

EDGE BASED 3D INDOOR CORRIDOR MODELING USING A SINGLE IMAGE

Ali Baligh Jahromi and Gunho Sohn

GeoICT Laboratory, Department of Earth, Space Science and Engineering, York University,
4700 Keele Street, Toronto, Ontario, Canada M3J 1P3
(baligh, gsohn)@yorku.ca

Commission III, WG III/4

KEY WORDS: Indoor Model Reconstruction, Geometric Reasoning, Hypothesis-Verification, Scene Layout Parameterization

ABSTRACT:

Reconstruction of spatial layout of indoor scenes from a single image is inherently an ambiguous problem. However, indoor scenes are usually comprised of orthogonal planes. The regularity of planar configuration (scene layout) is often recognizable, which provides valuable information for understanding the indoor scenes. Most of the current methods define the scene layout as a single cubic primitive. This domain-specific knowledge is often not valid in many indoors where multiple corridors are linked each other. In this paper, we aim to address this problem by hypothesizing-verifying multiple cubic primitives representing the indoor scene layout. This method utilizes middle-level perceptual organization, and relies on finding the ground-wall and ceiling-wall boundaries using detected line segments and the orthogonal vanishing points. A comprehensive interpretation of these edge relations is often hindered due to shadows and occlusions. To handle this problem, the proposed method introduces virtual rays which aid in the creation of a physically valid cubic structure by using orthogonal vanishing points. The straight line segments are extracted from the single image and the orthogonal vanishing points are estimated by employing the RANSAC approach. Many scene layout hypotheses are created through intersecting random line segments and virtual rays of vanishing points. The created hypotheses are evaluated by a geometric reasoning-based objective function to find the best fitting hypothesis to the image. The best model hypothesis offered with the highest score is then converted to a 3D model. The proposed method is fully automatic and no human intervention is necessary to obtain an approximate 3D reconstruction.

1. INTRODUCTION

People spend approximately 90% of their time indoors (U.S. EPA, 2015). However, unlikely in outdoor environment, not much spatial information of indoor space is available and thus human's indoor activities and related issues in health, security and energy consumption are difficult to be understood. Thus, with a rapid emergence of Building Information Model (BIM) and Building Science, providing semantically rich and geometrically accurate indoor models has recently gained more attention from the researchers. The generation of an indoor space 3D model needs a proper implementation of sensors as well as selecting a proper algorithm to reconstruct 3D models from the incoming data. This can help to accurately model the whole scene and contribute towards the efficiency of the reconstructed model later on. Considering the available data gathering techniques with respect to the sensors cost and data processing time, single images proved to be one of the reliable sources. Normally, single images can cover a limited field of view. Therefore, large scale environments may not be handled with a single image. However, they are still suitable for modeling the limited areas of indoor environments. In this paper modeling of indoor corridors using a single image is in focus.

The early attempts on understanding the scenes start by recovering vanishing points and camera parameters from an image using straight line segments (Kosecka and Zhang, 2002). Considering the Manhattan World Assumption, rectangular surfaces aligned with main orientations were detected using vanishing points (Kosecka and Zhang, 2005; Micusik et al., 2008). Top-down grammars were applied on line segments for

finding grid or box patterns which has rectangular pattern aligned with vanishing points (Han and Zhu, 2005). The statistical methods on image properties were used to estimate regional orientations and vertical regions "popup" considering the estimated orientations (Hoiem et al., 2005).

The relative depth-order of partial rectangular regions was inferred by considering their relationship and vanishing points (Yu et al., 2008). Parameterized models of indoor environments introduced which were fully constrained by specific rules to guarantee physical validity (Lee et al., 2009). Possible spatial layout hypothesis is sampled from collection of straight line segments but the method is not able to handle occlusions and fits room to object surfaces.

Statistical learning showed to be an alternative to rule-based approaches (Hoiem et al., 2005; Delage et al., 2006; Hoiem et al., 2007). Having a new image, the list of extracted features should be evaluated. The associations of these features with 3D attributes can be learned from training images. Therefore, the most likely 3D attributes can be retrieved from the memory of associations. The first method to integrate local surface estimates and global scene geometry used a single box to parametrize the scene layout (Hedau et al., 2009). Appearance based classifier was used to identify clutter and visual features were only computed from non-clutter regions. They used the structural learning approach to estimate the best fitting box to the image. Another approach similar to this has been proposed which does not need the clutter ground truth labels (Wang et al., 2010).

In recent years, some other approaches have been proposed for the extraction of 3D layout of rooms from single images (Hedau et al., 2010; Lee et al., 2010; Hedau et al., 2012; Pero et al., 2012; Schwing et al., 2012; Schwing and Urtasun, 2012; Schwing et al., 2013; Chao et al., 2013, and Zhang et al., 2014). Most of these approaches parameterize the room with a single box and assume that the room is aligned with the three orthogonal directions defined by vanishing points (Hedau et al., 2009; Wang et al., 2010; Schwing et al., 2013, and Zhang et al., 2014). Some of these approaches make use of the objects for reasoning about the scene layout (Hedau et al., 2009; Wang et al., 2010, and Zhang et al., 2014). On one hand, presence of objects can provide some physical constraints such as containment in the room and can be employed for scoring the room layout (Lee et al., 2010; Pero et al., 2012, and Schwing et al., 2012). On the other hand, the scene layout can be utilized for better detection of objects (Hedau et al., 2012, and Fidler et al., 2012).

