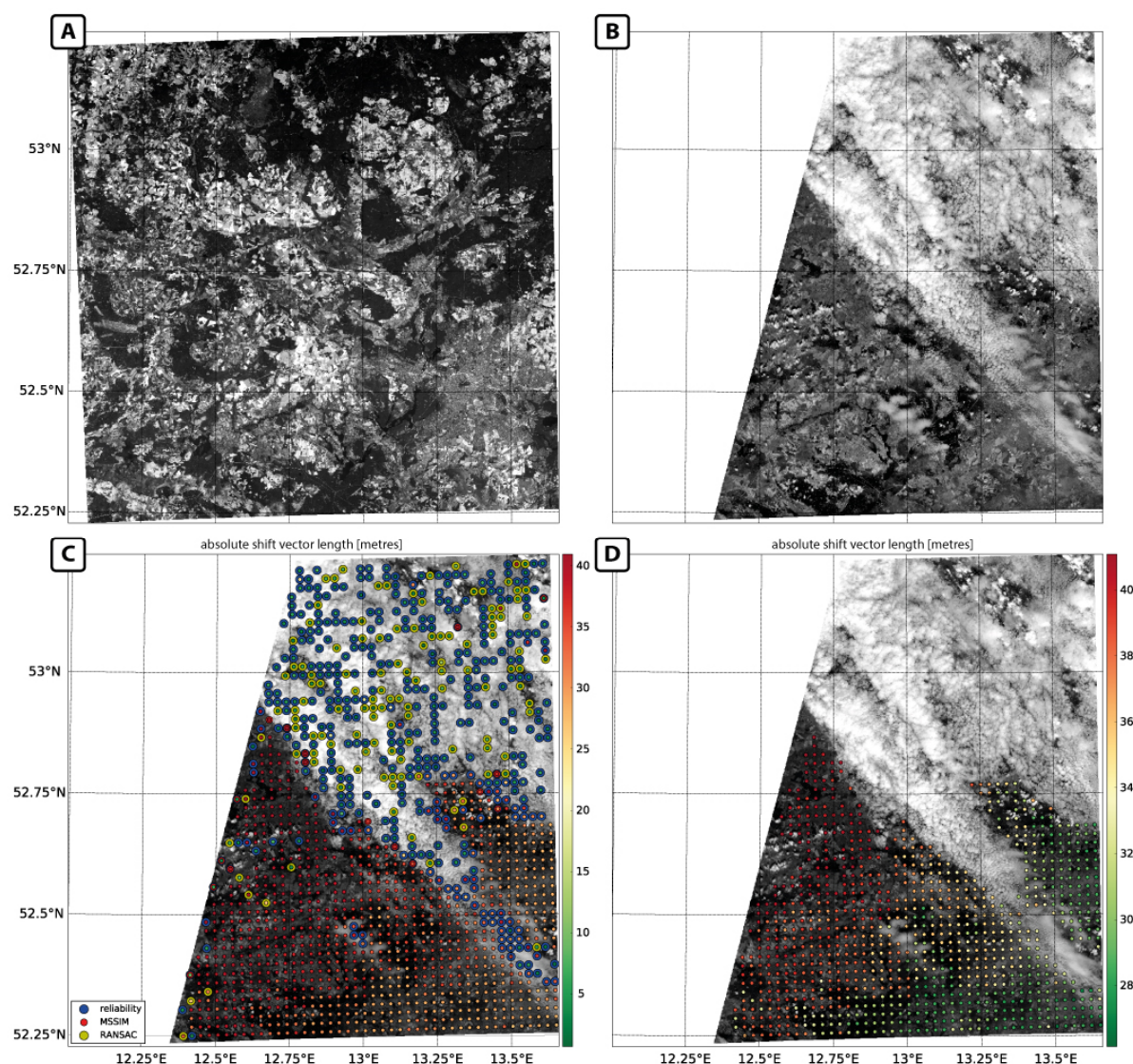
**Table 1.** Summary of example satellite datasets.

| Name | Reference/Target Image (R/T) | Sensor | Acquisition Date | Spatial Resolution (m) | Image Dimensions (Rows, Columns) | Processing Level |
|---|---|---|---|---|---|---|
| INTER1 | R | Landsat-8, band 2 | <multiple> | 15 | 7320, 7320 | 1T |
| | T | Sentinel-2A, band 3 | 29 May 2016 | 10 | 10,980, 10,980 | 1C |
| INTER2 | R | Landsat-8, band 5 | 1 June 2013 | 30 | 7541, 7721 | 1T |
| | T | RapidEye, band 5 | 23 April 2013 | 5 | 5000, 5000 | 3A |
| INTRA1 | R | Sentinel-2A, band 7 | 11 January 2017 | 20 | 5490, 5490 | 1C |
| | T | Sentinel-2A, band 8A | 11 January 2017 | 20 | 5490, 5490 | 1C |
| INTRA2 | R | TerraSAR-X, X-band | 10 February 2014 | 2.75 | 19,090, 15,636 | EEC |
| | T | TerraSAR-X, X-band | 1 January 2015 | 2.75 | 19,090, 15,636 | EEC |

