Environment interpretation for autonomous indoor navigation

Abhishek Kumar Tripathi
Image Understanding Group,
Uurmi Systems Pvt. Ltd.,
Hyderabad, India
Email: abhishekt@uurmi.com

Shanti Swarup
Image Understanding Group,
Uurmi Systems Pvt. Ltd.,
Hyderabad, India
Email: shanti@uurmi.com

***Abstract*—In this paper, indoor environment classification and interpretation algorithms are proposed. Proposed algorithms need low computation power and low payload thus enabling MAV to quickly react and navigate. Here indoor environment is classified in corridor, staircase and open space by using edge Gist descriptors and neural network classifier. Use of some predetermined threshold in each algorithm further increase the confidence of the classification. Detection of horizontal lines cluster and vanishing point is used for the navigation in staircase and corridor environments. Results demonstrate that proposed algorithms can interpret the indoor environment effectively.**

***Index terms*: Indoor environment, Gist descriptors, classification, MAV, vanishing point, edge linking.**

I. Introduction

The doctrine for intelligence gathering and counter measures are getting rapidly rewritten, driven by the available technology. Among all this, Micro Air Vehicles (MAVs) are rapidly carving a niche for themselves in the area of indoor reconnaissance and surveillance. The task could be locating an explosive inside a building, intelligence gathering by monitoring activities of suspects in a hideout or assessment of a hostage situation.

Navigating MAVs for outdoor surveillance involves defining a set of waypoints, flying in the direction of next waypoint while monitoring the state to correct for drift, if any. Adapting this model for indoor navigation is, at the minimum, extremely cumbersome; one must know the specific structure and dimensions, keep in mind commands achievable by the aircraft, and take care to avoid hitting walls or ground or frames while defining waypoints. In most cases, the indoor environment may simply not be known. The question that arises then is whether the MAV can be made intelligent enough to interpret the scene, understand the context, and navigate through it. There is a pressing need to improve situational awareness of MAVs using vision sensor, which the MAV will anyway carry for surveillance/reconnaissance purposes.

Building 3D structure from 2D image is challenging because reliable odometry is not always available, models have holes for textureless surfaces, and model creation-exploitation is computationally expensive causing unacceptable delays between perception and action cycles. These are just some of the significant challenges on robustness, computation requirements and even fundamental limitations of traditional methods when applied to real-life problems. This motivates the need to look for fundamentally new techniques to quickly understand the scene category and exploit it to navigate using simple rules. There are a number of approaches devoted to scene recognition for recognizing outdoor scenes [4], [8].

GIST features [7] are proposed specifically for scene recognition tasks. The GIST descriptor computes the output energy of a bank of 24 filters. The filters are Gabor-like filters tuned to 8 orientations at 4 different scales. The square output of each filter is then averaged on a 4×4 grid. Pavlopoulou et al [5] investigated the utility of human performance data on indoor-outdoor scene categorization in improving the generalization performance of a machine learnt indoor-outdoor classifier. In this experiments, authors compared two types of scene gist, image gist and edge gist. While scene gist is extracted from the original image, edge gist is extracted from the edge map of the image. The experiments revealed that edge gist characterizes indoor scenes far better than image gist. Indoor scenes consist of complex illumination phenomena such that natural lighting coming through windows, artificial lighting from various sources, multiple surfaces with different reflection properties. Further, they can be very cluttered. Edge gist features are more robust to such drastic changes in appearance than the image gist ones and result in improved classification rates.

L. Fei-Fei and P. Perona [1] proposed a novel approach to learn and recognize natural scene categories. In their approach, images of a scene are represented by a collection of local regions, denoted as codewords obtained from unsupervised learning. The results show satisfactory categorization performances on a large set of 13 categories of complex scenes. Lazebnik et. al [2] have proposed learning of quantization code-books by information loss minimization. Vogel et. al [3] present a semantic vocabulary for scene classification tasks wherein each image patch is labeled with a semantic label like sky, water, grass, etc. Garg et.al [6] explain the bag of words representation from a soft computing perspective. In their work they show that fuzzy and probabilistic codeword assignment improves the classification performance on a 15 scene dataset. Thus there are a number of works related to implementation of Scene classification which can be studied and modified for the implementation of classifying which environment the MAV is currently in and to follow different strategies for navigation

