# 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

**Abstract**

Mapping planar structure in vision-based SLAM can increase robustness and significantly improve efficiency of map representation. However, previous systems have implemented planar mapping as an auxiliary process on top of point based mapping, leading to delayed initialisation and increased overhead. We address this by introducing unified mapping based on a common parameterization in which both planar and point features are mapped directly, as and when appropriate according to scene structure. Specifically, no distinction is made between points and planes at initialisation - the 'best' representation emerges after matching has progressed, minimizing delay and making the detection of planar structure implicit in the method. We demonstrate the approach within an EKF monocular SLAM system and show its potential for efficient and robust mapping over large areas for both indoor and outdoor environments.

## 1 Introduction

This paper is concerned with efficient map representation in real-time simultaneous localisation and mapping using a single camera (monocular SLAM). That is, we are interested in tracking the 6-D pose of an agile camera (typically hand-held), whilst simultaneously mapping the 3-D structure of the surrounding scene, all in real-time. Rapid advances in monocular SLAM over recent years has yielded several systems which demonstrate impressive performance using a variety of processing frameworks [1, 3, 5, 6, 7, 8, 9, 10, 13, 16], with robust operation in the face of challenging motions and environments, including the ability to loop close and re-localise a lost camera .

However, the majority of these systems are based on mapping single, spatially distinct visual features, such as 3-D points or edgelets, with no encoded relationship between them. Although this approach has been effective, it remains an inefficient representation when higher order structure is present in a scene - features lying on such structures can be more efficiently parameterised in terms of those structures, hence reducing map size, memory requirements and computational load. This becomes important when attempting to increase map size, either for greater physical coverage or to increase map density. Planar structures in particular are prevalent in man-made environments and as previously noted in the structure from motion literature, this can be exploited in order to minimise redundancy of representation [2, 23]. The focus of the work described here is to do so in the context of a real-time filter-based visual SLAM algorithm.


BMVC 2010 doi:10.5244/C.24.43

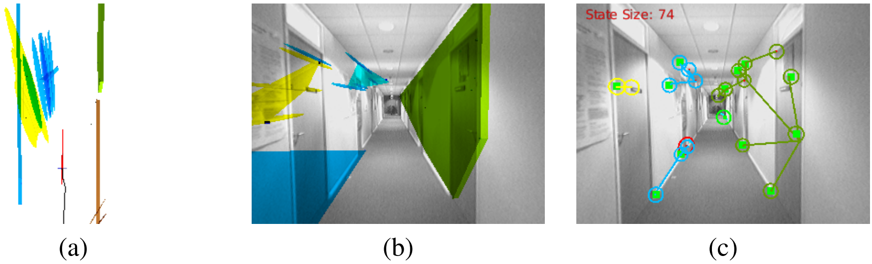



Figure 1: Planar features initialized with the proposed framework. (a) 3-D View of the world. (b) Projection of planar bounding box on the camera view. (c) 2-D Points Measurements in the Camera View, note measurements with a common plane are linked with a line.

As we summarise below, the use of planar structure in visual SLAM is not new. However, to date it has been divorced from single feature mapping, being regarded either as an add-on to supplement a point based map [11, 12, 14] or as an alternative to single feature mapping, as in [18, 21]. This is clearly unsatisfactory - ideally, planar mapping should sit alongside single feature mapping within a unified framework, with each being utilised as and when appropriate. To address this, we introduce a novel framework in which the mapping of points and planes is carried out using a common parameterisation, enabling either point or planar representations to emerge according to the underlying scene structure. This is achieved by also using common measurements - planes are defined in terms of sets of point features which are grown recursively from a common seed feature as mapping proceeds, with seeds that are unable to grow defaulting to single features. The approach has two important benefits: first, the initialisation of planar structure is not delayed, as it is in previous methods which require either the convergence of point features or the prior segmentation of planar regions; second, the detection of planar structure in the scene is implicit in the method, avoiding the need for a separate detection process. This second property is important, since it means that the algorithm only takes advantage of planar structure when it is safe to do so within the operation of the SLAM, hence avoiding the imposition of unrealistic planar constraints which could lead to inconsistency and instability.

