DEPTH-AWARE INDOOR STAIRCASE DETECTION AND RECOGNITION FOR THE VISUALLY IMPAIRED

Rai Munoz *Xuejian Rong* *Yingli Tian*
Dept. of Electrical Engineering
The City College of New York, CUNY
New York, NY 10031
rmunoz00@citymail.cuny.edu, {xrong, ytian}@ccny.cuny.edu

ABSTRACT

A mobile vision-based navigation aid is capable of assisting the visually impaired to travel independently, especially in unfamiliar environments. Despite many effective navigation algorithms having been developed in recent decades, accurate, efficient, and reliable staircase detection in indoor navigation still remains to be a challenging problem. In this paper, we propose an effective indoor staircase detection algorithm based on an RGB-D camera. The candidates of staircases are first detected from RGB frames by extracting a set of concurrent parallel lines based on Hough transform. The complement depth frames are further employed to recognize the staircase candidates as upstairs, downstairs, and negatives (i.e., corridors). A support vector machine (SVM) based multi-classifier is trained and tested for the staircase recognition with our newly collected staircase dataset. The detection and recognition results demonstrate the effectiveness and efficiency of the proposed algorithm.

***Index Terms*—** RGB-D Camera, Staircase Detection, Indoor Navigation, Visually Impaired

1. INTRODUCTION

There are about 25.2 million adult Americans (over 8%), who are blind or visually impaired based on the 2008 National Health Interview Survey[1]. In worldwide, 45 million are blind of the 314 million visually impaired people[2]. Independent travel in unfamiliar environments is well known to present significant challenges for individuals with severe vision impairment, therefore, it is important to address the increased potential risk of falling for the visually impaired, especially downstairs. Staircase detection should be an essential function for a navigation and wayfinding aid for visually impaired people. Over recent years, many assistant technologies have been developed to facilitate independent navigation, obstacle detection, and wayfinding tasks [1, 2, 3], but few of them can detect staircases.

In this paper, a robust and accurate indoor staircase detection approach is proposed, which is composed of the RGB image preprocessing, staircase candidate detection and validation, and staircase recognition. Generally, the preprocessing will extract edge information and then Hough transform is applied to detect a set of parallel lines with a series of geometric constraints to ensure potential staircase candidates.

The proposed algorithm can further distinguish escalators from stationary stairs by applying the optical flow-based tracking, which has not been addressed by previous methods. The stationary staircase candidates have the average of the lines' midpoints used as the reference in the depth frame to extract distance information of the staircase steps. The depth features are then fed into an SVM-based multi-classifier to differentiate upstairs, downstairs, and negative data.

In our prototype development and algorithm validation, a Google Project Tango Tablet [4] mounted at the chest position of the user is adopted due to the following advantages: 1) capability of capturing both RGB and depth images, 2) wide range field of view, and 3) low-cost, efficiency, and compact size w.r.t. the Microsoft Kinect incorporated with a regular computer. These features allow for a reliable and proficient video acquisition and testing of the algorithm across different indoor environments. For effectively conveying the detected and recognized results to the blind user, we also integrate the Text-to-Speech (TTS) and Speech-to-Text (STT) modules which could help the blind user perceive and interact with the surrounding unfamiliar environments via voice information.

The remainder of the paper is organized as follows: Section 2 summarizes the related work of staircase detection. Section 3 discusses the proposed staircase candidate detection algorithm and how to identify if the staircase candidate is moving (potential escalators). Section 4 discusses the experimental results. Section 5 concludes the paper and presents future work.

2. RELATED WORK

Staircases widely exist in man-made environments and as such remain a major problem in robot and human navigation. Many different kinds of sensors, like monocular [5, 2, 6, 7] and stereo cameras [3], or laser scanning [1] devices (e.g., LiDAR), have been used for detecting staircases.

