KUVA SPACE

SMARTER DATA FOR A STRONGER PLANET

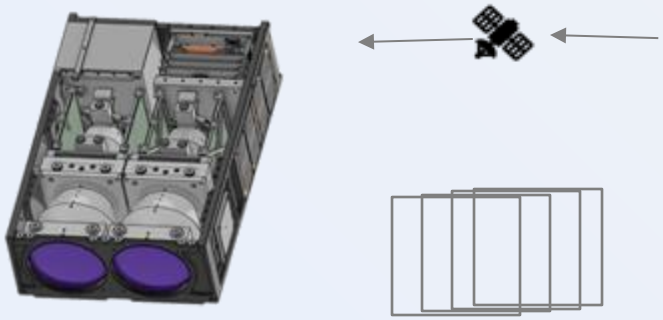
Band Alignment Benchmarking

Lennert Antson

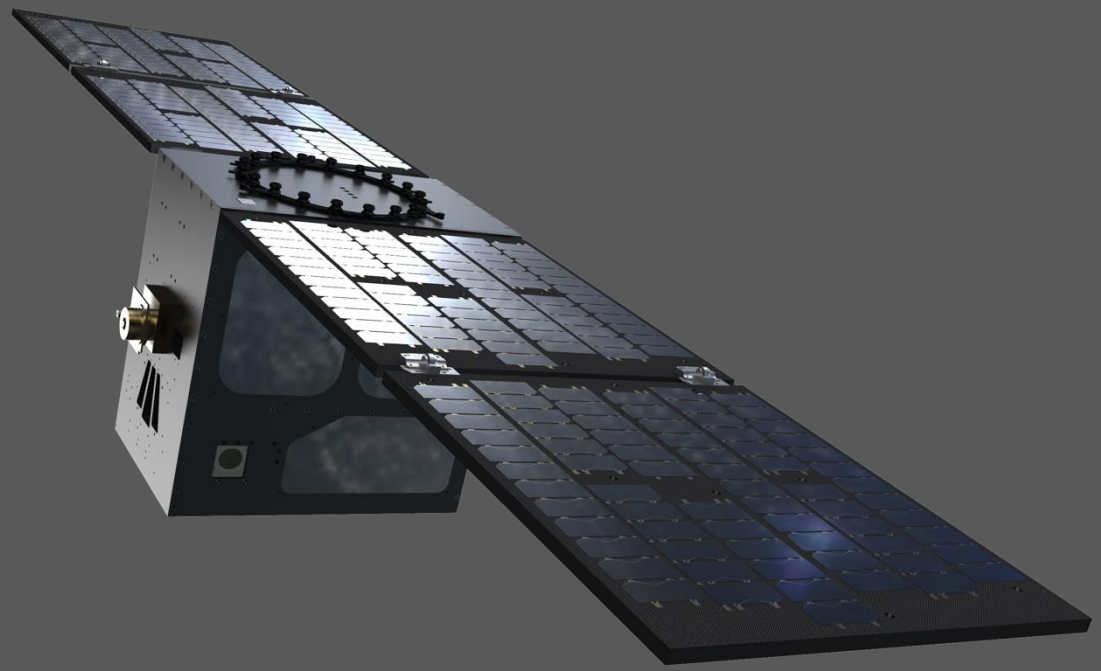
3rd WORKSHOP ON INTERNATIONAL COOPERATION IN SPACEBORNE IMAGING SPECTROSCOPY,
ESTEC, 13 Nov 2024

Challenges

- Kuva Space's Hyperfield payload is equipped with two or more hyperspectral snapshot cameras.
- Acquired images at a certain wavelength are misaligned with respect to each other.
- Therefore, alignment is mandatory to obtain a properly aligned hypercube to reconstruct the spectrum for each pixel in the image



Hyperfield-1 payload with two snapshot imagers (VIS+NIR).



Benchmark Objectives

- Align the images with subpixel accuracy
- Handle large spectral, contrast, and illumination changes, especially since images are acquired at different wavelengths.
- Handle large displacements due to translation and rotation

Deep Learning Based band alignment methods

Sparse keypoint detection and matching

- alike+NN [1]
- disk+lightglue [2]
- aliked+lightglue[3]
- superpoint +lightglue [4][5]
- superpoint +superglue [4][6]
- dedode [7]

Detects and matches distinctive key points across images, optimizing for speed and efficiency.

Semi-dense alignment

- loftr [8]
- eloftr [9]

Matches features over larger but not fully exhaustive regions, balancing detail and computational load.

Dense alignment

- RoMa [10]

Aligns all pixels or regions for maximum detail and contextual accuracy, ideal for complex and variable scenes.