We illustrate the technique within a standard extended Kalman filter (EKF) monocular SLAM algorithm, although we would emphasise that the approach could be adapted to other SLAM implementations - it is the principles and advantages of a unified mapping framework that we wish to demonstrate here. For the common representation of points and planes we utilise an adaptation of the inverse depth parameterisation (IDP) [17] and use a mix of correlation and descriptor based matching for the EKF measurements. In the following, Section 2 provides an overview of the method and implementation details are given in Section 3. Experimental results for both indoor (as shown in Fig. 1) and outdoor environments are then presented in Section 4. Both illustrate the potential of the method for efficient map representation, producing significant reductions in EKF state size compared with that for a points only approach whilst retaining robust operation, including camera relocalisation.

## 1.1 Related Work

Work on using planar structure in visual SLAM can be split into two categories: methods that extract planes from previously mapped points and those that map planar structure directly. In the former category, Gee *et al.* [11, 12] detect the presence of planar structure within sets of mapped points in an EKF framework and re-parameterise the filter state in order to exploit

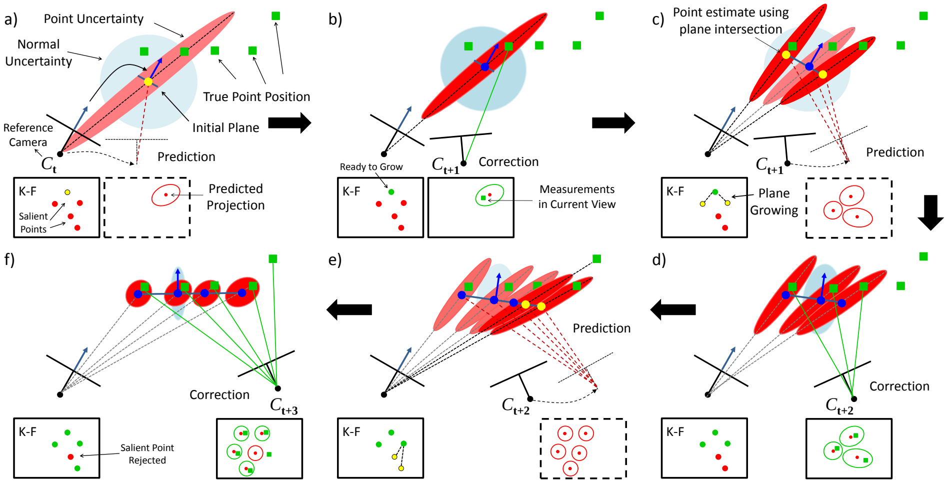



Figure 2: Undelayed plane intialisation by using the Inverse Depth Plane Parameterization and a 2-D point based measurement model.

the planarity constraint. This was extended by Martinez-Carranza and Calway [15], who combine appearance information with geometry to detect physical planes. They also demonstrate the use of adaptive measurement via the planar constraint to increase map density [14], analogous to the use of 'hallucinated correspondences' in [23]. An alternative approach is taken by Pietsch [18], who demonstrates direct mapping of planar structure within an iterative EKF based on pixel alignment, albeit with prior knowledge of plane location and extent. Similarly, Silveira *et al.* [22] also use an iterative alignment method to map planar structure within segmented regions using an optimisation framework for SLAM. By way of contrast to both of the above categories, Castle *et al.* [4] use prior knowledge of plane appearance gained off-line to detect the presence of known planes and hence insert planar geometry into an EKF state to constrain feature matching.

## 2 Unifying Point and Planar Mapping

We begin by giving an overview of the unified feature mapping which is the main contribution. We assume that we have a core SLAM algorithm, which in our case is an EKF, and that for each frame this seeks to initialise new features in regions not covered by the current map. Since our measurements for both points and planes will be based on matching local patches, potential features are those corresponding to patches with high local saliency.

