Plane-based detection of staircases using inverse depth

Titus Jia Jie Tang, Wen Lik Dennis Lui and Wai Ho Li
Monash University, Australia
titus.tang@monash.edu, dennis.lui@monash.edu, wai.ho.li@monash.edu

Abstract
Staircases are a common feature in urban environments. Nevertheless they often pose a navigational challenge to both visually impaired people and autonomous mobile systems. In this paper, we propose a plane-based approach to the problem of staircase detection using depth data in inverse depth coordinates. This forms a basis for our long term goal of assisting visually impaired people in navigating indoor environments. Our proposed algorithm iteratively uses Preemptive RANSAC in a segment-then-fit approach to detect steps of a staircase. This allows our algorithm to detect the presence of a staircase, identify its inclination, and model each step of the staircase as a plane model in 3D space. Experiments were conducted using a real world dataset of 121 images with manually labelled ground truth. Results shows a Type I and Type II error rate of approximately 1 and 5% respectively for the detection of staircases. Our algorithm runs at approximately 16 frames per second.

1 Introduction
Staircases are a common feature in urban environments. Despite their largely regular and predictable structure, both autonomous mobile systems and visually impaired people often encounter difficulties in identifying and navigating staircases. The ability to safely identify and navigate staircases is of particular importance as it commonly results in serious injury or damage to the visually impaired person or mobile robot from failure to do so.

A staircase can generally be described as a series of horizontal planes (steps) placed at regular intervals in space for the purpose of facilitating travel between different storeys of a building. In this paper, we define a staircase as consisting of at least three equally spaced steps as shown in Figure 1. This includes the ground plane on which the sensor is located. A staircase must have one of two directions of inclination: ascending or descending.

We are concerned with the detection of both ascending and descending staircases at close range (1 to 5 metres). The information gained from the detection of staircases is intended to be used for further scene understanding in future work. The final goal of assisting visually impaired people in detecting and safely navigating staircases forms a strong motivation for the work presented here.

Staircase detection is performed through the detection of the individual steps of a staircase. We model steps as horizontal planes in 3D space and detect them by using Preemptive RANSAC [Nister, 2005] on depth data from a Kinect sensor (Figure 2). We work in inverse depth coordinates as opposed to Euclidean coordinates as this better conforms to the sensor’s data [Lui et al., 2012]. Since using Preemptive RANSAC in a global search for planes in the image is expensive and inefficient, we iteratively apply a segment-then-fit methodology for the detection of individual steps of the staircase. Our methodology is presented in further detail in Section 2.

The presented algorithm is able to provide the end user or system with the following information:
- The presence of a staircase.
- The inclination of the detected staircase (ascending...
or descending).

- An estimate of the number of steps of a staircase in the field of view and a plane model best describing each detected step.

The above information extracted from a staircase in a scene can serve as a starting point for future work in scene understanding such as the identification of the first step of the staircase (for safe boarding onto the staircase), and detecting discontinuities and obstacles on the staircase.

1.1 Background

In the literature, work relating to the detection of staircases have been focused on one of two main applications:

1. Assistive navigation for visually impaired people [Molton, 1999; Se and Brady, 2000; Andersen and Seibel, 2001; Capi and Toda, 2011].

2. Autonomous mobile robots [Theeravithayangkura et al., 2008; Hesch et al., 2010; Ray et al., 2010; Mihankhah et al., 2009] and humanoids [Albert et al., 2001; Gutmann et al., 2004].

In the former, methodologies and experimentations are largely focused on the development of navigational aid prototypes which guide users as they navigate their environment. This involves communicating the relative orientation of a detected staircase to the visually impaired user either via audio or tactile feedback. While these cited systems succeed in orientating the user relative to staircases, none of them provide the user with top-level information regarding the type and structure of staircases necessary for more complex actions such as navigating the staircase itself.

The latter involves the development of step or staircase detection techniques to aid in robot stair-climbing behaviour. The developed stair-climbing techniques focus on the detection of specific aspects of a staircase (such as vertical edges, vertical planes, heading direction, etc) which are specific to the robot’s needs and available hardware. Demonstrated experiments show proof of concept in constrained environments.

