Citation: A. Pérez-Yus, G. López-Nicolás, J. J. Guerrero. Wide RGB-D for Scaled Layout Reconstruction. International Workshop on Lines, Planes and Manhattan Models for 3-D Mapping (LPM 2017). IROS Workshops, Vancouver, Canada, September 28, 2017.

Wide RGB-D for Scaled Layout Reconstruction

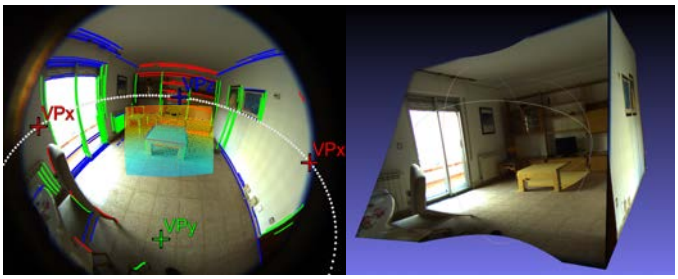
Alejandro Perez-Yus, Gonzalo Lopez-Nicolas, Jose J. Guerrero

One of the most important topics in computer vision and robotics has always been to perceive the 3D information from the scene. The advent of consumer RGB-D cameras has caused a great positive impact in the field. Unfortunately, these devices usually have a field of view (FOV) too narrow for certain applications, and it is necessary to move the camera to capture different views of the scene. Our goal is to be able to reconstruct the structure of the scene with scale in one single shot. To achieve this goal we propose to use a color camera with wide FOV to extend the depth information in a novel hybrid camera configuration composed by a depth and a fisheye camera (Fig. 1a). Once the cameras are calibrated [1], the system is capable of viewing over a 180° of color information where the central part of the image has also depth data (Fig. 1b). To our knowledge, this is the first time this configuration has been used, although the interest in such sensor pairing is clear in new devices e.g. Google Tango.

In particular, we propose to extend the 3D information in one single shot via spatial layout estimation. Our layout estimation method is based on line segments from the fisheye image, and provides scaled solutions rooted on the seed depth information. As a result, a final 3D scene reconstruction is provided (see Fig. 1c). The 3D room layout can be seamlessly merged with the original depth information to generate a 3D image with the periphery providing an estimation of the spatial context to the central part of the image, where the depth is known with good certainty. The collaboration between cameras is bidirectional, since the extension of the scene layout to the periphery is performed with the fisheye, but the depth information is used both to enhance the layout estimation algorithm and to scale the solution. A scheme of the whole algorithm is shown in Fig. 2.

In detail, the depth camera provides a region of the image with 3D data, from which an initial estimate of the Vanishing Points (VPs) and 3D planes can be recovered. We assume scenes are from a Manhattan World [2], and the VPs are used to retrieve the scene orientation to generate layout proposals. The 3D planes extracted are used to find the floor and provide scale, impossible to get otherwise with one single shot and no previous knowledge of the scene. Having scale has many advantages in this type of methods which usually have many heuristics. For instance, when tuning parameters

(a)



(b) (c)

Fig. 1: (a) Fields of view of our proposed system composed by a Fisheye and a RGB-D camera. (b) The depth information in the center is extended to the periphery combining information with the line segments that we use to extract the spatial layout of the scene. (c) At the end we obtain a 3D reconstruction of the scene with scale.

their values can be grounded in reality with real measurement units. Depth information is also used to filter hypotheses and reward line segments corresponding to planar intersections.

The line segments from the wide image are classified according to the the three Manhattan directions. The horizontal lines are projected either to the floor plane or the estimated ceiling plane to have the 3D segment position in the real world. Structural corner candidates are then looked for, by considering plausible and simple cases of line distribution. These corners are evaluated by our scoring function, so layout hypotheses are proposed by the probability of these corners to occur in the real world. Then, layout hypotheses are generated based on geometrically coherent wall distributions that do not contradict the initial depth information and the observable segments. The algorithm is able to work even under high clutter circumstances due to the combination of lines from both floor and ceiling (because of using a large FOV camera), but also because of our generation of Manhattan hypotheses that can estimate hidden corners to complete the layout. For the evaluation stage we

Instituto de Investigación en Ingeniería de Aragón (I3A), Universidad de Zaragoza, 50018, Zaragoza, Spain. alperez@unizar.es, gonlopez@unizar.es, josechu.guerrero@unizar.es

This work was supported by the projects DPI2014-61792-EXP and DPI2015-65962-R (MINECO/FEDER, UE), the grant BES-2013-065834 (MINECO).