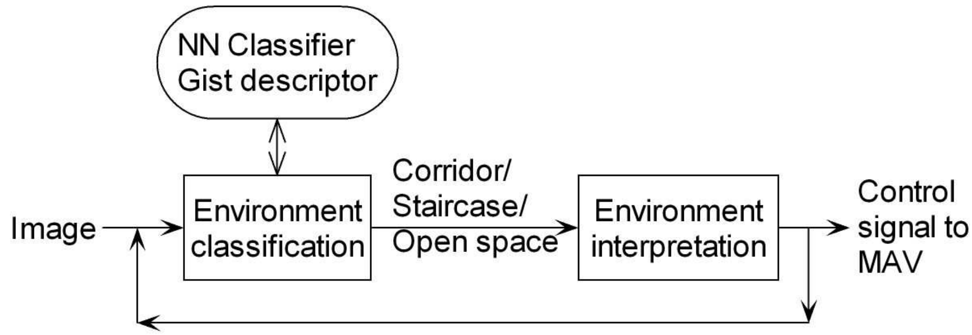


Fig. 1: Block diagram of proposed algorithm for scene classification and interpretation.

depending on the type of environment.

Our approach first involve coarsely classifying the sensed environment into different categories, such as stairs, corridors, areas with non-uniform height, small enclosed spaces, urban canyons, etc. These classified environments then processed to find pertinent perspective cues from the data that can in turn be utilized to localize the vehicle and navigate to its destination. We extracted a edge gist from the input imagery then used neural classifier to classify the image as one of several categories such as corridors, stairwell, indoor hall, room, etc.

Bills et al [4] used image gist descriptors as the image features for classification. Proposed algorithm uses image edge gist descriptors as the image features. It is noted in [5] that edge gist are more robust than the image gist. Unlike Bills et al, proposed algorithm has a refinement stage in the staircase detection algorithm. In this stage, perpendicular is drawn for each horizontal line. Number of the intersections with other horizontal lines are used to find the cluster of the lines which depict the staircase. Bills et al used Hough transform for the detection of straight lines in corridor vanishing point and staircase detection algorithm. However, in proposed algorithm first canny edge operator are used and to increase the accuracy edge linking for contours detection are used. These contours are then split into the line segments.

II. Proposed Approach

Proposed algorithm is composed of two stages, Environment classification and interpretation of environment. Block diagram of the proposed algorithm are shown in Fig.1

A. Environment classification

Indoor environments can be interpreted in terms of the arrangements of the straight lines. We first extracted the Gist descriptor of the edge map of the image by resizing the input image to 256×256. Then we have used Neural classifier to classify these images into three categories; corridor, staircase and open space. Gist descriptors measure the global distribution of line segments orientation in an image which is suitable for indoor environments since it can leverage the long lines found in indoor images.

B. Environment interpretation

One we knew the type of the environment next task is to interpretation of the environment.

1) Staircase: To navigate MAV in staircase environment, our goal is find the center of staircase. Staircase can interpreted as the collection of the horizontal straight lines. Canny edge operator with hysteresis threshold is used to generate the edge map. Straight lines segments are extracted for the edge map by linking edges into contours and then splitting the contours into straight lines segments. Then horizontal lines are extracted from these straight lines segments by examining the slope of the lines segments.

These horizontal lines are the potential candidates for the staircase with some false alarm. False alarm are reduced by finding the largest cluster of the horizontal straight lines. A line perpendicular to a horizontal line was drawn which would cut multiple horizontal lines. This was repeated for every horizontal line. A cluster of horizontal lines was formed using the perpendicular cutting maximum horizontal lines. All the horizontal straight lines of this cluster verify the location of the staircase.

Number of horizontal lines those are the candidates for the staircase justify the classify environment. It is noted that for staircase environment number of horizontal lines should be greater than a threshold. This threshold act as the confidence for the environment classification algorithm and increase the accuracy of the environment interpretation algorithm. Mean of the endpoints of the horizontal lines gives the center of the staircase.

Difference between the location of the center of the staircase and center of the image is used to generate the control signal for the MAV. This control signal keep the MAV in middle of the staircase and help to navigate.