Our common feature parameterisation has three components: the 6-D pose of a reference camera which defines a local coordinate system; the depth w.r.t. the reference camera of a seed 3-D point feature; and a 3-D normal vector. The mapping process starts by initialising the seed point feature within the local coordinate system and in subsequent frames seeks to initialise coplanar point features to associate with the seed and so estimate the normal vector. This process continues in order to 'grow' a planar feature within the map. If coplanar points cannot be found then the feature automatically defaults to a point feature corresponding to the original seed point. Hence no distinction is made between points or planes at initialisation, both are dealt with in a common framework.

A simple example illustrating the process is shown in Fig. 2. This shows a scene consisting of 5 points (green squares), four of which lie on a plane. First, the seed point is initialised corresponding to a salient patch (red dots) and defined within the local coordinate system given by the pose of a reference camera. As shown in Fig. 2a, we have a mean estimate for the seed (yellow square), with associated covariance indicating the depth uncertainty, and a large isotropic uncertainty on the normal vector since at this stage it is undetermined. In the next frame, we take a measurement for the seed point and update its position estimate as shown in Fig. 2b.

With a successful update, the seed becomes 'ready to grow' and neighbouring salient patches in subsequent frames are used to initialise associated feature points assuming a coplanar constraint as shown in Fig. 2c. Successful measurement of these and the seed in subsequent frames then allows an update of the normal and a corresponding reduction in its uncertainty as shown in Fig. 2d. Further points can then be initialised and added to the growing planar structure according to the coplanar constraint as shown in Fig. 2e, with the reference camera pose, seed position and normal vector then being updated with each set of new measurements.

Note that the size of the feature representation remains fixed throughout the above process, enabling point features to be parameterised in term of their common planar structure. Note also, as shown in Fig. 2f, any inconsistent measurements enable outlier point features to be rejected from the planar feature. As noted earlier, in the event that no coplanar points can associated with the seed, then the latter defaults to a single point feature which after convergence can be converted to a standard 3-D parameterisation as in existing systems [5, 6, 7].

# 3 Implementation

In this section we outline the implementation of the above mapping process defined within an EKF framework. Much of the latter mirrors that used in previous monocular SLAM systems, see e.g. [7], and thus we only provide a summary of its main components.

## 3.1 EKF State Representation

The state in the EKF is defined in terms of the current 6-D pose of the camera (3-D position and 3-D orientation) and a representation of the scene map. We denote this by $\mathbf{x}$ with its associated covariance $\mathbf{P}$, i.e.

$$\mathbf{X} = [\mathbf{v}, \mathbf{m}]^T \qquad \mathbf{P} = \begin{bmatrix} \mathbf{P_{vv}} & \mathbf{P_{vm}} \\ \mathbf{P_{mv}} & \mathbf{P_{mm}} \end{bmatrix} \tag{1}$$

where $\mathbf{v} = [\mathbf{r}^w, \omega^w, v^w, w^w]^T$ defines the camera position $\mathbf{r}^w$, its orientation in axis angles $\omega^w$ and associated velocities $v^w$ and $w^w$, all defined in the world coordinate system. The map $\mathbf{m} = [\mathbf{m}_1, \mathbf{m}_2, \ldots, \mathbf{m}_n]^T$ is defined in terms of a set of generalised features $\mathbf{m}_i$, which in our case is a mix of planar structures and single point features, defined in terms of the new common feature parameterisation (or standard 3-D parameterisation for converged single points).

Our common feature parameterisation is an adaptation of the Inverse Depth Parameterisation (IDP) which is now used routinely for single feature initialisation [17] and is similar in form to that used in [15, 18]. We denote it as the *Inverse Depth Planar Parameterization*

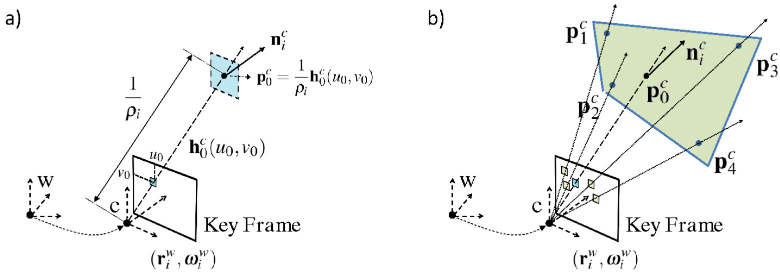



Figure 3: (a) Inverse depth planar parameterization; (b) 3-D points in the camera coordinate system obtained with the planar parameterization and the key frame.

