ACIVS ’07

A framework for scalable vision-only navigation

Siniša Šegvić, Anthony Remazeilles, Albert Diosi and François Chaumette

INRIA/IRISA, Campus de Beaulieu, F-35042 Rennes Cedex, France

CONTRIBUTION

- a novel approach to **large-scale** **vision-only** navigation

CONTRIBUTION

- a novel approach to **large-scale** **vision-only** navigation
- *hybrid* (topological-geometric) environment representation:
  1. environment partitioned into manageable subworlds
  2. the topology induced by subworld adjacency
  3. subworlds described by metric 3D reconstructions

CONTRIBUTION

- a novel approach to **large-scale** **vision-only** navigation
- *hybrid* (topological-geometric) environment representation:
  1. environment partitioned into manageable subworlds
  2. the topology induced by subworld adjacency
  3. subworlds described by metric 3D reconstructions
- "*use 3D vision as little as possible, but no less than that!*"
- **benefits**: scalability, occlusion tollerance, robust control

CONTRIBUTION

- a novel approach to **large-scale** **vision-only** navigation
- *hybrid* (topological-geometric) environment representation:
  1. environment partitioned into manageable subworlds
  2. the topology induced by subworld adjacency
  3. subworlds described by metric 3D reconstructions
- "*use 3D vision as little as possible, but no less than that!*"
- **benefits**: scalability, occlusion tollerance, robust control
- evaluated in *many* outdoor **experiments** (>100m)
  (automatic **real-time** mapping, autonomous path following)

AGENDA

- Introduction
- The proposed navigation framework
  (the vision, current implementation)
- The main components of the framework
  (mapping, localization, navigation)
- Experiments (3 minute video)
- Conclusion

INTRODUCTION

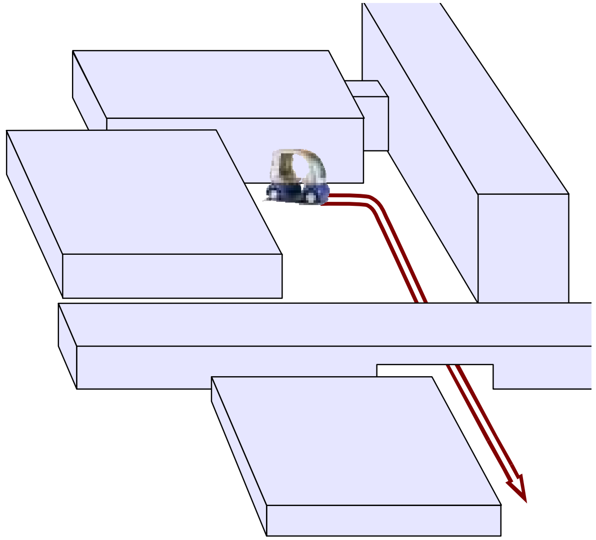
The **environment representation** (central issue in navigation):

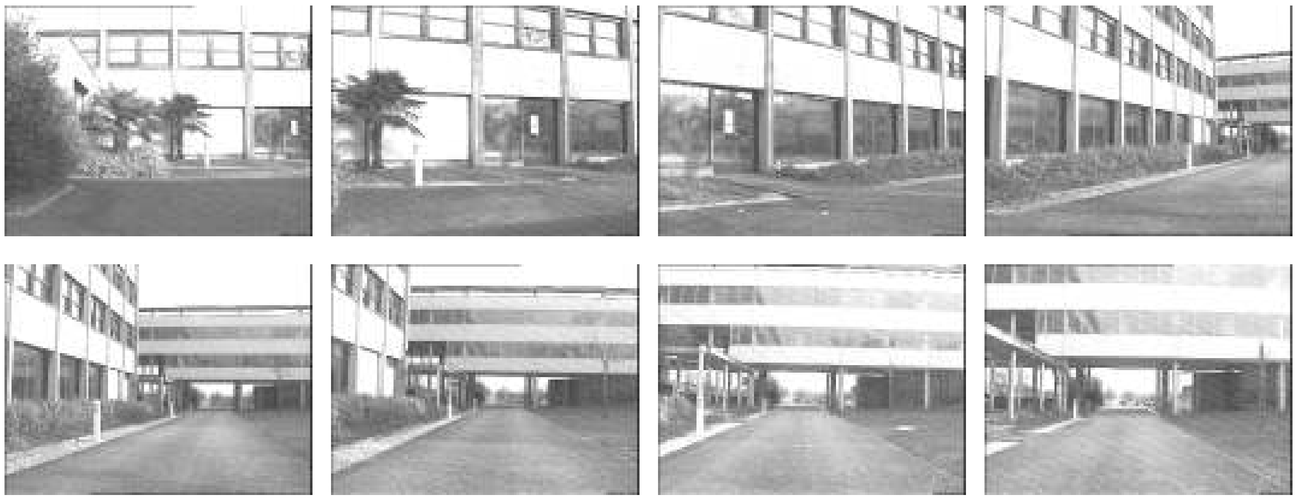
- the classic model-based approach:
  environment-centred (globally consistent geometry)
- the appearance-based alternative:
  sensor-centred (graph of key-images)
  - navigation with **relaxed** requirements (scalability!)
  - **simple** relation perception $\longrightarrow$ action (robust control!)
  - esp. for structured environments, constrained motion

Introduction

The **environment representation** (central issue in navigation):

- the classic model-based approach:
  environment-centred (globally consistent geometry)
- the appearance-based alternative:
  sensor-centred (graph of key-images)
  - navigation with **relaxed** requirements (scalability!)
  - **simple** relation perception → action (robust control!)
  - esp. for structured environments, constrained motion

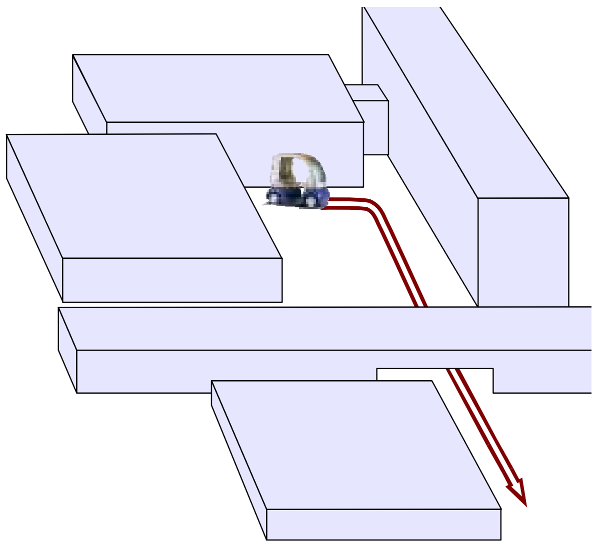




INTRODUCTION

The **environment representation** (central issue in navigation):

