Supplementary material
Efficient Initial Pose-graph Generation for Global SfM

Daniel Barath[1,2,3], Dmytro Mishkin[1], Ivan Eichhardt[2], Ilia Shipachev[1], and Jiri Matas[1]
[1]Visual Recognition Group, Faculty of Electrical Engineering, Czech Technical University in Prague
[2]MPLab, SZTAKI, Budapest [3]Department of Computer Science, ETH Zurich

1. Second Nearest Ratio Test with Pool Size-sensitive Threshold

This section supplement Section 3 in the main paper. The common way of filtering unreliable tentative correspondence is the second-nearest ratio test (*aka* SIFT ratio test or Lowe ratio test) [3, 2]. In this test, after the descriptor matching, tentative correspondences get rejected if their "best" match is not significantly closer than the second best one. Thus, correspondences are filtered if

$$\frac{\texttt{distance of 1st nearest neighbor}}{\texttt{distance of 2nd nearest neighbor}} > \gamma,$$

where γ is the SIFT ratio threshold. Parameter γ is typically set to $0.7 - 0.9$, the common default being 0.8.

When applying the proposed *Epipolar Hashing* algorithm to select a subset of candidate matches for each feature point, this SIFT ratio test is rendered almost completely ineffective without the adaptation of threshold γ. This is caused by the fact that *Epipolar Hashing* reduces the number of features in the pool from which the neighbors are selected, significantly, to $2 - 30$ on average in our experiments. Due to this small pool, the density of points and thus the distance to second nearest descriptor is increased. Therefore the second best one is unlikely to be almost as close as the best match. In such cases, the standard SIFT ratio test fails to filter incorrect correspondences. In other words, there are many false positive matches.

Let us assume that non-matching descriptors are randomly distributed *w.r.t* the query descriptor. Consequently, the more descriptors we have in the pool, the lower the distance to the closest ones to the query will be. Therefore, if an equally strict condition on the quality of the tentative correspondences is required, in terms of false positives, regardless the number of features detected, we need to adapt the SNN ratio test threshold γ based on the number of features in the pool.

The following experiment was run on each image pair from the HPatches-Sequences [1] dataset. First, 8000 SIFT features were detected in both images. For each feature point, the nearest neighbor (minimizing the SIFT descriptor distance) and "reference" second nearest neighbor (second nearest at 8000) were found using the full set of features in the other image. The second nearest neighbor was selected from a random p-sized subset of points (second nearest at x). We then calculated the distance ratio of the nearest and second nearest neighbors. The results were averaged over all features and image pairs. In Fig. 1, this ratio is plotted as a function of the pool size p from which the second nearest neighbor is selected. The dependence of the SNN ratio on

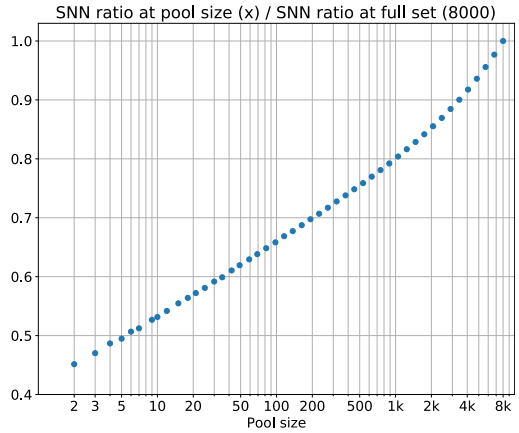

Figure 1: Dependence of Lowe's SNN ratio on the the descriptor pool size. Averaged over HPatches image pairs, 8000 SIFT features.

Table 1: The results of a global SfM [4] algorithm on scene Madrid Metropolis with and without adaptive second nearest distance ratio when applying the proposed Epipolar Hashing. The reported properties are: the number of views (2nd) and multi-view tracks (3rd) reconstructed by the global SfM procedure.

	# views	# tracks
w/o adaptive ratio test	136	9486
with adaptive ratio test	**282**	**29 665**

Figure 2: Example triplets of images and the found inlier correspondences used for calculating the values in Fig. 3. (**Orange**) Inlier correspondences between the 1st and 2nd images which are visible in the 3rd one. (**Green**) Correspondences which got good rank by the proposed method and are consistent with the ground truth epipolar geometry between the 2nd and 3rd images. (**Red**) Correspondences which got good rank and are inconsistent with the epipolar geometry. Significantly more "good" correspondences got good ranking than incorrect ones – the number of green points is higher than that of the red ones.

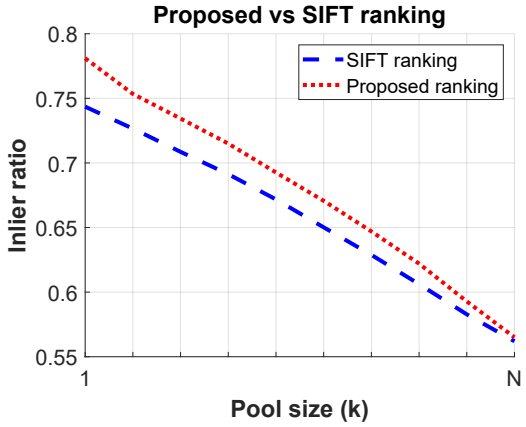

Figure 3: Inlier ratio in the first k correspondences when ordered by the proposed or SIFT rankings. The values are calculated from 500 randomly selected image triplets from the London Bridge dataset.

the feature pool size is almost linear in the log space. We use this dependence to correct the SNN ratio threshold – the default value of 0.9 for mutual SNN ratio [2] is multiplied by the y-value depending on the feature number in the pool. For example, for 5 features, the resulting threshold is 0.45.

