A Bayes Filter based Adaptive Floor Segmentation with Homography and Appearance Cues

Suryansh Kumar*
Robotics Research Center
IIIT-Hyderabad
k.sur46@gmail.com

Ayush Dewan
Robotics Research Center
IIIT-Hyderabad
pisces.aayush@gmail.com

K.Madhava Krishna
Robotics Research Center
IIIT-Hyderabad
mkrishna@iiit.ac.in

ABSTRACT

This paper proposes a robust approach for image based floor detection and segmentation from sequence of images or video. In contrast to many previous approaches, which uses a priori knowledge of the surroundings, our method uses combination of modified sparse optical flow and planar homography for ground plane detection which is then combined with graph based segmentation for extraction of floor from images. We also propose a probabilistic framework which makes our method adaptive to the changes in the surroundings. We tested our algorithm on several common indoor environment scenarios and were able to extract floor even under challenging circumstances. We obtained extremely satisfactory results in various practical scenarios such as where the floor and non floor areas are of same color, in presence of textured flooring, and where illumination changes are steep.

Keywords

Segmentation,Optical Flow,Homography.

1. INTRODUCTION

There has been a fair number of methods in literature that have attacked the problem of indoor floor segmentation. Some have taken a purely appearance based formulation through color and texture cues [12], while others have approached it primarily from a geometric/homographic standpoint [16]. Some others have more recently proposed methods that combine geometry and appearance cues [5]. This paper proposes a framework based on recursive segmentation and Bayesian propagation of class probabilities that results in robust floor extraction even in difficult conditions such as when floor and non-floor regions have similar or same color and texture, when the floor texture varies widely in a single view, where illumination varies significantly and when shadows are present. The keynote of the algorithm is its adaptability inherited through this framework that manifests in two beneficial ways:

1) A floor segmentation that improves with time especially in relation to the boundaries even when floor and non-floor regions possess same color and texture.
2) A floor segmentation that retains accuracy when the floor and surrounding structure changes such as when the robot turns around a corner or when the floor area ahead of the robot changes appreciably.

The algorithm by an apt combination of graph based segmentation due to [6] and homography cues. Using these cues an image is classified into equivocal and unequivocal segments. The equivocal segments or clusters are those which contains both kinds of features, those that satisfy the floor homography and those that do not. They also contain features whose satisfaction or non satisfaction of homography is ambiguous. Unequivocal or unambiguous clusters with coherent homography are to the extent segmented from these equivocal clusters by repeatedly (usually only once more) subjecting them to graph segmentation of [6]. The remaining ambiguous clusters are assigned probability values that are propagated temporally eventually leading to their decisive classification as belonging or not belonging to the floor. Indeed probabilities are computed through out the image and propagated but they find the maximum utility for clusters that are uncertain about their belongingness. We show comparative results portraying performance gain of the current method vis-a-vis recently proposed methods [11] in terms of segmentation accuracy. Such floor segmentation algorithms find immense use in vision based exploration, mapping and homography based SLAM systems.

2. LITERATURE SURVEY

As mentioned in the introductory section most previous approaches can be classified as belonging to appearance based or geometry based or those that combine both. In the case of appearance based approaches multiple visual clues from the environment are used for detection. Lorigo et al. [12] used combination of color and gradient histograms to distinguish free space from obstacles. Due to over reliance on color based descriptors their approach failed in an homogeneous environment. Unlike some previous work [12] which concentrated on determining empty navigable space in front of robot, Li and Birchfield [11] used combination of vertical edges, thresholding and segmentation to approximate a wall floor boundary and then classify horizontal edges which

*Corresponding author

ICVGIP '12, December 16-19, 2012, Mumbai, India

lie on that boundary. Their approach produced some good results and robustly dealt with specular reflection on floor which is common in indoor environment but failed when vertical edges were missing in the lower half of image, when there were no structural edges on side walls close to the robot. It also seemed less than efficient when the floor was textured. In geometry based approaches the ground plane constraint has been exploited, motion of pixels is put under microscope to see whether they match the values that would be expected if the points that lie on the floor. Jin and Li proposed a method for ground plane detection [16]. They used a monocular camera to calculate dominant homography between two images by classifying sparse feature points. Kim and et al. [3] used dense optical flow correspondence for calculating planes perpendicular to ground plane for obstacle detection and then calculating a path for visual navigation of robot. While Fazl and Tsotos [5] also used dense point correspondences but they combined it with stereo vision for calculating floor anomalies.

