Large Scale Vision-Based Navigation Without an Accurate Global Reconstruction

Siniša Šegvić, Anthony Remazeilles, Albert Diosi and François Chaumette

http://www.irisa.fr/lagadic, IRISA/INRIA Rennes, France

IRISA lagadic

CVPR 2007

Introduction

- We present a novel vision framework for scalable mapping and localization, enabling fault-tollerant feature-oriented **appearance-based navigation** in outdoor environments:
  - the mapping and navigation stages are considered separately, as an interesting and not completely solved problem;
  - the employed hierarchical environment representation has a graph of key-images at the top (**scalability**), and local 3D reconstructions at the bottom (**feature prediction**);
  - both mapping and localization framework components rely on a differential tracker featuring warp with isotropic scaling and affine contrast compensation [4].

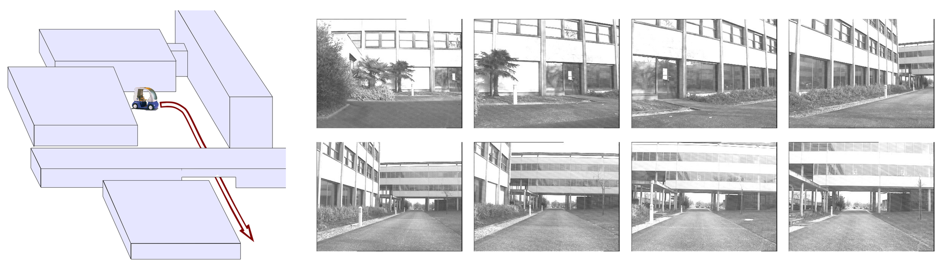

Fig. intro1. Appearance-based navigation: the navigation task (left), and the first eight key-images (right).

- The topological representations have been used by the robotic community, in the context of range sensors; we introduce this idea to computer vision, together with [2, 3].
- Although related to [2, 3], our work is novel since it better exploits the power of multi-view geometry techniques.
- We share the problem with [1], but employ different techniques and do not require global consistency:
  - by posing weaker requirements we obtain scalable real-time mapping and can close loops regardless of the drift;
  - experiments with **15000 landmarks** have been performed without any performance degradation.

Mapping

- The mapping component extracts point features from the learning sequence, constructs the environment graph and annotates its nodes and arcs with geometric information.

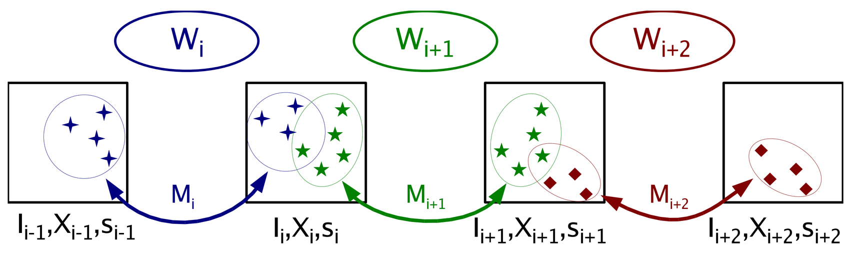

Fig. map1. The linear environment graph. Nodes contain images I_i, features X_i and scale factors s_i. Arcs contain match arrays M_i and the two-view geometries W_i.

- The devised automatic solution uses the tracker to find the stablest features in the current subrange of the sequence:
  - the features are tracked until the reconstruction error becomes high or the number of tracked points becomes low;
  - then the current frame is discarded, while the previous frame is registered as the new node of the graph.

