15 Years of Visual SLAM

Andrew Davison

Robot Vision Group and Dyson Robotics Laboratory
Department of Computing
Imperial College London
www.google.com/+AndrewDavison

December 18, 2015

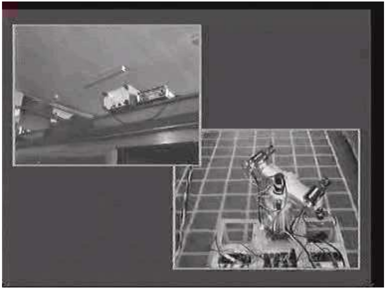
What Has Defined Visual SLAM for me?

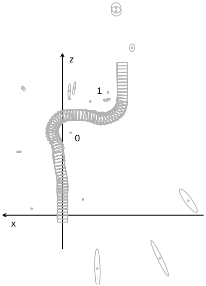
- Closed loop estimation, *predictive*, efficient.
- **Live demos!**
- Focus on a single visual sensor in a small area; drift-free, consistent localisation.
- Many possible applications easily apparent.
- Commodity hardware (cameras and processors); open source software.

- I believe that this research is evolving towards general real-time spatial perception (but that it's still SLAM!)

My Pre-2000 Visual SLAM Work

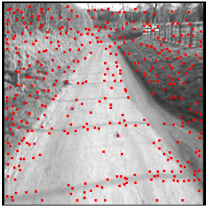
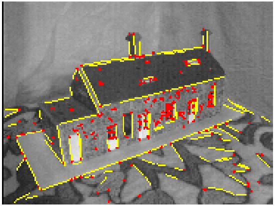
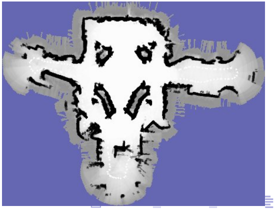
- SLAM with Active Vision (with David Murray, Oxford). 5Hz real-time loop on a 100MHz PC:
  Predict, move, measure, update.
- Generalised system at AIST, Japan and first SceneLib open source code.





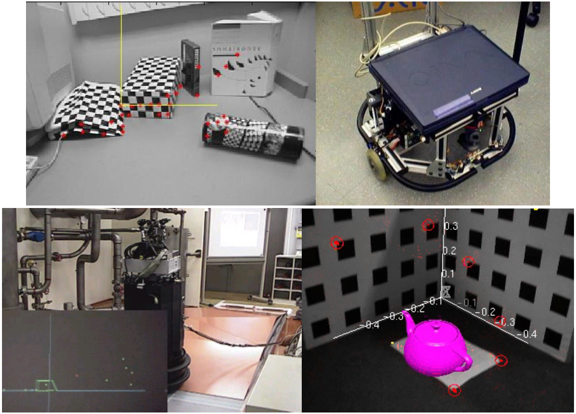
Earlier Inspirations and Building Blocks

- DROID (Harris, late 1980s, feature-based VO)
- Off-line SFM moving towards sequence processing (e.g. Fitzgibbon, Pollefeys).
- EKF SLAM with non-visual sensors (Durrant-Whyte, Leonard, etc.).
- Laser scan matching (e.g. Gutmann and Konolige).
- The mobile robotics community had almost completely turned away from vision.
- The computer vision community had almost completely turned away from real-time and robotics.

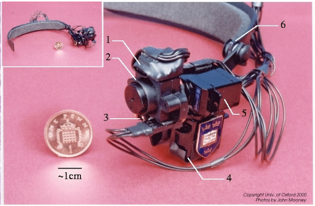




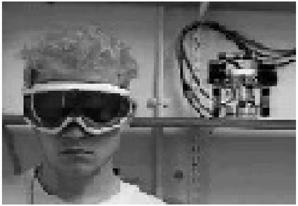
The Move to 3D Monocular SLAM

- Chiuso, Favaro, Jin, Soatto MfM sequential SFM 2000
- My work on 3D motion of a wheeled robot; experiments with general 3D tracking.