Data Set

- 213 Sentinel-2 L2A (2023) images sampled all across the globe
- Covering various scenes (heterogeneous, homogenous, snow, desert, vegetation, forests, coastal areas, ..)
- Multiple resolutions: 10m, 20m, and 30m
- Multispectral: spectral differences can be simulated

Mosaic of used Sentinel-2 images in the benchmark (left), and their location (up)

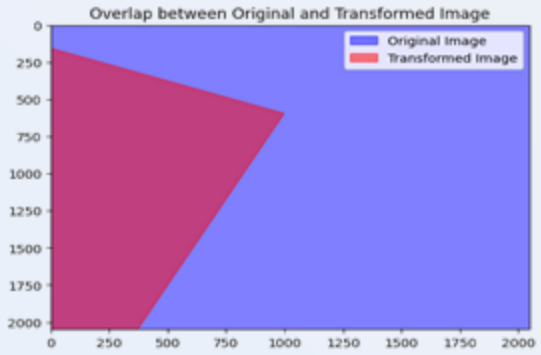
KUVA SPACE

Methodology

Generation of misalignment image pairs

Homography Sampling (100 homographies / S2-image)

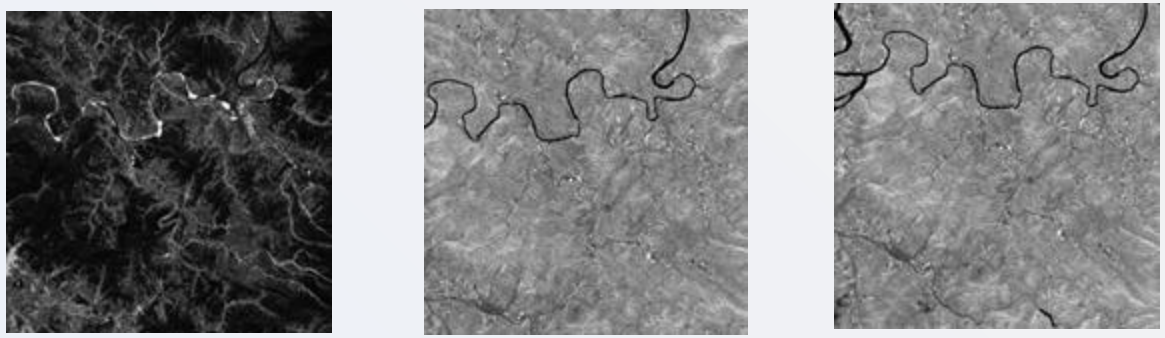
- Assuming an image shape of (2048, 2048)
- Randomly (uniformly) sample homography coefficients using the following ranges:
  - Translation in x, y: **[-1434; +1434]** (70% of 2048)
  - Rotation: **[-30°; +30°]**
  - Perspective coefficients in x, y: **[-5e-6; +5e-6]**
  - **Additional constraint** during random homography sampling:
    - The sampled homography must result in an **overlap** (IoU) between both images that **exceeds 25%**.



Randomly sampled homography resulting in 28.9% overlap.

Random Band Selection:

- Different bands are used to simulate spectral differences between the original and misaligned images.



The original image (e.g. B02) (left), the newly sampled band to simulate spectral differences (e.g. B08) (center), and the misaligned image (B08) using the randomly sampled homography (right).

KUVA SPACE

KUVA SPACE

SMARTER DATA FOR A STRONGER PLANET

Results

Results

Metric: mean_acc@px: mean accuracy based on the average distance between the original and re-aligned image corners below a specified pixel threshold.

	mean_acc@px → Methods ↓	@1px	@3px	@5px	@10px
Sparse keypoint detection & matching	alike+NN [1]	0.473	0.671	0.695	0.706
	disk+lightglue [2]	0.462	0.685	0.734	0.773
	aliked+lightglue[3]	0.513	0.801	0.846	0.879
	superpoint +lightglue [4][5]	0.671	0.862	0.895	0.920
	superpoint +superglue [4][6]	0.664	0.873	0.910	0.935
	dedode [7]	0.594	0.766	0.788	0.799
Semi-dense	loftr [8]	0.674	0.915	0.937	0.952
	eloftr [9]	**0.829**	**0.919**	**0.938**	**0.951**
Dense	RoMa [10]	**0.817**	**0.923**	**0.945**	**0.963**

