Full 3D layout reconstruction from one single 360º image

Clara Fernández Labrador, Alejandro Pérez Yus, Gonzalo López Nicolás, Jose J. Guerrero

Grupo de robótica, percepción y tiempo real (RoPeRT)
Instituto de Investigación en Ingeniería de Aragón (I3A)
Universidad de Zaragoza, Mariano Esquillor s/n, 50018, Zaragoza, Spain.
Tel. +34-976762707, e-mail: cfernandez@unizar.es

Abstract

We propose an entire pipeline which receives as input a 360º panorama and returns a closed, 3D reconstruction of the room faithful to its actual shape. We exploit deep learning combined with geometry to obtain structural lines, and thus structural corners, from which we generate final layout models assuming Manhattan world.

Introduction

Layout recovery of indoor scenes is an essential step for a wide variety of computer vision tasks and has recently received great attention from several applications like virtual and augmented reality, scene reconstruction, indoor navigation and SLAM.

First approaches facing this task used conventional images and made really big progresses based on geometric reasoning. However, these works present a strong constrain due to the limited field of view of traditional images, which conducts to obtain open geometries and simple box assumptions considering rooms to have just four walls. The challenge nowadays is to recover closed geometries without strong shape assumptions. With this motivation, a more recent research direction looks to extend the FOV by using catadioptric systems or fisheye cameras. But the real impact comes with omnidirectional 360º images, which allows to acquire the whole scene at once and hence, it is possible to exploit their wide FOV to generate closed room solutions based on the best consensus distributed around the scene. PanoContext [1] is the first work using panoramas to deal with the layout recovery problem but they also assume the room as a simple 3D box. Additionally they find bounding boxes of the main objects inside the room. On the other hand, in the last years, the research community started to face layout recovery problems with convolutional neuronal networks (CNN) achieving an outstanding success like in [2], where they train a CNN to extract the informative structural edges of indoor scenes ignoring those edges from clutter. Other deep learning works extract an estimation of the depth or/and surface normals from RGB images which also produces an interesting outcome for layout estimation [3]. The main drawback of these CNNs is that they are always focused on traditional images with limited FOV and therefore do not work well when used directly on panoramas. In this work we inspire ourselves in [1] and go ahead achieving more flexible geometries that are faithful to the actual shapes of the rooms by combining the accuracy provided by geometric reasoning with the exploitation of deep learning techniques.

Layout recovery

An overview of the proposed algorithm is shown in Fig.1. Initially, we detect lines and vanishing points by a RANSAC-based algorithm working directly with panoramas and achieving really similar results to the state of the art but also much faster. To do that, we adopt the Manhattan World assumption whereby there exist three dominant orthogonal directions *(x,y,z)*. In cluttered scenes is very difficult to know whether lines come from actual wall intersections or from other elements of the scene. Proceeding with all the lines has led to an intractable number of hypotheses during years. We propose to tackle this problem introducing a deep learning approach [2] that successfully extracts feature maps showing with higher probability the areas where structural edges are. In order to take advantage of this method, we split the panoramas in a set of overlapping images projected as conventional ones and run the CNN on them separately. To do this split, we propose to discretize the sphere with an algorithm based on the golden section spiral, which leads in an elegant, evenly distribution. We associate each extracted line to a score calculated as the sum of the corresponding probability values to the pixels it occupies in the edge map. In this way, we can work with a smaller subset of accurate lines (geometry) that belong to the main structure (deep learning). Intersecting these structural lines in pairs, we obtain candidate corners to generate the final layout. We classify corners following two criteria: 1) their position along the z axis: ceiling and floor candidate corners, 2) their position in the XY-plane: we divide the scene into four quadrants taking into account the horizontal VP and the center of the camera, so that $c \in \{q1|q2|q3|q4\}$. Manhattan World rooms always have an even number of walls and, in each quadrant, an odd number of corners. We generate layout hypotheses from our candidate corners and we obtain that the number of walls N_W that our algorithm is able to solve, depend on the group of corners randomly selected g_c at each iteration $N_W^{max} = 2(N_{g_c} - 1)$. We establish a minimum requirement for which there must be corners in at least three quadrants, thus the corner in the remaining quadrant can be estimated assuming closed Manhattan layouts, and there must be at least one corner of each hemisphere to estimate the height of the room. We proceed with the geometric reasoning in 2D, projecting the direction vectors of the ceiling corners into a reference plane as points and joining consecutive corners in clockwise order by choosing the union which produces alternatively oriented consecutive (x-y) walls. The optimal projection for the floor corners is found along their direction vector to estimate then a relative distance between ceiling and floor planes. Models satisfying Manhattan world (walls forming 90º), will be compared with a Normal Map (I^R). The hypothesis with higher pixel coincidence with the Normal Map is selected as final room layout. We obtain this Normal Map throughout another CNN approach [3] following the same process as before. Here we face more complex designs which will be faithful to the actual shapes of the rooms, introducing the possibility of estimating in-between hidden corners when required, *i.e.* when they are occluded by clutter or due to scene non convexity. In Fig.2 we show some final 3D room models.

