Visual SLAM

Part III. Advances in visual methods and alternative frameworks

Andrew Davison
Imperial College London

Andrew Calway   Walterio Mayol
University of Bristol

Improvements can be made

- On the computer vision side, by helping data association.
- On representations and parameterizations to enhance mapping while within real-time.
- On alternative frameworks for mapping.

For data association, at the beginning...

- Small (e.g. 11x11) image patches around salient points to represent features.
- Normalized Cross Correlation (NCC) to detect features.
- Small patches + accurate search regions lead to fast camera pose estimation.

However

- Simple patches are insufficient for large view point or scale variations.
- Small patches help speed but prone to mismatch.
- Search regions can't always be trusted (camera occlusion, motion blur).

Some Solutions

- Estimate the 3D structure of each patch [Molton et al. BMVC 2004].
- Use a richer 2D descriptor [Chekhlov et al. ISVC 2006].
- Don’t use patches [Eade and Drummond, BMVC 2006, Smith et al. BMVC 2006, Gee and Mayol ISVC 2006].

- **Problem:** Difficult to match simple 2D features under viewpoint changes.

- **A solution:** Estimate patches in 3D and use NCC for matching.

[Molton, Davison and Reid, BMVC 2004]

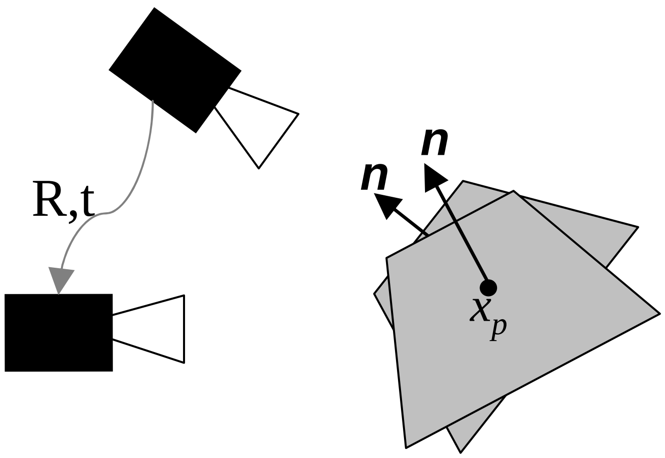
•Probabilistic method estimates surface normal using gradient-based image alignment.

•Assume patch around x_p is locally planar in 3D.

•Predict local appearance before measurement.

•Use measurement to update normal estimate.

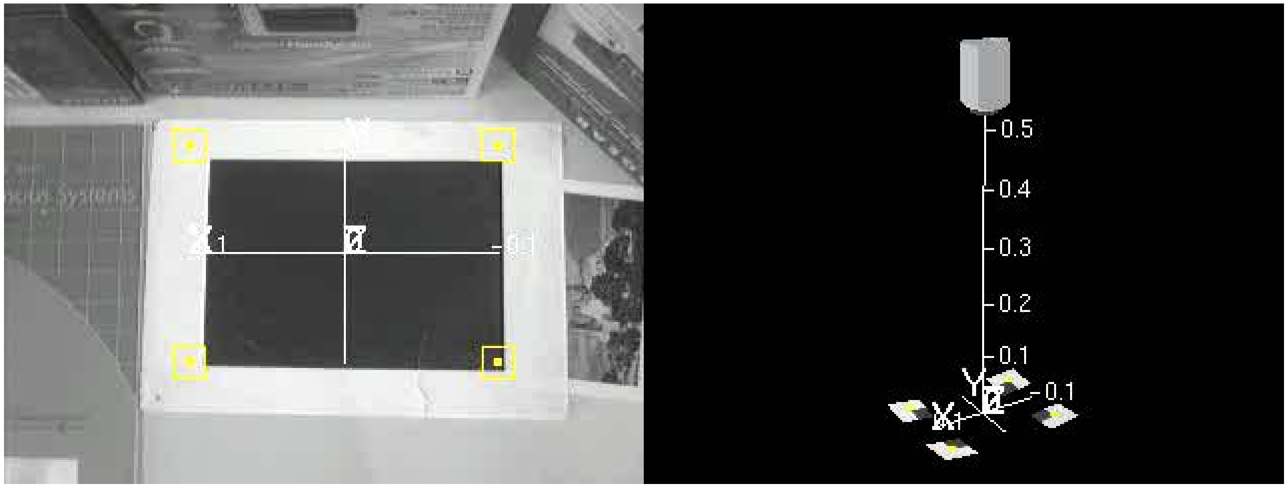
•Uses inverse compositional algorithm [Baker&Matthews 02]



Assume first sight fronto-parallel

Estimation of ***n*** kept apart from main SLAM

[Molton, Davison and Reid, BMVC 2004]



Video at http://www.doc.ic.ac.uk/~ajd/Movies/boxes.mpg

$$I_n(W_{n-1}(\mathbf{x})) = I_T(W_i(\mathbf{x};\mathbf{p})) \approx I_T(\mathbf{x}) + \nabla I_T \frac{\partial W_i}{\partial \mathbf{p}} \Delta\mathbf{p} \;; \qquad \mathtt{H} = \mathtt{CR}[\mathbf{n}^T \mathbf{x}_p \mathtt{I} - \mathbf{t}\mathbf{n}^T]\mathtt{C}^{-1}$$

- Find parameters $\mathbf{p}$ to warp template I_T to observation I_n .

- Warps W_n and W_i depend on camera parameters and patch normal $\boldsymbol{n}$.