Results at 10m resolution

mean_acc@px → Methods ↓	@1px	@3px	@5px	@10px
alike+NN	0.524	0.714	0.736	0.747
disk+lightglue	0.474	0.695	0.748	0.789
aliked+lightglue	0.521	0.817	0.859	0.892
superpoint +lightglue	0.701	0.893	0.926	0.945
superpoint +superglue	0.687	0.901	0.937	0.956
dedode	0.596	0.774	0.798	0.809
loftr	0.714	0.934	0.955	0.966
eloftr	**0.851**	**0.938**	**0.954**	**0.965**
RoMa	**0.814**	**0.926**	**0.949**	**0.968**

Results at 20m resolution

mean_acc@px → Methods ↓	@1px	@3px	@5px	@10px
alike+NN	0.440	0.610	0.637	0.653
disk+lightglue	0.452	0.666	0.719	0.760
aliked+lightglue	0.511	0.791	0.833	0.863
superpoint +lightglue	0.673	0.858	0.890	0.913
superpoint +superglue	0.667	0.866	0.898	0.918
dedode	0.539	0.720	0.747	0.760
loftr	0.653	0.889	0.919	0.939
eloftr	**0.749**	**0.849**	**0.875**	**0.895**
RoMa	**0.786**	**0.896**	**0.928**	**0.952**

Results at 30m resolution

Conclusions:

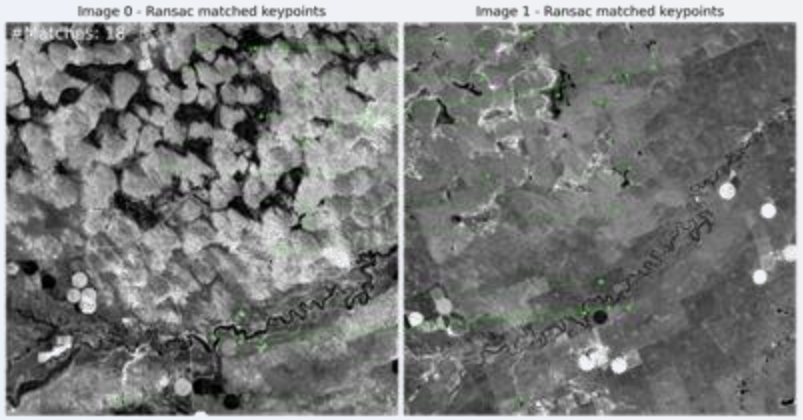
- Sparse keypoint detection and matching methods seem to underperform compared to dense and semi-dense methods
- Dense and semi-dense methods seem to depend on less specific spatial features, e.g. in homogeneous regions.
- Across all resolutions, RoMa and eloftr provide the best results on a sub-pixel level, being able to handle large spectral variations.

Results - Keypoint matches

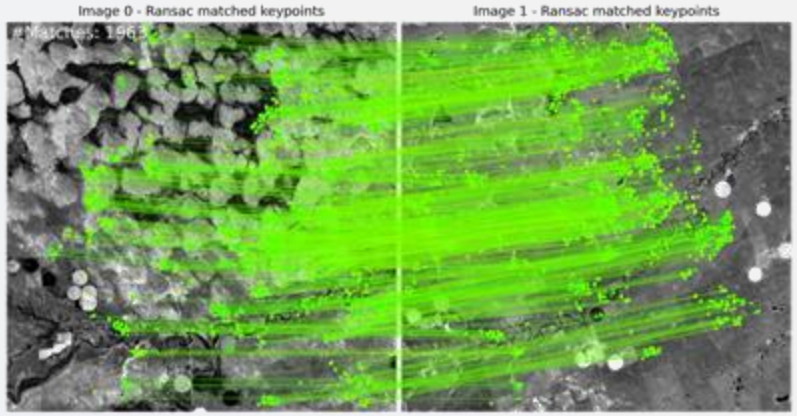
Image reference: S2B_MSIL2A_20230811T173909_N0509_R098_T13TGH_20230811T234353

- S2 at 10m resolution
- Left image: B02 (Blue)
- Right image: B08 (NIR)

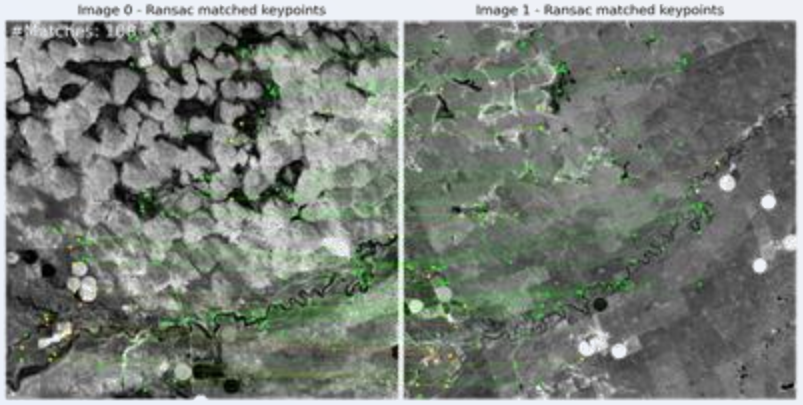
Very challenging image to align due to spectral differences



