Can You Trust Your Pose? Confidence Estimation in Visual Localization

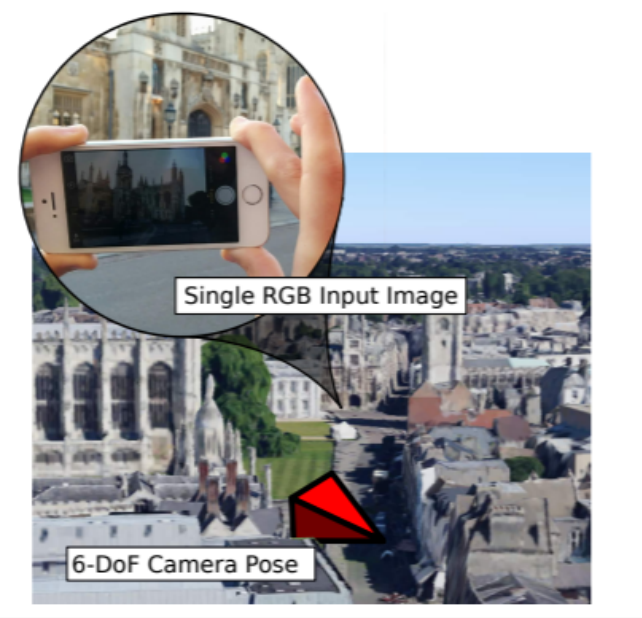
Luca Ferranti[1,2], Xiaotian Li[2], Jani Boutellier[1], Juho Kannala[2]

