Live Tracking and Reconstruction from Video

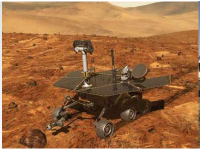
Andrew Davison
Department of Computing
Imperial College London

June 4, 2010

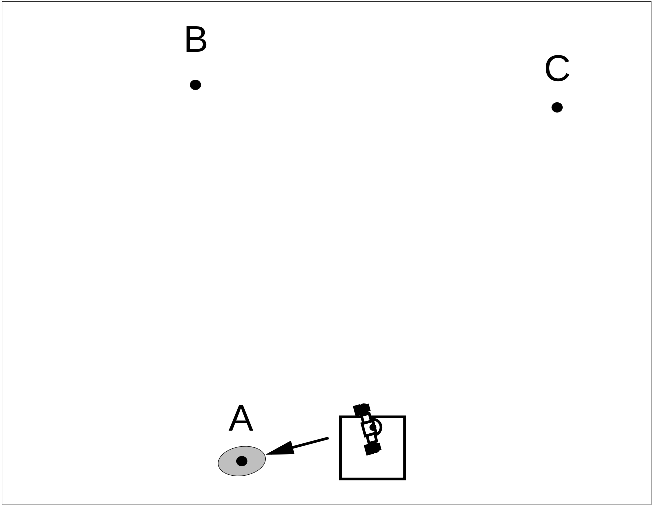
Robot Vision *in Real-Time*

Performance in robot vision is advancing *fast*. What are the reasons?

- Continued exponential increase in low-cost computer power.
- Bayesian probability theory: now widely agreed upon as the absolute framework for doing inference with real-world data.
- A wealth of well understood methods that *really work* are publicly available (well engineered algorithms or even code) and can be easily used to put systems together.

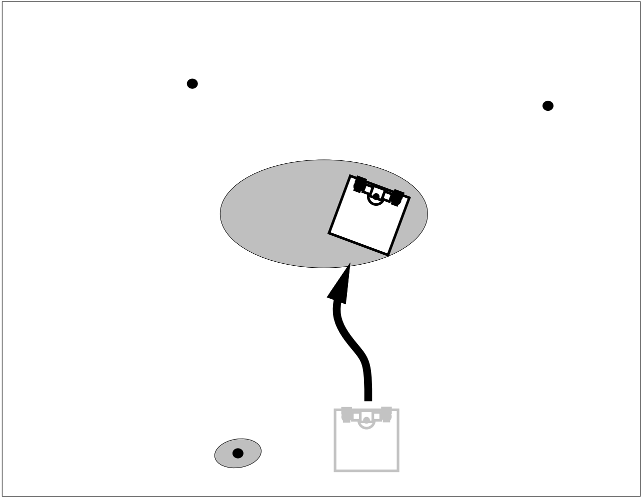




Simultaneous Localisation and Mapping



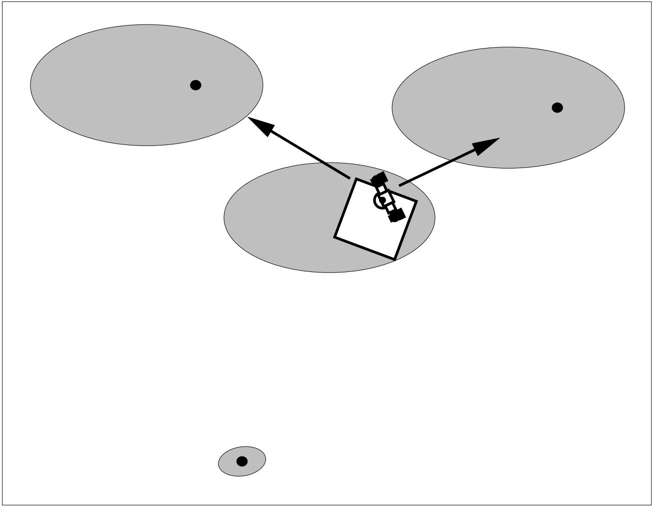
(a) Robot start (zero uncertainty); first measurement of feature A.

