EarthMatch: Iterative Coregistration for Fine-grained Localization of Astronaut Photography

Supplementary Material

7. Astronaut Photography Localization Metrics

Since the introduction of the Astronaut Image Matching Subset (AIMS) dataset in FMAP [53], various metrics have been proposed to best capture the astronaut photography localization challenge. FMAP itself uses a fixed set of reference images provided as part of AIMS, and for each astronaut photograph, there may be one or more matching reference images. FMAP measures average precision over these reference images. Such a metric emphasizes high recall matching, such that any query/reference pair with sufficiently overlapping extents should match.

FMAP uses a fixed, discrete set of inlier thresholds and at most 100 negatives in it's average precision calculation. Steerers [10] expands this calculation to a continuous inlier range, and uses all of the negatives in AIMS. This gives a more complete, but otherwise comparable, metric.

EarthLoc [8] introduces the concept of retrieval to astronaut photography localization, and brings a retrieval oriented metric to the task: Recall@N. This is the percent of astronaut photos in which one of the top N retrieved reference images is correct, with correctness defined as having non-zero overlap between the astronaut photo and reference image. Due to recasting APL as a retrieval, rather than a pairwise matching task, this metric is not directly comparable to the average precision of FMAP and Steerers.

In this work, we again shift the metric, this time to more closely align it with the downstream localization task. Since the end goal of APL is to find the correct location of an astronaut photograph on Earth, and our clearest indication of the correct place on Earth is the manually annotated center point, we use the number (or, equivanetly, percent) of photos where our predicted footprint contains the centerpoint as a metric, and denote this as "images correctly localized". This is a better measure of real world performance and allows us to compare methods as they would operate in deployment.

8. Qualitative results

In Fig. 8 we show examples of matchings with different models.

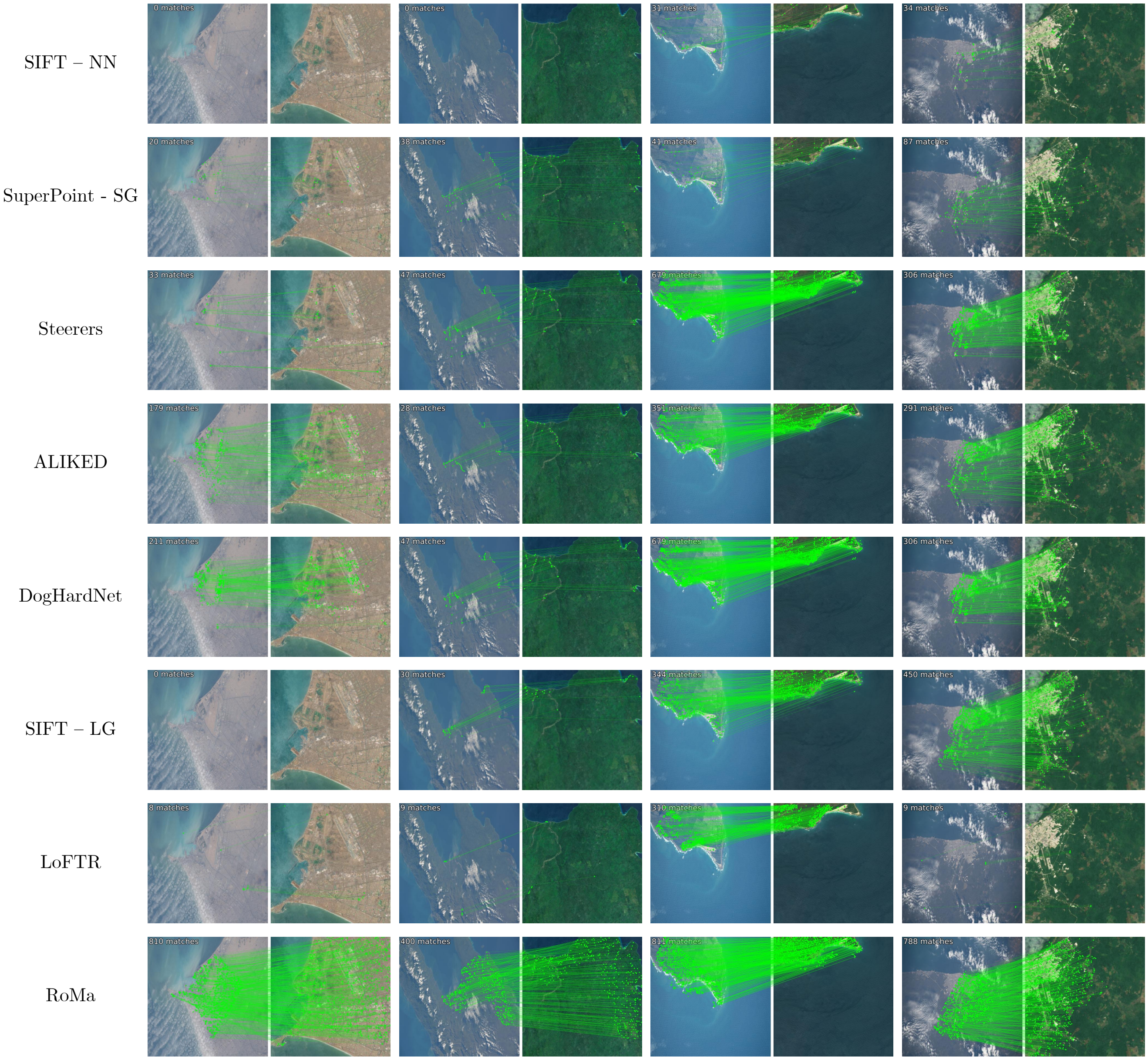

Figure 8. **Qualitative results** for a select number of matching methods proposed in our benchmark.