Peripheral Expansion of Depth Information via Layout Estimation with Fisheye Camera: Supplementary Material

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

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

In this supplementary material we provide additional details of some parts of the proposed algorithm described in the article. Specifically, in Section 1 we describe more thoroughly the heuristics behind the layout hypotheses generation process. Additionally, in Section 2 we show some additional results of our algorithm in different scenarios, some visualizations of the resulting 3D point clouds and an explanation of the content of the video included in the submission.

1 Layout hypotheses generation

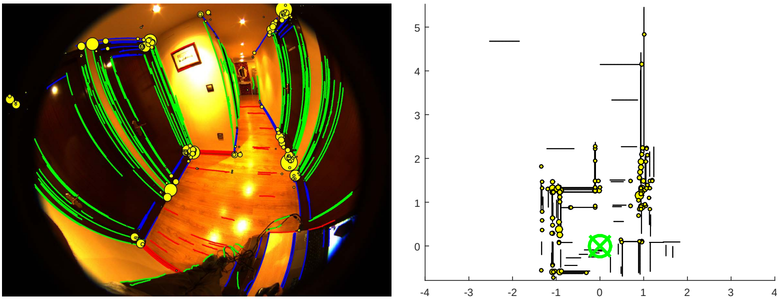
In this section we address the procedure to generate valid hypotheses given the information extracted from the camera system as described in the paper. This is a more detailed explanation of the Section 4.3 from the manuscript.

1.1 Description of the physically valid layouts

We define a layout as a set of corners in the 2D floor plane and the height of the ceiling (H_{ceil}). From previous stages we have information available of H_{ceil} and the corners present in the image and the line segments that define them (Fig. 1a). The line segments have been projected to either the floor or the ceiling (depending on lines being *upper* or *lower* lines, i.e. being above or below the horizon line), and as we have computed the floor and ceiling plane equation using the depth information, we know the real dimensions of the line segments and the 3D position of the corners with respect to the camera. From this we have the segments in a scaled 2D floor plane we call *XZ-plane* (Fig. 1b). We assume structural floor-ceiling symmetry, so corners are considered equally whether they come from the floor or the ceiling. As in most indoor environments the level of clutter is higher in the lower part of the scene, having corners from the ceiling allows to provide results in difficult environments.

Corners have been scored using the real dimensions of the line segments that define them and their distance to the intersection point, as described in the paper. Corners have higher score when:

1. They are supported by longer line segments.
2. Line segments are closer to the intersection points.
3. Corners have more lines supporting them (corners can be defined by 1, 2 or 3 lines).



(a) **(b)**

Fig. 1: (a) Relevant corners in the scene plotted over the fisheye image as yellow circles with diameter proportional to their score. (b) Same scene with the projection of the corners and line segments to the XZ-plane.

4. Line segments coincide with intersections of 3D planes from the depth information.

We assign a probability P to corners c of occurring in the real world:

$$P(c_i) = \frac{S_{c_i}}{\sum_{j=1}^{N_c} S_{c_j}}$$

where S_c is the score of a corner and N_c is the number of corners in the image. In Fig. 1 the corners are displayed as a yellow circle with diameter proportional to this probability.

To generate hypotheses we do not impose any condition about the shape of the scene in order to provide valid solutions to any kind of indoor environment. However, we consider a layout physically valid when it satisfies certain conditions:

– The walls must follow the Manhattan World convention, i.e. the structure of the world is organized according to three orthogonal directions $(\mathbf{m}_x, \mathbf{m}_y, \mathbf{m}_z)$. This means that, a wall directed in $\mathbf{m}_x$, must be followed by a wall directed in $\mathbf{m}_z$, and vice versa. An angular threshold is used to prevent possible deviations from Manhattan in real world constructions.
– We generate hypotheses of layouts only of the room the camera is in, designing closed wall distributions around the camera point.
– As the fisheye camera has a FoV greater than 180° and its slightly pointing downwards, it has a partial view of the scene behind the camera (e.g. in Fig. 1a, the wall behind the camera is partially visible at the bottom of the image). However, when the view from behind does not provide structural information, we assume the walls extend beyond the field of view towards the rear vanishing point in order to keep our layout closed.
– The layout hypotheses must not contradict the information from the depth camera, i.e. there cannot be a wall in front of the given depth map.

