Find My Astronaut Photo: Automated Localization and Georectification of Astronaut Photography

Alex Stoken[1]
alex.h.stoken@nasa.gov
[1]Jacobs Technology, NASA Johnson Space Center

Kenton Fisher[2]
kenton.r.fisher@nasa.gov
[2]NASA Johnson Space Center

Abstract

Astronaut photography from the International Space Station (ISS) forms one of the longest continuous remote sensing datasets of Earth and has facilitated a large body of research ranging from glacial surface area analysis to volcanic sediment delivery. Such studies are enabled by the geolocation and georectification of the imagery. Yet, localizing astronaut photography of Earth is a challenging and labor-intensive task, tempering the amount of research that can be performed. We present a method for automatically localizing these images named Find My Astronaut Photo, which makes this task feasible by casting the problem as a precision-oriented image similarity and matching exercise.

As the ISS orbits the globe, astronauts can view and photograph most locations on Earth, so there is no precomputable database of finite landmarks for image comparison. Therefore, we iteratively generate potentially similar images from geolocated satellite imagery on-demand and rely on an image matcher to discriminately detect overall similarity between these images and an astronaut photo.

We evaluate various image matching techniques to find methods which allow us to discretize and reduce our search space to a manageable size, and locate astronaut photographs with high precision and speed.

Find My Astronaut Photo has successfully geolocated over 30,000 photos to date, adding critical location information that increases the downstream utility of the Gateway to Astronaut Photography of Earth (GAPE) database. We also introduce AIMS, the Astronaut Imagery Matching Subset, a new real world evaluation dataset that joins the collection of challenging image matching benchmarks.

1. Introduction

Since the beginning of human spaceflight, astronauts have used handheld photography to share their unique perspective of Earth with the rest of the world. This remote sensing dataset contains over 4 million images that span more than 50 years, cover most of Earth's landmass, and provide a complimentary and unique dataset for researchers (Fig 1). The primary differentiator of astronaut observations is also one of its biggest benefits: the method in which the data is collected. In astronaut photography, a trained human is looking through a camera lens at a scene and interpreting the features and phenomena visible, which allows them to react to what they are seeing and adjust for better data collection in real time. While researchers are the primary requesters of imagery, astronaut photography is popular with the general public. The Gateway to Astronaut Photography of Earth (GAPE) receives 30 million visits per month, many of which utilize the image search function to look for photographs of specific locations. While the GAPE database contains photographs from all NASA human spaceflight missions, the majority have been taken from the International Space Station (ISS) due to the spacecraft's 22+ years of continuous crewed operations and the advent of digital camera systems. From low Earth orbit about 415 kilometers (250 miles) above the surface, wide swaths of the planet are visible at all times and even small changes in camera orientation can alter the location depicted in a photo by many kilometers. The microgravity environment and constantly

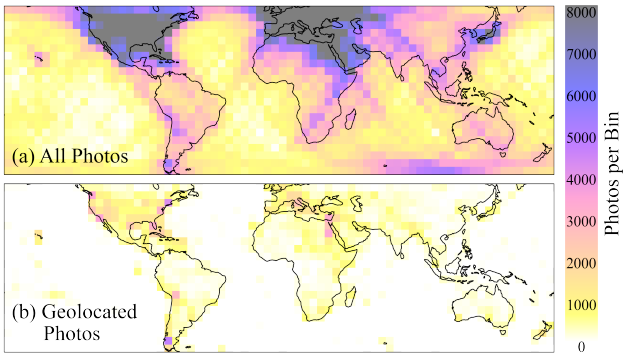

Figure 1. Distribution of (a) all astronaut photography by ISS location and (b) geolocated photographs by center point. Only a fraction of all imagery is geolocated. Bins are 10° squares.

changing orientation of the ISS with respect to the Earth means there is not even a canonical "up" with which to tentatively orient a photo.

Astronaut photography offers a unique combination of spatial resolution, temporal frequency, solar illumination, and look angle variation that traditional satellite imagery is unable to capture (Fig 2). However, the free-floating nature of the cameras, while providing significant benefit to data collection, inhibits the ability to easily determine the ground area shown within an image as the cameras do not record their orientation. While the position of the ISS is well known, the absence of camera pose information means that each image could have a ground center point (geographic coordinates of the center pixel) anywhere within a 2,000 kilometer radius of the ISS nadir point (location on the Earth directly under the ISS position). With image field of view varying from 10 to 2,500 kilometers depending on the camera lens used, finding the ground center point within this area requires considerable effort. Therefore, most of the images in the GAPE dataset lack the geographic information that would make them most useful to researchers and most accessible to the public.

A human operator can spend minutes to hours georeferencing a single photo. This time varies depending on the difficulty of the image and the uniqueness of its features. To this day, despite a team of dedicated citizen scientists localizing new photos each month, only 7% of all astronaut photography has a determined ground center point (Fig 1).

We present an image matching-based localization and georectification pipeline to automate ground center point determination in a timely manner and, further, produce fully georectified imagery. Our solution is designed to work in a uniquely "online" setting - while many image matching or retrieval benchmarks and applications focus on finding a best image given a fixed set of potential matches, our reference set changes per astronaut photo due to the storage and compute complexity of the search space - the entire surface of the Earth's landmass, at multiple scales. Reference imagery is derived from cloud-free composites of multispectral satellite sensor data and is visually distinct from DSLR-aquired astronaut photography, introducing further difficulty to the matching problem. Reference image generation is time intensive, accentuating the need for a high precision matcher that can enable "early stopping", the ability to end a search once a good match has been found, without having to check and rank all potential matches.