While purely appearance based approaches fail under homogeneity of appearance of floor and non-floor regions, geometric methods on the other hand are efficient at detecting features that constitute the floor. However geometry based approaches need extra cues to extract or segment the optimum boundary that encompasses the features belonging to the floor. In methods that encompasses both geometry and appearance, Lee et al. [10] used combination of ceiling and floor intensity edges with geometric constraints for interpretation of an indoor scene from single image and more recently Rituerto et al. [14] proposed a method of semantic label propagation in an indoor scene. They manually labelled the segments in first frame which is then propagated and updated through the sequence. Whereas in [3] ground plane estimation based on homography combined with region growing is used for segmentation.

The current approach differs in the way of using recursive segmentation to disambiguate areas that appear homogeneous both from geometry and appearance based perspectives, typically the region around the floor boundaries when the floor and non floor regions are of same color and texture. It also uses a probabilistic framework to temporally propagate the probabilities that leads to a segmentation that improves with time as well as adapt to changes.

3. SYSTEM OVERVIEW

Figure1 shows an overview of our approach. Using KLT feature detector and optical flow, feature correspondence between two frames I_n and I_{n-1} is calculated. These point correspondence are used for estimating homography error e_r. On the basis of homography error value e_r, detected features are classified as floor P_f, non-floor P_{nf}, ambiguous P_a using Bayes filter. The ambiguous segments are subject to further segmentation that tries to remove ambiguity, which eventually is removed temporally through the filter. These classified features are then used to select ground segments S_f.

4. GROUND PLANE DETECTION

In our approach we decided to exploit the fact that points lying on same plane will have a coherent motion pattern which will be different from other pixels in the image. We used optical flow for determining point correspondence between two successive frames and planar homography for estimation of ground plane.

4.1 Feature Detection and Matching:

For the calculation of reliable homography it is necessary to have a set of features in two different camera views. These features can either be dense i.e pixel to pixel correspondence as done in [5] or sparse feature detection as used in [16]. Calculation of dense point correspondence is computationally expensive while sparse method do not give enough feature correspondences using optical flow. So instead of calculating sparse features on whole image, we divided the image into 10x10 grid and then track features in each of the small boxes which gave us point correspondence dense enough to get good optical flow correspondences but still not as computationally expensive as in case of dense optical flow correspondences. Figure 2 shows all the three point correspondence discussed above.

We have used Kanade-Lucas-Tomasi feature tracker for feature detection which is based on initial method proposed by Lucas and Kanade [13] which was then further improved by Lucas and Tomasi [15].

4.2 Homography

Theoretically points which lie on same plane share a homography transformation. If points x and x' are coplanar then the transformation can be represented by the equation:

$$x' = Hx \tag{1}$$

Here x and x' represent homogeneous coordinate $(x, y, w)^T$ of a pixel in two different view while both represent the same feature. H is a 3x3 matrix which provides the transformation of pixel from one view to another. It is only defined up to a scale and therefore has 8 degrees of freedom. Since each correspondence gives 2 constraints, a minimum of 4 point correspondence is required for calculation although more than 4 correspondence helps in improving the accuracy of homography matrix.

For the initial computation of homography we assume that area in front of robot is floor and use only point correspondences that belongs to a specified area in the lower half of image. These point correspondences are given as input to RANSAC [7] algorithm for homography calculation. This homography matrix is used to determine the position of feature correspondences in second frame, detected in the first frame. Once homography of ground plane is established correspondences can be established by Homography for features on the ground plane. We compute the error between predicted correspondences obtained based on homography with those actually obtained by optical flow. This error is used in the computation of posterior beliefs of a feature belonging to the ground plane, which is explained later in section 5.

We used a Bayesian framework, for assigning a probability value to each pixel. This value represents its belongingness to floor and is initially set to 0.5. Since in every iteration only sparse set of features or pixels are available, only their probability value is updated on the basis of homography error value. This framework is discussed in 6. These pixels are then classified as floor P_f, non-floor P_{nf} or ambiguous P_a on the basis of recently updated probability value. The ambiguous points generally lie near floor obstacle boundary and they were mentioned in [16] during discussion on “virtual plane” problem. The floor(green), non-floor (red) and

