Onboard to Satellite Image Matching for Relocalization of the UAVs*

Tomasz Pogorzelski[1][0000-0002-8299-0727] and Karol Majek[2][0000-0002-1351-8496]

[1] Łukasiewicz Research Network – Institute of Aviation, 02-256 Warsaw, Poland
[2] Cufix, 05-825 Grodzisk Mazowiecki, Poland
karolmajek@cufix.pl

Abstract. In this paper, we study image-based localization as a component of a vision navigation system. We perform an experimental verification of image matching techniques on a real UAV flight in a rural area. Images from onboard camera are matched to the satellite images based on last known position. We find deep front-end SuperPoint with deep middle-end matcher SuperGlue better suited than the pretrained LoFTR for the image-based localization in feature-poor areas. The proposed component implementation allows UAVs for emergency localization based on single frame with error below 100m in a rural area.

Keywords: Image-based localization · UAV localization · Image matching.

1 Introduction and Related Work

We aim to ensure redundancy of satellite navigation in applications including acquisition of photographic materials with the use of UAVs [5]. The goal is to achieve a visual navigation component for the autopilot by matching the camera images to the orthophotomap for a large operation area. Image based localization for UAVs can be considered as full 6-dof [6] for indoor applications as well as for gps-denied outdoor environments [1]. Such approaches use conventional algorithms [7], either deep learning [7] or mixed [8]. For this purpose, the Mutual Information algorithm [12], image segmentation [3] and graph search are also often used [2]. For the purpose of image-based localization, hand-crafted image matching methods such as ORB [9] can be used. In contrast, newer approaches are based on neural networks and are learned from data during the training process. Use of neural networks is relevant to UAVs since availability of high performance processing on embedded devices such as Nvidia Jetson Xavier series or Intel Movidius. Feature detector and descriptor SuperPoint [4] was introduced as a deep front-end. SuperGlue introduced in [10] is a deep middle-end matcher which utilizes graph neural networks to solve assingment problem. Introduced recently, LoFTR is a detector-free method inspired by the SuperGlue matching [11] not limited to detected keypoints.

* Supported by an implementation doctorate registered at the Doctoral School of the Warsaw University of Technology and the Center for Unmanned Technologies, Łukasiewicz Research Network - Institute of Aviation.

2 Redundant Image Based Localization System for UAVs

To determine the absolute geographic location we use: the recently known position, the current image from the camera and the geo-referenced orthophotomap. The last known position is used to limit the the analyzed area - Region Of Interrest (ROI) in order to ensure the proper performance of the algorithm. Then the current camera image frame is compared with the cut ROI. The selected deep neural networks determine matches between images. Based on the best fit, a homography matrix, and a perspective quadrilateral are computed. The confidence of measurement depends on the number of detected points. The return is calculated from the middle of the top edge of the quadrilateral.

We compare learning-based feature detector and descriptor SuperPoint [4] with feature matcher SuperGlue [10], and detector-free LoFTR [11]. During system initialization, a rough start position is taken to limit the ROI (Region Of Interrest) search area from the whole map. Using the above algorithms, common points characteristic for the ROI and the current frame from the on-board camera are found. On the basis of these points, a homographic matrix is calculated, by means of which the perspective rectangle (blue) is calculated. Its geometric center is the location. Currently, the system analyzes each frame separately.

(a) SuperPoint + SuperGlue (b) LoFTR

Fig. 1: Visual comparison of image-based localization methods. Large rectangle is ROI used for localization. GPS trajectory in dark blue. Points in cyan - localization result. Cyan quadrilateral - represents field-of-view of the localized camera.

3 Experimental Results

The images generated by the system, based on a real UAV flight in a rural area, are shown below. For our input, conventional descriptor methods were only worked in simulated flight in Google Earth Studio, compared to a map with the same source. In order to get closer to the real conditions, a video was recorded made by a drone from a height of 120m above ground level. Flight logs were used for the Ground Truth reference. In the case of a real flight recording, the above descriptors found too few common points. Only the use of deep learning techniques allowed for correct matching in this case. This may be related to unfavorable lighting, or to significant differences between the current image and the historical one from Google maps. The season was similar, so the colors of the vegetation did not matter. Figure 1 shows the visualization of the perspective (camera point of view) in the form of a blue quadrilateral placed on the map. Its geometric center is a preliminary location estimation. It can be noticed that both algorithms work worse in the case of obtaining a homogeneous image.

It has been observed that the even distribution of characteristic points in the image has a significant influence on the accuracy of the location. The method based on the average distance between the nearest, neighboring points was used to measure the uniformity coefficient. This tells whether the data points are concentrated or scattered. The graph below shows the error and the average distance between points in consecutive video frames. On Figure 2 a correlation between the calculated location error and the average distance of the point from the nearest neighbor can be observed.