December 10, 2020

[1]University of Vaasa, [2]Aalto University

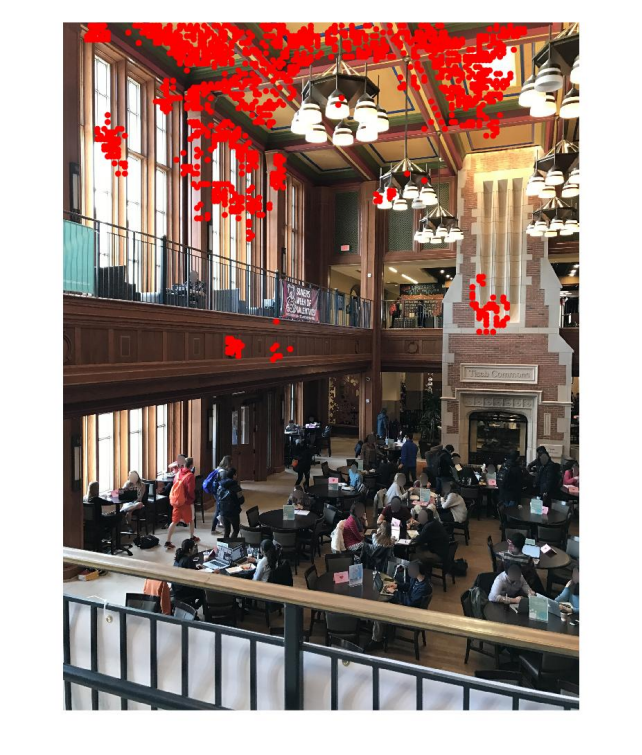
Introduction

Image-Based Localization – Overview



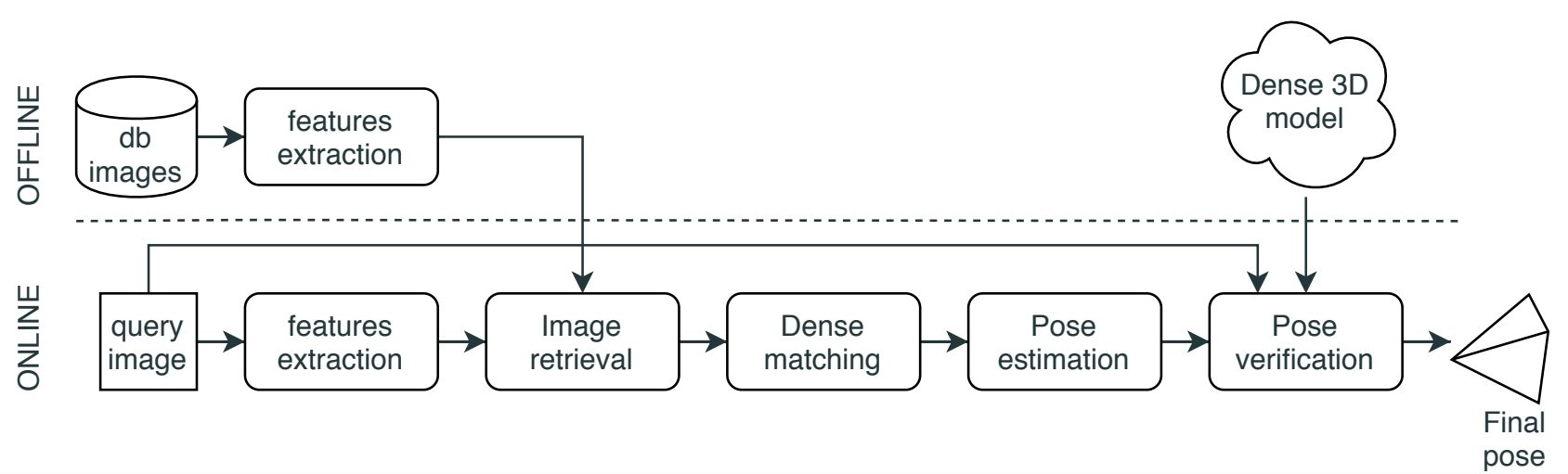
- Estimate camera position and orientation (shortly camera pose) from a given picture.
- Applications: odometry, augmented reality, autonomous vehicles
- requires knowledge of the environment, either through a 3D point cloud and/or through a large database of images.
- needs at least 3 2D-3D correspondences $\Rightarrow$ P3P-algorithm. Generally inside a RANSAC loop

Need for confidence estimation



- A wrong pose may still have a high number of inliers
- Example on the left: 1200 inliers, still camera position 12 m wrong
- Main factors causing failure: redundant patterns (brick walls, textureless surfaces), varying illumination conditions
- A more robust metric is needed
- Opposed to previous work, we want now to compare different query images

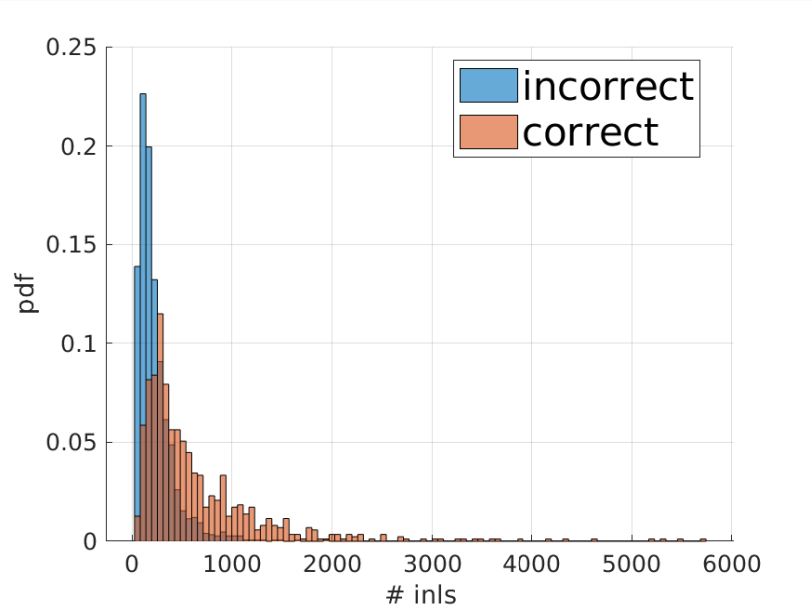
InLoc Pipeline



- **Image Retrieval:** 100 most similar database images are retrieved
- **Dense Matching:** for each query-database image pair, 2D-2D points correspondences are formed. These induce 2D-3D correspondences. Only 10 best pairs are kept.
- **Pose Estimation:** for each pair, camera pose is computed suing P3P-RANSAC, obtaining 10 candidate poses.
- **Pose Verification:** 10 candidate poses are reranked to choose the best candidate.