We offer an evaluation of methods for image matching and similarity detection in this setting, with special attention paid to the discriminability, scale robustness, and visible region overlap (covisibility) requirements of the methods. These properties are critical to the overall speed and reliability of the Find My Astronaut Photo pipeline and offer insight into important qualities for image matchers in a real world setting. We evaluate pretrained self-supervised embedding models, global feature based matchers, traditional and learned local feature matchers, and dense warp estimators across these criteria. Our key contributions are as follows: (1) the Find My Astronaut Photo pipeline, a method for geolocating and georectifying imagery from the high value astronaut photography of Earth database, (2) quantifying the performance of matchers for image similarity detection, (3) an investigation into the effects of scale, image size, and covisibility on a range of methods for image similarity detection in a challenging real world setting, and (4) the Astronaut Photography Image Matching Subset (AIMS), a new evaluation dataset for image matchers complete with ground truth ground center points and georectification data.

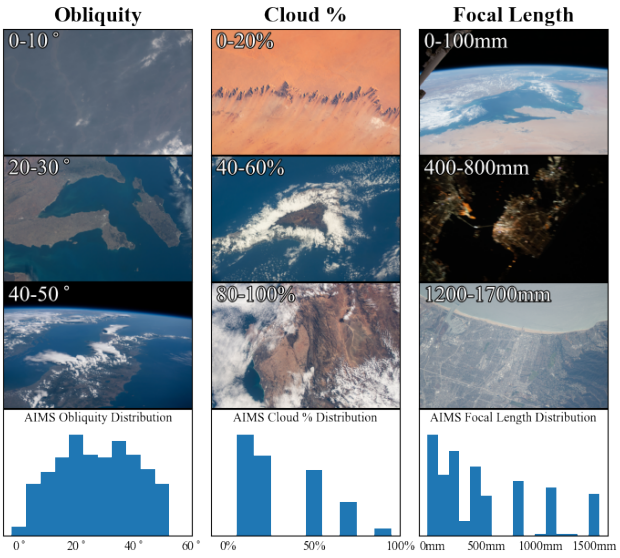

Figure 2. Astronaut Photos from GAPE. These images show the variation in the image set, particularly in camera focal length (reflected in field of view), cloud percentage, and obliquity (tilt). The final row shows the normalized distribution of these properties in the AIMS evaluation set (Section 3).

.

2. Related Works

Although localization and georectification are critical to the downstream use of astronaut photography, the problem has not been thoroughly studied. A recent work [30] focused on nighttime astronaut imagery, localizing to a rasterized street map based on the maximum number of feature matches over a fixed area and numerous rotations. Though similar in principle, our work focuses on the broader collection of daytime imagery, increases localization speed and breadth by using an adaptive search space and a discriminatory matcher, and is robust to blur, obliquity, and cloud

occlusion.

More generally, the problem of locating and georectifying astronaut photography lies at the intersection of visual place recognition (VPR), image matching, and image similarity. We review popular methods from each domain and discuss their application to astronaut photo localization.

2.1. Visual Place Recognition

On the surface, our task is an instance of visual place recognition, as we seek to identify the location of an image's content given only the image itself and a set of already localized images. Common challenges to VPR methods are also present in astronaut photography, including environmental appearance variation and, particularly in regions of the world lacking distinctive features like forests and deserts, perceptual aliasing [15]. Many VPR systems are retrieval-based [2, 25], first producing a global feature (embedding) for each database image and a query image, and using top-k filtering via a distance measure in embedding space before an optional second, local feature-based verification step re-ranks the top matches [16, 28]. Often the database images are processed and stored prior to use (offline), so they can be quickly compared with a new query that is processed online. Our problem differs from VPR in this key component - there is no perpetual database ("reference set") of images to compare against, so we cannot precompute a database of reference features. For each new astronaut photo, a reference set is generated from a satellite imagery repository in a online fashion, negating the benefits of precomputed features. For an embedding-based system to work well in our use case, we cannot rely on relative ranking for retrieval and instead require there to be a discriminating distance such that all images less than the distance are good matches and all greater are not.

2.2. Wide Baseline Image Matching

To that end, we turn toward pairwise similarity methods. In the pairwise domain, our task most closely aligns with wide baseline image matching [18]. This foundational area of computer vision aims to match images by generating correspondences between an image pair and then using these to align the images or estimate camera pose. This may be done as a part of a larger Structure from Motion (SfM) or Simultaneous Localization and Mapping (SLAM) pipeline. Variations in astronaut photography, especially when compared to the satellite imagery that forms our reference sets (Table 1), make this a multiple baseline problem.

Image matching is often accomplished via local feature matching. Classically, this was done in a detect-then-describe manner [22], employing handcrafted features. Later, learned features were introduced and gained popularity. Recently, detector-free, semi-dense matchers have shown considerable performance boosts in tasks that rely on accurate keypoint matching like the HPatches homography estimation benchmark [3], pose estimation [11, 21], and the Image Matching Challenge [18]. These matchers follow the style of the Local Feature TRansformer (LoFTR) [33], leveraging the benefits of self and cross attention between the images themselves or their features [5, 9, 17, 34, 35]. Alternatively, dense methods estimate a warp between two images and extract sparse keypoints from that warp [14]. Pairwise matchers are typically evaluated by their performance on downstream tasks, and not as methods for similarity detection itself. While we do use keypoints for homography estimation in the latter half of our pipeline, we primarily focus on using pairwise matchers and their keypoints to infer general similarity between images. We call this *similarity detection*, as it addresses whether two images depict the same scene, despite potentially confounding variations like seasonality, occlusion, obliquity, and more. Using the same principles that have shown success in these challenges and benchmarks, we define a keypoint-based similarity measure instead of using the points for alignment or pose estimation.

2.3. Other Image Similarity Methods

Image similarity determination is the essence of our task, but it is a less-developed subfield. Recently the Image Similarity Challenge [12] introduced a benchmark dataset. The challenge is split into two tracks: descriptor (global feature/retrieval based) and matching (pairwise comparison). Challenge results show the overall performance benefit of pairwise approaches and the power of incorporating local features for high precision matching.