Overall, applications related to the detection and navigation of staircases involve a balance of trade offs between system robustness and quality of information obtained. From the point of view of developing an assistive aid for staircase navigation by visually impaired people, current systems in the literature could benefit from more intelligent algorithms that are able to provide users with a broader understanding of the nature of their environment.

1.2 Contributions

Our system differs from or improves on existing work in the following aspects:

- Current plane-related approaches such as [Gutmann et al., 2004; Theeravithayangkura et al., 2008] focuses on the identification of specific parts of a staircase directly related to the robot’s needs in order to navigate the staircase. Such approaches are robotspecific and may not be suitable for use in other systems. Similar to [Delmerico et al., 2012], our work takes a more top-level approach to the general detection of staircases and its individual steps and thus allows implementation in a broader range of applications.

- [Delmerico et al., 2012] uses range discontinuities in depth images from a Kinect for the detection of the edges of steps. This requires low-lying viewing angles of staircases (<1m) which are more suited to many mobile robots. Our approach uses the horizontal planes of steps for the detection of staircases and thus requires the viewing of staircases from a greater height. Such an approach is geared towards applications for visually impaired people with head or chest mounted sensors, or humanoid robots.

- With the exception of [Hesch et al., 2010; Fair and Miller, 2001], no other existing system known to the authors explicitly deals with the problem of detecting descending staircases. Both cited papers are geared towards the alignment and traversal of staircases by low-lying wheeled robots using two separate methodologies for dealing with ascending and descending staircases. They do not perform top-level modelling of the structure of staircases. Our system uses a single unifying approach that is able to detect both ascending and descending staircases.

- By contrast with vision-based methods such as [Molton, 1999; Se and Brady, 2000; Ray et al., 2010] which uses Canny edge detection, Hough transforms and parallel line clustering for the detection of staircases, our approach uses only depth and accelerometer data. While we do not discount the usefulness of vision-based methods of staircase detection, our
approach is independent of shadows and visual textures on staircases, and does not depend on sufficient human-visible illumination in the environment.

- By contrast with approaches such as [Hub et al., 2005] which guide visually impaired people based on a priori knowledge of the environment, our approach performs detection without any apriori knowledge of the location and type of staircases. We instead perform detection based on a few geometrical constraints unique to staircases.

- To the best of the authors’ knowledge, no datasets of depth images of staircases with labelled ground truth have been published in previous work. We collected our own dataset of images of staircases and are working towards publishing it online. This dataset consists of colour and depth images, and accelerometer readings from a Kinect.

The rest of this paper is structured as such: In the next section, we provide details of our methodology which includes working in inverse depth coordinates, plane modelling, and an overview of our algorithm, including our implementation of Preemptive RANSAC. Section 3 introduces experiments performed and presents results. Finally in Section 4, we provide a discussion on the experimental results and consider future work.

2 Methodology

In this section, we first provide an overview of working in inverse depth coordinates, including the modelling of planes in inverse depth for use in Preemptive RANSAC. We then provide details of our algorithm design and elaborate on our implementation of Preemptive RANSAC.

2.1 Inverse depth coordinates

Having knowledge of the sensor’s intrinsic parameters, the inverse depth coordinates of a point $P(u,v,q)$ is calculated from image pixel coordinates as follows:

$$u = \frac{i - C_u}{f_u} \quad (1)$$
$$v = \frac{j - C_v}{f_v} \quad (2)$$

In the direction of the column and row index respectively, $i$ and $j$ are the pixel coordinates in the image, $C_u$ and $C_v$ are the principal point offsets, $f_u$ and $f_v$ are the focal lengths.