Confidence Measure

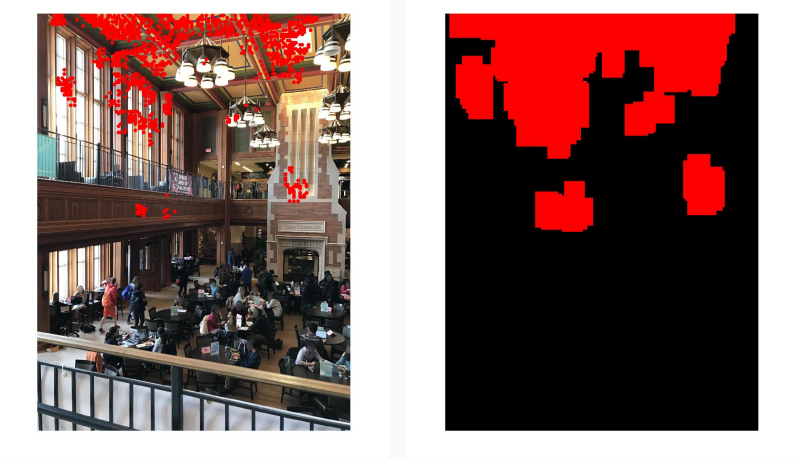
Number of Inliers

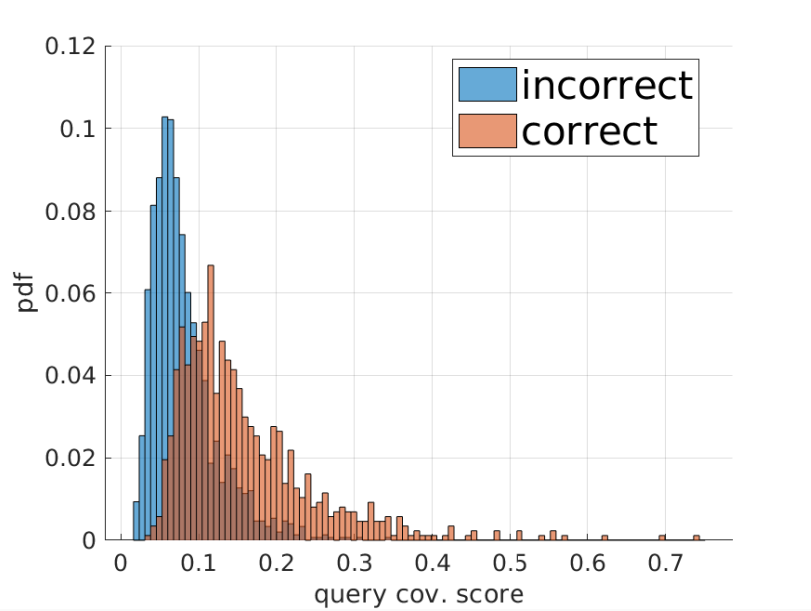


- The inliers count is the standard confidence measure, the higher, the more likely the pose is correct
- **However, that's not the case**
- Redudandancy in the environment is particuarly tricky when using the inliers count.

Inliers Spatial Distribution

- **Idea:** The estimate is more reliable, if the inliers cover a wider area in the picture.
- A *coverage map*, visualizing the inliers spread, is computed and used to compute a *coverage score*. This is done both for query and database images.