In this paper, we tackle the problem of indoor space modeling from a single image through middle-level perceptual organization. We search for layout that can be translated into a physically plausible 3D model. Considering the Manhattan Rule Assumption, we adopt the stochastic approach to sequentially generate many physically valid layout hypotheses from line segments. Each generated hypothesis will be scored for finding the one that best matches the detected line segments. Finally, the best created hypothesis will be converted to a 3D model.

The main contribution of the proposed method is providing an approach to create layout of indoor corridors in a hybrid way using both detected line segments and virtually generated rays from vanishing points. This method is beneficial for two main reasons. First, the hybrid way of generating scene layout provides a realistic solution when dealing with objects or occlusions in the scene. Moreover, it is well-suited to describe most corridor spaces. It outperforms the methods which only use virtual rays for layout creation, since these rays are deviating from the true layout in long corridors due to the inaccuracy of the estimated vanishing points. Also, this method outperforms the other methods that use only actual line segments for generating the scene layout due to their inability to handle occlusions. Second, we propose a scoring function to score the created layout hypotheses. This function considers the volumetric aspect of the created hypotheses along with their correspondences to real edges, and compatibility to the orientation map. This scoring function finds the most fitting solution in a linear way. In the following section an overview of the proposed method will be provided.

2. OVERVIEW

Normally, the generation of 3D indoor space models can be achieved through two different approaches; top-down and bottom-up. On one hand, top-down approaches are very much deterministic in employing strong prior; therefore they can be robust to the missing data problem. An example of this approach is the work presented by Hedau et al. (2009). On the other hand, in bottom-up approaches perception forms by data and they make use of weak prior; therefore the created model could be more flexible. An example of this could be the method presented by Lee et al. (2009). In this paper the proposed method is more inclined to the top-down approaches, since it considers the indoor scene to have a cubic formation. However,

the main innovation of this method is that it does not restrict the indoor scene to only one box. Hence, it relaxes the strong one box prior by letting indoor scene to be comprised of multiple boxes. Therefore, the occluding spaces can easily be modelled by the proposed method.

Figure 1, shows the workflow of the proposed method; 1) Edges are extracted and grouped into straight line segments. 2) Lines are grouped based on parallelism, orthogonality, and convergence to common vanishing points. 3) Hypothetical cubic layouts are sequentially formed by intersecting hypothetical structural planes. These structural planes are created using detected line segments in the image space and virtual rays of vanishing points if necessary. 4) The best fitting layout hypothesis is selected using the linear scoring function. 5) The best fitting indoor space layout hypothesis is converted to 3D model. In the following sub-sections more details about the proposed method will be presented.

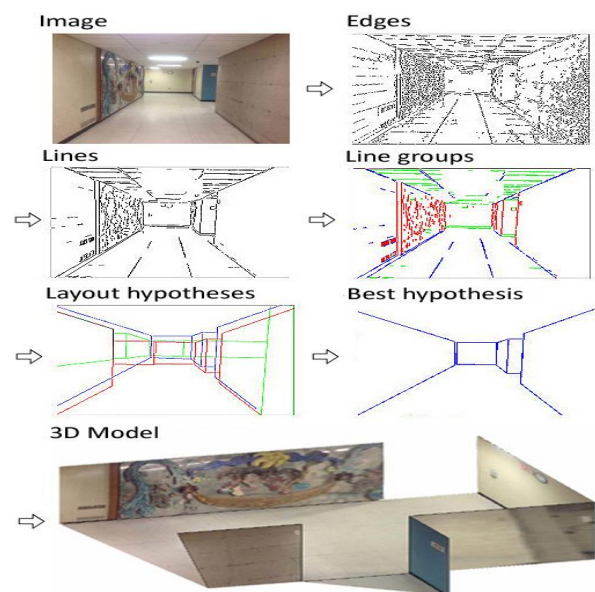


Figure 1. The proposed method detects edges and groups them into lines, and then line groups. It makes layout hypotheses using vanishing points and scores them using a linear scoring function, and finally converts the best hypothesis into 3D.

2.1 Line Grouping

Normally, many edge pixels can be extracted from a single image. The intention is to link the extracted edge pixels into straight line segments based on predefined criteria. Moreover, the straight line segments can be grouped into line groups based on their orientation. The straight line segment orientation can be identified based on its convergence into a vanishing point. It should be noted that in most of the manmade structures there are bunch of parallel lines which can provide orthogonal vanishing points (Kosecka and Zhang 2002, and Denis et al., 2008). Vanishing points are valuable for camera calibration (Kosecka and Zhang 2002; Cipolla et al., 1999; Caprile and Torre 1990, and Tardif 2009), estimation of rotation angles (Kosecka and Zhang 2002; Antone and Teller 2000, and Denis et al. 2008), and more importantly 3D reconstruction (Parodi

and Piccioli 1996, and Criminisi et al., 2000). In order to find vanishing points, different methods of straight line clustering are available (Bazin et al., 2012). There are four main categories for these methods based on: 1) Hough Transform (HT), 2) Random Sample Consensus (RANSAC), 3) Exhaustive Search on some of the unknown entities, and 4) Expectation Maximization (Bazin et al., 2012).