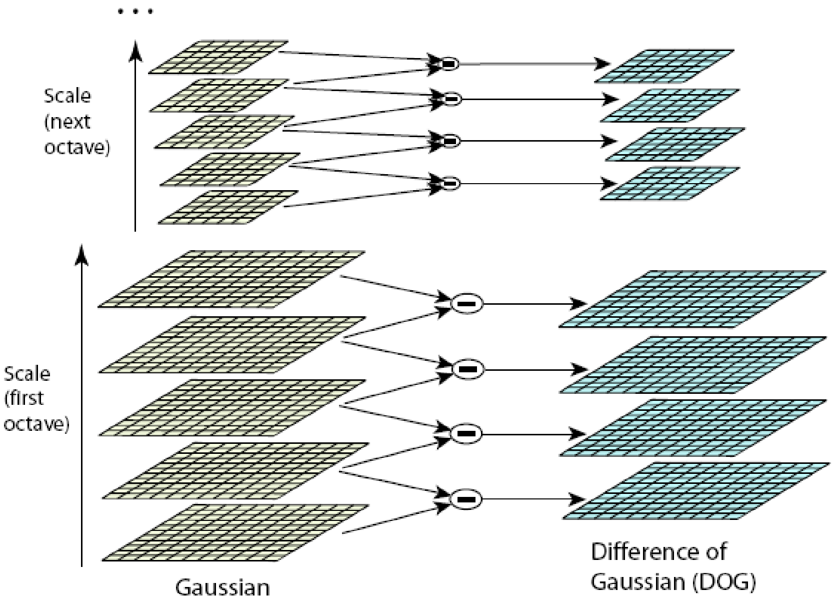
- Assess patch's normal $\boldsymbol{n}$ via $\mathbf{p}$ through homography H relating views.

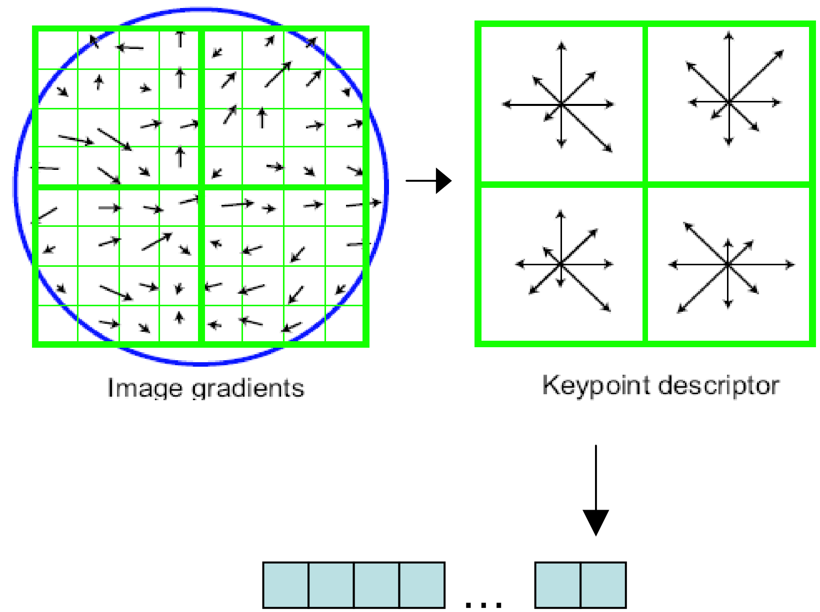
- Several important considerations: keeping the original camera position within SLAM state, resampling issues, etc.

•**Problem:** camera pose estimates can become invalid after occlusion or rapid motion resulting in possible mismatches.

•**A solution:** Use a richer feature descriptor e.g. SIFT.

Recall SIFT [Lowe, IJCV 2004]





128 elements vector

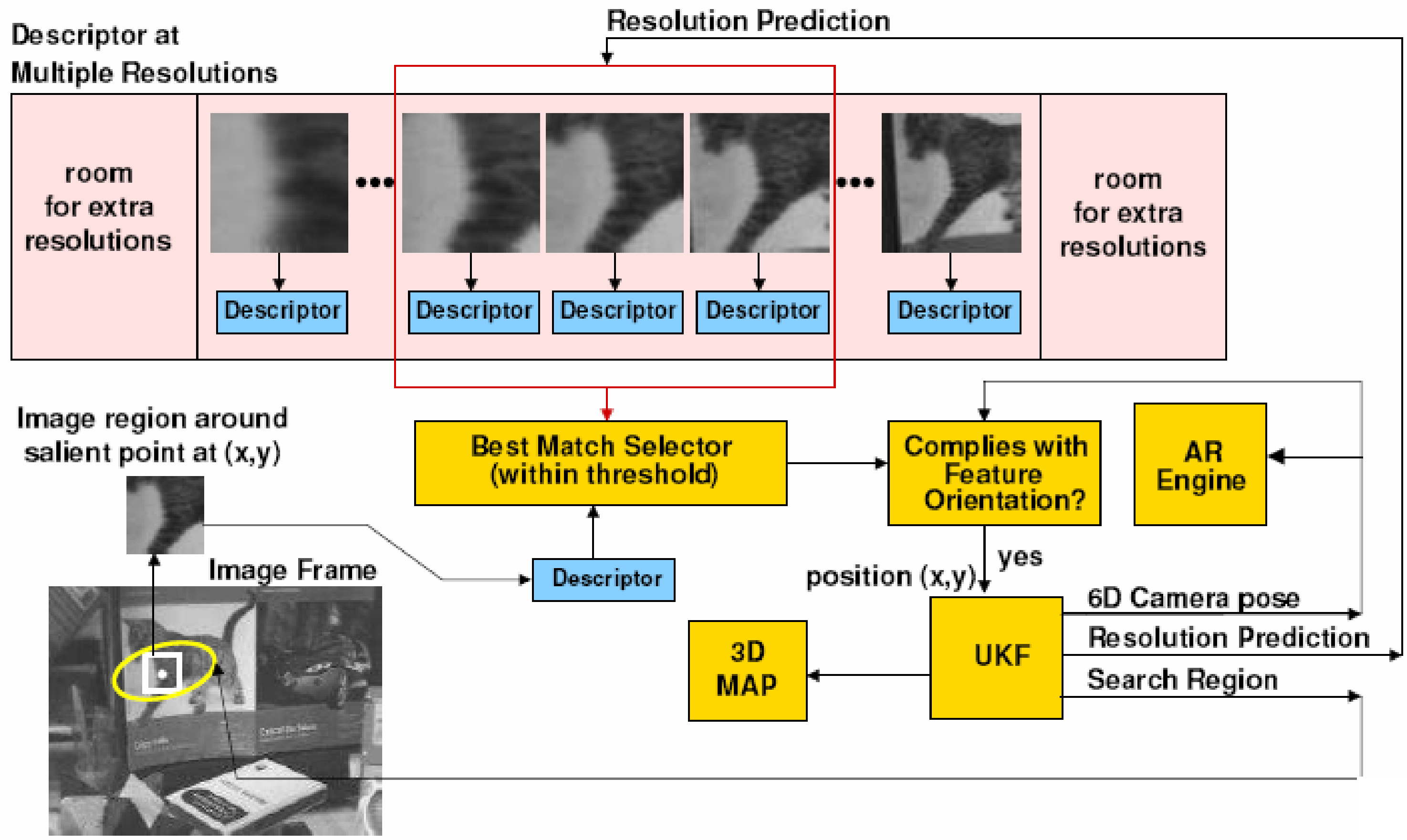
Find maxima in scale space to locate keypoint.