Assuming the same origin in both inverse depth and Euclidean coordinate frames, points in one space are related to the other as follows:

$$u = \frac{x}{z}, \quad v = \frac{y}{z}, \quad q = \frac{1}{z} \quad (3)$$

or in homogeneous coordinates

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \frac{1}{z} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \quad (4)$$

2.2 Plane and error modelling in inverse depth

Given the Euclidean form of a plane equation as:

$$Ax + By + Cz + D = 0 \quad (5)$$

we divide each term by the $z$ and $D$ parameters to give us

$$\frac{Ax}{Dz} + \frac{By}{Dz} + \frac{C}{D} + \frac{1}{z} = 0 \quad (6)$$

Using the relationship between inverse depth and Euclidean coordinates from Equation 4, we get

$$\alpha u + \beta v + \gamma = q \quad (7)$$

$\alpha$, $\beta$ and $\gamma$ form the infinite plane model in UVQ space and can be written as

$$\alpha = -\frac{A}{D}, \quad \beta = -\frac{B}{D}, \quad \gamma = -\frac{C}{D} \quad (8)$$

The deviation, $e_i$ of a point, $i$ from fitting a plane, $P$ can then be written as

$$e_i = |\alpha u_i + \beta v_i + \gamma p_i - q_i| \quad (9)$$

The deviation, $e_i$ of each point of interest is compared to a threshold, $T$ to identify inliers in the Preemptive RANSAC procedure. As detailed in our previous work [Tang et al., 2011; Lui et al., 2012], working in inverse depth coordinates as opposed to Euclidean coordinates results in an isotropic and homogeneous error model which better conforms to the Kinect’s depth data. Consequently, such an error model allows for error thresholds in the Preemptive RANSAC procedure, $T$ to be set constant regardless of measurement distances. Such a constant threshold allows for a simplified plane fitting methodology and is thus a main motivation for working in inverse depth coordinates. The value of $T$ is determined empirically by the degree of trade off between detected plane robustness and stability and would vary depending on the application of choice. Also note that our plane models are infinite, that is, planes have no boundaries. Plane segmentation is a topic of discussion in future work.
2.3 Algorithm design

Figure 3 provides an overview of our staircase detection algorithm. Our algorithm begins by performing ground plane detection and then uses the ground plane as a reference plane in the detection of subsequent step of the staircase. This is based on the assumption that all traversable staircases must be connected to the ground plane in which the sensor (user) is located.

We model staircases as a structure consisting of multiple parallel horizontal planes at fixed intervals in space. The detection of planar surfaces in a scene thus forms a fundamental component of our staircase detection methodology. In this paper, we perform plane detection using a form of Preemptive RANSAC as proposed by [Nister, 2005]. This is a variation to the original RANSAC methodology developed by [Fischler and Bolles, 1981]. Using Preemptive RANSAC (or any variation of RANSAC) in a global search for steps in the image is expensive and inefficient as steps may be as small as tens of pixels wide and occupying a tiny fraction of the image. We instead constrain the search space to 3D volumes representing potential locations of steps of a staircase. Such an approach is taken to limit the range of points used in our Preemptive RANSAC procedure. This segment-then-fit methodology is used iteratively in the detection of subsequent steps of the staircase and is described further in the following subsections.

Identification of ground plane

Our algorithm begins by locating the ground plane in the scene. Knowledge of the model of the ground plane in the scene acts as a starting point for our search for the steps of a staircase. This is based on the assumption that staircases that are of interest from the perspective of navigation would begin from the ground plane regardless of staircase inclination. Having an estimate of the height of the sensor from the ground plane and the orientation of the gravity vector obtained from the Kinect’s accelerometer, we are able to form a crude estimate of the ground plane’s model. We then identify points in the image which are within a height tolerance range, $H_{\text{tol}}$ from this initial plane model. This crude segmentation results in an initial pool of points which are subsequently used in Preemptive RANSAC for hypothesis generation and voting to determine the final model for the ground plane. Our implementation of Preemptive RANSAC is described in detail in Section 2.4.

Detection of steps on the staircase

We begin by identifying points which are $H_{\text{step}} \pm H_{\text{tol}}$ distance away from the ground plane. $H_{\text{step}}$ is an estimate of the height of a step; $H_{\text{tol}}$ is a tolerance range described in the previous section. This first segmentation provides us with two sets of points - above and below the ground plane. We perform Preemptive RANSAC on each set of points to determine the plane model for the first step of the staircase (if it exists in the scene). The position of the first step relative to the ground plane also provides knowledge of the inclination of the staircase.