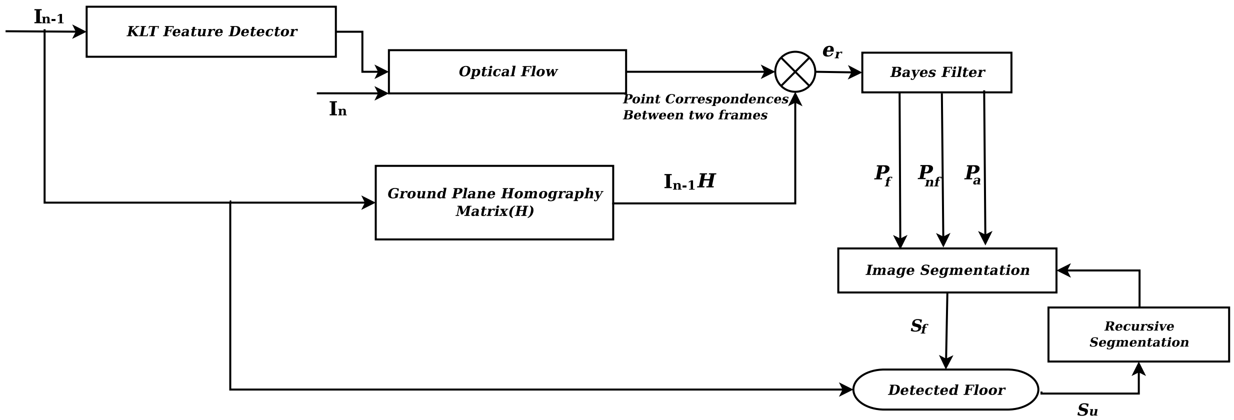

Figure 1: Flowchart for proposed method of floor segmentation

Figure 2: (a):Sparse feature correspondence.(b):Dense feature correspondence.(c):Proposed modified feature correspondence

ambiguous points (blue) are shown in Figure 3.

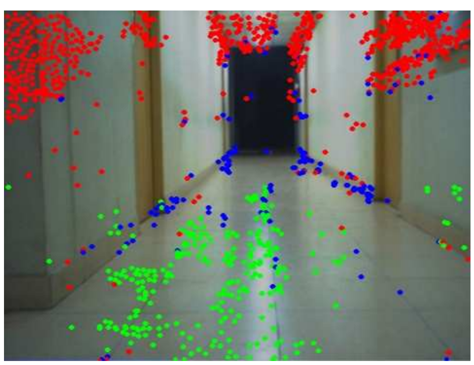

Figure 3: Green points represent floor points,red points represent non-floor points and blue points represent ambiguous points

5. FLOOR EXTRACTION

After ground plane detection and probability update we only have discrete set of pixels whose belongingness to floor is known. To estimate the boundary between floor and non floor area as well as to estimate the complete floor area that encompasses those discrete pixels the following is done. The image is divided into segments and an entire segment is classified as floor if majority of features in that segment are floor features/pixels. This is further explained below.

5.1 Segmentation

The segmentation algorithm proposed in [6]. They used graph based representation of images for estimating boundary between neighbouring pixels. The image is divided into small segments on the basis of color{R,G,B}. These segments are further classified into three classes called floor, non-floor and as ambiguous segments. All those segments which only have floor pixels are classified as floor segments and those which only have non-floor pixel are classified as non-floor and those segments which only have ambiguous points is classified ambiguous segment. Those segments which have points from multiple classes are left unclassified. Unclassified segments are those that have features which belong to both classes in tangible numbers. Ambiguous segments are those which dominantly consist of features whose posterior beliefs are equivocal/ambiguous about their belongingness to floor or non floor areas. Using recursive segmentation an unclassified segment is divided into small segments using the same segmentation algorithm but with a stringent threshold in comparison to threshold used for initial segmentation. The small segments are then classified into floor, non-floor or ambiguous. Figure4 illustrates the process of segmentation, classification of segments into multiple classes and recursive segmentation.

Classification of pixels on the basis of probability is discussed in section 6. These probability values are propagated in subsequent frames which leads to disambiguation of ambiguous pixels into floor and non-floor classes. This classification causes division of an ambiguous segment into floor and non-floor segment. This also decreases the probability of a segments being classified as ambiguous. Figure 5 illustrates probability propagation of ambiguous pixels and how an ambiguous segment is classified as floor in subsequent frames.

6. PROBABILISTIC FRAMEWORK

We find the homography H_n, between the current and previous view, I_n and I_{n-1} by RANSAC. The points in the previous image, I_{n-1}, whose probability of belonging to the ground plane is very high are considered as candidate points for RANSAC. Their correspondences in I_n, are obtained through the modified optical flow algorithm discussed in section 4. For the very first image I_0, the points which lie within a small trapezium erected from the base of the image are considered belonging to the ground, this provides for bootstrapping. Let $x_{\{i,1\}}, x_{\{i,2\}}.....x_{\{i,n-1\}}, x_{\{i,n\}}$, be the image coordinates of the same 3D point in images $I_1, I_2....., I_{\{i,n-1\}}, I_{\{n\}}$. We denote the prior belief as

$$\hat{Bel}(x_{i,n}) = P(x_{i,n}/u_n, H_{n-1}, u_{n-1}, H_{n-2}, ...H_1) \qquad (2)$$

The above belief is the probability of pixel $x_{i,n}$ belonging to ground based on all homography till the previous pair of images I_{n-1} and I_{n-2} and all motion model including u_n, which is obtained from optical flow detection I_n and I_{n-1}.