## 4. Results and Discussion

### 4.1. Inter-Sensoral Use Cases

Figure 4 demonstrates reference (A) and target image (B) of sample use case "INTER1" and the calculated tie point grid before (C) and after (D) removing false-positives. Colour coding visualizes detected absolute shift vector lengths in metres with a scale reaching from green (tie points with small displacements) to red (large displacements). The satellite images cover an area of approximately 110 by 110 km of Brandenburg, Germany, that is dominated by agriculture and an urban agglomeration in the southeast (Berlin, Potsdam). For the reference image, a 15-m sharpened, cloud-free Landsat-8 composite [41] has been used. Approximately 50% of the target image (Sentinel-2A) is covered by thick clouds. A cloud mask was intentionally excluded during the calculation of tie points to demonstrate the effectiveness of false-positive detection. These false-positives are highlighted in Figure 4c for the filtering techniques "reliability", "MSSIM" and "RANSAC" as explained in Section 2.3.2. As can be observed in the upper part of Figure 4c, reliability filtering identified the most false-positives (381 TPs), followed by RANSAC (123 TPs) and MSSIM (17 TPs). Tie points marked as false-positives are almost exclusively located at cloudy positions, where reliable calculation of image displacements is truly impossible due to the opaqueness of the clouds. In addition, some TPs have been rejected at the western image corner of the target image. Especially at these positions, the size of the matching windows is automatically reduced to avoid artificial image corners within the matching window (see Section 2.2.5). Consequently, the accuracy of detected displacements decreases due to the smaller window size, leading to rejection by subsequent validation algorithms. The remaining tie points (Figure 4d) show a clear gradient of absolute shift lengths from southeast to northwest that are already apparent from visual comparisons of the two input images. The gaps within the tie point grid can, on the one hand, be explained by filtered false-positives and, on the other hand, by positions where no valid match could be found after reaching the maximum number of iterations (see Section 2.3.2). Detected displacements vary between 26 m and 41 m corresponding to 2.6 to 4.1 Sentinel-2A pixels and relate to the fact that a global cloud-free reference image for Sentinel-2 is not yet available and that errors in the Global Land Survey (GLS) framework corrupted the rectification of Landsat-8 data [25]. After image warping, the remaining spatial displacements have been quantified in a range between 2.1 m and 6.4 m (2%/98% quantiles without false-positives) corresponding to 0.14 to 0.43 pixels of the reference image spatial resolution, which represents a quite satisfactory result considering that no cloud mask has been provided.
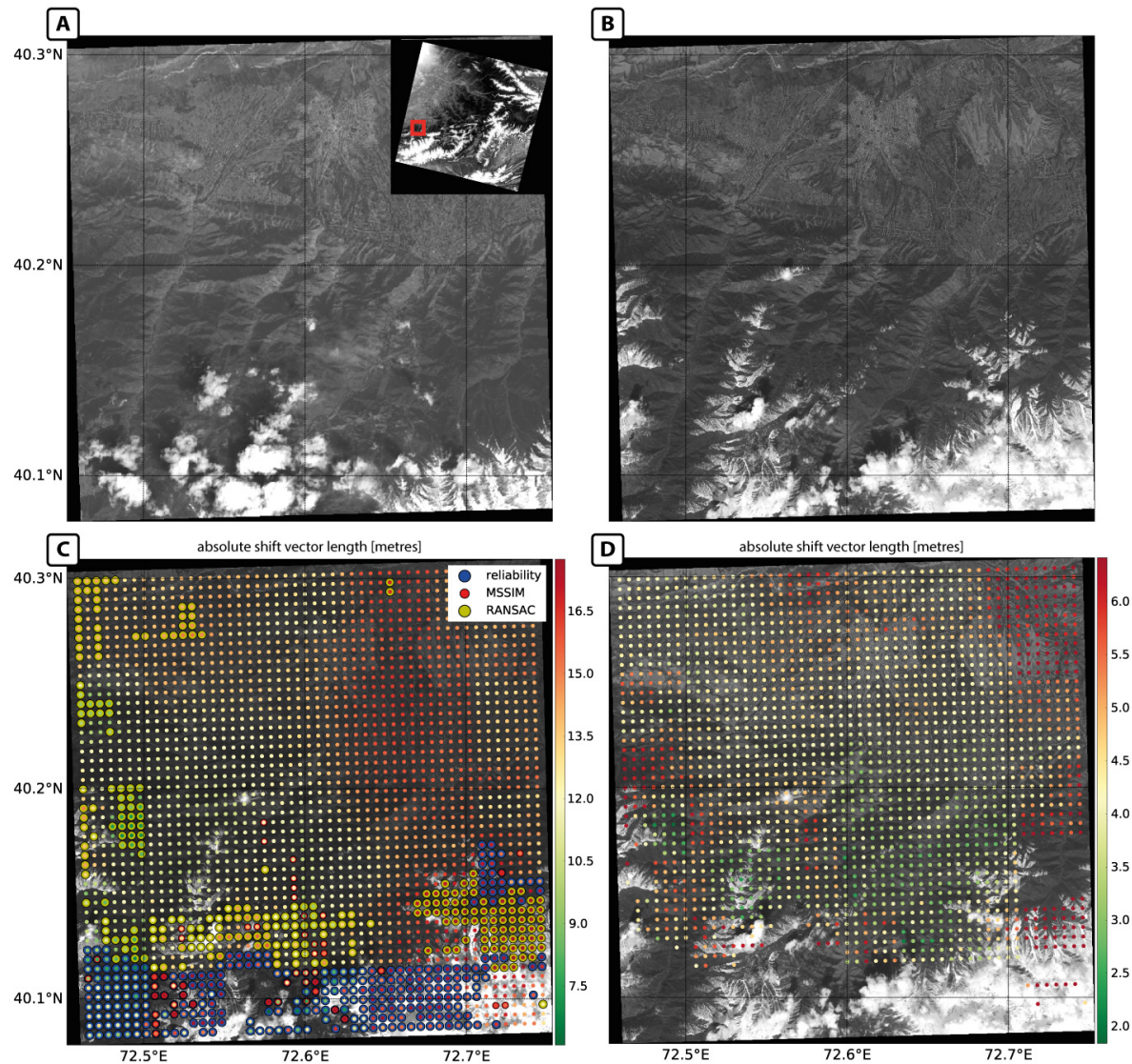
**Figure 4.** Use case INTER1: (**A**) Reference image (Landsat-8); (**B**) target image (Sentinel-2A); (**C**) calculated tie point grid including false-positives (absolute shift vector length in metres); and (**D**) calculated tie point grid used for image warping (absolute shift vector length in metres).