Finally, similarity can also be viewed as generalization of overlap region estimation. Some recent works aim to estimate covisibility, or overlap regions between images, via specialized embeddings [10] or box regression [26]. Yet, these methods struggle with the more nuanced features in our imagery and are not well suited to confidently predict "no overlap" for truly non-overlapping imagery.

3. Dataset

The Gateway to Astronaut Photography of Earth dataset contains over 4 million astronaut photographs and serves as an open repository of remote sensing imagery for researchers and the public. The imagery is acquired by astronauts in space with handheld cameras, in contrast to the automated and highly controlled satellite imagery acquisition process, where sensor pose is well-known. Thus, astronaut imagery contains a high degree of variability compared to its satellite counterparts (Fig 2). In all cases, the conditions of astronaut photography increase the difficulty of localization whether by human or automated means (Table 1). Astronaut photography is used extensively for scientific research [1, 19, 20, 23, 27, 31] as well as for disaster response by NASA and other organizations [32]. In both cases, end

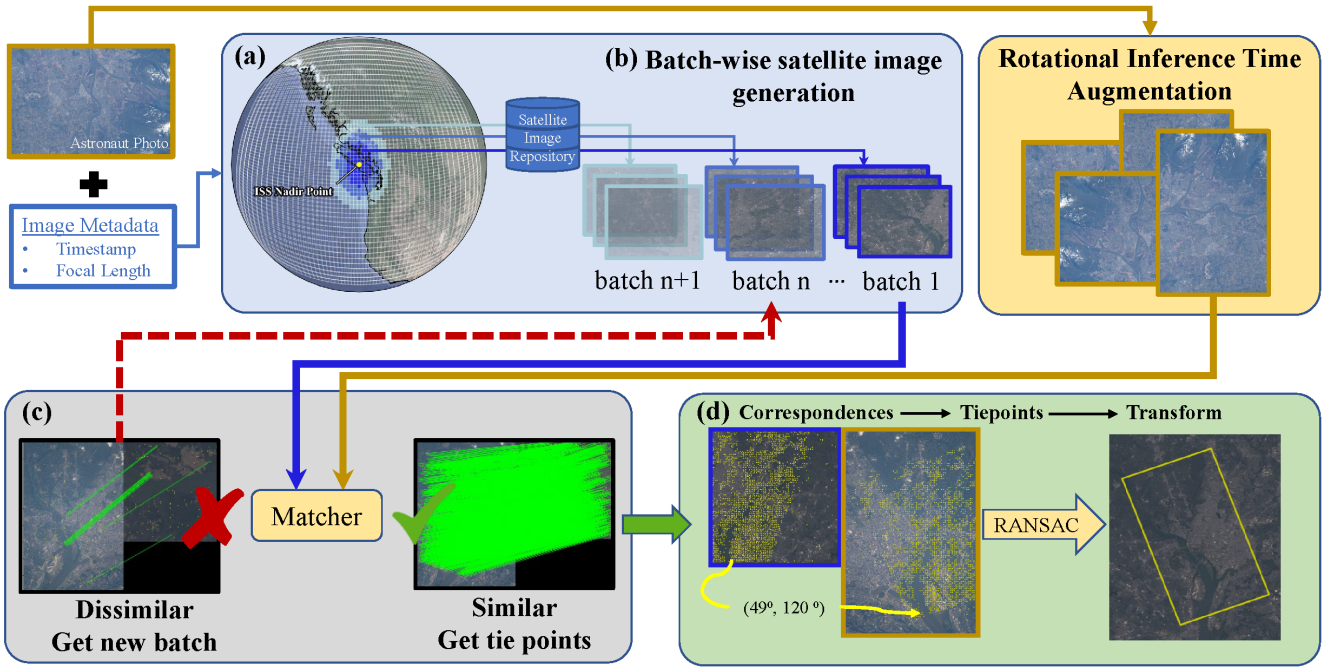

Figure 3. The Find My Astronaut Photo Pipeline. **(a)** From image metadata, we determine the ISS nadir point and discretize the Earth area visible from the ISS into a grid. We batch and prioritize grid patches by proximity (see Fig 5) **(b)** We produce satellite images covering the corresponding area of each grid patch from a satellite repository. **(c)** The astronaut photo and satellite image are put through a matcher. If the pair meets the match criteria, we move to (d), else return to (b) to generate more images. In **(d)**, high confidence geometrically verified keypoint matches are used as tie points to georectify the astronaut photo.

Condition	Astronaut Photographs	Satellite Imagery
Look Angle	**Variable (nadir to oblique)**	Fixed (near-nadir)
Illumination	**Day AND Night**	Day OR Night
Orientation	**Variable**	North Up
Field of View	**10-2,500 km**	Fixed per sensor
Seasonality	**Multiseasonal**	**Multiseasonal**

Table 1. Comparison of acquisition conditions between imagery types. More difficult condition bolded. Variable look angle and orientation make camera pose unknown.

users require geolocated and georectified imagery, a time consuming, manual process that must be done per image.

To fairly evaluate automated localization methods, we select 323 images with expert location and georectification information already established. We call this evaluation set AIMS - the Astronaut Image Matching Subset. AIMS has variability (Fig 2, bottom row) across cloud percentage per image (occlusion), focal length (field of view/scale), orientation (North angle), and tilt (obliquity, higher as distance from the ISS nadir point increases). Change across each of these axes poses a challenge to matching techniques so understanding performance in these settings is critical to characterizing methods. We publish the AIMS test set alongside an evaluation protocol for future benchmarking.

