

MULTI-SATELLITE IMAGE ALIGNMENT OVER LARGE AREAS WITH FEATURELESS REGIONS

Constantine J. Roros, Rahul Deshmukh, Avinash C. Kak

(croros, deshmun5, kak)@purdue.edu

Robot Vision Lab, Purdue University, School of Electrical and Computer Engineering, West Lafayette, IN, USA

KEY WORDS: Image Alignment, Photogrammetry, DSM, Classification, Change Detection

ABSTRACT:

There is a great deal of interest in fusing together the information provided by different satellites for real-time change detection on the surface of the earth. For detecting important types of changes on the ground, it is necessary to inject geometry into the data provided by low-res satellites using multi-view imaging satellites that record each point on the ground from multiple perspectives. Combining these perspectives and generating a DSM (Digital Surface Model) gives us the geometry needed for a more meaningful analysis of satellite data. Before such an analysis can be carried out, it is necessary to align the images from all available satellites. Automatic image alignment, however, requires features on the ground that can be identified and correctly matched across different images using computer vision algorithms. While such features are common in urban areas, that is not always the case in predominantly rural areas that present a more-or-less uniform texture to the sensors. In this paper we present methods for automatic identification and alignment of featureless regions. Featureless regions are identified using point spread maps, which are a byproduct of DSM generation. The subsequent strategy for aligning featureless regions depends on the proportion of featureless regions to feature-rich regions. If most of the AOI (Area of Interest) is feature-rich, we ignore featureless regions when estimating inter-satellite image alignment parameters and apply those parameters to the entire AOI. Finally, we present a technique to propagate and fuse parameters from feature-rich regions to featureless regions.

1. INTRODUCTION

Monitoring the changes on the surface of the earth, and doing so in real-time to the extent possible, has acquired new importance on account of the proliferation of earth imaging satellites. The changes of interest include those that are man-made for civilian and military applications and those that may be caused by natural phenomena such as flooding, earthquakes, forest fires, hurricanes, etc. Since different satellites yield information at different spatial resolutions and in different spectral bands, ideally a change detection framework would fuse together all of the evidence available from all the satellites before drawing inferences regarding the changes on the ground. An important component of this data fusion would be the high-precision geometry provided by the multi-view high-res satellites for understanding the changes more accurately, both spatially and semantically. As to how one can extract high-precision geometry from multi-view high-res satellite images, it could be through the sort of algorithms presented in (Comandur and Kak, 2021). And, as to the need for such geometry, we will just repeat what has now become a common refrain among algorithm developers for satellite images: Without the geometry, you would not be able to tell the difference between a change taking place in a large parking lot from a similar change taking place on the roof of a Walmart store.

Before one can attempt a fusion of multi-satellite data in the sense described above, it is necessary to align all the images from the different satellites over a given AOI (Area of Interest). In the work described in this paper, our goal is to align the low-res Landsat and Sentinel images with the multi-view high-res Maxar images.

A key step in the multi-satellite image alignment work described in this paper is the construction of a DSM (Digital Surface

Model) from the multi-view high-res images. A DSM indicates height values at each lat/long point on the ground at a higher resolution than, say, the 30m SRTM DEM (Shuttle Radar Topography Mission Digital Elevation Model). We use the DSM to orthorectify with high precision the bundle-adjusted¹ off-nadir high-res images from Maxar. Subsequently, the high-res orthorectified images are aligned with the low-res Landsat and Sentinel images in our MuRA (Multi Resolution Alignment) framework (Deshmukh et al., 2023) as described in the rest of this paper.

The process described above works quite well for AOIs that are of primarily urban areas. The main reason as to why the urban areas produce better results than the more rural areas is that the DSM's over the latter tend to be noisy. The DSM generation logic is based on extracting keypoints from the high-res Maxar images, the keypoints being the high-variance pixels in the images (that is, the pixels where the data is varying maximally in all directions). The keypoints may be extracted using one of a large number of algorithms now available for that purpose, including the more recently proposed deep-learning based solutions. We have obtained best results with the algorithm FAST+VGG (Rosten and Drummond, 2006, Simonyan and Zisserman, 2015) in which FAST is used for keypoint extraction and the VGG neural network for estimating the descriptor vectors associated with the keypoints. In urban areas, the keypoints extracted by such algorithms correspond to different types of corners on building roofs and facades, road intersections, road markings, etc. However, in the more rural areas, especially when there are plowed fields and wooded regions in the areas, the same keypoint extractors yield a random assortment of points. Subsequently, when these points are fed into a ste-

¹ The Appendix at the end of the paper presents further information on how we carry out bundle adjustment of the multi-view Maxar images.

reconstruction algorithm for generating the DSM, one ends up with 3D noise, as we will show in this paper. The bottom line is that the noisy DSMs constitute unreliable height values and those, in turn, result in poorly orthorectified imagery, making multi-satellite image alignment difficult. In the rest of this paper, we will refer to those portions of an AOI where the DSM is noisy as the “featureless regions”.