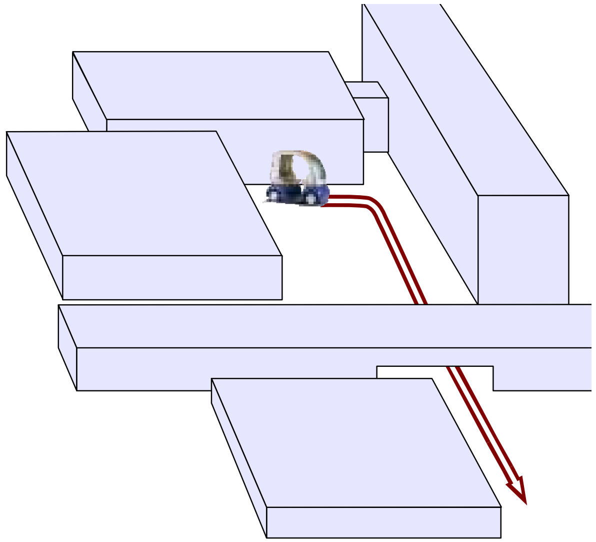
- the classic model-based approach:
  environment-centred (globally consistent geometry)
- the appearance-based alternative:
  sensor-centred (graph of key-images)
  - navigation with **relaxed** requirements (scalability!)
  - **simple** relation perception $\rightarrow$ action (robust control!)
  - esp. for structured environments, constrained motion



INTRODUCTION

The **environment representation** (central issue in navigation):

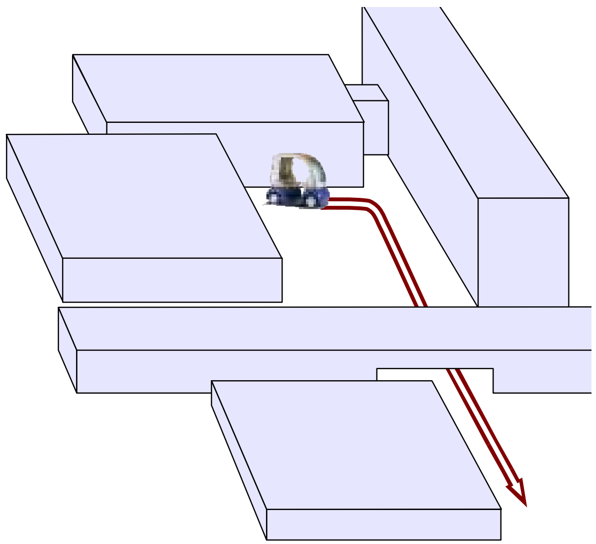
- the classic model-based approach:
  environment-centred (globally consistent geometry)
- the appearance-based alternative:
  sensor-centred (graph of key-images)
  - navigation with **relaxed** requirements (scalability!)
  - **simple** relation perception → action (robust control!)
  - esp. for structured environments, constrained motion



INTRODUCTION

The **environment representation** (central issue in navigation):

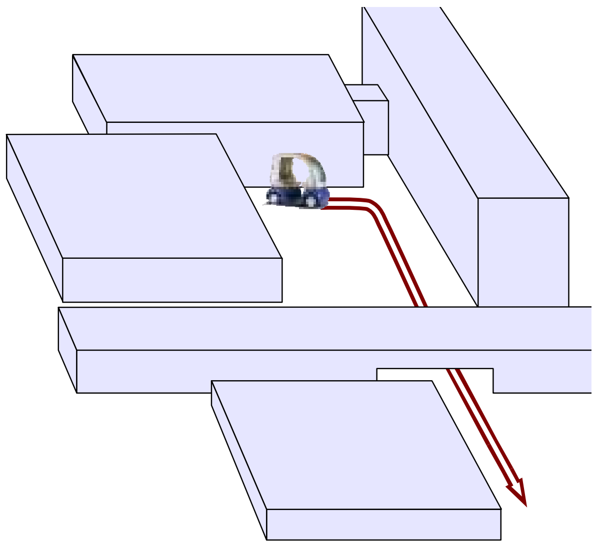
- the classic model-based approach:
  environment-centred (globally consistent geometry)
- the appearance-based alternative:
  sensor-centred (graph of key-images)
  - navigation with **relaxed** requirements (scalability!)
  - **simple** relation perception $\rightarrow$ action (robust control!)
  - esp. for structured environments, constrained motion



INTRODUCTION

The **environment representation** (central issue in navigation):

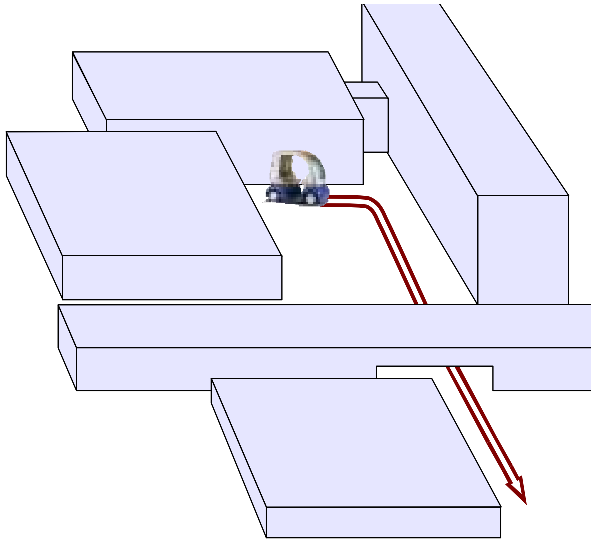
- the classic model-based approach:
  environment-centred (globally consistent geometry)
- the appearance-based alternative:
  sensor-centred (graph of key-images)
  - navigation with **relaxed** requirements (scalability!)
  - **simple** relation perception $\rightarrow$ action (robust control!)
  - esp. for structured environments, constrained motion



INTRODUCTION

The **environment representation** (central issue in navigation):