Around keypoint, build invariant local descriptor using gradient histograms.

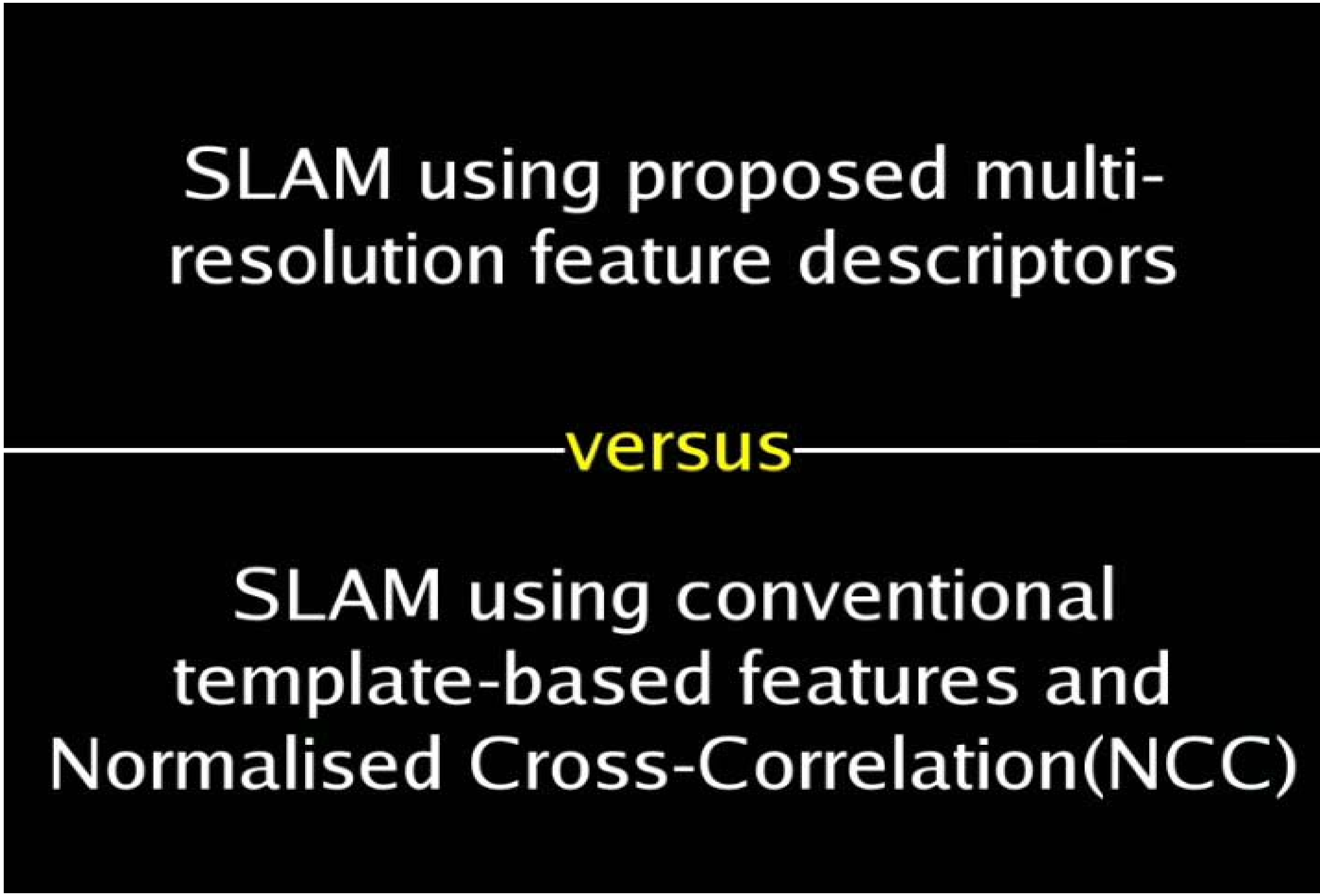
If for tracking, this may be wasteful!

[Chekhlov, Pupilli, Mayol and Calway, ISVC06/CVPR07]

•Uses **SIFT**-like descriptors (histogram of gradients) around Harris corners.
•Get scale from SLAM = “predictive SIFT”.



[Chekhlov, Pupilli, Mayol and Calway, ISVC06/CVPR07]



Video at http://www.cs.bris.ac.uk/Publications/attachment-delivery.jsp?id=9

- Able to relocate after camera total occlusion + moderate “kidnapping”.

- **Problem:** Objects may lack rich texture.