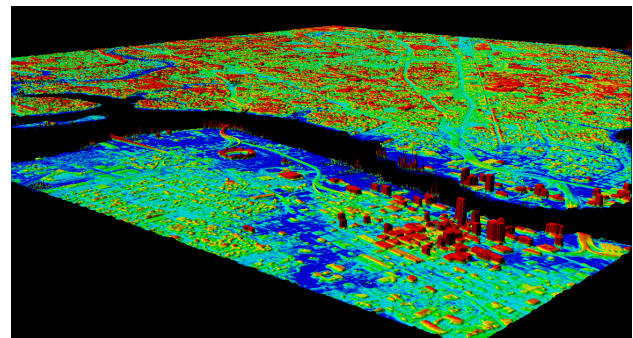
As an example, in Figure 1 we show DSMs generated for an urban area rich with features in Jacksonville, FL and for comparison show in Figure 2 DSMs generated for a more rural area of fields next to a town in Demmin, Germany. In Figure 2a, the lack of correctly matched features in the field resulted in noisy spikes that dominate the height range when visualizing the DSM, preventing the buildings in the adjacent town from being displayed as clearly as the buildings. This portion of Demmin is a smaller part of a larger 800 km² AOI that contains both towns with good features and featureless countryside. When presented with such an AOI, the issues are twofold: (1) We must determine what portions of the AOI have reliable DSM (i.e., without noisy spikes) that can produce good quality ortho-images. (2) Once the reliable portions of the AOI have been identified, we need to estimate alignment parameters from those portions and apply those parameters to the rest of the AOI. In this paper, we present an extension of the MuRA framework that automatically identifies regions of an AOI with trustworthy DSMs and propagates the alignment parameters estimated from the feature-rich to the featureless regions.

The rest of our paper is outlined as follows: First we briefly discuss previous work for multi-satellite image alignment. In Section 3 we explain the MuRA framework. Then we present our method for detecting featureless regions in Section 4 and our alignment parameter propagation scheme in Section 5. This section explains what exactly is meant by propagating alignment parameters and how parameters are fused when there are multiple sets of parameters to choose from. Finally, in Section 6, we apply our updated framework to the previously mentioned 800 km² AOI in Germany and compare relative alignment accuracy with and without the featureless region logic using USGS’ GAASS (Geometric Alignment and Assessment) metric.

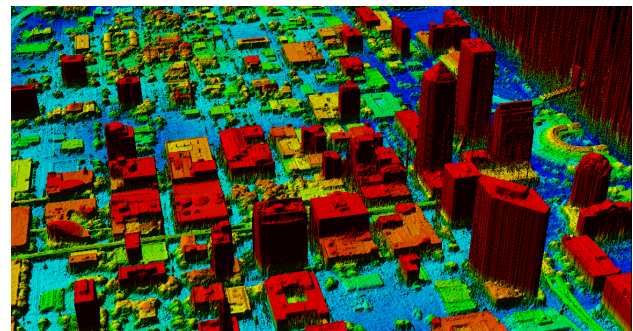
2. RELATED WORK

Over the years there have been several works focused on multi-satellite image alignment. The common thread in all of these methods is to compare either features or intensities between a reference image and an unaligned image. Based on this comparison, a transformation is estimated and applied to the unaligned image.

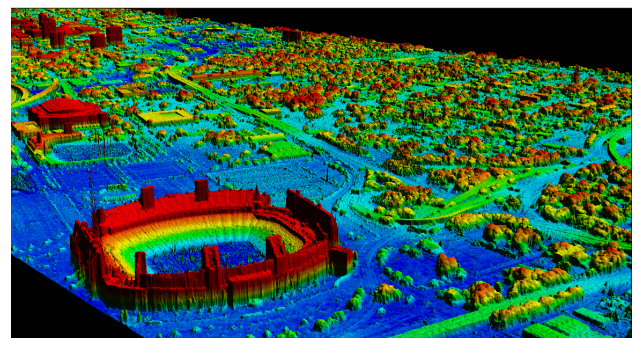
ARRSI (Automatic Registration of Remote-Sensing Images) (Wong and Clausi, 2007) detects features of interest using a phase congruence model (Kovesi, 1999) to align Landsat 7 and Spaceborne Imaging Radar-C/X-Band Synthetic Aperture Radar (SIR-C/X-SAR). NASA’s AROP (Automated Registration and Orthorectification Package) (Gao et al., 2009) uses cross-correlation to align Landsat and ASTER images by estimating a polynomial model corresponding to either a shift, affine, or quadratic correction. Similar to AROP, the work in (Behling et al., 2014) also uses cross-correlation to align ASTER and SPOT images to Landsat reference images, but only estimates a shift correction for more robust and efficient alignment. The framework described in (Yan et al., 2016) aligns pairs of Landsat and



(a) A 100 km² DSM generated from 29 Maxar images at (downsampled) 1.5 m resolution



(b) Details from a 2.5 km² DSM generated from 28 Maxar images at 0.25m resolution

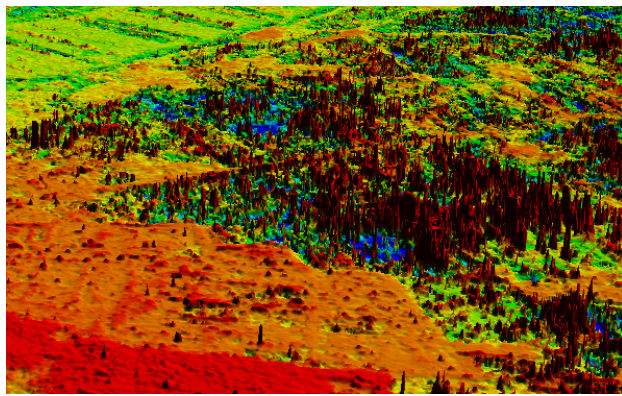


(c) Details from another 2.5 km² DSM generated from 28 Maxar images at 0.25m resolution

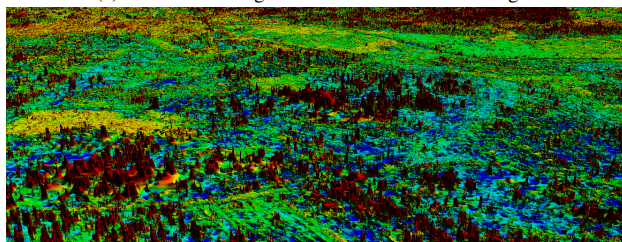
Figure 1. DSMs from an AOI in downtown Jacksonville, FL, USA. We typically generate large-area DSMs at 1.5m resolution and small-area DSMs at 0.25m resolution.

