VISUAL LOCALIZATION USING SPARSE SEMANTIC 3D MAP

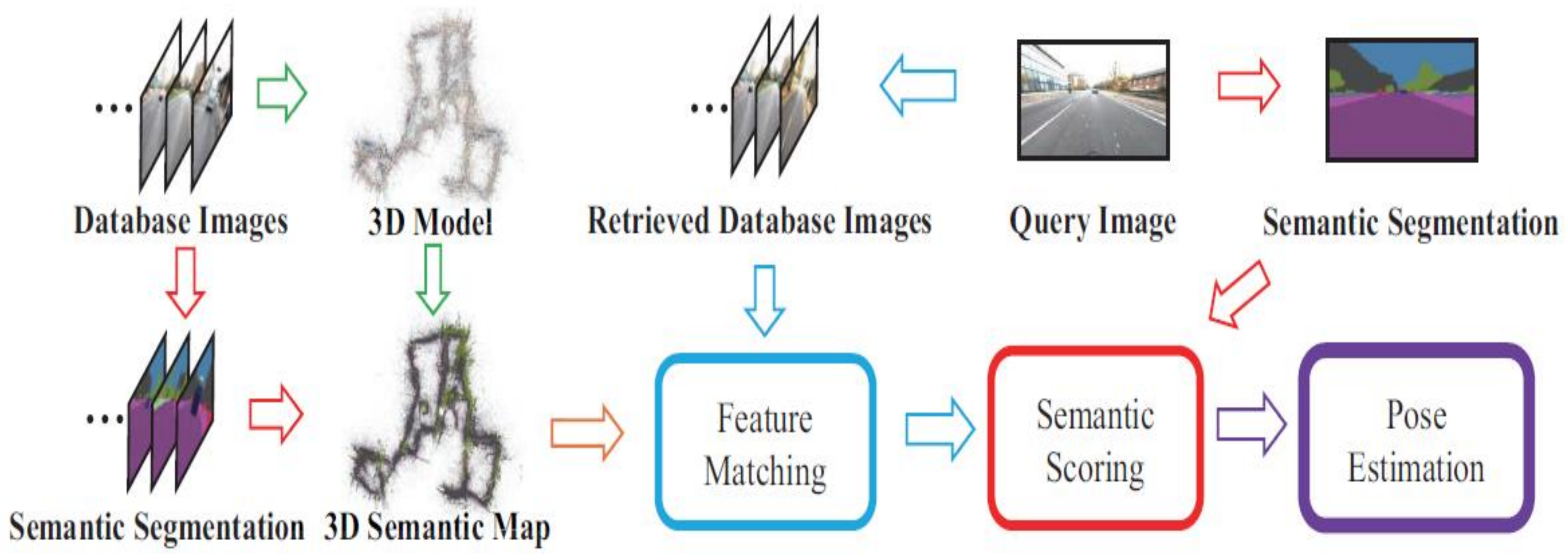
Tianxin Shi[1,2] Shuhan Shen[1,2] Xiang Gao[1,2] Lingjie Zhu[1,2]

1. NLPR, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, P. R. China 2. University of Chinese Academy of Sciences, Beijing 100049, P. R. China

Introduction

Accurate and robust visual localization under a wide range of viewing condition variations is the key component for many computer vision applications. Under these changing conditions, most traditional methods would fail to locate the camera. In this paper we present a visual localization algorithm that combines traditional methods with semantic information to improve localization results.