(IDPP). It is defined in terms of a 10-D vector $\mathbf{m}_i = [\mathbf{r}_i^w, \omega_i^w, \mathbf{n}_i^c, \rho_i]^T$, where $\mathbf{r}_i^w$, and $\omega_i^w$ denote the position and orientation of the reference camera in world coordinates, $\mathbf{n}_i^c$ denotes the 3-D normal vector in the reference camera coordinate system and $\rho_i$ is the inverse depth of the seed point. The parameterisation is illustrated in Fig. 3a. Note that we use a redundant non-unit 3-D vector for the normal representation, which avoids potential singularities, compared with a polar representation, for instance.

Based on the above formulation, we now consider in detail the steps involved in the unified mapping process described in Section 2, namely the initialisation of new features, the associated measurement model and the observation and update steps within the EKF.

## 3.2 Initialization

At a given a time step, a salient point $(u_0, v_0)$ in the current image is used to initialize a new seed point. For this, the state vector is augmented with $\mathbf{m}_i = [\mathbf{r}_i^w, \omega_i^w, [0,0,1]^T, \hat{\rho}_i]$, where $\mathbf{r}_t$ and $\omega_t$ are the position and orientation estimates of the current camera pose. The vector $[0,0,1]^T$ initializes the normal in the perpendicular direction to the camera plane and $\hat{\rho}_t$ is initialized as for conventional IDP [17]. Outside of the state vector, the current image is associated with the new feature $\mathbf{m}_i$ as a *key frame* along with the initial pixel position $(u_0, v_0)$ which defines the direction of the seed point $\mathbf{p}_0^c$ as shown in Fig. 3a. The state covariance $\mathbf{P}$ is augmented to include the initial normal covariance $\mathbf{R_n}$, which is taken to be diagonal with large values, i.e. it is undetermined at this stage, and the variance of the inverse depth $\sigma_\rho^2$, which is initialized as in IDP [17]. Thus, the new covariance $\mathbf{P}^{new}$ is given by:

$$\mathbf{P}^{new} = \mathbf{J} \begin{pmatrix} \mathbf{P} & 0 & 0 \\ 0 & \mathbf{R_n} & 0 \\ 0 & 0 & \sigma_\rho^2 \end{pmatrix} \mathbf{J}^T \qquad \mathbf{J} = \left( \begin{array}{cccc|c} & \mathbf{I} & & & 0 \\ \hline \mathbf{I}_{6\times 6}, & 0, & \ldots, & 0, & \mathbf{I}_{4\times 4} \end{array} \right) \tag{2}$$

## 3.3 Measurement Model

The measurement model consists of two steps: (i) construct a set of 3-D points in the world coordinate system using the predicted state of the planar structure $\hat{\mathbf{m}}_i = [\hat{\mathbf{r}}_i^w, \hat{\omega}_i^w, \hat{\mathbf{n}}_i^c, \hat{\rho}_i]$ ; and (ii) project the set of 3-D points onto the current predicted camera plane to obtain a set of predicted 2-D measurements.