Simultaneous Localisation and Mapping



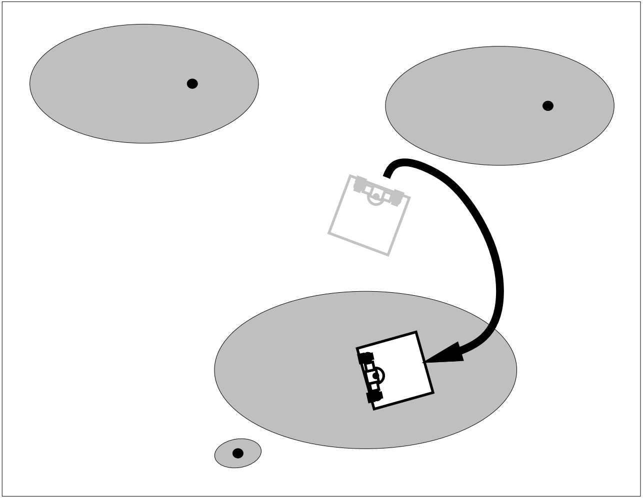
(b) Robot drives forwards (uncertainty grows).

Simultaneous Localisation and Mapping



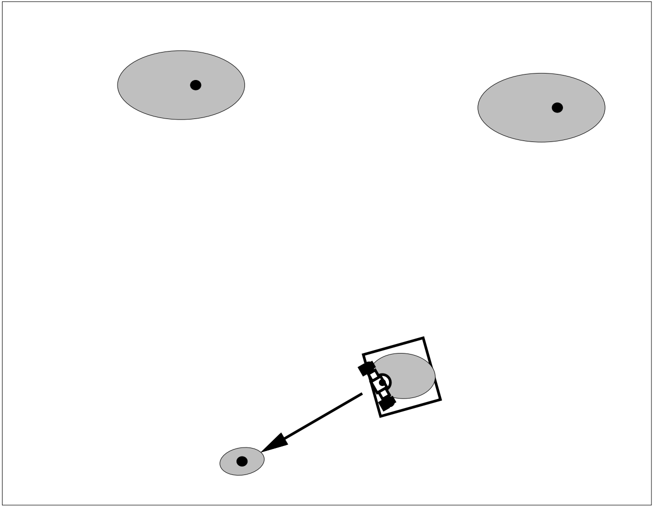
(c) Robot makes first measurements of B and C.

Simultaneous Localisation and Mapping



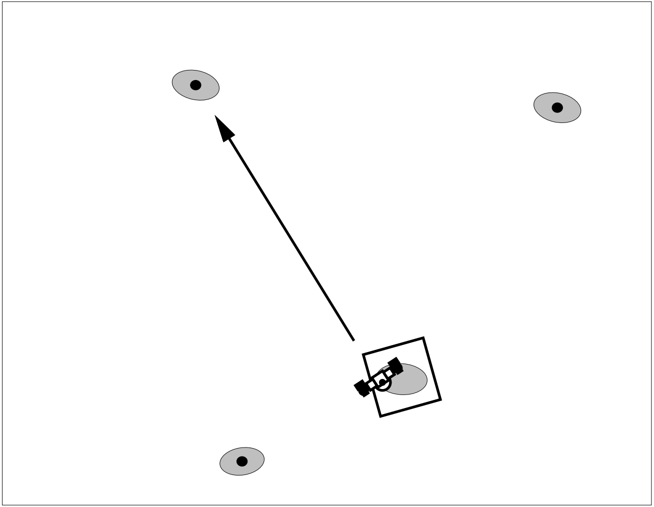
(d) Robot drives back towards start (uncertainty grows more)

Simultaneous Localisation and Mapping



(e) Robot re-measures A; *loop closure*! Uncertainty shrinks.