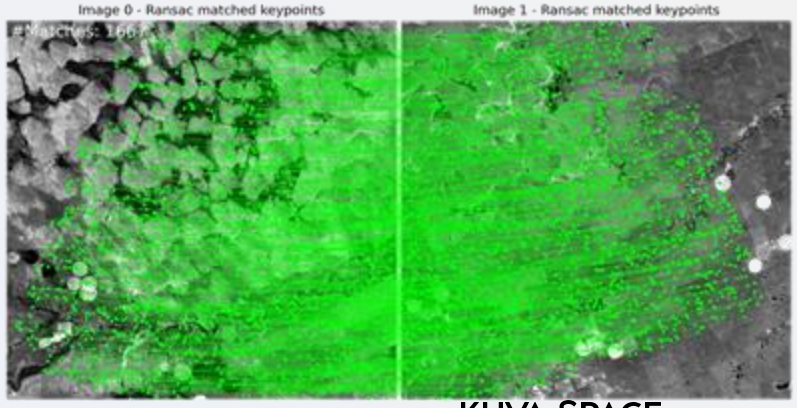
ALIKE+NN



eloftr



Superpoint+LightGlue



RoMa

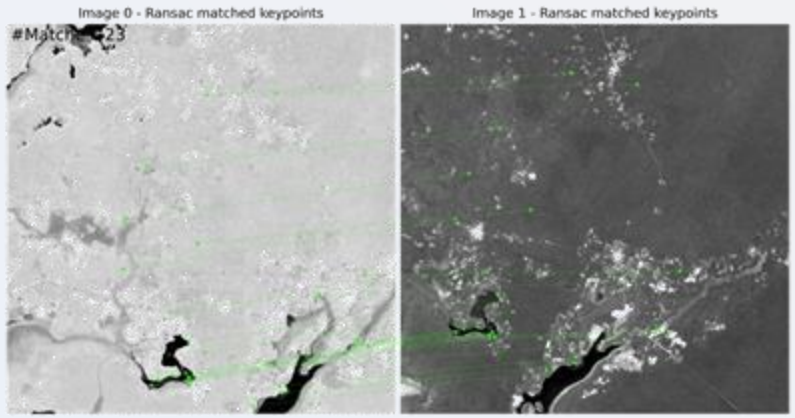
KUVA SPACE

Results - Keypoint matches

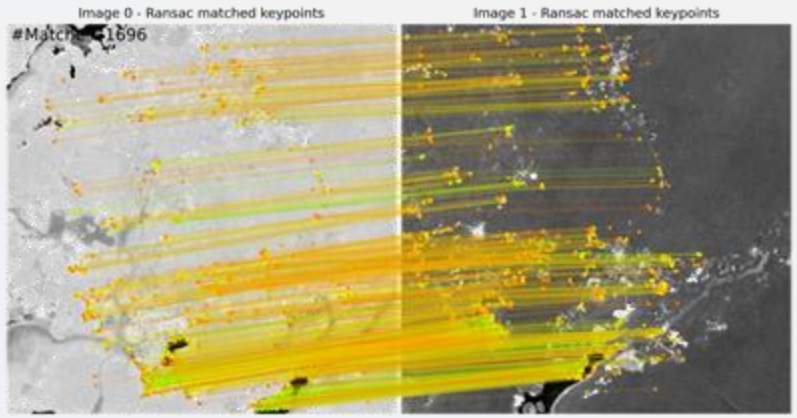
Image reference: S2A_MSIL2A_20230213T162401_N0509_R040_T16QCG_20230213T221655

- S2 at 10m resolution
- Left image: B08 (NIR)
- Right image: B03 (Green)

