A Constant-Time Efficient Stereo SLAM System

Christopher Mei[1], Gabe Sibley[2], Mark Cummins[2], Paul Newman[2] and Ian Reid[1]

[1]Active Vision Group,
[2]Mobile Robotics Group,
University of Oxford

Simultaneous Localisation and Mapping

Goal of SLAM :

- to estimate the trajectory of a sensor in a iterative fashion,
- build a representation of the environment.

What makes SLAM challenging ?

- computationally expensive (the complexity grows with the size of the environment),
- requires robust data fusion algorithms.

A constant time stereo SLAM system

This talk will describe...

- a continuous relative representation (CRR) providing a way to represent precisely the **local** environment and improve accuracy at loop closure **without** global estimation,
- low-level vision tasks: image pyramids, quadtree feature selection, scale invariance through stereo, relocalisation and sub-pixel minimisation to provide improved accuracy and robustness.

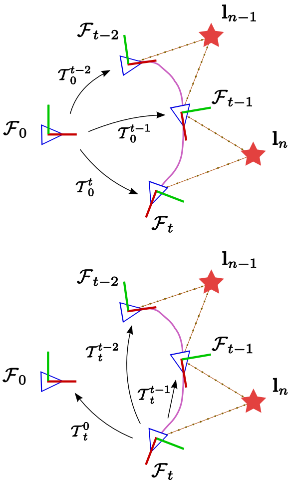
... to obtain a **constant-time** precise and robust stereo SLAM system.

Outline

Outline

1. World representations
2. Stereo Processing
3. Experimental results
4. Conclusion

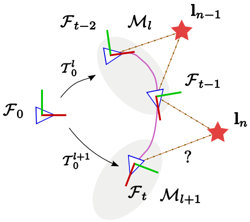
Global representations

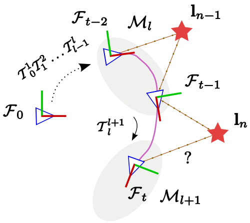


Global Representations

- “standard” representation: robot poses and landmarks are represented in a global frame,
- robocentric representation: global frame centred on the current robot position. Improves the consistency of EKF-SLAM [Castellanos et al IFAC 2004]

Sub-mapping

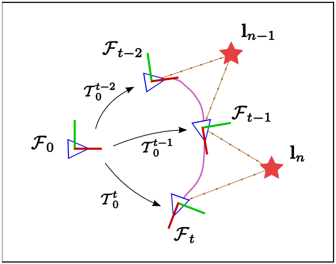


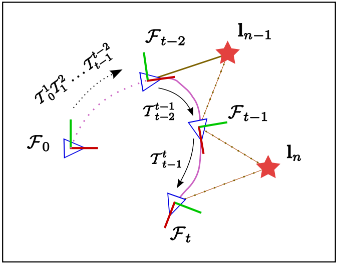


Sub-maps

- sub-maps: robot poses and landmarks are grouped in smaller maps. Each sub-map uses a global or relative representation ([Estrada *et al.* ITRO 2005] [Bosse *et al.* IJRR 2004]),
- reduces the complexity and improves consistency,
- limitations: sharing information between sub-maps is a difficult problem in particular when updating the estimates after loop closure.

Continuous Relative Representation

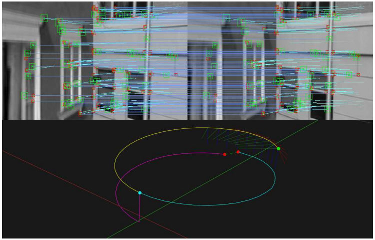


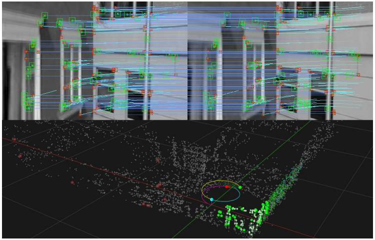


Changing the representation has a **profound** impact on the mapping process. In particular, the maximum likelihood cost functions for the global representation and CRR are **different**:

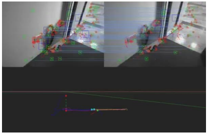
See [Sibley *et. al* RSS 2009]

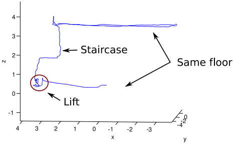
Example with a loop closure

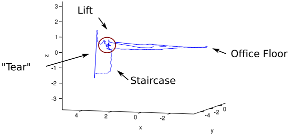




Change of perspective







Before loop closure

After loop closure

Change of perspective

- a robot should be able to evolve in an environment with non observable ego-motion,
- a challenge is to find the places where such events occur (lifts, trains, planes, ...)