Now, from law of total probability

$$\hat{Bel}(x_{i,n}) = \sum_i P(x_{i,n}/u_n, x_{i,n-1}......H_1) \\ .P(x_{i,n-1}/H_{n-1}, u_{n-1}, ...H_1) \qquad (3)$$

And upon invoking standard Markov assumption give

$$\hat{Bel}(x_{i,n}) = \sum_i P(x_{i,n}/u_n, x_{i,n-1}) \\ .P(x_{i,n-1}/H_{n-1}, u_{n-1}, ...H_1) \qquad (4)$$

The posterior given by

$$Bel(x_{i,n}) = P(x_{i,n}/u_n, H_n......H_1) \qquad (5)$$

includes the current homography computation in belief update. This can be reduced using usual Bayesian procedures to

$$Bel(x_{i,n}) = P(x_{i,n}/H_n)\hat{Bel}(x_{i,n-1}) \qquad (6)$$

From (4) one also identify the second term on the right hand side of (4)as the posterior belief at $n-1$. Thus (4) recurses to

$$\hat{Bel}(x_{i,n}) = \sum_i P(x_{i,n}/u_n, x_{i,n-1})Bel(x_{i,n-1}) \qquad (7)$$

This homography based sensor model $P(x_{i,n}, H_n)$ is computed as $P(x_{i,n}, H_n) = C/||x_{i,n} - H_n x_{i,n-1}||$ as the norm of standard innovation term between the expected location of $x_{i,n-1}$ in I_n given by $H_n x_{i,n-1}$ and the actual sensed model from correspondence, $x_{i,n}$. The motion model is simple computed as a binary model i.e

$$P(x_{i,n}/u_n, x_{i,n-1}) = \begin{cases} 1 \; if \;\; x_{i,n-1} \; corresponds \;\; to \;\; x_{i,n} \\ \qquad\qquad from \;\; optical \;\; flow \\ 0 \; if \;\; x_{i,n-1} \;\; do \; not \; corresponds \\ \qquad\qquad to \;\; x_{i,n} \;\; from \; optical \; flow \end{cases} \qquad (8)$$

A similar set of analogous belief states can be computed for x_i, not belonging to the ground plane and the beliefs are finally normalized such that their sum is unity. This propagation of beliefs across images results in tracked features getting their association or non-association to the ground with progressively increasing confidence lending itself to sharper boundaries of the floor in the image. Essentially the Bayesian propagation is one such formal way of taking into account the temporal statistics of image features, which in essence gives a higher probability of the image feature belonging to the floor if it is.

7. RESULT

This section will summarize the performance of our algorithm. The images are captured by Logitech Quick Pro web-cam mounted on P3AT mobile robot platform. The whole algorithm is implemented in C++ using OpenCV vision library on a 2.4Ghz Core 2 Duo processor. To check the performance and robustness we tested our algorithm on three different indoor environment scenarios. Figure 7(a) is a sample image from a homogeneous environment. In this image color of floor and wall is same and base of the wall is covered with floor tiles. Figure 7(e) is a sample image of a room environment and Figure 7(i) is sample image from an environment in which floor tiles are of multiple colors.

The performance of algorithm is analyzed by calculating precision P_r and detection rate d_r values. Of all the pixels classified as floor, what fraction of it is actually floor is signified by precision values and of all the pixels which represent floor, what fraction of it is detected is signified by detection rate.

$$P_r = \frac{TruePositive}{TruePositive + FalsePositive} \qquad (9)$$

$$d_r = \frac{TruePositive}{TruePositive + FalseNegative} \qquad (10)$$

For quantitative analysis we used sequence of 50 images from each of the three datasets. Detection rate for initial frames is low but as probability propagates, detection rate starts to increase. Once probability value stabilizes, detection rate value saturates. The change in detection rate percentage for a sequence of frames can be seen in graph shown in Figure 6. One can see that the detection rate saturates at very high percentages within very few frames.

Figure 7(a)-(d) shows output results for homogeneous environment, and based on our survey we have not found such results portrayed when floor and non floor areas have the same color. Figure7(a) is output result for first frame. Increase in detection rate can be seen in next two images and last image on right(fig7(d)) shows output result after detection rate has saturated. Figure 7(e)-(h) shows similar output results for a room environment. Figure 7(i)-(l) shows results where texture and color varies across the floor.