Method Overview



Sparse Semantic Map

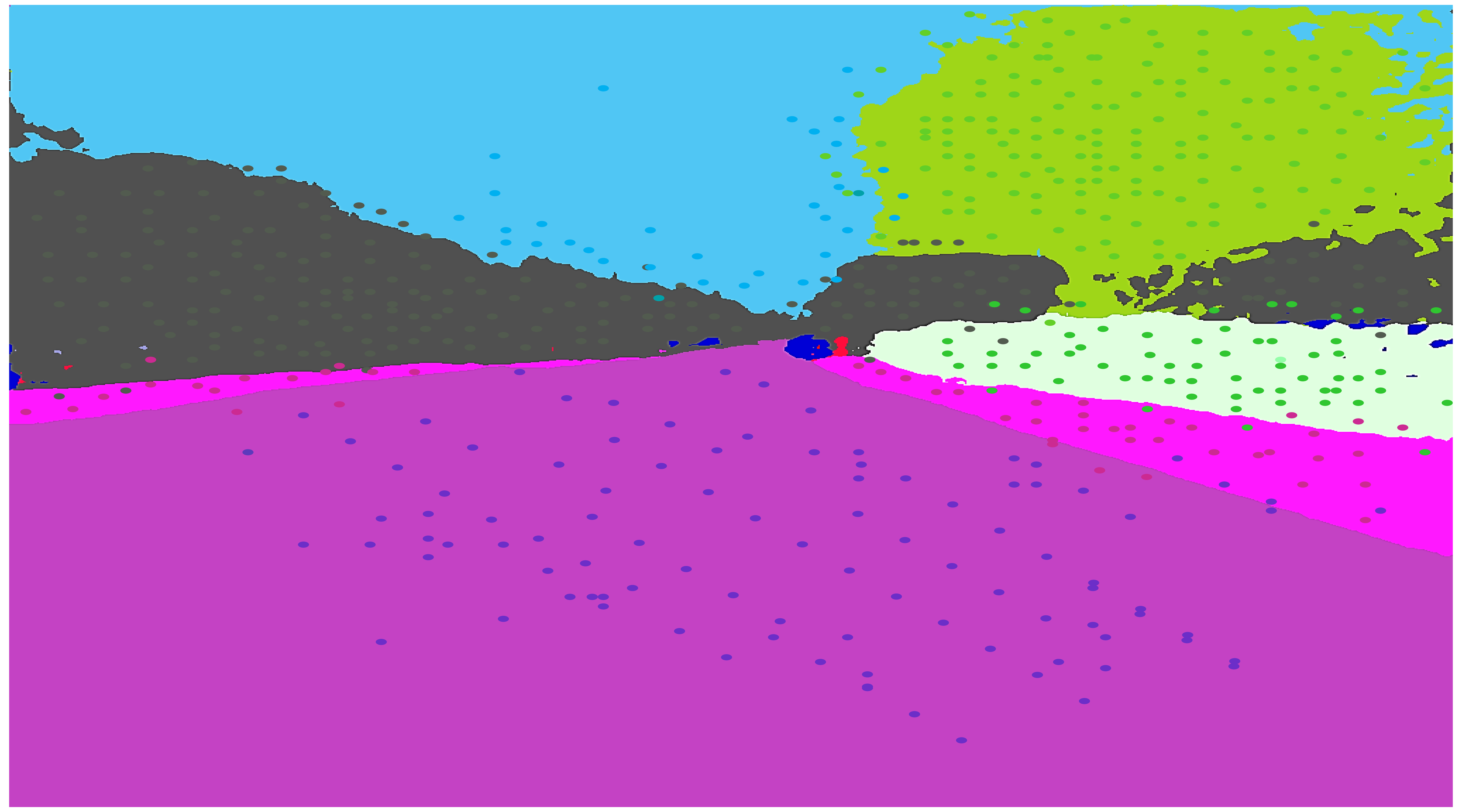
We first run a regular SfM pipeline using all database images to construct a sparse 3D model of the scene. After SfM, the location as well as the visible images of each 3D point are obtained. Given semantic segmentation for all database images, we assign each 3D point a semantic label by maximum voting with reprojection pixel labels in all its visible images.

Image Retrieval

In the image retrieval step, we use the classical image retrieval method provided by colmap which using vocabulary tree with spatial re-ranking to obtain top-k ranked database images.

Projection

We use one retrieved image at a time for feature matching with query image. Through 2D-2D matches and 2D-3D correspondences provided by SfM, we can obtain 2D-3D matches between query image and 3D map. Then, these 2D-3D matches are used to recover a temporary query pose by applying a PnP solver. Given semantic segmentation about query image, we only project the visible 3D points into query image by the estimated temporary query pose to check the number of consistent semantic points.