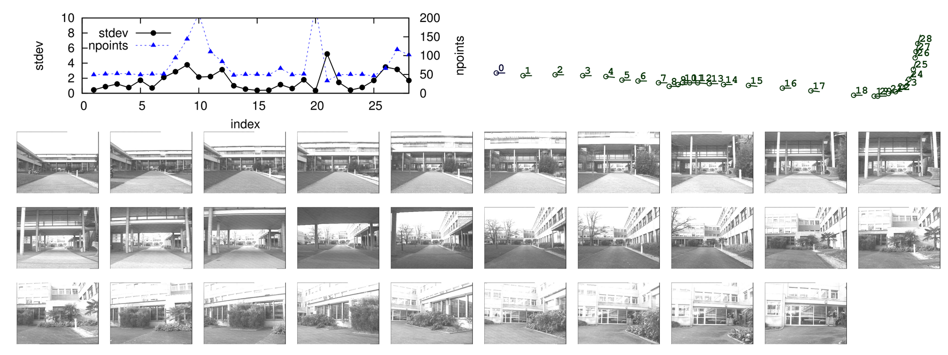

Fig. map2. Mapping results (1900 images along 150 m): feature counts $|M_i|$ and reprojection errors $\sigma(W_i)$ (left), reconstructed camera poses (right), 29 key-images (bottom).

- Disjoint parts of the env. graph (circular sequences, general tracking failures) are connected by wide-baseline matching.

Localization

- **Initialization**: wide-baseline matching the current image with the two key-images incident to the actual arc:
  - the correspondences are used to recover the three-view geometry (3vg), and locate the mapped features.

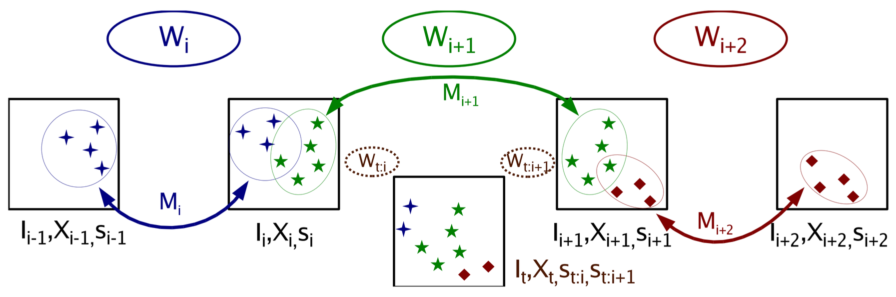

Fig. loc1. Three groups of features considered for tracking.

- **Feature prediction**: the tracked features are used to estimate two-view geometries $W_{t:i}(I_i, I_t)$ and $W_{t:i+1}(I_{i+1}, I_t)$:
  - from these we recover 3vgs (I_t, I_{i+1}, I_{i+2}) and (I_t, I_i, I_{i+1}) by a decomposed approach.

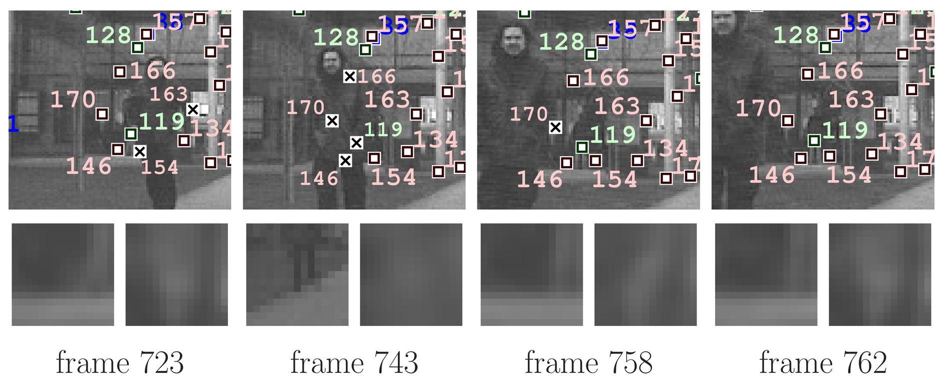

Fig. loc2. Feature prediction in action. Tracked features and rejected projections are designated with squares and crosses. The bottom row shows the references and optimized warps for the features 146 (left) and 170 (right).