We have made an attempt to consider multiple environment scenarios and test our algorithm on them. Our algorithm successfully dealt with homogeneous environment on which color based segmentation algorithm generally fails and feature tracking also becomes difficult. We successfully tested our algorithm on textured floor, a drawback of [11]. We have also presented results for a room environment. Figures of 7 shows that how the probability of a pixel to be floor changes over the iterations. Performance of our algorithm is summarized in table 1. For different environments considered the detection rate for initial frames (d_{ri}) is around

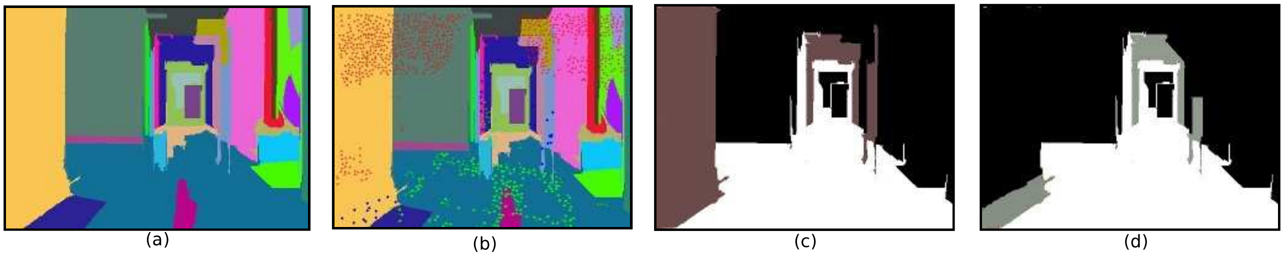

Figure 4: (a):Segmentation of the image using [6].(b):Floor(Green),non-floor(Red) and ambiguous points(Blue) selects their corresponding segments.(c):Classification of segments into floor(White),non-floor(Black),while other segments remains unclassified(Brown).(d) Recursive segmentation divides unclassified segment into non-floor(black) and ambiguous (gray).

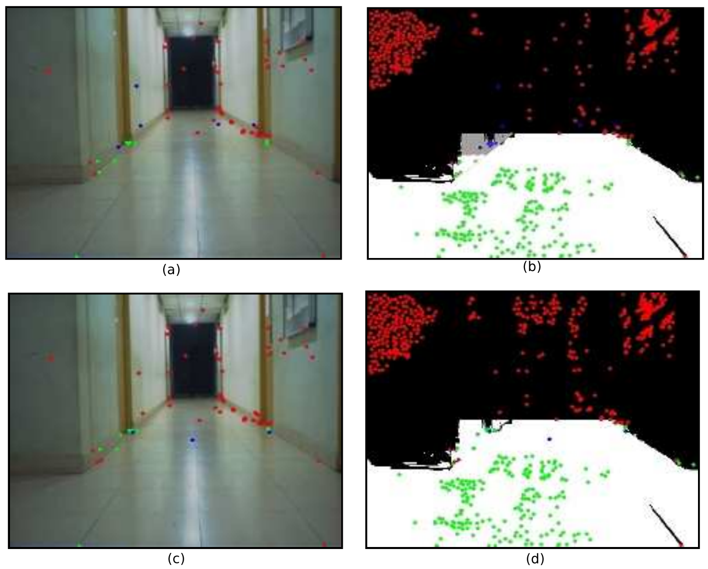

Figure 5: (a)&(c)Ambiguous points(blue) is classified into floor(green) and non-floor(red) points as probability propagates in sequence of images.(b)&(d)An ambiguous segment merges into floor in subsequent frames.

Table 1: QUANTITATIVE ANALYSIS

	$d_{ri}\%$	$d_{rs}\%$	$P_r\%$
Room	63.64	96.65	99.07
Homogeneous	65.50	95.05	98.92
Textured Floor	75.32	95.04	98.72

65% and once probability saturates (d_{rs}) this value climbs to 95%. The algorithm also has a high precision rate (P_r) of 98%. The detection rate and precision values quantitatively and qualitatively as well by visual inspection of the Figures of 7 and Figures of 9. we conclude the robustness of algorithm to tackle variety of indoor situations. The algorithm processes two frames of resolution 480x360 in 3 seconds to produce the final result.

8. APPLICATION

The above mentioned algorithm has been extended into an elementary exploration technique. Using the floor information and vertical edges, possible navigable paths are estimated. A grid is mapped on lower half the image for tracking the spread of floor. The vertical line segments are determined in sequence of images using Canny edge detector [1], Hough transformation [4] and some morphological image processing techniques. When floor bends around a