[1] http://www.cdc.gov/nchs/nhis/nhis_ad.htm
[2] http://tinyurl.com/who-blindness

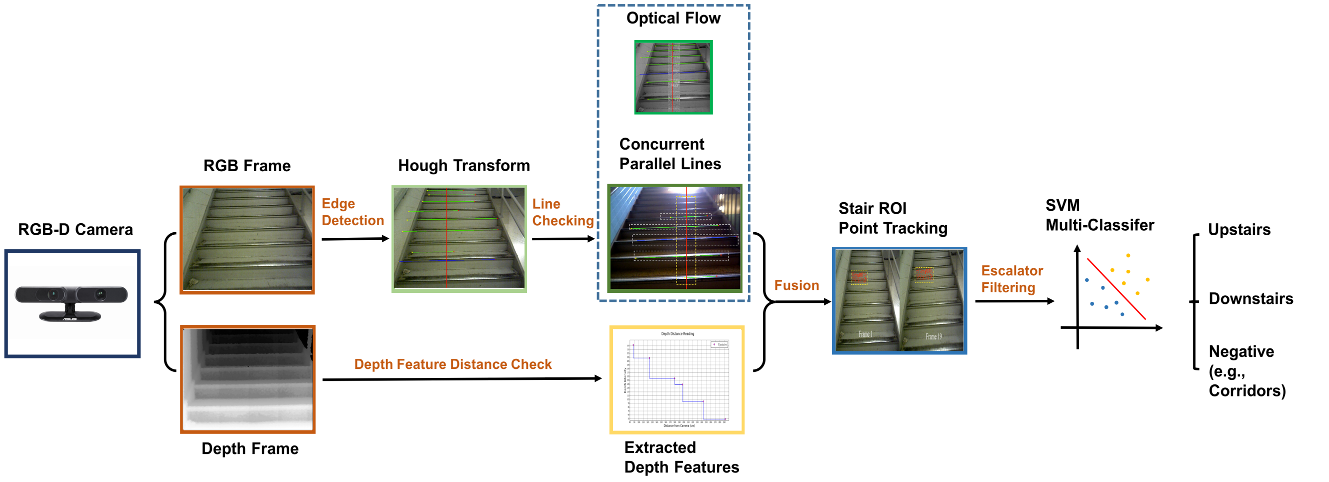


Fig. 1: Flowchart of the proposed algorithm for staircase detection and recognition

Lee and Kim [8] optimally detected the staircase in real time through utilizing a stereo camera attached to a pair of glasses along with a vest containing feedback effectors. Through extracting ground floor estimation and temporal consistency information from the stereo images, they obtained favorable results in detecting staircases, but only focused on detecting the location of the staircase without classifying downstairs and upstairs. Kim et al. [9] suggested incorporating a stereo camera into the white cane and using actuators for guidance and distance feedback. The cane utilizes hand gestures and the visual information delivery assistant (VIDA) to identify the object and portray the distance information via actuator feedback. It is useful in providing user-selected information, but its semi-autonomous design does not take into account the obstacles presented in motion.

The existing algorithms for staircase detection are still far from satisfactory and can be significantly improved. Recently, RGB-D cameras (e.g., Microsoft Kinect) and RGB-D mobile devices (e.g., Google Project Tango Tablet), are widely used in the fields of computer vision and robotics, for the capability of capturing both color and depth information of the environment simultaneously, and generating corresponding video streams. In staircase detection to assist blind people, the depth sensor can provide distance information to blind users. Moreover, the infrared-based depth perception is robust to the environment textures or low illumination environments. In paper [10], the authors modeled a stair's tread and rise through using cloud points, scene segmentation, and geometry constraints. Wang et al. [11, 12] proposed an RGB-D image-based detection approach of stairs, pedestrian crosswalks, and traffic signs, which achieves decent detection rate in the staircase detection, but cannot handle the escalator detection.

In this paper, we present an efficient algorithm to detect staircases and recognize the types of staircases in using RGB-D videos. A new benchmark staircase detection dataset is captured from the RGB-D camera mounted at the chest position. As indicated in [13], the chest position promotes desirable features according to body motion and social acceptability.