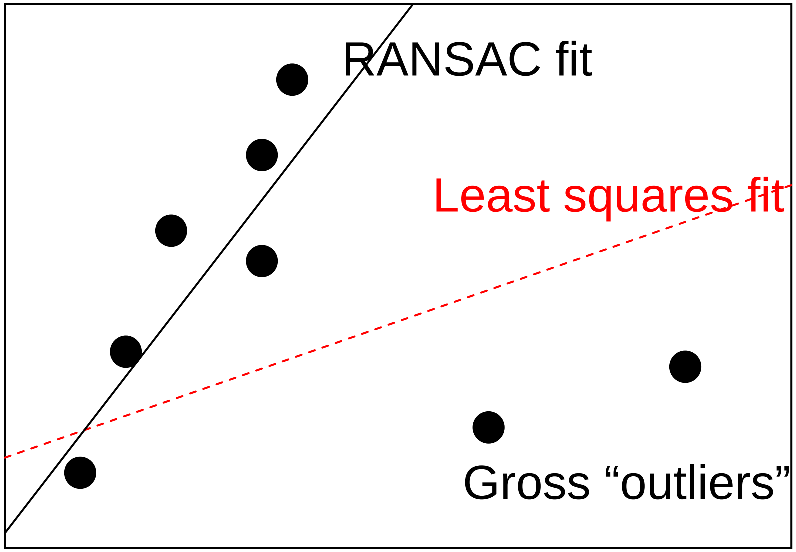
- **A solution:** Use edge information.

Or “2006 the year we lived at the *edge*”

Edges have less distinctiveness compared to textured patches -> use a robust estimator e.g. RANSAC, MLESAC, etc.

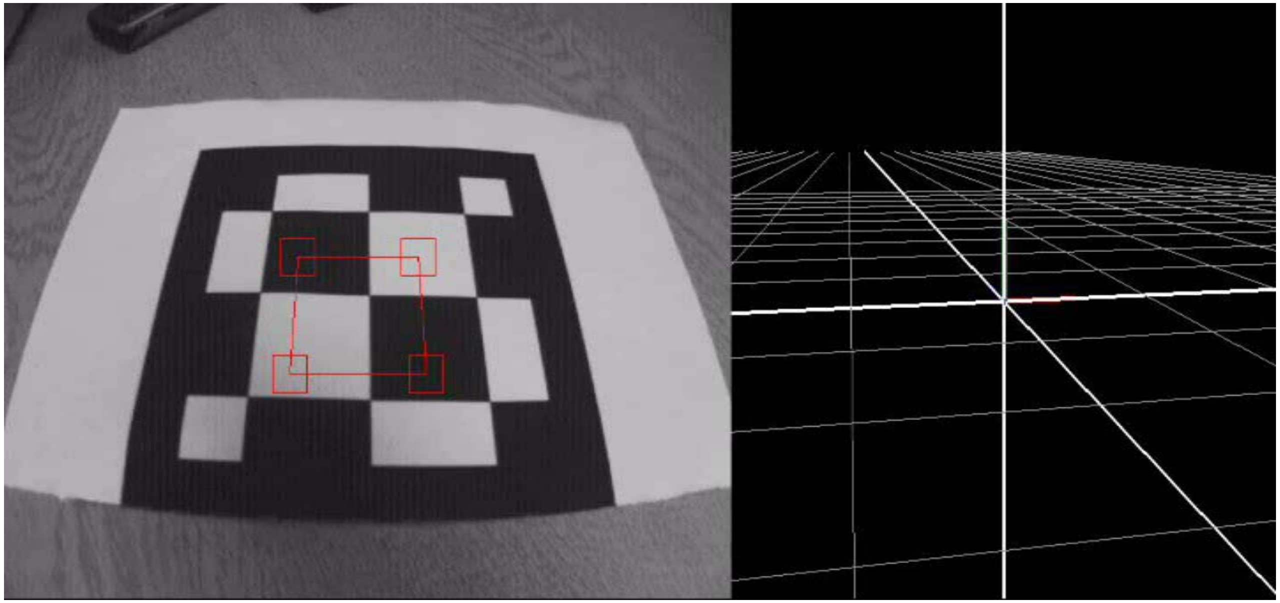
Recall RANSAC [Fischler and Bolles 1981]

Random Sampling and Consensus



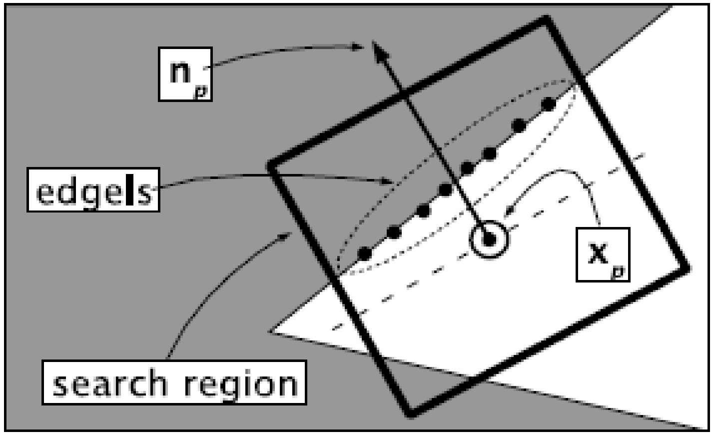
- Select random sample of points.
- Propose a model (hypothesis) based on sample.
- Assess fitness of hypothesis to rest of data.
- Repeat until max number of iterations or fitness threshold reached.
- Keep best hypothesis and potentially refine hypothesis with all inliers.

