6D Relocalisation for RGBD Cameras Using Synthetic View Regression

Andrew P. Gee
http://www.cs.bris.ac.uk/~gee
Walterio Mayol-Cuevas
http://www.cs.bris.ac.uk/~wmayol

Visual Information Laboratory
University of Bristol
Bristol, UK

With the advent of real-time dense scene reconstruction from handheld RGBD cameras [3], one key aspect to enable robust operation is the ability to relocalise in a previously mapped environment or after loss of measurement. Tasks such as operating on a workspace, where moving objects and occlusions are likely, require a recovery competence in order to be useful. For RGBD cameras, this must also include the ability to relocalise in areas with reduced visual texture.

Approaches from both the point cloud and monocular camera literature can be used for relocalisation on these types of densely reconstructed maps. Local feature-based methods extract distinctive geometric [4] or visual [6] features from the map and match them to features extracted from the current camera view of the world to estimate pose. In contrast, view-based methods construct geometric [5] or visual [1] descriptors for complete views of the map and match these to the current camera view.

This paper describes a view-based method for relocalisation of a freely moving RGBD camera in small workspaces. In contrast to related methods [1, 2], this method combines intensity and depth information from synthetic RGBD images to estimate full 6D pose at framerate using a regression framework.

The relocalisation problem can be formulated as a minimisation problem, where the goal is to find the set of camera pose parameters $\mathbf{x} = [\mathbf{t}, \ln(\mathbf{q})] \in \mathbb{SE}_3$, that minimises the distance measure

$$\mathbf{x} = \arg\min_{\hat{\mathbf{x}}} \|\mathbf{I}_0 - \mathbf{I}(\hat{\mathbf{x}}, \mathcal{M})\|, \tag{1}$$

where $\mathbf{t}$ is a 3D position vector, $\mathbf{q}$ is a quaternion representing rotation, $\mathbf{I}(\hat{\mathbf{x}}, \mathcal{M})$ is the synthetic view generated from the map $\mathcal{M}$ at pose $\hat{\mathbf{x}}$, and $\mathbf{I}_0$ is the true RGBD image from the camera. The j-th RGBD image $\mathbf{I}_j = [\mathbf{u}_j, \mathbf{v}_j, \rho_j, \mathbf{c}_j]$ is composed of n pixels, where $[u_{ji}, v_{ji}]$ are image coordinates, ρ_{ji} is the depth value, and c_{ji} is the grey intensity of the i-th pixel.

We treat the estimation of $\mathbf{x}$ in Eq. 1 as a general regression problem over a set of m synthetic views $\mathbf{I}_j$ and their poses $\mathbf{x}_j$, for $j = 1 \ldots m$. Using the Nadaraya-Watson estimator, we can approximate the camera pose $\tilde{\mathbf{x}}$ from the set of synthetic views as

$$\tilde{\mathbf{x}} = \frac{\sum_{j=1}^{m} \mathbf{x}_j K(\|\mathbf{I}_0 - \mathbf{I}_j\|/h)}{\sum_{j=1}^{m} K(\|\mathbf{I}_0 - \mathbf{I}_j\|/h)}, \tag{2}$$

where K is a kernel function centred on each sample with bandwidth h. In this case, we opt for a Gaussian kernel, such that

$$\tilde{\mathbf{x}} = \frac{\sum_{j=1}^{m} \mathbf{x}_j d_j}{\sum_{j=1}^{m} d_j}, \tag{3}$$

$$d_j = exp\left(-\frac{1}{n\alpha}\sum_{i=1}^{n}\left(\frac{(c_{0i} - c_{ji})^2}{\sigma_{ci}^2} + \frac{(\rho_{0i} - \rho_{ji})^2}{\sigma_{\rho i}^2}\right)\right), \tag{4}$$

where σ_c and σ_ρ are vectors of standard deviations in the intensity and depth computed per pixel over all of the sample views $\mathbf{I}_j$, for $j = 1 \ldots m$, and α is a scaling factor that controls the smoothness of the regression. The estimate $\tilde{\mathbf{x}}$ is therefore a normalised weighted sum, where the contribution of each sample view is determined by the normalised Euclidean distance between the sample view and the current camera view.

One key difference between our work and previous relocalisation systems is that, enabled by the recovered 3D map, we can generate novel synthetic views that have not been visited by the system during mapping and that are considered likely poses where relocalisation will be needed. This enhances the power of the sampling used by the regression framework but introduces the issue of knowing which views should be generated.

Here we have adopted the approach of randomly sampling poses around a pre-defined trajectory. For each of the m sampled synthetic views, a pose on the trajectory is randomly selected and a random Gaussian perturbation with 10° and 5.0cm standard deviation is applied. Synthetic views are resampled if fewer than 50% of the pixels intersect with the map. During the sampling process, the statistics for σ_c and σ_ρ, required by the regression algorithm, are also calculated and stored. It is also trivial to extend the set of synthetic views online, for example if the camera is tracked into a location not covered by the initial trajectory.

Figure 1: Examples of synthetic view regression relocalisation on four different test sequences. Images show ground-truth camera view (upper rows) and synthetic view generated from relocalised pose (lower rows).

The performance of the system is demonstrated in the paper by a comparison against visual and geometric feature matching relocalisation techniques and tested on sequences with moving objects and minimal texture. Some relocalisation results for the different test sequences are shown in Fig. 1. The results in the paper show that the method is both fast (< 80ms) and accurate (< 10cm, < 10° median error) and able to cope with small changes to the environment and low texture workspaces. The most common failure mode occurs when the camera moves to a viewpoint outside the set of synthetic view samples.

Our conclusion is that view-based relocalisation using synthetic RGBD images provides a feasible and useful alternative to slower, feature-based methods in small workspaces. This is particularly true in areas with low texture or low geometry, where the use of visual or geometric features alone is prone to failure, and in scenarios with continuously moving cameras, where the time required to relocalise is critical.

[1] M. Felsberg and J. Hedborg. Real-time view-based pose recognition and interpolation for tracking initialization. *Journal of Real-Time Image Processing*, 2(2-3):103–115, 2007.

[2] G. Klein and D. Murray. Parallel tracking and mapping for small AR workspaces. In *Proc. IEEE and ACM Int. Symp. on Mixed and Augmented Reality*, 2007.

[3] R.A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A.J. Davison, P. Kohli, J. Shotton, S. Hodges, and A. Fitzgibbon. KinectFusion: real-time dense surface mapping and tracking. In *Proc. Int. Symp. on Mixed and Augmented Reality (ISMAR)*, 2011.

[4] R.B. Rusu, N. Blodow, and M. Beetz. Fast point feature histograms (FPFH) for 3d registration. In *Proc. IEEE Int. Conf. on Robotics and Automation (ICRA)*, May 2009.

[5] R.B. Rusu, G. Bradski, R. Thibaux, and J. Hsu. Fast 3d recognition and pose using the viewpoint feature histogram. In *Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, October 2010.

[6] B. Williams, G. Klein, and I. Reid. Real-time SLAM relocalisation. In *Proc. IEEE Int. Conf. on Computer Vision (ICCV)*, 2007.