Estimating Structure of Rooms from Full-view Image

Hanchao Jia
Tottori University
Tottori, Japan
m09t3009@faraday.ele.tottori-u.ac.jp

Shigang Li
Tottori University
Tottori, Japan
li@ele.tottori-u.ac.jp

1. Introduction

Recent years have seen a growing interest in indoor scene understanding from a single image. Most proposed methods tackle this problem based on the pinhole camera model using conventional images. On the other hand, some researches consider the images captured by omnidirectional cameras with wide field of view, for example, catadioptric sensors [1, 2], and fisheye cameras [3]. The papers argue that omnidirectional vision can provide more information than conventional vision which only has a small part of view. However, we notice that though the fields of view are enlarged to a hemisphere, they remain involve just partial scene. That is to say, we cannot obtain the complete structure of a room where camera is in by any means, because there is no corresponding information in the image. For example, catadioptric images lost the ceiling of rooms, and fisheye images only catch half of the scene. In essence, the models in all existing methods so far obey the similar geometric constrains, and the spatial layouts recovered refer to incomplete structures. Here, we describe these cases as the "open" geometry, which implies that some parts of the space are lost.

Can we have other more specific and comprehensive models, which enable computers to predict entire structure, as well as recognize the whole room by only one image? It seems what we have to do is to break through the limitation of the open geometry. We pay our attention to full-view images. Nowadays, 360-degree panorama display became easily obtained and widely used in various aspects. Some examples are given in Figure 1. A famous application in practice refers to Google Street View. The feature of these full-view images can be included as that you are able to enjoy the entire scene at the camera point without losing any information. In another word, the spatial layout recovered from the images is complete, constructing a "close" space. Thus, we impose an analogous description to present geometric constrains of this model, called "close" geometry. A full-view image based on close geometry allows minimizing the possibility of fatal line detection caused by occlusions and partial views. Suppose an extreme case, when an occluding object covers the whole conventional image, it is impossible to carry out any estimation. However, we may still be able to reconstruct the room regarding the conventional image as a part of our full-view image.

In this work, we advocate a new model based on close geometric constrains to explore indoor scene understanding from a single full-view image. A novel method is also given to estimate the structure of rooms by searching for the structure which fits the generated corners best.

Figure 1: Examples of full-view images. Top: Two half scenes captured by fisheye camera. Bottom: A panorama view.

2. A model for indoor scene

Generally, it is convenient to turn the images captured by various types of central cameras into an equivalent sphere. Therefore, we employ sphere model to handle full-view images, which describes a 360-degree panorama view without considering specific visual sensors. It provides several important properties. A lines segment in 3D environment is converted to a part of great circle in the sphere space. Vanishing points of parallel lines always lie in the sphere.

Thanks to the entire view, we are able to observe complete "ceiling", "walls", "floor" all the time. Though a full-view image can also be segmented as some independent images, which followed the open geometry like conventional ones, our model obeys more strict close geometry. The main additional differences are, first, boundaries between walls and ceiling or floor are completely visible as long as we see by ourselves at the camera location. Walls go around making a circle, no longer left-to-right. That means walls are equivalent without a start or an end, avoiding the trouble of poor estimation on the edge. Second, the ceiling-floor symmetry means there is exactly only one ceiling plane corresponding with one floor plane with constant ceiling height. We modify the planar homology used in [4] to infer the image locations of ceiling points by the corresponding ones in floor plane, or vice versa. Another criterion is addressed among "wall" labels. The vanishing points of horizontal lines must be located at the walls with different orientation. Obviously, for the conventional images, this close geometry is not always guaranteed because of limited field of view.

3. Estimating structured layout

We start from a collection of line segments and vanishing points under the Manhattan world assumption. Any method capable of extracting line segments and optimizing the vanishing points could be applied.

Due to the ceiling-floor symmetry, either a point in the ceiling-wall boundary or in the floor-wall boundary is sufficient to specify both. Without loss of generality, we choose to present two corresponding boundaries only by the floor-wall boundary. In a full-view image, the longest horizontal line always lies in floor-wall boundary, if it is not occluded too heavily to be visible. According to this assumption, we may obtain a preliminary structure including the longest horizontal line and the ones share the same ending. The purpose of this step is to enhance the possibility of layout estimation integrated with the following processing. However, it is not vitally necessary. We set a safe threshold for checking the length of lines to ensure what we choose really belongs to the floor-wall boundary. In the case of failing check, this step will be omitted and have no impact on following spatial layout estimation.

Then, we generate a set of corners from vertical lines, since detection of vertical lines is more robust and less susceptible to noise than horizontal lines. We define a corner as the intersection of horizontal and vertical lines (extended line) at their ending point, or ending point of a vertical line itself. We denote these corners as a set $Q(q)$. A segment of the floor-wall boundary W is determined by two key corners and an orientation. Thus we can represent one structure of indoor room G as a closed floor-wall boundaries $(W_1, W_2, \ldots, W_n)$.

We come up with a novel scheme of searching for the best fitting hypothesis only using corners, since corners provide us desirable local geometric reasoning as what we define. Considering the representation of indoor structure in our paper, we draw a conclusion that a structure hypothesis is feasible if all of its key corners are feasible, and the feasibility is dependent on the corners. Given a structure $G(W)$ and a corner set Q, a score function $C_q(q_i, G)$ indicates the fitness of corner q_i to the structure G. Total score $C(G)$ of a hypothesis is computed by summing up every score of corners within the boundaries. The process of searching for the best fitting hypothesis is converted into a maximization problem. More formally, we have

$$G^* = \arg\max(C(G)) \quad (1)$$

Thus, we can obtain the entire structure by maximize (1) to greatest score and label the surface of indoor rooms.

4. Experiments and conclusions

We cope with this indoor scene understanding problem by combining the process of deciding walls and pursuing the maximal score together, which makes our algorithm much efficient and robust. We test our approaches on several full-view images captured from different sensors. Figure 2 shows the experiment result of Figure 1. We can see most parts are recovered well except some parts of ceiling-wall boundaries. This is due to the computational accuracy of line detection and vanishing points. The results indicate that the proposed method (close geometric constraints) performs well to estimate the entire structure of indoor scene from a single full-view image.

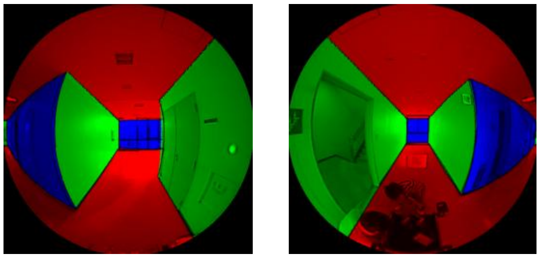

Figure 2: The result of structure estimation. We project the full-view image in the sphere model onto two planes for easily seeing.

References

[1] Ozisik. N.D., López-Nicolás. G. and Guerrero. J.J, "Scene structure recovery from a single omnidirectional image," In ICCV Workshops, pp. 359–366, 2011.

[2] Jason Omedes, Gonzalo López-Nicolás and José Jesús Guerrero, "Omnidirectional Vision for Indoor Spatial Layout Recovery," Frontiers of Intelligent Autonomous Systems 2013, pp. 95-104, 2013.

[3] Hanchao Jia and Shigang Li, "Estimating the Structure of Rooms from a Single Fisheye Image," In ACPR Workshops, pp. 818–822, 2013.

[4] Alex Flint, Christopher Mei, David W. Murray and Ian D. Reid, "A Dynamic Programming Approach to Reconstructing Building Interiors," In Proc. ECCV, 2010.