[Eade and Drummond, BMVC2006]



Video at http://mi.eng.cam.ac.uk/~ee231/bmvcmovie.avi

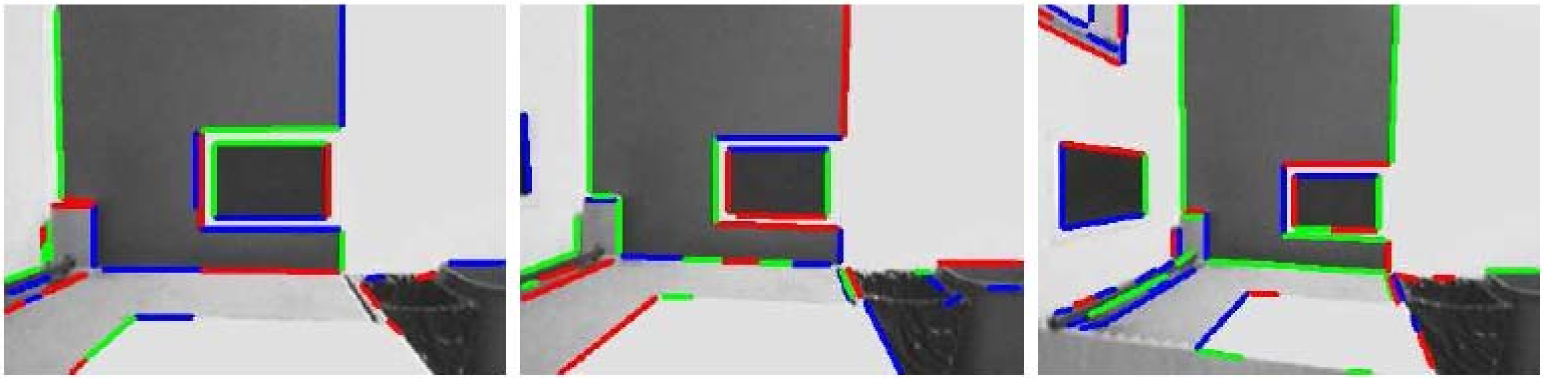
Edglets:



- Locally straight section of gradient Image.
- Parameterized as 3D point + direction.
- Avoid regions of conflict (e.g. close parallel edges).
- Deal with multiple matches through robust estimation.

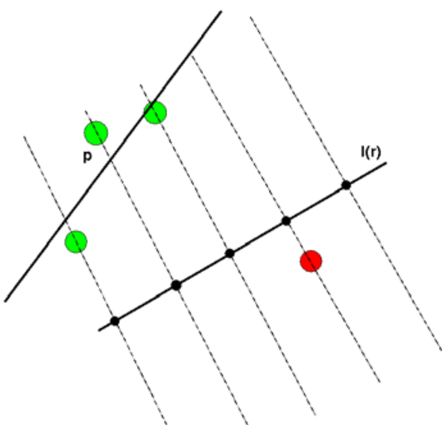
[Gee and Mayol, ISVC 2006]

Detect lines

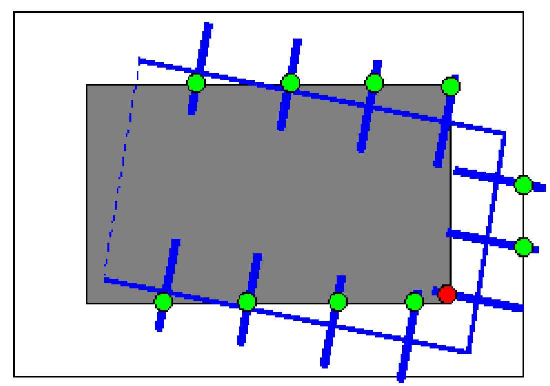


Fast Conic extraction by tracing undistorted edge image using [Gates et al, 05]

Measure

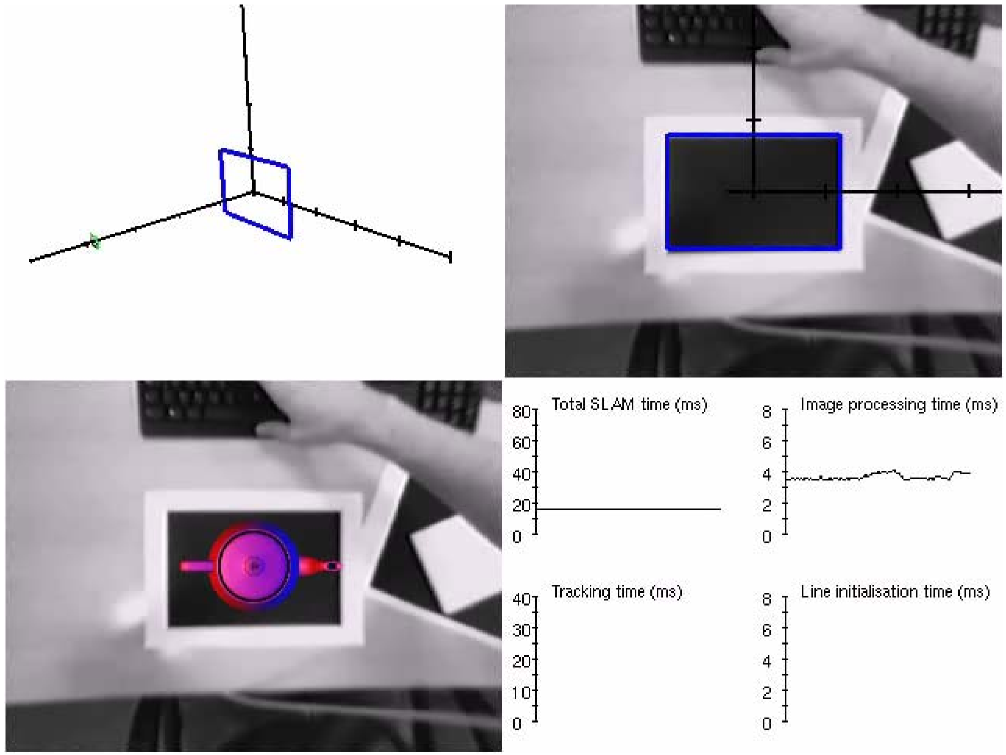


Search orthogonal to projected line.
Robust edge point detection.