For the step (i), assume that we have a set of salient points in the key frame which have been associated with a planar structure. Each point $j$ has an associated pixel position $(u_j, v_j)$

that defines a bearing ray $\mathbf{h}_j^c(u_j, v_j)$ in the normalized reference camera plane. Each one of the rays will intersect the planar surface producing a set of 3-D points, as shown in Fig. 3b. To find the 3-D position where the intersection occurs the plane intersection equation can be formulated by using a known point on the plane and its normal. The latter is provided by the predicted normal $\hat{\mathbf{n}}_i^c$ and the seed point $\mathbf{p}_0^c$ provides the former, constructed from the predicted inverse depth $\hat{\rho}_i$ and the ray $\mathbf{h}_0^c$ corresponding to the initial pixel $(u_0, v_0)$ as shown in Fig. 3a. Since the plane pose is defined in the reference camera coordinate system then the plane-ray intersection equation is greatly simplified, i.e. for each ray $\mathbf{h}_i^c$, the intersection is given by:

$$\mathbf{p}_j^c = \frac{1}{\rho_i}\left[\frac{\mathbf{h}_0^c(u_0,v_0)^T \cdot \mathbf{n}_i^c}{\mathbf{h}_j^c(u_j,v_j)^T \cdot \mathbf{n}_i^c}\right]\mathbf{h}_j^c(u_j, v_j) \tag{3}$$

where each point $\mathbf{p}_j^c$ is defined in the reference camera coordinate system. To convert to the world coordinate system we use the predicted translation $\hat{\mathbf{r}}_i^w$ and rotation matrix $\mathbf{R}^w(\hat{\omega}_i^w)$ of the reference camera as follows:

$$\mathbf{p}_j^w = \mathbf{R}^w(\hat{\omega}_i^w)\mathbf{p}_j^c + \hat{\mathbf{r}}_i^w \tag{4}$$

For step (ii), we can use the standard perspective projection model into the current predicted camera pose:

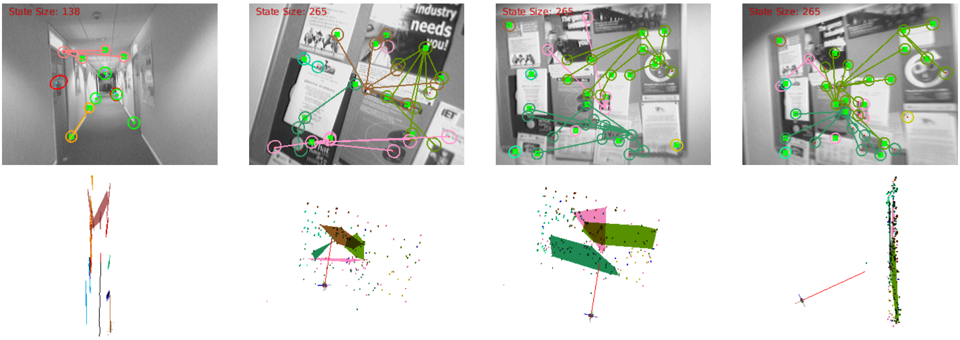
$$\mathbf{z}_j = \mathbf{h}(\mathbf{v}, \mathbf{p}_j^w) = \prod(\mathbf{R}^c(\omega^w)(\mathbf{p}_j^w - \mathbf{r}^w)) \tag{5}$$

where $\mathbf{r}^w$ is the predicted translation component and $\mathbf{R}^c(\omega^w)$ is the rotation matrix of the current predicted camera $\mathbf{v}$; $\prod$ denotes pin-hole projection for a calibrated camera with additive zero-mean Gaussian noise and pixel noise covariance $R$.

Before the observation is performed, the EKF requires the calculation of Jacobians for the measurement model in 5 which can be taken from the already available Jacobians for the point measurement model. Thus, the Jacobian w.r.t. the current camera $\mathbf{v}$ remains the same whilst the one w.r.t. the planar feature $\hat{\mathbf{m}}_i$ is calculated with the chain rule: $\frac{\partial \mathbf{h}}{\partial \hat{\mathbf{m}}_i} = \frac{\partial \mathbf{h}}{\partial \mathbf{p}_j^w}\frac{\partial \mathbf{p}_j^w}{\partial \hat{\mathbf{m}}_i}$ in a similar fashion to the IDP approach [17].