Subsequent steps of the staircase in the scene are detected in a similar fashion as discussed above and summarised below:

1. Filter for points $H_N \pm H_{\text{tol}}$ distance away from the ground plane. $H_N = H_{\text{step}} \times N$th step of the staircase. In other words, all steps are found relative to the ground plane; the incorrect detection of any intermediate step on the staircase would not affect the detection of subsequent steps.

2. Perform Preemptive RANSAC to determine the step plane model.

Figure 4 provides snapshots of the algorithm’s output as it searches for and models steps of a staircase.

2.4 Preemptive RANSAC for plane detection

Our implementation of Preemptive RANSAC is summarised in Figure 5 and detailed in the following steps.

1. Select and filter a subset of points for use in plane hypothesis generation in step 2 by performing the following:
Figure 4: Images showing the internal workings of the staircase detection algorithm. The highlighted regions in the first row of images represent points used in Preemptive RANSAC to search for each step (Blue: ground plane. White: steps). The highlighted regions in the second row of images represent inliers of the final accepted step model. The algorithm progresses from left to right. Image pair 5 is the final output. Images best seen in colour. Note that inliers of the final step model often include image points which do not represent any part of the physical step. This is a result of modelling steps as infinite planes without considering step boundaries. As discussed in future work, we aim to filter for such points, for example through the use of local patch normals.

I Randomly select a point from the initial pool of points from the depth image chosen to be used in the RANSAC voting procedure.

II Calculate the local patch normal and accept if it is parallel to the ground plane.

III Repeat the above steps for the required number of points.

2. From the set of accepted points, generate N plane hypotheses - N is proportional to the number of accepted points. Visibly larger planes in the scene consist of a larger number of points and thus require more plane hypotheses to increase the likelihood of determining the best possible fit, and vice versa. We therefore relate the value of N to plane size to allow for the robust detection of both large and small planes while reducing computational cost where possible.

3. Discard plane hypotheses that are not parallel to the ground plane.

4. For each of the remaining hypotheses, and while at least one hypothesis remains, perform the following:

   I Subsample a set of points from the initial pool of points for use in voting.

   II Perform voting on each of the remaining hypotheses - A point is deemed to be an inlier to a plane hypothesis if its deviation from the plane is less than the threshold, T defined in the previous section.

   III Discard hypotheses that have no possible chances of being selected - All hypotheses are compared with the current best performing hypothesis in terms of votes received. Hypotheses that have no chance of obtaining more votes than the best performing hypothesis within the remaining voting cycles are discarded.

   IV Discard the worst performing hypothesis.

5. The last remaining hypothesis is accepted as the final plane model if the number of votes received is more than a minimum threshold.

Our above proposed methodology provides multiple venues for the filtering of noisy or poor fitting points and planes before they are used for subsequent generation of plane hypotheses and in the voting stages of the algorithm. This greatly improves on the robustness and cost of the algorithm.

3 Results

3.1 Data collection

We collected our own dataset of images of staircases in the Monash University Clayton campus using a Kinect. This dataset consists of pairs of both colour and depth images (640 by 480 pixels), and accelerometer readings. The Kinect was mounted on a helmet and worn by the experimenter as seen in Figure 6. This was to mimic head-mounted sensors of a navigational aid which are a likely design choice for future prototypes.

Images of staircases were taken with the experimenter standing between 1 to 5 meters from the first step of a staircase, and his head tilted in the general direction of the first step. Where possible, images were compared with ground truth and noted for the inclination of the
staircase and the number of steps seen on the staircase. This dataset consists of 65 image pairs of staircases - 38 pairs (58%) ascending and 27 pairs (42%) descending. Some examples of colour images from this dataset are shown in Figure 7. Note the variation in lighting conditions, stair surface material, and viewing angle and distance from the staircase.