Here, straight line segments were extracted in the image space using LSD method (Grompone von Gioi et al., 2010). LSD method can be used on digital images for line segment extraction and it is a linear-time Line Segment Detector which can provide sub-pixel accurate results without tuning the parameters. The original idea of LSD is coming from Burns, Hanson, and Riseman's method (Burns et al., 1986), which makes use of a validation approach based on Desolneux, Moisan, and Morel's theory (Desolneux et al., 2000; and Desolneux et al., 2008).

After the extraction of straight line segments, recovering vanishing points is possible using RANSAC. In this approach two straight line segments will be randomly selected and intersected to create a vanishing point hypothesis and then count the number of other lines (inliers) that pass through this point. The drawback of RANSAC is that it does not guarantee the optimality of its solution by considering the maximum intersecting lines as inliers. Here we follow Lee et al. (2009) to find three orthogonal vanishing points. In Lee et al. (2009) the coordinates of the RANSAC solution are fine-tuned using non-linear optimization with the cost function proposed in (Rother 2000). Having found the three orthogonal vanishing points, the available line segments can be grouped into four different classes. Three of these classes are represented by the estimated vanishing points. The last class contains the line segments which are not related to the estimated vanishing points.

2.2 Layout Hypotheses Creation

Considering a single image of an indoor scene, the scene complexity may be very high to be recognized and to be modeled. Therefore, we tried to simplify the indoor scene as much as possible. For example, to modify the structure of an indoor scene, walls would be at the primary interest rather than windows or doors. Following the Manhattan rule assumption, the structure of the incoming indoor model should have a cube like formation. If the indoor scene is not bounded to only one room or one corridor, then there must be a key cube in the scene and some other side cubes which are intersecting with the key cube to form the scene layout. Therefore, the whole structure of a single model would be created based on a single cube or the integration of different single cubes. Consequently, vertical walls in the scene can only have 2 different orientations (facing the camera or being almost parallel to the camera line of sight, in case of having a vanishing point inside the image space), and floor plane and ceiling would have the same orientation. In other words, we are allowed to define 3 different surface planes in the scene which in the Cartesian coordinate system they might belong to: a) X-Y Plane, b) X-Z plane, and c) Y-Z plane.

Hedau et al. (2009) proposed a method for creation of a single box layout hypothesis by sampling pairs of rays from two furthest orthogonal vanishing points on either side of the third vanishing point. They evenly spaced the image with these vanishing point rays. However, the position of the sampling rays is dependent on the estimated coordinates for vanishing

points. Hence, this approach may not provide acceptable results when dealing with long corridors. Therefore, in the proposed approach the layout is not going to be created completely by sampling rays from vanishing points. The sampling rays will only be created if their presence is necessary for completing the process of layout hypothesis creation. In other words, these sampling rays will be employed if their presence is justified by the actual line segments. For example, the created ceiling plane can provide some information about the formation of the floor plane. Hence, sampling rays of the vanishing points can be employed to complete this formation.

In the proposed method the scene layout will be sequentially created. The whole structures of the scene (for example corridors) can be presented by cubes which are intersected to each other. For example in Figure 1, the scene layout is created by the integration of three different cubes. The camera in standing in the key cube at the time of exposure while there are two other cubes (accessory hall ways) locating at the right and left side of the key cube. Here, the key cube hypothesis is generated first. More formally, let $L_x = \{l_{x,1}, l_{x,2}, \dots, l_{x,n}\}$ and $T_x = \{t_{x,1}, t_{x,2}, \dots, t_{x,n}\}$ be the set of actual line segments and virtually generated rays of orientation x , where $x \in \{1, 2, 3\}$ denotes one of the three orthogonal orientations. A "key corridor layout hypothesis" H is created by intersecting selected lines from L_x and T_x where the minimum number of selected line segments from L_x is 4, and the total number of all lines needed for this creation is 8. Figure 2(b), shows the creation of a key cube hypothesis through intersection of solid and dashed lines which are representing actual line segments and virtually generated rays respectively.

Algorithm 1: Generating key cube hypotheses

```

Set  $H1 \leftarrow \emptyset$ , where  $H1$  is the set of key corridor hypotheses;
for all pair of line segments  $(li, lj, lg)$  which are below
    horizon do
        if  $li$  on the left  $\wedge$   $lj$  on the right side of the image  $\wedge$ 
             $li$  and  $lj$  have overlap at  $Vy \wedge lg$  belongs to  $Vx$  then
                intersect  $li$  with  $lg$  and  $lj$  with  $lg$  and
                add floor plane  $[\pi (li, lj, lg)]$  to  $Hp$ 
        end if
    end for
for all pair of line segments  $(ln, lm, lc)$  which are above
    horizon do
        if  $ln$  on the left  $\wedge$   $lm$  on the right side of the image  $\wedge$ 
             $ln$  and  $lm$  have overlap at  $Vy \wedge lc$  belongs to  $Vx$  then
                intersect  $ln$  with  $lc$  and  $lm$  with  $lc$  and add
                ceiling plane  $[\pi (ln, lm, lc)]$  to  $Hq$ 
        end if
    end for
for all  $\pi_i \in Hp$  and  $\pi_j \in Hq$  do
        if  $\pi_i \wedge \pi_j$  corners can be connected through virtual
            rays of  $Vz$  then
                add scene with 1 key cube  $(\pi_i$  and  $\pi_j)$  to
                 $H1$ 
        end if
    end for
return  $H1$ 

```

As mentioned above, in the proposed method line segments are randomly selected and intersected to form the key cube (major scene layout). It should be noted that very short line segments will be ruled out and hypothesis creation will be started by line

segments with longer length. The overall process is described in Algorithm 1. In this algorithm the general workflow of generating key cube hypothesis has been described. This will lead to the generation of many different hypothetic key cubes in the scene. It should be noted that only physically valid cubic hypotheses will be accepted in this process. Therefore, the number of created key cube hypotheses will be reduced to some extent.