2) Corridor: To navigate MAV in corridor environment is it required to find the vanishing point. Vanishing point locates the end of the corridor. In corridor environment long straight lines are converge to the vanishing points. Similar to the staircase, straight lines are extracted by canny edge operator and edge linking method as discussed earlier.

For vanishing point detection, we found the intersection between each pair of straight lines. To reduce the computation, We have removed near horizontal and vertical lines for this calculation because near horizontal and vertical lines can not converge at vanishing point. We have subdivided the image into predetermined number of rectangular grid. The grid having maximum number of intersections is considered to valid candidate for having vanishing point. This grid is referred as vanishing point grid. Weighted average of all intersection points present in vanishing point grid represent the appropriate location of the vanishing point. Here weights depends upon the distance of the point from the center of the grid.

Number of lines intersecting in the vanishing point grid justify the classify environment. It is noted that for corridor environment number of lines intersecting on vanishing point should be greater than a threshold. This threshold act as confidence for our algorithm.

Location of the vanishing point is used for the generation of the control signal for the MAV. Control signal should be such that vanishing point always remain in middle of the image when MAV is moving towards it.

III. Simulation & Results

Simulation is carried out in MATLAB 7.12.0 (R2011a) environment on a system with a 3.29 GHz Intel(R) Core(TM) CPU and 3.40 GB of RAM, running on MS XP.

A. classifier

For training, testing and validation we have used 275 images (139 corridor, 80 staircase and 56 open space images) downloaded through the web. Confusion matrix and receiver operating characteristic (ROC) of the neural network classifier id shown in Fig.2. Results confirm the $> 90\%$ accuracy of the classifier. Accuracy of the classifier is increased by the confidence threshold of the environment interpretation algorithm. For experiments, we have also captured few images from the forward facing camera of our MAV (see Fig.3).

B. staircase

Result of the staircase detection is shown in Fig.3. Results depicts the efficiency of each of the steps of the algorithm. Fig.3a shows the original staircase image. Fig.3b shows the all lines segments (shown in red color). Fig.3c shows the detection of the horizontal lines according to their slope. All horizontal lines are shown in green color. Fig.3d shows the separation of the staircase line segments from the horizontal lines segment. This step uses a clustering algorithm as discussed in section II-B1. All staircase line segments are shown in blue color. Here green and blue vertical lines show the location of the staircase and middle of the image respectively. Some more results are shown in Fig.4.

C. Corridor

Results of the vanishing point detection algorithm are shown in Fig.5. Here, intersection between the line segments are shown in green cross, vanishing point grid is shown in blue and vanishing point location is marked in blue circle. Results verify that the proposed algorithm is able to detect vanishing point effectively.

IV. Conclusion

This paper is aimed at identifying key structural elements from the scene viz. corridors, staircases and open space, defining intuitive navigation rules for these elements, and turning these rules into commands executable by the autopilot. The proposed methods are a combination of innovative image processing and vision algorithms, heuristics and nature-inspired approaches for navigation through each structure. These algorithms require low complexity and enabling the MAV to quickly navigate in indoor environments. Proposed algorithms require only a light weight camera which is a good solution for the payload restriction of the MAV.

Our research effort will make novel contributions in the areas of scene class identification, and will provide an efficient system that can run in real time computing platform.

References

[1] L. Fei-Fei and P. Perona. A Bayesian hierarchical model for learning natural scene categories. In cvpr, pages 524-531, 2005.

[2] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching text for recognizing natural scene categories. In cvpr, pages 2169-2178, 2006.

[3] J. Vogel and B. Schiele. Semantic model of natural scenes for content-based image retrieval. Intl. J. Computer Vision, 72:133-157, 2007.

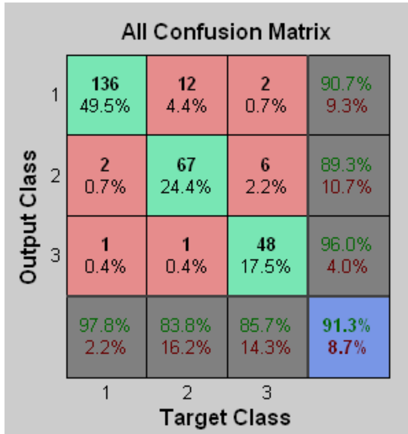
[4] Cooper Bills, Joyce Chen, Ashutosh Saxena,Autonomous MAV Flight in Indoor Environments using Single Image Perspective Cues, In International Conference on Robotics and Automation (ICRA), 2011.

