Body-Relative Navigation Guidance using Uncalibrated Cameras

Olivier Koch

PhD Thesis Defense

21-Jan-2010

Thesis Committee: Prof. Bill Freeman, Prof. Rob Miller, Prof. Seth Teller
Problem Statement

Human-oriented navigation in GPS-denied environments

- Finding your way in unknown environments with no prior map and no external source of localization

Soldiers in the field

Hazmat teams

Visually impaired
Problem Statement

Rely on internal sensing to provide navigation guidance:

- Accurate
- Spatially extended
- Temporally consistent
- Complex, cluttered, dynamic environments

Soldiers in the field

Hazmat teams

Visually impaired
Related Work

Simultaneous Localization and Mapping (SLAM)
• Build map and localize agent in the map at the same time

Visual odometry over 650m trajectory (Civera, 2009)

Laser-based SLAM, 30x30m (Grisetti, 2006)
Related Work

Simultaneous Localization and Mapping (SLAM)
• Build map and localize agent in the map at the same time

Visual odometry over 650m trajectory (Civera, 2009)

Laser-based SLAM, 30x30m (Grisetti, 2006)

Limitations of SLAM
• Sensitive to degenerate user motions
• Sensitive to ambiguous world configurations
• Challenging in dynamic environments
• Requires full sensor calibration

Full metric reconstruction overkill for human navigation?
Contributions

• Sensor suite for vision-based navigation

• Body-relative navigation guidance using uncalibrated cameras
  • Topological mapping & loop closure
  • Localization & Rotation guidance

• Application to ground robot navigation

• User Study
System Overview

Why vision?

- Light, cheap, compact
- Rich information

Why uncalibrated cameras?

- Intrinsic calibration is tedious
  (center of projection, focal length, distortion parameters)
- Extrinsic calibration is hard
  (sensor-to-body transformation, 6DOF/camera)
- Calibration subject to change in real conditions
Outline

- Sensor suite for vision-based navigation
- Body-relative navigation guidance using uncalibrated cameras
  - Topological mapping & loop closure
  - Localization & Rotation guidance
- Application to ground robot navigation
- User Study
Method Overview

**EXPLORATION (FIRST VISIT)**

Place graph

- Node = location in the world
- Edge = physical path btw nodes
- Arbitrary 3D environments

**NAVIGATION (SUCCESSIVE VISIT)**

Localization (node estimation)  
Rotation guidance

*Body-relative navigation guidance using uncalibrated cameras*, O. Koch, S. Teller, ICCV 2009
Body-relative navigation guidance using uncalibrated cameras, O. Koch, S. Teller, ICCV 2009
Place Graph

- Sparse representation of the exploration path
Place Graph

- Sparse representation of the exploration path

<table>
<thead>
<tr>
<th>World features</th>
<th>Node</th>
</tr>
</thead>
</table>

1/21/2010 Olivier Koch
Place Graph

• Sparse representation of the exploration path
Place Graph Generation

• Variation in environment appearance is captured by feature matching
Place Graph Generation

• Variation in environment appearance is captured by feature matching
Place Graph Generation

- **Energy function** $\Psi$ measures visual similarity.

$$\Psi(F_1, F_2) = \frac{1}{|F_1| + |F_2|} \cdot \left( |F_1| + |F_2| - 2 \cdot |\Phi(F_1, F_2)| + \sum_{(f_1^i, f_2^j) \in \Phi(F_1, F_2)} \|f_1^i - f_2^j\|_2 \right)$$

$F_1, F_2$: feature sets  
$\Phi$: feature matching function  
penalty for unmatched features  
average L2 distance in feature space

F1, F2: feature sets  
$\Phi$: feature matching function

$0 < \Psi < \Psi_0$  
$\Psi_0 < \Psi < 1$
**Place Graph Generation**

Live video stream

Energy Function

- Features
- Matched features
Method Overview

Live video stream

Place Graph Generation

EXPLORATION

Place graph:

0 2 3
1 4

NAVIGATION

Local Node Estimation

Rotation Guidance

Loop Closure Detection
Loop Closure Detection

**Approach**

1. generate loop closure hypothesis using visual similarity between nodes
2. validate hypothesis using temporal consistency
Bag-of-words

- Bag-of-words: model for data representation

- John will fly to San Francisco next week.
- Mary would like to fly to the moon with John.