Model based tracking [Harris, 1992]

[Gee and Mayol, ISVC 2006]



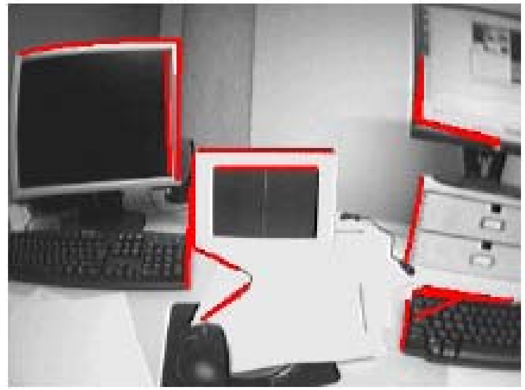
Video at http://www.cs.bris.ac.uk/Publications/attachment-delivery.jsp?id=14

- Straight edges extracted in 2D.
- Camera pose and 3D edge initialization using model-based tracking.
- RANSAC used for dealing with occlusions.

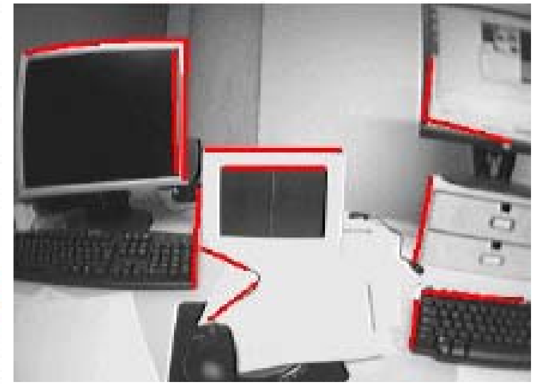
[Smith, Reid and Davison, BMVC2006]

Detect edges

FAST corners + Sobel



Lines longer
than 30pixels



After removing
Overlapping lines

In undistorted image:

•Detect FAST corners [Rosten and Drummond, 2005].

•Quickly verify there is an edge between two corners by bisecting checks.

•Remove overlapping lines.

•To measure a line, also use normal to projected line as in [Harris 1992].

[Smith, Reid and Davison, BMVC2006]

Video at http://www.doc.ic.ac.uk/~ajd/Movies/SRD06_LineSLAM.avi

- Combines points and lines.

- Straight edges extracted in 2D, points represented as 2D textures using [Molton et al. 04].

Issues with edges:

- These systems show it is possible to use 3D edges in real-time visual SLAM.
- But extracted edges are high-contrast.
- More work is needed to extract “fainter” reliable edges and to characterize them.

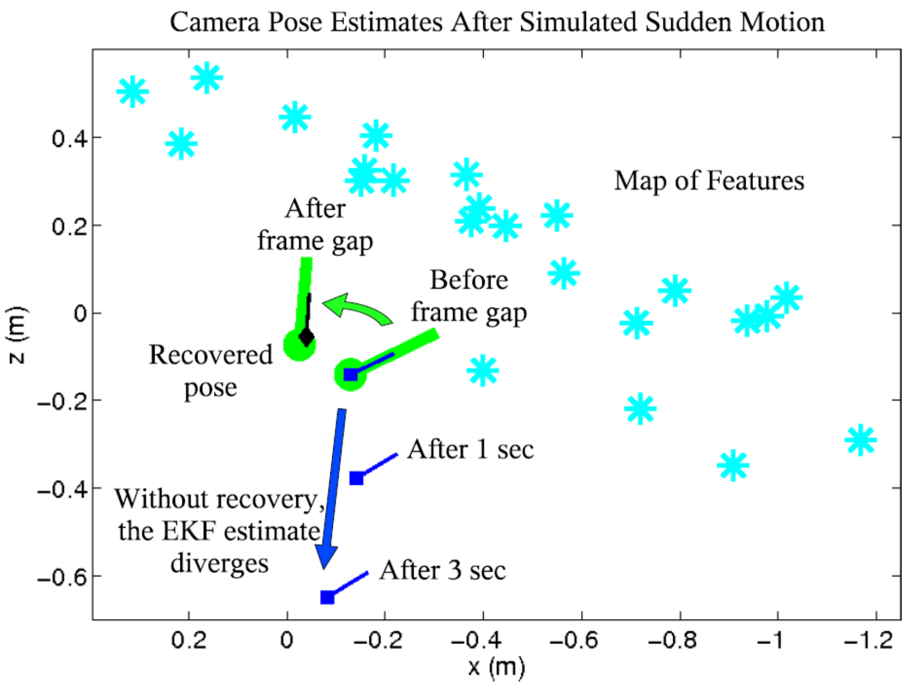
OK nice but…

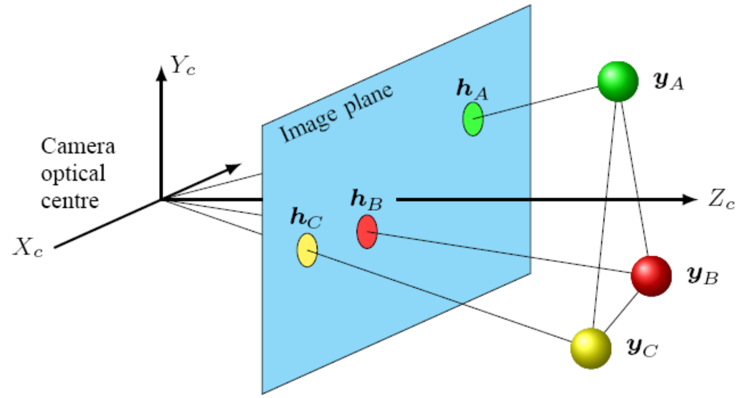
- Having rich descriptors or even multiple kinds of features may still lead to wrong data associations (mismatches).

- If we pass to the SLAM system every measurement we think is good it can be catastrophic.

- Better to be able to recover from failure than to think it won’t fail!