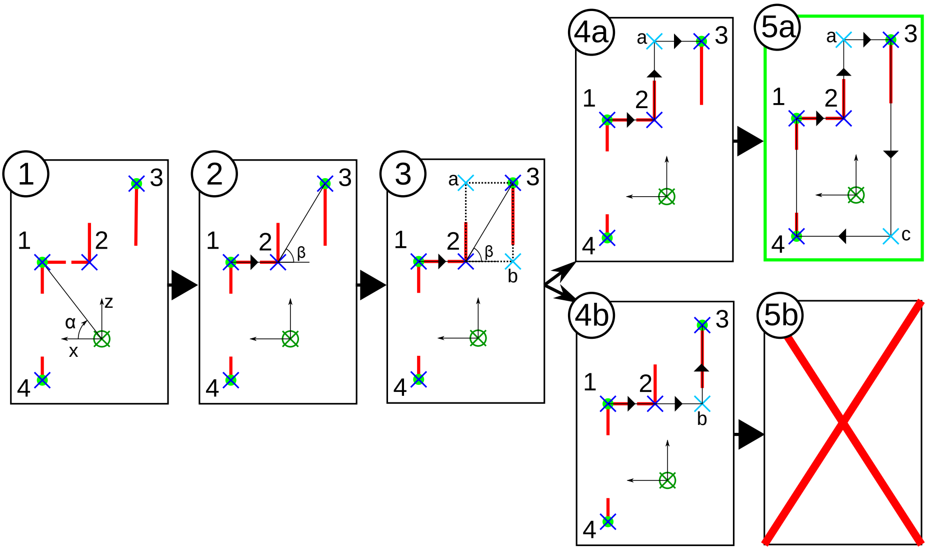


Fig. 2: Example of layout generation in the XZ-plane given a pre-selected corners from our set (in blue, with their respective red horizontal contours and a green circle when there is a vertical line). The camera position is the green circled cross. Detailed explanation of the procedure is provided in the text.

– It also must not contradict the information given by the line segments of the corners.

Though we enforce the initial layout design to be closed even if it requires to perform certain assumptions, the final solution only extends to where the field of view of the fisheye camera reaches.

1.2 Algorithm to generate hypotheses

The explanation of this section can be followed using the graphical sample case from Fig. 2. In the general case, our algorithm looks for a number of hypotheses by iterating following these steps:

1. Using the probability from Eq. 1.1, we randomly choose a number of corners from the set to generate a hypothesis. As the view of the scene is not complete and we do not impose any condition about the shape of the room, the number of corners to select cannot be fixed, and therefore it must be randomly chosen every time a hypothesis is generated.
2. The selected corners are projected to the XZ-plane and ordered clockwise considering the angle α as shown in the Fig. 2 (1).
3. The walls from the layout are going to be created joining every corner with the following one. The angle β between corners is observed to verify if the walls are oriented according to the Manhattan convention (Fig. 2 (2)):
   – If it is closer than an angular threshold to 0, $\frac{\pi}{2}$, $-\frac{\pi}{2}$ or π, then it is accepted as valid as it is (case between corners 1 and 2 in Fig. 2).