4. Approach

We formulate the localization, or ground center point finding, of astronaut photos as a cross-domain image similarity and matching problem. Assuming that images that encompass much of the same real world area will be visually similar, we can determine an astronaut photo's approximate location by finding a visually similar geolocated satellite image. Generating a satellite image and running an image matcher is a resource-heavy process so we use the coarse-scale geometry of the spacecraft-Earth system to grid off the entire region of the Earth's surface area visible from the ISS into a set of candidate patches (Fig 3a). Each patch covers a different portion of the Earth's surface, with extent determined by its location with respect to the camera and the camera intrinsics, such that a patch closer to the ISS covers a smaller field of view than a patch further from the ISS, just as an image taken of that area would.

4.1. Reference Image Generation

We iteratively generate a satellite reference image from publicly available Landsat data products [37] by extracting an image corresponding to the ground region defined by each patch (Fig 3b). We use cloud-free Landsat 8 median-composites from multiple years of data, which introduces a multi-temporal aspect to the matching process, as cloud

free pixels could come from an image from any season or year. While off nadir astronaut photos show obliquity characteristics, all satellite imagery used is nadir, so satellite images corresponding to what would be an oblique astronaut photo are still "flat", nadir-like images. Astronaut photos are taken using commercial off-the-shelf DSLR cameras and therefore do not collect radiance values like Landsat and other satellite imagers. Multispectral Landsat radiance values must be converted into RGB and brightness adjusted for use in matching. Brightness adjustment is done uniformly, though it is an imperfect adjustment due to other confounding factors such as atmospheric dust. This difference in sensors and resulting image characteristics between an astronaut photo query and satellite reference image led us to consider this a cross-domain matching problem.

4.2. Similarity-based localization

We batch patches by proximity to the ISS nadir point, such that the closest patches are attempted first. We parallelize image generation within a batch, and run the batch through an image matcher (Fig 3b). Given a highly discriminant, high precision image matcher that only activates on true positives, we can use early stopping upon a positive match and do not need to continue searching the entire space of candidate patches. We evaluate matchers for this property in Section 5.

Additional desired properties in a matcher are rotation invariance, as satellite images are typically north up, and astronaut photos have no canonical orientation. Many otherwise strong matchers do not meet this qualification, so for these we repeat the matching procedure for every 90° rotation and take the best score across orientations.

To determine match score when using a keypoint matcher, we take initial correspondences from the matcher (tentatives) and apply geometric verification. We estimate a homography between images using OpenCV MAGSAC [4, 6] with inlier threshold of 5 pixels and 100,000 iterations. We use the number of resulting correspondences (inliers) in our scoring metric. For a given matcher, we empirically determine a threshold number of inliers for a positive image match (Section 5.1). An image pair's match score is the maximum inliers over the tested rotations, normalized to the threshold. Scores over 1.0 indicate a positive match.

4.3. Georectification

After finding a positive corresponding satellite image, we georectify the astronaut photo (Fig 3d). This process assigns a location on the Earth (latitude/longitude coordinate pair) to each pixel in the image. To georectify, we determine a transformation that warps the image to the Earth - a pixel-to-world coordinate transformation. A set of N tie points are determined to represent N coordinates, and a homography is computed from these points. Due to the curvature of the Earth, georectifying off-nadir imagery benefits from an even dispersion of points across the source image. Based off the location of inliers in the matching candidate satellite image, which may have an incomplete overlap with the astronaut photo (Fig 5, inset), we generate another satellite image specifically for georectification. This image encapsulates an estimated extent of the astronaut photo, and produces better distributed tie points. We select well-dispersed, high-confidence correspondences which connect geolocated satellite image pixels with astronaut photo pixels, and use them to determine the pixel-to-geographic coordinate transform. The geolocated, georectified photo (Fig 4) is now searchable and ready for downstream use.

This approach is only feasible if using a matcher with high precision. If we cannot determine a repeatable, discriminative threshold for correctness, then the size of the search space drives the runtime of this approach to near infeasibility. Maximum per image runtime depends on the number of candidates, with higher focal lengths requiring more candidates to cover the visible Earth area. Accounting for this, the average maximum runtime is 7.65 minutes per photo. When applied to the collection of over 4 million photos, this procedure would take over 58 compute year equivalents without early stopping. In practice, early stopping reduces the average runtime by about 60%. Thus, we evaluate matching methods to find one with this property.

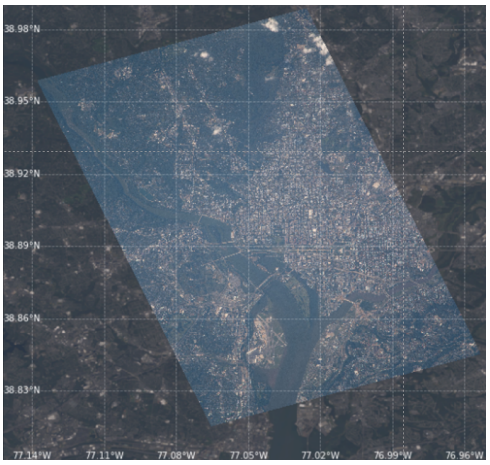

Figure 4. Fully georectified astronaut photo ISS065-E-241885, located and transformed via the Find My Astronaut Photo pipeline.

5. Experiments

We tailor our experiments toward quantifying matcher performance as an image similarity detector, as well as testing for other desirable criteria to speed up our approach. We evaluate matchers in the Find My Astronaut Photo search setting, and ensure we capture the broad range of variations present in astronaut photography by using AIMS (Sec-