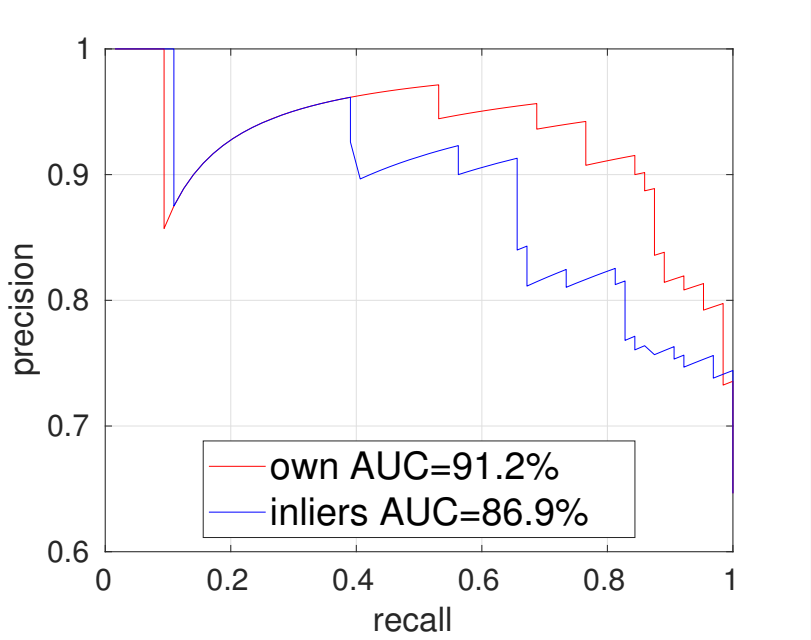


Learning a Confidence Measure

- We have three scalar parameters: inliers count x_1, query image coverage score x_2 and database image coverage score x_3.
- To compute our confidence measure, we use logistic regression, i.e. $\gamma = \mathrm{logsig}\left(b + \sum_{i=1}^{3} w_i x_i\right)$
- A simple model was chosen because
  - We want our model to work on different datasets and thresholds, hence we want to avoid overfitting to a specific dataset
  - Confidence estimation is an additional step to the pipeline, with our simple model the added computation time is neglettable

Results

Confidence Estimation in InLoc



- We changed the labels of test data changing error threshold and evaluated our model again

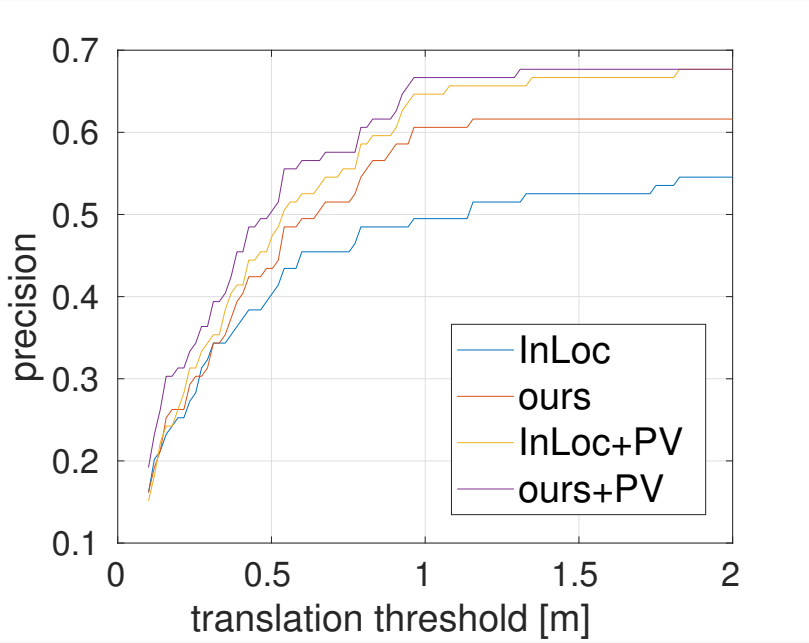
Error Threshold	AUC (inls count)	AUC (our)
1.5 m, 10°	87.1%	91.8%
1 m, 10°	86.9%	91.1%
0.5 m, 10°	68.7%	76.8%
0.25 m, 10°	51.8%	56.8%

