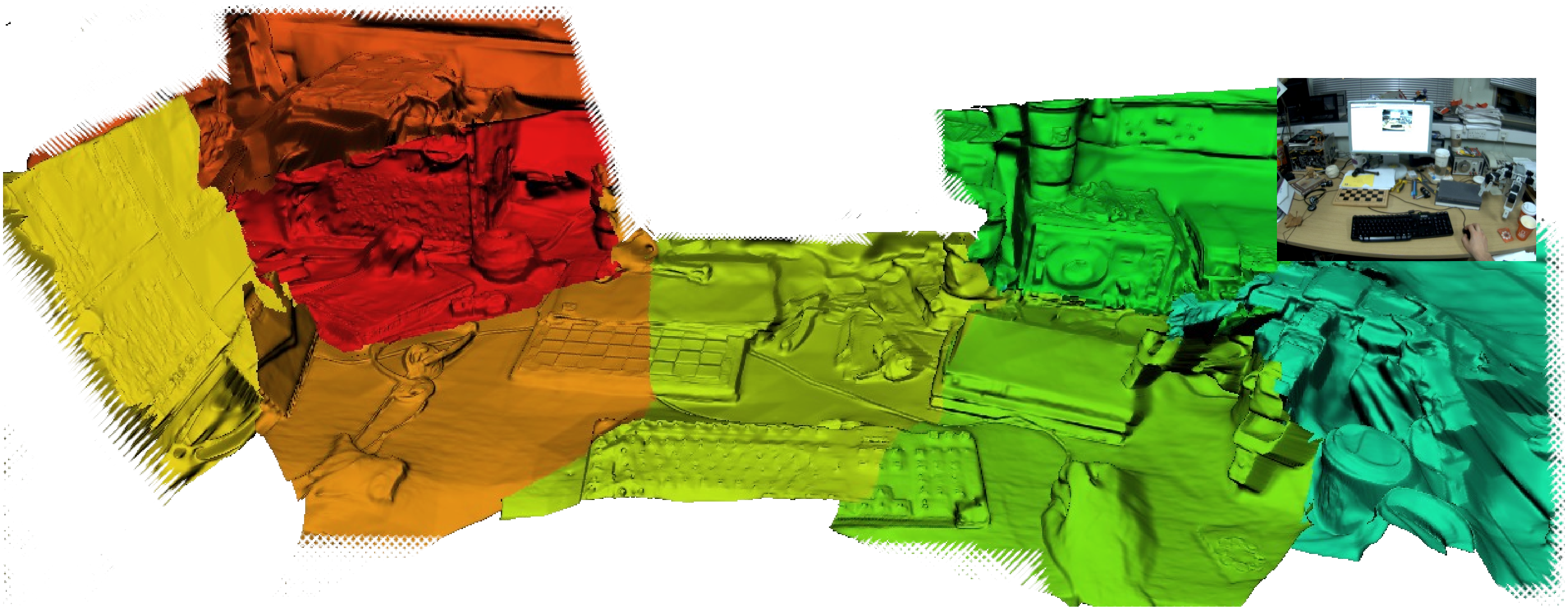
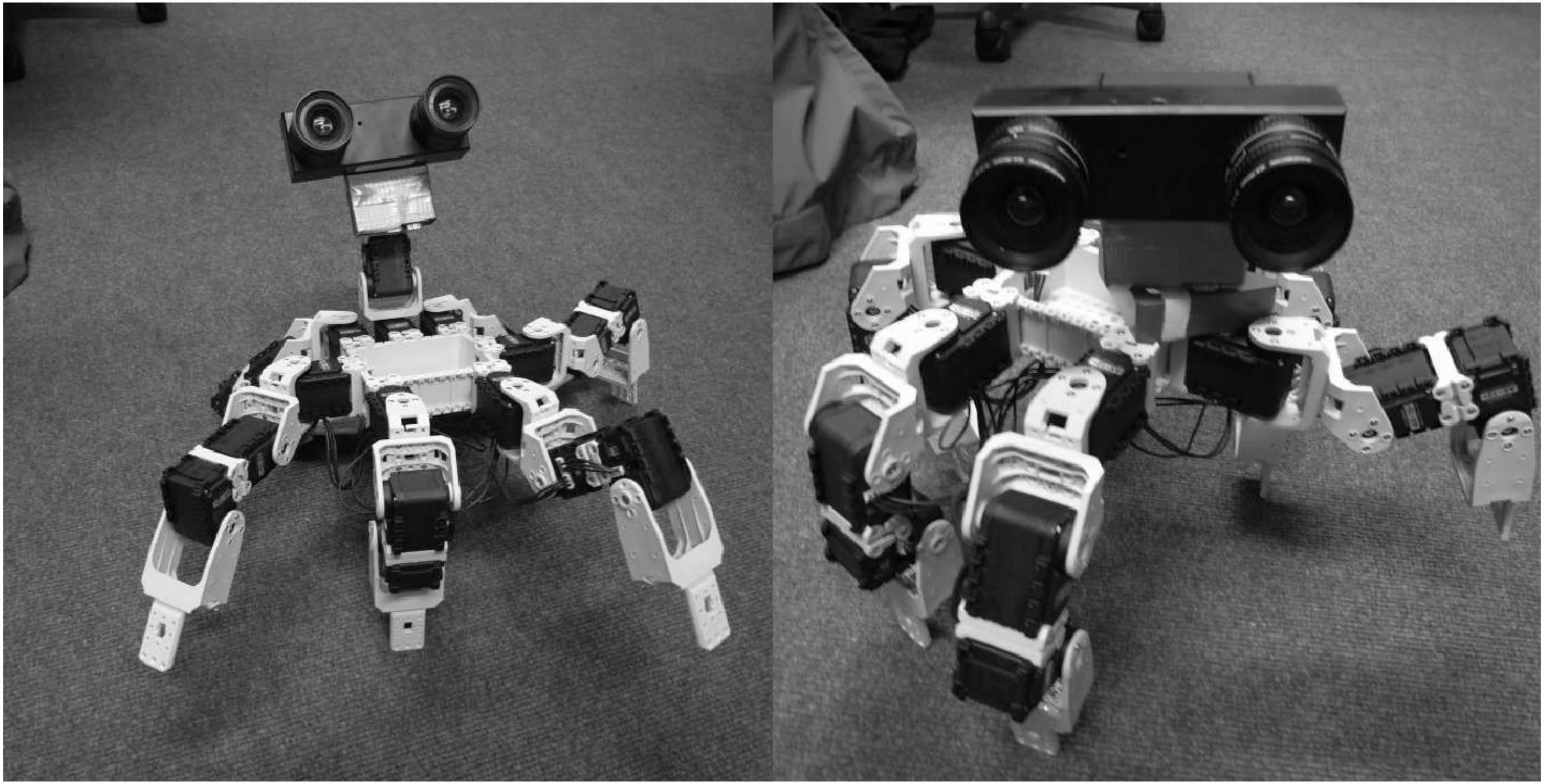
Live dense reconstruction using a single camera



