

# DPPTAM: Dense Piecewise Planar Tracking and Mapping from a Monocular Sequence

Alejo Concha and Javier Civera  
I3A, Universidad de Zaragoza  
{alejocb, jcivera}@unizar.es

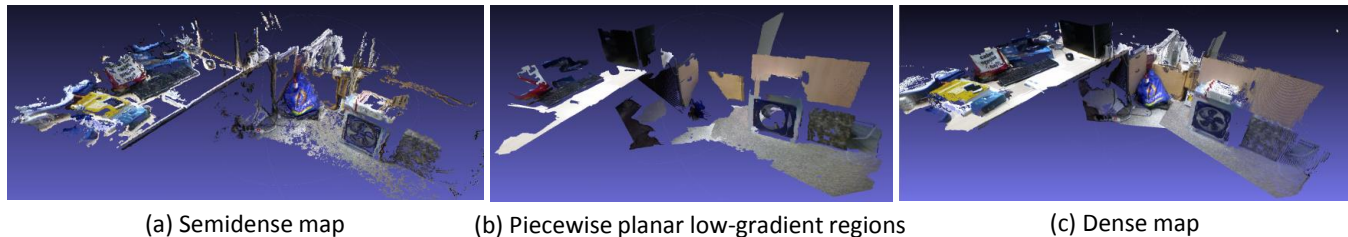


Fig. 1: Illustrative results of our demo. We estimate a semidense 3D map from a monocular sequence and reconstruct low-gradient areas assuming they are piecewise planar.

**Abstract**—Our demo is a direct monocular SLAM algorithm that estimates a dense reconstruction of a scene in real-time on a CPU. Highly textured image areas are mapped using standard direct mapping techniques [1], that minimizes the photometric error across different views. We make the assumption that homogeneous-color regions belong to approximately planar areas. Our contribution is a new algorithm for the estimation of such planar areas, based on the information of a superpixel segmentation and the semidense map from highly textured areas.

## I. INTRODUCTION

One of the key pieces of any virtual or augmented reality system is the 3D estimation of the surrounding scene and the pose of the device from sensing data, sequentially and in real-time. This is also an essential component of an autonomous robots and has been usually denoted with the acronym SLAM—Simultaneous Localization and Mapping. The monocular camera stands out as one of the most convenient sensors for several reasons.

One of the hardest challenges in monocular SLAM is the estimation of a *fully dense* map of the imaged scene. Pixels in textureless areas cannot be reliably matched across views and standard 3D reconstructions from monocular SLAM are limited to areas of high photometric gradients.

Our research starts in [2], [3] modelling the environment with 3D points for high-gradient areas and 3D planes for low-gradient areas. The assumption made is that image areas with low color gradients are mostly planar; which is met in most indoors and man-made scenes. Low-gradient image areas are segmented using superpixels.

## II. OVERVIEW

In our approach, the camera is tracked in real time at video frequency by minimizing the photometric error between the

high-gradient pixels of the current frame and the reprojection of the corresponding map points.

A semidense map is estimated from a sparse set of selected keyframes. This map is used to register the current camera in a global reference frame; and hence it should be estimated at a high rate.

Finally, a dense map is estimated from the same set of keyframes but at a slower rate. This dense map can be used for realistic augmentation or robotic navigation. The regularization that produces fully dense maps can be very demanding and a GPU is needed to do it in real-time, limiting its use to high-end devices. Our proposal is to leverage scene priors, specifically the Manhattan and piecewise planar structures in man-made scenes, to reduce the complexity of the map estimation. Some illustrative results of our algorithms can be seen in figure 1. The maps in this figure have been estimated in real-time in a CPU. The results can be better appreciated in the video of the footnote link <sup>1</sup>.

## ACKNOWLEDGMENT

This research was funded by the Spanish government with the projects IPT-2012-1309-430000 and DPI2012-32168

## REFERENCES

- [1] J. Engel, T. Schöps, and D. Cremers, “LSD-SLAM: Large-scale direct monocular slam,” in *Computer Vision—ECCV 2014*. Springer, 2014, pp. 834–849.
- [2] A. Concha and J. Civera, “Using superpixels in monocular SLAM,” in *IEEE International Conference on Robotics and Automation*, Hong Kong, June 2014.
- [3] A. Concha, W. Hussain, L. Montano, and J. Civera, “Manhattan and piecewise-planar constraints for dense monocular mapping,” in *Robotics: Science and Systems*, 2014.

<sup>1</sup><http://webdiis.unizar.es/~jcivera/videos/iros15submission.mp4>