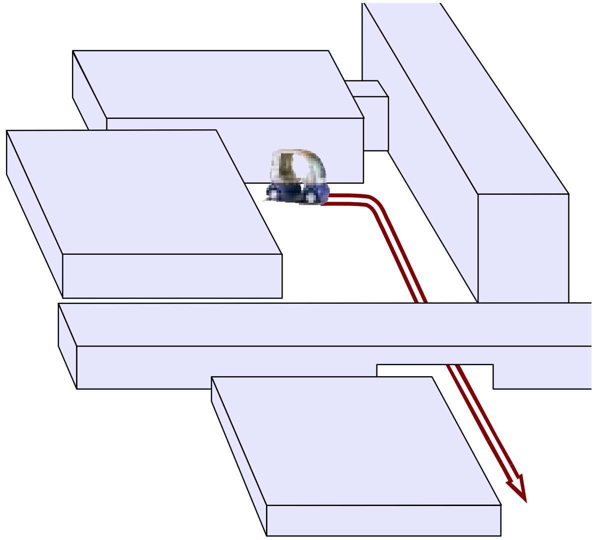
- the classic model-based approach:
  environment-centred (globally consistent geometry)
- the appearance-based alternative:
  sensor-centred (graph of key-images)
  - navigation with **relaxed** requirements (scalability!)
  - **simple** relation perception $\rightarrow$ action (robust control!)
  - esp. for structured environments, constrained motion

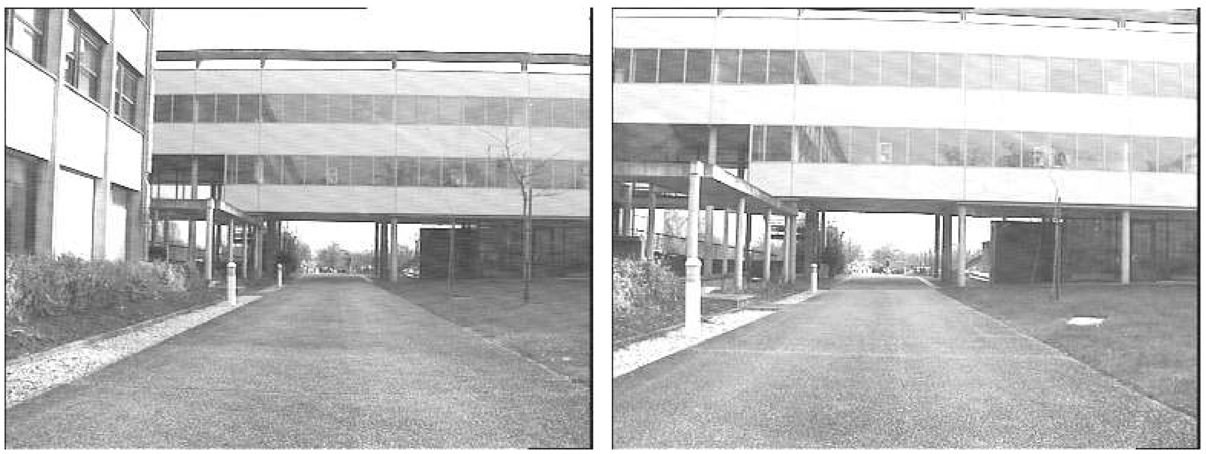


INTRODUCTION

The **environment representation** (central issue in navigation):

- the classic model-based approach:
  environment-centred (globally consistent geometry)
- the appearance-based alternative:
  sensor-centred (graph of key-images)
  - navigation with **relaxed** requirements (scalability!)
  - **simple** relation perception $\rightarrow$ action (robust control!)
  - esp. for structured environments, constrained motion

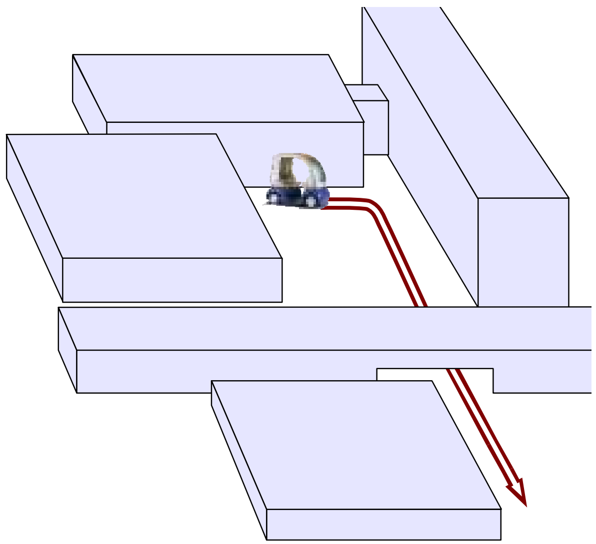




INTRODUCTION

The **environment representation** (central issue in navigation):

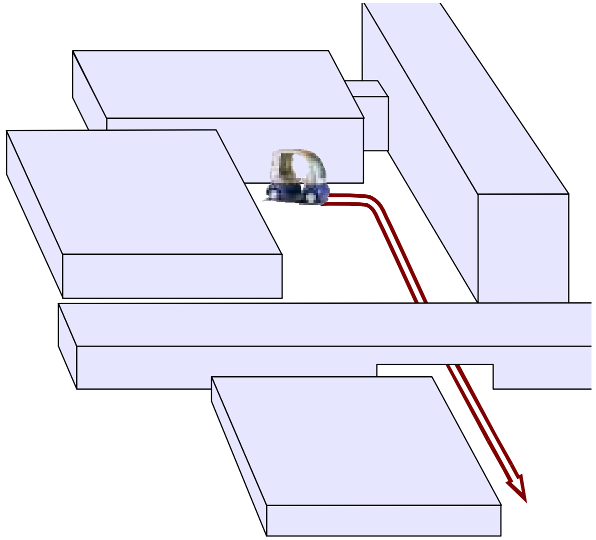
- the classic model-based approach:
  environment-centred (globally consistent geometry)
- the appearance-based alternative:
  sensor-centred (graph of key-images)
  - navigation with **relaxed** requirements (scalability!)
  - **simple** relation perception → action (robust control!)
  - esp. for structured environments, constrained motion