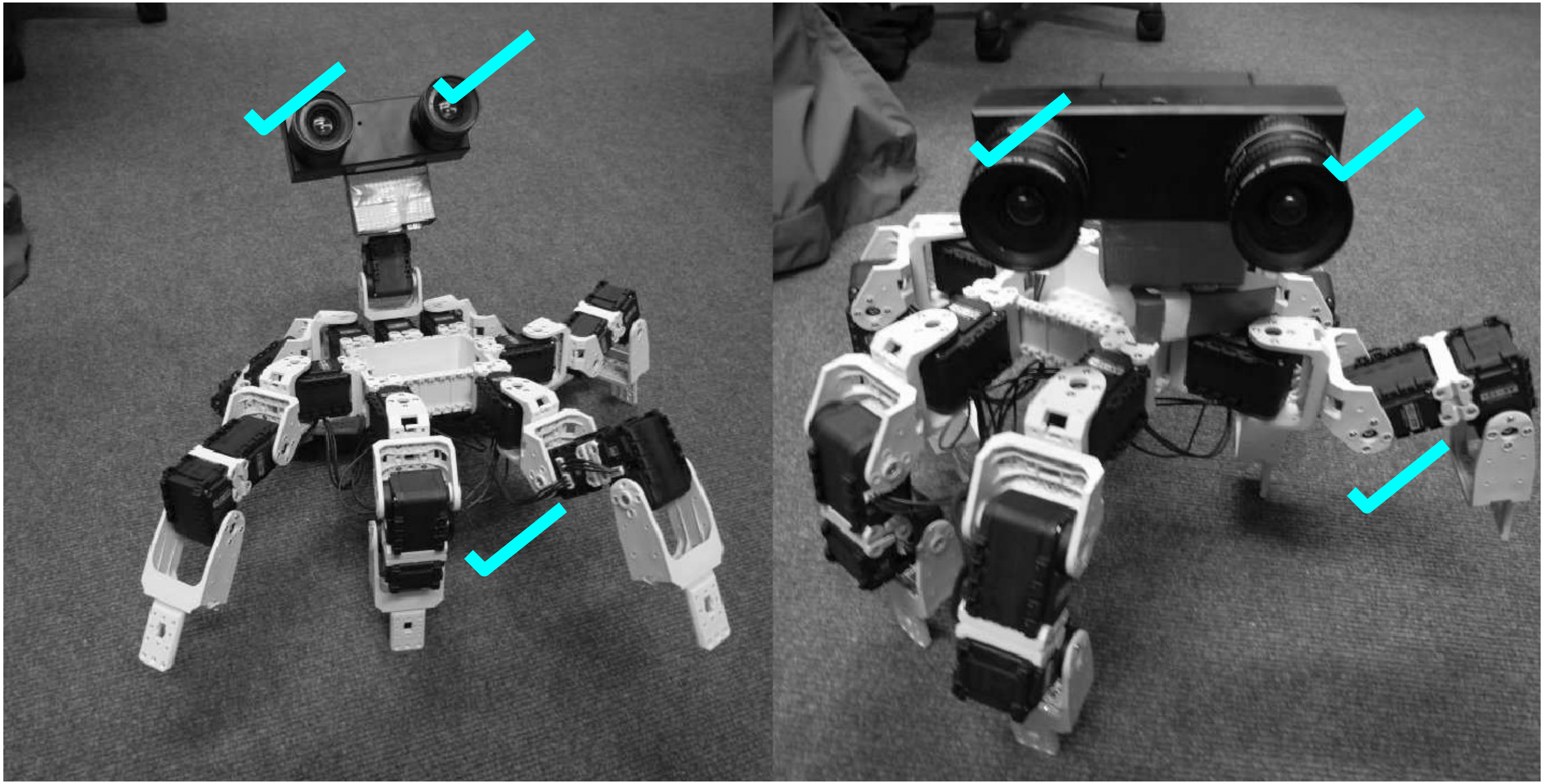
Richard A. Newcombe
[rnewcomb]@doc.ic.ac.uk

Motivation



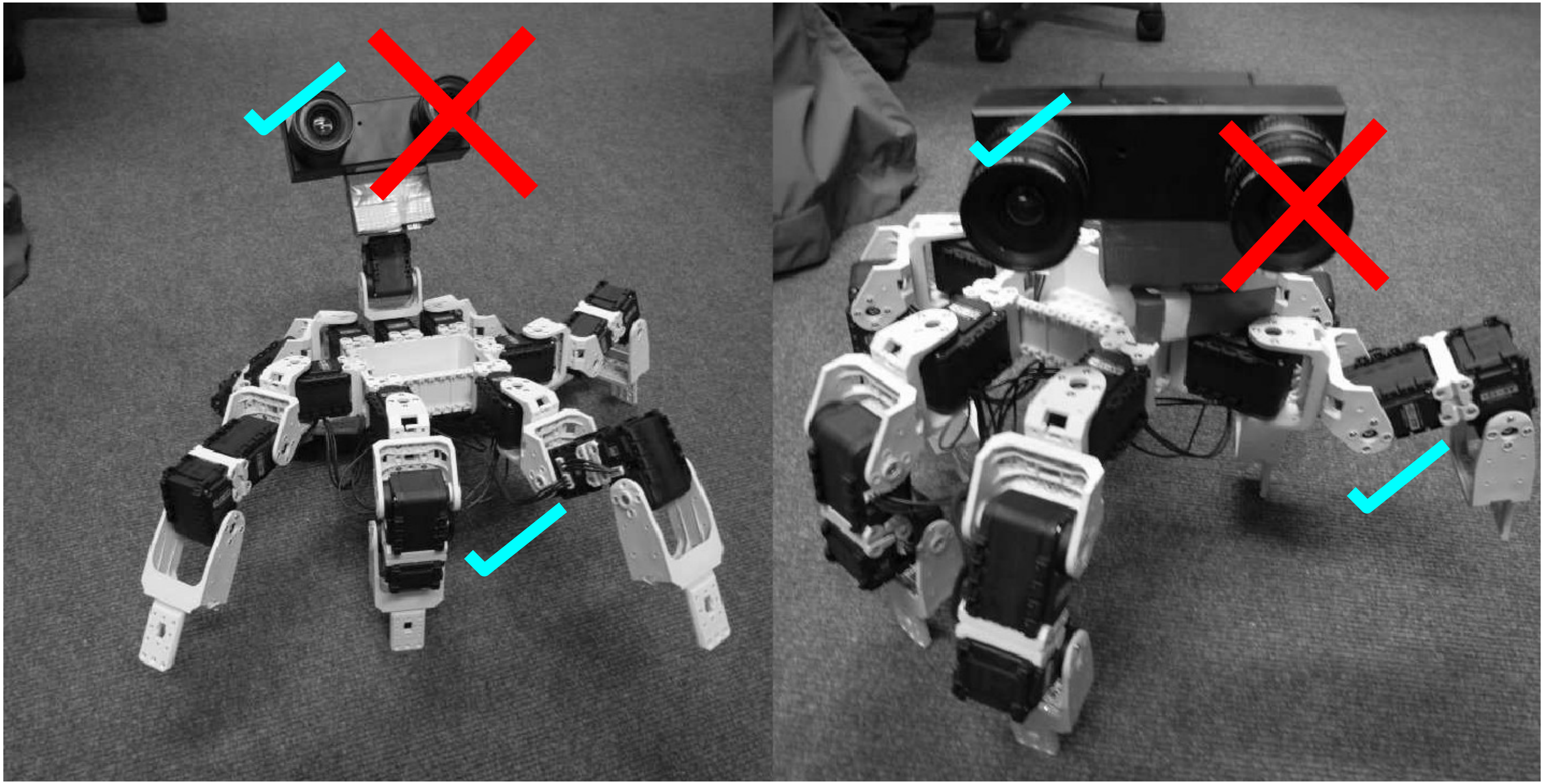
- Within the model based robotics paradigm
- Live reconstruction of scenes for physical prediction
  - Starting with **surface** geometry of (static) scenes.