Simultaneous Localisation and Mapping



(f) Robot re-measures B; note that uncertainty of C also shrinks.

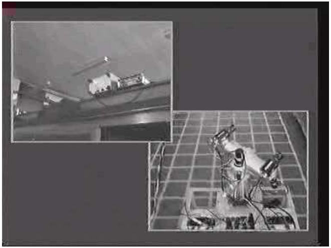
SLAM with First Order Uncertainty Propagation

$$\hat{\mathbf{x}} = \begin{pmatrix} \hat{\mathbf{x}}_v \\ \hat{\mathbf{y}}_1 \\ \hat{\mathbf{y}}_2 \\ \vdots \end{pmatrix} \quad , \quad \mathrm{P} = \begin{bmatrix} \mathrm{P}_{xx} & \mathrm{P}_{xy_1} & \mathrm{P}_{xy_2} & \ldots \\ \mathrm{P}_{y_1x} & \mathrm{P}_{y_1y_1} & \mathrm{P}_{y_1y_2} & \ldots \\ \mathrm{P}_{y_2x} & \mathrm{P}_{y_2y_1} & \mathrm{P}_{y_2y_2} & \ldots \\ \vdots & \vdots & \vdots & \end{bmatrix}$$

- Camera pose and map stored in single state vector and updated on every frame via a single Extended Kalman Filter.
- Full PDF over robot and map parameters represented by a single multi-variate Gaussian.

SLAM Using Vision: First Steps

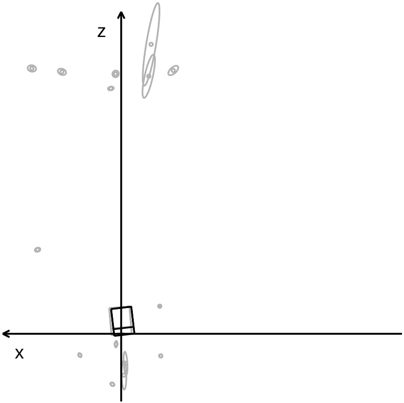
- Fixating active stereo measuring one feature at a time.
- 5Hz real-time processing (100MHz PC!).

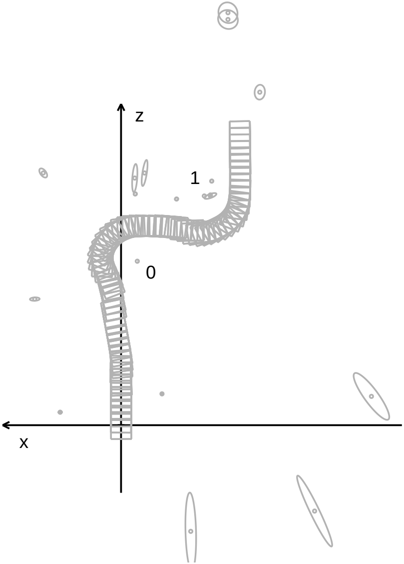


Davison and Murray, ECCV 1998, PAMI 2002.

SLAM Using Active Stereo Vision

Probabilistic Map Results





Monocular SLAM

- Can we still do SLAM with a single unconstrained camera, flying generally through the world in 3D?

- 30Hz or higher operation required to track agile motion.
- Salient feature patches detected once to serve as long-term visual landmarks.
- Landmarks gradually accumulated and stored indefinitely.