Sentinel images by constructing a Gaussian image pyramid, extracting features at each level using the Förstner corner detector (Förstner and Gülch, 1987) and performing least squares matching. EMMOSI (Liu et al., 2017) aligns images from the Gaofen and Ziyuan3 satellites by finding initial feature matches with a modified SIFT operator known as Uniform Robust (UR)-SIFT, then “propagating” (i.e., detecting additional) matches using geometric correspondences and probabilistic relaxation.

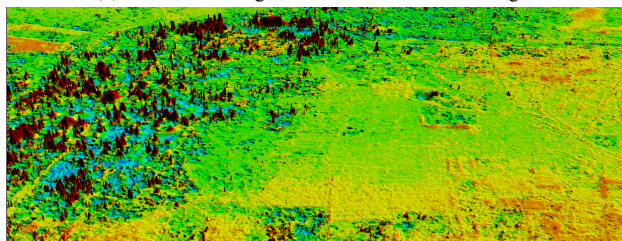
The method proposed in (Skakun et al., 2017) uses phase correlation (Foroosh et al., 2002) for estimating alignment parameters between Landsat and Sentinel. Phase correlation involves transforming images to the frequency domain, then calculating alignment parameters using the cross-power spectrum of the frequency domain images. There is also a comparison in (Skakun et al., 2017) of different methods of modeling the alignment transformation: polynomial models, radial basis functions, and non-linear random forests. The AROSICS frame-



(a) DSM for tile 1 generated from 19 Maxar images



(b) DSM for tile 3 generated from 27 Maxar images



(c) DSM for tile 8 generated from 24 Maxar images

Figure 2. Noisy DSMs for 90 km² tiles (as labelled in Figure 6) of an AOI in Demmin, Germany that were generated from Maxar images at (downsampled) 1.5 m resolution

work (Automated and Robust Open-Source Image Coregistration Software) in (Scheffler et al., 2017) also uses the concept of phase correlation to align Landsat and Sentinel images. This framework has the option to estimate alignment parameters from an “image subset” within the AOI and apply them to the entire AOI. This subset region size is manually specified, and the location is chosen based on which portions of the reference and unaligned image overlap (assuming these two images do not have the same footprint). The framework described in our paper detects subset regions for parameter estimation automatically based on DSM noise and includes additional case by case logic for propagating and fusing alignment parameters. In order to understand this logic, we must first review MuRA.

3. MULTI-SATELLITE IMAGE ALIGNMENT

The MuRA framework (Deshmukh et al., 2023) takes inspiration from AROP (Gao et al., 2009) to align lower-resolution satellite images to an image from a higher-resolution satellite. Figure 3 shows a diagram of the steps of MuRA. When explaining MuRA, we refer to the lower resolution images we want to align, typically from Landsat or Sentinel, as warp images and the higher-resolution image we are aligning to, typically from Maxar or Planet, as the base image. While it is technically possible to use Landsat or Sentinel as the base image

for MuRA, doing so results in much worse alignment accuracy than if Maxar is used. This is because Maxar images are rich with geometric information, so we can perform intra-satellite alignment of these images with much higher absolute accuracy than with Landsat or Sentinel. MuRA aligns warp images to the

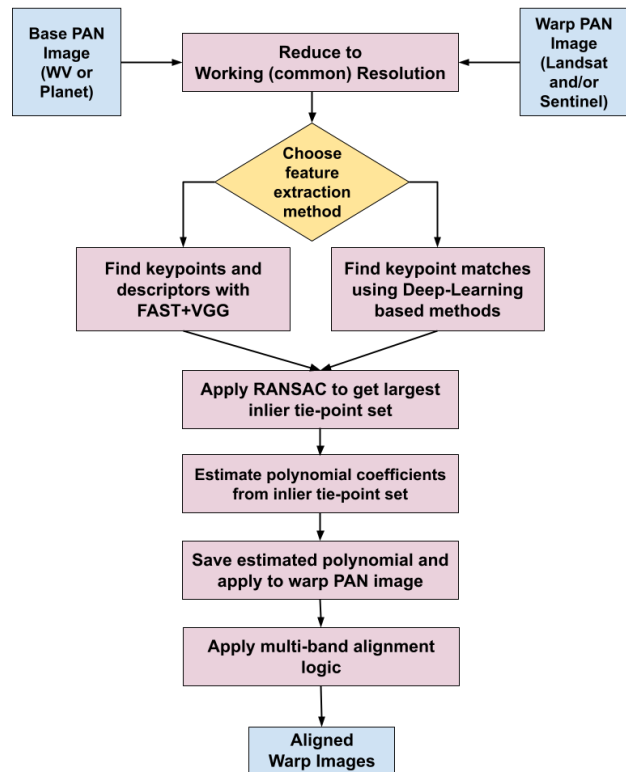


Figure 3. Overview of MuRA framework

base image by estimating and applying a warp polynomial. The polynomial can be chosen to apply a shift, affine, or quadratic correction based on the polynomial order. In our experience, a shift correction provides the best relative accuracy (based on the GAASS metric) for aligning the warp images. Formally, for each pixel coordinate in a warp image (\hat{x}_w, \hat{y}_w) , we want to estimate (in the 0-order case) the polynomial coefficients a_1 and b_1 such that

$$\begin{aligned} \hat{x}_w &= a_1 + x_b \\ \hat{y}_w &= b_1 + y_b \end{aligned} \quad (1)$$

Where (x_b, y_b) are coordinates in the base image. While the polynomial is estimated using images from the Panchromatic band, it can be applied to imagery of other bands.