Motivation



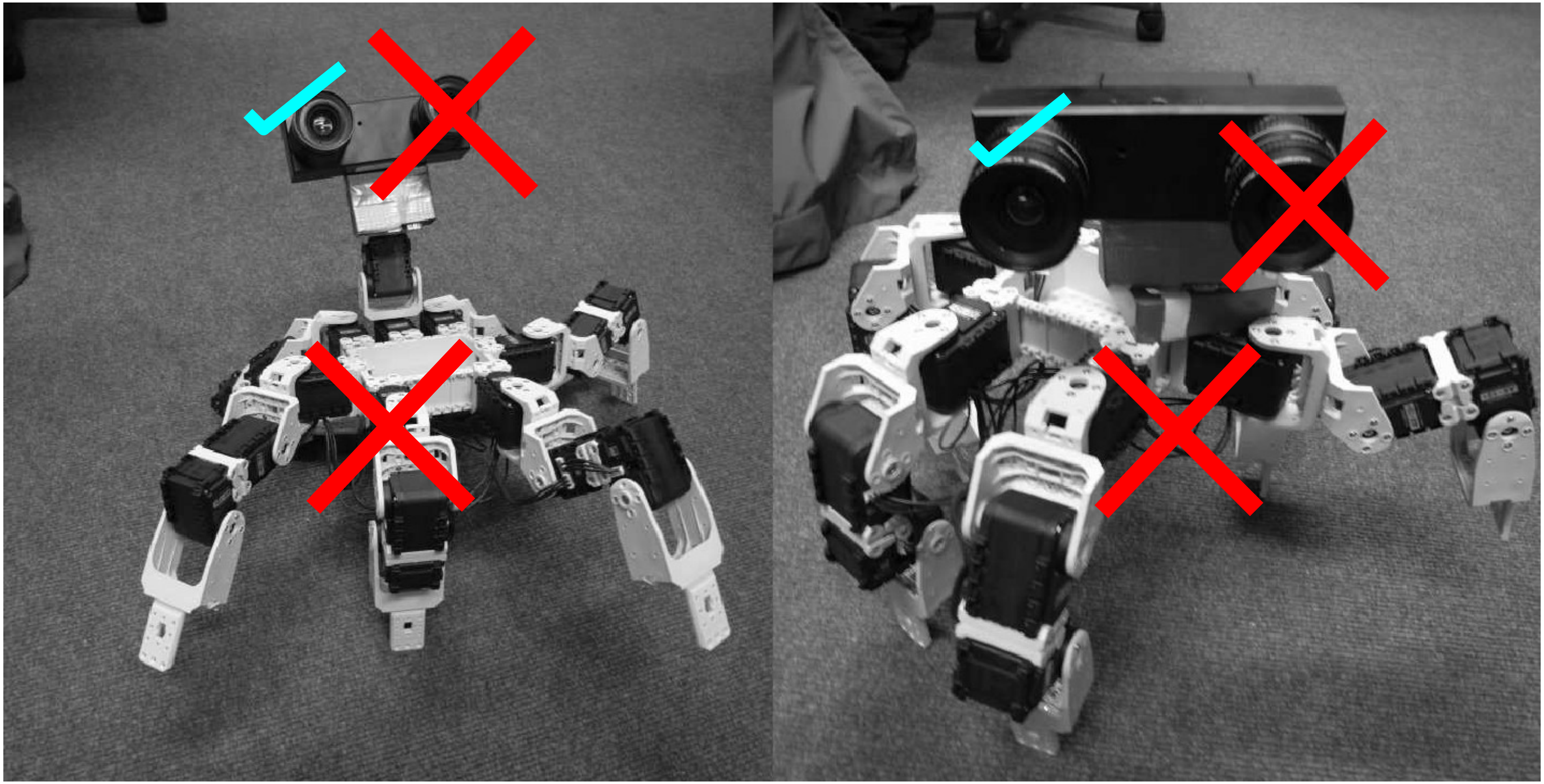
- What is the limit of what can be from a multisensory robot platform with multiple passive cameras?

Motivation



- What is the limit of what can be inferred from a single embodied (moving) passive camera?