Conclusions

Our idea of exploiting deep learning combined with geometry allows to work directly with structural lines to create more efficient algorithms that tackle the layout estimation problem with less iterations and more accuracy. We also propose a new evaluation approach, Normal Map, alternatively to classical approaches. Last but not least, we handle flexible closed geometries not limited to 4-wall boxes which has a high relevance if it is used in a real room-navigation system. Experimental results are conducted within two public datasets. For more details, [4]. (Supported by Projects DPI2014-61792-EXP and DPI2015-65962-R (MINECO/FEDER, UE)).

REFERENCIAS

[1]. ZHANG, Y., SONG S., TAN P., and XIAO J. PanoContext: A whole-room 3D context model for panoramic scene understanding. *In European Conference on Computer Vision*, p. 668–686, 2014.

[2]. MALLYA, A., and LAZEBNIK, S. Learning informative edge maps for indoor scene layout prediction. *In International Conference on Computer Vision,* p. 936–944, 2015.

[3]. EIGEN, D., and FERGUS, R. Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture. In *International Conference on Computer Vision*, p. 2650–2658, 2015

[4]. FERNANDEZ, C., PEREZ, A., LOPEZ, G., GUERRERO, J.J. Layouts from Panoramic Images with Geometry and Deep Learning. *Submitted 2018.*

Figure 2. Layout estimations handling different geometries.

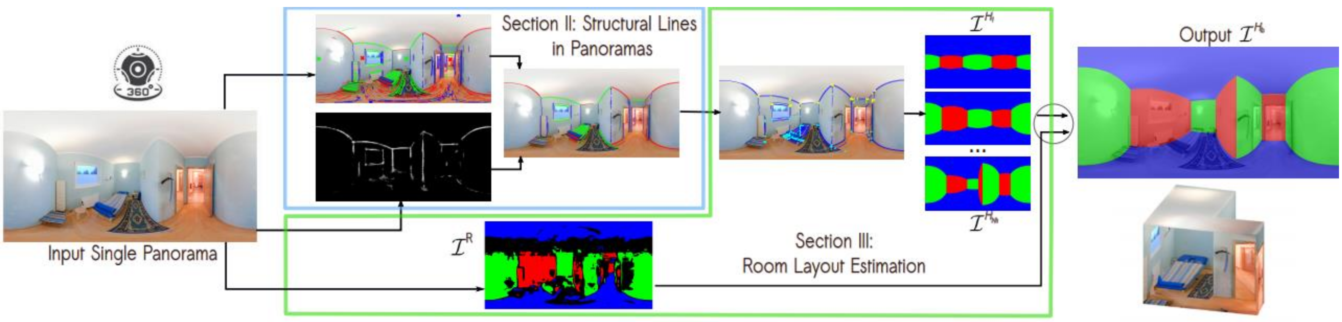

Figure 1. Proposed pipeline. Starting from a single spherical panorama, we exploit the combination of geometry (accurate lines) and deep learning (edge map) to recover the main structure of the room, achieving 3D complex layouts.