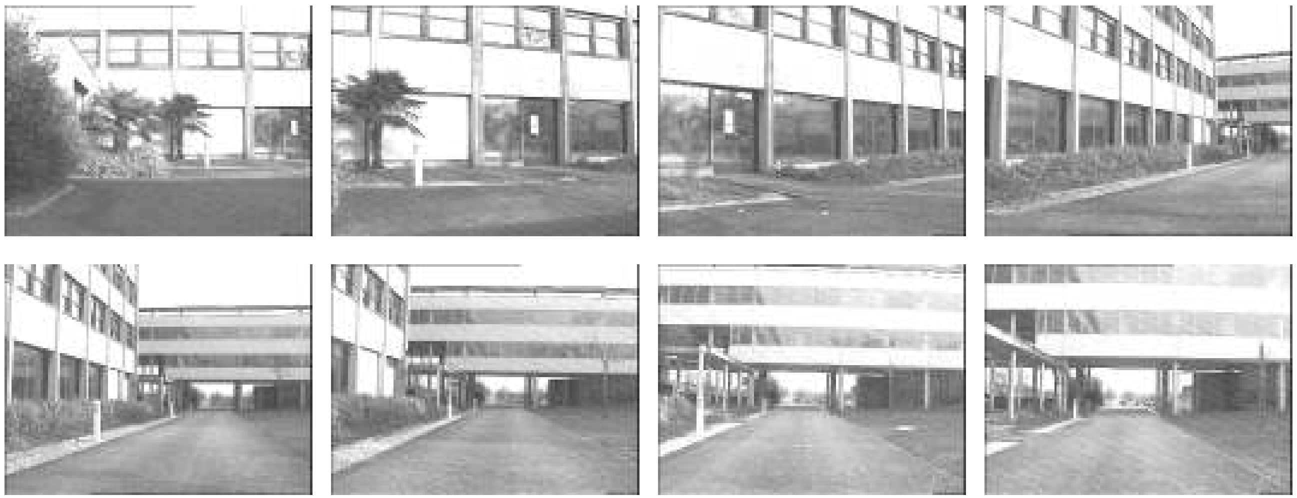


INTRODUCTION

The **environment representation** (central issue in navigation):

- the classic model-based approach:
  environment-centred (globally consistent geometry)
- the appearance-based alternative:
  sensor-centred (graph of key-images)
  - navigation with **relaxed** requirements (scalability!)
  - **simple** relation perception → action (robust control!)
  - esp. for structured environments, constrained motion





INTRODUCTION (2)

More on the employed **appearance-based** scheme:

- feature-oriented: relies on **correspondences** between neighbouring key-images
  - Harris point features, **custom** differential tracker

INTRODUCTION (2)

More on the employed **appearance-based** scheme:

- feature-oriented: relies on **correspondences** between neighbouring key-images
  - Harris point features, **custom** differential tracker
- the map constructed from a learning sequence

INTRODUCTION (2)

More on the employed **appearance-based** scheme:

- feature-oriented: relies on **correspondences** between neighbouring key-images
  - Harris point features, **custom** differential tracker
- the map constructed from a learning sequence
- **mapping** and **navigation** heavily rely on the tracker: **few** association errors, **fast** implementation

Introduction (2)

More on the employed **appearance-based** scheme:

- feature-oriented: relies on **correspondences** between neighbouring key-images
  - Harris point features, **custom** differential tracker
- the map constructed from a learning sequence
- **mapping** and **navigation** heavily rely on the tracker: **few** association errors, **fast** implementation
- augment appearance-based map with local 3D geometries to enable feature prediction (**point transfer**):
  - **recovery** from tracking errors (occlusion tollerance!)
  - **smooth** introduction of new features

INTRODUCTION (2)

More on the employed **appearance-based** scheme:

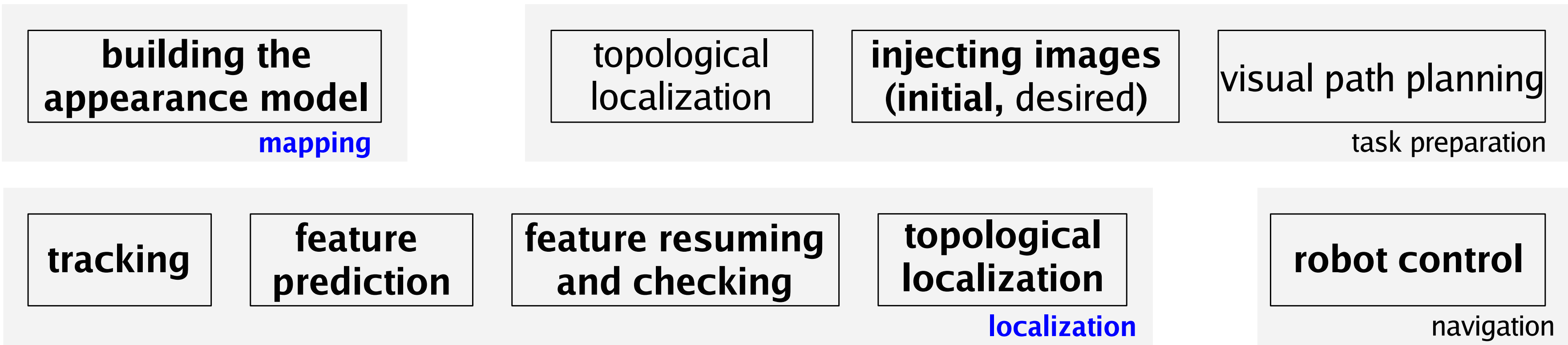
- feature-oriented: relies on **correspondences** between neighbouring key-images
  - Harris point features, **custom** differential tracker
- the map constructed from a learning sequence
- **mapping** and **navigation** heavily rely on the tracker: **few** association errors, **fast** implementation
- augment appearance-based map with local 3D geometries to enable feature prediction (**point transfer**):
  - **recovery** from tracking errors (occlusion tollerance!)
  - **smooth** introduction of new features
- vectors between the **current** and the **desired** feature positions used to steer the robot (visual servoing)

The navigation framework

The envisioned application: **robotic taxi** in pedestrian zones

The main **four** components:

