Recognition of Wall Surfaces from Monocular Imagery for SLAM Applications in Indoor Environments

Karthik Mahesh Varadarajan and Markus Vincze
{kv, mv}@acin.tuwien.ac.at
ACIN, TU Wien

***Abstract*—Recognition of wall surfaces and other structural line segment features is an important aspect of LASER based Simultaneous Location and Mapping (SLAM) applications. Traditional indoor structural environment modeling algorithms employ schemes such as clustering of point clouds for parameterization and identification of the wall surfaces that serve as features to localize and map the environment of interest. RANSAC based line fitting and Spike detection are two common approaches in this regard. Alternatively, extensions to feature based stereo have also been used, mainly focusing on 3D line descriptions, along with techniques such as half-plane detection, real-plane or facade reconstruction, plane sweeping etc. Noise in the range data (from either LASER or stereo), especially in low texture regions, accidental line/plane grouping under lack of cues for visibility tests, presence of depth edges or discontinuities that are not visible in the 2D image and difficulties in adaptively estimating metrics for clustering can hamper efficiency of practical systems. In order to counter these issues, we propose a novel framework to use monocular imagery of the scene fusing 2D local and global features such as edges, texture and regions to detect wall surfaces and then use the resulting wall segmentation masks to guide the range data clustering. The focus of this paper is on the novel wall surfaces detection algorithm. Accuracy of recognition of wall surfaces is estimated with respect to a number of different material surface types, found in typical indoor environments. Testing on a representative dataset yielded recognition rates in excess of 95%, at 97% wall detection accuracy.**

I. Introduction

Recognition of wall surfaces and other structural line segment features is an important aspect of LASER based Simultaneous Location and Mapping (SLAM) applications. Traditional indoor 3D structural environment modeling algorithms employ schemes such as clustering of dense point clouds for parameterization and identification of the 3D surfaces. RANSAC based plane fitting [1] and Spike detection are common approaches in this regard. Alternatively, extensions to feature based stereo have also been used, mainly focusing on 3D line descriptions, along with techniques such as half-plane detection, real-plane or facade reconstruction, plane sweeping etc. Pioneering work in this regard is attributed to Baillard et al. [2,3] and Zisserman et al. [4]. Other important works include facade detection and multi-level regeneration by Lee - Nevatia [5]. Recent efforts at plane grouping based on PCA and visibility tests include [6] and [7]. The literature in building 3D line descriptor based structure analysis is also quite vast. Recent articles include Hausdorff measure based grouping [8] and model based recognition [9]. However, the performance of most of these techniques rapidly degrade in the presence of high amounts of noise (in range data such as stereo) under conditions of low illumination and in regions of low-texture or sparse features. Furthermore, accidental line/plane grouping (due to shelves/ cupboards), especially under lack of cues for visibility tests, presence of depth edges or discontinuities that are not visible in the 2D image and difficulty in adaptively estimating metrics for clustering can hamper efficiency of practical systems for door/doorway detection. On the other hand, traditional high quality laser [10] or panoramic camera based [11, 12] (multi-view) room modeling (often using piecewise planar modeling [13], triangulation [14] or space carving [15]) are often impractical for cost-effective domestic robots.

In order to counter these issues, we propose a novel framework to use monocular imagery of the scene fusing 2D local and global features such as edges, texture and regions to detect wall surfaces and then use the resulting wall segmentation masks to guide the range data clustering. The focus of this paper is on the novel wall surfaces detection algorithm. Accuracy of recognition of wall surfaces is estimated with respect to a number of different material surface types, found in typical indoor environments.

II. Overview

This paper offers the following novel contributions. Firstly, this paper demonstrates efficient indoor wall segmentation eliminating issues of over-segmentation caused by highlights and shadows, typical of conventional segmentation algorithms. Secondly, this paper offers a novel texture analysis algorithm that helps identify wall surfaces. Performance tests of the texture analysis algorithm, evaluated with respect to typical textured material

surfaces found in indoor environments shows high accuracy levels exceeding 95%.