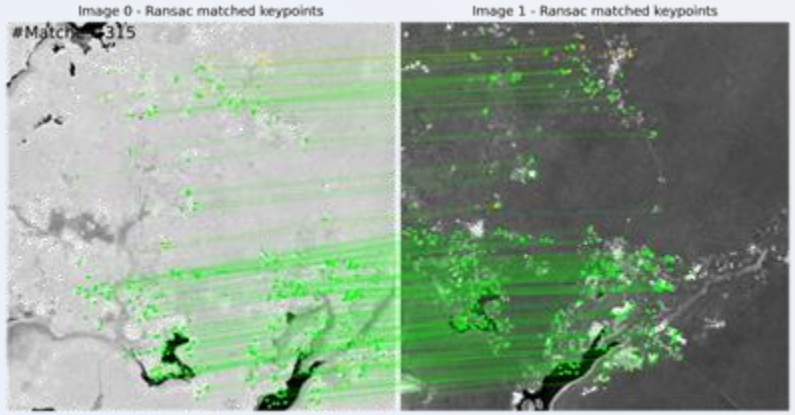
Very challenging image to align due to spectral differences



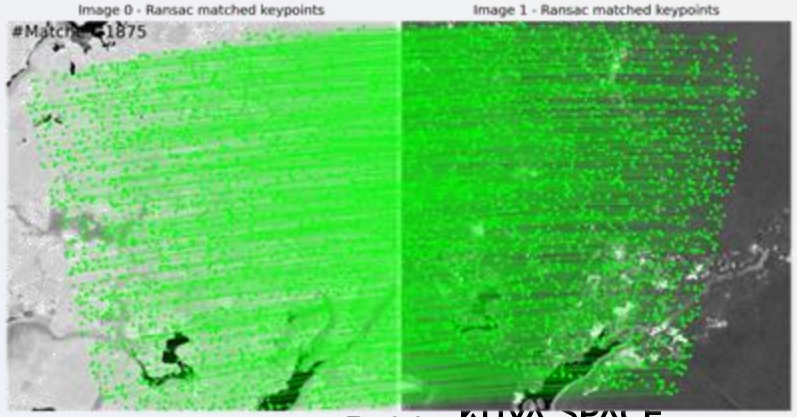
ALIKE+NN



eloftr



Superpoint+LightGlue



RoMa

Questions

1. **How much “instrument agnostic” are the current state of the art algorithms? What is the level of customization/tuning/auxiliary data required to be able to run them on a different instrument?**

   These deep learning-based keypoint detection and matching algorithms have not been trained specifically on remote sensing data, yet they still perform quite well. They have been trained on millions of natural RGB images acquired with a variety of cameras, optical systems, sensors. Often, these algorithms are built on robust, pre-trained foundation models like the Vision Transformers from DINOv2.

   There are challenges to consider, such as the varying wavelengths at which remote sensing images are captured. Additionally, achieving "sub-pixel" accuracy is highly dependent on image resolution. These models are typically trained on relatively small images, with dimensions of 480x480 or 768x768 pixels. While further fine-tuning of the models may enhance performance, this has not yet been tested.

1. **Can the AI based approaches really become instrument agnostic? If new training dataset are required and slightly different models are obtained, is it not one-model-one-sensor situation?**

   Models can always be fine-tuned or improved when new training datasets become available. By including more data into the training process, I do believe that for e.g. alignment algorithms, more-or-less instrument agnostic methods can be developed.

References

1. Zhao, Xiaoming et al. “ALIKE: Accurate and Lightweight Keypoint Detection and Descriptor Extraction” (https://arxiv.org/abs/2112.02906)
2. Tyszkiewicz, Michał J. et al. “DISK: Learning local features with policy gradient” (https://arxiv.org/abs/2006.13566)
3. Zhao, Xiaoming et al. “ALIKED: A Lighter Keypoint and Descriptor Extraction Network via Deformable Transformation” (https://arxiv.org/abs/2304.03608)
4. DeTone, Daniel et al. “SuperPoint: Self-Supervised Interest Point Detection and Description” (https://arxiv.org/abs/1712.07629)
5. Lindenberger, Philipp et al. “LightGlue: Local Feature Matching at Light Speed” (https://arxiv.org/abs/2306.13643)
6. Sarlin, Paul-Edouard et al. “SuperGlue: Learning Feature Matching with Graph Neural Networks” (https://arxiv.org/abs/1911.11763)
7. Edstedt, Johan et al. “DeDoDe: Detect, Don't Describe -- Describe, Don't Detect for Local Feature Matching” (https://arxiv.org/abs/2308.08479)
8. Sun, Jiaming et al. “LoFTR: Detector-Free Local Feature Matching with Transformers” (https://arxiv.org/pdf/2104.00680)
9. Wang, Yifan et al. “Efficient LoFTR: Semi-Dense Local Feature Matching with Sparse-Like Speed” (https://zju3dv.github.io/efficientloftr/files/EfficientLoFTR.pdf)
10. Edstedt, Johan et al. “RoMa: Robust Dense Feature Matching” (https://arxiv.org/abs/2305.15404)