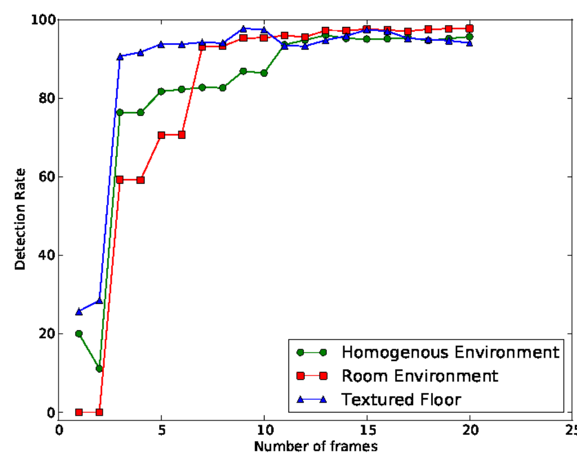

Figure 6: Detection Rate for different datasets

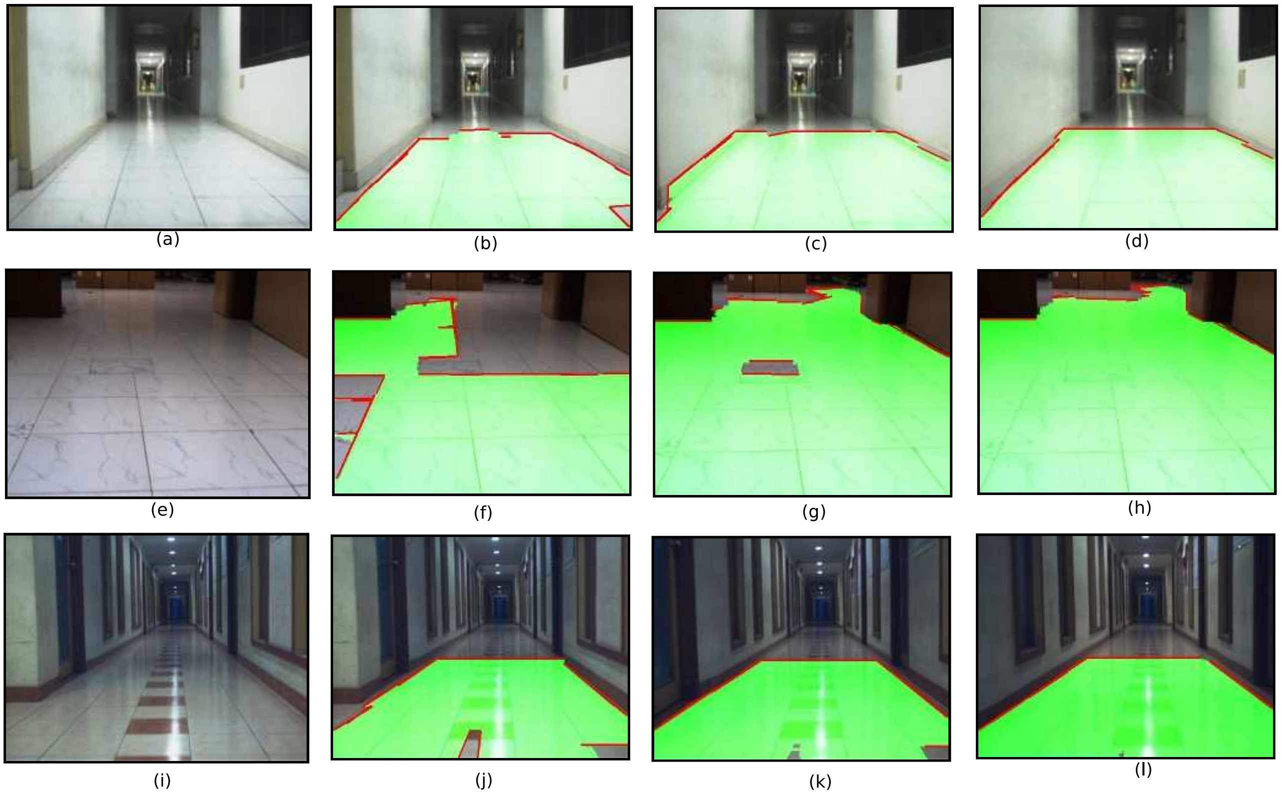

Figure 7: Detection of floor improves as probability propagates and pixels gain confidence (a)-(d)Homogeneous wall and floor color.(e)-(h) Room.(i)-(l)Textured Floor.