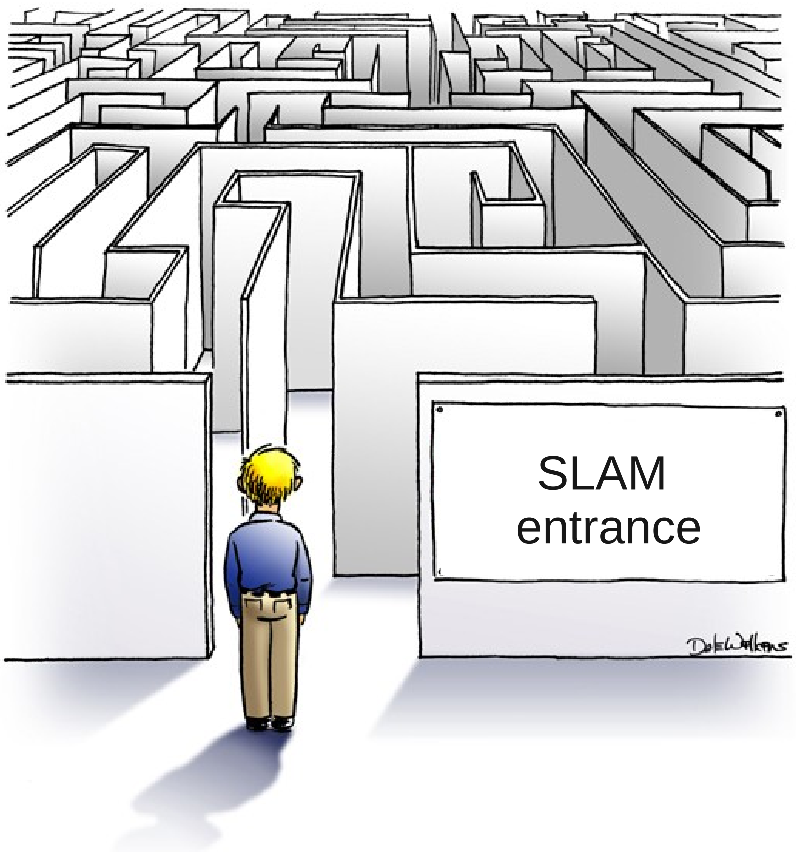
General Motivation



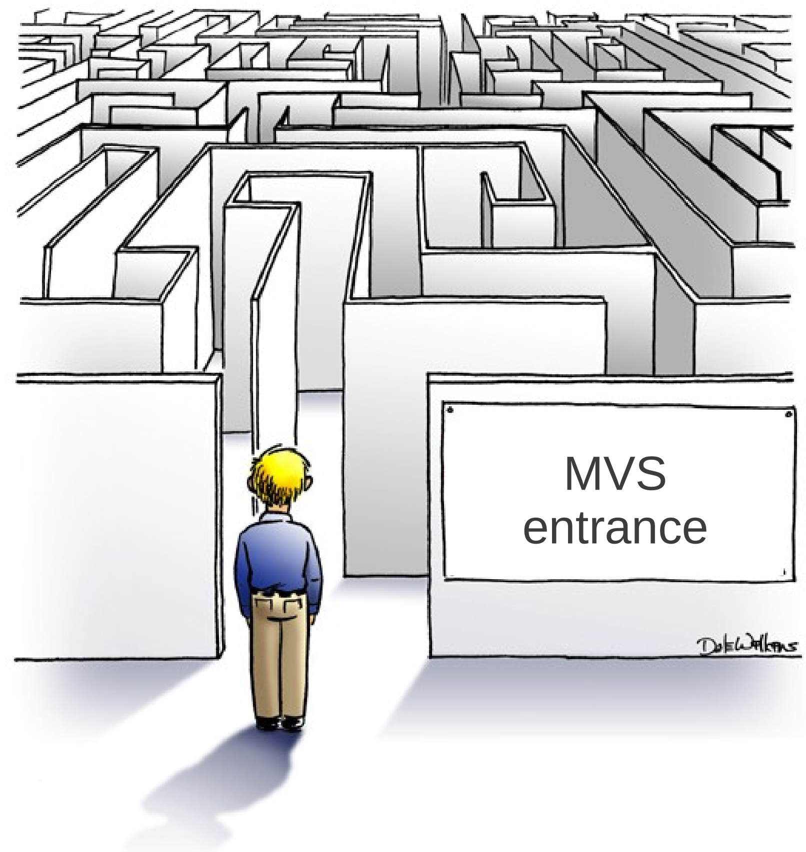
- What is the limit of what can be inferred from a single embodied (moving) passive camera?

In search of *a* path to live dense reconstruction...

Dense model out.



...



Passive images in.

Our (robotics) motivation, live incremental reconstruction

- Would like an dense estimate of the scene geometry that does not require batch processing of all views
  - A robot might need the reconstruction before it can plan where next to move
  - If incrementally estimated the dense geometry can provide useful information to the tracker (occlusion and normals of points)
  - A live pipeline allows the user in the loop to improve reconstruction by inspecting current result.

Our (robotics) motivation, live incremental reconstruction

- Associated **benefits** and **new challenges** over the standard MVS datasets:
- lots of data from a real time camera;
- small baseline between frames;
- user/robot in the loop to get better data if needed
- but the data will have motion blur and camera poses might not be perfect and may not be a global solution.

Talk Outline

Part 1

- From sparse structure and motion to dense correspondences
  - Dense surface geometry not utilised in the tracking loop

(Quick) Part 2 new work

- Towards fully dense SLAM with a single camera
  - No explicit feature tracking in mapping or tracking: Dense geometry back in the tracking loop

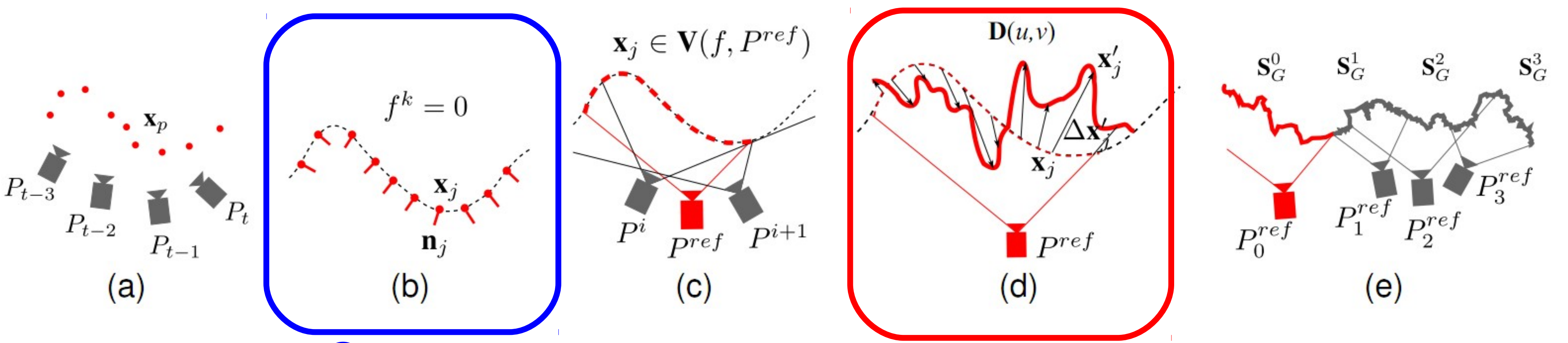
Modern real time structure from motion systems

Camera tracking and point clouds

- **Monoslam** [Davison et al ~ 01] – first demonstration of a realtime structure from motion system. Probabilistic joint estimation of point structure and camera motion.
- **PTAM** [Klein and Murray 07] – much improved camera tracking and denser point maps by splitting the tracking and mapping steps.
- Conclusion: that we can obtain live high quality camera motion but we still need surfaces not point clouds.
- *Enabling technology to allow development of live dense reconstruction systems*

System Overview (CVPR 2010)