Modelling an Agile Camera

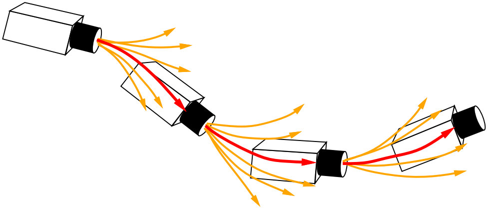
Camera state representation: 3D position, orientation, velocity and angular velocity:

$$\mathbf{x}_v = \begin{pmatrix} \mathbf{r}^W \\ \mathbf{q}^{WR} \\ \mathbf{v}^W \\ \omega^R \end{pmatrix}$$

Each feature state is a 3D position vector:

$$\mathbf{y}_i = \begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix}$$

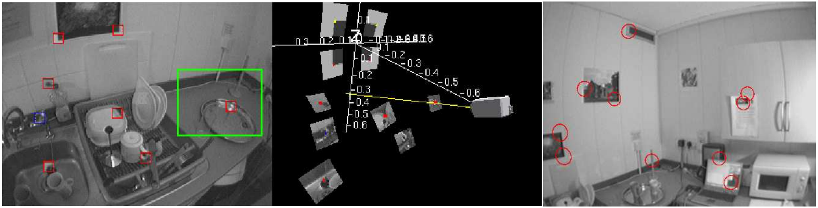
Prediction Step: A 'Smooth Motion' Model



Assume bounded, Gaussian-distributed linear and angular acceleration.

$$\mathbf{f}_v = \begin{pmatrix} \mathbf{r}^W_{new} \\ \mathbf{q}^{WR}_{new} \\ \mathbf{v}^W_{new} \\ \omega^R_{new} \end{pmatrix} = \begin{pmatrix} \mathbf{r}^W + (\mathbf{v}^W + \mathbf{V}^W)\Delta t \\ \mathbf{q}^{WR} \times \mathbf{q}((\omega^R + \boldsymbol{\Omega}^R)\Delta t) \\ \mathbf{v}^W + \mathbf{V}^W \\ \omega^R + \boldsymbol{\Omega}^R \end{pmatrix}$$

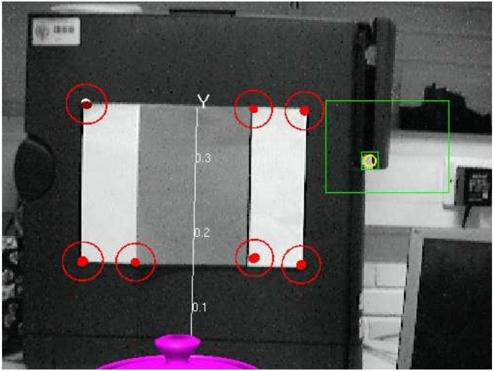
Measurement Step: Image Features and Active Search



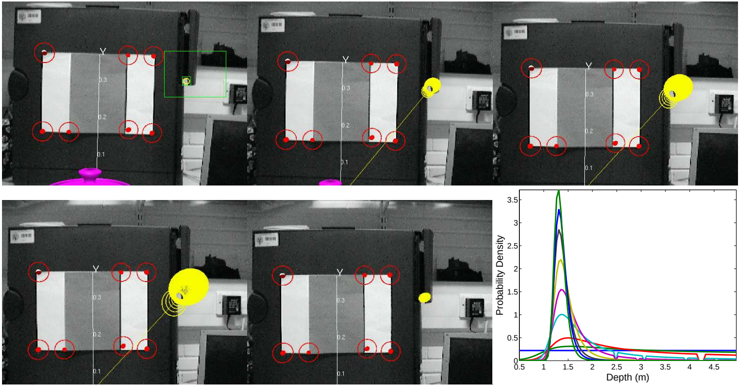
- Salient feature patches detected to serve as visual landmarks.
- Uncertainty-guided active search within elliptical regions.