Another 56 Kinect image pairs of urban environments without staircases were collected to check for false positives. Some example colour images are shown in Figure 8. We are working towards open sourcing both positive and negative datasets and intend to improve on the datasets over time.

3.2 Algorithm performance

Table 1 summarises our experimental results in terms of Type I (false positive) and Type II (false negative) errors. We tested our algorithm with the following metrics:

1. Staircase detection - As defined in our introduction, the detection of a staircase requires the detection of at least 3 steps (including the ground plane).

2. Staircase inclination (ascending or descending) - The inclination of a staircase is only evaluated given the positive detection of a staircase in an image. This being the case, tests for staircase inclination are a subset of tests for staircase detection.

The above evaluations were conducted on both positive and negative datasets. We tested each dataset 100 times due to the stochastic nature of the Preemptive RANSAC component of our algorithm. We also tested our algorithm on sequences of images taken from a head mounted Kinect to allow for a qualitative understanding of the performance of our algorithm in its intended application. Please refer to the video attached with this paper.

Our algorithm is single threaded and implemented using OpenCV libraries on Ubuntu. The algorithm run time is proportional to the number of steps detected. In this sense, time taken per detected step is a more accurate evaluation of the speed of the algorithm compared to overall frame rate as the number of steps in a scene changes from image to image. Our algorithm runs at \( \sim 13 \) ms per detected plane (or approximately 16 frames per second on the presented dataset) on an Intel i7 2.93 GHz processor. All experiments were conducted using the same parameter set.

4 Discussion

While a best effort is given in detecting the steps of a staircase, the algorithm does not guarantee the detection of all visible steps in the scene. Figure 9 shows the detection rate of steps in accordance to their physical position in a staircase. Steps located further away from the sensor have a lower chance of being detected. This
Figure 7: Example colour images from the staircase dataset collected using a Kinect. The images were collected under varying lighting conditions, staircase surfaces, and viewing angles and distances from staircases.

Figure 8: Example colour images from the negative dataset collected using a Kinect.

<table>
<thead>
<tr>
<th>Evaluation metric</th>
<th>Total tests</th>
<th>Type I error (%)</th>
<th>Type II error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Staircase detection</td>
<td>12300</td>
<td>1.02</td>
<td>5.07</td>
</tr>
<tr>
<td>Inclination - Ascending</td>
<td>6003</td>
<td>2.33</td>
<td>1.65</td>
</tr>
<tr>
<td>Inclination - Descending</td>
<td>6003</td>
<td>2.52</td>
<td>1.10</td>
</tr>
</tbody>
</table>

Table 1: A summary of the performance of the algorithm on both positive and negative datasets. We tested each dataset 100 times due to the stochastic nature of the Preemptive RANSAC component of our algorithm. These repetitions are reflected in the total tests cited in the table above. The inclination of a staircase is only tested for given the detection of a staircase. Tests for inclination are thus a subset of tests for staircase detection.
Figure 9: A plot of the step position in a staircase (moving away from the sensor beginning from the ground plane as step 0) versus true positive detection rate. The further the step is from the sensor, the lower the chances of it being detected. Our system has a detection range of up to about 9 steps.

is attributed to be the result of the following:

- Larger uncertainty and error for distant measurements result in less accurate plane hypotheses.
- Further steps of the same size are represented by less pixels in the image and therefore have a lesser chance of having sufficient good points for use in the Preemptive RANSAC procedure.
- We discard all points having distance measurements larger than 5m as they are often too noisy for meaningful use (quantisation error is $\sim 8\text{cm}$ at a distance of 5m). This practically limits the number of steps that can be detected by the algorithm (about 9 steps). The number of steps available in any single image from our dataset (ground truth) varies from 5 to 21 steps.