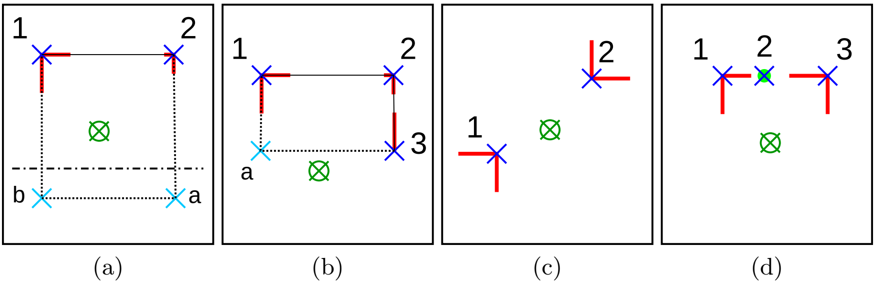


Fig. 3: Special cases in the layout hypotheses generation process. (a) The corners selected do not generate a closed layout behind the camera. (b) The layout is closed leaving the camera outside. (c) Given the contours of the line segments that define the corners, joining corners does not make physical sense. (d) Corner 2 is not a structural corner.

  - If it is not (angle between corners 2 and 3 in Fig. 2), two additional corners (a and b) are created as shown in Fig. 2 (3).
4. In case two additional corners are defined, the generation of layout goes on with consecutive corners in separate branches, as the cases (4a) and (4b) in Fig. 2.
5. At any point the line segments composing a corner can invalidate a layout generation branch. For instance, in (4b) the wall from corner 2 to corner b goes in direction X, but there is a line segment that defines corner 2 in direction X as well that goes in opposite direction. Solution (4a) matches the line segments from corner 2 perfectly.
6. Continue in every branch until the layout is closed or the solution is invalid (Fig. 2 (5b)). Finally, in Fig. 2 (5a) it can be seen how the layout is completed by defining an additional corner c as performed before, and no line segments contradict the wall distribution.

One of the keypoints of this method is that hidden corners can be estimated using the Manhattan assumption, even if there is no visible evidence of the presence of the corner in the image (e.g. in Fig. 3 corners a and c were not detected but its definition provides a physically coherent closed Manhattan layout). This means that the algorithm can handle heavy occlusions and still provide physically coherent results.

Occasionally the selected corners meet special conditions that require specific treatment. These are based on physical coherence and can be used to automatically discard hypotheses without further evaluation:

- If there is no information from behind the camera, the walls are extended until the horizon line in the back to form a closed solution, as shown in Fig. 3a. To keep layouts at a reasonable size we fix the distance of the horizon at $D_H = 10$ m.
- The camera must be inside the layout. The layout from Fig. 3b would be automatically discarded.

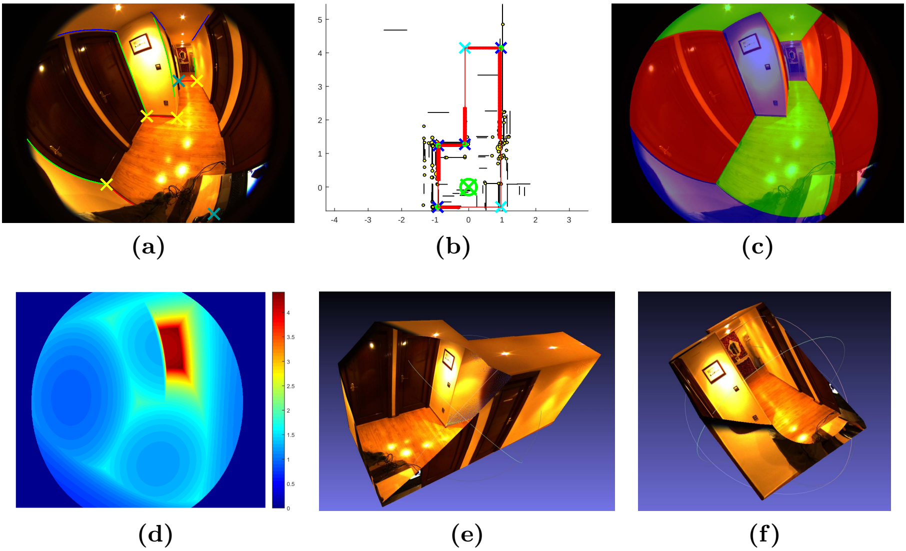


Fig. 4: (a) Layout hypotheses example with its original corners in yellow with their line segments shown and the additional corners in light blue. (b) XZ-plane with the layout overlaid. (c) Colored wall-floor-ceiling distribution of the hypotheses. (d) Corresponding depth map of the hypotheses with scale in meters. (e) and (f) Different views of the corresponding 3D point cloud.

