# Fast Image-Based Localization using Direct 2D-to-3D Matching - Supplementary Material

Torsten Sattler, Bastian Leibe, Leif Kobbelt
RWTH Aachen University
{tsattler@cs, leibe@umic, kobbelt@cs}.rwth-aachen.de

In this supplementary material we present further details on the results of the experiments carried out in the paper. The material consists of three parts. The first part (Tables A-C) gives additional information for the Rome and Vienna dataset that complement Table 2 of the paper. The second part (Figure 1) details the effect of $N_t$ on the inlier ratios of the registered query images and in the last part (Table E) we examine the experiment of trying to register query images from the other datasets against a model.

As mentioned in the paper, the behavior of tree-based and vocabulary-based prioritized search (VPS) methods is similar on all datasets. Therefore, only detailed results for the Dubrovnik dataset were given in Table 2 of the paper. Detailed results of the experiments for the Rome and Vienna dataset can be found in Tables B and C. Besides using the FLANN library [3] for kd-tree search, we also experimented with the ANN library written by David M. Mount and Sunil Arya [1]. Results for using ANN with different limits on the maximal number of leaf nodes visited during the search can be found in Tables A- C for the three datasets. Although ANN also is able to register significantly more images than the indirect method proposed by Li *et al*. [2], it performs slightly worse than FLANN.

We mentioned in the paper that the average time spent on pose estimation inside a RANSAC loop does not change significantly for choosing a value $N_t \geq 100$ for the threshold on the maximal number of correspondences (*c.f*. Table 3 in the paper, listed for reference as Table D in the supplementary material). From this observation we deduced that the choice of $N_t$ does not affect the average inlier ratio for the registered images too much. We concluded that the threshold $N_t$ on the number of correspondences does not bias the correspondence search towards false correspondences. To support this conclusion, we analyze the effect of different values for $N_t$ on the distribution of inlier ratios over the registered query images, with $R$ set to $0.2$. Figure 1 shows these distributions, obtained by averaging the histograms of the inlier ratios over $10$ repetitions of the experiment, for the different datasets and the VPS methods *all descriptors* and *integer mean per vw*. Although the distributions are slightly shifted towards lower inlier ratios, there is little difference when choosing a finite value for $N_t$ on both the Dubrovnik and the Rome dataset. While there is a noticeable difference between the distributions for the Vienna dataset, the difference is only significant for higher inlier ratios where it has a smaller impact on the runtime of RANSAC. The results confirm our conclusion that the choice of $N_t$ does not bias the correspondence search towards false correspondences.

The last part of the supplementary material deals with the experiment of trying to register query images from other datasets against a 3D model. While no image from another dataset could be registered, significantly fewer correspondences are found for these images as can be seen in Table E. As evident by the mean number of correspondences found, $12/N > R = 0.2$ holds for query images from other datasets, with $N$ being the number of correspondences found. As a consequence, the RANSAC-based pose estimation is accelerated and thus the rejection time reduced. Notice that the mean number of correspondences found for rejected images is lower than the threshold $N_t = 100$ and thus the prioritized search considers all features in the query image. The slight deviations in the mean rejection time for images from another dataset are caused by the random nature of RANSAC. The number of rejected images fluctuates for images from the same dataset (*c.f*. Table D). Therefore, the deviation in the rejection time for these images is also affected by the time spent on the linear search.

## References

[1] S. Arya, D. M. Mount, N. S. Netanyahu, R. Silverman, and A. Y. Wu. An optimal algorithm for approximate nearest neighbor searching fixed dimensions. *J. ACM*, 45(6):891–923, 1998.

[2] Y. Li, N. Snavely, and D. P. Huttenlocher. Location recognition using prioritized feature matching. In *ECCV*, 2010.

[3] M. Muja and D. G. Lowe. Fast approximate nearest neighbors with automatic algorithm configuration. In *VISAPP*, 2009.