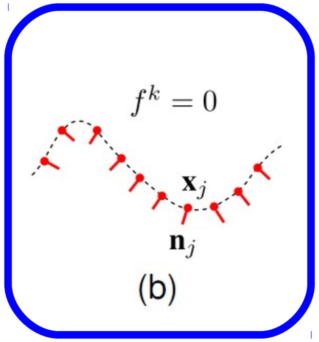
- (a) Utilise state of the art in real-time single camera SLAM.
- (b) ***For each each new key frame obtain a dense coarse reconstruction by globally fitting a function to the bundle adjusted point cloud***
- (c) Chose a bundle of frames (images and poses)
- (d) ***Obtain a depth map for a reference frame in the bundle***
- (e) Place the depth map into the global frame



Coarse geometry

Dense correspondences and surface

Coarse geometry



Coarse textured model from point cloud

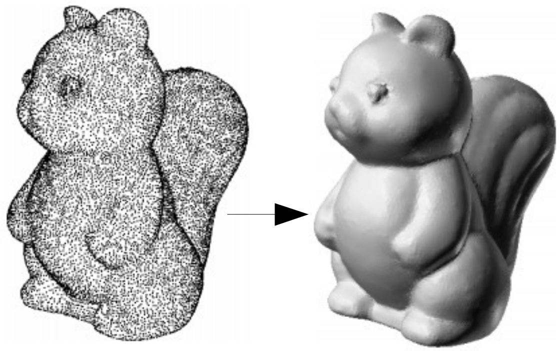
- Use the currently available point cloud to construction a coarse mesh using a fast surface fitting method

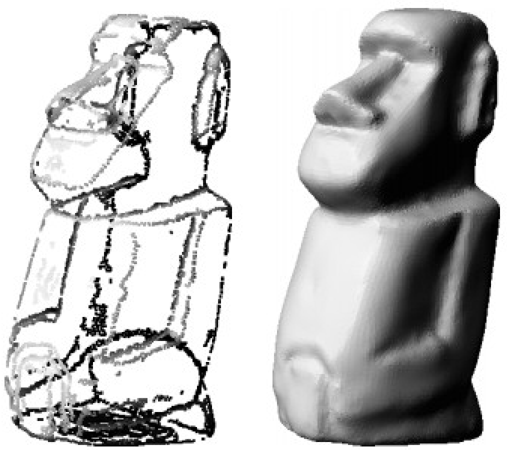
Why?

- We can use the coarse model to obtain a predicted view or correspondence field between two images.
- Can be useful in itself

Fitting a global function to the sparse point cloud

- Most 3D scattered data interpolation methods are designed for densely sampled sets.
- The point data obtained from PTAM is very sparse
- Feature samples exist only at high contrast corners of textured areas of the image
- We need a global function fitting to interpolate across large empty spaces.





Samples from Ohtake et al 2003

Multi-scale Compactly Supported Basis Functions (1)

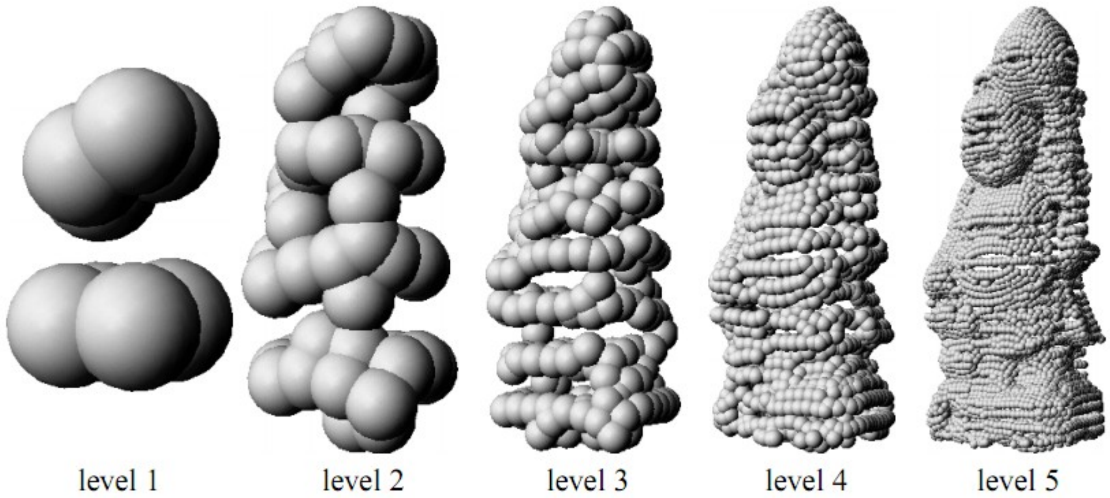
- A standard technique to function fitting is to define the dense surface solution implicitly and represent the solution as a sum of RBFs.
- Globally defined solution using RBFs with infinite extent enable large unsampled spaces to be interpolated
  - lead to a dense linear system that is prohibitively expensive to solve for large #points.
- Local function fitting leads to a sparse system, but has limited interpolation capabilities.

Multi-scale Compactly Supported Basis Functions (2)

- We use MSCSBF interpolation [Ohtake et al] that use “function valued” compact basis functions [H.Wendland 95] defined over multiple scales:

$$f^k(\mathbf{x}) = f^{k-1}(\mathbf{x}) + o^k(\mathbf{x}) \qquad (k = 1, 2, \ldots, M),$$

- Good global solution with local function fitting over multiple scales.
- Where a coarse level solution is the offsetting function for the next finer level: enables interpolation of data across large empty regions



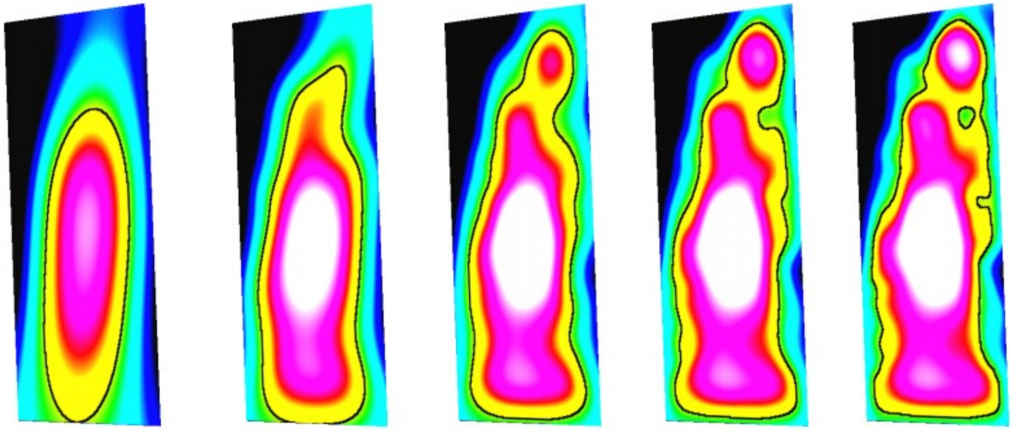
Hierarchical octree samples and interpolation function showing 0 level set from Ohtake et al 2003

Multi-scale Compactly Supported Basis Functions (2)

Each scale utilises the well defined CSBF ϕ weighted by a local quadric fit $g(\mathbf{x})$ of the local node mean point and normal: sparse system to solve.