- Given that the contours that define the corners are known, in order to join two consecutive corners they have to make physical sense. In Fig. 3c there is an example where two corners cannot be joined even with the inclusion of an additional corner.
- Consecutive corners must form walls alternatively in X and in Z direction. In Fig. 3d corner 2 is an spare corner that does not add any information to the layout, and, therefore, it can be suppressed.

In Fig. 4 there is an example of a layout hypotheses from the scene from Fig. 1, similar to the one from Fig. 2. In Fig. 4a the original corners (in yellow) and the line segments that define them have been displayed in the image along with the additional corners (in light blue). In Fig. 4b the solution has been plotted over the XZ-plane. The resulting wall distribution colored is shown in Fig. 4c. As the XZ-plane is scaled and the H_{ceil} have been estimated we can generate a 3D depth map of the scene (Fig. 4d). We can compare the result with the depth map provided by the depth camera, and use that comparison to filter hypotheses that contradict that information (i.e. there cannot be walls in front of the given depth map). The depth map can be used to recover the 3D point cloud of the complete layout, as it can be seen in Fig. 4e and Fig. 4f.

2 Additional results

Apart from the results shown in the paper, here we show a few others from different scenes or perspectives in Fig. 5 and Fig. 6, ordered depending the type of scene (corridor, bedroom, living room, other). In most cases, our algorithm provides layout solutions that fit well to the real solutions. However, there are some failure cases due to the difficulty of some scenes, where there are severe occlusions or bad lightning conditions. The most notable are:

- **Corridor scenes:** In the first scene, the occluding wall of the right is slightly out of place in SS and OM because some spurious corners appear due to reflexes in the floor. In the second scene, there are not so many corners found at the end of the corridor, which is missed in all but the OM case.
- **Living Room:** The SE case in the second scene has one of the corners in the laptop. In the last two scenes, the excessive furnishing generates spurious corners that cause overly complex layouts.
- **Bedrooms:** In the first scene only the SS case is able to find a good solution. In the third scene it is the SS case also the only one to find the wall behind the camera.
- **Other:** In the first scene, the clutter prevents the H_{ceil} to be properly obtained, causing layout proposals which generates a frontal wall going from the ceiling to a rug. In the second case, the bad lightning prevents the lines of the ceiling to be extracted.

However, in most cases there is at least one of the three proposed solutions which matches reasonably well the reality. Even in the cases where some failures occur, the high level solution is usually a good approximation of the scene.

In Fig. 7, for each case we show the fisheye image with the depth map from the depth camera overlaid, and one of the obtained 3D layout. From the first and third column it can be observed how the field of view of the depth camera is very small compared to the more than 180° of FoV we can get from the fisheye camera. With our algorithm we are able to retrieve scaled spatial layouts as the ones shown in columns 2 and 4, showing the potential of such hybrid camera system composed by depth camera (providing scale and certainty in the central part of the image) and a fisheye camera (providing large field of view to extend depth information). Note that the obtained 3D layout recovers the color information from the fisheye camera, and that the layout obtained contains only the high level information of the structure of the room. This makes objects in the scene to appear projected to the floor or walls, having a distorted look from any other perspective than the viewpoint of the fisheye camera. As the solutions we propose are scaled, we can merge the layout 3D reconstruction with the original depth point cloud to add the three-dimensional information of the central part of the image and therefore include some objects from the scene.

In the video attached to this submission, besides a brief visual summary of our method, we include some 3D visualizations from different scenes with the original and the layout point cloud merged to make the complete 3D reconstruction of the scene.