[Williams, Smith and Reid ICRA2007]



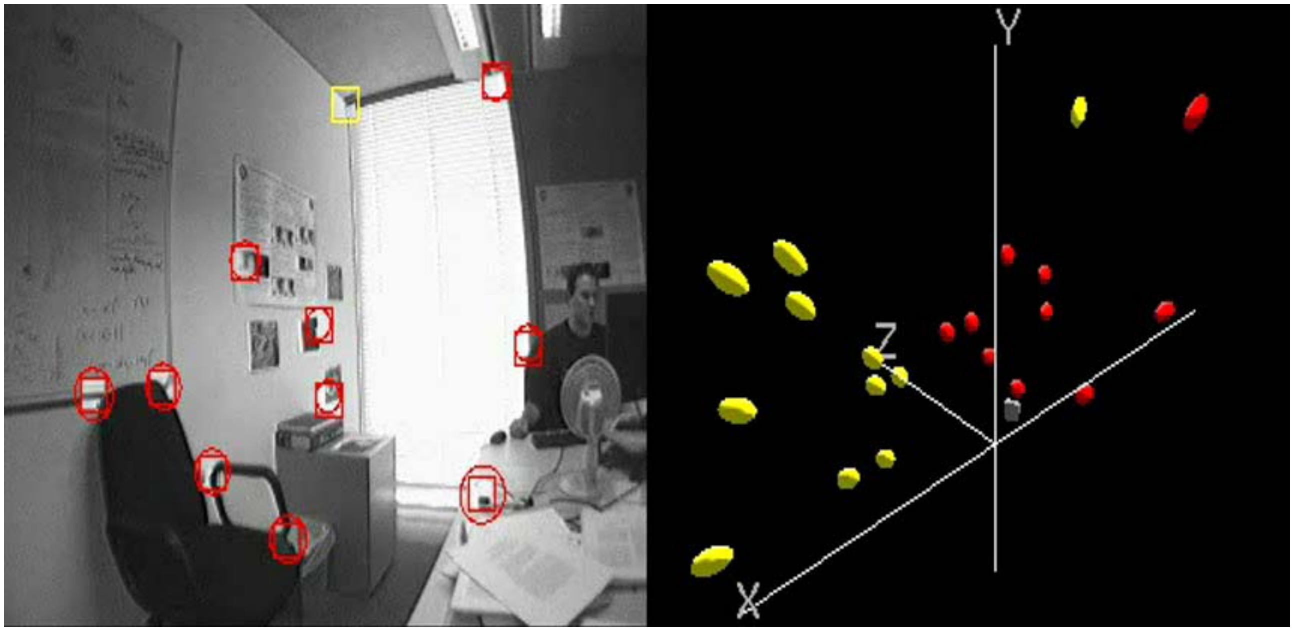


Use 3 point algorithm -> up to 4 possible poses. Verify using Matas' $T_{d,d}$ test.

•Camera relocalization using small 2D patches + RANSAC to compute pose.

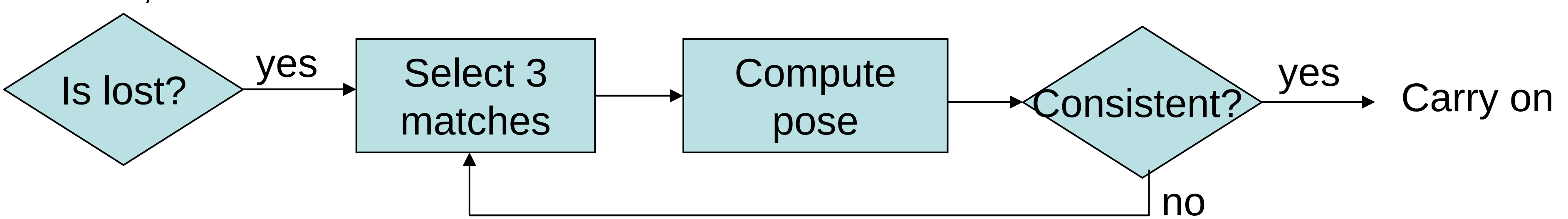
•Adds a “supervisor” between visual measurements and SLAM system.

[Williams, Smith and Reid ICRA2007]



Video at http://www.robots.ox.ac.uk/ActiveVision/Projects/Vslam/vslam.04/Videos/relocalisation_icra_07.mpg

In brief, while within real-time limit do:



Also see recent work [Williams, Klein and Reid ICCV2007] using randomised trees rather than simple 2D patches.

Submapping

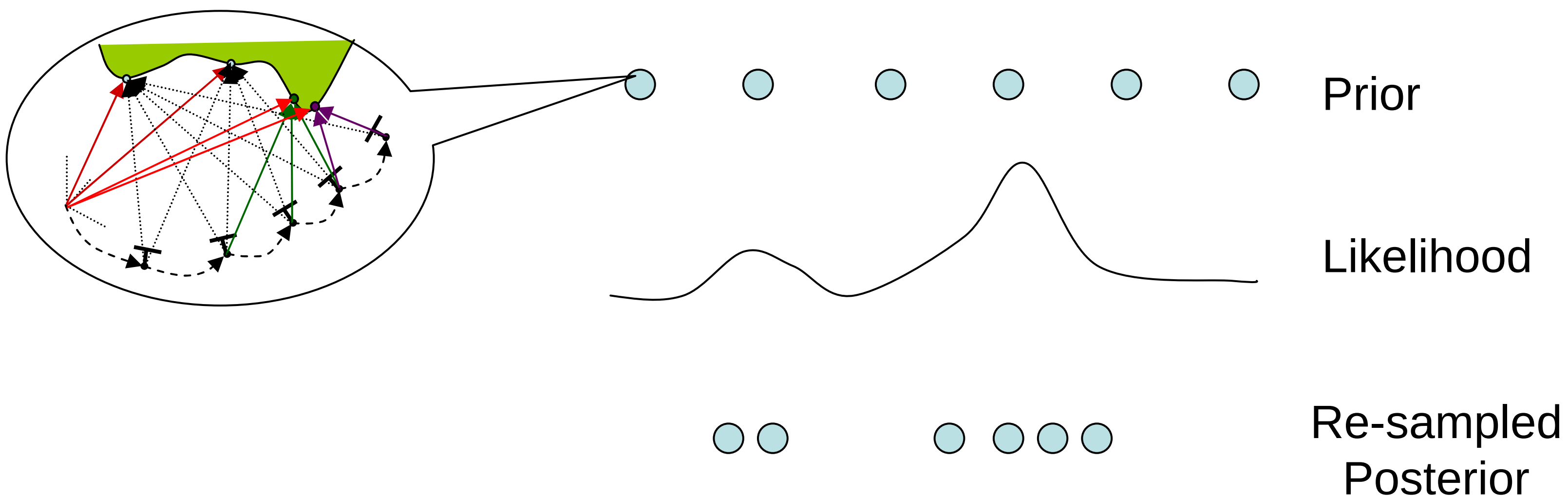
Partitions problem in smaller maps able to be handled in real-time by the EKF.