## 3.4 Observation and Update

The observation for each planar structure is the projection of the set of associated feature points as defined by eqn (5). These are sought about their predicted positions derived from the filter as illustrated in Fig. 2. Individual search regions about the predictions for each point are determined in the usual manner by projecting the covariances of the feature estimates through equations (3)-(5). For each point, an observation is obtained by applying a template matching process on all salient points within the predicted search region in the current frame.This is based on regions of size $11 \times 11$ pixels extracted from the key frame and normalised cross correlation, with warping applied according to the current mean estimate of the normal vector. Note therefore that the method automatically provides a degree of view invariance due to the use of the planarity constraint. The best matches for all the planar points are then used to update the filter and hence the estimates of the reference camera, seed point and normal vector. As indicated earlier, individual points for which sufficiently good matches cannot be found over several frames are regarded as outliers and removed from the planar structure as shown in Fig. 2f.




(a) (b) (c) (d)

Figure 4: Camera views (top) and 3-D views for the indoor sequence: (a) Examples of plane and point feature estimates, the latter shown as single green ellipses and the former as connected point features in the camera view and bounding boxes in the 3-D view. (b)-(d) New point features associated with existing planar features allow increased map density without increasing in the overall state size.

To allow robust SLAM operation we also implemented a relocalisation mechanism to allow recovery should tracking fail. For this, we adopted the approach of Chekhlov *et al.* [8], using spatial gradient descriptors to characterise a subset of selected points associated with each planar feature. This enables rapid relocalisation, helped by the large number of points available due to the efficient map representation provided by the planar representations.

# 4 Experiments

The proposed method has been tested under real time operation in several indoor and outdoor scenarios. We provide two representative examples to demonstrate the effectiveness of the method. Both were obtained using a hand-held calibrated camera with an image size of 320x240 pixels and a wide angle lens having an $81^o$ horizontal FOV. Salient points were obtained by using the Shi and Tomasi salient point detector [20] in combination with FAST and non-maximal suppression [19].

The aim of the experiments was to demonstrate that the approach is effective in (i) undelayed initialisation of planar features alongside the usual point initialisation; (ii) growing those structures as and when appropriate whilst maintaining a stable hybrid map; and (iii) significantly reducing the size of the map, and hence the state space, within the EKF. We have therefore selected test sequences in which the camera motion is fast and in which features are only in view for very limited periods. With points only mapping, such cases result in large numbers of features being initialised, resulting in a large map representation. Moreover, the point feature estimates do not converge before moving out of view, hence making the alternative forms of SLAM based planar mapping inappropriate since they rely on point convergence prior to initialising planar features.

## 4.1 Indoor Sequence

The first sequence was taken by a person moving along a corridor whilst keeping the camera pointing ahead. The corridor was approximately 35 meters long with a single right angled

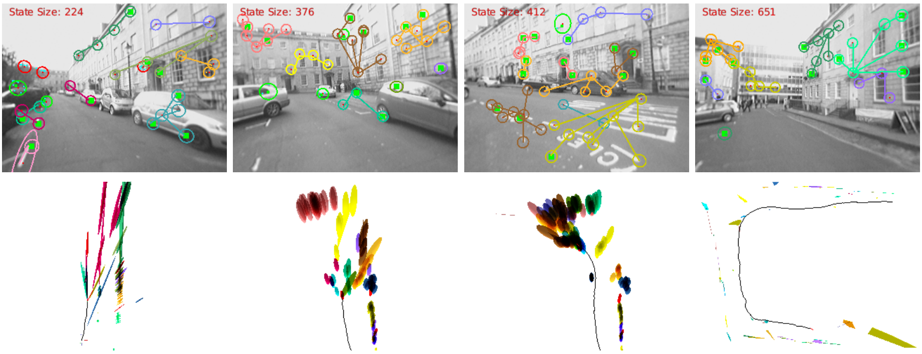



Figure 5: Camera views (top) and 3-D views for the outdoor sequence showing estimated point and planar features.