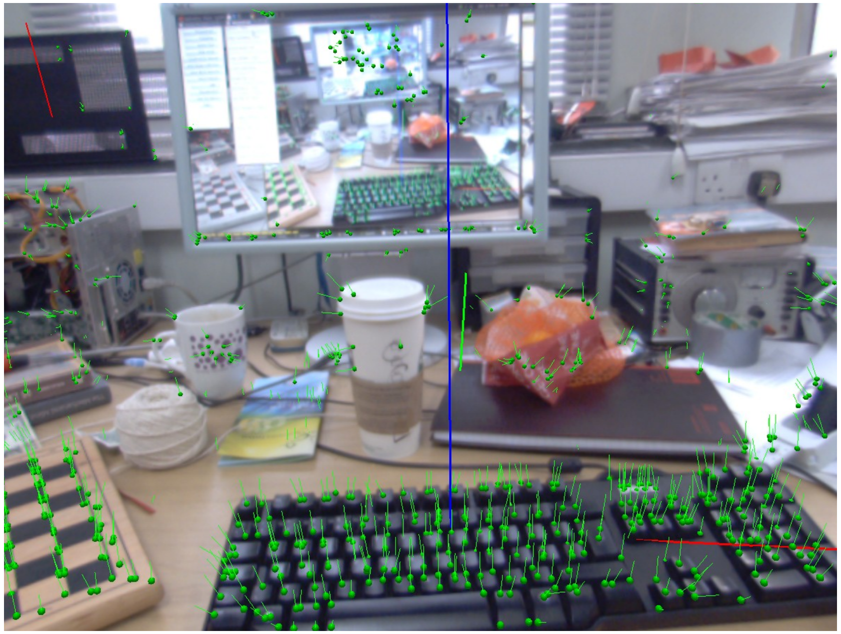
$$o^k(\mathbf{x}) = \sum_{\mathbf{p}_i^k \in \mathscr{P}^k} \left[g_i^k(\mathbf{x}) + \lambda_i^k \right] \phi_{\sigma^k}(\|\mathbf{x} - \mathbf{p}_i^k\|).$$

- Fast solution using gradient descent

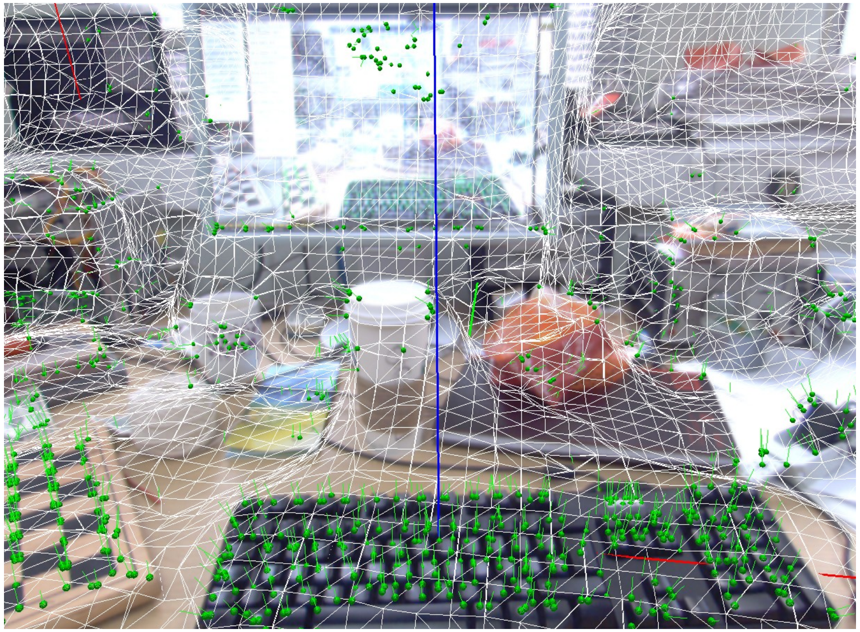


Hierarchical octree samples and interpolation function showing 0 level set from Ohtake et al 2003

MSCSRBF with PTAM point clouds



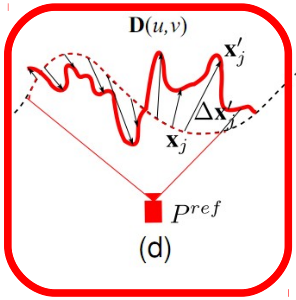
Bundle adjusted point cloud input from PTAM



Function fit with polygonisation of the surface level set using Bloomenthal's method.

- Alternatives include full tetrahedralisation of the point clouds using the visibility constraints [Qi Pan et al '09].

Dense correspondences



Obtain dense correspondences

- Dense correspondences between the reference and several target frames are used to obtain a per pixel point estimate by minimising the per pixel re-projection error.
- Our estimated correspondence field allows us to initialise a coarse to fine optical flow algorithm to give high accuracy dense correspondences.

Why

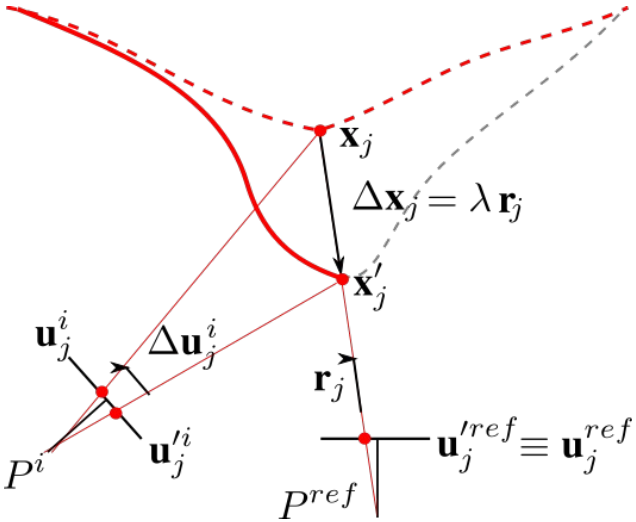
- We are interested in obtaining a reconstruction of surfaces even in low textured areas. In this case the binary point features used in the SLAM system will not initialise or track.

