

# 1 Recognition Using Visual Phares

In their paper, the authors propose an extension of their previous work, from estimating the surface labels of a room from an image to recover the spatial layout of a cluttered room. The major contribution of the paper is an algorithm estimating the full spatial properties of an image of a room while being robust to clutter. The authors noted that the proposed algorithm is robust to clutter because of the fact that it parameterizes the rooms. The second reason why the algorithm is robust is that the clutter is explicitly modeled in the algorithm. Along with the proposed algorithm, the authors provide one possible application of it.

## 2 Approach

The proposed approach contains two steps iteratively employed to recover the spatial properties of a room image. The first step of the algorithm is to parameterize the room being investigated with a 3-D box. In order to model the room with a 3-D box, line segments are extracted, and the segments are then employed to estimate the vanishing points. After the line segment detection step, all of the candidate vanishing points are chosen from the intersection points of the lines. (I believe it would be a more coherent read for me if the authors provided the framework they used for the line segment extraction.) The candidate vanishing points are then evaluated based on a Hough-like vote-based system in which, for each line segment the middle point, length and the angle between the middle point of the line segment are used in the voting function. The vanishing points not only provide the orientation of the box fit to the room, but they also set hard constraints for the geometry of the room which finalize the parameterization step. The authors also provide a learning framework which outputs confidence levels for the box layouts which are fit for the room. The second major step in the paper is to estimate the surface labels which are based on the acquired box layout. In order to label the surfaces, the image is oversegmented to extract superpixels. A boosted decision tree classifier is used to get a likelihood score for each pixel based on the visual cues, namely color, texture and edges, and the vanishing points. For each superpixel, the pixel-level likelihood scores are integrated to acquire the label confidence scores. An interesting idea is to use the surface labels as a feedback for new 3-D box estimations to increase the accuracy. The authors claim that using these two major steps, box parameterization of the room and surface label estimation, in a recursive fashion makes the algorithm robust to clutter.

## 3 Conclusions

It is reasonable to assume that man-made structures can be characterized with perpendicular lines; however, it is not necessarily the case for all the indoor images. If one considers darkening curtains and structural supports, it is obvious that they do not form the spatial properties of the room. Given their strong visual appearance, the proposed algorithm may poorly estimate the spatial properties of the room. Another point that I would like to mention is that room lighting and wallpaper texture can be used to improve the performance of the algorithm in estimating the shape of the rooms. In the box parameterization step, the greedy search technique is used to choose the vanishing points amongst candidates; although the greedy search is known for being susceptible to convergence at the local optima points.