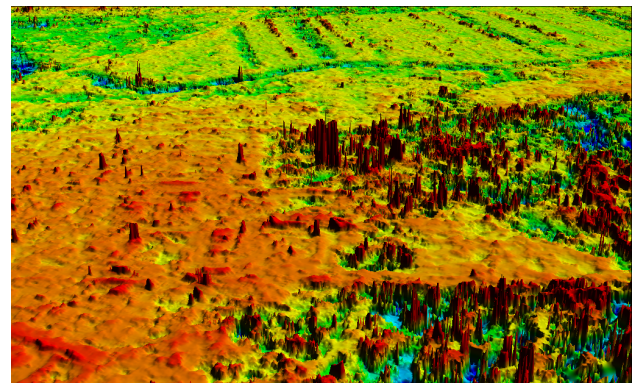
The steps of estimating the warp polynomial are as follows. First the base and warp images are reduced to a common “working” resolution. Then keypoints are identified and matched in both images using an algorithm for feature extraction and matching such as SIFT (Lowe, 1999) or FAST+VGG (Rosten and Drummond, 2006, Simonyan and Zisserman, 2015). The idea is that the feature extraction and matching algorithm of choice can be used in MuRA on a plug-n-play basis. Once we match features in both images, we use RANSAC to find the largest inlier tiepoint set. This inlier tie point set is used to estimate the coefficients for the warp polynomial by solving a linear least square problem for a_1 and b_1 that minimizes the reprojection error for (\hat{x}_w, \hat{y}_w) .

The MuRA framework is usually able to align Landsat warp images to a Maxar base image with subpixel relative accuracy for AOIs that are rich in features. AOIs that are richer in features or have good quality ortho base images are more likely to have a larger inlier tiepoint set than those that are relatively featureless. Since the polynomial coefficients are estimated using linear least squares, a larger inlier tiepoint set would result in polynomial coefficients that yield higher alignment accuracy. When an AOI has a featureless regions, the resulting ortho-images have artifacts that can lead to false matches and therefore a smaller inlier tiepoint set. In the next section, we describe a method for detecting featureless regions that would yield poor quality ortho-images. By detecting these regions, we can subsequently apply MuRA only to portions of the AOI that would provide the largest inlier tiepoint sets.

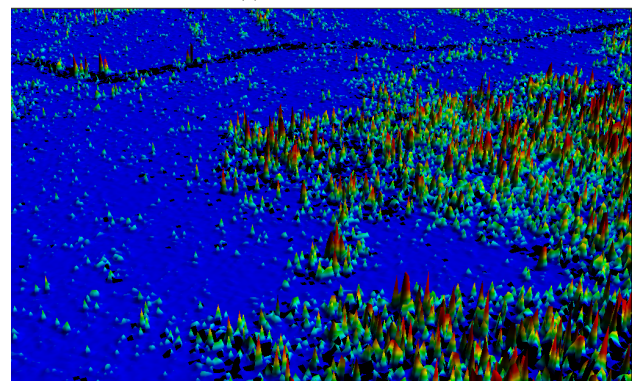
4. AUTOMATICALLY DETECTING FEATURELESS REGIONS

To determine what regions in an AOI MuRA should estimate polynomial coefficients from, we need to detect what regions of the AOI are featureless. We consider a region featureless if the DSM generated for that region has noisy spikes like the DSM shown in Figure 2. When performing stereo matching in featureless regions, we are more likely to have erroneous matches due to the homogeneity of textures. These erroneous matches can result in incorrect Z-values during triangulation, hence the resulting spikes in the DSM. Detecting featureless regions then becomes an issue of how to automatically find these noisy spikes in a DSM. Fortunately, a byproduct of DSM generation known as a point spread map can indicate where these spikes lie. In Figure 4, notice that the spiky portions of the DSM (Figure 4a) correspond to higher point spread values in Figure 4b and featureless fields in Figure 4c.

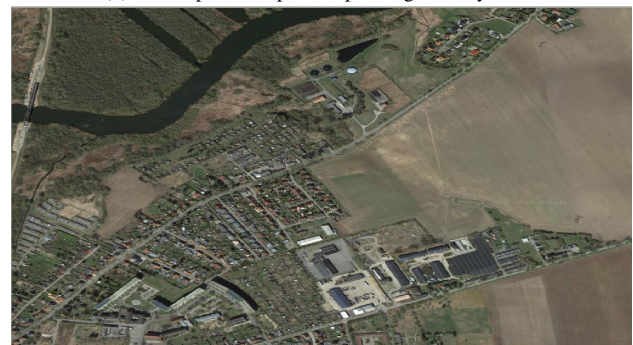
In order to explain what a point spread map is, we must first briefly describe the DSM generation process. Given a set of multi-view images we form stereo pairs such that each image in the pair has sufficiently different viewing angle to make triangulation possible (Patil et al., 2019a). These pairs are then rectified using an approach similar to that of (de Franchis et al., 2014), where we can assume an affine camera model for sufficiently small image tiles. We use the dense stereo matching algorithm tMGM (Patil et al., 2019b) to produce a disparity map for finding dense correspondences for the stereo pair. These dense correspondences are needed to perform triangulation, which results in a point cloud. We will have a point cloud for each stereo pair, which are then fused into a single point cloud. We can assume that at each lat/long location in a given AOI, we will have a collection of points from this fused point cloud. To get the final height value for the DSM at a given lat/long location, we take the median Z-value of the top N highest points at that location (N is a predetermined hyperparameter). The point spread value at that lat/long location is given by taking the standard deviation (instead of the median) of the Z-values of those points. Therefore, a point spread map indicates the standard deviation of height values at any location in a DSM, and effectively measures the uncertainty for the height values. Since we want to classify regions of the AOI as featureless or feature-rich, we use Otsu's method to threshold the point spread map, yielding a binary mask for the entire AOI.