Simple least squares depth map given multiple correspondences

- Minimise sum of L2 norm re-projection errors for each individual pixel.

$$E_j = \sum_{i=1}^{n} \|\mathbf{P}^i(\mathbf{x}_j + \lambda_j \mathbf{r}_j) - \mathbf{u}_j\|_2^2$$

- Gradient descent by linearising around each λ to obtain a linear least squares problem



GPU friendly
Dense correspondences

- Most accurate correspondences are obtained using global optimisation methods that utilise a point wise data term with a spatial regularisation
- Assuming pixel value constancy data term* a general regularised flow field problem is:

$$\min_{\boldsymbol{u}} \left\{ \int_{\Omega} \psi(\boldsymbol{u}, \nabla \boldsymbol{u}, \ldots)\, d\Omega + \lambda \int_{\Omega} \phi\left(I_0(\boldsymbol{x}) - I_1(\boldsymbol{x} + \boldsymbol{u}(\boldsymbol{x}))\right)\, d\Omega \right\}$$

- *Note: the data term can be used on pre-processed input data (i.e. a structure texture decomposition)
- Original variational formulation for dense small displacement field by Horn and Shunck 1981 The ψ=L2 on $\mathrm{grad}(\boldsymbol{u})$ ϕ=L2 on the linearised data term.

TV-L1 optic flow

- Much work has gone into obtaining a robust, discontinuity preserving, solution by utilising the L1 norm on both the data and regularisation terms.
- We will give an overview of the solution we've employed from Zach, Pock, Cremers et al.
- Setting $\phi(x) = |x|$ and $\psi(\nabla \boldsymbol{u})) = |\nabla \boldsymbol{u}|$
- We have theTV-L1 energy

$$\min_{\boldsymbol{u}} \left\{ \int_\Omega |\nabla \boldsymbol{u}| \, d\Omega + \lambda \int_\Omega |I_0(\boldsymbol{x}) - I_1(\boldsymbol{x} + \boldsymbol{u}(\boldsymbol{x}))| \, d\Omega \right\}$$

- Use the linearised data term

$$\rho(\boldsymbol{u}) = I_1(\boldsymbol{x} + \boldsymbol{u}_0) + \langle \nabla I_1, \, \boldsymbol{u} - \boldsymbol{u}_0 \rangle - I_0(\boldsymbol{x})$$

TV-L1 optic flow Solution

- Quadratic splitting [Aujol et al] allows the data and regularisation term to be split:

$$\min_{\boldsymbol{u},\boldsymbol{v}} \left\{ \int_\Omega \sum_d |\nabla u_d| \, d\Omega + \frac{1}{2\theta} \sum_d \int_\Omega (u_d - v_d)^2 \, d\Omega + \lambda \int_\Omega |\rho(\boldsymbol{v})| \, d\Omega \right\}$$

- The first part is a well understood TV-L2 (ROF) model. Can be solved exactly using projected gradient via its dual formulation.
- The second part involves a point wise optimisation that be solved exactly too.
- Optimisation is now over 2 fields u and v. Both are a sum of convex functions
- Modern convex optimisation provides optimal solutions

* Rudin, L., Osher, S., and Fatemi, E. (1992).

Coarse to fine solution

- To ensure the linearised data term

$$\rho(\boldsymbol{u}) = I_1(\boldsymbol{x} + \boldsymbol{u}_0) + \langle \nabla I_1, \boldsymbol{u} - \boldsymbol{u}_0 \rangle - I_0(\boldsymbol{x})$$

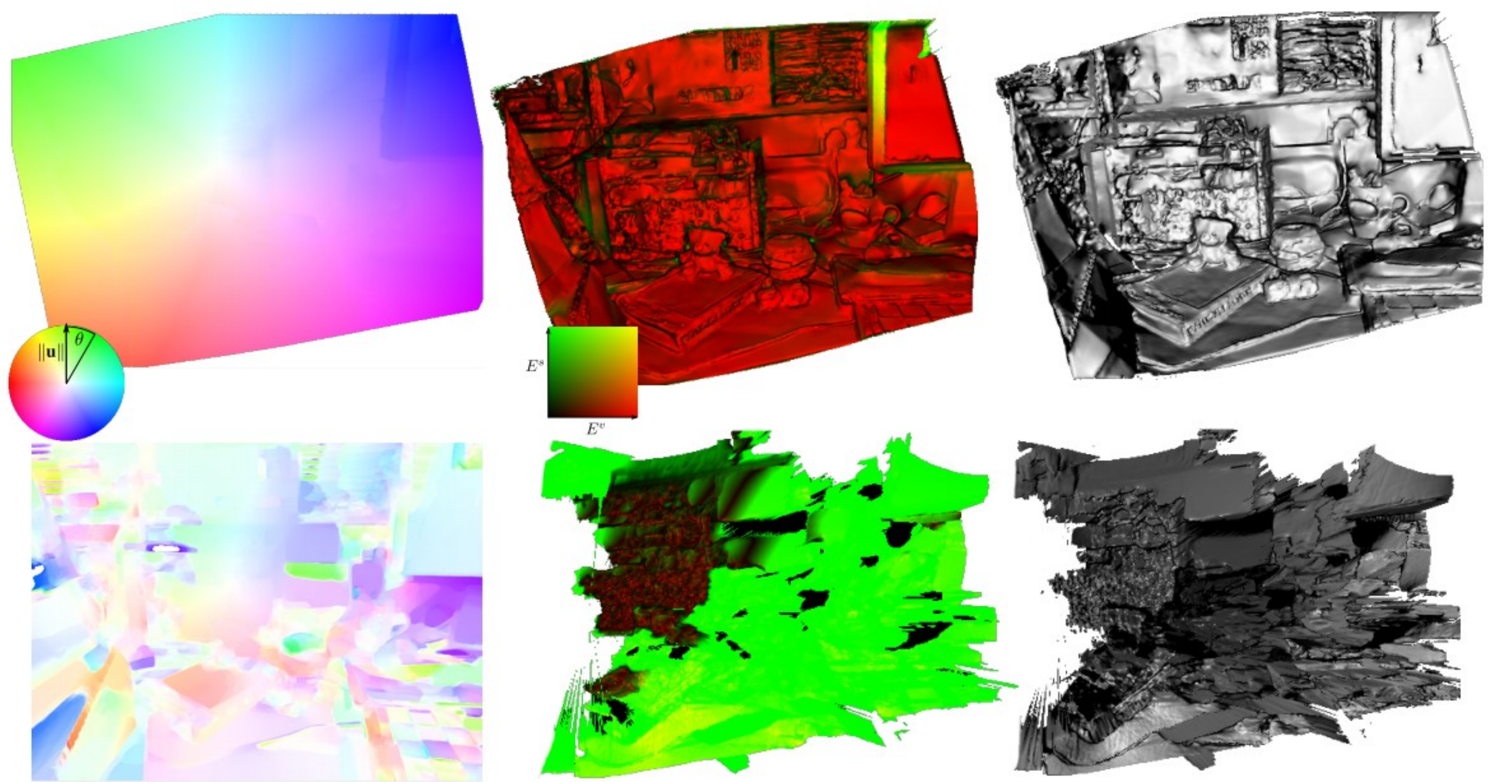
- is meaningful for larger displacements a coarse to fine solution is employed.
- Problems with a coarse to fine warping scheme?
- Initialise flow field with coarse geometry flow: Set the initial flow field to the predicted flow field computed from the coarse model given the estimated camera motion.