| | Method | Dubrovnik: registered images: # reg. images | # corr. | inlier ratio | corr. [s] | RNSC. [s] | total [s] | rejected: time [s] |
|---|---|---|---|---|---|---|---|---|
| tree-based | flann 50 leaves | 789 | 545.3 | 0.60 | 1.29 | 0.34 | 1.63 | 8.92 |
| | flann 100 leaves | 793 | 581.8 | 0.62 | 1.64 | 0.39 | 2.04 | 11.54 |
| | flann 200 leaves | 794 | 609.9 | 0.64 | 2.40 | 0.31 | 2.71 | 12.21 |
| | **flann 300 leaves** | **795** | 620.1 | 0.64 | 3.11 | 0.30 | **3.40** | **14.45** |
| | flann 500 leaves | 794 | 629.1 | 0.65 | 4.55 | 0.21 | 4.76 | 23.86 |
| | ann 50 leaves | 768 | 432.4 | 0.56 | 1.53 | 0.68 | 2.21 | 11.18 |
| | ann 100 leaves | 779 | 488.7 | 0.59 | 2.07 | 0.46 | 2.53 | 4.27 |
| | ann 200 leaves | 786 | 541.2 | 0.62 | 3.04 | 0.37 | 3.41 | 8.91 |
| | ann 300 leaves | 790 | 568.5 | 0.63 | 3.97 | 0.53 | 4.51 | 9.16 |
| | ann 500 leaves | 792 | 598.6 | 0.63 | 5.72 | 0.32 | 6.04 | 11.64 |
| VPS | **all descriptors** | **785** | 537.0 | 0.64 | 0.66 | 0.15 | **0.81** | **2.19** |
| | mean | 774 | 472.9 | 0.62 | 1.08 | 0.53 | 1.61 | 2.36 |
| | medoid | 762 | 412.5 | 0.62 | 0.50 | 0.34 | 0.84 | 1.58 |
| | mean per vw | 782 | 523.8 | 0.64 | 0.99 | 0.32 | 1.31 | 5.25 |
| | **integer mean per vw** | **783** | 523.0 | 0.64 | 0.56 | 0.31 | **0.87** | **5.35** |
| | medoid per vw | 778 | 480.7 | 0.63 | 0.52 | 0.14 | 0.66 | 4.34 |
| P2F [2] | | 753 | - | - | - | - | 0.73 | 2.70 |

Table A: Comparison of different direct 2D-to-3D matching approaches on the Dubrovnik dataset. We list the number of successfully registered images (#reg. images), as well as the average time needed to register / reject an image. We furthermore report for successfully registered images the average number of correspondences (#corr.), the average inlier ratio, the average time needed to compute the correspondences (corr.) and the average time RANSAC needs to estimate the pose (RNSC.).

| | Method | Rome: registered images: # reg. images | # corr. | inlier ratio | corr. [s] | RNSC. [s] | total [s] | rejected: time [s] |
|---|---|---|---|---|---|---|---|---|
| tree-based | flann 50 leaves | 978 | 462.0 | 0.72 | 2.09 | 0.58 | 2.66 | 5.36 |
| | flann 100 leaves | 978 | 494.1 | 0.75 | 2.39 | 0.44 | 2.82 | 8.52 |
| | flann 200 leaves | 981 | 518.4 | 0.76 | 2.98 | 0.51 | 3.49 | 5.69 |
| | **flann 300 leaves** | **983** | 526.8 | 0.77 | 3.59 | 0.38 | **3.97** | **6.27** |
| | flann 500 leaves | 985 | 535.0 | 0.77 | 4.83 | 0.45 | 5.28 | 3.28 |
| | ann 50 leaves | 970 | 350.5 | 0.67 | 1.99 | 0.55 | 2.54 | 8.07 |
| | ann 100 leaves | 976 | 397.3 | 0.71 | 2.39 | 0.36 | 2.75 | 7.57 |
| | ann 200 leaves | 979 | 442.0 | 0.74 | 3.20 | 0.49 | 3.70 | 8.28 |
| | ann 300 leaves | 981 | 465.6 | 0.75 | 4.01 | 0.42 | 4.43 | 12.50 |
| | ann 500 leaves | 980 | 493.9 | 0.76 | 5.40 | 0.53 | 5.92 | 14.52 |
| VPS | **all descriptors** | **979** | 493.9 | 0.75 | 1.00 | 0.53 | **1.53** | **4.07** |
| | mean | 972 | 405.4 | 0.75 | 1.65 | 0.48 | 2.13 | 1.28 |
| | medoid | 961 | 347.5 | 0.74 | 0.71 | 0.34 | 1.05 | 3.74 |
| | mean per vw | 976 | 475.3 | 0.75 | 1.64 | 0.58 | 2.23 | 6.50 |
| | **integer mean per vw** | **976** | 474.0 | 0.75 | 0.74 | 0.59 | **1.33** | **5.92** |
| | medoid per vw | 972 | 424.0 | 0.75 | 0.75 | 0.42 | 1.17 | 7.27 |
| P2F [2] | | 921 | - | - | - | - | 0.91 | 2.93 |

Table B: Details for the Rome dataset to complement Table 2 of the paper.