Overall, the performance of our presented algorithm is on par with recently published papers on staircase detection such as [Wang and Wang, 2009; Hernandez and Jo, 2010]. [Wang and Wang, 2009] uses Real AdaBoost for staircase detection and demonstrates a zero false positive detection rate and an accuracy of approximately 70% based on their own performance metric and dataset of 555 colour images. However, their proposed method only detects the presence or absence of a staircase and does not provide any other information relating to the pose and structure of detected staircases. Visual segmentation of outdoor staircases by [Hernandez and Jo, 2010] has an average Type I and Type II error rate of approximately 5% based on a dataset of only six images. Their methodology is dependant on the detection of vanishing points of staircases as well as regions in images with high concentrations of horizontal lines. This requires staircases to be in full view and head-on to the sensor; whereas our proposed approach is able to detect both wholly and partially seen staircases with varying approach angles given the detection of the ground plane as demonstrated in our experiments.

Both example papers cited here are vision-based and thus dependant on lighting conditions for accurate detection of staircases. Our system’s robustness in handling variations in lighting conditions (apart from the presence of interfering infrared light, such as sunlight) makes it a suitable tool for indoor assistive navigation. Figure 10 provides an example of how the system works in low visible illumination conditions.

On the other hand, the performance of our proposed algorithm is conditional on the following criteria as mentioned throughout this paper:

- The structure of staircases to be detected should not vary too greatly from our generic model of a staircase defined in this paper. This requires staircases to have a constant step height and be connected to the ground plane on which the sensor is positioned.
- Part of the ground plane should be within the field of view of the sensor.
- The approximate height of the sensor from the ground plane provided to the system should be within the height tolerance range, $H_{tol}$ of the true height. This is generally a valid assumption for a walking person and most mobile robots. Given that our system is designed to operate at the height of a human adult, it should be suitable for full-sized humanoid platforms as well as many telepresence robots.
- The horizontal planar surfaces of steps should be visible; the algorithm does not work for low-lying robots.
- The location of staircases need to be within the working distance of the depth sensor used (up to 10 metres for the Kinect sensor).

5 Future work

Note in the examples provided in this paper (Figures 4 and 10) that inliers of a plane representing a detected step often include points that are obviously not part of the step itself. This is due to the fact that steps are modelled as infinite planes and indiscriminately take as inliers any point within the error tolerance range of the plane model. Future work aims at developing a method for the filtering of points not belonging to a step. This is expected to enable us to:

- Identify the bounded 3D model of a plane representing a step in the real world.
• Taking into account the repetitive nature of the structure of a staircase - predict the possible locations of further steps not in the field of view.

• With the above two points in mind - identify and filter for planes that do not belong to a staircase.

In addition to the above, our current implementation of the proposed algorithm uses a constant value for the estimated height of each step of a staircase, $H_{\text{step}}$. This poses a problem if the actual height of steps deviate from our estimate. We intend, in future work, to allow for an adaptive step height estimate based on the height of steps already detected in the image.

Effectively communicating the output of the system to the end user is an important aspect of any navigational aid for the visually impaired. Audio and tactile cues are two of the most common forms of feedback provided by navigational aids for the visually impaired as seen in the literature. Specific to the problem of staircase detection, the following information may be deemed useful to a visually impaired person:

• The presence of a staircase.

• The direction of the staircase relative to the visually impaired user.

• The inclination of the staircase.

• The approximate height of steps on the staircase.

• The approximate distance of the first step of the staircase from the visually impaired user.

All the above information can be estimated from our existing framework. The contents of a final navigational prototype, however, would depend on user testing and feedback. We intend to conduct human factors experiments in the future upon development of our navigational aid prototype for the visually impaired.

6 Conclusion

This paper has presented a plane-based approach to staircase detection. The algorithm detects steps of a staircase iteratively relative to the ground plane using Preemptive RANSAC on depth data obtained from a Kinect. Such an approach allows not only for the detection of the presence of a staircase, but also allows us to identify the inclination of the staircase and model each detected step of the staircase as a plane model in 3D space. Experiments involved evaluating the algorithm’s performance using ground truth labelled datasets. A Type I and Type II error rate of approximately 1 and 5% respectively is achieved for the detection of staircases. Our algorithm runs at approximately 16 frames per second.

Acknowledgement

This work was supported by the Australian Research Council Special Research Initiative in Bionic Vision and Sciences (SRI 1000006).

References