Automatic Map Management

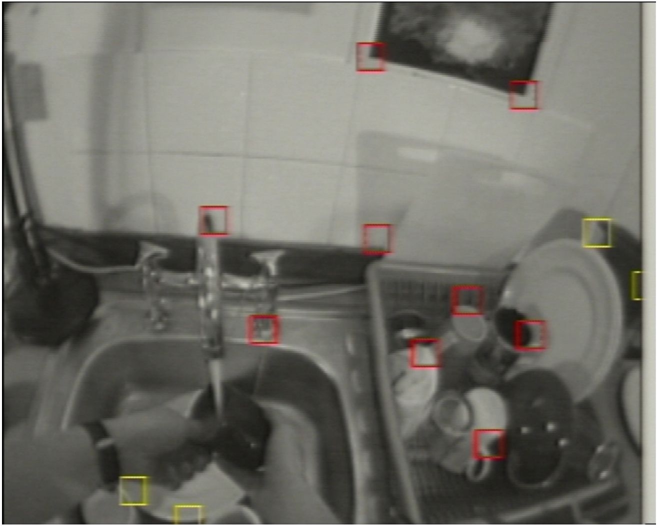
- Initialise system from a few known features.
- Add a new feature if number of measurable features drops below threshold (e.g. 10).
- Choose salient image patch from search box not overlapping existing features.



Monocular Feature Initialisation with Depth Particles



MonoSLAM

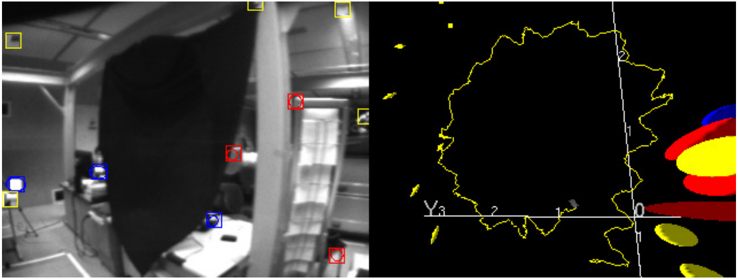


Davison, ICCV 2003; Davison, Molton, Reid, Stasse, PAMI 2007.

Application: HRP-2 Humanoid at JRL, AIST, Japan

- Small circular loop within a large room
- No re-observation of ‘old’ features until closing of large loop.

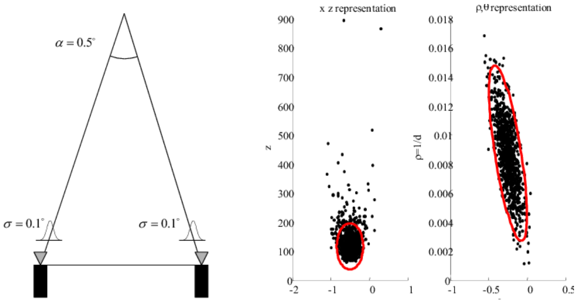
HRP2 Loop Closure



(Davison, Stasse, *et al.*, PAMI 2007)

Dealing with Distant Features

- In low parallax stereo reconstruction:



- Monte Carlo simulation reveals high Gaussianity in ρ, θ space where ρ is *inverse depth*.

Unified Inverse Depth Parameterisation for Monocular SLAM

A scene 3D point i is defined by the state vector:

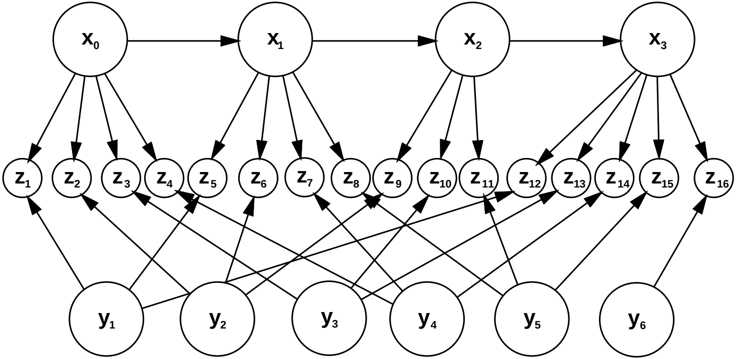
$$\mathbf{y}_i = \begin{pmatrix} x_i & y_i & z_i & \theta_i & \phi_i & \rho_i \end{pmatrix}^\top$$

which models a 3D point located at:

$$\begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix} + \frac{1}{\rho_i}\mathbf{m}\left(\theta_i, \phi_i\right)$$

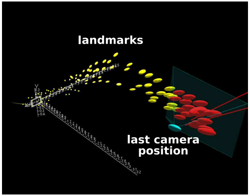
- Montiel, Civera, Davison, RSS 2006.