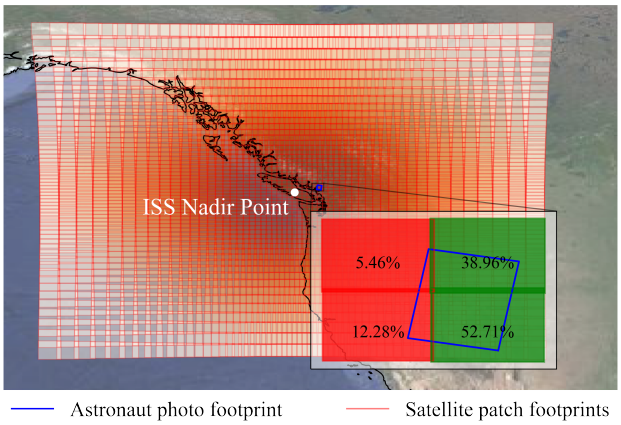

Figure 5. **Main** Visit priority for satellite image patches. Darker patches are explored first. Historically, proximity to the ISS nadir point is a good heuristic to follow for fast localization. Footprint areas range from 1,025 km^2 at center to 33,154 km^2 at edge. **Inset** Zoom of astronaut photo with all intersecting satellite patches and their respective overlap percent. Patches with overlap >25% are labeled correct (green), else labeled incorrect (red).

tion 3). Due to the method of search space discretization (Fig 5), there are multiple "good" candidate patches (high overlap with the astronaut photo) and each of these good patches will not have exactly the same extent as the astronaut photo. Thus, for our evaluation set, we also track nominal target overlap percent per candidate. This is the percent of the astronaut photo that is nominally (not accounting for occlusion via clouds, spacecraft hardware, etc.) covered by the satellite image patch. We label any candidate with overlap $>$ 25% as a "correct" patch. This value is empirically derived by what minimum overlap aligns with human judgment of the images representing the same "place".

We examine models from the image matching, retrieval, and similarity domains. We divide the models into four categories (1) detector-based local feature matchers (SIFT [22], SuperPoint+SuperGlue [29], D2-Net [13]) (2) detector-free local feature matchers (LoFTR [33], SE2-LoFTR [5], Aspanformer [9], Matchformer [35], Patch2Pix [39]) (3) pretrained global embeddings (NetVLAD [2], GeM-Net [24, 25], SwAV [7], DINO [8], Barlow Twins [38]) and (4) dense similarity/overlap predictors (DKM [14]). For embedding models, we evaluate them under "early stopping" conditions, determining whether there is a distance that divides the embedding space into matching and non-matching images, instead of using traditional retrieval metrics like correctness of top-k candidates. For local feature matching models, we threshold on the number of geometrically verified correspondences (inliers). We investigate whether a threshold exists to separate correct and incorrect images. Finally, for image similarity models, we evaluate over predicted overlap percent, with 25% as the minimum for a positive match. We refer to this "early stopping" criteria as discriminability.

5.1. Discriminability

In the astronaut photography localization setting with early stopping, a false positive from a non-discriminant matcher can completely interrupt image localization by concluding a search before the correct patch is encountered. To test discriminability, we compute match scores between the astronaut photo and satellite images generated from the gridding procedure in Section 4. On average, three correct and 46 incorrect satellite images are produced per astronaut photo in the AIMS set. We plot the distribution of match scores and calculate average precision over various thresholds. For each global feature matcher, we normalize the distances so that they can be readily compared. We evaluate each matcher in its best configuration and take the threshold that produces the most desirable point on a Precision-Recall curve. We additionally measure and seek to minimize the number of false positives per query.

5.2. Scale

Scale robustness is critical to reducing the number of candidates in the search grid. We nominally generate candidate patches to align with the approximate native spatial resolution (m/pixel) of the astronaut photo so that geographic features are of a similar pixel size in both images. Covering the entire visible area of the Earth at native scale can require evaluating up to 40,000 candidates per astronaut photo, particularly for high focal length images. This number can be significantly reduced by generating candidate patches that have a larger spatial extent than the astronaut photo, but this increases the burden on the matcher to perform well across scale variations as alike features would no longer have similar pixel size. Alternatively, the astronaut photo could be downsampled to match the reduced spatial resolution (see Section 5.3 for effects of this adjustment). This motivates our investigation into a scale robust matcher.

We examine how similarity detection performance on AIMS changes across patches that are 1.0, 1.5, and 2.0 times the spatial extent of the astronaut photo. Additionally, we take the best case scenario for matching, where the satellite image's extent is the minimum rectilinear bounding box of the astronaut photo, and compute the number of matches between these scaled images and the astronaut photo.

5.3. Image Size

One of the main concerns with Find My Astronaut Photo is the per-photo runtime. While this is in theory driven by the number of candidates in the search grid, in practice it can be reduced for a given hardware configuration

by working with smaller images. Smaller images allow for larger batch sizes and faster matching, yet contain less information to match with. Some features discernible in a large, high resolution image will disappear once that image is downsampled. To find the minimum size requirement for productive matching, we revisit the scenario from scale robustness testing (Section 5.2) and match the best-case patch at different image sizes. We resize both the satellite image and astronaut photo equally (maintaining relative scale) until matching is impossible. Image sizes are chosen to align with required dimensions for CNN feature extraction.

6. Results

Experiments with the AIMS dataset illustrate how different image similarity techniques perform in the context of a real world application. We review each in turn.

	Average Precision					
AIMS Split	SE2-L	Aspan	MF	SP-SG	D2-Net	DKM
1.0x scale	**0.61**	0.48	0.56	0.49	0.39	0.38
1.5x scale	**0.52**	0.33	0.44	0.50	0.43	0.32
2.0x scale	0.25	0.21	0.21	**0.41**	0.37	0.25
Low cloud	**0.62**	0.50	0.56	0.52	0.40	0.42
High cloud	0.49	0.36	**0.54**	0.31	0.28	0.14
North up	**0.62**	0.47	0.55	0.56	0.47	0.47
All other	**0.51**	0.09	0.06	0.14	0.06	0.04

Table 2. Average precision on important splits of AIMS.

6.1. Discriminability Results

We first analyze discriminability by measure of average precision over thresholds. Fig 6 depicts this in the rotation augmented, 1.0x scale multiplier setting. Generally, there is lower intra-class variance than inter-class variance in precision-recall space. This is particularly true among LoFTR-style models, suggesting that iterations on this style are incremental. Detector-free local features are the most discriminant class, followed by detector-based local features. Global embedding models, when using distance as a threshold instead of rank, admit too many false positives. They do not perform in the same regime as local feature matchers and we hereafter focus only on the more performant model classes.