Advantages and limitations of the CRR

Advantages

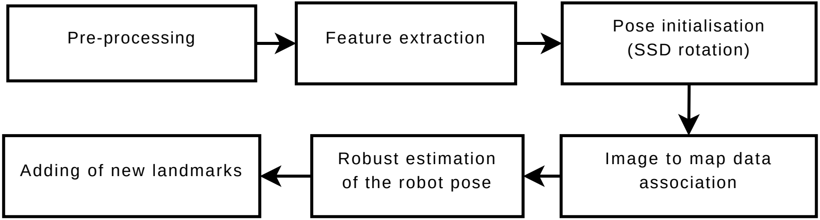
- simplifies loop closure and data association,
- the precision at loop closure is the same as during exploration.

Limitations

- to obtain a coherent global Euclidean representation, global estimation is required (with a complexity dependent on the size of the map), **but** is this often acquired?

Outline

Stereo SLAM processing steps



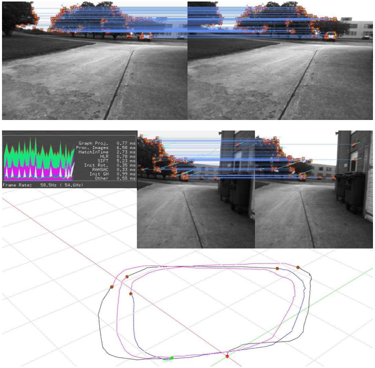
Loop closure is done using a bag of words approach: FabMap [Cummins and Newman IJRR 2008].

Components of the SLAM system

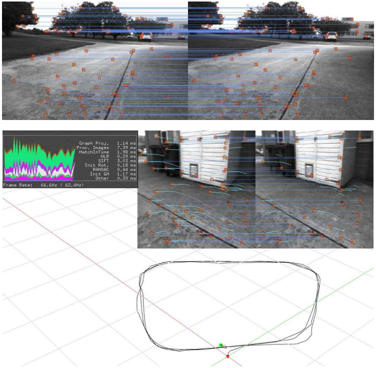
Processing steps that greatly effect the system's performance

- Pose initialisation by image-based rotation estimation ([Mei *et. al* ITRO 2008]),
- Better pose conditioning by spreading features using quadtrees,
- Sub-pixel minimisation for each landmark to image matching to improve precision,
- Efficient scale invariance by using the landmark distance ("true scale").

Improving conditioning using quadtrees



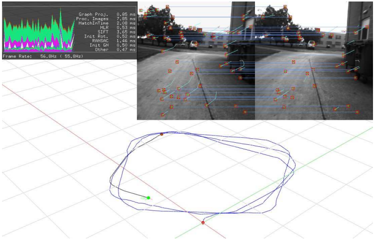
Without quadtrees



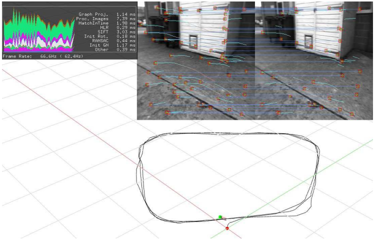
With quadtrees

Improving precision through sub-pixel minimisation

Applying sub-pixel minimisation (eg. Kanade-Lucas-Tomasi tracker) greatly improves precision and data association.