Our main contributions are, 1) an end-to-end (raw RGB-D video frames in, staircase category, location, and orientation out) real-time staircase detection and recognition approach is proposed and achieves the state-of-the-art results. 2) A new benchmark staircase detection dataset is collected which consists of more than 100 staircases, including extensive staircase shapes and appearances. 3) An optical flow-based stair tracking algorithm is proposed to distinguish the escalators from the stationary staircases which have never been addressed by previous methods.

3. RGB-D IMAGE-BASED STAIRCASE DETECTION

The staircase is defined as a set of stairs and its supporting structures, with each having its own variety of steps. In our design, we are inclined towards implementing the algorithm to detect the orientation of staircases with homogeneous stair treads and steps.

As shown in Figure 1, firstly, the edge maps of acquired RGB images are generated using the Sobel edge detector. Afterward, staircase candidates with a set of concurrent parallel lines are proposed using Hough transform, which are further validated via the depth channel, and tracked across the frame sequence using the average midpoint of the detected lines. Optical Flow features extracted along the midpoint line are introduced to identify if the staircase is moving. In a stationary staircase image, the 480 feature vector obtained along the midpoint line in the depth image is fed into the SVM multi-classifier used to categorize the staircases as either upstairs, downstairs, or negatives, with 3-fold cross validation on randomly divided subsets of the data. In addition, this feature vector provides the distance between the user and staircase steps.

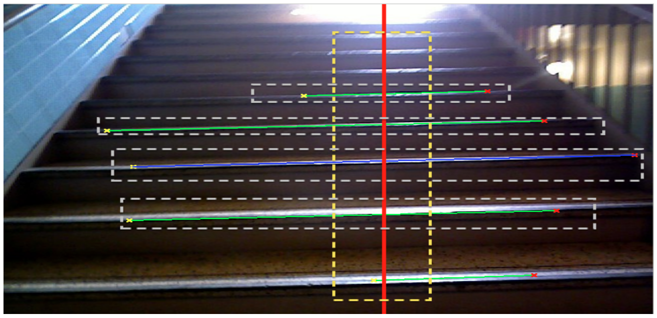


Fig. 2: Example of staircase candidate detection on RGB channel with a set of parallel lines. Yellow box represents midpoint check results while the white boxes represent overlap tests results on potential stair candidates.

3.1. Staircase Step Candidate Detection

The structure of staircases appears as a set of concurrent parallel lines in images. After applying the histogram equalization and Sobel edge detection, Hough transform is employed to the extracted edge map to detect potential staircase edge lines. These lines are further validated if they belong to a set of concurrent parallel lines, which prevent noises from unexpected lines and promote a more robust detection of the stairs. These edge points, represented by (x_i, y_i), form a line represented in slope-intercept form $y = ax + b$, where a represents the slope of the line while b represents the y-intercept. In our system implementation, to better parameterize a set of lines, the following equation is used:

$$r + y \cdot \sin\theta + x \cdot \cos\theta = 0, \tag{1}$$

where θ represents the angle of the line relative to a horizontal alignment, while r represents the line's length from the Hough edge endpoints.

As seen in Figure 1, the staircases have been recorded in the RGB and Depth channel. In order to promote more effective detection, a set of rules and constraints has been applied on the channels to screen for potential staircase candidates.

- Line length r should be between 100 and 500 pixels.
- Number of unfiltered Hough Transform lines in each RGB frame, $h_{line} \geq 2$
- Angle range should be between $\theta \in [85°, 90°] \cup [-85°, -90°]$ respectively.
- The average midpoint horizontal direction range limit uses x_{lower} and x_{upper} as constant pixel bounds for the candidates' midpoint x_{mid} in the $x_{mid} - x_{lower} \leq x_{mid} \leq x_{mid} + x_{upper}$ bounds
- Dynamic vertical direction line limit utilizes the upper and lower bound decreasing coefficients, α and β, and the growing number of lines fitting the limit, n_{line}, to decrease the bounds and prevent overlapping of lines. $y_{upperlimit} - (\alpha * n_{line}) \geq |y_{avg_i} - y_{avg_{i-1}}| \geq y_{lowerlimit} - (\beta * n_{line})$.
- Depth distance limit is to maintain distance detection range, $|D_{local}| \leq |D_{limit}|$. D_{local} denotes the distance difference from each independent line with the closest line to user in the current frame, while D_{limit} represents the depth distance difference limit measured in the initial frames.