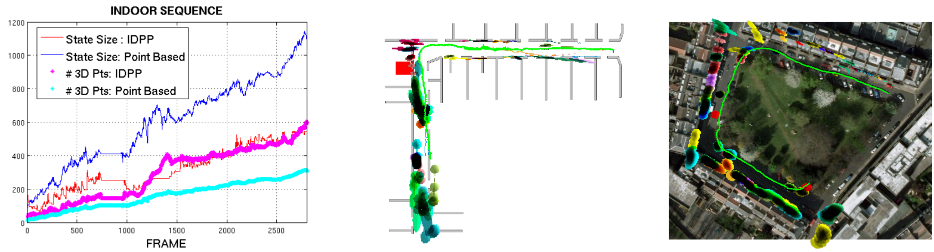
turn. The person was moving at a good walking pace and thus salient points on the planar walls on either side move quickly in and out of view. The scene is not particularly well textured and there are frequent changes in illumination due to the overhead lighting.

Figure 1 and the first column in Fig. 4 show examples of point and planar features initialised and tracked in different sections of the sequence. The planar features are shown both by bounding boxes in the 3-D external view (Fig. 1a and the bottom row of Fig. 4), the projection of those boxes into the current camera view (Fig. 1b), and connected point features as shown in Fig. 1c and the top row of Fig. 4. For example, in Fig. 1c, a large planar feature has been grown corresponding to the right-hand side of the corridor and currently consisting of 8 connected point features. Note also the small number of isolated point features within the hybrid map.

Figures 4c-d provide a good example of the ability of the method to keep down the size of the map representation. At this point in the sequence, the person has stopped to view the right-hand corridor wall and moves the camera closer. To maintain tracking new features need to be initialised into the map. With a points only system this would necessarily lead to a large increase in the state size. In contrast, with the proposed approach, the new features are linked to existing planar features via the respective key frame, resulting in no increase in the size of the map representation (the EKF state size is shown in the top-left of the figures). This is a further example of adaptive measurement for increasing map density which was previously introduced in [14].

To assess the significance of the state size reduction, we compared performance against a standard points only method. The variations in state size and the total number of point features initialised over the sequence for both approaches are plotted in Fig. 6a. This illustrates several significant outcomes. First, the hybrid method IDPP requires a significantly smaller state size (around 50%). Second, the state size reduction starts as soon as the map begins to be built, illustrating that planar features are being introduced with minimal delay in contrast to previous methods. Third, it does this whilst at the same time initialising a greater number of point features, thus maintaining greater stability in the filter. This is achieved because the latter are successfully associated to planar features, as anticipated.

To illustrate the overall mapping performance, Fig. 6b shows the final map and the associated uncertainties of the point and planar features overlaid on a schematic plan of the corridor. Apart from some drift at the end, which is to be expected without loop closure, this shows good alignment of the two maps. Notably it does this whilst operating in realtime at above 20 fps throughout the sequence. This is in comparison to the points only case in which




(a) (b) (c)

Figure 6: (a) State size and number of 3-D points obtained with the hybrid IDPP and a points-only method for the indoor sequence. (b) Estimated map and uncertainties overlaid on a schematic plan for the indoor sequence. (c) Estimated map and uncertainties overlaid on an aerial view for the outdoor sequence.

the large state size compromises the filter and the processing rate drops to below 1 fps by the end of the sequence. Of course, this is with a single filter and the use of a sub-mapping mechanism, as in [10] for instance, would aid in this respect. Nevertheless the comparison is useful to indicate the significant degree of saving that can be made by utilising planar structure. Moreover, we would argue that delaying the introduction of sub-maps is beneficial in terms of minimising the associated overhead of managing sub-maps and that maintaining efficient map representation is therefore of key importance.

### 4.2 Outdoor Sequence with a Long Walk

In a similar manner to the indoor sequence, the outdoor sequence was captured by a person pointing the camera ahead and walking with normal speed for about 230 meters around an urban square. Again, large numbers of salient points on either side move quickly in and out of camera view. Figure 5 shows examples of initialised tracked point and planar features in both the camera and 3-D views. Note that the system successfully detects planes corresponding to many planar structures within the scene, on the road, walls, buildings and cars. Figure 6c shows the estimated map overlaid on an aerial view of the square obtained from Google Maps. The total state size for this sequence was of 820 with a density of 765 planar points for the hybrid method which again maintained a processing rate above 20 fps. In contrast, for the points only method the map size increased quickly producing an early drop of the frame rate. At the end of the sequence the state size was 2232 with 671 mapped points with the frame rate falling below 1 fps.