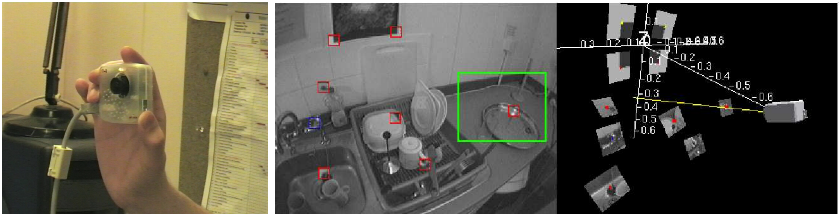
Key Applications for Single Sensor SLAM





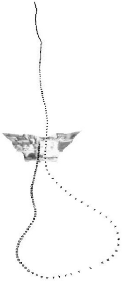
- Low-cost robotics.
- Agile robotics (e.g. MAV).
- Smartphone/personal/wearable.
- AR/VR inside-out tracking; gaming.

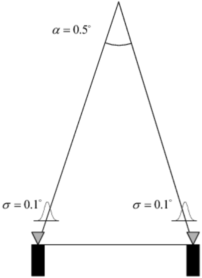
MonoSLAM: Sparse Feature-Based SLAM (2003)

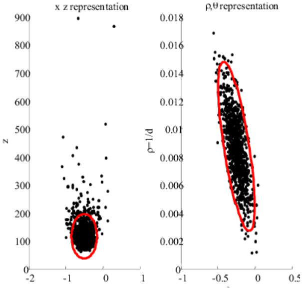


- EKF estimation; sparse map of high quality features; tight measurement loop with active prediction. Solid 30FPS performance on a laptop. Collaboration with Ian Reid, Nick Molton, Walterio Mayol and others.
- Live demos at ICCV 2003, ISMAR 2003, CVPR 2004, BMVC 2004, many others.
- Thanks particularly to Walterio Mayol and ISMAR for pushing me to demo it.

Intermediate Years

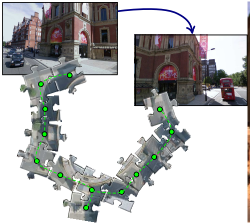


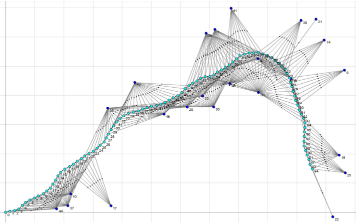




- 2003/4 Nister Visual Odometry (joint CVPR 2005 Tutorial).
- 2003 Jung and Lacroix aerial SLAM.
- 2005 Pupilli and Calway (particle filter) + other Bristol work.
- 2005 Robert Sim RBPF visual SLAM.
- 2006–2008 with Montiel, Civera *et al.*Zaragoza Inverse depth features and better parameterisation.

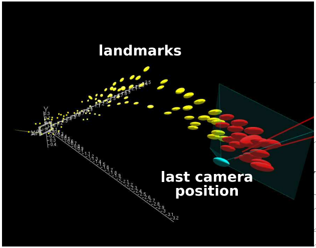
Towards Large Scale Consistent Mapping

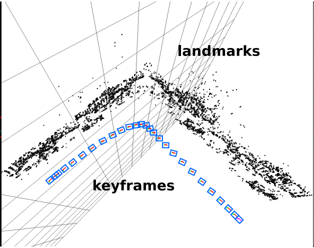


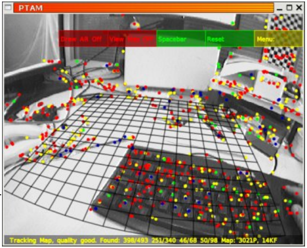


- 2006 Ho and Newman; then Cummins and Newman FAB-MAP: image retrieval for loop closure detection.
- 2006 SLAM Summer School: real joining of graph/BA optimisation methods into SLAM; particular Dellaert and Konolige.

Big Improvements in Small Local Monocular SLAM



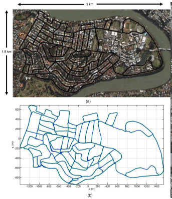
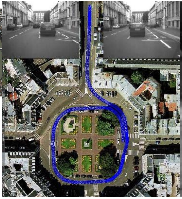
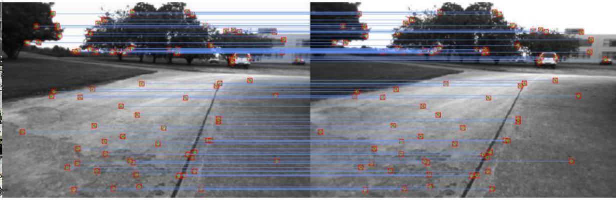