Algorithm 2: Generating side cube hypotheses

```

Set  $H2 \leftarrow \emptyset$ , where  $H2$  is the set of corridor hypotheses having
side cubes;
for all line segment  $li$  which belongs to  $Vx \wedge$  vertical
plane  $\pi j$  which belongs to  $Hj \wedge Hj \in H1$  do
    if  $li$  is inside  $\pi j$  then
        connect the end points of  $li$  to  $\pi j$  borders
        through virtual rays of  $Vz$  to make a
        quadrilateral ( $\pi'j$ )  $\wedge$  add scene with side cube
        ( $\pi'j$ ) to  $Hj$ 
    end if
Set  $H2 \leftarrow H2 \cup Hj$ 
end for
return

```

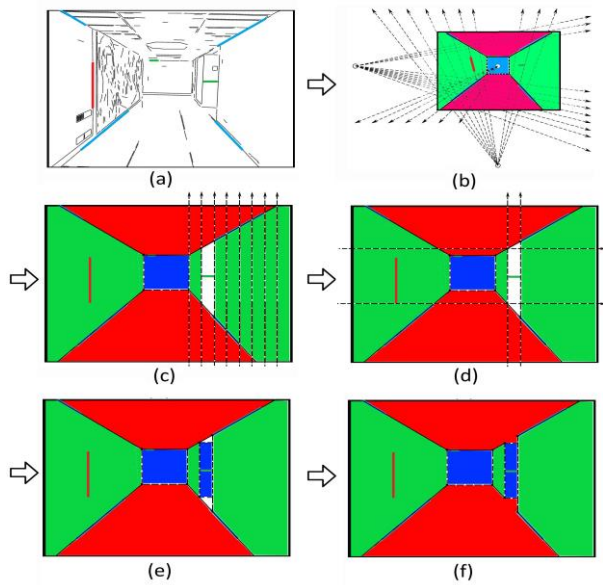


Figure 2. Creating Layout hypothesis; a) Line segments are randomly selected. b) Key cube generation by intersecting line segments (solid lines) and created virtual rays (dashed lines) using vanishing points. c) Virtual rays partitions the vertical plane. d) More partitioning using virtual rays. e) Identifying the plane facing the camera (blue plane). f) Identifying floor and ceiling regions in the created partition (white triangles in “e”).

Having created the key cube hypotheses, the presence of the side cubes (minor cubes on the sides of the key cube) will be examined. If any of the key cube side walls contain a line segment which is supposed to be perpendicular to it, then it can be treated as a hint for having a side cube in the scene. This process is described in Algorithm 2. In this algorithm the general workflow of generating side cube hypotheses has been described. Following this algorithm, many hypothetical cubes

may be generated on the sides of each key cube hypothesis. It should be noted that in this process duplicate side cube hypothesis will be deleted and also overlapping hypothesis will be merged. Moreover, only physically valid hypothesis will remain in the pool. Therefore, the number of valid hypothesis will be reduced and the remaining ones are the final scene layout hypotheses. Figure 2, describes the core part of this process intuitively.

2.3 Evaluating Layout Hypotheses

As mentioned in the previous section, in the proposed method the complete scene layout hypothesis is sequentially created through generation and integration of cubic structures. This process is performed in the image space using classified line segments and virtual rays of vanishing points. Following this rational, many scene layout hypotheses will be created. Therefore, the created hypotheses must undergo an evaluation process for selection of the best fitting hypothesis.

In order to perform the evaluation process, a linear scoring function is defined to score each hypothesis individually. Given a set of created layout hypotheses in the image space $\{h_1, h_2, \dots, h_n\} \in H$, we wish to do the mapping $S: H \rightarrow R$ which is used to define a score for the automatically generated candidate layouts for an image. It should be noted that the proposed scoring function must take some independent factors into consideration. Here, the expected value of the proposed scoring function “ S ” can be decomposed into the sum of three different components, which characterize different qualities of the created hypothesis. These components, together encode how well the created layout hypothesis represents the corridor scene in the image space. We thus have

$$S(h_i) = w_1 * S_{volume}(h_i) + w_2 * S_{edge}(h_i) + w_3 * S_{omap}(h_i) \quad (1)$$

where h_i = candidate hypothesis
 S = scoring function
 S_{volume} = scoring function for volume
 S_{edge} = scoring function for edge correspondences
 S_{omap} = scoring function for orientation map
 $w_{1,2,3}$ = weight values

As it can be seen in the above equation, the outcome of three different functions are combined together to create the proposed scoring function. Here, each function is focusing on a specific factor. These factors are: a) volume, b) edge-correspondences, and c) orientation map (Lee et al., 2009). These factors are the representatives of different qualities of the created layout hypothesis. Considering these factors, three different functions can be defined to score each quality of the hypothesis. The final score of a candidate hypothesis will be defined by summing the outcomes of these three functions. The above weight values are considered equal in the implementation procedure. However, the optimization of these weight values will be considered in the future work.