| | Method | Vienna # reg. images | # corr. | inlier ratio | corr. [s] | RNSC. [s] | total [s] | rejected time [s] |
|---|---|---|---|---|---|---|---|---|
| | | registered images | | | | | | rejected |
| tree-based | flann 50 leaves | 218 | 447.3 | 0.60 | 1.16 | 1.36 | 2.52 | 8.84 |
| | flann 100 leaves | 219 | 463.2 | 0.63 | 1.62 | 0.62 | 2.25 | 4.64 |
| | flann 200 leaves | 219 | 470.8 | 0.67 | 2.49 | 0.12 | 2.61 | 2.93 |
| | **flann 300 leaves** | **220** | 471.5 | 0.68 | 3.34 | 0.09 | **3.44** | **2.72** |
| | flann 500 leaves | 220 | 472.3 | 0.69 | 4.99 | 0.07 | 5.06 | 3.65 |
| | ann 50 leaves | 210 | 399.6 | 0.52 | 1.31 | 2.54 | 3.85 | 15.94 |
| | ann 100 leaves | 215 | 417.9 | 0.57 | 1.89 | 2.00 | 3.89 | 9.23 |
| | ann 200 leaves | 218 | 439.9 | 0.62 | 2.94 | 0.60 | 3.54 | 4.96 |
| | ann 300 leaves | 217 | 454.5 | 0.64 | 3.93 | 0.59 | 4.53 | 4.53 |
| | ann 500 leaves | 218 | 465.8 | 0.67 | 5.80 | 0.12 | 5.92 | 4.81 |
| VPS | **all descriptors** | **211** | 414.7 | 0.59 | 0.42 | 1.41 | **1.83** | **9.95** |
| | mean | 210 | 370.7 | 0.57 | 0.65 | 1.40 | 2.05 | 9.19 |
| | medoid | 203 | 339.6 | 0.55 | 0.34 | 1.89 | 2.23 | 9.40 |
| | mean per vw | 212 | 399.1 | 0.58 | 0.59 | 1.87 | 2.46 | 6.87 |
| | **integer mean per vw** | **211** | 395.0 | 0.58 | 0.32 | 1.71 | **2.02** | **7.59** |
| | medoid per vw | 211 | 375.3 | 0.57 | 0.35 | 1.46 | 1.81 | 8.25 |
| P2F [2] | | 204 | - | - | - | - | 0.55 | 1.96 |

Table C: Details for the Vienna dataset to complement Table 2 of the paper.

| | $N_t$ | all descriptors # reg. | linear search [s] | RANSAC [s] | total [s] | integer mean per vw # reg. | linear search [s] | RANSAC [s] | total [s] |
|---|---|---|---|---|---|---|---|---|---|
| Dubrovnik | 50 | $778.90 \pm 1.52$ | 0.04 | 0.05 | $0.23 \pm 0.00$ | $775.80 \pm 1.48$ | 0.03 | 0.05 | $0.21 \pm 0.00$ |
| | **100** | $\mathbf{783.90 \pm 1.60}$ | 0.10 | 0.08 | $\mathbf{0.31 \pm 0.01}$ | $\mathbf{782.00 \pm 0.82}$ | 0.08 | 0.08 | $\mathbf{0.28 \pm 0.01}$ |
| | 150 | $783.90 \pm 1.10$ | 0.16 | 0.08 | $0.36 \pm 0.01$ | $781.80 \pm 1.40$ | 0.12 | 0.08 | $0.32 \pm 0.01$ |
| | 200 | $784.40 \pm 1.26$ | 0.20 | 0.08 | $0.40 \pm 0.01$ | $782.50 \pm 1.35$ | 0.15 | 0.08 | $0.35 \pm 0.01$ |
| | $\infty$ | $784.60 \pm 1.17$ | 0.47 | 0.08 | $0.68 \pm 0.01$ | $782.50 \pm 1.08$ | 0.34 | 0.08 | $0.54 \pm 0.01$ |
| Rome | 50 | $972.00 \pm 1.41$ | 0.06 | 0.02 | $0.18 \pm 0.00$ | $971.30 \pm 1.25$ | 0.05 | 0.02 | $0.16 \pm 0.00$ |
| | **100** | $\mathbf{976.90 \pm 1.29}$ | 0.15 | 0.05 | $\mathbf{0.29 \pm 0.00}$ | $\mathbf{974.60 \pm 1.65}$ | 0.11 | 0.05 | $\mathbf{0.25 \pm 0.00}$ |
| | 150 | $977.80 \pm 1.32$ | 0.23 | 0.06 | $0.39 \pm 0.01$ | $976.50 \pm 1.51$ | 0.17 | 0.06 | $0.33 \pm 0.01$ |
| | 200 | $979.20 \pm 1.75$ | 0.30 | 0.07 | $0.46 \pm 0.01$ | $976.90 \pm 1.52$ | 0.22 | 0.07 | $0.38 \pm 0.00$ |
| | $\infty$ | $980.10 \pm 0.88$ | 0.81 | 0.07 | $0.98 \pm 0.00$ | $976.90 \pm 1.20$ | 0.57 | 0.07 | $0.74 \pm 0.00$ |
| Vienna | 50 | $200.40 \pm 1.26$ | 0.02 | 0.13 | $0.28 \pm 0.01$ | $199.10 \pm 1.20$ | 0.02 | 0.10 | $0.26 \pm 0.01$ |
| | **100** | $\mathbf{207.70 \pm 1.06}$ | 0.06 | 0.30 | $\mathbf{0.50 \pm 0.02}$ | $\mathbf{206.90 \pm 0.88}$ | 0.05 | 0.28 | $\mathbf{0.46 \pm 0.02}$ |
| | 150 | $208.20 \pm 0.92$ | 0.09 | 0.30 | $0.52 \pm 0.03$ | $207.90 \pm 0.74$ | 0.07 | 0.29 | $0.50 \pm 0.03$ |
| | 200 | $208.80 \pm 1.23$ | 0.11 | 0.29 | $0.54 \pm 0.04$ | $208.20 \pm 1.14$ | 0.08 | 0.30 | $0.52 \pm 0.03$ |
| | $\infty$ | $207.90 \pm 1.29$ | 0.24 | 0.27 | $0.65 \pm 0.03$ | $208.20 \pm 0.42$ | 0.17 | 0.28 | $0.59 \pm 0.03$ |