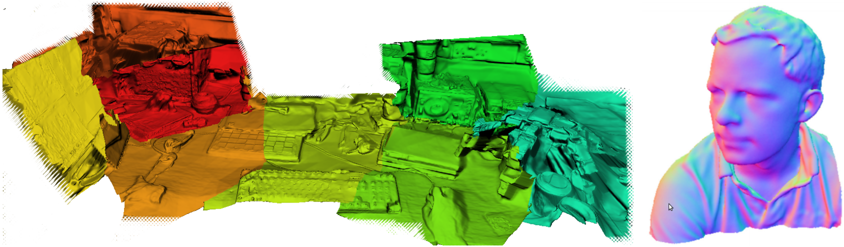
- 2007 Relocalisation in MonoSLAM (Williams, Klein, Reid).
- 2007 PTAM, Klein and Murray.
- 2007 Eade and Drummond, information filter method.
- MonoSLAM clearly beaten by PTAM!

Visual SLAM Becomes Well Defined; some Important Innovations

- 2008 IEEE Transactions on Robotics special issue on visual SLAM (edited by Neira, Leonard, Davison)
- 2007 RatSLAM, Milford and Wyeth
- 2007 Comport, Dense visual odometry
- 2009 R-SLAM, relative bundle adjustment, Mei, Sibley, Cummins, Reid, Newman *et al.*

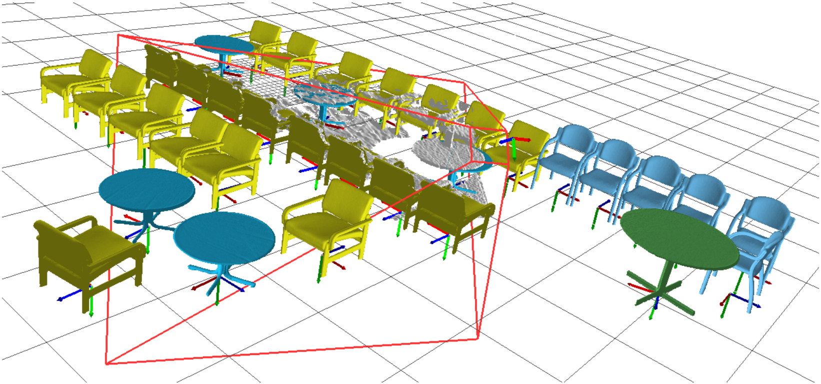
Dense SLAM Begins



- Around 2010, GPGPU enables real-time regularised dense reconstruction; PTAM tracking for Richard Newcombe's Live Dense Reconstruction with a Moving Camera paper.
- Dense tracking, DTAM (Dense Tracking and Mapping).
- 2010, Kinect opens the era of commodity high quality depth cameras, and KinectFusion leads to many other dense SLAM systems.
- Dense maps are ripe for semantic labelling and this is now starting to happen excitingly.

Towards Pure Object-Level SLAM

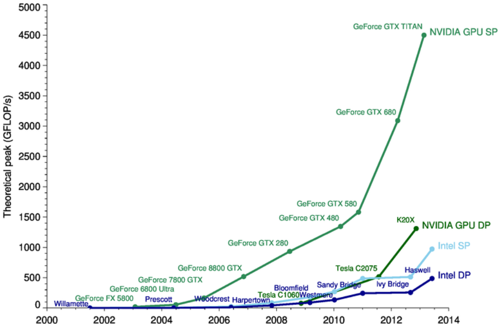
- SLAM++ (Salas-Moreno *et al.*2013): bring object recognition to the front of SLAM, and directly build a map at that level to benefit from strong predictions immediately.



- Predict, measure, update will be even stronger with object or even whole scene priors.

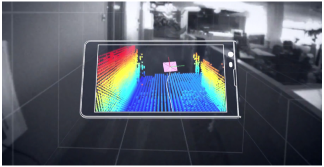
Brute Force Vision

- Rising processing allows increasingly computationally expensive computer vision algorithms to be brought into play in robot vision.
- Bundle adjustment; image retrieval; MVS regularised dense reconstruction; random forests, CNN and MRF.
- However... real applications need low power, compactness and real-world robustness.

