Efficiently Increasing Map Density in Visual SLAM Using Planar Features with Adaptive Measurements

José Martínez-Carranza
http://www.cs.bris.ac.uk/~carranza

Andrew Calway
http://www.cs.bris.ac.uk/~andrew

Department of Computer Science
University of Bristol, UK

The visual simultaneous localisation and mapping (SLAM) systems now in widespread use are based on localised point features [2, 4, 5]. Although effective in many respects, the approach has limitations when considering the density and efficiency of map representation. With a dense population of features, camera tracking can be robust, able to withstand significant occlusion and large changes in camera viewpoint. But this comes at a high computational cost, typically increasing quadratically with the number of features.

In this work we propose increasing map density by building in higher-order structure in the form of planar features. An important and novel aspect of the work is the manner in which the planar features are updated and used to localise the camera. We base our approach on an extended Kalman filter (EKF) monocular SLAM system developed by Chekhlov *et al*. [3]. This provides real-time estimates of the 3-D pose of a calibrated camera whilst simultaneously mapping the scene in terms of point based features.

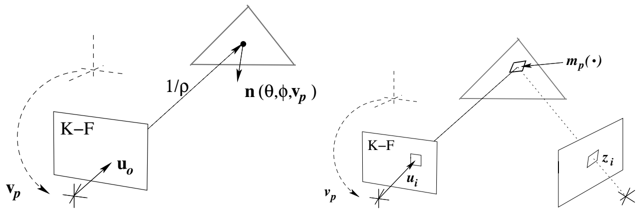
In order to incorporate planar structure into the real-time monocular SLAM we carry out three steps: **detection** of planar structure in the scene; **insertion** of planar features into the map; and **adaptive measurement** of the features. To apply the principle of adaptive measurement it is essential that planar features inserted into the map correspond to actual planar structure in the scene. For this we employ the method proposed by Martínez-Carranza and Calway [6], which uses an appearance model to detect planes defined by subsets of mapped point features (at least three points).

Having detected planar features in the scene these are inserted into the map using a suitable representation within the filter state. This has two components: plane parameterisation and the reference camera. The plane is defined by $\mathbf{y}_p = (\theta, \phi, \rho)$, where (θ, ϕ) defines the unit normal of the plane in polar coordinates in the reference camera and ρ is the inverse depth of the plane centre along the ray defined by $\mathbf{u}_o$, with the latter being stored at initialisation of the plane, see figure 1a. Insertion of the reference camera is done by augmenting the state with a copy of the current pose, i.e. $\mathbf{v}_p = \mathbf{v}$, and with initialised plane parameters $\mathbf{y}_p$ derived from the pose and the subset of mapped points which define the plane.

The reference camera serves two purposes: it references the plane in the SLAM coordinate system (with the associated uncertainties) and enables subsequent measurement of the planar feature using region based matching with respect to the current frame (key frame). To facilitate the latter the key frame image is also stored in the system. As illustrated in figure 1b, measurements for a planar feature are therefore assumed to take the following form:

$$\mathbf{z}_i = \mathbf{h}(\mathbf{v}, \mathbf{m}_p(\mathbf{v}_p, \mathbf{y}_p, \mathbf{u}_o, \mathbf{u}_i)) + \mathbf{w} \tag{1}$$

where $\mathbf{m}_p(\mathbf{v}_p, \mathbf{y}_p, \mathbf{u}_o, \mathbf{u}_i)$ is the point on the plane which projects to the salient point $\mathbf{u}_i$ in the key frame and $\mathbf{h}$ denotes perspective projection. The noise term $\mathbf{w}$ is assumed to be $N(\mathbf{0}, \mathbf{R})$.



(a) (b)

Figure 1: Parameterisation and measurement of planar features.

(a) (b) (c)

Figure 2: Planar features and adaptive measurement: (a) Adaptive measurements allow the selection of points for matching according to the camera pose, ensuring that predicted matches lie within the current camera view, whether at a distance (view A) or near to the scene (view B); (b-c) Selected point matches seen through the camera for the far and near cameras.

Having introduced planar features into the map, they can be updated and used to localise the camera via appropriate measurement. We base this on matching sets of salient points between the key frame and the current frame. Measurements are obtained by correlating small regions about salient points in the key and current frames, using predicted search regions from the filter measurement covariance and with correction for perspective distortion based on the mean planar structure. These are then used to update the filter state, and hence the plane parameters and camera pose, via the usual EKF equations [1].

Adaptive selection of the salient points in the key frame allows the measurements to be tailored according to the current state. The benefits of adaptive measurement are illustrated in figure 2 where the adaptation takes account of camera position. An example of a real time run is depicted in figure 3. Images show an external view of the camera and map, with planar features incorporated, and views through the camera showing projected planar triangulations and selected measurement points (blue circles).

Therefore, we conclude that the use of a higher order structure such as planar features corresponding to the physical world can be exploited efficiently through the adaptive measurements. The adaptive selection increases the effective density of the map whilst maintaining a fixed state size in the filter. The result is a computationally efficient system capable of robust localisation and mapping over a wide range of camera views.

Figure 3: Examples of typical runs of visual SLAM with planar features.

[1] Y. Bar-Shalom, T. Kirubarajan, and X.R Li. *Estimation with Applications to Tracking and Navigation*. 2002.

[2] A.J Davison, I. Reid, N. Molton, and O. Stasse. Monoslam: Real-time single camera slam. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 29(6):1052–1067, 2007.

[3] D.Chekhlov, W.Mayol-Cuevas, and A.Calway. Appearance based indexing for relocalisation in real-time visual slam. In *Proc. British Machine Vision Conf*, 2008.

[4] E. Eade and T. Drummond. Scalable monocular slam. In *Proc. Int Conf on Computer Vision and Pattern Recognition*, 2006.

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

[6] J. Martínez-Carranza and A. Calway. Appearance based extraction of planar structure in monocular slam. In *Scandinavian Conference on Image Analysis*, July 2009.