(a) DSM with noise



(b) Point Spread Map corresponding to noisy DSM



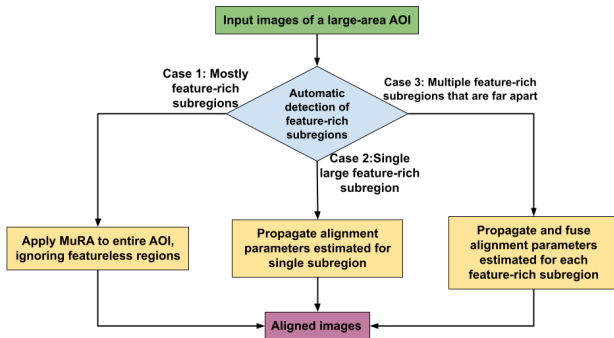
(c) Google Earth image corresponding to noisy DSM

Figure 4. Comparison of a noisy DSM and its point spread map. Also shown is the corresponding Google Earth image showing that the noisy portion of the DSM is a field

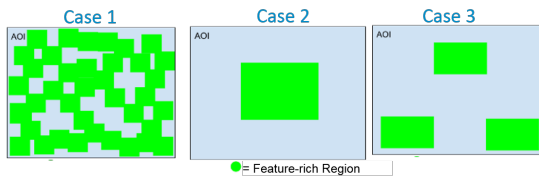
5. A FRAMEWORK FOR ALIGNING FEATURELESS REGIONS

Once the featureless regions of an AOI have been detected, estimating polynomial coefficients from the feature-rich regions is relatively straightforward. We can use our binary mask to apply MuRA only to the portions with usable DSM. However, the method by which we apply these alignment parameters (i.e., polynomial coefficients) to the featureless regions is a bit more complicated. Therefore we break down our featureless region alignment procedure into three different strategies (Figure 5a) used for three different AOI cases (Figure 5b).

For case 1, if AOI is mostly feature-rich, we apply MuRA to the entire AOI and simply ignore featureless regions using the mask. If there is a single region that has a usable DSM (case2), then we only apply MuRA to that region. The alignment para-



(a) Overall framework for automatic detection and alignment of featureless regions



(b) Hypothetical AOIs where each case logic is appropriate

Figure 5. The strategy for aligning featureless regions of an AOI depends on how much of the AOI is feature-rich and how concentrated those feature-rich regions are

parameters are then propagated out to the featureless regions by simply applying these parameters to the entire AOI. Case 3 is when there are multiple regions that are feature-rich, but are too spread out to ignore featureless regions like in case 1. In such a case, we would estimate alignment parameters only from the feature-rich regions as in case 2, but when propagating these parameters we would need to fuse multiple sets of polynomial coefficients to align featureless regions.

To align a given point in the featureless region of an AOI using the case 3 logic, we apply an aggregate polynomial calculated using a weighted average of all polynomial coefficients estimated from feature-rich regions. The weighting of each set of polynomial coefficients is based on how much we “trust” those alignment parameters for aligning a given point. The factors we consider for our level of trust include how close the point we are aligning is to each feature-rich region and the average point spread value of each of these regions. The reasoning for this is that a point is more likely to be accurately aligned using parameters that were estimated from a region in close proximity to that point. However, if that region has a relatively noisy DSM, its alignment parameters may not be as accurate as, say, parameters from a region that is further away but has a much cleaner DSM. For a point \mathbf{x} that we want to align, we compute its alignment parameters $\{a_i^x\}, \{b_i^x\}$ from the parameters of r feature-rich regions as follows:

$$\begin{aligned} \{a_i^x\} &= \sum_{j=1}^r \{a_i^j\} \frac{1}{p^j d^j} \\ \{b_i^x\} &= \sum_{j=1}^r \{b_i^j\} \frac{1}{p^j d^j} \end{aligned} \quad (2)$$

Where d^j is the normalized distance from center of the j th feature-rich region to \mathbf{x} and p^j is that region’s average point spread value.



Figure 6. A Google Earth image of the 800 km² AOI in Germany split into numbered tiles used for our experiments.

6. EXPERIMENTAL RESULTS AND DISCUSSION

To demonstrate the improvements of this extension to the MuRa framework, we aligned the previously mentioned 800 km² test AOI in Germany both with and without a mask. For DSM generation, the AOI was split into nine 90 km² tiles (as visualized in Figure 6 with Google Earth imagery) of Maxar (WorldView-3) imagery that was downsampled to 1.5m resolution.