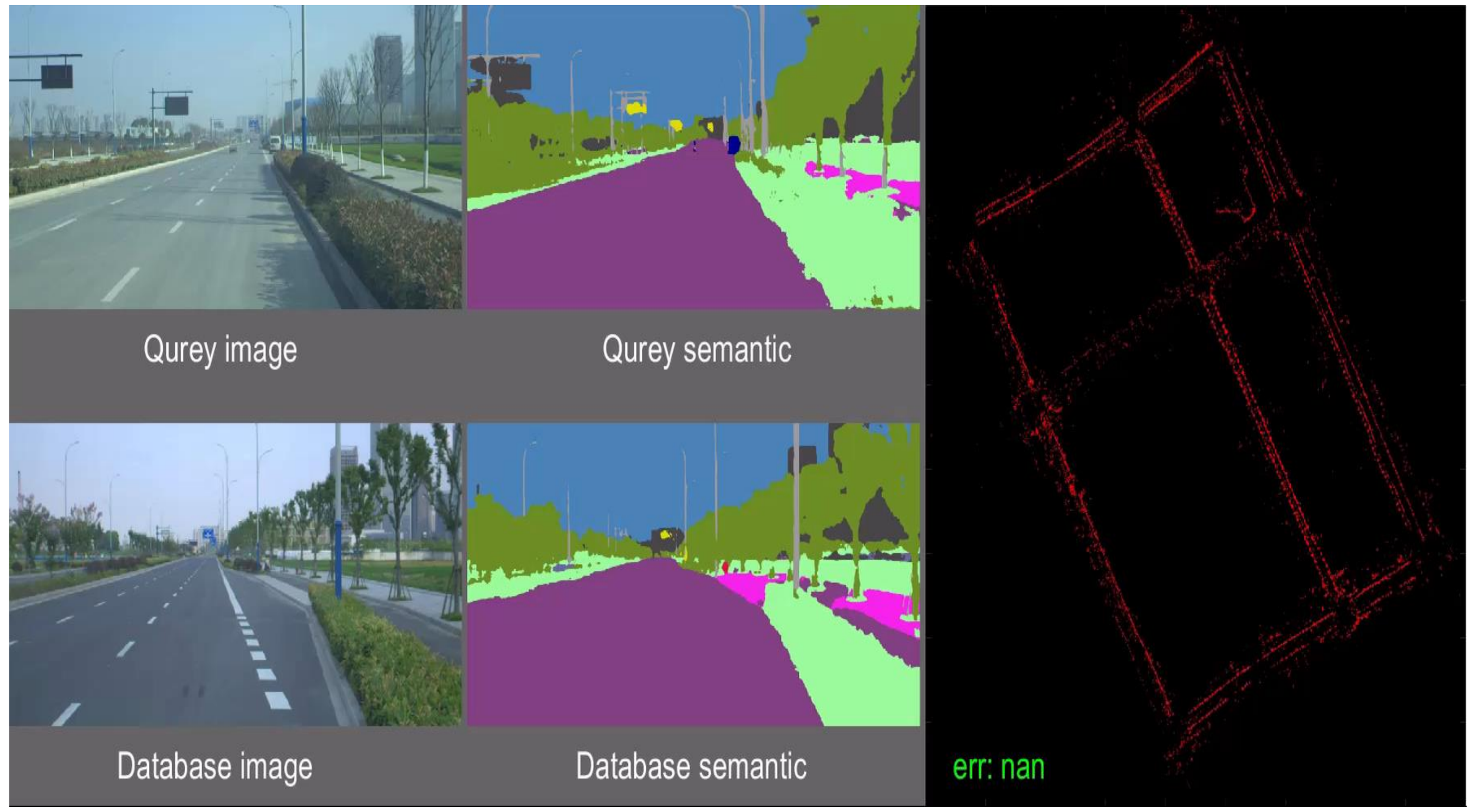


We count the number of 3D points whose labels are the same as their projections in the query image. We use the number as the semantic score of the current retrieved image. Then, we can assign each retrieved image a semantic score which equals to the consistent semantic number of projections. The retrieved database images with high semantic scores can be considered as correct retrieved images, while those with low scores means erroneous retrieved images to some extent.

Weighted RANSAC + PNP

- We put 2D-3D matches produced by all retrieved database images together to run a final PnP solver, inside a RANSAC loop.
- The 2D-3D matches produced by the same retrieved database image are assigned a same score which equals to the semantic score of the retrieved database image.
- We normalize each score by the sum of scores of all 2D-3D matches and use the normalized score as a weight p for RANSAC's sampling.

A demo for our method



References

[1] Carl Toft, Erik Stenborg, Lars Hammarstrand, Lucas Brynte, Marc Pollefeys, Torsten Sattler, and Fredrik Kahl, "Semantic match consistencyfor long-term visual localization," in European Conference on Computer Vision. Springer, 2018, pp. 391–408.

[2] Torsten Sattler, Will Maddern, Carl Toft, Akihiko Torii, Lars Hammarstrand, Erik Stenborg, Daniel Safari, Masatoshi Okutomi, Marc Pollefeys, Josef Sivic, et al., "Benchmarking 6dof outdoor visual localization in changing conditions," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 8601–8610.