- **Maintaining the topological location**: very important since the success of tracking and robot control relies on it:
  - the employed criterion: $\langle -\mathbf{R_{i+1}}^\top \cdot \mathbf{t_{i+1}}, \mathbf{t_{t:i+1}} \rangle < 0$.

Mapping and localization on a circular sequence

- We consider the learning sequence `loop-clouds`, taken along a *circular* path of approximately 50 m.
- Fig. loop1 illustrates the mapping sensitivity w.r.t. parameters: (i) minimum count of features n, (ii) maximum allowed reprojection error σ, and (iii) the tracking RMS threshold R:
  - the presence of node 0' indicates that the cycle at the topological level has been closed by wide-baseline matching;
  - the right-most map in Fig. loop1 was deliberately constructed using suboptimal parameters: our navigation approach works regardless of the accumulated drift.
- Fig. loop2(left) illustrates the capability of the localization component to traverse a topological cycle:
  - the navigation sequence was obtained during two rounds roughly along the same circular physical path;
  - the first round was used for mapping (`loop-clouds`, cf. Fig. loop1) while the localization is performed along the combined sequence, involving two complete rounds;
  - the localization was successful in both rounds.
- Fig. loop2(right) shows results on sequence `loop-sunlight` acquired along a similar circular path in bright sunlight:
  - despite the different imaging conditions (cf. Fig. loop3), the localization was successful except in arcs 10, 11 and 12;
  - the illumination difficulties were aggravated by a featureless tree and a considerable curvature of the learning path;
  - Fig. loop3 shows the localization after the reinitialization in arc 13, where the buildings behind the tree begin to appear.

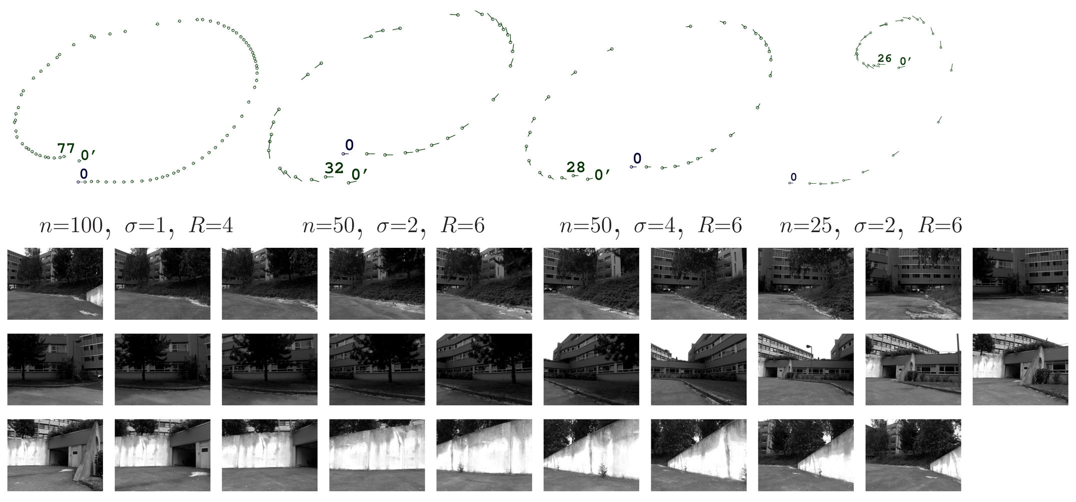

Fig. loop1. Mapping results on the circular sequence `loop-clouds`.