See [Clemente et al. RSS 2007]

Other frameworks: FastSLAM

Rao-Blackwellized Mapping:

- Keep multiple hypothesis in the form of “particles”.
- Each particle represtents a camera trayectory.
- Each particle keeps its own map.
- Each particle “offsprings” based on the likelihood of its map.



Other frameworks: FastSLAM

•Recall $\mathbf{P}$ (covariance matrix) in EKF SLAM scales to the square of number of features.

•In Particle Filters, the increase is linear (in relation to number of particles).

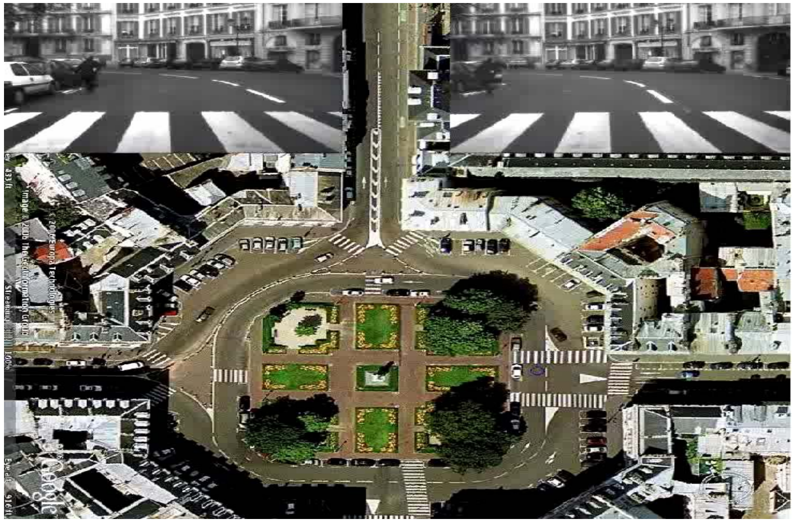
•Though keeping a full map for each particle impacts computational demands.

For reference See:

• Eade and Drummond. *Scalable Monocular* SLAM, CVPR, 2006.

•Stachniss and Burgard, *Rao-Blackwellized Particle Filters and Loop Closing* IROS'05 Tutorial.

Other frameworks: concurrent SFM approaches



A. Comport *et al.*

Video at http://www-sop.inria.fr/icare/personnel/andrew.comport/videos/visualodometry/Versailles_rond_all.mpg

Some recent examples:

- A.I. Comport, E. Malis, P. Rives. Accurate Quadri-focal *Tracking for Robust 3D Visual Odometry*. In IEEE Int. Conf. on Robotics and Automation, ICRA'07.

- E. Royer, M. Lhuillier, M. Dhome and J.M. Lavest, *Monocular vision for mobile robot localization and autonomous navigation* International Journal of Computer Vision, 74(3) , 2007.

- D. Nister, Preemptive RANSAC for Live Structure and Motion Estimation, *Machine Vision and Applications*, Volume 16, Issue 5, Dec 2005, Pages 321–329, 2005.

New stuff:

- Extracting high-level structure:
  - See Gee *et al.* BMVC2007
- AR applications:
  - Klein and Murray ISMAR 2007 (not SLAM but impressive!)
  - Chekhlov *et al.* ISMAR 2007
- Enhancements:
  - Williams, Klein and Reid ICCV 2007 (online randomized trees for SLAM)
  - Castle *et al.* ICRA 2007 (links recognition with SLAM)
  - Eade and Drummond, ICCV 2007 (graph SLAM)

Software tools:

- http://www.doc.ic.ac.uk/~ajd/Scene/index.html

<MonoSLAM code for Linux, works out of the box>

- http://www.openslam.org/

<for SLAM algorithms mainly from robotics community>

- http://www.robots.ox.ac.uk/~SSS06/

<SLAM literature and some software in Matlab>

Recommended intro reading:

- Yaakov Bar-Shalom, X. Rong Li, Thiagalingam Kirubarajan, *Estimation with Applications to Tracking and Navigation*, Wiley-Interscience, 2001.

- Hugh Durrant-Whyte and Tim Bailey, Simultaneous Localisation and Mapping (SLAM): Part I The Essential Algorithms. *Robotics and Automation Magazine*, June, 2006.

- Tim Bailey and Hugh Durrant-Whyte, Simultaneous Localisation and Mapping (SLAM): Part II State of the Art. *Robotics and Automation Magazine*, September, 2006.

- Andrew Davison, Ian Reid, Nicholas Molton and Olivier Stasse MonoSLAM: Real-Time Single Camera SLAM, IEEE Trans. PAMI 2007.

Some Challenges

- Deal with larger maps.
- Obtain maps that are task-meaningful (robotics, AR, metrology).
- Use different feature kinds on an informed way.
- Benefit from other approaches such as SFM but keep efficiency.
- Incorporate semantics and beyond-geometric scene understanding.

Part III Comments

- SLAM problem mathematically *solved*.

- Computer vision implementations still improving: *Feature description, kinds of features, parameterizations and robustness*.

- But demonstrators are already working.

Thank you!

Please check our homepages later for a copy of the slides