For the second inter-sensoral use case (INTER2), the algorithm was applied to a RapidEye-5 (target image, Figure 5a) and a Landsat-8 scene (reference, Figure 5b). The overlap area covers a 25 by 25 km mountainous region in western Kyrgyzstan. The image pair has been chosen to show co-registration performance in case of greatly varying ground sampling distances (5 vs. 30 m) and in the context of large altitudinal differences, especially in the southern part of the overlap area. Again, a cloud mask has been omitted. The resulting tie point grid (Figure 5c) reveals a non-linear pattern of spatial displacements with absolute shift lengths between 9.6 m and 16.7 m (2%/98% quantiles without false-positives) corresponding to up to 3.3 RapidEye pixels. This could also be confirmed visually. False-positives have mainly been identified in the most southern part of the RapidEye image, which can, on the one hand, be linked to cloud and cloud shadow coverage in both input images and alternatively to steep slopes. Such mountainous regions are particularly prone to orthorectification inaccuracies that cause geometric artefacts in the input data, leading to a higher probability of outliers within the tie point grid. However, after image warping, the geometric displacements were reduced to 2.1 to 6.5 m—a result within the sub-pixel range with respect to the 6.5-m native ground sampling distance of RapidEye during acquisition, which is later resampled

to 5-m pixel size during provider-side pre-processing [42]. Moreover, the tie point grid has been calculated at a 30 m resolution, since co-registration is always performed at the pixel size of the lower resolved input dataset (see Section 2.2.4). Therefore, a co-registration accuracy with residual shifts up to 6.5 m constitutes a result far in the sub-pixel domain of Landsat-8. With respect to the detected uneven pattern of spatial displacements, a global correction is insufficient for achieving an acceptable co-registration result.



**Figure 5.** Use case INTER2: (**A**) Reference image subset (Landsat-8, band 5), small figure demonstrates full extent; (**B**) target image (RapidEye-5, band 5); (**C**) calculated tie point grid before correction (absolute shift vector length in metres) including false-positives; and (**D**) calculated tie point grid after correction (absolute shift vector length in metres) excluding false-positives.

### 4.2. Intra-Sensoral Use Cases

As a sample use case for quantifying misregistrations between multiple bands of the same multi-spectral satellite dataset (use case INTRA1), a recent Sentinel-2A granule covering 110 by 110 km of northwestern Poland was chosen (Figure 6a). The proposed algorithm was applied to bands 7 (reference image, 783 nm) and 8A (target image, 865 nm), both with a spatial resolution of 20 m. For a more detailed visualization, a very dense grid with a mesh width of 30 pixels (500 m on ground)

resulting in more than 33,000 TPs was calculated. The results are shown in Figure 6b (absolute shift length) and 6c (shift direction).
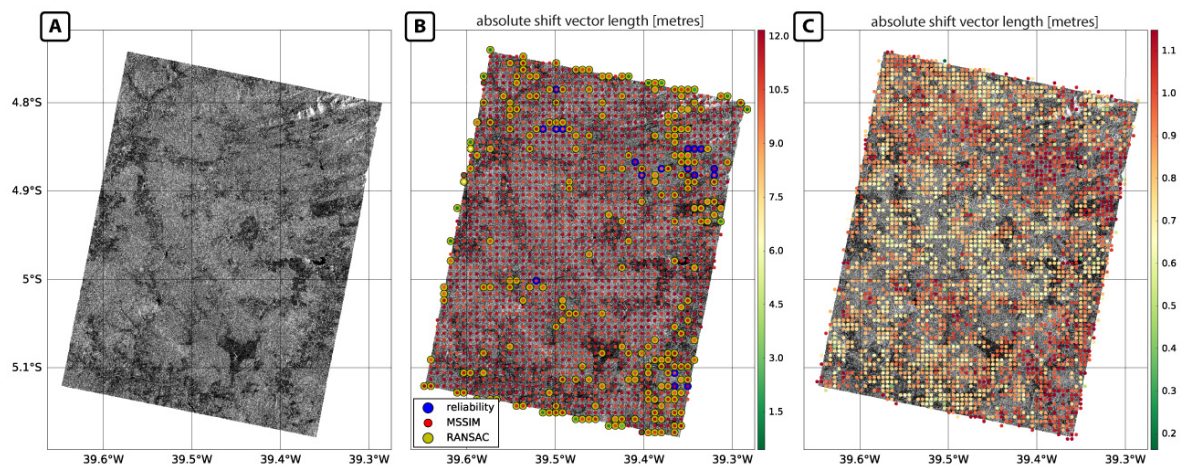


**Figure 6.** Use case INTRA1. (**A**) Target image (Sentinel-2A, band 8A); (**B**) calculated tie point grid (absolute shift vector length in metres); and (**C**) calculated tie point grid (shift direction in degrees from north).