The images used for evaluating the developed algorithms have been obtained from indoor environments under a variety of lighting conditions, with different wall, floor types as well as other material artifacts. The major issues in detection of walls in indoor environments arise from over segmentation in areas with highlights or shadows. Conventional segmentation algorithms are full gradient algorithms that do not distinguish between gradients arising from reflectance/ shading effects caused by diffuse or specular lighting from fixtures on the wall or from windows. These algorithms also do not analyze gradients arising from material properties of the objects differently from the gradients due to shading. As a result, shading effects can cause significant over-segmentation as demonstrated in Figure 1. In this paper, we present a gradient classification algorithm that is robust to such effects. Once the gradients have been classified, efficient segmentation is obtained using a simple edge based scheme. Finally, a region selection algorithm is presented which categorizes regions in the image as wall surfaces or otherwise. This algorithm uses the following assumptions for the wall modeling/ hypothesis:

Wall Modeling

Walls are typically characterized by

1. Homogeneous regions or areas with regular texture, usually with high numeric intensity values.
2. Largest single color regions in a given scene, especially under conditions of no large occluding obstacles in the vicinity.
3. Hold pixels with the farthest visible range information on planes parallel to the ground plane.
4. Frequent loss of homogeneity in color values owing to lighting and shading effects.

III. ALGORITHM

A. *Color Pre-processing*

The color images obtained from conventional cameras are pre-processed to eliminate gross errors. Locations of dead or noisy sensor pixels are pre-determined and the intensity at these locations approximated by nearest-neighbor filling. As a pre-processing step, the noise in the color image is reduced using a bilateral filter that preserves salient gradient values and hence sharp edges that are crucial for algorithms in the following stages of processing, including 2D segmentation.

B. *Intrinsic Reflectance Gradients Extraction*

The gradients of the filtered color image are estimated and these gradients are decomposed into shading and reflectance components. The shading component captures the lighting and shadows in the scene while the reflectance component captures the distinction in the material surfaces. This step is helpful to eliminate the highlights and shadow patterns created by light fixtures typically mounted on walls. Since walls are the primary focus of this SLAM enabling framework, it is beneficial to use reflectance components as they are devoid of gradients pertaining to highlight and shadow artifacts, thus representing the wall faces as true homogenous surfaces. The algorithm we employ is based on the intrinsic image extraction algorithm developed by Weiss [19] and extended by Tappen [20]. In the presented framework, gradients in the intensity channel of the color image are classified as 'shading' or 'reflectance' gradients by modeling an asymptotic linear color variation across neighboring pixels. The formulation for intrinsic image extraction [20] is

$$I(x,y) = S(x,y) \times R(x,y) \quad (1)$$

where $S(x,y)$ is the shading image, $R(x,y)$ is the reflectance image and $I(x,y)$is the input image defined in the dimensions x and y. Using a logarithmic transformation and applying multiple scale selective gradient/ derivative filters f_x, f_y we have the gradient images F_x and F_y, the (x,y) components of which can be classified as shading if the color pixels satisfy the constraints $c_{x+1} = \alpha c_x$ and $c_{y+1} = \alpha c_y$ respectively and as reflectance otherwise. The component images can be reconstructed as

$$C(x,y) = g * [(f_x(-x,-y) * F_{cx}) + (f_y(-x,-y) * F_{cy})] \quad (2)$$

where, * represents convolution, $F_{cx}$and $F_{cy}$are component (shading/ reflectance) gradients and g is obtained from

$$g * [(f_x(-x,-y) * f_x(x,y)) + (f_y(-x,-y) * f_y(x,y))] = \delta \quad (3)$$

The shading and reflectance components as defined by equation (2) are shown in Fig. 1C and 1D. In our framework, the reflectance image gradients F_{cx} and F_{cy} are used directly in the segmentation process.

