Unifying Planar and Point Mapping in Monocular SLAM

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

Planar features in filter-based Visual SLAM systems require an initialisation stage that delays their use within the estimation. In this stage, surface and pose are initialised either by using an already generated map of point features [2, 3] or by using visual clues from frames [4]. This delay is unsatisfactory specially in scenarios where the camera moves rapidly such that visual features are observed for a very limited period.

In this paper we present a unified approach to mapping in which points and planes are initialised alongside each other within the same framework. The best structure emerges according to what the camera observes, thus avoiding delayed initialisation for planar features. To do this we use a similar parameterisation to the one used for planar features in [3, 4]. The *Inverse Depth Planar Parameterisation* (**IDPP**), as we call it, is used to represent both planes and points. This IDPP is also combined with a point based measurement model where the planar constraint is introduced. The latter allows us to estimate and grow a planar structure if suitable, or to estimate a 3-D point if visual measurements do not support the constraint. The IDPP contains three main components: (1) A reference camera (**RC**); (2) the depth w.r.t. the RC of a seed 3-D point on the plane; (3) the normal of the plane.
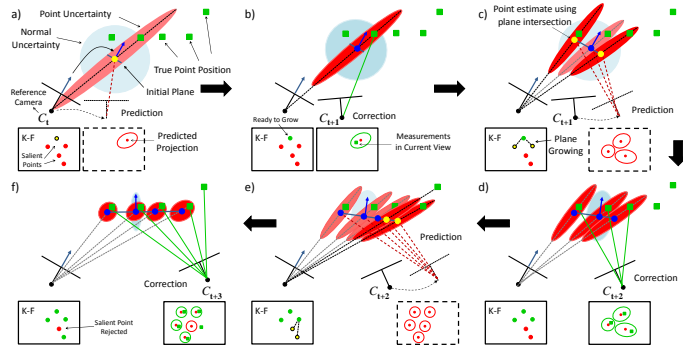

Figure 1: Undelayed plane intialisation by using the proposed framework.

Figure 1 shows the implicit initialisation process for a planar feature starting out as a single point, the seed point, that will evolve to a planar structure. Initially, the depth of the seed point and normal are quite uncertain, Fig. 1a. After the first correction, the plane contains a single point whose measurement contributes to reduce depth uncertainty, Fig. 1b. For the next step the system grows the plane by randomly selecting two salient points on the *Key Frame* (**KF**) attached to the RC, Fig. 1c. The selected pixels produce rays to infinity intersecting the plane estimate, thus producing 3-D points and their corresponding uncertainties. Therefore, the plane contains three 3-D points which can be measured to effectively correct the current camera pose, RC, depth and normal of the plane as shown in the Fig. 1d. The same process is carried out again as Fig. 1e-f illustrates, also showing a case where a salient point selected from the KF does not fulfill the planar constraint.
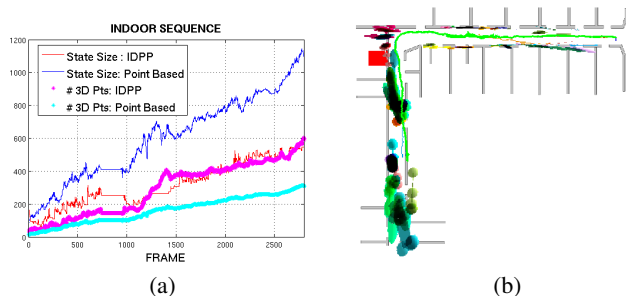

Figure 2: (a) State size and number of 3-D points obtained with IDPP and a points-only method for an indoor sequence, Fig 3. (b) Estimated map and uncertainties overlaid on a schematic plan for the same sequence.
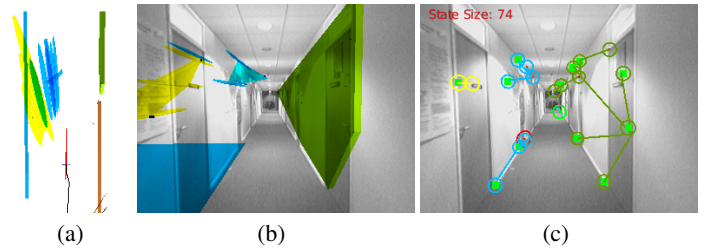

Figure 3: Planar features initialized with the proposed framework. (a) 3-D View of the world. (b) Projection of planar bounding box on the camera view. (c) 2-D Points Measurements in the Camera View, note measurements with a common plane are linked with a line.

To carry out the observation a template matching process is performed as for conventional point based mapping. Individual 3-D points within the planar feature are represented by templates of 11x11 pixels extracted from the KF. Each template is warped according to the plane estimate thus providing view invariance. Additionally, a subset of the 3-D points associated with the plane are characterised with spatial gradient descriptors as in [1] in order to recover from tracking failure. Hence, this hybrid representation allows fast camera tracking and robust camera relocalisation.

The proposed framework has been tested in several indoor and outdoors scenarios where the camera performs fast motion and always points ahead. The experiments demonstrate the benefits of the approach in terms of undelayed initialisation of planar features along side conventional point initialisation and efficient state representation. In fact, the latter allowed mapping of larger areas with a single filter under real time operation.

Figure 2a shows a comparison in the number of points and state size generated for an indoor sequence depicted in Fig. 3 for our approach and the standard point based mapping. A schematic plan of the corridor is shown in Fig. 2b where the final map and camera trajectory (in green) are overlaid.

Figure 4 shows an outdoor experiment where the camera moves approx. 230 meters. This experiment illustrates the significant degree of saving since the planarity of the scene is exploited to produce dense scene representation within a more compact state size, thus allowing real time operation even when a single filter was used. The final map and camera trajectory are overlaid on an aerial view in the same figure.

Figure 4: Outdoors example where the camera moved approx. 230 meters approximately. The system operated with a single filter.

[1] 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.

[2] A. Gee, D. Chekhlov, W.W Mayol, and A. Calway. Discovering planes and collapsing the state space in visual slam. In *Proc. British Machine Vision Conf*, 2007.

[3] J. Martinez-Carranza and A. Calway. Efficiently increasing map density in visual slam using planar features with adaptive measurement. In *British Machine Vision Conference*, September 2009.

[4] T. Pietzsch. Planar features for visual slam. In *In Proc. German Conference on Artificial Intelligence (KI 2008)*. Springer, September 2008.