[5] C. Pavlopoulou, S. X. Yu, "Indoor-Outdoor Classification with Human Accuracies: Image or Edge Gist?" , IEEE CVPR Workshop on Advancing Computer Vision with Humans in the Loop, San Francisco, California, June 2010.

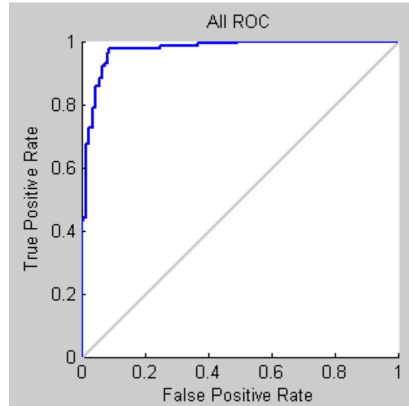
[6] Vinay Garg, Sreekanth Vempati and C.V. Jawahar Bag of visual words: A soft clustering based exposition Proceedings of 3rd National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics, ISBN 978-0-7695-4599-8, pp.37-40 15-17 Dec. 2011, Hubli, India.

[7] Aude Oliva, Antonio Torralba, "Modeling the shape of the scene: a holistic representation of the spatial envelope",International Journal of Computer Vision, Vol. 42(3): 145-175, 2001.

[8] A. Quattoni, and A.Torralba Recognizing Indoor Scenes IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2009.



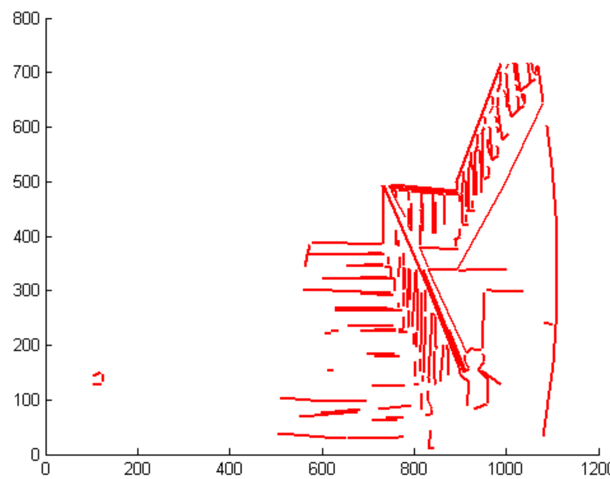
(a)



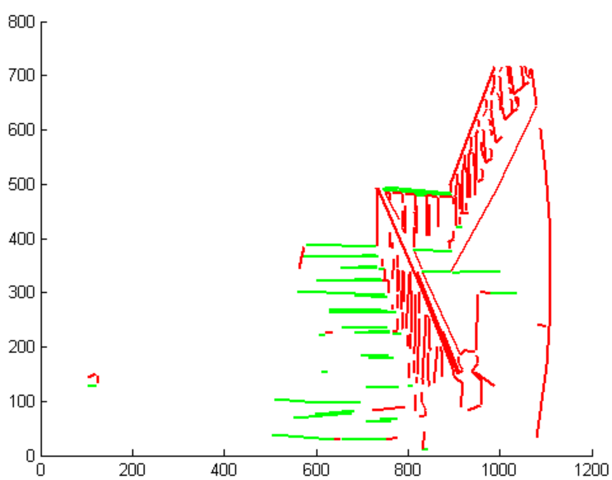
(b)

Fig. 2: (a) Confusion matrix and, (b) Receiver operating characteristic (ROC) of the classifier.