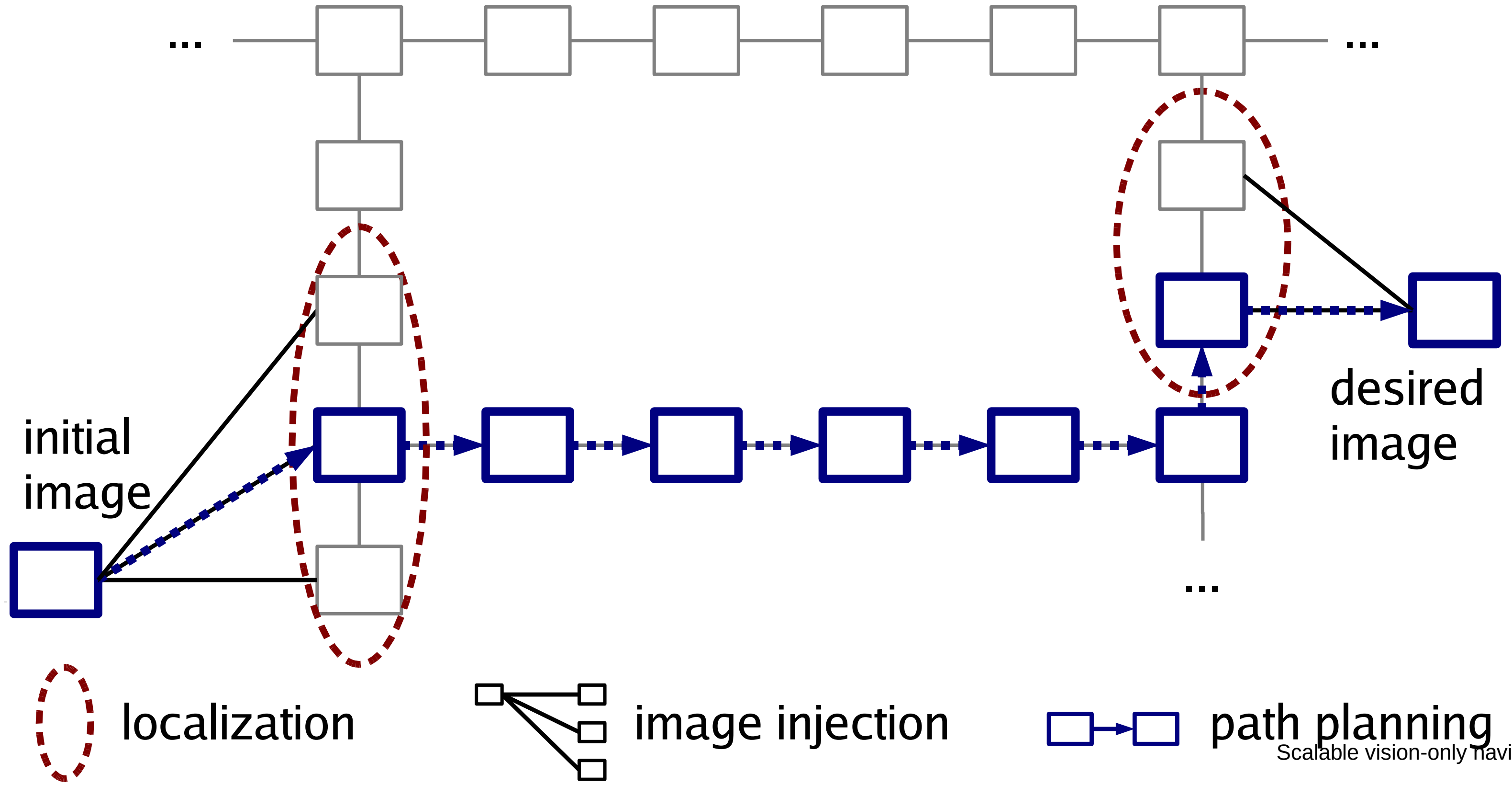
- *mapping* (performed off-line, entire environment)
- *task preparation* (input: the next navigation task)
- *localization* (coarse topological, fine-grained through tracking)
- *navigation* (robot control, visual servoing)



THE NAVIGATION FRAMEWORK (2)

Task preparation produces the linear graph used for navigation

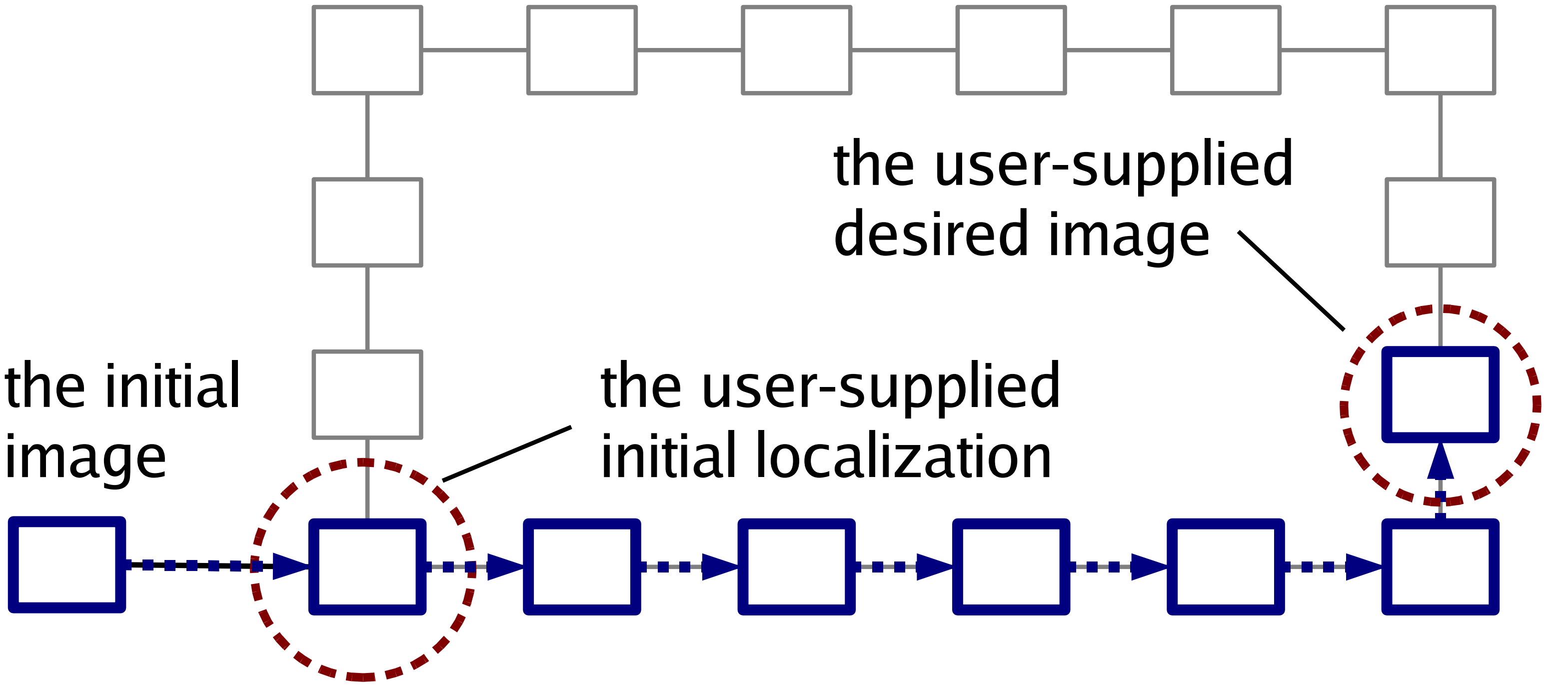
- global topological **localization** (CBIR, user-supplied)
- **injecting** the two images into the topology (w-b matching)
- **path planning** (shortest path in the environment graph)