Confidence Estimation for More Accurate Poses

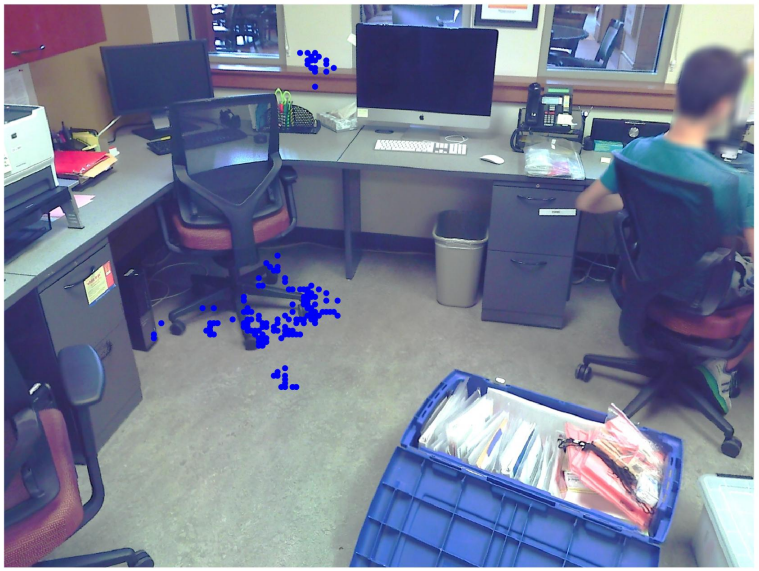
- In the experiments so far our confidence estimator was *after* the pipeline and used to compare different query images.
- Can our estimator be included *inside* the pipeline to choose the best candidate of a query image?

Accuracies at 1 m, 10° error threshold.

	InLoc	Ours
Without PV	49.5%	60.6%
With PV	64.7%	66.7%



Confidence Estimation for More Accurate Poses

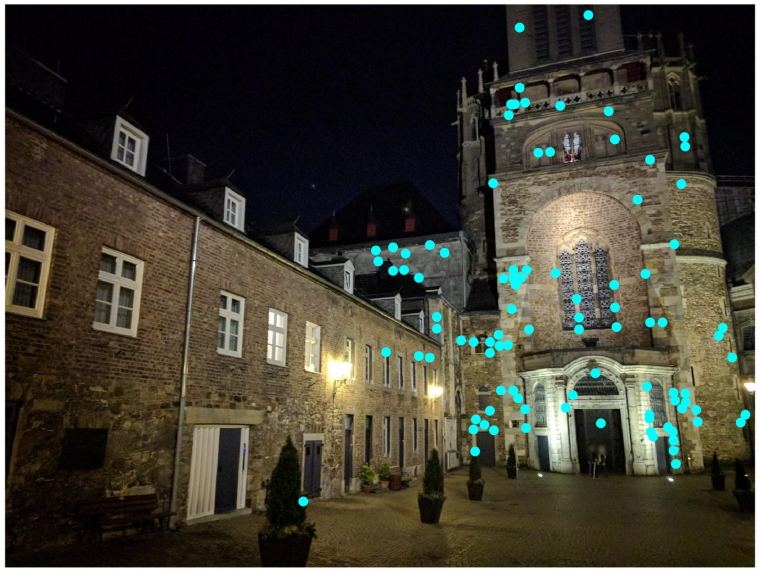


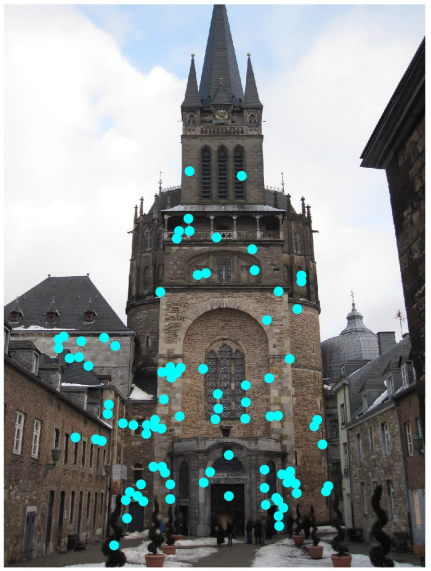
Generalization to Different Datasets: Aachen