VISUAL LOCALIZATION USING SPARSE SEMANTIC 3D MAP

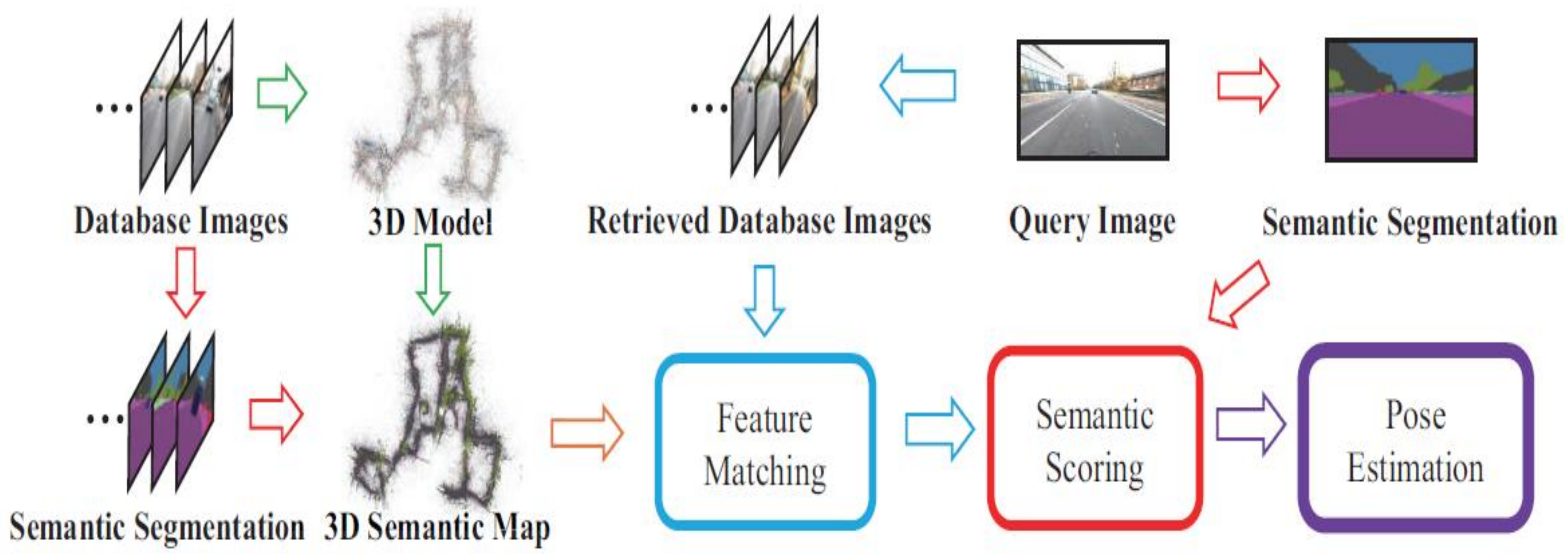
Tianxin Shi[1,2] Shuhan Shen[1,2] Xiang Gao[1,2] Lingjie Zhu[1,2]

1. NLPR, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, P. R. China 2. University of Chinese Academy of Sciences, Beijing 100049, P. R. China

Introduction

Accurate and robust visual localization under a wide range of viewing condition variations is the key component for many computer vision applications. Under these changing conditions, most traditional methods would fail to locate the camera. In this paper we present a visual localization algorithm that combines traditional methods with semantic information to improve localization results.