Detected alignment errors vary between 1.3 m and 6.6 m and are thus within the mission requirements of 0.3 pixels at 3 σ confidence level regarding band-to-band geometric performance of Sentinel-2A [43]. Characteristic co-registration error patterns follow the flight line direction and are likely also related to the individual Sentinel-2 detector arrays and with the sensor's orbit track and movements in space (Figure 6b,c). These patterns arise in the form of block-like error clusters with a width of approximately 25 km that occur exactly in the direction of instrument overpass; they are due to slight geometric misalignments between neighbouring Sentinel-2 detectors. Since a single detector involved in the acquisition of a 20-m resolved Sentinel-2A band has a width of 1296 pixels [44], the mapped distance on the ground equals 25.9 km (including detector overlap). Furthermore, the calculated tie point grid reveals wave-like patterns in a distance of 7 to 9 km that appear perpendicular to the flight line. This can be addressed to micro-vibrations of the sensor during its overpass in space, which is also supported by the calculated shift directions shown in Figure 6c. These directions show similar patterns, depending on the image position and the respective detector array. Due to the already excellent co-registration of the analysed Sentinel-2 bands, a correction of the detected displacements has been omitted, considering the alteration of pixel values induced by resampling.

Apart from the optical satellite data, the proposed algorithm has been successfully utilized for multi-temporal co-registration of more than 80 TerraSAR-X radar (X-band) satellite images [45]. One example (use case INTRA2) is shown in Figure 7. The scene covers a rural area of 32 by 42 km of the province Ceará, northeastern Brazil. Calculated spatial displacements vary between 10.4 m and 12.2 m (detected false-positives excluded) corresponding to 3.8 to 4.4 TerraSAR-X pixels (Figure 7b). Sub-pixel co-registration is reached after image warping with a maximum remaining shift of 0.4 pixels (1.1 m on ground, Figure 7c). Visual inspections after shift correction confirmed excellent co-registration results.

**Figure 7.** Use case INTRA2: (**A**) Target image (TerraSAR-X, X-band); (**B**) calculated tie point grid before correction (absolute shift vector length in metres) including false-positives; and (**C**) calculated tie point grid after correction (absolute shift vector length in metres) excluding false-positives.
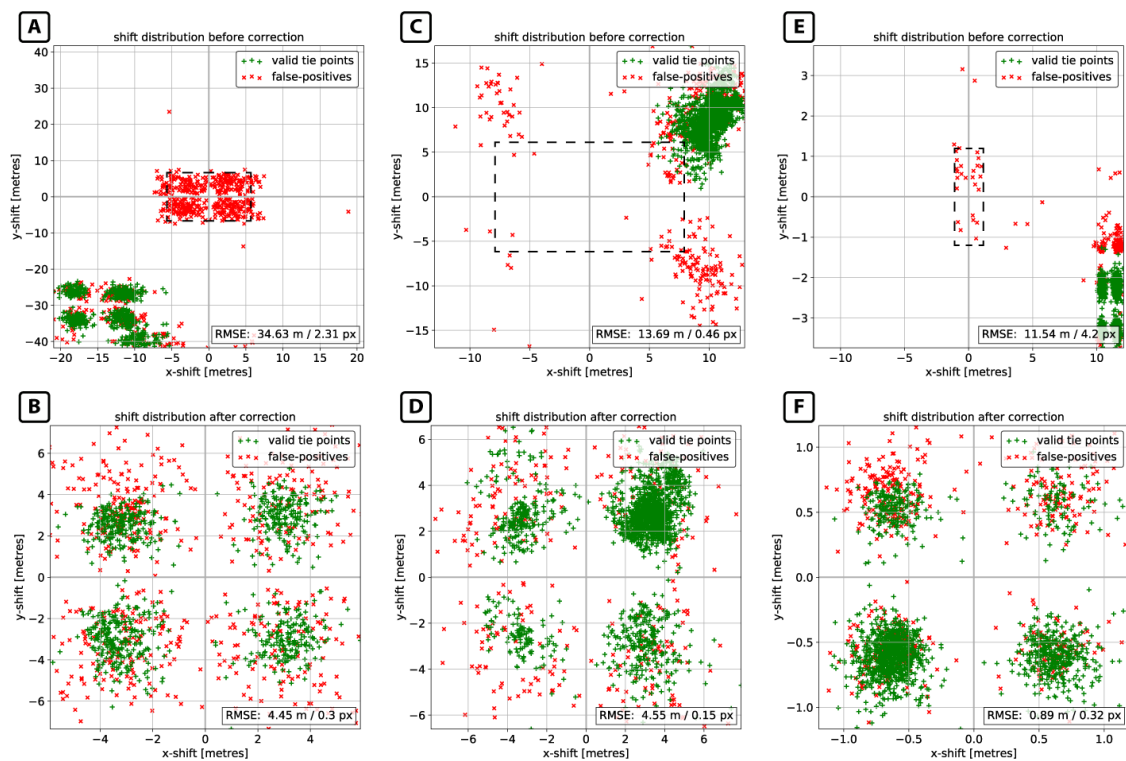
## 5. Algorithm Performance

### 5.1. Co-Registration Performance

A comparison of root mean square (RMSE) errors of detected displacements before and after the correction was conducted to quantitatively assess the overall performance of the proposed co-registration workflow. The intra-sensoral use case on Sentinel-2 band-to-band co-registration (INTRA1) was omitted here, because a correction of the detected sub-pixel shifts has not been considered to be beneficial (Section 4.2). Figure 8 demonstrates the detected shift distributions before and after correction for the use cases INTER1, INTER2 and INTRA2. Green markers show valid TPs, whereas red markers stand for false-positives. The indicated RMSE values were calculated based on all absolute X/Y shifts within the whole tie point grid, after the exclusion of false-positives.