THE NAVIGATION FRAMEWORK (3)

The simplifications in the actually implemented system:

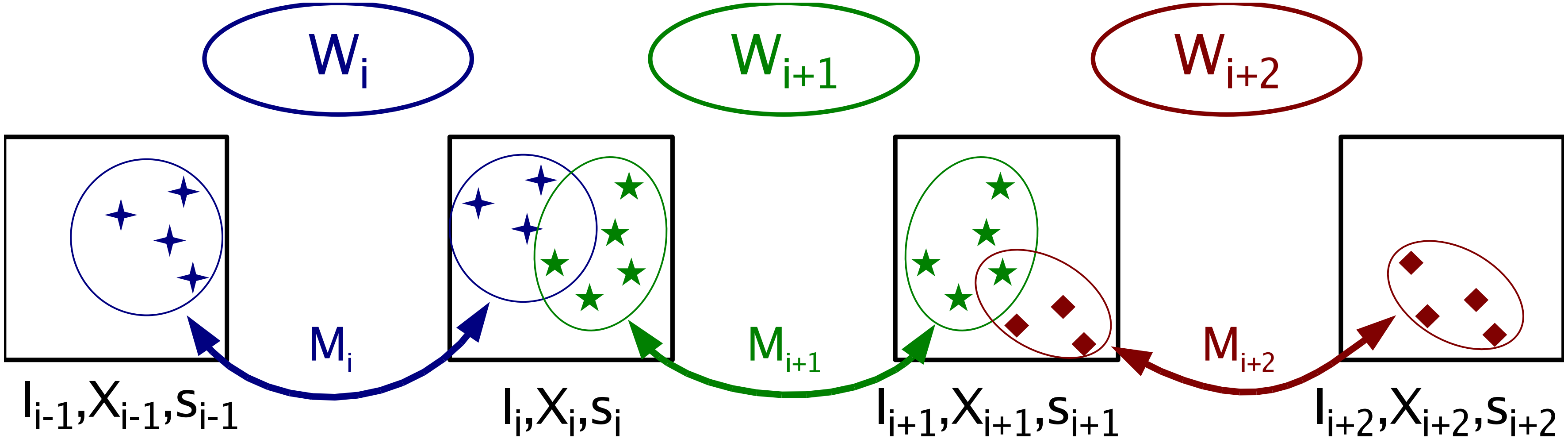
- **no branches** in the topology
- user supplied **initial** topological location
  (nevertheless, the injection of the inital image is required)
- the **goal** corresponds to a key-image



Mapping

The task: construct the **map** from the **learning sequence** $\{I_j^{LS}\}$

- form the environment graph: extract key-images $\{I_i\}$
- find correspondences M_i between (I_i, I_{i+1}) $\forall$ i
- precompute two-view geometries $W_i(I_{i-1}, I_i)$, $s_i(I_{i-1}, I_i)$



- nodes contain key-images I_i, features X_i and scale factors s_i
- arcs contain match arrays M_i and two-view geometries W_i

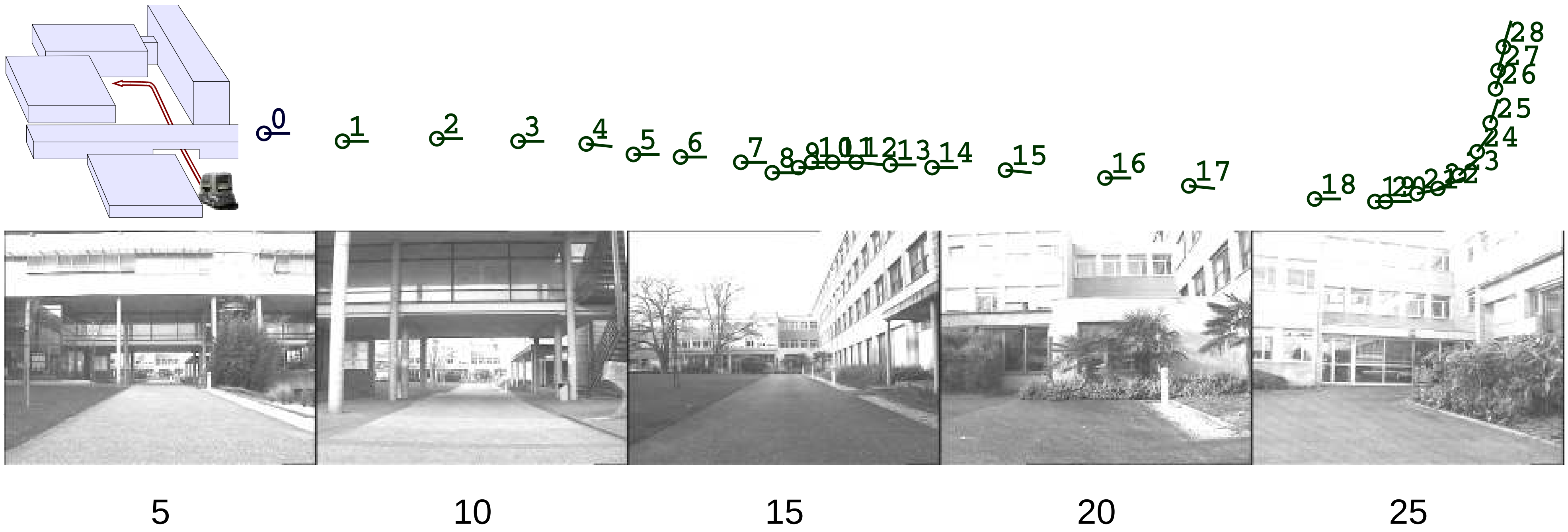
MAPPING (2)