These rules are applied on all parallel lines (the green lines while the blue line is the longest), with the average midpoint line represented by the red line in Figure 2. If the candidate satisfies the depth distance constraint, then it will manifest in the depth image as a blue point. All constraints are used in the initial detection of the staircase orientation. However, the depth distance, line, and vertical direction constraints are exempted during the user's movement towards the staircase.

3.2. Escalator and Stationary Staircase Identification

During the staircase candidate detection, sparse optical flow is applied to the RGB images to track the candidates along the midpoint line. The Horn-Schunck method is adopted for the optical flow estimation. The gradient of the optical flow measured from the user initially standing still is utilized to determine the user's movement and then filter it from the staircases' movement. Similar to the depth distance constraint mentioned before, the optical flow relies on the video frames to obtain the optical flow vector and the derivatives of the image intensities. Using these values, the method can fabricate flow vectors along the red midpoint line, represented by white markers in Figure 5.

In addition to the optical flow, the Kanade-Lucas-Tomasi (KLT) point tracking algorithm is employed [14], which uses the Shi-Tomasi corner detection system to compute eigenfeatures of a Region of Interest (ROI). Through using an image patch area denoted by (u, v) and shifting it by (x, y), the weighted sum of square distances $\chi(x, y)$ between the image patches is approximated in matrix form by the equation below, where $w(u, v)$ denotes the weights used, A represents the Harris tensor matrix, and I represents the image used.

$$\chi(x, y) \approx w(u, v) A \begin{pmatrix} x \\ y \end{pmatrix} \tag{2}$$

A is determined by:

$$A \approx \begin{bmatrix} \langle I_x^2 \rangle & \langle I_x I_y \rangle \\ \langle I_x I_y \rangle & \langle I_y^2 \rangle \end{bmatrix} \tag{3}$$

The eigenfeatures are established from the minimum of the eigenvalues, $\min(\lambda_1, \lambda_2)$, of the A matrix as it allows for faster computation. Using these corner features as input, the point tracking algorithm determines the KLT optical flow path. The 150×50 pixel rectangular ROI is selected around a single staircase step candidate in the first frame, and the KLT point tracker will continue to track these feature points across the video frames. Using the position of features from the last frame, the class of staircase can be deduced based on the relationship below,

$$|\frac{\sum(y_{last}(i) - y_{first}(i))}{n \times features}| \geq y_{thresh}, \tag{4}$$

where $y(i)$ denotes the vertical direction of the feature and n denotes the total number of features present. If the equation is satisfied, then the staircase is potentially an escalator. If it is not, the staircase is stationary and will have the 480 feature midpoint vector used in the SVM classification. The purpose of this check is to notify the user of the different risks in climbing either staircase. An example of escalator detection and feature displacement is demonstrated in Figure 3.

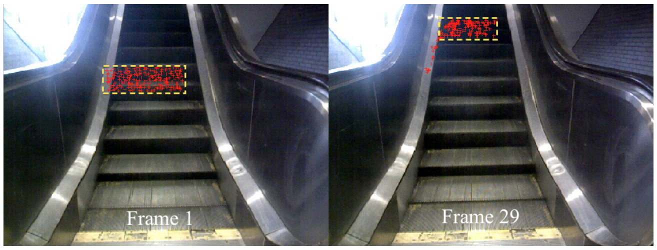


Fig. 3: Montage of Tracking Feature implemented on an upstairs escalator. The yellow box represents the region of interest selected while the red markers represent the eigenfeatures extracted from the stair.