The shift distribution demonstrates two major TP clusters for use case INTER1 (Sentinel-2/Landsat-8; Figure 8a). The obvious separation of valid and false-positive TPs in the image frame can be explained by cloudy versus clear-sky positions of the target image (see Figure 4). TPs at cloudy positions have been reliably identified as false-positives by the proposed combination of outlier detection techniques. All valid TPs show stable results regarding the determined shift direction (5.8 degrees standard deviation). After correction (Figure 8b), the RMSE of spatial shifts could be reduced from 34.63 m to 4.45 m (2.31 to 0.3 pixels of the lower resolved reference image).

The shift distribution of use case INTER2 (RapidEye-5/Landsat-8, Figure 8c) reveals shift vectors pointing to opposing directions, but colour coding confirms that directional outliers could be safely identified as false-positives. The large variance of detected valid shifts comes along with the uneven pattern of displacements visualized in Figure 8. The initial RMSE of 13.69 m decreased to 4.55 m, corresponding to 0.15 pixels of the lower resolved 30 m reference image after image warping.

The intra-sensoral use case INTRA2 (TerraSAR-X multi-temporal co-registration, Figure 8e,f) demonstrates a shift distribution without remarkable variation in terms of shift direction and absolute shift vector length. TPs that significantly differ from the primary cluster were marked as false-positives. After shift correction, an overall RMSE of 0.89 m (0.32 TerraSAR-X pixels) was achieved compared to an initial RMSE of 11.54 m (4.2 TerraSAR-X pixels).

**Figure 8.** Distribution of detected geometric shifts before and after correction. RMSE in pixel units always refers to the spatial resolution during image matching: (**A**,**B**) use case INTER1—Sentinel-2/Landsat-8; (**C**,**D**) use case INTER2—RapidEye-5/Landsat-8; and (**E**,**F**) use case INTRA2—TerraSAR-X. The dashed boxes in (**A**–**C**) indicate the axis limits of (**B**–**F**) for easier comparison of geometric shifts before and after correction.
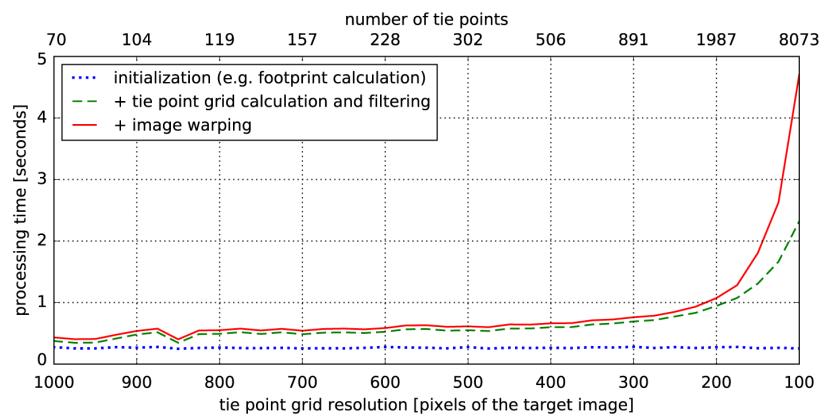
We nevertheless note that the corrected images still show directional variations of remaining shifts that are particularly evident in Figure 8f. However, the respective X/Y shifts clearly do not exceed sub-pixel domain precision with respect to the spatial resolution during image matching. Experiments based on more than 40 TerraSAR-X datasets (not presented here) show no dependency on the regular distribution of tie points within the dense grid. Moreover, no correlation to the number of tie points and the matching window size could be determined. Further research is needed to completely clarify the origin of these patterns. Thus far, we attribute them to our assumption that initial misregistration patterns can be modelled by a more or less affine transformation. This preserves straight lines and planes but only allows scaling, translation and rotation operations [13]. Consequently, the remaining shifts with directional variations are to be expected since this assumption is respected during detection of false-positives and during calculation of transformation parameters for image warping. Assuming a more complex transformation model, e.g., a higher polynomial one, may reduce these effects.

## 5.2. Computational Efficiency

The computational efficiency of the proposed algorithm was analysed by a benchmark based on use case INTER1 (Sentinel-2/Landsat-8) executed on a server with 32 CPU cores (2.3 GHz) and 256 GB working memory running an Ubuntu Linux system. For the benchmark presented here (Figure 9), we performed local co-registration for INTER1 (Section 4.1) with varying tie point grid resolutions between 1000 and 100 pixels of the target image, creating an increasing number of tie points from 70 up to 8073 TPs (Figure 9, second x-axis). Time measurements have been divided into: (1) initialization of the co-registration workflow, including disk access, image footprint detection and coordinate grid equalization; (2) tie point grid calculation and filtering; and (3) image warping and shift correction.

The initialization is always executed in single-CPU processing, whereas the remaining procedure uses all available CPUs. A co-registration with up to nearly 2000 TPs can be performed within one minute total processing time (Figure 9). Further time can be saved, especially by using fewer TPs for co-registration, which may be sufficient for simple kinds of misregistration patterns, e.g., when using only 100 TPs the processing time decreases to approximately 30 s whereas the proportion of workflow initialization amounts to more than 50%. There is a time overhead in multi-processing originating from additional multi-processing specific background tasks (Python object serialization processes), which indicates that, in single-processing, less time per CPU core is needed to process the same number of TPs.



**Figure 9.** Benchmark results of the proposed co-registration workflow based on use case INTER1 with varying tie point grid resolution.