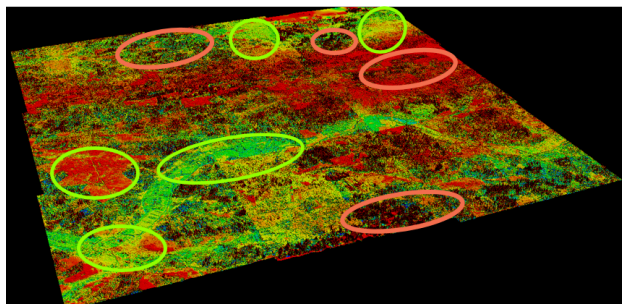
Figure 7. The mask generated from the point spread maps for the Germany AOI. The white regions indicate the featureless areas that need to be masked out

The mask generated for this AOI is shown in Figure 7, and the effect of this mask on removing noise in the DSM can be seen in Figure 8. Notice that all of the red circled regions indicating noisy spikes in Figure 8a are masked out in Figure 8b. To highlight how much noise is removed in the DSM, in Figure 9 we zoomed into several tiles.

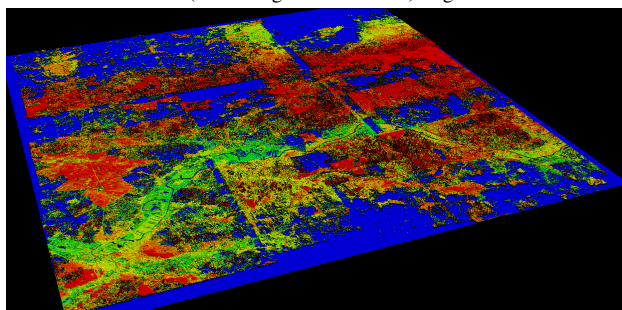
Once we orthorectified the available Maxar images, we aligned 275 Landsat 8 images. We used case 1 logic, where MuRa is applied to the entire AOI with the featureless regions masked out, for this AOI since most of it is feature-rich as indicated by Figure 7. In general, case 1 is the most common case we encountered in AOIs we align with MuRa. To evaluate the relative alignment accuracy of our framework, we used the GAASS metric developed by USGS. This metric operates on pairs of aligned images closest in date recorded, with one image labelled the “reference image” and the other labelled the “comparison image”. Each of these images are split into 100 × 100 corresponding patches. Then for each pair of corresponding patches, we compute the NCC (Normalized Cross Correlation)

Tile #	Tile 1		Tile 2		Tile 3		Tile 4		Tile 5		Tile 6		Tile 7		Tile 8		Tile 9	
Alignment Type	No Mask	Mask	No Mask	Mask	No Mask	Mask	No Mask	Mask	No Mask	Mask	No Mask	Mask	No Mask	Mask	No Mask	Mask	No Mask	Mask
CE90 Mean	6.977	5.793	7.674	6.971	6.995	6.035	8.079	6.545	6.957	5.648	8.991	6.234	3.911	3.911	13.371	8.0465	7.772	6.692
CE90 Std-Dev	7.839	4.313	6.501	4.606	6.451	4.502	8.07	4.031	6.653	4.18	9.046	4.671	N/A	N/A	9.716	3.195	8.639	5.693
CE90 Median	4.377	4.377	6.435	6.435	4.741	4.741	5.169	5.169	4.547	4.547	4.404	3.801	3.911	3.911	11.268	7.993	4.547	4.547
Percent Masked	N/A	0.214	N/A	0.401	N/A	0.324	N/A	0.354	N/A	0.354	N/A	0.34	N/A	0.551	N/A	0.484	N/A	0.507

Table 1. GAASS Metrics for the same set of images for masked and non-masked cases



(a) Original DSM with noisy rural areas circled in red areas with reliable DSM (urban or good rural areas) in green.

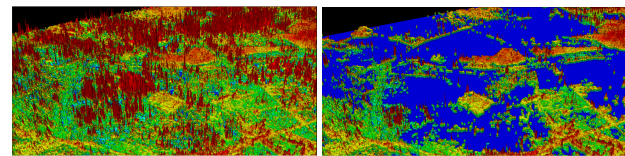


(b) Masked DSM. Blue regions indicate masked-out areas.

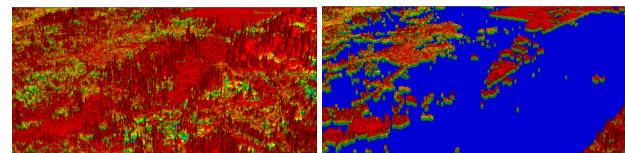
Figure 8. All tiles for the noisy DSM of 800 km² AOI in Germany before and after the noise mask is applied.

over a range of X and Y offsets by sliding the patch on the comparison image. For the set of NCC values in each sliding direction, we fit a quadratic function as shown in Figure 10. The minimum of this function in each direction is our error in meters. The error over all patches is computed based on the CE90 (circular error in the 90th percentile) (Subcommittee, 1998). We compute the mean, standard deviation, and median of the CE90 over all images pairs evaluated for relative accuracy.