Dictionary: {1: John, 2: will, 3: fly, 4: like, 5: to, 6: moon, 7: Francisco, 8: Mary}
- [1, 1, 1, 0, 1, 0, 1, 1, 0]
- [1, 0, 1, 1, 1, 0, 0, 1]

- Image="document", image feature="word"
- Applications
  - object categorization (generative/discriminative models)
  - object segmentation, localization
Loop Closure Detection

- Node="document", image feature="word"
Loop Closure Detection

- Similarity matrix
Loop Closure Detection

**Live video stream**

**Place Graph**

**Similarity Matrix**
Method Overview

Live video stream

Place Graph Generation

EXPLORATION

Place graph

Local Node Estimation

NAVIGATION

Loop Closure Detection

Rotation Guidance
Local Node Estimation

Discrete recursive Bayesian filtering

\[ p(x_k \mid z_k) : \text{probability of state } x_k \text{ while observing } z_k \text{ at time } k \]

1. **Prediction step**

\[ p(x_{k+1} \mid z_k) = \sum_k p(x_{k+1} \mid x_k)p(x_k \mid z_k) \]

2. **Observation step**

\[ p(x_{k+1} \mid z_{k+1}) = \lambda p(z_{k+1} \mid x_{k+1})p(x_{k+1} \mid z_k) \]
1. **Prediction step**

\[ p(x_{k+1} \mid x_k) \] : motion model

2. **Observation step**

\[ p(z_{k+1} \mid x_{k+1}) \] : observation model
Local Node Estimation

Live video stream

Place Graph

Localization

Node Estimation PDF
Method Overview

Live video stream

Place Graph Generation

Place graph

0
1
2
3
4

Local Node Estimation

Loop Closure Detection

Rotation Guidance

EXPLORATION

NAVIGATION
Relative orientation problem

Determine the orientation of the user at time \( t' \) relative to the orientation at time \( t \).

- **Time \( t \)**
  - Diagram showing orientation at time \( t \).
- **Time \( t' > t \)**
  - Diagram showing orientation at time \( t' \) compared to \( t \).
Rotation Guidance

Idea

Learn the relationship between user rotation and feature matches across cameras.

\[ \text{time } t \]

\[ \text{time } t' > t \]
Feature matching with one camera

Assumption: features are uniformly distributed in image space.

\[ \alpha: \text{rotation angle} \]
\[ f: \text{camera field of view} \]
Feature matching with two cameras

Assumption: cameras have the same field-of-view.
Feature matching with $n$ cameras

Match matrix $(n \times n)$
Learning the match matrix from training

System rotating in place at constant speed $\omega$

- feature observed on camera $i$ at time $t$
(same) feature observed on camera $j$ at time $t + \Delta t$
Learning the match matrix from training

Repeat for all features and all pairs of image frames (time sampling)

Match matrix $(n \times n)$
Rotation Guidance

Match Matrix

Live video stream

Close-up view of matrix cell
Relative orientation problem

Determine the orientation of the user at time $t'$ relative to the orientation at time $t$. 

*Rotation angle $\alpha$?*
Orientation from a single correspondence

\[ Rotation \ angle \ \alpha: \ \alpha = \tau(h_{ij}, \sigma_{ij}) \]

Estimate of \( \alpha \):

\[ \bar{\alpha} = h_{ij} \]

Error:

\[ \epsilon = \bar{\alpha} - \alpha = \tau(0, \sigma_{ij}) = \tau(0, \sigma) \]
Orientation from $N$ correspondences

Estimate of $\alpha$:
$$\bar{\alpha} = \frac{1}{N} \sum_{i,j} h_{ij} \quad (N \text{ observations})$$

Error:
$$\epsilon = \bar{\alpha} - \alpha = \frac{1}{N} \sum_{i,j} \tau(0, \sigma)$$

CLT
$$N \xrightarrow{} \infty \quad \epsilon \sim N(0, \sigma_N) \quad \sigma_N = \frac{\sigma}{\sqrt{N}}$$

$f = 90^\circ \quad \sigma^2 = f^2/6 \quad \sigma = 36.7^\circ \quad N = 200 \quad \sigma_N = 2.6^\circ$
Rotation Guidance

**Training**
- Learn match matrix from training data
- Once and for all for a given camera configuration
- Arbitrary environment

Match matrix \((n \times n)\)

**Navigation**
- Use match matrix to estimate *relative* user orientation

Match matrix \((n \times n)\)

Rotation angle \(\alpha\)?
Rotation Guidance: Video

Live video stream

Localization

Features
Rotation Guidance

- Place graph node
- Test location
- Exploration path (notional)
- Rotation guidance

10m

1/21/2010 Olivier Koch 43
Method Overview

Live video stream → Place Graph Generation → Place graph

EXPLORATION

NAVIGATION

Local Node Estimation → Rotation Guidance → Loop Closure Detection
Rotation Guidance evaluation using IMU-derived ground truth (30 sec. sequence)

\[\mu = 0.7^\circ, \quad \sigma = 2.5^\circ, \quad \text{max} = 8^\circ\]
Method Evaluation