The number of inliers is a popular heuristic for determining whether an image contains the same scene. Aspanformer [9] uses 25 inliers as a threshold for keeping images for InLoc evaluation. Figure 6 illustrates the trade-off space for different inlier thresholds. Aspanformer reaches high precision with fewer inliers than other LoFTR-style models, though SE2-LoFTR has the highest average precision (Table 2) due to it's superior recall at lower inlier thresholds. Of the detector-based methods, SuperPoint+SuperGlue performs best but still lags behind detector-free methods due to lower recall at low inlier thresholds.

LoFTR-style models also achieve the highest recall for a given number of false positives (Fig 6). For detector-free models, recall can exceed 60% before a single false positive per astronaut photo is encountered. This is a promising indicator for the Find My Astronaut Photo application. All matchers are extremely sensitive to rotation with the exception of SE2-LoFTR, which is designed [36] specifically for rotation equivariance. On average, models perform 7 times better when matching an image with the closest 90° rotation to its true orientation.

6.2. Scale Results

We find the number of inlier matches, even in optimal conditions, drops with inter-image scale variation (Fig 7, bar plot). For detector-free matchers, the drop is significant. Scale variation has a similar negative impact on average precision in AIMS evaluation (Fig 7, lineplot). For AIMS, where correct matches rarely have 100% overlap, fewer matches on partially overlapping patches means a loss in discriminating power for a particular threshold, as true positives can no longer exceed the threshold. Table 2 illustrates the pitfalls of LoFTR style models over scale change. It's noteworthy that SE2-LoFTR maintains performance when jumping between 1.0 and 1.25 scale factors. Non-LoFTR family models generalize better over scale.

6.3. Image Size Results

Detector-free methods are notably sensitive to input image size (Fig. 8). This can be attributed to architectural choices and constraints as well as to the reduction in distinctive image features as additional downsampling occurs. Many image sizes were impermissible to certain models. Once a permissible size was found (dimensions multiple of 16 or 32), there is a linearly decreasing trend in the number of inliers with respect to decreasing image size. However, there is a peak in inlier quantity at 578 pixels, which is similar to the training size of 640 for SE2-LoFTR. We did not examine the performance of different image sizes on the downstream AIMS task, but it is possible that the discriminating power of detector-free matchers is reduced at non-optimal, but potentially still permissible input sizes. Certainly, there is a minimum image size for good similarity detection around 300 pixels square. Images below this size contain too few inlier matches for discriminability.

6.4. AIMS Results

We gain further insight from performance across splits of AIMS (Table 2). SE2-LoFTR shines throughout, but other LoFTR variants display certain desirable properties. Matchformer performs best under occlusion in the high cloud split. Detector-based models, though weaker in best-

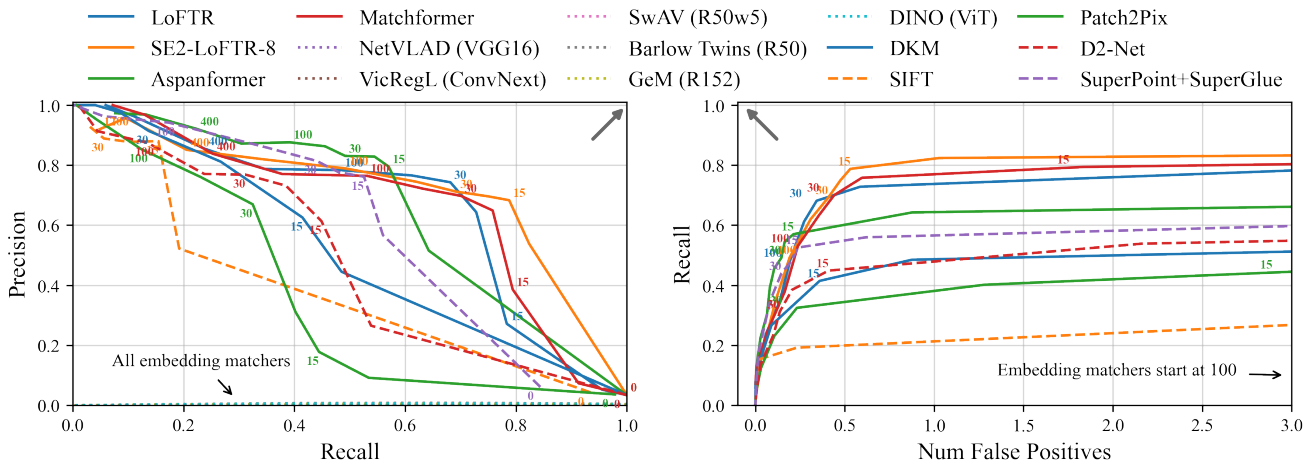

Figure 6. Precision-Recall and Recall vs False Positives Per Query in rotation augmented, 1.0x scale multiplier setting. Gray arrows indicate direction of better performance. Linestyle indicates method class: detector-free (solid), detector-based (dashed), embedding (dotted). Threshold values are annotated.

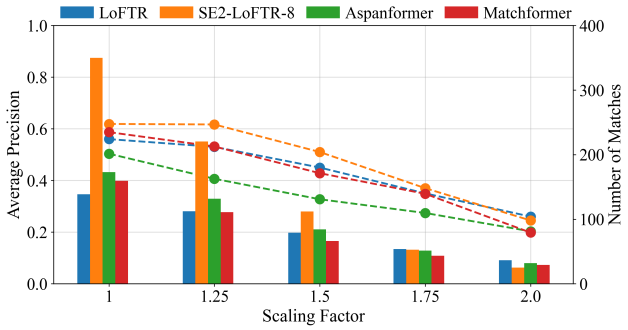

Figure 7. Scale robustness results. Average precision line plot (left axis), number of matches in ideal case bar plot (right axis). Performance drops significantly beyond a 1.25 scaling factor.