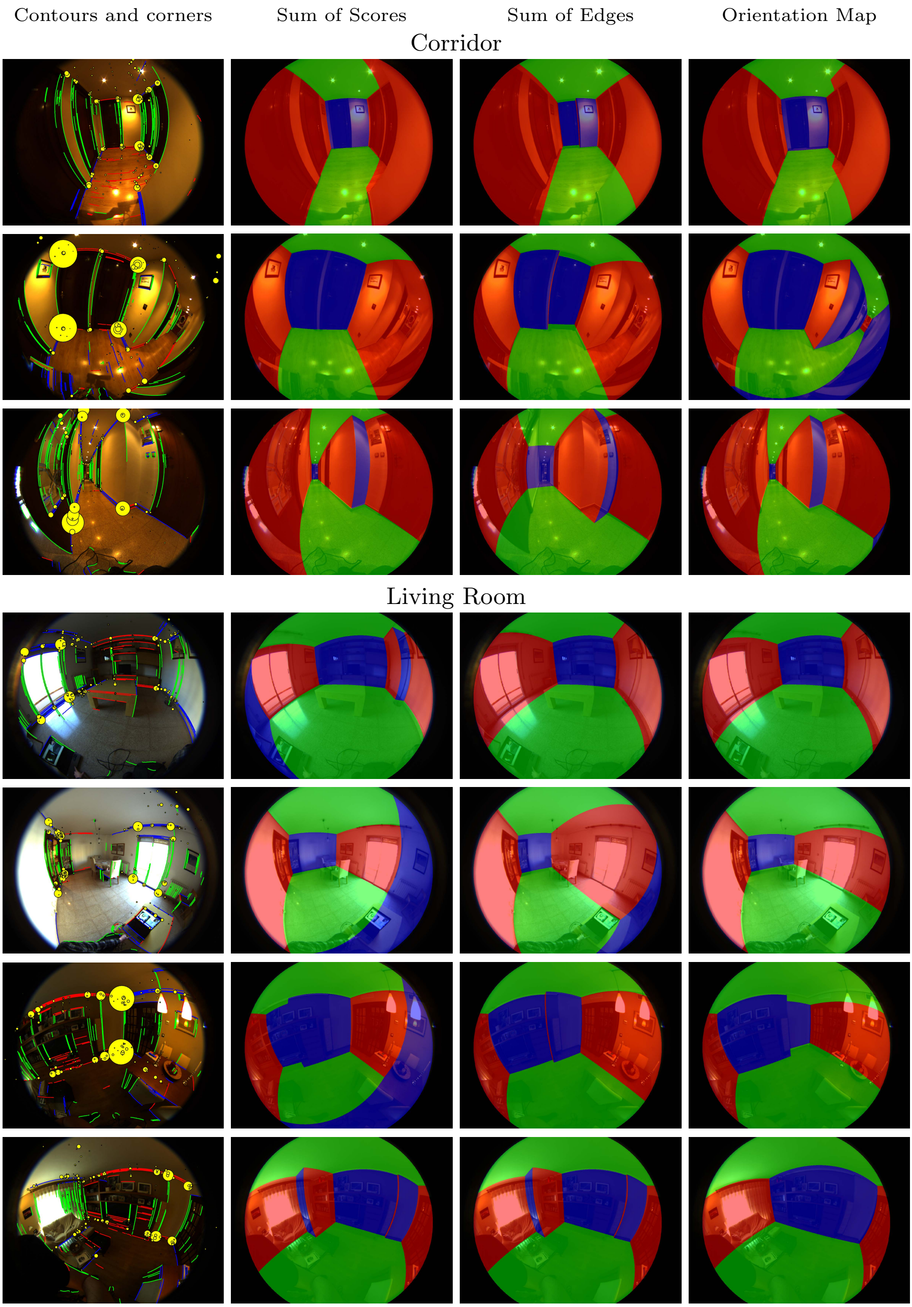


Fig. 5: Examples of results from corridor and living room images from our set with best layout proposal for each method.