Lee et al. (2010) imposed some volumetric constraints to estimate the room layout. They model the objects as solid cubes which occupy 3D volumes in the free space defined by the room walls. Following the same rational, here the containment constraint is taken into consideration which dictates that every

object should be contained inside the room. We interpret this constraint as the search for the maximum calculated volume among all of the created layout hypotheses. Therefore, we decide to give a higher score to the layout hypothesis which has a larger volume. In other words, the layout hypothesis which covers a larger area is more probable to contain all of the objects in the room. Hence, the scoring function gives the highest volume score (score one) to the layout hypothesis which has the largest valid volume. Also, it gives the minimum volume score (score zero) to the layout hypothesis which has the smallest valid volume. Hence, the calculated score of a candidate layout hypothesis will be a positive real number between zero and one. The volume score of a candidate hypothesis (h_i) can be calculated from the following equation.

$$S_{Volume}(h_i) = \frac{V_i - V_{Min}}{V_{Max} - V_{Min}} \quad (2)$$

where S_{Volume} = scoring function for volume
 h_i = candidate layout hypothesis
 V_i = calculated volume for hypothesis h_i
 V_{min} = minimum calculated volume among all of the created layout hypotheses
 V_{max} = maximum calculated volume among all of the created layout hypotheses

Considering the above equation, the other two functions which score edge-correspondences quality of the created layout hypothesis and the compatibility of the created layout hypothesis to the orientation map are also defined in the same way. Hence, the defined function gives the highest edge-correspondences score to the layout hypothesis which has the maximum positive edge-correspondences to the actual detected line segments. Here, the positive edge correspondences is defined by counting the number of edge pixels which are residing close enough to the borders of the created layout hypothesis. Therefore, the layout hypothesis which has the biggest number of detected edges close enough to its borders will get the highest score from the proposed function (S_{edge}).

The compatibility of the created layout hypothesis to the orientation map is calculated pixel by pixel. The created layout hypothesis will provide specific orientations to each pixel in the image, and the orientation map is also conducted the same task. Therefore, by comparing these two (pixel by pixel) the compatibility between the created layout hypothesis and the orientation map can be calculated. Hence the number of pixels which get the same orientation from the created layout hypothesis and the orientation map are going to be counted. The proposed function (S_{omap}) gives the highest orientation map score to the layout hypothesis which has the most pixel-wise compatibility to the orientation map.

Considering these three functions, each hypothesis will be examined individually, and gets score based on the above mentioned factors. As mentioned before, the incoming scores will be normalized based on the maximum and minimum incoming values. The normalized scores will be integrated and the hypothesis with the maximum score will be selected as the best fitting hypothesis. Finally, the best fitting hypothesis will be converted to 3D following the method presented by Lee et al.

(2009). The only assumption made here is that all units of metrics are in camera height, i.e., the distance of the camera to the floor should be measured perpendicular to the floor and it equals 1. This is only because the absolute distances cannot be measured from a single image.

3. EXPERIMENTS

We have collected 53 various single images (corridor scenes) taken from different indoor locations at York University campus area. The images are 3264 x 2448 in size and have been taken by a smart phone (Apple iPhone 4s). Moreover, data set has metadata file which is associated with each image. It should be noted that in some of the scene frames different objects are included which obstruct the view of the scene frame. In order to prepare a ground truth, each image in the data set has manually labeled. Here, MATLAB software has been utilized to identify the exact coordinates of a line segment's end points in the 2D image coordinate system. Eventually, the set of line segments and planes which conform to the 3D orthogonal frame of the scene has been identified and the ground truth orientation for every pixel has been labeled, ignoring the occluding objects. The average percentage of pixels that have the correct orientation for each image is 87%. Also, 81% of the images had less than 20% misclassified pixels. However, only 27% of the images had less than 5% misclassified pixels. It should be noted that when objects partially occlude the floor-wall boundary, the underlying layout structure could still be recovered.



Figure 3. Examples of the created layouts which can be successfully convert to 3D.

Qualitatively, around 58% of the images returned acceptable layouts. It should be noted that even when floor-wall boundary was partially occluded by the objects or could not be detected through middle-level perceptual organization, the scene layout was successfully recovered in some images (Figure 3). When the actual line segments cannot be detected from the image, virtual rays created through vanishing points can play the same role as the actual line segments. In these cases, the key cube hypothesis will be created using both actual line segments and virtual rays. It should be noted that virtual rays cannot be

always helpful, especially when the corridor length is very large. In these cases the estimated vanishing points may not have sufficient accuracy. Therefore, the created virtual rays will be deviated from the real borders as the ray gets closer and closer to the camera.

As mentioned above, when the real boundary is occluded, the actual line segments detected from the ceiling-wall boundary along with the virtual rays of vanishing points could help identifying the underlying scene layout. However, there are some failure cases which are mostly because of inability to identify orthogonal vanishing points, detection of wrong line segments on glass surfaces or waxed floors, misaligned boundaries, no lines supporting down the corridor or fully occluded floor-wall boundaries. Some failure cases are shown in Figure 4.