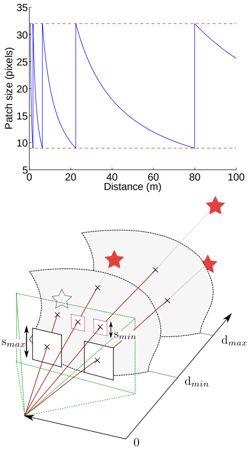


Without sub-pixel minimisation



With sub-pixel minimisation

Efficient scale invariance

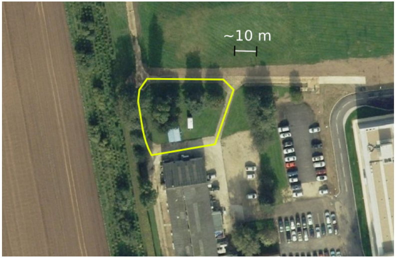


- Scale invariance in SIFT: obtained by looking for a maximum score response at different scales in a DoG pyramid, this is an expensive process,
- Alternative: choose a fixed 3-D template size for computing the descriptors, ("true scale").
- Problem: the camera image size is insufficient to represent large changes in depth,
- Solution: split the 3-D space in regions ("bands") where the 3-D templates have the same size.

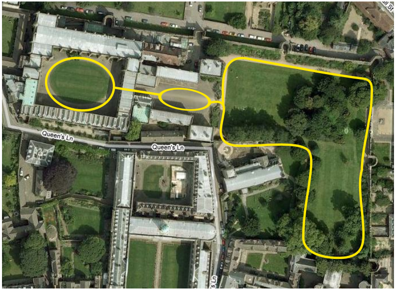
Outline

Experimental results

Begbroke Science Park (1.1km)



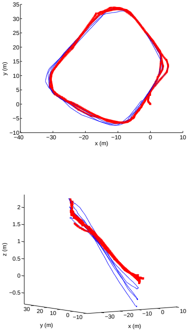
New College (1.8km)



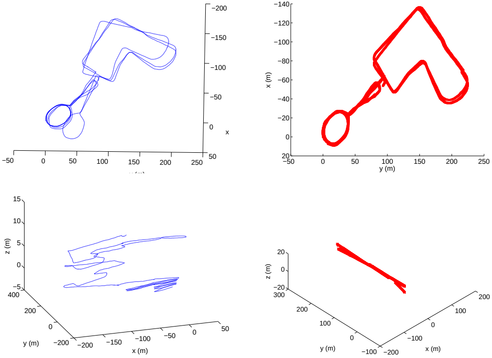
Importance of loop closure on precision (1/2)

Trajectory estimates with/without loop closure.

Begbroke



New College



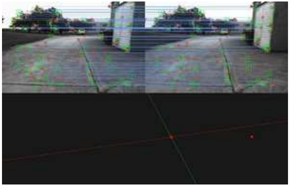
Importance of loop closure on precision (2/2)

	Begbroke	New College
Distance Travelled	1.08 km	2.26 km
Frames Processed	23K	51K
Reprojection Error Min/Avg/Max	0.003 / 0.17 / 0.55 pixels	0.03 / 0.13 / 1.01 pixels
Accuracy without loop closure	$\sim$1m in (x-y) plane, $\sim$1m in z	$\sim$15-25m in (x-y) plane, $\sim$15m in z
Accuracy with loop closure	$\sim$1cm in (x-y) plane, $\sim$1cm in z	$\sim$10cm in (x-y) plane, $\sim$10cm in z

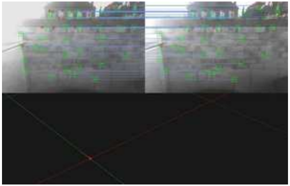
Results on difficult sequences

Large changes in view-point

Blur



Lens Flare and Saturation



Outline

Advantages and limitations

Advantages

- constant-time computation,
- robust: relocalisation and loop closure,
- adapted to large scale exploration: the only limit is the memory and not the computational resources.

Limitations

- loop closing is based on appearance and only triggered in distinctive environments: **no guarantee** of unicity or bounds on memory can be given,
- the visualisation of the map can be difficult for a user, imposing Euclidean constraints violates the constant-time performance guarantee.

Future work

Future work

- detecting non-observable ego-motion between Euclidean maps (eg: lifts, cars, trains, ...)
- long-term mapping: how can we detect change from simple occlusions and tracking failures?

[A demo will be running during tomorrow's poster session.]