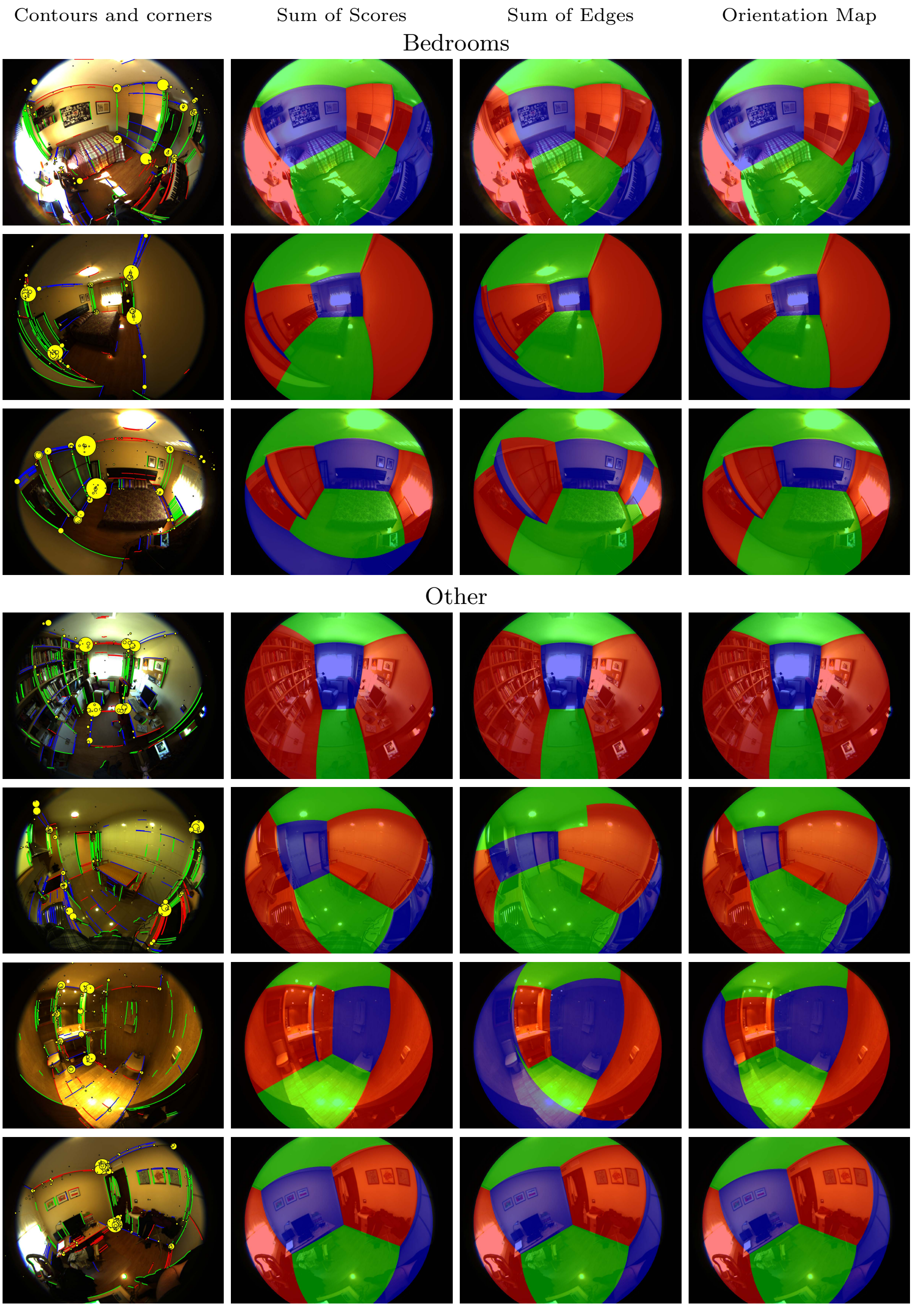


Fig. 6: Examples of results from bedroom and other (office, kitchen, bathroom) images from our set with best layout proposal for each method.