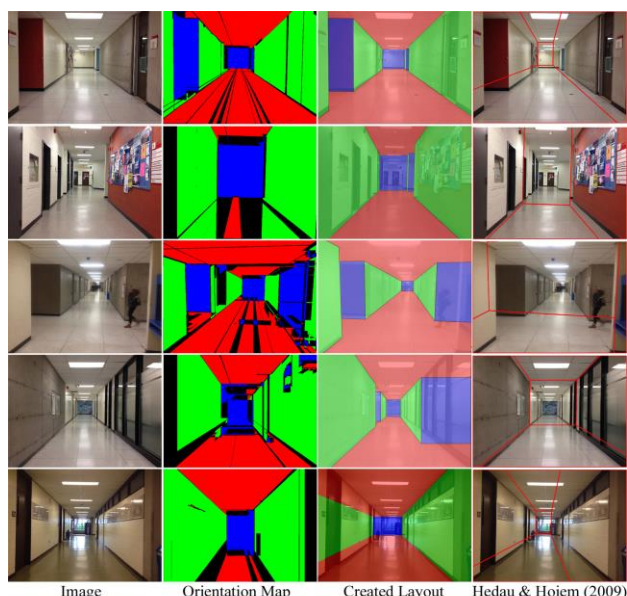


Figure 4. Example of the failure cases due to the wrong corridor depth estimation or misplacement of side cubes.

Considering Figure 4, the created layout hypotheses are deviated from the ground truth. The most conspicuous problems in the above images are: a) wrong depth estimation for the key cube hypothesis, b) wrong side cube estimation. Although the algorithm could manage to select the correct number of cubes in most of the images, the proposed edge based layout creation method could not filter out inaccurate edges (edges detected on the glass surfaces). Also, the proposed scoring function was not precise enough for selecting the best hypothesis in some cases. Consequently, the estimated depth of the key cube was wrong. Hence, the selection of independent factors of scoring function and their correspondent weights must be optimized in an adoptive way (this is an ongoing research).

For each image quantitative tables were produced to examine the out coming results. Sample tables are presented here (Tables 1 and 2) which are presenting the quantitative results of the created layout in Figure 5. Table 1, reveals the orientation difference between the estimated layout and the ground truth layout. This table can be used for evaluating the overall performance of the generated layout. Here, a comparison

between the ground truth orientation and the orientation suggested by the created layout is performed. It should be noted that this comparison is accomplished based on measuring a pixel to pixel correspondence. Therefore, if two correspondent pixels on the ground truth image and the created layout image having the same orientation, then it shows that the proposed method could correctly estimate the layout orientation at that pixel.

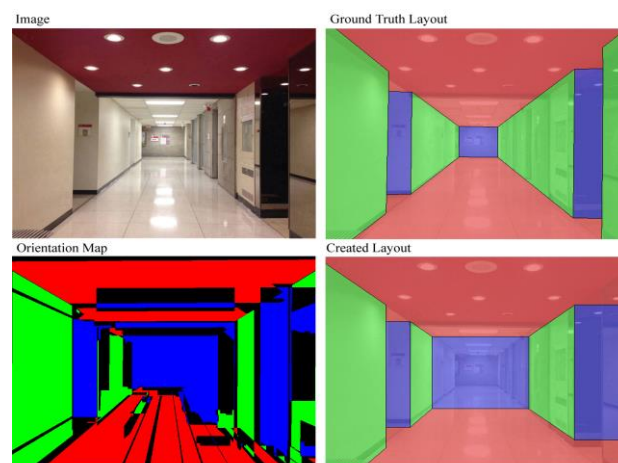


Figure 5. The created layout and the ground truth layout depicted in the image space.

Each image pixel is allowed to accept only one orientation out of three (O1, O2 or O3). The orientations are colorized by “Red”, “Green”, or “Blue” in Figure 5. Table 1, reveals the pixel to pixel orientation correspondences for the created layout in Figure 5.

Figure 5	Floor	Ceiling	Front Walls	Right Walls	Left Walls
Floor	172223	0	28448	0	4613
Ceiling	0	265663	20131	0	0
Front Walls	3478	4262	122469	1483	1060
Right Walls	15833	21729	87519	78750	0
Left Walls	0	11271	23933	0	211481

Table 1. Pixel to pixel correspondences based on orientation.

Table 1 reveals valuable information. The (i, j)-th entry in this table represents the number of pixels with ground truth label i which are estimated as label j, over the test image. As it can be seen in this table, floor, ceiling and wall estimates are partially correct and some specific regions were wrongly oriented. This can be explained by the dependence of the method on the creation of the true key cubic hypothesis (slightly deviated in Figure 5) and also the major impact of the scoring function in the selection of the best hypothesis. If the key cube hypothesis is wrongly estimated at the first step, the method could not correct this false estimation and will end up in awkward result. Therefore, a true estimation of the key cube provides a very strong condition to the success of the method.