Table D: Table 3 from the paper, listed for reference.

| Query Images from | Dubrovnik avrg. #corr. | Dubrovnik avrg. time [s] | Rome avrg. #corr. | Rome avrg. time [s] | Vienna avrg. #corr. | Vienna avrg. time [s] |
|---|---|---|---|---|---|---|
| all descriptors | | | | | | |
| Dubrovnik (800 images) | $\mathbf{37.79 \pm 17.84}$ | $\mathbf{2.22 \pm 0.26}$ | $12.47 \pm 7.69$ | $1.16 \pm 0.01$ | $25.48 \pm 14.47$ | $0.75 \pm 0.01$ |
| Rome (1000 images) | $18.23 \pm 12.85$ | $0.59 \pm 0.00$ | $\mathbf{39.62 \pm 29.75}$ | $\mathbf{1.90 \pm 0.10}$ | $22.50 \pm 15.21$ | $0.68 \pm 0.00$ |
| Vienna (266 images) | $28.58 \pm 20.05$ | $1.38 \pm 0.00$ | $18.51 \pm 12.81$ | $1.31 \pm 0.00$ | $\mathbf{42.72 \pm 28.85}$ | $\mathbf{2.40 \pm 0.06}$ |
| integer mean per vw | | | | | | |
| Dubrovnik (800 images) | $\mathbf{35.84 \pm 17.76}$ | $\mathbf{1.70 \pm 0.18}$ | $12.52 \pm 7.65$ | $0.86 \pm 0.02$ | $25.01 \pm 14.08$ | $0.66 \pm 0.01$ |
| Rome (1000 images) | $17.67 \pm 12.36$ | $0.48 \pm 0.00$ | $\mathbf{39.21 \pm 28.95}$ | $\mathbf{1.66 \pm 0.10}$ | $22.24 \pm 15.16$ | $0.60 \pm 0.00$ |
| Vienna (266 images) | $27.78 \pm 19.49$ | $1.22 \pm 0.03$ | $18.48 \pm 12.86$ | $1.01 \pm 0.00$ | $\mathbf{41.71 \pm 28.80}$ | $\mathbf{2.43 \pm 0.08}$ |

Table E: Rejection performance for choosing $N_t = 100$ and $R = 0.2$. Since no query image belonging to a different dataset could be registered, we do not report the number of rejected images but only the average number of correspondences for images that were rejected (and the standard deviation in the number of correspondences) as well as the average time for rejecting an image and the deviation of that mean rejection time. Both methods find significantly fewer correspondences for query images from other datasets. As a result, the initial inlier ratio $12/N$ is substantially higher than $0.2$, with $N$ being the number of correspondences found for a query image. Therefore, RANSAC needs fewer steps to reject an image.