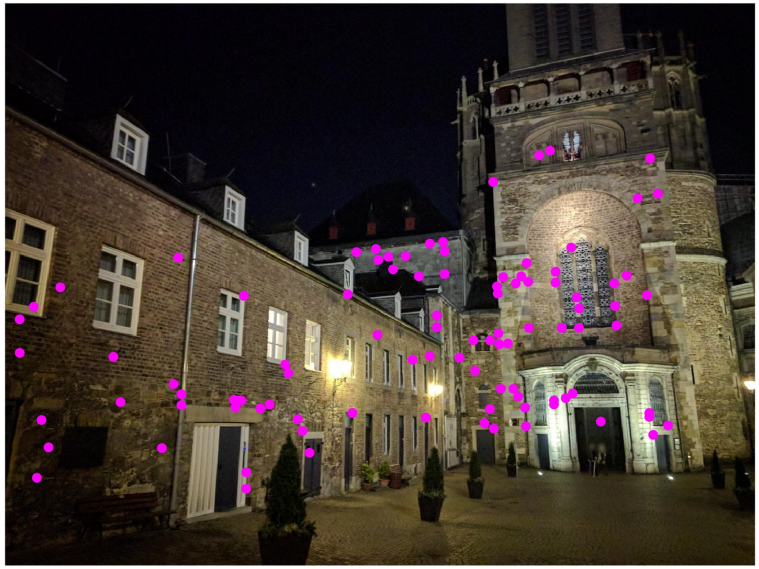
- A different dataset is considered: Aachen, containing day and night scenes from Aachen city.
- The pipeline produces 10 candidate poses and chooses the best pose using the number of inliers. We show ranking with our confidence measure gives better accuracies. (**Note!** our method was trained with InLoc at a fixed threshold, Aachen is completely unforeseen)
- Accuracies evaluated at 3 different thresholds.

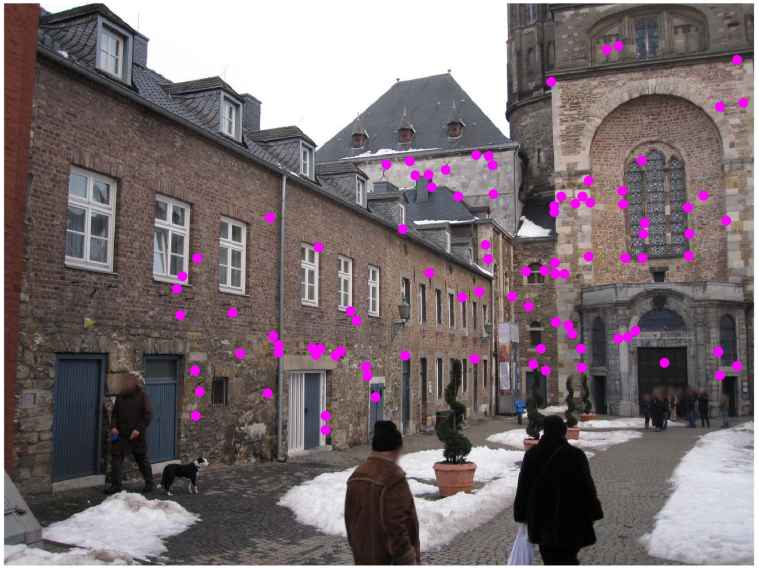
Scene	Baseline accuracies [%]	our accuracies [%]
Day	70.9/81.9/91.6	71.4/83.0/91.6
Night	32.7/43.9/64.3	36.7/45.9/64.3

Generalization to Different Datasets: Aachen









Conclusions

Conclusions

- We formalized the confidence estimation problem.
- Opposed to previous work, we aim at comparing different query images
- We proposed a standardizes metric to quantify the confidence of the estimated pose
- We showed our metric is independent of the dataset or error threshold used for training
- We showed our metric can also be used to achieve more accurate camera poses

Questions?