Method Overview



Sparse Semantic Map

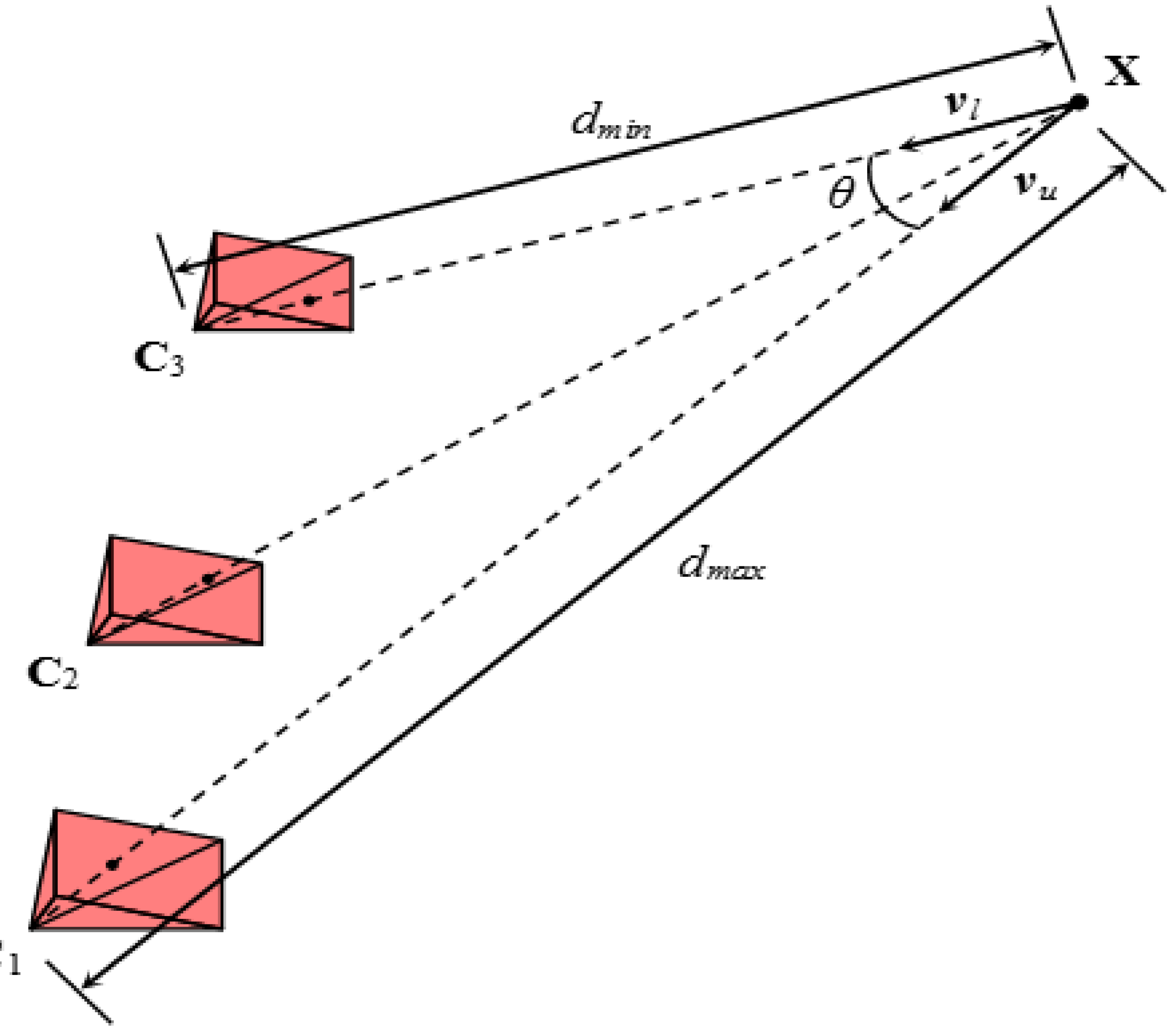
We first run a regular SfM pipeline using all database images to construct a sparse 3D model of the scene. After SfM, the location as well as the visible images of each 3D point are obtained. Given semantic segmentation for all database images, we assign each 3D point a semantic label by maximum voting with reprojection pixel labels in all its visible images.

Image Retrieval

In the image retrieval step, we use the classical image retrieval method provided by colmap which using vocabulary tree with spatial re-ranking to obtain top-k ranked database images.

Projection

We use one retrieved image at a time for feature matching with query image. Through 2D-2D matches and 2D-3D correspondences provided by SfM, we can obtain 2D-3D matches between query image and 3D map. Then, these 2D-3D matches are used to recover a temporary query pose by applying a PnP solver. Given semantic segmentation about query image, we only project the visible 3D points into query image by the estimated temporary query pose to check the number of consistent semantic points.