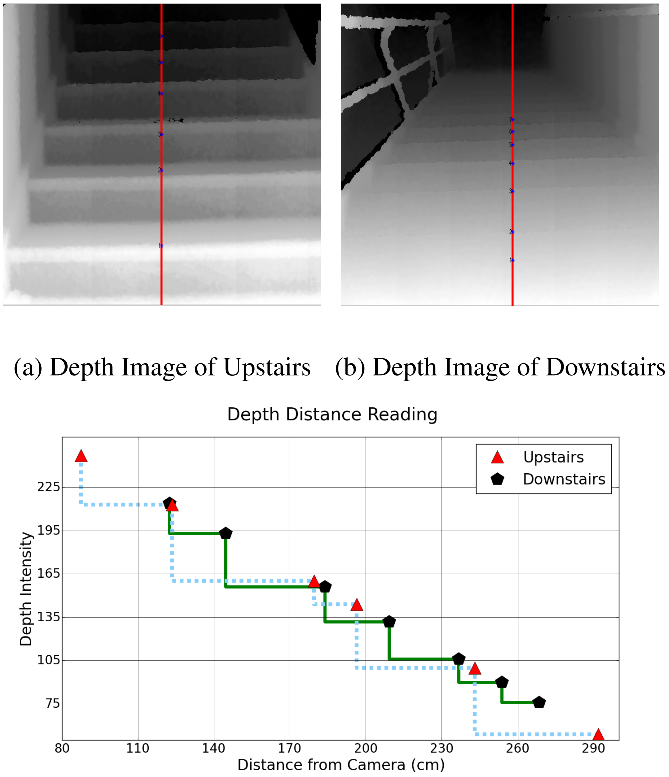
3.3. Staircase Distance Estimation

In order to estimate the distance between the user, staircase, and the size of each stair step, the depth channel is employed. The resolution of the depth images is 640×480 pixels with the effective distance range from 0.8 meters to 3.5 meters according to [15]. Each depth image is represented as an image with intensities $[0, 255]$ as seen in the depth images in Figure 4. The darker intensities indicate farther distances while the brighter intensities indicate closer distances.

The distance of each potential stair step from the user is calculated using a linear correlation between a step's midpoint depth intensity and the effective camera distance range. The stair steps have been plotted as star points in the image along the red midpoint line. The intensities from the star points render the distance in centimeters and are plotted accordingly in Figure 4. The red triangles on the blue dash line plot describe the staircase steps detected in the upstairs while the black pentagons on the green line denote the downstairs. As shown in Figure 4, the upstairs demonstrate a larger distance change between steps compared to the downstairs. The upstair steps are also more distinguishable than downstair steps.

In consideration of these particular characteristics, the depth distance constraint balances the overall stair step detection distance range. The difference in distance between the nearest and farthest step is evaluated across an initial number of frames. The overall average difference is used as a control value for detection in future frames.

By utilizing this control value, the detection can be further improved by reducing any noise produced from outside the distance range of the staircase. In our implementation, the overall control value range of the different staircases varies from 100 cm to 300 cm but never exceeds the threshold of 330 cm. For the stationary staircase classification, the intensity values along the red midpoint line are extracted. The dimensions of the derived depth feature vector are 480×1 and the corresponding distances of the vector are input into the SVM multi-classifier.



(a) Depth Image of Upstairs (b) Depth Image of Downstairs

(c) Depth Image Distance Graph

Fig. 4: Examples of depth images for upstairs, downstairs, and the distance calculation.

3.4. Recognizing Directions of Staircases

Some of the detected staircase candidates may be negative samples (i.e. non-staircases.) To recognize upstairs, downstairs, and negatives, we train a Support Vector Machine (SVM) based classifier with a radial basis function (RBF) kernel [16]. The depth feature vector is fed into the SVM classifier as input. We implement the SVM as a multi-classifier for 3 classes: upstairs, downstairs, and negatives. For each class, one classifier is trained by considering the samples of this class as positives while the rest are as negatives. Cross-validation among 3 subsets of the training data is conducted to prevent the biasing of the dataset and possible over-fitting.