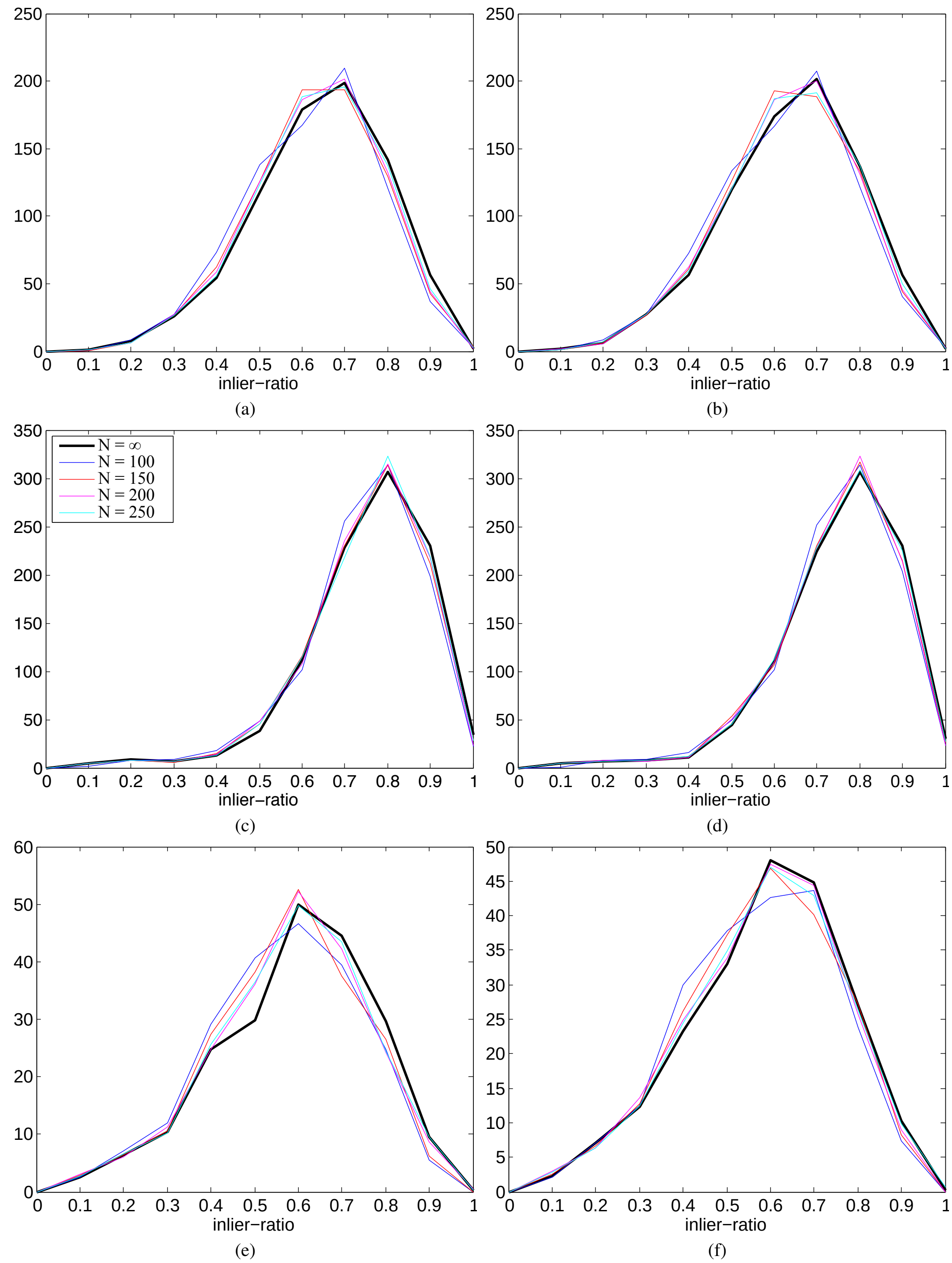


Figure 1: Effect of $N_t$ on the inlier ratios. The figures show the distribution of inlier ratios for the registered query images for Dubrovnik using *all descriptors* (a) and *integer mean per vw* (b), for Rome using *all descriptors* (c) and *integer mean per vw* (d), and for Vienna using *all descriptors* (e) and *integer mean per vw* (f). The results confirm the conclusion made in the paper that the threshold $N_t$ on the number of correspondences does not bias the correspondence search towards false correspondences.