Large-scale rotation ground-truth using upward images

\[ \mu = 2.5^\circ \quad \sigma = 10.7^\circ \]
Method Evaluation

Large-scale localization ground-truth

![Graph showing node estimation and groundtruth over time and node ID]
Outline

• Sensor suite for vision-based navigation

• Body-relative navigation guidance using uncalibrated cameras
  • Topological mapping & loop closure
  • Localization & Rotation guidance

• Application to ground robot navigation

• User Study
Ground Robot Navigation

- Demonstrate approach on robot equipped with low-level obstacle avoidance capability
- Collect extended ground-truth data

Two-wheel drive – IMU – Hokuyo LIDAR – 4-camera rig

*Ground robot navigation using uncalibrated cameras*, Koch, Walter, Huang, Teller, ICRA 2010
1. Localization & high-level guidance using vision
2. “Greedy” local obstacle avoidance using laser

Ground robot navigation using uncalibrated cameras, Koch, Walter, Huang, Teller, ICRA 2010
Ground Robot Navigation

Loop closure – 24 minutes – 290 m
MIT Stata Center 3rd floor

- Exploration path (ground-truth)
- Loop closure
Ground Robot Navigation

3 missions  74 min – 380 m
MIT Stata Center  3rd floor

- Mission A: V, W, X
- Mission B: X, Y, Z
- Mission C: V, W, Z, V
Ground Robot Navigation

Distance to original path (m)

![Distance to original path graph]

Rotation guidance error (deg.)

![Rotation guidance error graph]
Ground Robot Navigation

*Loop closure – 26 minutes – 400 m*

MIT Stata Center 1st floor

- Outdoor path (notional, no mapping)
- Indoor path (ground-truth)
- Loop closure
Ground Robot Navigation
Using Uncalibrated Cameras

O. Koch, M. Walter, A. Huang, S. Teller
ICRA 2010
Outline

• Sensor suite for vision-based navigation

• Body-relative navigation guidance using uncalibrated cameras
  • Topological mapping & loop closure
  • Localization & Rotation guidance

• Application to ground robot navigation

• User Study
User Study

Experiment Setup
• Follow directions to a series of unknown target destinations in the building.

Effectiveness (qualitative metrics)
• User survey

Efficiency (quantitative metrics)
• # successful missions
• # unclear guidance events
• Speed ratio

Tablet PC User Interface
User Study

9 users, 2.5 hours, 6 km of exploration, 59 missions

**Quantitative Analysis (log data)**

<table>
<thead>
<tr>
<th></th>
<th>Worst</th>
<th>Mean</th>
<th>Best</th>
</tr>
</thead>
<tbody>
<tr>
<td>Success rate</td>
<td>50%</td>
<td>70%</td>
<td>100%</td>
</tr>
<tr>
<td>Speed ratio</td>
<td>0.37</td>
<td>0.51</td>
<td>0.61</td>
</tr>
<tr>
<td>Unclear guidance</td>
<td>1.5 min.</td>
<td>3 min.</td>
<td>Never</td>
</tr>
<tr>
<td>User lost</td>
<td>3 min.</td>
<td>5 min.</td>
<td>Never</td>
</tr>
</tbody>
</table>

**Qualitative Analysis (survey)**

**Positive Feedback**

High-level guidance and non-metrical information are crucial.

**Suggested improvements**

More advanced turn-by-turn directions (Google Maps)
Compass/turn-by-turn inconsistency.
Drawing directions on images.
Simplify GUI.
Exploration Mission

23 minutes - 1.1 km
Indoor/outdoor
Dynamic scenes
**User Study**

**Navigation Mission**

32 minutes

# user lost: 3
# unclear guidance: 8
Speed ratio: 0.70
User Study

False positive

Ambiguous world configuration
Failure Modes & Limitations

- Featureless environments
- Visually repetitive environments
- Non-isotropic feature distributions
- Ambiguous world configurations
- “One-shot” loop closure events
Contributions

• Sensor suite for vision-based navigation

• Body-relative navigation guidance using uncalibrated cameras
  • Topological mapping & loop closure
  • Localization & Rotation guidance

• Application to ground robot navigation

• User Study
Future Work

• Incorporate metrical constraints in the place graph
• Account for spatial relationship of features
• Bound expansion of visual vocabulary
• User interface
• Application to the visually impaired and in natural environments
Thank you

• Seth Teller
• Bill Freeman, Rob Miller
• The Draper Laboratory
• TIG, Bryt (Bradley), Ron (Wiken)

The 33x neighborhood
Questions?

Repeat for all features and all pairs of image frames (time sampling)

Match matrix \((n \times n)\)

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ETA 00:01:10
DIST 95 m.

78% confidence

Report problem