## 5 Conclusion

We have presented a novel algorithm for mapping in monocular SLAM which unifies the estimation of point and planar features. The key contribution is the use of the IDPP parameterisation which allows the features to be dealt with in a common framework, making no distinction at initialisation and allowing planar or point structure to emerge according to the underlying scene. This allows fast and recursive introduction of planar structure, unlike previous methods, and avoids additional processing overhead, plane fitting, etc. Experiments demonstrate the potential of the method to give efficient map representation when operating over large areas for both indoor and outdoor environments. Future work will include further investigations into the consistency of the approach through simulations, looking at incorporating the approach within a sub-mapping framework and the use of visual clues to guide the selection of salient points when growing planar structures.

**Acknowledgements:** This work was partially supported by CONACYT Mexico under the grant 189903. We would like to thank Andrew Gee and Walterio Mayol-Cuevas for valuable discussions and advice.

# References

[1] M. Agrawal and K. Konolige. FrameSLAM: From bundle adjustment to real-time visual mapping. *IEEE Transactions on Robotics*, 24(5), October 2008.

[2] A. Bartoli and P. Sturm. Constrained structure and motion from multiple uncalibrated views of a piecewise planar scene. *International Journal of Computer Vision*, 52(1): 45–64, 2003.

[3] B.Williams, G.Klein, and I.Reid. Real-time slam relocalisation. In *Proc Int. Conf. Computer Vision*, 2007.

[4] R. O. Castle, D. J. Gawley, G. Klein, and D. W. Murray. Towards simultaneous recognition, localization and mapping for hand-held and wearable cameras. In *Proc. Int. Conf. Robotics and Automation*, 2007.

[5] D. Chekhlov, M. Pupilli, W. Mayol, and A. Calway. Robust real-time visual slam using scale prediction and exemplar based feature description. In *Proc. Int Conf on Computer Vision and Pattern Recognition*, 2007.

[6] A.J. Davison. Real-time simultaneous localisation and mapping with a single camera. In *Proc. International Conference on Computer Vision*, volume 2, pages 1403–1410. Citeseer, 2003.

[7] 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.

[8] 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.

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

[10] Ethan Eade and Tom Drummond. Monocular slam as a graph of coalesced observations. In *Proc. 11th IEEE International Conference on Computer Vision*, Rio de Janeiro, Brazil, October 2007.

[11] 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.

[12] A.P Gee, D.Chekhlov, A.Calway, and W.Mayol-Cuevas. Discovering higher level structure in visual slam. *IEEE Trans. on Robotics*, 24(5):980–990, 2008.

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

[14] 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.

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

[16] C. Mei, G. Sibley, M. Cummins, P. Newman, and I. Reid. A constant time efficient stereo slam system. In *BMVC*, 2009.

[17] J.M.M. Montiel, J. Civera, and A.J. Davison. Unified inverse depth parametrization for monocular slam. In *Robotics: Science and Systems Conf.*, 2006.

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

[19] E. Rosten and T. Drummond. Machine learning for high-speed corner detection. In *Proc. Euro Conf on Computer Vision*, 2006.

[20] Jianbo Shi and Carlo Tomasi. Good features to track. In *1994 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'94)*, pages 593 – 600, 1994.

[21] G F. Silveira, E. Malis, and P. Rives. Real-time robust detection of planar regions in a pair of images. In *International Conference on Intelligent Robots and Systems*, pages 49–54, 2006.

[22] G. F. Silveira, E. Malis, and P. Rives. An efficient direct approach to visual slam. *IEEE Transactions on Robotics*, 24(5):969–979, 2008.

[23] R. Szeliski and P. Torr. Geometrically constrained structure from motion: Points on planes. *3D Structure from Multiple Images of Large-Scale Environments*, pages 171–186, 1998.