Mapping **iteration** starts by locating prominent features in I_0^{LS}

The features tracked **until** either:

- reprojection error σ(W(I_0^{LS},I_j^{LS})) too high (quality), **or**
- number of tracked features too low (robustness)

Then: $I_{j-1}^{LS} \rightarrow$ the new key-image, **continue** from I_{j-1}^{LS}



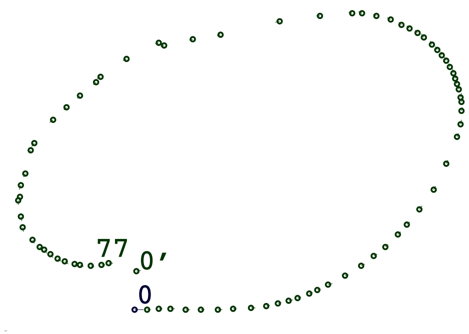
MAPPING (3)

Note: disjoint graph parts often can be merged by w-b matching:

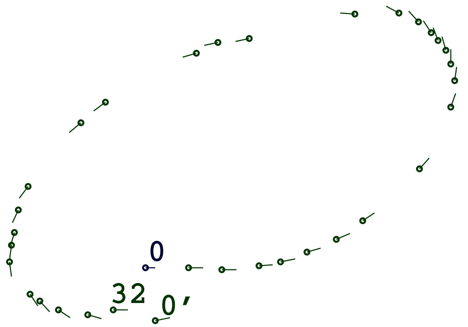
- to complete the mapping of a circular sequence
- to recover from a general tracking failure

Mapping results on a **circular** sequence, with different (n, σ, R):

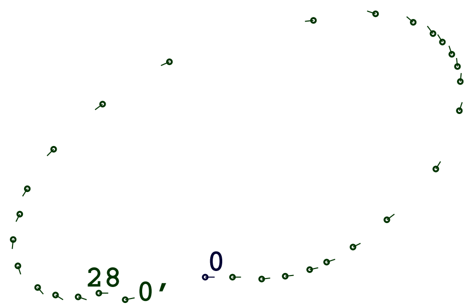
- node 0' indicates that the **cycle** has been successfully closed
- the rightmost map constructed with suboptimal parameters: the approach works regardless of the accumulated **drift**