Figure 1. Intrinsic Image Extraction and Segmentation (A) Input color image (B) Segmentation using the standard Felzenszwalb-Huttenlocher (FH) graph based algorithm – demonstrates high clutter in regions of the left wall with lighting changes (C) Shading intrinsic image (D) Reflectance intrinsic image – note that C and D (obtained by inversion of input image gradients classified as shading or reflectance respectively) (E) Segmentation on the input image using a low complexity multi-scale full gradient edge analysis scheme (F) Segmentation on the input image with the same scheme using reflectance-only gradients – shows superior performance in wall regions affected by lighting changes in comparison with full-gradient image segmentation schemes such as the graph based FH. Similar values of gradient and region size thresholds were used for all three segmentation scenarios.

C. 2D Reflectance Gradient based Segmentation

Using the gradients F_{cx} and F_{cy} obtained in the previous stage, segmentation is carried out using a low complexity multi-scale edge analysis scheme. The scheme links edges found at various scales (by analysis of reflectance only gradients) using proximity and similarity measures to form enclosed regions or segments. The choice of the segmentation algorithm is based on the goal of meeting real-time constraints for deployments on robots, which excludes the possibility of using algorithms like the Felzenszwalb-Huttenlocher (FH) graph based algorithm. It can be seen from Fig. 1B, 1E and 1F that the output of the proposed 'reflectance gradient only' segmentation approach is superior to traditional full-gradient algorithms like FH (with gradients as grid graph edge weights) and the variant of the multi-scale edge analysis algorithm operating on full-gradients, given the given context of wall detection with lighting and shading changes. The performance is comparable in regions devoid of lighting changes. Any possible over-segmentation in regions of high texture does not affect the output since these regions are unlikely to be wall surfaces. The segmented regions are then subjected to a region selection algorithm to select walls and wall-like structural surfaces.

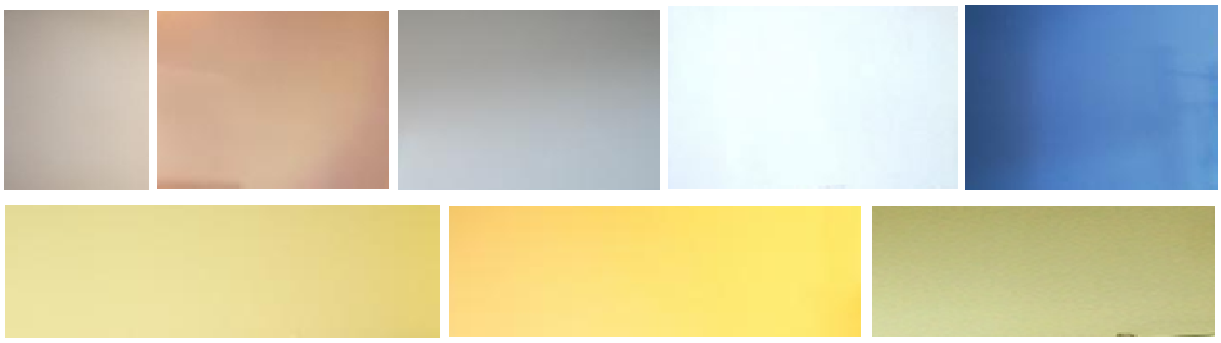

Figure 2. Sample image texture chips used for training/ weight estimation for wall detection – Positive training examples

D. Wall Region Isolation using Texture Analysis

The characteristic features of walls such as low texture, high homogeneity, large pixel spans and representations using high gray-scale intensity values are used in region selection. The current framework employs 2 levels of thresholds (hard and soft) on measures of entropy (E), homogeneity (H), uniformity energy (U), correlation (R), contrast (C) and other constraints based on the Grey Level Co-occurrence Matrix (GLCM) to select wall-like surfaces. Based on a number of positive and negative training examples (Fig. 2 and Fig. 3), weighting factors were estimated using a regression approach. Estimated soft threshold values, along with the assigned confidence values of the measures on condition conformance (in brackets) for the two-class separation (positive wall classification) are H>0.99 (1.0), C<0.0275 (1.0), R>0.9 (0.9), U>0.6 (0.3) and E<5.5 (0.8). R can also be undefined or slightly negative under near perfect surface homogeneity and reducing slightly below this threshold for walls with rough

texture such as in the case of visible brick layouts. U is unreliable and varies with the lighting changes; ideally 1.0 under no lighting change or natural lighting but drops to 0.3 under artificial lighting/ large lighting changes.