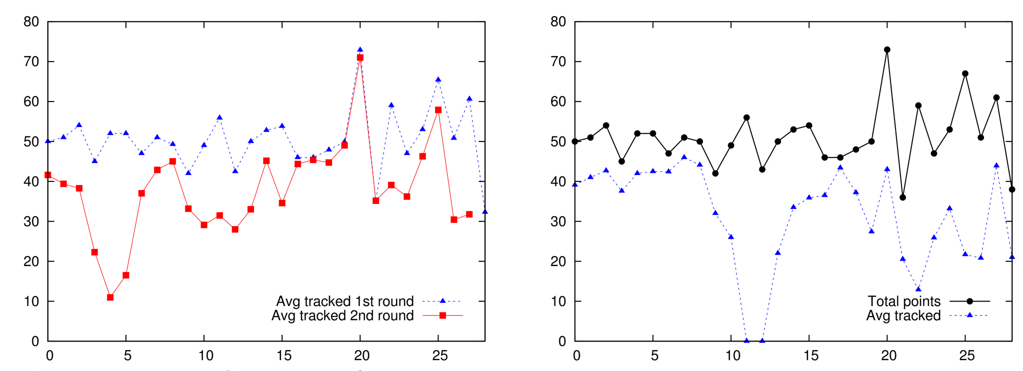

Fig. loop2. Counts of tracked points while localizing on `loop-clouds` (2 rounds, left) and `loop-sunlight` (1 round, right).

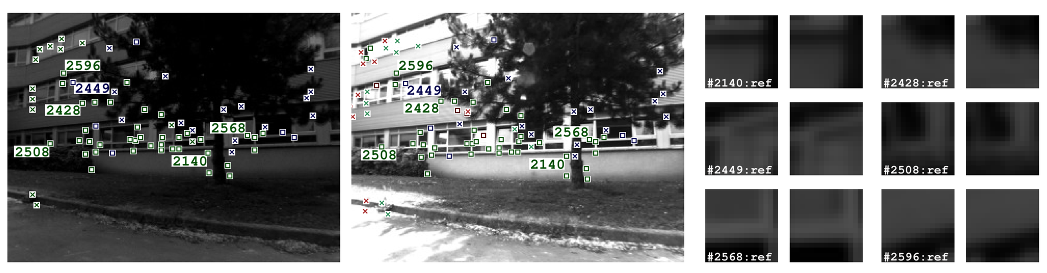

Fig. loop3. Tracking features from key-image `loop-clouds`:12 (left) in `loop-sunlight`:#509 (middle). Enlarged features are on the right.

Conclusions

- The poster presents a novel framework for scalable feature-oriented appearance-based navigation.
- The framework combines 3D prediction with 2D navigation:
  - the global topological map ensures unlimited scalability;
  - the local geometric representation enables recovery from tracking failures through feature prediction;
  - the desired feature positions in the next key-image allow simple and effective 2D navigation by visual servoing.
- The framework allows large-scale navigation without requiring a geometrically consistent global view of the environment:
  - in the experiment with a circular path, the navigation proceeds regardless of the extent of the drift;
  - the framework is applicable in interconnected environments, where global consistency may be difficult to enforce.
- Very encouraging results in realistic experiments:
  - real-time navigation in public areas with other moving objects and moderate-to-large changes in imaging conditions;
  - difficulties: illumination variations, featureless areas, sharp turns, nearby vegetation.
- The main performance bottleneck is CPU power:
  - most of the processing time is spent within the multi-resolution point feature tracker [4];
  - the mapping and localization throughput on a notebook with CPU equivalent to P4@2Ghz is 5 Hz and 7 Hz (cca 1 m/s) on 320 × 240 gray–level images;
  - a working system using only small images suggests that vision-based autonomous transportation is getting close.

Real-time navigation results

- The robot control has been imlemented by a simple visual servoing scheme; the steering angle ψ is determined as: $\psi = -\lambda(\overline{x}_t - \overline{x}^*)$, where $\lambda \in \mathcal{R}^+$.
- We show an experiment on the map shown in Fig.nav1, offering a variety of driving conditions: narrow sections, slopes and driving under a building (see [5] for more experiments):
  - the speed was set to 30 cm/s in turns, otherwise 80 cm/s,
  - five reinitializatioins were required, as shown in Fig. nav1,
  - between A and B the robot drove over 740 m despite the occlusions such as is shown in Fig. nav2.