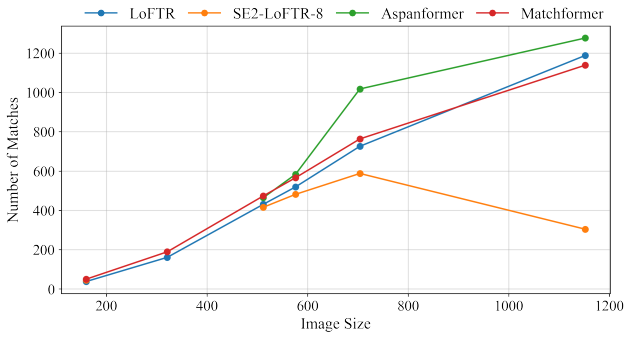

Figure 8. Detector-free matcher image size robustness. Models seem to have preferential input sizes, and generally number of matches decreases with image size.

case scenarios, show significant upside in scale robustness. Furthermore, SuperPoint+SuperGlue does better than some detector-free matchers in optimal, North up conditions.

Based on these experiments, we choose SE2-LoFTR as our matcher for Find My Astronaut Photo. Its rotation equivariance eliminates the need for inference time augmentation and of the detector-free methods, it is most robust to scale variation. We set the inlier threshold for a positive match at 30, the highest value that keeps the average number of false positives under one.

7. Conclusion

We introduce the astronaut photography localization problem, an image matching based solution called Find My Astronaut Photo, and a corresponding evaluation dataset, AIMS. We additionally conduct an empirical investigation of image matchers for similarity detection, focusing on discriminability, scale robustness, and image size robustness.

We find that recent advances in detector-free matchers enable sufficient image similarity performance to efficiently power the Find My Astronaut Photo pipeline with early stopping. This allows for automated geolocation and georectification of large portions of the Gateway to Astronaut Photography of Earth, making the invaluable collection more accessible to researchers and the general public. Using the settings found in Section 6, we have georectified over 30,000 images in 10 months of operation.

Finally, we release the AIMS evaluation set in the hope that it will inspire future work in image matching and similarity detection. It serves both as a high-value application and a challenging benchmark for image matching due to the natural variations that arise with astronauts taking photographs of the Earth from the International Space Station.

References

[1] Serge Andréfouët, Julie A Robinson, Chuanmin Hu, Gene C Feldman, Bernard Salvat, Claude Payri, and Frank E Muller-Karger. Influence of the spatial resolution of seawifs, landsat-7, spot, and international space station data on estimates of landscape parameters of pacific ocean atolls. *Canadian Journal of Remote Sensing*, 29(2):210–218, 2003. 3

[2] Relja Arandjelovic, Petr Gronát, Akihiko Torii, Tomás Pajdla, and Josef Sivic. Netvlad: CNN architecture for weakly supervised place recognition. *CoRR*, abs/1511.07247, 2015. 3, 6

[3] Vassileios Balntas, Karel Lenc, Andrea Vedaldi, and Krystian Mikolajczyk. Hpatches: A benchmark and evaluation of handcrafted and learned local descriptors. In *CVPR*, 2017. 3

[4] Daniel Barath, Jiri Matas, and Jana Noskova. MAGSAC: marginalizing sample consensus. In *Conference on Computer Vision and Pattern Recognition*, 2019. 5

[5] Georg Bökman and Fredrik Kahl. A case for using rotation invariant features in state of the art feature matchers. In *CVPRW*, 2022. 3, 6

[6] G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000. 5

[7] Mathilde Caron, Ishan Misra, Julien Mairal, Priya Goyal, Piotr Bojanowski, and Armand Joulin. Unsupervised learning of visual features by contrasting cluster assignments. 2020.

[8] Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *Proceedings of the International Conference on Computer Vision (ICCV)*, 2021. 6

[9] Hongkai Chen, Zixin Luo, Lei Zhou, Yurun Tian, Mingmin Zhen, Tian Fang, David Mckinnon, Yanghai Tsin, and Long Quan. Aspanformer: Detector-free image matching with adaptive span transformer. In *ECCV*, 2022. 3, 6, 7

[10] Ying Chen, Dihe Huang, Shang Xu, Jianlin Liu, and Yong Liu. Guide local feature matching by overlap estimation. In *AAAI*, 2022. 3

[11] Angela Dai, Angel X. Chang, Manolis Savva, Maciej Halber, Thomas Funkhouser, and Matthias Nießner. Scannet: Richly-annotated 3d reconstructions of indoor scenes. In *Proc. Computer Vision and Pattern Recognition (CVPR), IEEE*, 2017. 3

[12] Matthijs Douze, Giorgos Tolias, Ed Pizzi, Zoë Papakipos, Lowik Chanussot, Filip Radenovic, Tomás Jenícek, Maxim Maximov, Laura Leal-Taixé, Ismail Elezi, Ondrej Chum, and Cristian Canton-Ferrer. The 2021 image similarity dataset and challenge. *CoRR*, abs/2106.09672, 2021. 3

[13] Mihai Dusmanu, Ignacio Rocco, Tomas Pajdla, Marc Pollefeys, Josef Sivic, Akihiko Torii, and Torsten Sattler. D2-Net: A Trainable CNN for Joint Detection and Description of Local Features. In *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019. 6

[14] Johan Edstedt, Ioannis Athanasiadis, Mårten Wadenbäck, and Michael Felsberg. DKM: Dense kernelized feature matching for geometry estimation. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2023. 3, 6

[15] Sourav Garg, Tobias Fischer, and Michael Milford. Where is your place, visual place recognition? *CoRR*, abs/2103.06443, 2021. 3