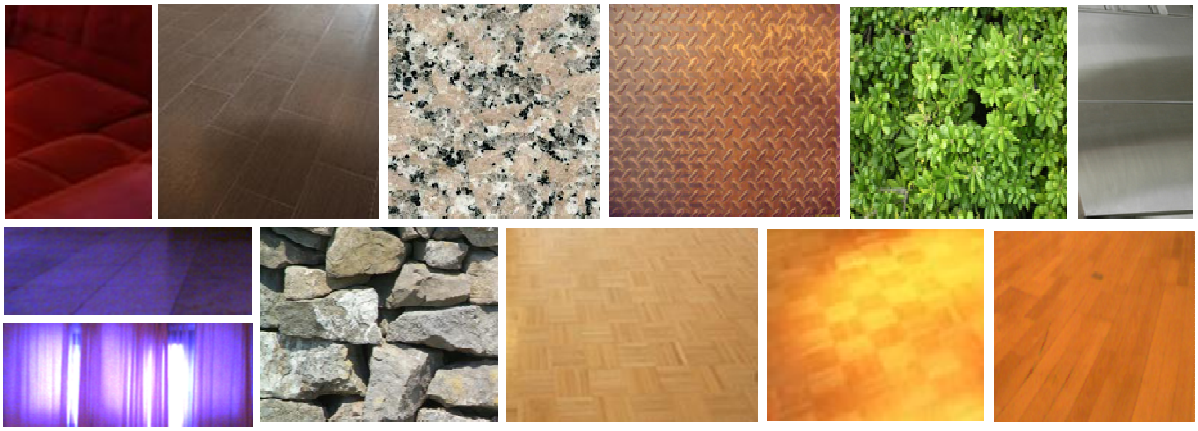

Figure 3. Sample image texture chips used for training/ weight estimation for wall detection – Negative training examples

The hard threshold values are H>0.96 (0.8), C<0.0475 (0.7), R>0.85 (0.6), U>0.3 (0.1), E<7.0 (0.5). The surface is classified as a wall if the aggregate confidence value exceeds 3.0 out of a maximum of 4.0. An alternative approach to learn these weights based on in situ adaptive learning in the environment of interest based on SVM was also evaluated.

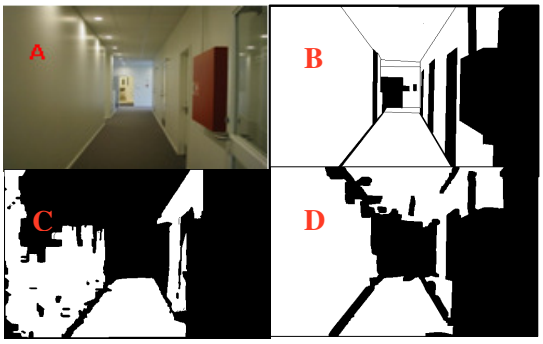

Figure 4. Segmentation and Regions of Interest Selection (A) Input color image (B) Ground Truth (Manual Segmentation) for walls and wall-like regions (floors, ceiling etc.) (C) Results using FH (mislabeled pixels: 70292) (D) Results using our framework (mislabeled pixels: 32069) – note the correspondence to Fig. 1B and 1F respectively, mislabeled pixels are estimated by XOR logical gating with the ground truth.