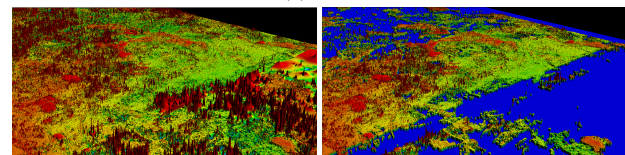
For a fair comparison between using a mask and not using one for MuRA, we compare GAASS metrics only for the same set of images in Table 1. While the median CE90 is comparable for both the mask and non-masked cases, the mean and standard deviation CE90 are usually much lower. While GAASS metrics are compared for the same set of images, the most striking improvement when using the case 1 logic is the number of successfully aligned images. When we say that an image is successfully aligned, that means that MuRA is able to estimate alignment parameters for that image. If MuRA is unable to find any inlier tiepoints between the base and warp image, then it cannot estimate alignment parameters. We think using the mask increases the number of successfully aligned images because it prevents MuRA from detecting bad keypoints in the featureless regions of an AOI and as a result reduces the likelihood of false matches. As shown in Table 2, the case 1 logic with the mask allows MuRA to align many more Landsat images than without the mask. For the purpose of change detection, having



(a) Tile 3



(b) Tile 5



(c) Tile 8

Figure 9. A sampling of DSM tiles before (left) and after (right) the mask is applied

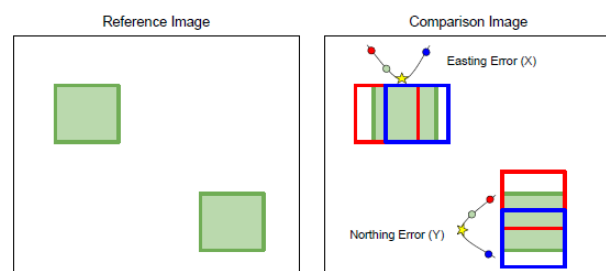


Figure 10. Diagram showing how the GAASS metric is computed between two images.

good coverage (i.e., a large number of aligned images) in a time series is important.

	Tile 1	Tile 2	Tile 3	Tile 4	Tile 5	Tile 6	Tile 7	Tile 8	Tile 9
Mask	275	233	275	275	274	273	245	265	168
No Mask	78	40	82	47	43	55	8	16	32

Table 2. A comparison of the number of Landsat images (out of 275 total) that can be successfully aligned when using a mask (i.e., case 1 alignment logic) and not using a mask during alignment parameter estimation.

7. CONCLUSION

This paper presents an extension of the MuRA framework for aligning the high-res and low-res satellite images over AOIs that contain featureless regions, these being regions that give rise to significant levels of noise in the DSMs. Using a thresholded point spread map, we can automatically detect which regions of the AOI are featureless. Once these regions are detected we can

use one of three possible strategies to align the AOI based on how much of the AOI is featureless. In the experiments shown in this paper, MuRA automatically detects and masks out the featureless regions of a large AOI in Germany, calculates the alignment parameters from the non-masked areas, and extends these parameters to the entire AOI. With this logic, MuRA is able to align many more images with this extension, while having comparable if not better relative alignment accuracy. Therefore the addition of alignment logic for featureless contributes meaningful improvement to MuRA. For future work, we plan on conducting experiments to evaluate the alignment logic with AOIs that fall under case 3. We also want to further evaluate the alignment accuracy of this extended framework, potentially calculating the absolute alignment accuracy using GCPs (Ground Control Points).

8. ACKNOWLEDGEMENTS

This research is based upon work supported in part by the Office of the Director of National Intelligence (ODNI, Intelligence Advanced Research Projects Activity (IARPA), via Contract #2021-21040700001. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes not withstanding any copyright annotation therein.

Satellite imagery used in DSMs © 2014–2021 Maxar Technologies.