*5.3. Performance Comparison to State of the Art Co-Registration Workflows*

Various aspects have to be considered when comparing the overall performance of the AROSICS framework with existing state of the art co-registration workflows (Section 1). From the pure perspective of registration accuracy, the implemented algorithm based on Foroosh et al., 2002 [19] achieves accuracies better than $1e^{-3}$ pixel in most scenarios [22], which is an excellent result compared with, e.g., 2D Gaussian peak fit [46], achieving $1e^{-1}$ pixel under optimal conditions [22]. However, this may be easily outperformed by other techniques, e.g., those proposed by Averbuch and Keller, 2002 [47] or Rogass et al., 2013 [22], allowing for accuracies up to $1e^{-6}$ pixel [22]. However, these values only apply for global co-registration at high SNR conditions. In case of local co-registration featuring image warping based on affine transformation parameters, deviations from the affine model, e.g., caused by topographic effects, decrease registration accuracy. Nonetheless, the intention of the proposed algorithm is not primarily to outperform existing co-registration accuracies but rather to achieve a high level of automation and robustness, coupled with generic and operational applicability while reliably reaching local co-registration accuracy in the sub-pixel range. For local co-registration, the AROSICS framework achieved root mean square errors of 0.3, 0.15 and 0.32 pixels (use cases INTER1, INTER2 and INTRA2, Section 5.1) using of an affine transformation model. Yan et al., 2016 [18] achieved 0.286, 0.302 and 0.303 pixels for three Landsat-8 and Sentinel-2A image pairs in case of affine transformation and showed that a polynomial transformation could not significantly improve these results. Regarding computational speed in case of global co-registration, Rogass et al., 2013 [22] compared the implemented phase correlation [19] method with: (1) their own co-registration algorithm; (2) a Fourier based approach proposed by Averbuch and Keller, 2002 [47]; and (3) with 2D Gaussian peak fit [46]. The approach of Foroosh et al., 2002 [19] outperformed them. However, the computational load for local co-registration mainly depends on the required number of tie points for co-registration and is therefore directly connected to the complexity of misregistration.

## 6. Conclusions

We presented a workflow, called AROSICS, for robust and automated detection and correction of local and global geometric displacements between multi-sensoral and multi-temporal satellite images in the sub-pixel range. The algorithm is based on a phase correlation approach that we extended by various additional steps to: (1) ensure appropriate co-registration inputs independently from sensor and data internal characteristics; (2) compute dense tie point grids, while handling acquisition specific conditions such as cloud coverage or land cover dynamics; and (3) effectively identify and filter false-positives before applying the correction. The results demonstrate that the proposed algorithm safely handles high degrees of heterogeneity between the input images and delivers co-registration results in the sub-pixel domain for both optical and radar satellite images. Registration accuracy always depends on the spatial resolution of the input images, the complexity of misregistration patterns, the similarity of the input images in terms of surface coverage dynamics, the signal-to-noise ratio of the input images and the chosen user-specified parameters. The presented algorithm has been implemented in Python as open-source software (GPLv3 license) and is now available for download at https://gitext.gfz-potsdam.de/danschef/arosics. Simple usability is achieved by an automated software installation routine and support for various raster image formats compatible with the Geospatial Data Abstraction Library (GDAL) [29]. Moreover, it requires a minimum of obligatory input parameters, offers an easy-to-use command line interface and a corresponding Python application programming interface (API) and can thus be easily implemented into existing remote sensing and big data workflows. The algorithm is especially designed for multi-core computing and offers various possibilities for reducing computational cost, e.g., by excluding no-data areas or cloud occlusions. Until now, it has been successfully utilized for multi-sensoral co-registration of more than 8500 Sentinel-2 images of Germany (with Landsat-8 as spatial reference) [48], for multi-temporal co-registration of approximately 80 TerraSAR-X acquisitions [45] and approximately 150 Landsat-5 and Landsat-7 images and for various further multi-sensoral and multi-temporal test cases proving the operational applicability of the proposed algorithm.

In summary, the innovation of the AROSICS framework is particularly seen in its robustness against the typical registration difficulties of multi-sensoral and multi-temporal satellite data, and its generic and operational applicability to a wide range of optical and radar remote sensing applications embedded into an easy-to-use, open-source software package that is deployable across different platforms and can be implemented into existing processing workflows, even without Python-specific programming skills. This enables easy improvements to the data basis for subsequent analyses such as monitoring or parameter mapping applications and to increase the temporal resolution of time series by including accurately co-registered multi-sensoral data. In the context of automated big data processing systems that aim to offer compatibility to various multi-sensoral and multi-temporal satellite data, the proposed algorithm represents an easy way to effectively realize fully automatic detection and correction of image misalignments with sub-pixel accuracy.

Future work may include the extension of the existing co-registration framework to image datasets that do not have geocoding information such as medical image data or not georeferenced raster geodata, e.g., scan stripes of imaging laboratory spectrometers. At present, artificial geocoding information has to be associated to such datasets to perform co-registration using AROSICS. Automation could be based on the global approach presented here for computing a first co-registration guess, which is then refined by applying the local co-registration approach. Apart from that, further robustness against rotational misregistrations would increase the applicability to use cases where image rotation is an issue. For that purpose, phase correlation would no longer suffice and another rotation-invariant core co-registration algorithm, e.g., one based on Fourier–Mellin transform [20,34], would need to be implemented.