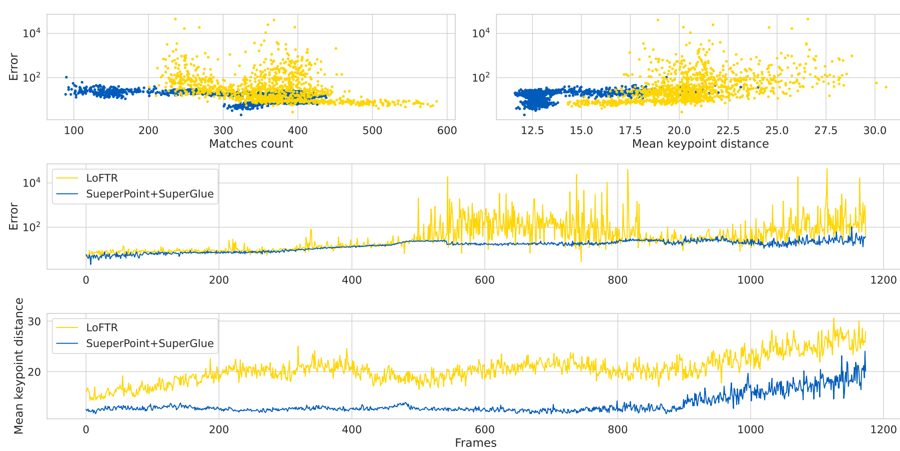

Fig. 2: 2D Error in meters related to the number of good matches and mean keypoint distance for SuperPoint with SuperGlue and LoFTR.

4 Conclusion and Future Work

We compared deep learning-based image matching methods for the purpose of UAV localization. Results show smaller error for keypoint-based method SuperGlue in favor of keypoint-free LoFTR pretrained in this setup. All correct matches were used to calculate the homography matrix. In the future, it is planned to select only those with the greatest probability of correctness. Fine-tuning LoFTR on aerial images should improve the results. The proposed image-based localization module works on a single frame against a large ROI. Therefore, localization error on raw outputs is too large for UAV navigation purposes and should be fused with IMU and visual odometry to achieve a good accuracy in gps-denied environments. The described work concerns the analysis of a short video recording on a small area, however, it allows for further orientation of the work in an optimal way.

References

1. Ashraf, S., Aggarwal, P., Damacharla, P., Wang, H., Javaid, A.Y., Devabhaktuni, V.: A low-cost solution for unmanned aerial vehicle navigation in a global positioning system–denied environment. International Journal of Distributed Sensor Networks **14**(6), 1550147718781750 (2018)
2. Chiu, H.P., Das, A., Miller, P., Samarasekera, S., Kumar, R.: Precise vision-aided aerial navigation. IEEE International Conference on Intelligent Robots and Systems pp. 688–695 (09 2014). https://doi.org/10.1109/IROS.2014.6942633
3. Choi, J., Myung, H.: Brm localization: Uav localization in gnss-denied environmentsbased on matching of numerical map and uav images (08 2020)
4. DeTone, D., Malisiewicz, T., Rabinovich, A.: Superpoint: Self-supervised interest point detection and description. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops. pp. 224–236 (2018)
5. Kacprzak, M.: Zastosowanie odbiorników gps i inercyjnych jednostek pomiarowych w akwizycji materiału zdjeciowego z wykorzystaniem bezzałogowych statków latajacych (bsl) pp. 267–276 (2016)
6. Lim, H., Sinha, S.N., Cohen, M.F., Uyttendaele, M.: Real-time image-based 6-dof localization in large-scale environments. In: 2012 IEEE conference on computer vision and pattern recognition. pp. 1043–1050. IEEE (2012)
7. Mughal, M.H., Khokhar, M.J., Shahzad, M.: Assisting uav localization via deep contextual image matching. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing **14**, 2445–2457 (2021). https://doi.org/10.1109/JSTARS.2021.3054832
8. Nassar, A., Amer, K., ElHakim, R., ElHelw, M.: A deep cnn-based framework for enhanced aerial imagery registration with applications to uav geolocalization pp. 1594–159410 (2018). https://doi.org/10.1109/CVPRW.2018.00201
9. Rublee, E., Rabaud, V., Konolige, K., Bradski, G.: Orb: An efficient alternative to sift or surf. In: 2011 International conference on computer vision. pp. 2564–2571. Ieee (2011)
10. Sarlin, P.E., DeTone, D., Malisiewicz, T., Rabinovich, A.: Superglue: Learning feature matching with graph neural networks. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 4938–4947 (2020)
11. Sun, J., Shen, Z., Wang, Y., Bao, H., Zhou, X.: Loftr: Detector-free local feature matching with transformers. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 8922–8931 (2021)
12. Yol, A., Delabarre, B., Dame, A., Dartois, J.E., Marchand, E.: Vision-based absolute localization for unmanned aerial vehicles pp. 3429–3434 (2014). https://doi.org/10.1109/IROS.2014.6943040