3.5. Speech-based User Interaction

After completing the staircase detection and recognition process, we further implement the Text-to-Speech (TTS) module to convey the results to blind users, including the information of the step distance, staircase direction, the number of remaining steps, and etc. The Android built-in speech synthesis engine is adopted in our system to transform the predefined phrases and processing results to the voice output, which provides adaptive navigation support to the blind users.

The CMU Sphinx [17] speech recognition engine is further employed to receive the voice commands from the user, including but not limited to, pausing, resuming, stopping, and restarting the staircase detection processing. The effectiveness of the TTS and STT modules has been validated in the experiments, and proven to significantly boost the practicability of our proposed algorithm and system.

4. EXPERIMENTAL RESULTS

4.1. RGB-D Camera Data Collection

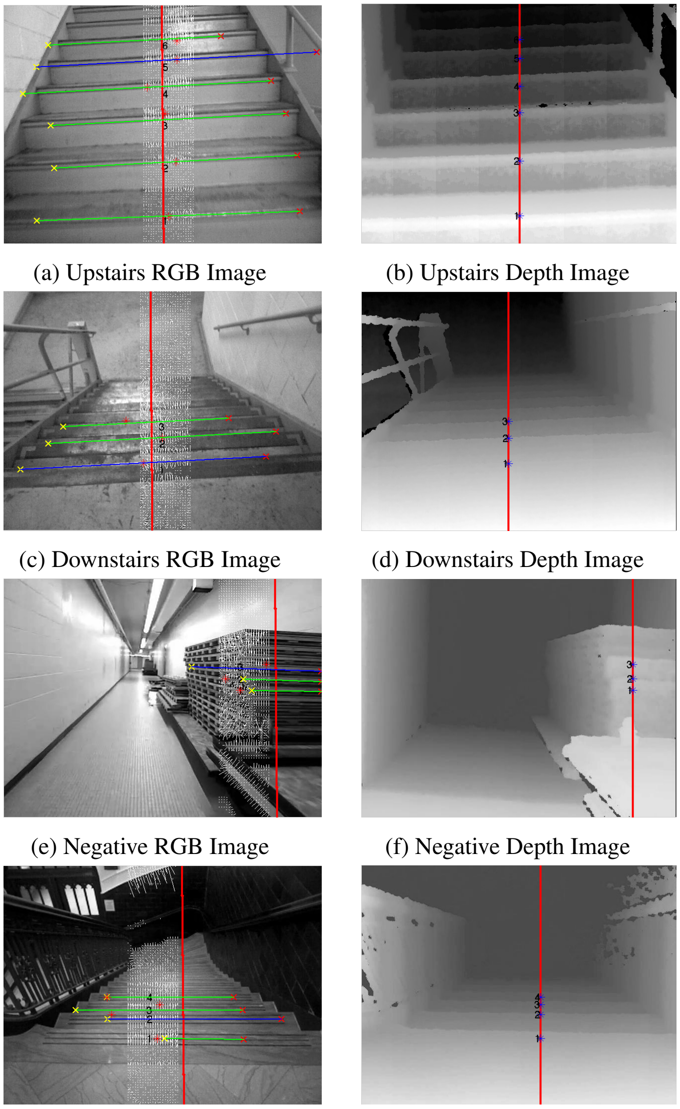
To evaluate the proposed algorithm, we collect a database of staircases including upstairs, downstairs, and negatives by using an RGB-D camera in a variety of indoor environments. In our system, the camera is mounted in the chest position-facing front. The videos of RGB and depth data are simultaneously captured at 30 FPS of staircases. The staircases of the database are captured in a variety of indoor environments including office buildings and homes. For each staircase, a video for approximately 10 seconds is recorded for stair detection and classification. The video is partitioned into frames at 10 frames per second to be used for analysis. The videos without staircases are collected at negative data including objects such as corridors, bookcases or ladders. Examples are displayed in Figure 5. Our database contains frames captured from 115 upstairs, 111 downstairs, and 120 negative data.