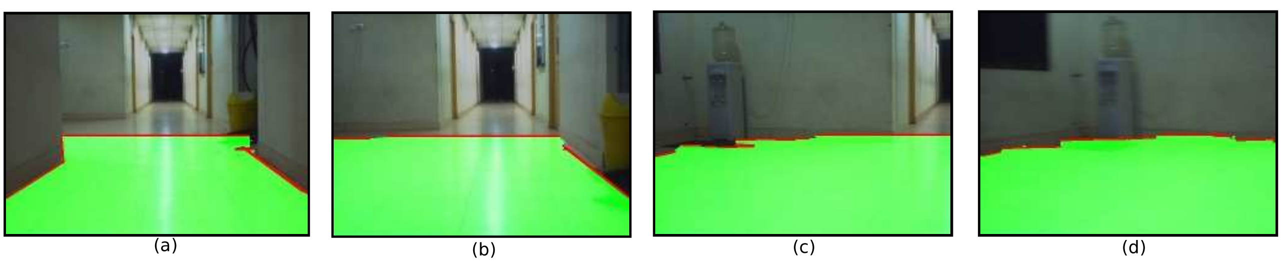

Figure 8: (a)-(d)Detection results when a robot is turning

vertical edge the algorithm detects a potential path to explore along the direction. Figure 11 shows the result on two pair of images. In each pair, image on left shows shows detected floor, vertical line segments and grid map. Direction of a navigable path is shown using arrows in right image.

9. CONCLUSION

This paper presented a robust floor segmentation algorithm that used both appearance and geometric cues dovetailed into a Recursive Bayes Filter formalism. The filter enables to maintain accurate segmentation even as the floor appearance changes and also obtains a precise boundary between floor and no floor areas. The cornerstone of this effort is the efficacy of the segmentation even in areas where floor texture changes, where floor and non floor areas are of same color, where the robot rotates into a new view, in presence of varying illumination and over extremely long sequences. High precision and detection rate on diverse datasets provides evidence of the efficiency and robustness of our approach and based on these quantitative evaluations we find that the current method shows performance gain over very recent approaches reported in literature. The above method can be optimized for an autonomous exploration and hence one can map the ground plane by projecting the ground plane coordinate to the world coordinate frame up to scale, and therefore it paves a path for VSLAM.

10. ACKNOWLEDGMENT

We would like to thank the Dept. of Information Technology to have funded this work through the grants made available by the National Program on Perception Engineering - Phase 2.