IV. RESULTS

On a representative data set of 80 image chips of various material textures found indoors, such as wood, tile, brick, rock, vegetation, carpet, cloth, curtain, steel, bronze, tree bark, granite etc., besides painted wall surfaces, the simple classifier achieved a classification rate of 95%, with a wall detection rate of 97.87%. False alarms were caused due to white curtains, steel and floor tiles that were 'wall-like' or homogenous. The features were robust to wall colors and surface roughness. Objects such as uniformly colored doors and cupboard doors had confidence measures close to that of walls. Since these surfaces can also be helpful in mapping, they can used in further analysis. Thresholds on pixel spans of the surfaces (> I_w*I_h/15, where, I_w is image width and I_h is image height) and average gray-scale intensity (> 100/255), can further help reduce the detected segments to the set of primary room surfaces. Use of SVM to adaptively estimate these thresholds based on in situ training in a specific environment of deployment of the SLAM system of the robot produced wall detection accuracies of about 95%. Fig. 4 demonstrates the results of the segmentation and region selection approach and compares the performance with the output of the region selection in combination with the FH segmentation algorithm. Note that weights were selected to permit identification of floors, in addition to ceiling and walls for this specific test. While the number of mislabeled pixels is 70292 with FH, this number is less than half (at 32069) for our framework. The region masks thus obtained can be used for refining range feature points. Fig.5 presents additional results under a variety of scenarios in a given home environment of interest with dim spot lighting. It can also be seen from Fig. 5 that the approach works quite robustly, irrespective of lighting and shading effects on walls and under heavy clutter.

Figure 5. Results of segmentation (row 2) and wall detection (row 3) for various input scenarios (row 1). In the segmentation output, segments are colored uniquely while detected wall regions in the wall detection output are in white.

V. Future Work

Integration of the wall detection algorithm with a SLAM system based on Stereo or LASER forms the next phase work. Such a system can be expect to improve the robustness of detection and tracking of features necessary for SLAM, resulting in an efficient 3D indoor mapping system. Implementation of algorithm on GPU for rapid operation along with the SLAM system is the future goal of this effort.

VI. References

[1] A. Bartoli, "A random sampling strategy for piecewise planar scene segmentation", Computer Vision and Image Understanding, 2007.
[2] C. Baillard, C. Schmid, et al., "Automatic line matching and 3D reconstruction of buildings from multiple views", ISPRS 1999.
[3] C. Baillard and A. Zisserman."A plane-sweep strategy for the 3D reconstruction of buildings from multiple images". ISPRS, 2000.
[4] C. Schmid and A. Zisserman."Automatic line matching across views". CVPR, 1997.
[5] S.C.Lee and R. Nevatia."Interactive 3D Building Modeling Using a Hierarchical Representation", ICCV2003.
[6] JK. Lee, S. Ahn et al., "A Prospective Algorithm for Real Plane Identification from 3D Point Clouds and 2D Edges", ICHIT '08.
[7] JK. Lee, S. Ahn et al., "Visibility-Based Test Scene Understanding by Real Plane Search", Advances in Visual Computing, Springer, 2008.
[8] C. Guerra et al."Line-based object recognition using Hausdorff distance from range images to molecular secondary structures" IVC05
[9] SH. Chang, S. Lee, D. Moon, et.al,"Model based 3D Object Recognition using Line Features", ICAR 2007.
[10] P. Biber, et al., "3D Modeling of Indoor Environments by a Mobile Robot with a Laser Scanner and Panoramic Camera", IROS 2004.
[11] P. Doubek, T. Svoboda, "Reliable 3D reconstruction from a few catadioptric images", OMNIVIS 2000.
[12] M. Nevado, et al. "Obtaining 3D models of indoor environments with a mobile robot by estimating local surface directions", RAS 2004.
[13] A. Dick, R. Cipolla et al. "Combining Single View Recognition & Multiple View Stereo for Architectural Scenes", ICCV 2001
[14] DD. Morris, T. Kanade, "Image Consistent Surface Triangulation", CVPR 2000.
[15] K Kutulakos, "Approximate N View Stereo", - Lecture Notes in Computer Science, 2000
[16] AC. Murillo, J. Kosecka, et al.,"Visual door detection integrating appearance and shape cues", RAS 2008.
[17] Z. Chen, S.T. Birchfield, "Visual Detection of Lintel-Occluded Doors from a Single Image", CVPRW 2008.
[18] R. Munoz-Salinas, E. Aguirre, M. Garcıa-Silvente, "Detection of doors using a genetic visual fuzzy system for mobile robots", Autonomous Robots, 2006.
[19] Y. Weiss, "Deriving intrinsic images from image sequences", ICCV, 2001.
[20] MF. Tappen, WT. Freeman, EH. Adelson, "Recovering intrinsic images from a single image", PAMI 2005.