REFERENCES

- Behling, R., Roessner, S., Segl, K., Kleinschmit, B., Kaufmann, H., 2014. Robust Automated Image Co-Registration of Optical Multi-Sensor Time Series Data: Database Generation for Multi-Temporal Landslide Detection. *Remote Sensing*, 6(3), 2572–2600. <https://www.mdpi.com/2072-4292/6/3/2572>.
- Comandur, B., Kak, A. C., 2021. Semantic Labeling of Large-Area Geographic Regions Using Multiview and Multidate Satellite Images and Noisy OSM Training Labels. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14, 4573–4594. <https://ieeexplore.ieee.org/document/9380895>.
- de Franchis, C., Meinhardt-Llopis, E., Michel, J., Morel, J.-M., Facciolo, G., 2014. An automatic and modular stereo pipeline for pushbroom images. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, II-3, 49–56. <https://www.isprs-ann-photogramm-remote-sens-spatial-inf-sci.net/II-3/49/2014/>.
- Deshmukh, R., Roros, C. J., Kashyap, A., Kak, A. C., 2023. An Aligned Multi-Temporal Multi-Resolution Satellite Image Dataset for Change Detection Research. <https://arxiv.org/abs/2302.12301>.
- Foroosh, H., Zerubia, J., Berthod, M., 2002. Extension of phase correlation to subpixel registration. *IEEE Transactions on Image Processing*, 11(3), 188–200.
- Förstner, W., Gülch, E., 1987. A fast operator for detection and precise location of distinct points, corners and centres of circular features. *Proc. ISPRS intercommission conference on fast processing of photogrammetric data*, 6, Interlaken, 281–305.
- Gao, F., Masek, J. G., Wolfe, R. E., 2009. Automated registration and orthorectification package for Landsat and Landsat-like data processing. *Journal of Applied Remote Sensing*, 3(1), 033515.
- Grodecki, J., Dial, G., 2003. Block adjustment of high-resolution satellite images described by rational polynomials. *Photogrammetric Engineering & Remote Sensing*, 69(1), 59–68.
- Kovesi, P., 1999. Image features from phase congruency. *Videre: Journal of computer vision research*, 1(3), 1–26.
- Liu, Y., Mo, F., Tao, P., 2017. Matching Multi-Source Optical Satellite Imagery Exploiting a Multi-Stage Approach. *Remote Sensing*, 9(12). <https://www.mdpi.com/2072-4292/9/12/1249>.
- Lourakis, M. I., Argyros, A. A., 2009. SBA: A software package for generic sparse bundle adjustment. *ACM Transactions on Mathematical Software (TOMS)*, 36(1), 1–30.
- Lowe, D. G., 1999. Object recognition from local scale-invariant features. *Proceedings of the seventh IEEE international conference on computer vision*, 2, Ieee, 1150–1157.
- Patil, S., Comandur, B., Prakash, T., Kak, A. C., 2019a. A New Stereo Benchmarking Dataset for Satellite Images. *arXiv:1907.04404*. <http://arxiv.org/abs/1907.04404>.
- Patil, S., Prakash, T., Comandur, B., Kak, A., 2019b. A Comparative Evaluation of SGM Variants (including a New Variant, tMGM) for Dense Stereo Matching. <https://arxiv.org/abs/1911.09800>.
- Rosten, E., Drummond, T., 2006. Machine learning for high-speed corner detection. *In European Conference on Computer Vision*, 430–443.
- Scheffler, D., Hollstein, A., Diedrich, H., Segl, K., Hostert, P., 2017. AROSICS: An Automated and Robust Open-Source Image Co-Registration Software for Multi-Sensor Satellite Data. *Remote Sensing*, 9(7). <https://www.mdpi.com/2072-4292/9/7/676>.
- Simonyan, K., Zisserman, A., 2015. Very deep convolutional networks for large-scale image recognition. *International Conference on Learning Representations*.
- Skakun, S., Roger, J.-C., Vermote, E. F., Masek, J. G., Justice, C. O., 2017. Automatic sub-pixel co-registration of Landsat-8 Operational Land Imager and Sentinel-2A Multi-Spectral Instrument images using phase correlation and machine learning based mapping. *International Journal of Digital Earth*, 10(12), 1253–1269. <https://doi.org/10.1080/17538947.2017.1304586>. PMID: 32021650.
- Subcommittee, F. G. C., 1998. Geospatial positioning accuracy standards part 3:national standard for spatial data accuracy.
- Wong, A., Clausi, D. A., 2007. ARRSI: Automatic Registration of Remote-Sensing Images. *IEEE Transactions on Geoscience and Remote Sensing*, 45(5), 1483–1493.
- Yan, L., Roy, D. P., Zhang, H., Li, J., Huang, H., 2016. An Automated Approach for Sub-Pixel Registration of Landsat-8 Operational Land Imager (OLI) and Sentinel-2 Multi Spectral Instrument (MSI) Imagery. *Remote Sensing*, 8(6). <https://www.mdpi.com/2072-4292/8/6/520>.

A. APPENDIX

This appendix presents an overview of our probabilistic framework for carrying out bundle adjustment of the multi-view Maxar images. This is a critical step in the generation of DSMs from Maxar images. High resolution satellite images use Rational Polynomial Coefficient (RPC) camera model ($\mathcal{P}_{rpc}(X) : \mathbb{R}^3 \mapsto \mathbb{R}^2$) which maps a 3D world point ($X \in \mathbb{R}^3$) to its corresponding pixel location ($x \in \mathbb{R}^2$) in the image. As demonstrated in (Grodecki and Dial, 2003), the RPC camera model can effectively be parameterized with only bias correction parameters ($\delta \in \mathbb{R}^2$) in image space coordinates to carry out bundle adjustment of cameras, we denote the parameterized camera as $\mathcal{P}_{rpc}(\delta, X)$.

We carry out bundle adjustment of Maxar data by formulating a Maximum A Posteriori (MAP) problem to estimate the bias correction parameters. When we are presented with a set of N images ($\{I_i\}_{i=1}^N$) with corresponding RPC cameras ($\{\mathcal{P}_{rpc}^{(i)}\}_{i=1}^N$) and a set of inlier tiepoints ($\{\mathcal{M}_{ij}\} \forall 1 \leq i < j \leq N : \mathcal{M}_{ij} = \{x_k^{(i)} \leftrightarrow x_k^{(j)} | x_k^{(i)} \in I_i, x_k^{(j)} \in I_j\} \forall 0 \leq k \leq |\mathcal{M}_{ij}|\}$), we associate a putative world point with every tiepoint ($x_k^{(i)} \leftrightarrow x_k^{(j)} \leftrightarrow X_k^{(ij)}$). The MAP problem can be converted into a minimization of regularized reprojection error as shown in Equation 3 to compute bias correction parameters for all images with zero-mean Gaussian prior.

$$\{\delta_i^*, \{X_k^{*(ij)}\}\} = \underset{\{\delta_i, \{X_k^{(ij)}\}\}}{\operatorname{argmin}} L_{rep} + \lambda L_{reg} \quad (3)$$

Where L_{rep} and L_{reg} are the reprojection error and regularization term respectively given by Equation 4 and λ is the weight of regularization term which is inversely proportional to the covariance of the Gaussian prior.

$$L_{rep} = \sum_{1 \leq i < j \leq N} \sum_{k=1}^{|\mathcal{M}_{ij}|} \left(\|x_k^{(i)} - \mathcal{P}_{rpc}^{(i)}(\delta_i, X_k^{(ij)})\|_2^2 + \|x_k^{(j)} - \mathcal{P}_{rpc}^{(j)}(\delta_j, X_k^{(ij)})\|_2^2 \right) \quad (4)$$

$$L_{reg} = \sum_{i=1}^N \|\delta_i\|_2^2$$

For our formulation we choose $\lambda = 0.5$ and solve the above minimization problem using sparse bundle adjustment similar to what is described in (Lourakis and Argyros, 2009).