SLAM as a Bayesian Network

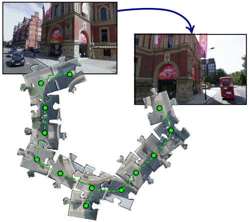


(See 'Probabilistic Robotics', Thrun, Burgard and Fox, MIT Press 2005.)

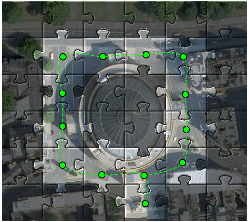
General Components of a Scalable SLAM Algorithm



Local Metric



Place Recognition



Global Optimisation

Local Metric Estimation: ‘Visual Odometry’

- Civera *et al.*, IROS 2009 (monocular EKF ‘forgetting filter’).
- High feature count provides local accuracy.

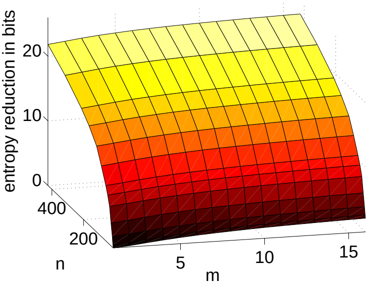
Global Topological: 'Loop Closure Detection'

- Angeli *et al.*, IEEE Transactions on Robotics 2008.

Optimisation: 'Pose Graph Relaxation'

- Keyframe-based spherical mosaicing, Lovegrove and Davison, 2010.
- Local tracking relative to keyframes with parallel global optimisation.

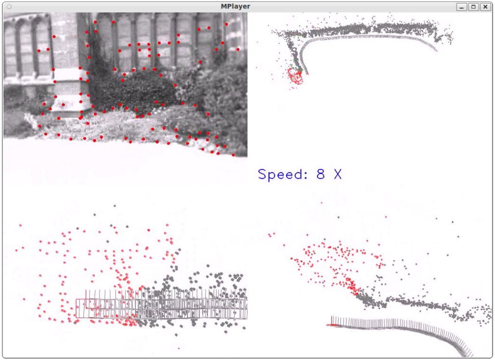
Real-Time Monocular SLAM: Why Filter?



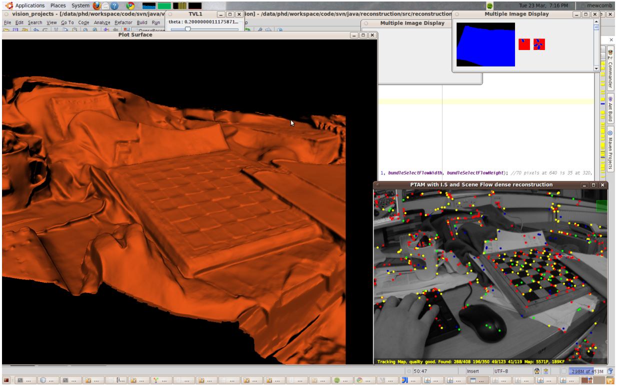
- Hauke Strasdat, J. M. M. Montiel and Andrew J. Davison, ICRA 2010.
- A comparison: filtering vs. keyframes + optimisation for monocular SLAM in terms of accuracy and computational cost.
- A clear winner with modern computing resources: keyframes + optimisation.

Large Scale Monocular SLAM using Optimisation

Scale Drift-Aware Large Scale Monocular SLAM (Strasdat, Montiel, Davison, Robotics: Science and Systems 2010).