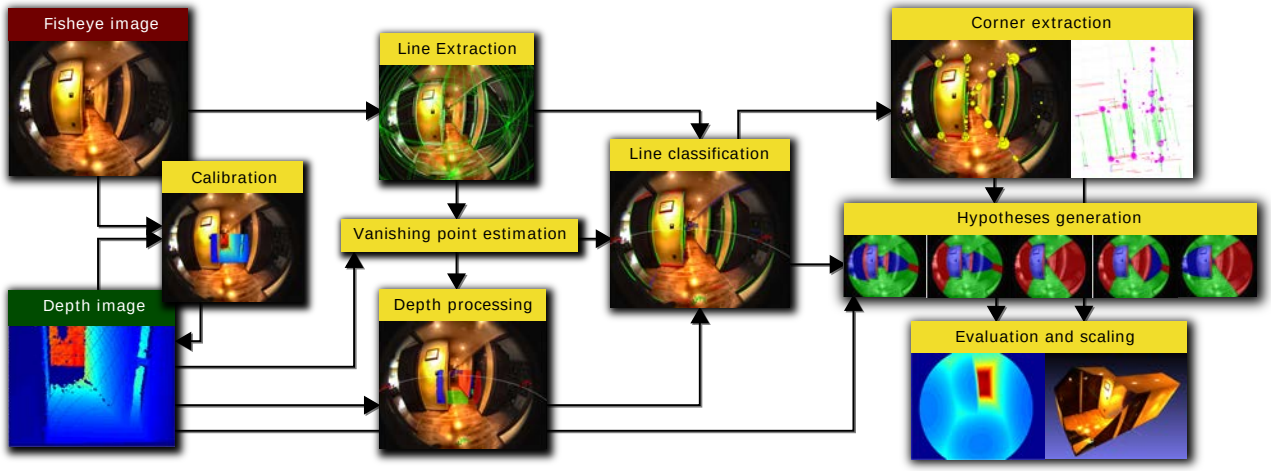


Fig. 2: Block diagram of the main stages of the algorithm, starting with the initial Fisheye and Depth images and finishing with the result of the evaluation.

propose three fast alternative methods whose performance is comparable to well known state of the art method [3].

Preliminary results were already presented in [4]. Experiments using a new dataset with real images already showed promising results about both the proposed algorithm and the camera configuration. In particular, they showed that our method gets good results even when only a few hypotheses are drawn, and that having depth information helps notably in the layout extraction. We show a summary of the algorithm and a few reconstructed layouts in a video[1]. In particular, for each resulting layout it can be seen how the estimated point cloud is able to extend the initial depth information up to 180°. Since the presentation of [4] we have improved some stages in the algorithm in both performance and efficiency, added some additional features and performed new experiments with another device. In particular, some of our recent improvemets include: *i)* a new evaluation method which presents similar results as with the Orientation Map [3] with much less required computation; *ii)* a scaling procedure for those cases where the floor plane is not found in the image; *iii)* a pre-filtering of hypotheses; and *iv)* new experiments with these new features and with a new dataset of images taken from the Google Tango developer kit. In Fig. 3 we show some experiments with this device, showing that our method achieves great performance in commercial systems too.

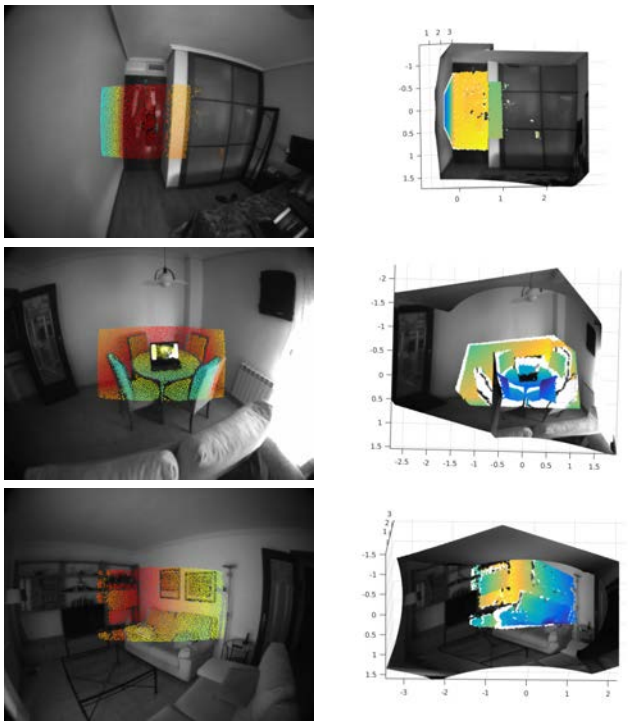


Fig. 3: Examples of 3D reconstructions with a Google Tango. Left: the fisheye image with the depth points projected in variable colors. Right: the resulting 3D point cloud with the initial 3D point cloud also displayed in color to show how the resulting cloud has been scaled and fits accordingly.

References

[1] A. Perez-Yus, G. Lopez-Nicolas, and J. J. Guerrero, "A novel hybrid camera system with depth and fisheye cameras," in *IAPR International Conference on Pattern Recognition (ICPR)*, 2016, pp. 2789–2794.

[2] J. M. Coughlan and A. L. Yuille, "Manhattan world: Compass direction from a single image by bayesian inference," in *IEEE International Conference on Computer Vision (ICCV)*, vol. 2, 1999, pp. 941–947.

[3] D. C. Lee, M. Hebert, and T. Kanade, "Geometric reasoning for single image structure recovery," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009, pp. 2136–2143.

[4] A. Perez-Yus, G. Lopez-Nicolas, and J. J. Guerrero, "Peripheral expansion of depth information via layout estimation with fisheye camera," in *European Conference on Computer Vision (ECCV)*. Springer, 2016, pp. 396–412.

[1] https://www.youtube.com/watch?v=nQYvhAhvv6U