Evaluation of the 3D reconstructed layout is possible too. 3D reconstruction is performed following the proposed approach in Lee et al. 2009. Here, three different parameters (λ_x , λ_y , λ_z) are defined for each cubic part of the layout in the object space. Considering an arbitrary 3D coordinate system in the object space, two different cubes (reconstructed in this coordinate system) can be compared using these three parameters. For example, λ_x could be defined as the width of the 3D reconstructed layout divided by the width of the ground truth layout. With the same rationale λ_y and λ_z could be defined as the ratio of length and height of the 3D reconstructed layout to the length and height of the ground truth layout. In table 2, width, length and height of the created layout are compared to the ground truth layout in the 3D space.

Figure 5	λ_x	λ_y	λ_z
Key Cube	1.1219	0.3599	1.1934
Right Cube	1.6164	2.0545	0.9976
Left Cube	0.9233	0.9494	0.9460

Table 2. Ground truth layout and created layout scale ratios.

It can be seen in the above table that the reconstructed side cube length at the right side of the layout is almost 2 times longer than the ground truth. Therefore, this table can give a better understanding about the quantitative performance of the proposed method.

Figure 6, shows the scale ratios between the reconstructed key cubic layouts and their ground truth layouts in 9 different images. These images have almost the same scene complexity, so that the comparison of their reconstructed layouts is possible. Notice that scene complexity by itself is a subjective term which may oppose confusion. Therefore, to avoid a possible confusion scene complexity is defined as a function of four major factors. These factors are: a) Type of scene layout or the number of structural planes, b) Presence of objects, c) Presence of occlusions, and d) Image depth.

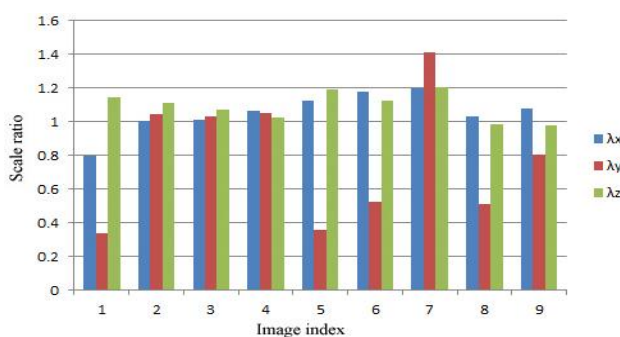


Figure 6. Scale ratios between the 3D reconstructed ground truth layout and the created layout.

As it can be seen in Figure 6, the proposed method was more successful in the estimation of scene layout width and height (λ_x and λ_z are close to 1) over the images. However, it has more problems in estimation of the true length of the corridors (λ_y is not close to 1). This is a very critical issue and it has to be scrutinized carefully (this is also an ongoing research). However, a typical explanation for this may directly emerge

from the applied equal weights in scoring function. Therefore, more experiments have to be done on this subject in the future to optimize the weights for the scoring function elements.

4. CONCLUSION

This paper focuses on 3D modeling of indoor corridors using single image. In general, modeling is not an easy task and it involves with major problems. These problems may directly inherit from the method itself and the adopted data gathering technique. Here, the proposed world model is adopted by considering the Manhattan rule assumption which simplifies the structure of the indoor layouts. However, the incoming model is not restricted to only 1 box and it can easily handle the presence of accessory hall ways and occlusions. This feature is the main advantage of this method compare to the previously proposed ones. Here, the stochastic approach is adopted, and the proposed method makes use of both actual line segments and the virtually generated rays to effectively create the scene layout. The experimental results showed that the proposed method is able to create scene layout hypotheses even if the objects are occluding some parts of the floor-wall or ceiling-wall boundaries. Also, a linear edge correspondence objective function is modified to score the created hypotheses and find the best fitting hypothesis to the image. The proposed method has shown that, by random selection of line segments, and by using some prior knowledge of indoor space, the 3D layout of an indoor corridor can be successfully recovered from a single image. A very interesting future problem would be to optimize the weight values of the proposed scoring function and moreover utilize the recovered layout to improve other integrated layouts and step towards complete indoor space modeling.

ACKNOWLEDGEMENTS

This research was supported by the Ontario government through Ontario Trillium Scholarship, NSERC Discovery, and York University. In addition, the authors wish to extend gratitude to Miss Mozhdeh Shahbazi and Mr. Kivanc Babacan, who consulted on this research.

REFERENCES

- Antone, M. and Teller, S., 2000. Automatic recovery of relative camera rotations for urban scenes. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 282–289.
- Bazin, J.C., Seo, Y., Demonceaux, C., Vasseur, P., Ikeuchi, K., Kweon, I. and Pollefeys, M., 2012. Globally Optimal Line Clustering and Vanishing Point Estimation in Manhattan World. In: Proceedings of 25th IEEE Conference in Computer Vision and Pattern Recognition, pp. 638–645.
- Burns, J. B., Hanson, A. R., Riseman, E. M., 1986. Extracting Straight Lines. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 8, no. 4, pp. 425–455.
- Caprile, B., and Torre, V., 1990. Using Vanishing Points for Camera Calibration. In: International Journal of Computer Vision, vol. 4, no. 2, pp. 127–140.
- Chao, Y.W., Choi, W., Pantofaru, C., and Savarese, S., 2013. Layout estimation of highly cluttered indoor scenes using geometric and semantic cues. In: Proceedings of the