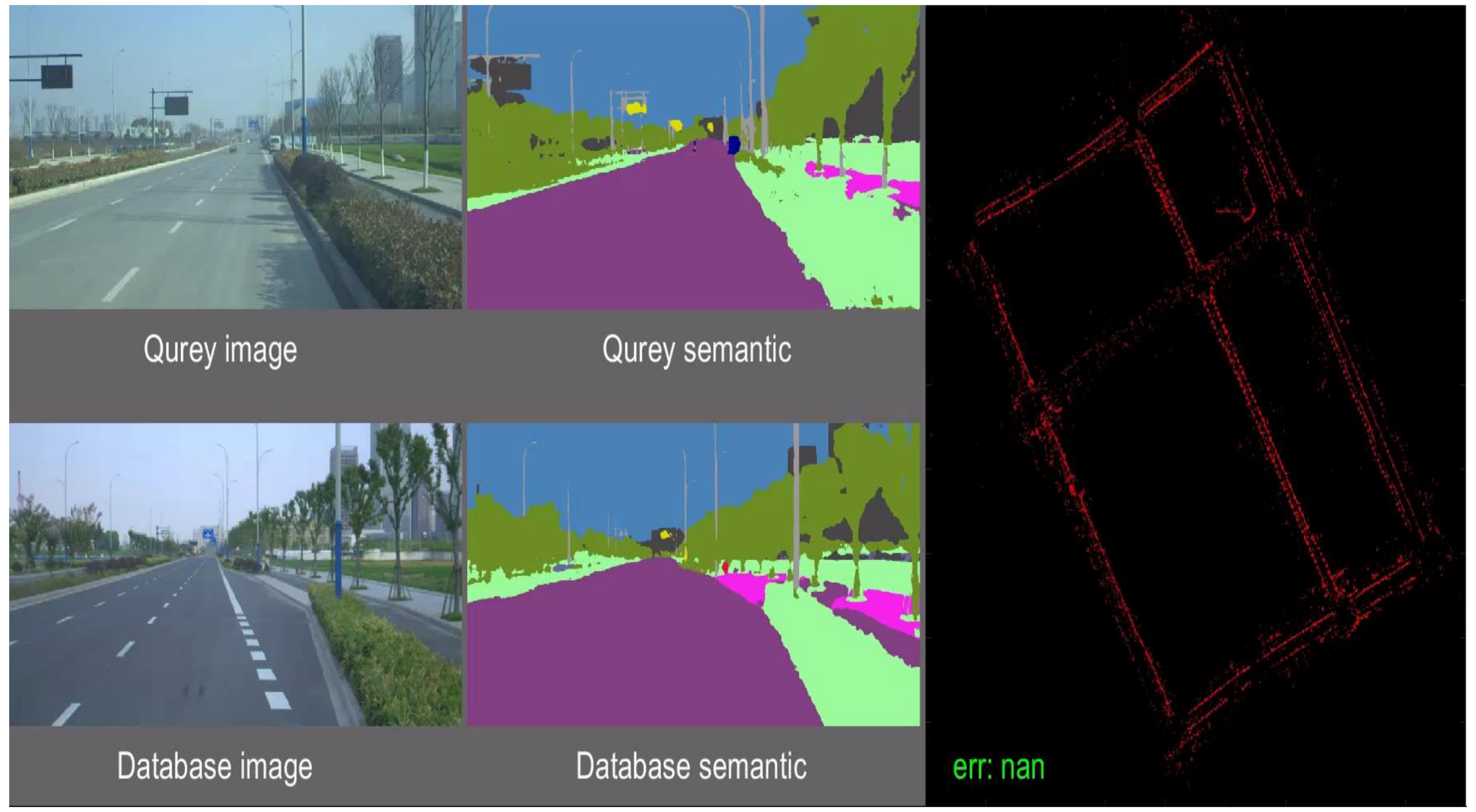
$$d_{\min} < \|v\| < d_{\max} \ , \ \angle(v, v_m) < \theta$$

$$v = C_Q - X \qquad v_m = \frac{1}{2}(v_l + v_u)$$

Weighted RANSAC + PNP

- We put 2D-3D matches produced by all retrieved database images together to run a final PnP solver, inside a RANSAC loop.
- The 2D-3D matches produced by the same retrieved database image are assigned a same score which equals to the semantic score of the retrieved database image.
- We normalize each score by the sum of scores of all 2D-3D matches and use the normalized score as a weight p for RANSAC's sampling.

A demo for our method



References

[1] Carl Toft, Erik Stenborg, Lars Hammarstrand, Lucas Brynte, Marc Pollefeys, Torsten Sattler, and Fredrik Kahl, "Semantic match consistencyfor long-term visual localization," in European Conference on Computer Vision. Springer, 2018, pp. 391–408.

[2] Torsten Sattler, Will Maddern, Carl Toft, Akihiko Torii, Lars Hammarstrand, Erik Stenborg, Daniel Safari, Masatoshi Okutomi, Marc Pollefeys, Josef Sivic, et al., "Benchmarking 6dof outdoor visual localization in changing conditions," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 8601–8610.