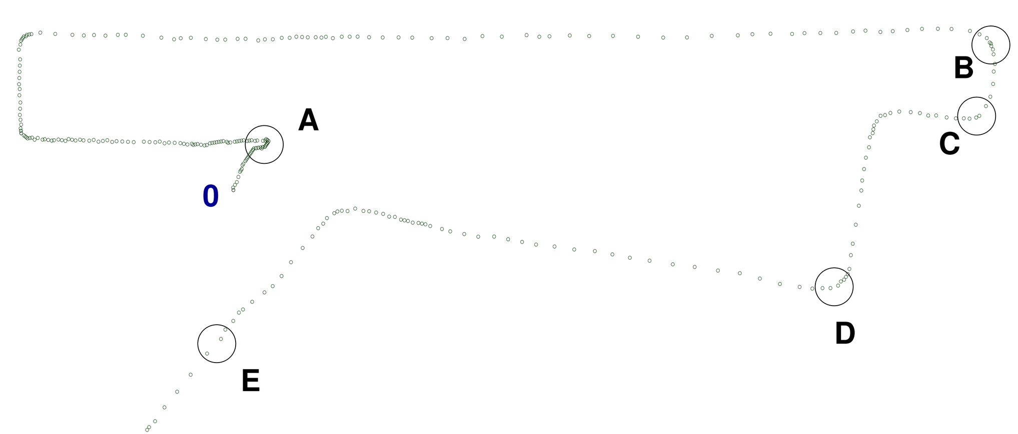

Fig. nav1. Graph of 320 nodes mapping a 1.1 km reference path.

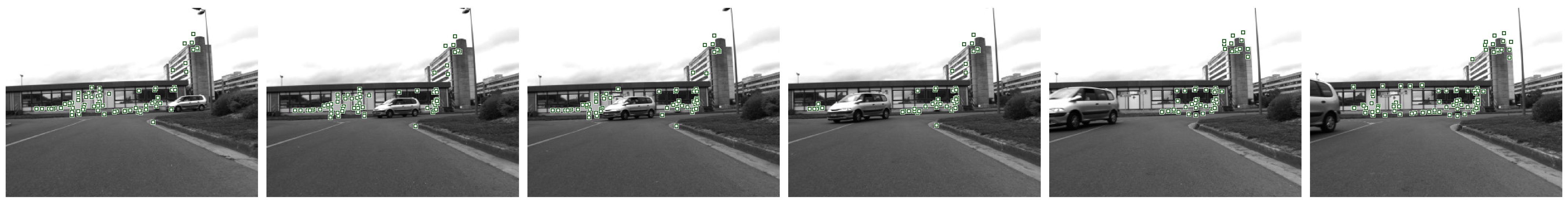

Fig. nav2. Navigation along the map from Fig. nav1. The features re-appear after being occluded and disoccluded by a moving car.

Selected references

[1] Royer, E. et al: Localization in urban environments: Monocular vision compared to a differential GPS sensor. In: Proc. of CVPR. Volume 2., Washington, DC (2005) 114–121

[2] Chen, Z., Birchfield, S.T.: Qualitative vision-based mobile robot navigation. In: Proc. of ICRA, Orlando, Florida (2006)

[3] Goedemé, T. et al: Omnidirectional vision based topological navigation. Int. J. Comput. Vis. (2007) to appear.

[4] Šegvić, S., Remazeilles, A., Chaumette, F.: Enhancing the point feature tracker by adaptive modelling of the feature support. In: Proc. of ECCV, Graz, Austria (2006) 112–124

[5] Diósi, A et al: Experimental evaluation of an urban visual path following framework. In: Proc. of IFAC Symposium on IAV, Toulouse, France (2007)

1. This work has been supported by the French national project Predit Mobivip, by the project Robea Bodega, and by the European MC IIF project AViCMaL.

2. A video with more recent results is available at: `http://www.irisa.fr/lagadic/video/CycabNavigation.mov`

3. This poster was prepared with LaTeX, using `a0poster.cls` and `pstricks.sty`