## References

1. Townshend, J.R.G.; Justice, C.O.; Gurney, C.; McManus, J. The Impact of Misregistration on Change Detection. *IEEE Trans. Geosci. Remote Sens.* **1992**, *30*, 1054–1060. [CrossRef]
2. Dai, X.; Khorram, S. The effects of image misregistration on the accuracy of remotely sensed change detection. *IEEE Trans. Geosci. Remote Sens.* **1998**, *36*, 1566–1577.
3. Liang, J.; Liu, X.; Huang, K.; Li, X.; Wang, D.; Wang, X. Automatic Registration of Multisensor Images Using an Integrated Spatial and Mutual Information ( SMI ) Metric. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 603–615. [CrossRef]
4. Sundaresan, A.; Varshney, P.K.; Arora, M.K. Robustness of Change Detection Algorithms in the Presence of Registration Errors. *Photogramm. Eng. Remote Sens.* **2007**, *73*, 375–383. [CrossRef]
5. Behling, R.; Roessner, S.; Segl, K.; Kleinschmit, B.; Kaufmann, H. Robust Automated Image Co-Registration of Optical Multi-Sensor Time Series Data: Database Generation for Multi-Temporal Landslide Detection. *Remote Sens.* **2014**, *6*, 2572–2600. [CrossRef]
6. Swain, P.H.; Vanderbilt, V.C.; Jobusch, C.D. A quantitative applications-oriented evaluation of thematic mapper design specifications. *IEEE Trans. Geosci. Remote Sens.* **1982**, *75*, 370–377. [CrossRef]
7. Le Moigne, J.; Netanyahu, N.S.; Eastman, R.D. *Image Registration for Remote Sensing*, 1st ed.; Cambridge University Press: Cambridge, UK, 2011.
8. Dalmiya, C.P.; Dharun, V.S. A Survey of Registration Techniques in Remote Sensing Images. *Indian J. Sci. Technol.* **2015**, *8*, 1–7. [CrossRef]
9. Zitová, B.; Flusser, J. Image registration methods: A survey. *Image Vis. Comput.* **2003**, *21*, 977–1000. [CrossRef]
10. Chen, H.; Arora, M.K.; Varshney, P.K. Mutual information based image registration for remote sensing data. *Electr. Eng. Comput. Sci.* **2003**, *24*, 3701–3706. [CrossRef]
11. Long, T.; Jiao, W.; He, G.; Zhang, Z. A fast and reliable matching method for automated georeferencing of remotely-sensed imagery. *Remote Sens.* **2016**, *8*, 56. [CrossRef]
12. Dawn, S.; Vikas, S.; Sharma, B. Remote Sensing Image Registration Techniques: A Survey. In *Image and Signal Processing. ICISP 2010. Lecture Notes in Computer Science*; Elmoataz, A., Lezoray, O., Nouboud, F., Mammass, D., Meunier, J., Eds.; Springer: Berlin/Heidelberg, Germany, 2010; Volume 6134, pp. 103–112.
13. Brown, L.G. A survey of image registration techniques. *ACM Comput. Surv.* **1992**, *24*, 325–376. [CrossRef]
14. Gianinetto, M. Automatic Co-Registration of Satellite Time Series. *Photogramm. Rec.* **2012**, *27*, 462–470. [CrossRef]
15. Wong, A.; Clausi, D.A. ARRSI: Automatic registration of remote-sensing images. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 1483–1492. [CrossRef]
16. Leprince, S.; Barbot, S.; Ayoub, F.; Avouac, J.P. Automatic and Precise Orthorectification and Coregistration and Subpixel Correlation of Satellite Images, Application to Ground Deformation Measurements. *IEEE J. Geosci. Rem. Sens.* **2007**, *45*, 1529–1558. [CrossRef]
17. Gao, F.; Masek, J.G.; Wolfe, R.E. Automated registration and orthorectification package for Landsat and Landsat-like data processing. *J. Appl. Remote Sens.* **2009**, *3*, 33515.
18. Yan, L.; Roy, D.; Zhang, H.; Li, J.; Huang, H. An Automated Approach for Sub-Pixel Registration of Landsat-8 Operational Land Imager (OLI) and Sentinel-2 Multi Spectral Instrument (MSI) Imagery. *Remote Sens.* **2016**, *8*, 520. [CrossRef]

19. Foroosh, H.; Zerubia, J.B.; Berthod, M. Extension of phase correlation to subpixel registration. *IEEE Trans. Image Process.* **2002**, *11*, 188–199. [CrossRef] [PubMed]

20. Chen, Q.; Defrise, M.; Deconinck, F. Symmetric Phase-Only Matched Filtering of Fourier-Mellin Transforms for Image Registration and Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **1994**, *16*, 1156–1168. [CrossRef]

21. Heid, T.; Kääb, A. Evaluation of existing image matching methods for deriving glacier surface displacements globally from optical satellite imagery. *Remote Sens. Environ.* **2012**, *118*, 339–355. [CrossRef]

22. Rogass, C.; Segl, K.; Kuester, T.; Kaufmann, H. Performance of correlation approaches for the evaluation of spatial distortion reductions. *Remote Sens. Lett.* **2013**, *4*, 1214–1223. [CrossRef]

23. Joglekar, J.; Gedam, S.S. Area Based Image Matching Methods—A Survey. *Int. J. Emerg. Technol. Adv. Eng.* **2012**, *2*, 130–136.