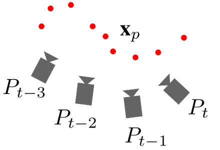
Live Dense Reconstruction with a Single Camera



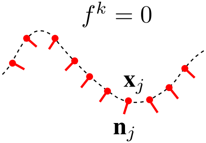
(Newcombe, Davison, CVPR 2010)

- During live camera tracking, perform dense per-pixel surface reconstruction.
- Relies heavily on GPU processing for dense image matching.
- Runs live on current desktop hardware.

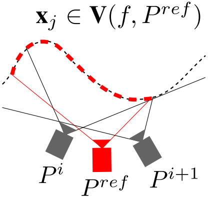
Live Dense Reconstruction with a Single Camera



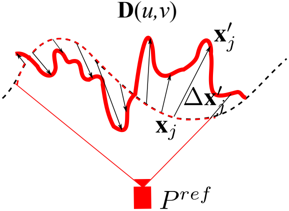
Point Cloud



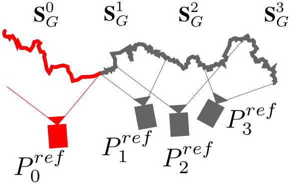
Base Surface



Bundle Matching



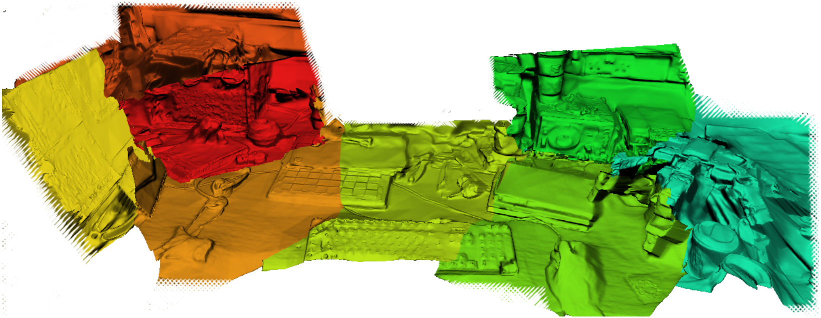
Dense Depth Map



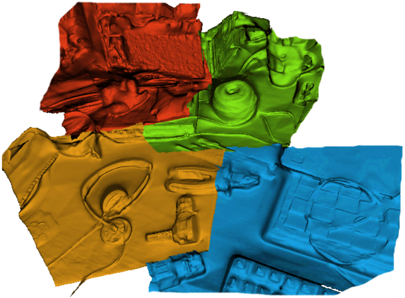
Depth Map Stitching

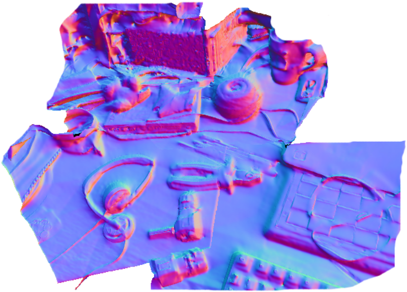
Live Dense Reconstruction with a Single Camera

- Multiple depths maps stitched live into single desktop model.

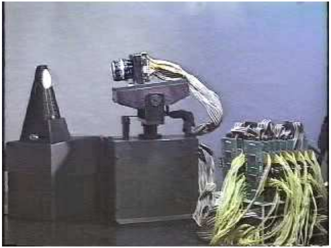


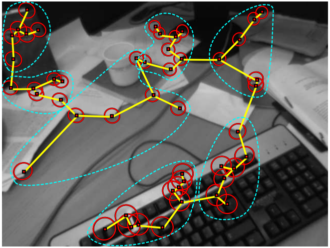
Live Dense Reconstruction with a Single Camera





Active Matching for Super-Efficient Tracking





(Scalable Active Matching: Handa, Chli, Strasdat, Davison, CVPR 2010)

- Many systems work well if the update rate can be kept high, because knowledge of continuity to permits local search: *tracking*.
- Active methods: updating online probabilistic estimates to drive sequential decisions.

SLAM for Scene Segmentation and Understanding

- Keypoint clustering and video segmentation, Angeli and Davison 2010.