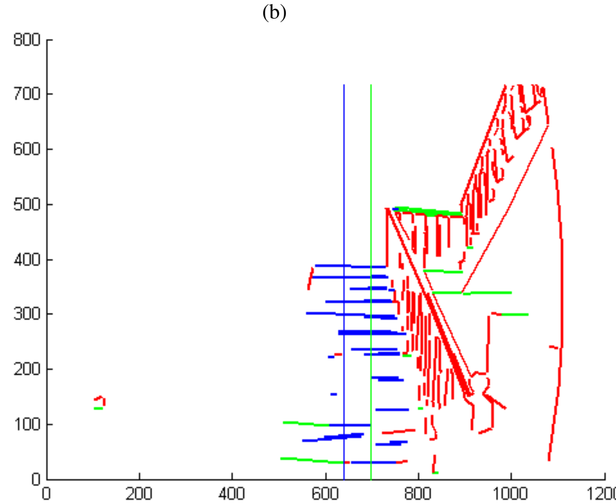
(a)



(b)



(c)



(d)

Fig. 3: (a) Original staircase image, (b) all line segments detection, (c) horizontal line segments detection, (d) staircase line segments. Here green and blue vertical lines show the location of the staircase and middle of the image respectively.

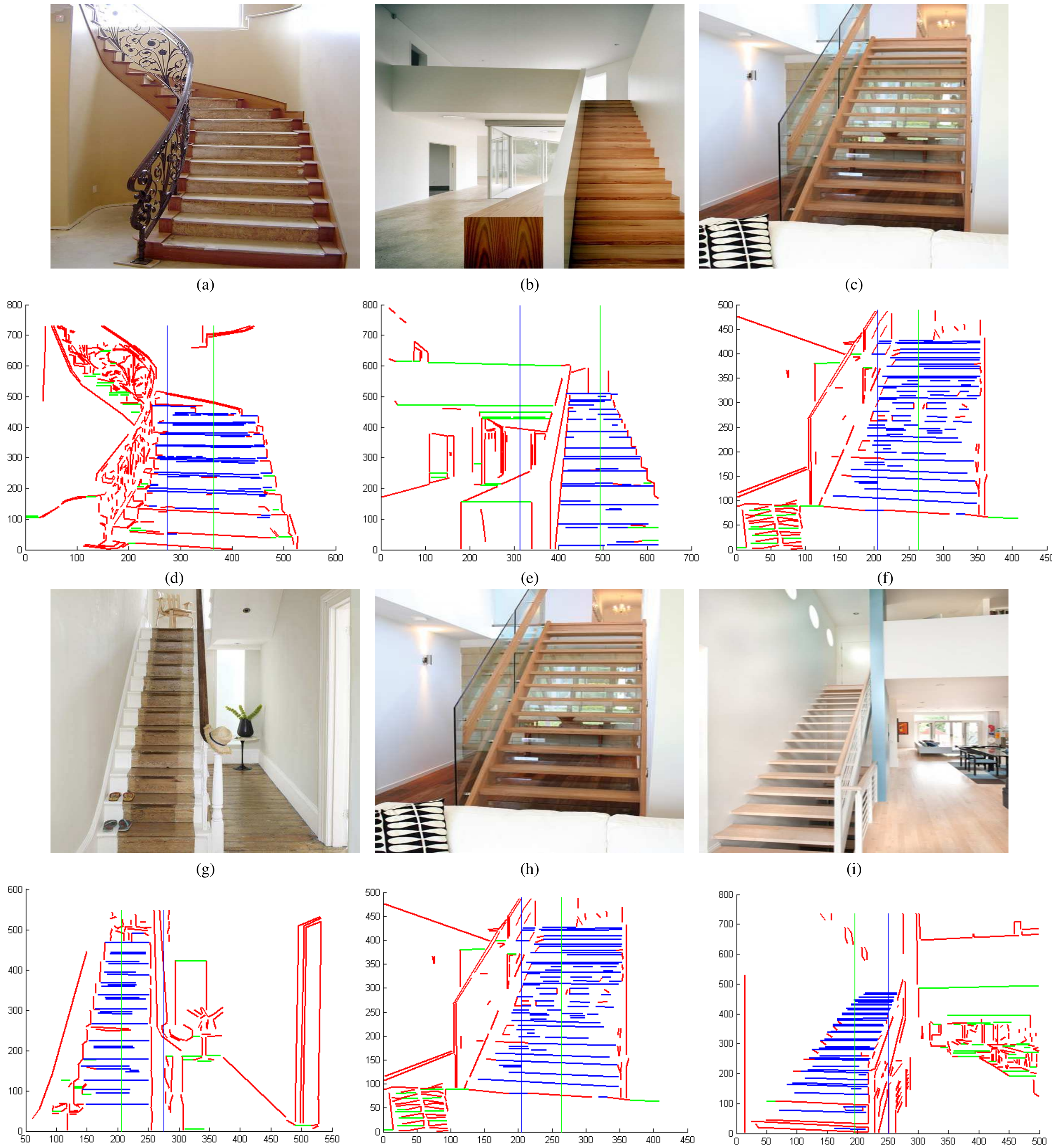
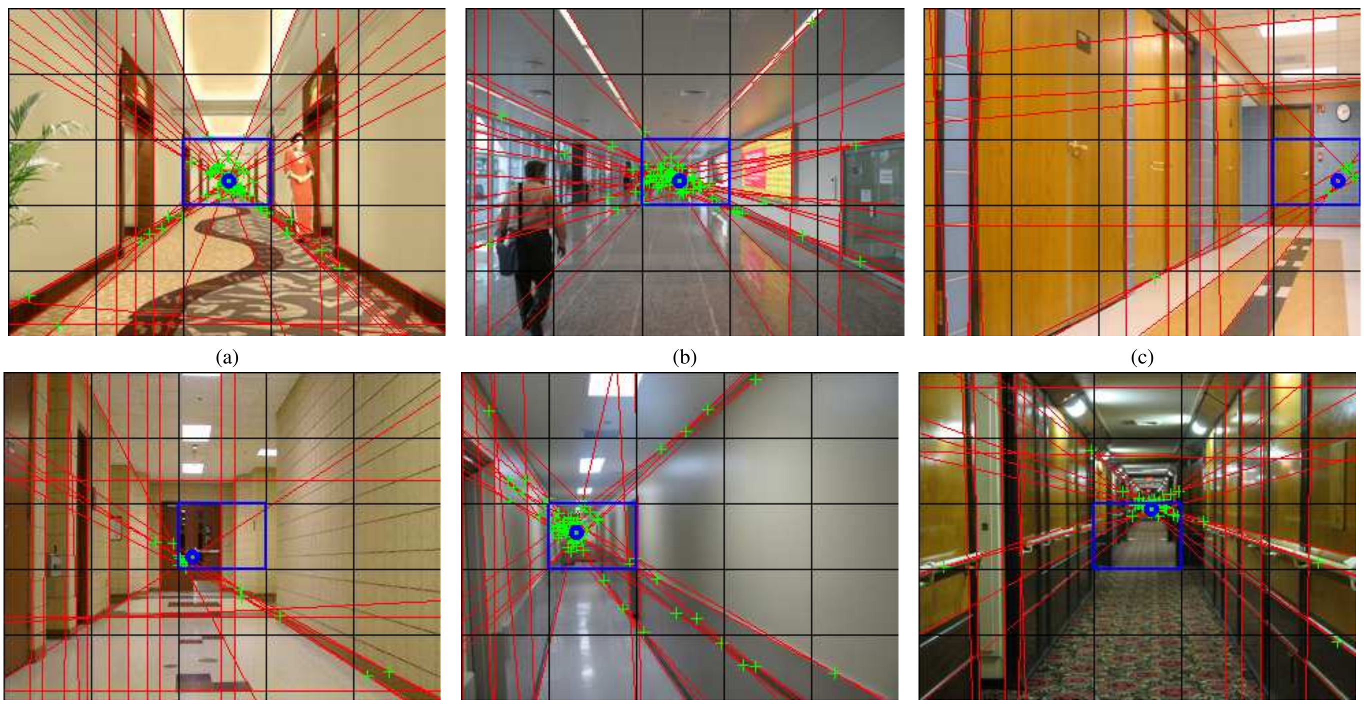


Fig. 4: First and third row show the staircase image and second and fourth row show the corresponding staircase detection.



(a) (b) (c)

(d) (e) (f)

Fig. 5: Some examples of the vanishing point detection algorithm.