Example reconstruction results with and without adaptive ratio test on scene Madrid Metropolis are shown in Table 1. It can be seen that the adaptive ratio test is extremely important in Epipolar Hashing. The large difference in the number of reconstructed views is caused by the following phenomenon. The proposed A*-based algorithm first attempts to efficiently connect a new image to the pose-graph by extending tracks using Epipolar Hashing. If this process produces seemingly sufficient number of correspondences, full descriptor-based matching does not take place. A high number of false positives in the Epipolar Hashing process leads to an incorrect decision that full matching is not needed and an incorrect pose is obtained from the false positive matches, which are all, by construction, consistent with the initial estimated epipolar geometry, which is incorrect or very imprecise.

2. Adaptive Correspondence Ranking

This section supplement Section 4 in the main paper. In order to compare the effect of the proposed correspondence re-ranking strategy, we selected 500 image triplets from the London Bridge dataset, see Fig. 2 for examples. The images in each triplet were selected randomly but in a way to ensure that they have a commonly visible area. For each triplet, the epipolar geometry was estimated between the first two images by standard RANSAC. Next, for estimating the relative pose between the second and third images, the correspondences were ordered either by the proposed re-ranking strategy or by their SIFT scores. Finally, we measured the inlier ratio in the sets consisting of the first k correspondences, $k \in [1, N]$. Fig. 3 plots the inlier ratio, averaged over the 500 tests, as a function of the pool size k. The proposed algorithm leads to a better ordering than exploiting the SIFT scores – its inlier ratio is higher among the first k correspondences.

3. Pose-Graph Traversal

Comparison on scene Alamo from the 1DSfM dataset [5] of the proposed A* and breadth-first traversals are shown in Fig. 4. The top plot shows the cumulative distribution functions of the processing times in seconds. It can be seen that if $\lambda < 1$, *i.e.* the weighting parameter of the heuristic, the A* traversal is significantly faster than breadth-first. Note that $\lambda = 1$ corresponds to the case when maximizing the similarity to the destination node along the path is turned off in the heuristic and, thus, the traversal does not aim at finding the destination node.

The bottom plot shows the run-time in seconds of the tested traversals as the function of the graph size. It can be seen that, if $\lambda < 1$, the run-time to find a path from the source to the destination nodes is significantly lower than by using the breadth-first traversal. If we assume that the relative poses in the pose-graph are reasonably accurate and, thus, set λ to a small value, *e.g.* 0.4, the run-time of finding a path by A* is 22.4 times lower than the time of breadth-first even in the full graph consisting of 47648 vertices.

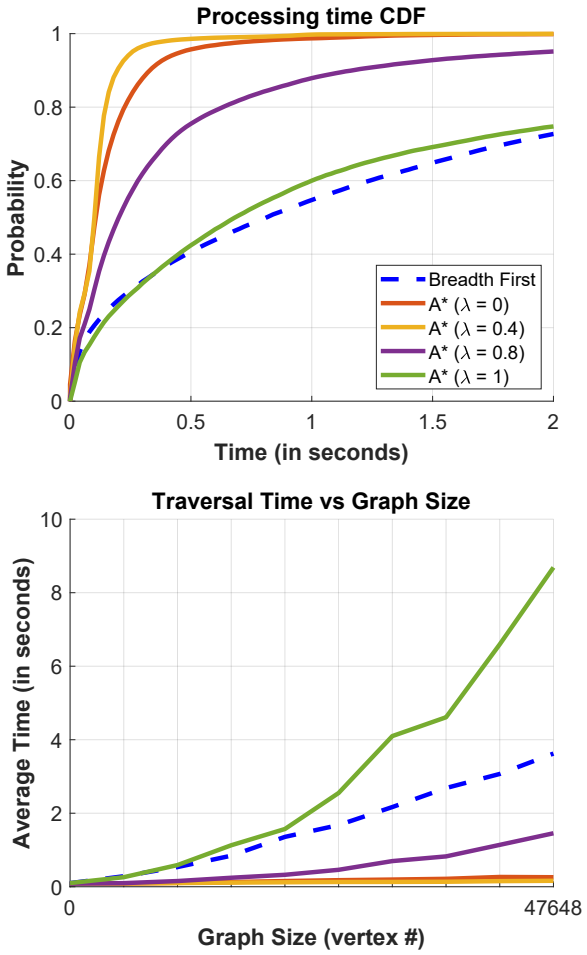

Figure 4: (**Top**) The cumulative distribution functions of the processing times in seconds of the proposed A* and breadth-first traversals on scene Alamo from the 1DSfM dataset [5]. (**Bottom**) The run-time in seconds of the tested traversals as the function of the graph size.

References

[1] Vassileios Balntas, Karel Lenc, Andrea Vedaldi, and Krystian Mikolajczyk. HPatches: A benchmark and evaluation of handcrafted and learned local descriptors. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. 1

[2] 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. *International Journal of Computer Vision*, 2020. 1, 2

[3] David Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004. 1

[4] Chris Sweeney. Theia multiview geometry library. http://theia-sfm.org. 1

[5] K. Wilson and N. Snavely. Robust Global Translations with 1DSfM. In *European Conference on Computer Vision*, pages 61–75, 2014. 3