- International Conference on Image Analysis and Processing, pp. 489-499.
- Cipolla, R., Drummond, T., Robertson, D., 1999. Camera calibration from vanishing points in images of architectural scenes. In: Proceedings of British Machine Vision Conference, 13-16 September, Nottingham, UK, pp. 382-391.
- Criminisi, A., Reid, I., and Zisserman, A., 2000. Single view metrology. *International Journal of Computer Vision*, vol. 40, no. 2, pp. 123-148.
- Delage, E., Lee, H., and Ng, A. Y., 2006. A dynamic Bayesian network model for autonomous 3D reconstruction from a single indoor image. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2418-2428.
- Denis, P., Elder, J.H., and Estrada, F.J., 2008. Efficient edge based methods for estimating Manhattan frames in urban imagery. In: Proceedings of the 10th European Conference on Computer Vision, Part II, pp. 197-210.
- Desolneux, A., Moisan, L., Morel, J. M., 2000. Meaningful Alignments. *International Journal of Computer Vision*, vol. 40, no. 1, pp. 7-23.
- Desolneux, A., Moisan, L., Morel, J.M., 2008. From Gestalt Theory to Image Analysis. *Interdisciplinary Applied Mathematics*, vol. 35. no.2, pp. 183-206.
- Fidler, S., Dickinson, S., and Urtasun, R., 2012. 3D object detection and viewpoint estimation with a deformable 3D cuboid model. In: Advances in Neural Information Processing Systems, pp. 611-619.
- Gioi, R. G., Jakubowicz, J., Morel, J. M., Randall, G., 2010. LSD: A Fast Line Segment Detector with a False Detection Control. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 4, pp. 722-732.
- Han, F., and Zhu, S.C., 2005. Bottom-up/top-down image parsing by attribute graph grammar. In: Proceedings of the IEEE International Conference on Computer Vision, vol. 2, pp. 1778-1785.
- Hedau, V., Hoiem, D., and Forsyth, D., 2009. Recovering the spatial layout of cluttered rooms. In: Proceedings of the 12th IEEE International Conference on Computer Vision, pp. 1849-1856.
- Hedau, V., and Hoiem, D., 2010. Thinking inside the box: using appearance models and context based on room geometry. In: Proceedings of the European Conference on Computer Vision, pp. 1-14.
- Hedau, V., Hoiem, D., and Forsyth, D., 2012. Recovering free space of indoor scenes from a single image. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2807-2814.
- Hoiem, D., Efros, A., and Hebert, M., 2005. Geometric context from a single image. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 654-661.
- Hoiem, D., Efros, A. A., & Hebert, M., 2007. Recovering surface layout from an image. *International Journal of Computer Vision*, vol. 75, no. 1, pp. 151-172.
- Kosecka, J., Zhang, W., 2002. Video compass. In: Proceedings of the European Conference on Computer Vision, pages 657 – 673.
- Kosecka, J., and Zhang, W., 2005. Extraction, matching, and pose recovery based on dominant rectangular structures. *Computer Vision and Image Understanding*, vol. 100, pp. 274-293.
- Lee, D.C., Hebert, M., Kanade, T., 2009. Geometric reasoning for single image structure recovery. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2136-2143.
- Lee, D.C., Gupta, A., Hebert, M., Kanade, T., 2010. Estimating spatial layout of rooms using volumetric reasoning about objects and surfaces. In: Advances in Neural Information Processing Systems, pp. 1288-1296.
- Micusik, B., Wildenauer, H., Kosecka, J., 2008. Detection and matching of rectilinear structures. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1-7.
- Parodi, P., Piccioli, G., 1996. 3D shape reconstruction by using vanishing points. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, pp. 211-217.
- Pero, L., Bowditch, J., Fried, D., Kermgard, B., Hartley, E., and Barnard, K., 2012. Bayesian geometric modeling of indoor scenes. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2719-2726.
- Rother, C., 2000. A new approach for vanishing point detection in architectural environments. In: Proceedings of 11th British Machine Vision Conference, pp. 382-391.
- Schwing, A. G., Hazan, T., Pollefeys, M., Urtasun, R., 2012. Efficient Structured Prediction for 3D Indoor Scene Understanding. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2815-2822.
- Schwing, A. G., Urtasun, R., 2012. Efficient Exact Inference for 3D Indoor Scene Understanding. In: Proceedings of the European Conference on Computer Vision, pp. 299-313.
- Schwing, A. G., Fidler, S., Pollefeys, M., Urtasun, R., 2013. Box in the box: Joint 3d layout and object reasoning from single images. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 353-360.
- Tardif, J. P., 2009. Non-iterative approach for fast and accurate vanishing point detection. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1250-1257.
- U.S. Environmental Protection Agency, 2015. The inside story: a guide to indoor air quality. http://www.fusionsvc.com/The_Inside_Story.pdf (accessed July 2015).
- Wang, H., Gould, S., Koller, D., 2010. Discriminative learning with latent variables for cluttered indoor scene understanding. In: Proceedings of the European Conference on Computer Vision, pp. 435-449.
- Yu, S., Zhang, H., Malik, J., 2008. Inferring spatial layout from a single image via depth-ordered grouping. In: Proceedings of the IEEE Workshop on Perceptual Organization in Computer Vision, pp. 1-7.
- Zhang, Y., Song, S., Tan, P., & Xiao, J., 2014. PanoContext: A whole-room 3D context model for panoramic scene understanding. In: Proceedings of the European Conference on Computer Vision, pp. 668-686.