Result: Initialising the flow field with the dense prediction

- Resulting least squares depth map for correspondences with and without the dense prediction initialisation of the flow field.
- Despite using the coarse to fine scheme there is often too much rotational velocity between frames with agile camera to compute correspondences.

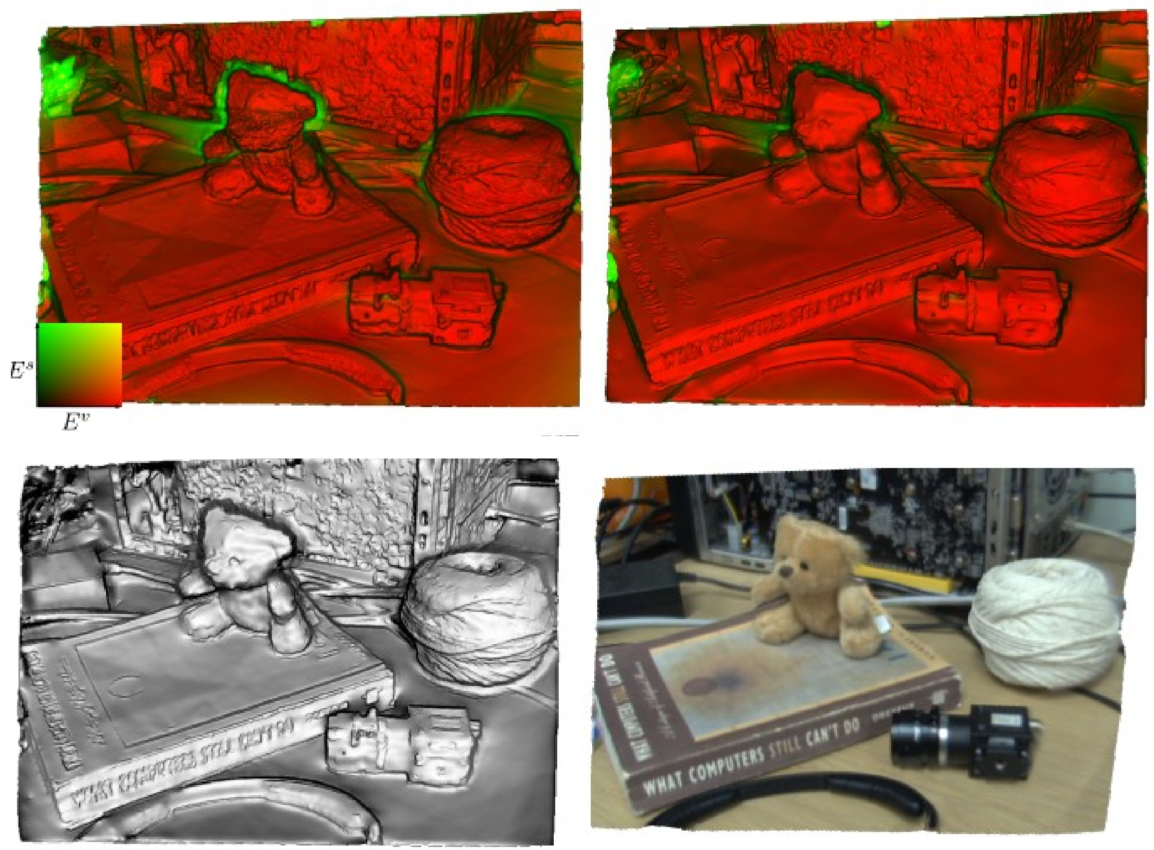
Correspondence field with Coarse initial geometry estimate.

Standard coarse to fine optic flow.



Iterating the solution

- Example result using 4 target images (4 flow fields generated), first solution shown from the base mesh initialisation followed by a second iteration using reinitialisation.
- Least squares residual colouring and visibility brightness.



Augmented reality application

- Reconstructing the desktop, and playing a simple car game.
- Dense surface reconstruction allows occlusion boundaries and physics (point clouds don't!)
- Simple ray-cast physics vehicle

Part 1 Conclusions and future

- This work was a departure from our usual monocular SLAM work where we work on binary associated point based feature tracking.
- The (projected) gradient descent methods and point-wise data terms are trivial to implement on **modern GPU technology.**

Quick Part 2 overview

- [Newcombe, Lovegrove and Davison '11] New work:
- Dense tracking and mapping: given the dense textured surface, align current live camera frame by full image allginment (2 ½D Lucas-Kanade style optimisation)
  - No explicit point feature matching.
  - For 6DOF using all pixels in the image makes a massively over determined system with increased robustness to fast motion.

Quick Part 2 overview

- [Newcombe, Lovegrove and Davison '11] New work:
- Exact Data term search in photometric cost volume: we are now utilising the idea that the convex regularisation can be used with any sampled data term [Steinbrücker et al]
  - No explicit correspondences between frames.
  - We can use 100s of images for a single depth map.
  - No need to linearise the photometric error term, so no coarse to fine warping needed.

Further work

- A single moving camera is a very rich sensor!
- We are forced to **model and utilise prior knowledge** about the scene.
- We are now moving towards modelling more in the scene
  - Lighting
  - Surface material properties.

References

G. Klein and D. W. Murray. **Parallel tracking and mapping for small AR workspaces.** In Proceedings of the International Symposium on Mixed and Augmented Reality (IS-MAR), 2007

Y. Ohtake, A. Belyaev, and H.-P. Seidel. **A multi-scale approach to 3D scattered data interpolation with compactly supported basis functions.** In Proceedings of Shape Modeling International, 2003.

J. Bloomenthal. **An implicit surface polygonizer**. In Graphics Gems IV, pages 324–349. Academic Press, 1994.

T. Pock **Fast Total Variation for Computer Vision**. PhD thesis, Graz University of Technology, January 2008.

A. Wedel, T. Pock, C. Zach, H. Bischof, and D. Cremers. **An improved algorithm for TV-L1 optical flow.** In Proceedings of the Dagstuhl Seminar on Statistical and Geometrical Approaches to Visual Motion Analysis, 2009.

C. Zach. **Fast and high quality fusion of depth maps**. In Proceedings of the International Symposium on 3D Data Processing, Visualization and Transmission (3DPVT), 2008

Q. Pan, G. Reitmayr, and T. Drummond. **ProFORMA:Probabilistic feature-based on-line rapid model acquisition.**In Proceedings of the British Machine Vision Conference(BMVC), 2009

M. Pollefeys, et al **Detailed real-time urban 3D reconstruction from video**. International Journal of Computer Vision (IJCV), 78(2-3):143–167, 2008.

Acknowledgements

- We thank Thomas Pock for discussions in understanding the primal-dual approaches and other useful practicalities of computing the variational solution.
- Also members of the robot vision Lab including Steven Lovegrove and Margarita Chli for useful discussions during the development of this work