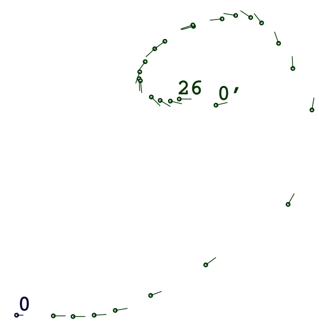
n=100, σ=1, R=4



n=50, σ=2, R=6



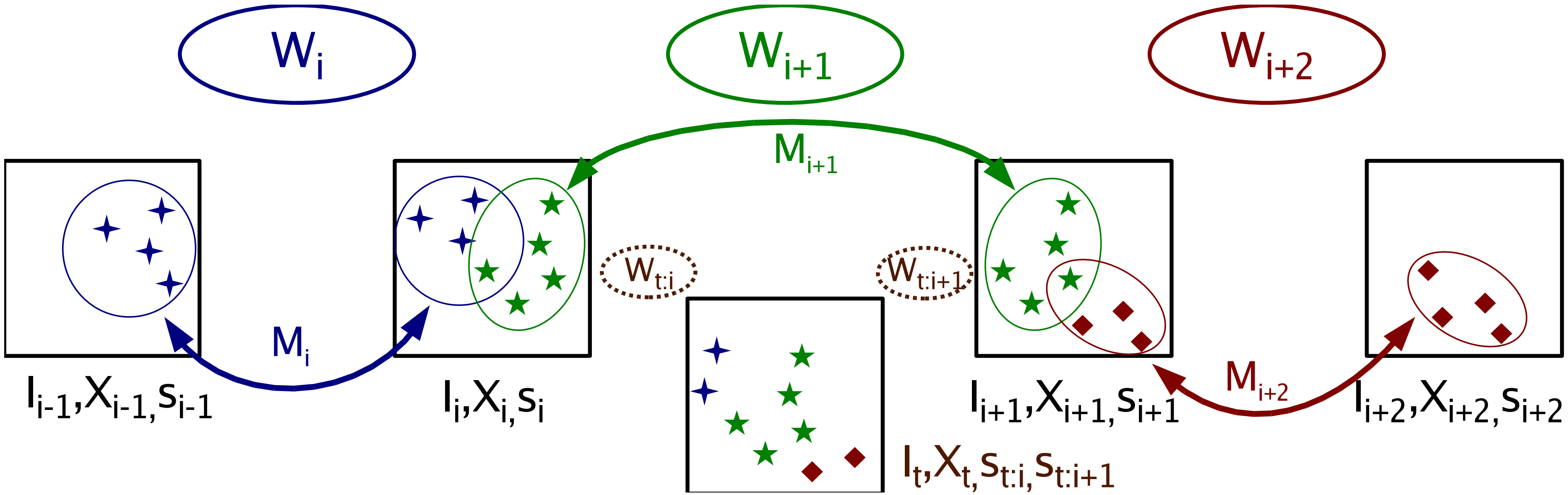
n=50, σ=4, R=6



n=25, σ=2, R=6

Localization

- features from the closest three arcs are **tracked**

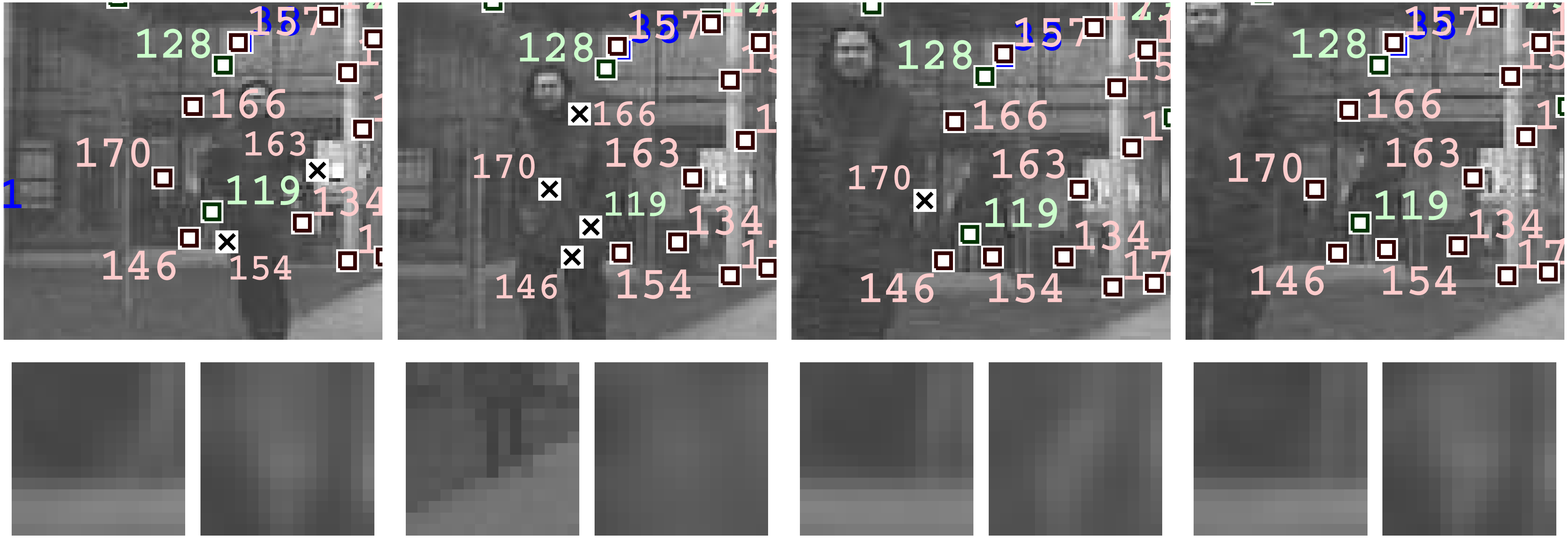


- from tracked features estimate $W_{t:i}(I_i, I_t)$ and $W_{t:i+1}(I_{i+1}, I_t)$
  - ... recover 3vgs (I_t, I_{i+1}, I_{i+2}) and (I_t, I_i, I_{i+1})
  - ... and **predict** occluded and previously invisible features
- the criterion for **maintaining** the topological location:

$\langle -\mathbf{R_{i+1}}^\top \cdot \mathbf{t_{i+1}}, \mathbf{t_{t:i+1}} \rangle < 0.$

LOCALIZATION (2)

Feature prediction and tracking resumption are **critical** ingredients of the localization component



frame 723 | frame 743 | frame 758 | frame 762

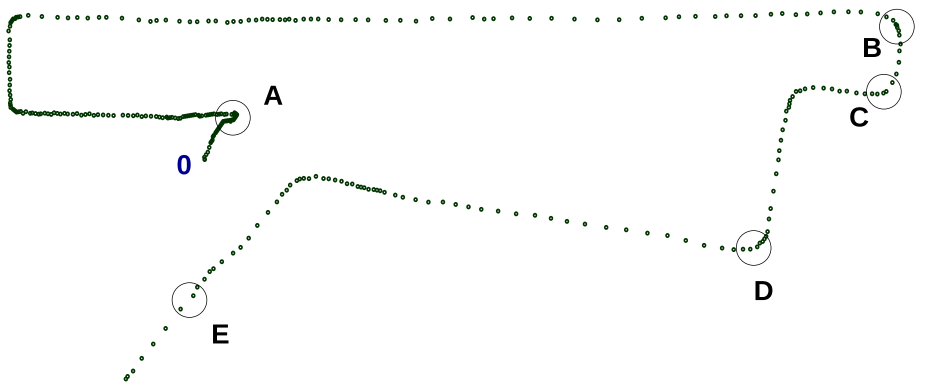
- top: current image, tracked features □, rejected projections ×
- bottom: optimized warps for #146 and #170

NAVIGATION

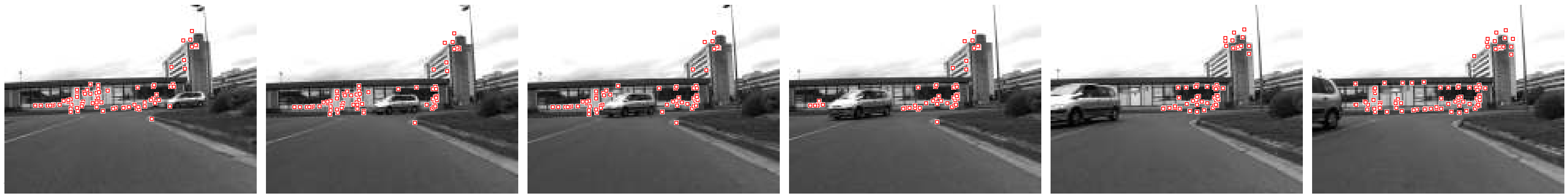
The steering angle ψ determined as: $\psi = -\lambda\left(\overline{x}_t - \overline{x}^*\right), \lambda \in \mathcal{R}^+$

A **large-scale** experiment:

- 320 key-images along a 1.1km learning path
- narrow sections, slopes, driving under a building



- the robot smoothly drove 740 m despite the clutter (A-B)
- five reinitializations in total were required (A-E)



EXPERIMENTS

- a short video
- for more experiments please see articles presented at IAV07 and IROS07 (Diosi et al.)

DISCUSSION

The main **points**:

- scalable navigation with a single perspective camera
- reliable real-time performance on public driveways
- the employed techniques:
  - differential tracking (mapping, localization)
  - calibrated point transfer (feature prediction)
  - wide-baseline matching (injecting the inital image)




Discussion (2)

Conclusions:

- **scalability** achieved by relaxing the global geometric consistency requirement
- **tollerance** to clutter and moderate illumination changes
- **real-time** mapping and navigation (5Hz, 2 m/s) on a notebook computer (2Ghz)

Future work:

- improve performance in sharp turns by active vision
- extend the mapper to improve from previous experience
- deal with more complex topologies

Thank you!

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