[16] Stephen Hausler, Sourav Garg, Ming Xu, Michael Milford, and Tobias Fischer. Patch-netvlad: Multi-scale fusion of locally-global descriptors for place recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14141–14152, 2021. 3

[17] Dihe Huang, Ying Chen, Shang Xu, Yong Liu, Wenlong Wu, Yikang Ding, Chengjie Wang, and Fan Tang. Adaptive assignment for geometry aware local feature matching, 2022. 3

[18] Yuhe Jin, Dmytro Mishkin, Anastasiia Mishchuk, Jiri Matas, Pascal Fua, Kwang Moo Yi, and Eduard Trulls. Image matching across wide baselines: From paper to practice. *CoRR*, abs/2003.01587, 2020. 3

[19] Bradley Johnson. Detecting impervious cover with artificial lighting in astronaut photography from the international space station. 07 2020. 3

[20] Tatiana Khromova, Gennady Nosenko, Stanislav Kutuzov, Anton Muraviev, and Ludmila Chernova. Glacier area changes in northern eurasia. *Environmental Research Letters*, 9(1):015003, jan 2014. 3

[21] Zhengqi Li and Noah Snavely. Megadepth: Learning single-view depth prediction from internet photos. In *Computer Vision and Pattern Recognition (CVPR)*, 2018. 3

[22] David G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2):91–110, Nov. 2004. 3, 6

[23] Jon J. Major, Daniel Bertin, Thomas C. Pierson, Álvaro Amigo, Andrés Iroumé, Héctor Ulloa, and Jonathan Castro. Extraordinary sediment delivery and rapid geomorphic response following the 2008–2009 eruption of chaitén volcano, chile. *Water Resources Research*, 52(7):5075–5094, 2016. 3

[24] F. Radenović, G. Tolias, and O. Chum. CNN image retrieval learns from BoW: Unsupervised fine-tuning with hard examples. In *ECCV*, 2016. 6

[25] F. Radenović, G. Tolias, and O. Chum. Fine-tuning CNN image retrieval with no human annotation. *TPAMI*, 2018. 3, 6

[26] Anita Rau, Guillermo Garcia-Hernando, Danail Stoyanov, Gabriel J. Brostow, and Daniyar Turmukhambetov. Predicting visual overlap of images through interpretable non-metric box embeddings. In *European Conference on Computer Vision (ECCV)*, 2020. 3

[27] Alejandro Sanchez de Miguel, Jaime Zamorano, Martin Aubé, Jonathan Bennie, Jesús Gallego, Francisco Ocaña, Donald Pettit, William Stefanov, and Kevin Gaston. Colour remote sensing of the impact of artificial light at night (ii): Calibration of dslr-based images from the international space station. 08 2021. 3

[28] Paul-Edouard Sarlin, Cesar Cadena, Roland Siegwart, and Marcin Dymczyk. From coarse to fine: Robust hierarchical localization at large scale. In *CVPR*, 2019. 3

[29] Paul-Edouard Sarlin, Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. SuperGlue: Learning feature matching with graph neural networks. In *CVPR*, 2020. 6

[30] Peter Schwind and Tobias Storch. Georeferencing urban nighttime lights imagery using street network maps. *Remote Sensing*, 14(11), 2022. 2

[31] Maria Shahgedanova, Gennady Nosenko, Tatyana Khromova, and Anton Muraveyev. Glacier shrinkage and climatic change in the russian altai from the mid-20th century: An assessment using remote sensing and precis regional climate model. *Journal of Geophysical Research: Atmospheres*, 115(D16), 2010. 3

[32] W. L. Stefanov and C. A. Evans. Data collection for disaster response from the international space station. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-7/W3:851–855, 2015. 3

[33] Jiaming Sun, Zehong Shen, Yuang Wang, Hujun Bao, and Xiaowei Zhou. LoFTR: Detector-free local feature matching with transformers. *CVPR*, 2021. 3, 6

[34] Shitao Tang, Jiahui Zhang, Siyu Zhu, and Ping Tan. Quadtree attention for vision transformers. *ICLR*, 2022. 3

[35] Qing Wang, Jiaming Zhang, Kailun Yang, Kunyu Peng, and Rainer Stiefelhagen. Matchformer: Interleaving attention in transformers for feature matching. In *Asian Conference on Computer Vision*, 2022. 3, 6

[36] Maurice Weiler and Gabriele Cesa. General E(2)-Equivariant Steerable CNNs. In *Conference on Neural Information Processing Systems (NeurIPS)*, 2019. 7

[37] Michael A. Wulder, Thomas R. Loveland, David P. Roy, Christopher J. Crawford, Jeffrey G. Masek, Curtis E. Woodcock, Richard G. Allen, Martha C. Anderson, Alan S. Belward, Warren B. Cohen, John Dwyer, Angela Erb, Feng Gao, Patrick Griffiths, Dennis Helder, Txomin Hermosilla, James D. Hipple, Patrick Hostert, M. Joseph Hughes, Justin Huntington, David M. Johnson, Robert Kennedy, Ayse Kilic, Zhan Li, Leo Lymburner, Joel McCorkel, Nima Pahlevan, Theodore A. Scambos, Crystal Schaaf, John R. Schott, Yongwei Sheng, James Storey, Eric Vermote, James Vogelmann, Joanne C. White, Randolph H. Wynne, and Zhe Zhu. Current status of landsat program, science, and applications. *Remote Sensing of Environment*, 225:127–147, 2019. 4

[38] Jure Zbontar, Li Jing, Ishan Misra, Yann LeCun, and Stéphane Deny. Barlow twins: Self-supervised learning via redundancy reduction. *arXiv preprint arXiv:2103.03230*, 2021. 6

[39] Qunjie Zhou, Torsten Sattler, and Laura Leal-Taixe. Patch2pix: Epipolar-guided pixel-level correspondences. In *CVPR*, 2021. 6