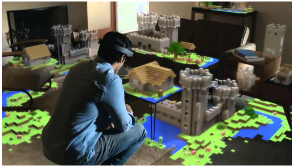


Modern Systems

Dyson 360 Eye



Google Project Tango



Microsoft HoloLens

- Positioning and reconstruction now rather mature. . . though I'd say it's still rather premature to call even that solved.
- Quality open source systems: LSD-SLAM, ORB-SLAM, SVO, KinectFusion, ElasticFusion.
- Commercial products and prototypes: Google Tango, Hololens, Dyson 360 Eye, Roomba 980.
- But SLAM continues. . . and evolves into generic real-time 3D perception research.

Modern Research Themes

- As algorithms, sensors and processors co-evolve and vision becomes an increasingly important driver, what do we imagine commodity systems of 2025+ will be capable of?
- Which research areas will we 'bring into' SLAM and how will be integrate them with SLAM's closed loop character?
- (My popular science book 'Robot Vision' hopefully finished and published soon!)

Chapter 6

SLAM with a Fast-Moving Single Camera

We have looked in detail at the 'SLAM' problem faced by a robot moving through an unknown world; and showed that for a ground-based robot such as a floor cleaner a combination of odometry and outward looking visual sensing over a wide field of view can make a good quality feature map. As multiple features are repeatedly measured from different robot positions, enough information is gained to enable estimation via probabilistic filtering of both the positions of the features in the world and the motion of the robot.

But in fact SLAM can work in much tougher conditions than this, and a big focus of my own research work has been on finding the limits. We will stick with computer vision as the main source of data but pare things right down to the case of a single camera which captures video continously as it moves through the world in a general, rapid manner — perhaps carried by a running person or a rapidly jumping or flying robot. Can we estimate motion and build a map in real-time from only this camera, without needing to know to what it is attached?

Getting a good result in this situation is the key to opening up all manner of applications for SLAM, and some of which are within fields we would not initially think of as robotics. Recently, there is great renewed interest in human-centric computer vision for wearable devices. Google

183

Dyson 360 Eye

engadget REVIEWS FEATURES GUIDES VIDEOS GALLERIES PUBLIC ACCESS GAMING ENGADGET LIVE

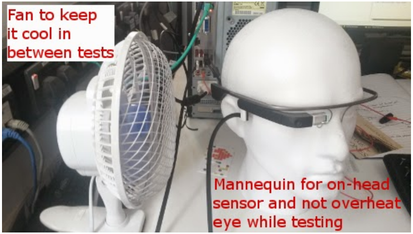
AdChoices

Dyson's $1,200 robotic vacuum is expensive, but also the best

by Mat Smith | @thatmatsmith | November 20th 2015 At 1:43pm

- Announced September 2014; now on sale in Japan; around the world soon.

The Need for Efficiency in Advanced Real-Time Vision



- Real applications need low power requirements, compactness and real-world robustness.
- Current GPUs run at 100s of Watts.

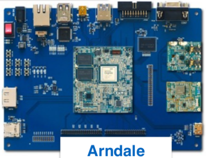
The Need for Efficiency in Advanced Real-Time Vision

TITAN
4998 GFLOPS
< 400 W

GTX 870M
2827 GFLOPS
< 100 W

TK1
404 GFLOPS
< 20 W

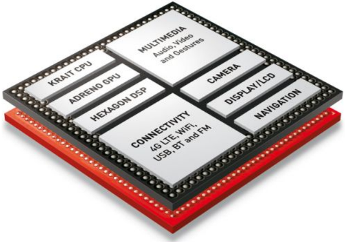
ODROID
170 GFLOPS
< 10 W

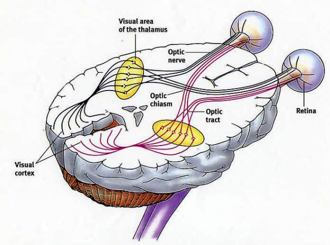


Arndale
87 GFLOPS
< 5 W

- We need 1000x power efficiency for truly capable always-on tiny devices; or to do much more with larger devices.

Embedded Vision 10 Years from Now





- Smartphone system-on-chip technology will provide the template for low power smart devices — and computer vision will be a major driver.
- CPUs, GPUs and increasingly specialised application-specific 'ASIC' chips.
- But how does the human brain achieve always-on, dense, semantic vision in 10W?
- I believe that the long-term way forward is to bring sensors, algorithms and processors together.