24. Tong, X.; Ye, Z.; Xu, Y.; Liu, S.; Li, L.; Xie, H.; Li, T. A Novel Subpixel Phase Correlation Method Using Singular Value Decomposition and Unified Random Sample Consensus. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 4143–4156. [CrossRef]

25. Storey, J.; Roy, D.P.; Masek, J.; Gascon, F.; Dwyer, J.; Choate, M. A note on the temporary misregistration of Landsat-8 Operational Land Imager (OLI) and Sentinel-2 Multi Spectral Instrument (MSI) imagery. *Remote Sens. Environ.* **2016**, *186*, 121–122. [CrossRef]

26. Yoon, Y.T.; Eineder, M.; Yague-Martinez, N.; Montenbruck, O. TerraSAR-X precise trajectory estimation and quality assessment. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 1859–1868. [CrossRef]

27. Young, N.E.; Anderson, R.S.; Chignell, S.M.; Vorster, A.G.; Lawrence, R.; Evangelista, P.H. A survival guide to Landsat preprocessing. *Ecology* **2017**, *98*, 920–932. [CrossRef] [PubMed]

28. Bracewell, R.N. *The Fourier Transform and Its Applications*; McGraw-Hill: New York, NY, USA, 1965.

29. Open Source Geospatial Foundation. GDAL Raster Formats. Available online: http://www.gdal.org/formats_list.html (accessed on 19 April 2017).

30. Luan, X.; Yu, F.; Zhou, H.; Li, X.; Song, D.; Wu, B. Illumination-robust area-based stereo matching with improved census transform. In Proceedings of the 2012 International Conference on Measurement, Information and Control (MIC), Harbin, China, 18–20 May 2012; Volume 1, pp. 194–197.

31. Frigo, M.; Johnson, S.G. The design and implementation of FFTW3. *Proc. IEEE* **2005**, *93*, 216–231. [CrossRef]

32. Frigo, M.; Johnson, S. *The Fastest Fourier Transform in the West. Technical Report MIT-LCS-TR-728*; MIT Lab for Computer Science: Cambridge, MA, USA, 1997.

33. Kuglin, C.D.; Hines, D.C. The phase correlation image alignment method. In Proceedings of the IEEE Conference on Cybernetics and Society, Banff, AB, Canada, 1–4 October 1975; pp. 163–165.

34. Keller, Y.; Averbuch, A.; Israeli, M. Pseudo-polar based estimation of large translations rotations and scalings in images. *IEEE Trans. Image Process.* **2005**, *14*, 12–22. [CrossRef] [PubMed]

35. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef] [PubMed]

36. Avanaki, A.N. Exact global histogram specification optimized for structural similarity. *Opt. Rev.* **2009**, *16*, 613–621. [CrossRef]

37. Fischler, M.A.; Bolles, R.C. Random Sample Consensus: A Paradigm for Model Fitting with Applicatlons to Image Analysis and Automated Cartography. *Commun. ACM* **1981**, *24*, 381–395. [CrossRef]

38. Hast, A.; Nysjö, J.; Marchetti, A. Optimal RANSAC—Towards a repeatable algorithm for finding the optimal set. *J. WSCG* **2013**, *21*, 21–30.

39. Toldo, R.; Fusiello, A. Automatic estimation of the inlier threshold in robust multiple structures fitting. In *Image Analysis and Processing—ICIAP 2009. ICIAP 2009. Lecture Notes in Computer Science*; Foggia, P., Sansone, C., Vento, M., Eds.; Springer: Berlin/Heidelberg, Germany, 2009; Volume 5716, pp. 123–131.

40. Subbarao, R.; Meer, P. Beyond RANSAC: User independent robust regression. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, New York, NY, UAS, USA 17–23 July 2006; Volume 2006.

41. Earthstar Geographics LCC. TerraColor® Global Landsat Satellite Images of Earth. Available online: http://www.terracolor.net/index.html (accessed on 19 April 2017).

42. Chander, G.; Haque, M.O.O.; Sampath, A.; Brunn, A.; Trosset, G.; Hoffmann, D.; Roloff, S.; Thiele, M.; Anderson, C. Radiometric and geometric assessment of data from the RapidEye constellation of satellites. *Int. J. Remote Sens.* **2013**, *34*, 37–41. [CrossRef]

43. Clerc, S. Sentinel-2 Data Quality Report Issue 1 (November 2015). *Ref. S2-PDGS-MPC-DQR, European Space Agency (ESA)*. Available online: https://earth.esa.int/documents/247904/685211/Sentinel-2+Data+Quality+Report+Issue+01+%28November+2015%29/74295984-60d0-479e-b6ea-f04e25a84f5a?version=1.2 (accessed on 22 April 2017).

44. Gascon, F.; Thépaut, O.; Jung, M.; Francesconi, B.; Louis, J.; Lonjou, V.; Lafrance, B.; Massera, S.; Gaudel-Vacaresse, A.; Languille, F.; et al. Copernicus Sentinel-2 Calibration and Products Validation Status. *Remote Sens.* **2017**, *9*, 584. [CrossRef]

45. Zhang, S.; Foerster, S.; Medeiros, P.; Waske, B. Monitoring water extent in macrophyte covered reservoirs in NE Brazil using TerraSAR-X time series data. (submitted).

46. Nobach, H.; Honkanen, M. Two-dimensional Gaussian regression for sub-pixel displacement estimation in particle image velocimetry or particle position estimation in particle tracking velocimetry. *Exp. Fluids* **2005**, *38*, 511–515. [CrossRef]

47. Averbuch, A.; Keller, Y. FFT based image registration. In Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing, Orlando, FL, USA, 13–17 May 2002; pp. IV-3608–IV-3611.

48. Helmholtz-Centre Potsdam—GFZ German Research Centre for Geosciences. GFZ Time Series System for Sentinel-2. Available online: http://www.gfz-potsdam.de/gts2 (accessed on 19 April 2017).