4.2. Experimental Results

The constraint values were preselected before applying the series of candidate tests. For the average midpoint constraints, the x_{lower} and x_{upper} bounds were set to 100 pixels. In the dynamic vertical line limit, the optimal α coefficient is set to 12 pixels per line and β to 2 pixels per line with the $y_{upperlimit}$ assigned 120 pixels and $y_{lowerlimit}$ assigned 30 pixels. The depth distance constraint $|D_{limit}|$ is set by the initial frames, but has a default value of 210 pixels, in case no candidates are detected, while $|D_{local}|$ is set in each individual frame.

In the training set, there are 66 downstairs, 70 upstairs, and 73 negative training samples. These samples were randomized when selected for the subset classification of the cross-validation phase, but each sample was used only once to prevent any biasing in the dataset. The testing set has 45 downstairs, 45 upstairs, and 47 negative samples, resulting in 137 samples total for the classification stage. As demonstrated in Table 1, our system achieves an average accuracy at 92.70% for staircase recognition with accuracies of 95.56% for upstairs, 88.89% for downstairs, and 93.62% for negatives respectively.

The algorithm actively processes RGB-D images to detect potential step candidates and notifies user throughout navigation. Once the SVM multi-class model classifies either an up or down staircase, the algorithm implements the additional features such as optical flow and KLT stair tracker. The algorithm uses an auditory feedback system to relay the stair distance in centimeters and estimated the amount of steps.



(a) Upstairs RGB Image
(b) Upstairs Depth Image
(c) Downstairs RGB Image
(d) Downstairs Depth Image
(e) Negative RGB Image
(f) Negative Depth Image
(g) Negative RGB Spiral Image
(h) Negative Depth Spiral Image

Fig. 5: Examples of staircase images of our database.

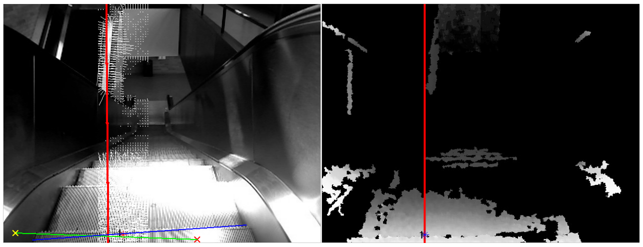
Without optimization, the current algorithm implemented in Matlab can process 2 frames per second, on a single core of an Intel Dual-Core 2.9 GHz processor without GPU acceleration. This computation can be speeded up when further optimized with C++. The proposed staircase detection and recognition algorithm doesn't rely on any dataset-specific tuning, therefore could be easily incorporated to improve the performance and user friendliness of current existing blind navigation systems.

Table 1: Experimental Results of Staircase Recognition

Class	Upstairs	Downstairs	Negative	Accuracy
Upstairs	43	0	2	95.56%
Downstairs	3	40	2	88.89%
Negative	3	0	44	93.62%

4.3. Limitations

The proposed pipeline effectively promotes uniformity between consecutive frames with histogram equalization, and enhances the performance of horizontal edge detection via edge filters fusion on RGB-D channels. However, it is insufficient to prevent certain potential limitations. Since the proposed staircase detection algorithm is based on both of the RGB and depth images captured by an RGB-D camera, the algorithm will fail if the environment is too dark (inadequate RGB data) or too bright (inadequate infrared data). For very dark environments, good quality edge maps cannot be extracted from RGB frames. For very bright environments, the distance cannot be correctly estimated due to the overexposed depth frames. Figure 6 illustrates one corrupted RGB-D image pair captured under very bright sunlight. The proposed algorithm handles the problem of extreme illumination by introducing parallel planes generation via inverse depth and accelerometer data [18], or depth image reconstruction [19]. The Hough transform also fails to recognize the spiral stairs due to its change in step orientation. Using the Horn-Schunck optical flow and calculating the average degree of the vector angles, the image can be rotated accordingly for better line detection.