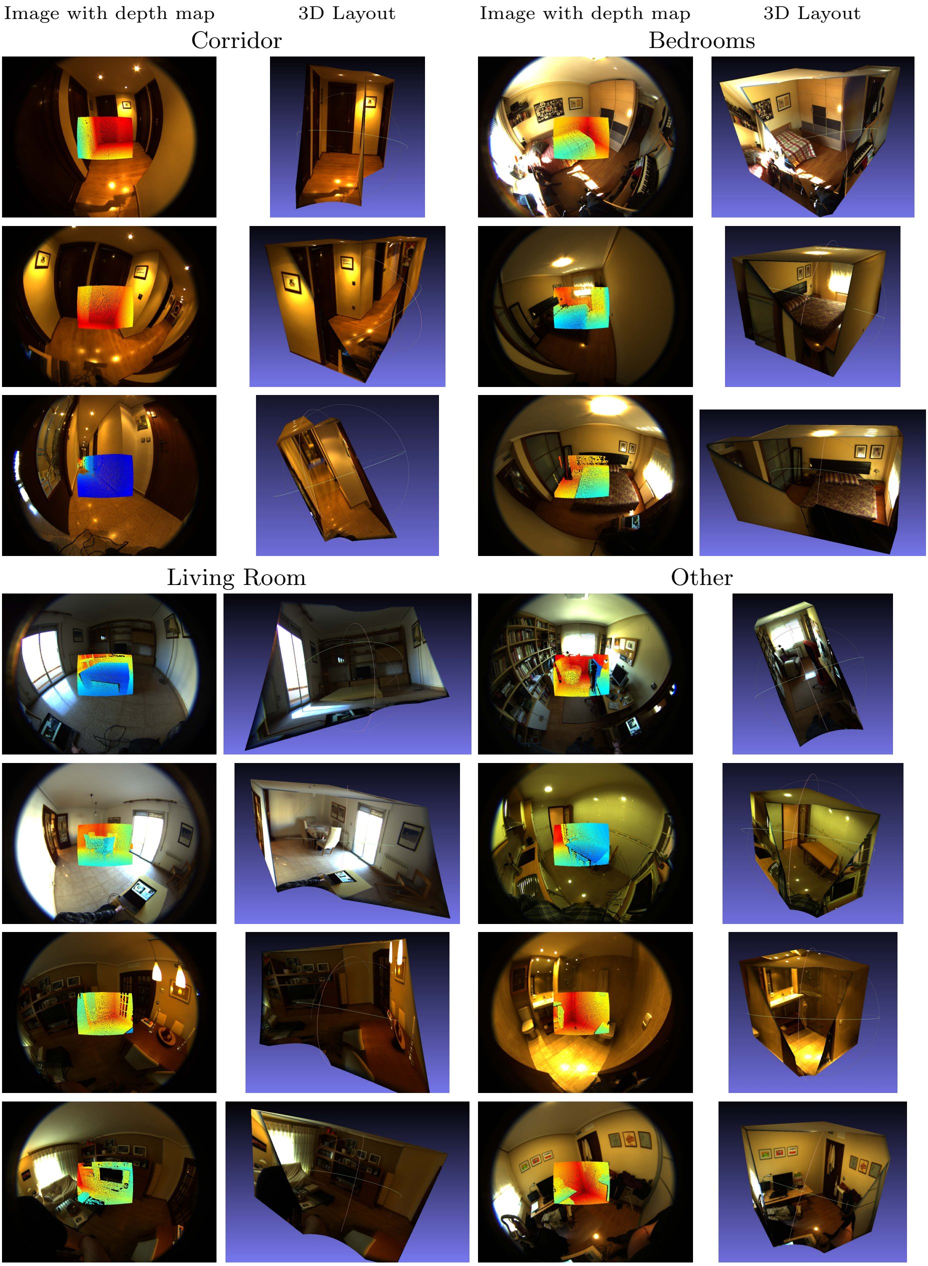


Fig. 7: Pair of images of fisheye images with the depth information from the depth camera overlaid and the 3D layouts we are able to retrieve corresponding to the cases shown in Fig. 5 and Fig. 6.

Acknowledgments

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