11. REFERENCES

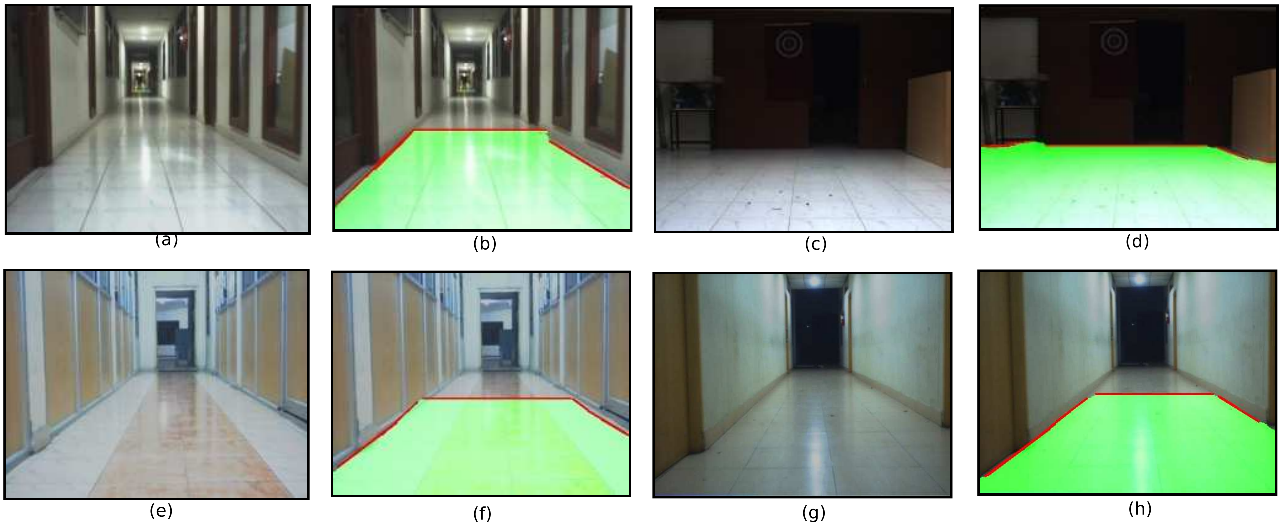

Figure 9: Results on different environments

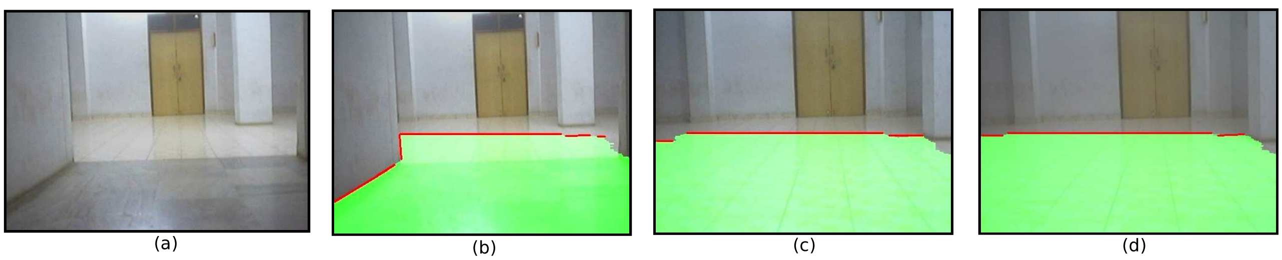

Figure 10: Output result when texture of floor changes

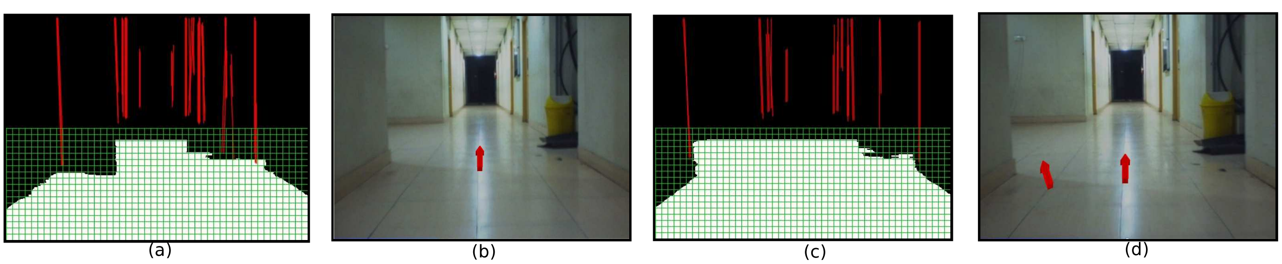

Figure 11: Images in left column shows detected floor,vertical line segments and grid map

[1] J. Canny. A computational approach to edge detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (6):679–698, 1986.

[2] D. Conrad and G. DeSouza. Homography-based ground plane detection for mobile robot navigation using a modified em algorithm. In *Robotics and Automation (ICRA), 2010 IEEE International Conference on*, pages 910–915. IEEE, 2010.

[3] X. Cui, Y. Kim, and H. Kim. Floor segmentation by computing plane normals from image motion fields for visual navigation. *International Journal of Control, Automation and Systems*, 7(5):788–798, 2009.

[4] R. Duda and P. Hart. Use of the hough transformation to detect lines and curves in pictures. *Communications of the ACM*, 15(1):11–15, 1972.

[5] E. Fazl-Ersi and J. Tsotsos. Region classification for robust floor detection in indoor environments. *Image Analysis and Recognition*, pages 717–726, 2009.

[6] P. Felzenszwalb and D. Huttenlocher. Efficient graph-based image segmentation. *International Journal of Computer Vision*, 59(2):167–181, 2004.

[7] M. Fischler and R. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.

[8] R. Hartley, A. Zisserman, and I. ebrary. *Multiple view geometry in computer vision*, volume 2. Cambridge Univ Press, 2003.

[9] Y. Kim and H. Kim. Layered ground floor detection for vision-based mobile robot navigation. In *Robotics and Automation, 2004. Proceedings. ICRA'04. 2004*

IEEE International Conference on, volume 1, pages 13–18. IEEE, 2004.

[10] D. Lee, M. Hebert, and T. Kanade. Geometric reasoning for single image structure recovery. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 2136–2143. IEEE, 2009.

[11] Y. Li and S. Birchfield. Image-based segmentation of indoor corridor floors for a mobile robot. In *Proceedings of the IEEE/RSJ Conference on Intelligent Robots and Systems (IROS)*. Citeseer, 2010.

[12] L. Lorigo, R. Brooks, and W. Grimsou. Visually-guided obstacle avoidance in unstructured environments. In *Intelligent Robots and Systems, 1997. IROS'97., Proceedings of the 1997 IEEE/RSJ International Conference on*, volume 1, pages 373–379. IEEE, 1997.

[13] B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proceedings of the 7th international joint conference on Artificial intelligence*, 1981.

[14] J. Rituerto, A. Murillo, and J. Kosecka. Label propagation in videos indoors with an incremental non-parametric model update. In *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*, pages 2383–2389. IEEE, 2011.

[15] C. Tomasi and T. Kanade. *Detection and tracking of point features*. School of Computer Science, Carnegie Mellon Univ., 1991.

[16] J. Zhou and B. Li. Robust ground plane detection with normalized homography in monocular sequences from a robot platform. In *Image Processing, 2006 IEEE International Conference on*, pages 3017–3020. IEEE, 2006.