(a) Corrupt RGB Image (b) Corrupt Depth Image

Fig. 6: Escalator with corrupted depth image due to excessive brightness.

5. CONCLUSIONS AND FUTURE WORK

In this paper, we have proposed a staircase detection algorithm to assist visually impaired people in unfamiliar environments. The proposed algorithm is evaluated with our collected staircase dataset and achieves recognition accuracies of 88.89% for upstairs, 95.56% for downstairs, and 93.62% for negatives. Further improvement of the algorithm will focus on expanding current dataset, user interface study, and evaluation by blind subjects.

6. ACKNOWLEDGEMENTS

This work was supported in part by NSF grants EFRI-1137172, IIP-1343402, and FHWA grant DTFH61-12-H-00002.

7. REFERENCES

[1] S. Oßwald, A. Hornung, and M. Bennewitz, “Improved proposals for highly accurate localization using range and vision data,” In *IROS*, 2012. 1

[2] D. Herńandez and K. Jo, “Stairway tracking based on automatic target selection using directional filters,” In *FCV*, 2011. 1

[3] X. Lu and R. Manduchi, “Detection and localization of curbs and stairways using stereo vision,” In *ICRA*, 2000. 1

[4] Google Project Tango Developers, “Project Tango Development Kit,” [Online]. https://store.google.com/product/project_tango_tablet_development_kit. Accessed: May 11, 2016. 1

[5] Y. Cong, X. Li, J. Liu, and Y. Tang, “A Stairway Detection Algorithm based on vision for UGV stair climbing,” In *ICNSC*, 2008. 1

[6] J. Hesch, G. Mariottini, and S. Roumeliotis, “Descending-stair detection, approach, and traversal with an autonomous tracked vehicle,” In *IROS*, 2010. 1

[7] S. Se and M. Brady, “Vision-based detection of staircases,” In *ACCV*, 2000. 1

[8] T. Lee, T. Leung, and G. Medioni, “Real-time staircase detection from a wearable stereo system,” In *ICPR*, 2012. 2

[9] D. Kim, K. Kim, and S. Lee, “Stereo camera based virtual cane system with identifiable distance tactile feedback for the blind,” *Sensors*, 2014. 2

[10] A. Pérez-Yus, G. López-Nicolás, and J. Guerrero, “Detection and modelling of staircases using a wearable depth sensor,” In *ECCV Workshop*, 2014. 2

[11] S. Wang and Y. Tian, “Detecting stairs and pedestrian crosswalks for the blind by RGB-D camera,” In *BIBM Workshop*, 2012. 2

[12] S. Wang, H. Pan, C. Zhang, and Y. Tian, “RGB-D image-based detection of stairs, pedestrian crosswalks and traffic signs,” *JVCIR*, 2014. 2

[13] W. Mayol-Cuevas, B. Tordoff, and D. Murray, “On the choice and placement of wearable vision sensors,” *IEEE TSMC*, 2009. 2

[14] C. Tomasi and T. Kanade, “Shape and motion from image streams under orthography: a factorization method,” *IJCV*, 2012. 3

[15] C. Kerl, “Odometry from RGB-D cameras for autonomous quadrocopters,” *Master’s thesis, TU Munich, Germany*, 2012. 4

[16] C. Chang and C. Lin, “Libsvm: A library for support vector machines,” *ACM TIST*, 2011. 4

[17] L. Gu and R. M. Stern, “Speaker segmentation and clustering for simultaneously-presented speech,” in *Interspeech*, 2009. 5

[18] W. Lui T. Tang and W.Li, “Plane-based detection of staircases using inverse depth,” In *Australasian Conf. on Robotics and Automation*, 2012. 6

[19] J. Lu and D. Forsyth, “Sparse depth super resolution,” In *CVPR*, 2015. 6