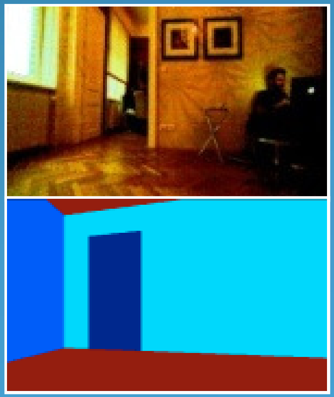
Functional Room Detection and Modeling using Stereo Imagery in Domestic Environments



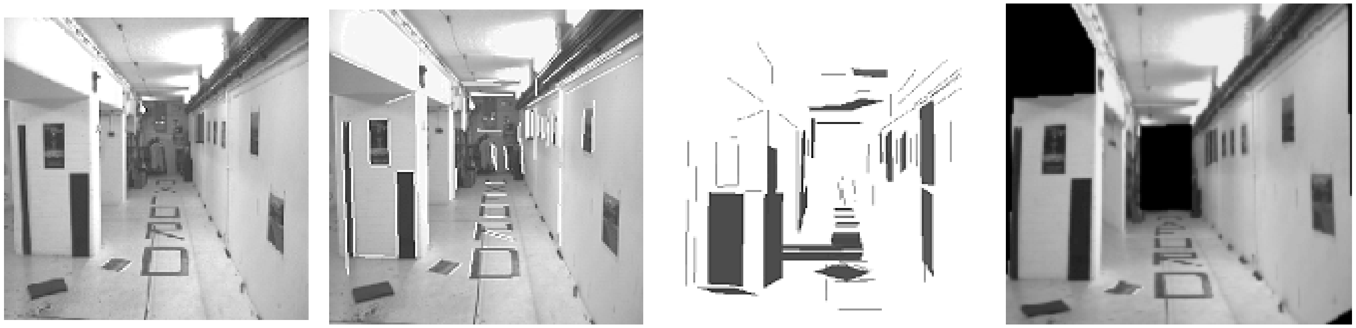
Karthik Mahesh Varadarajan, Markus Vincze
Vision for Robotics, ACIN
TU Wien, Austria.
{kv, mv}@acin.tuwien.ac.at

Domestic Robots – Requirements and Constraints

- Detection of room structure, doorways and connectivity between functional units in homes is crucial for place learning and task related navigation esp. for floor/ wall cleaning robots etc.
- In situ functional detection and classification of rooms in indoor environments – important in training and map building stages of deployment of robotic assistants
- Traditional place learning methods do not perform functional room or unit identification
  - Explicit user labeling of places as well as map editing
  - Feature based methods to detect typical objects and hypothesize room functionality/ learn places based on localization of these objects (RobotVision @ ImageCLEF challenge)
  - Unsuitable in dynamic environments or in unoccupied/ unfurnished homes
- Functional semantic definitions for rooms

Current Indoor Structural Modeling and Doorway Detection Scenario using EO

- 3D surface characterization by clustering point clouds, typically using RANSAC or its extensions
- Feature based multi-frame/ stereo using 3D line descriptions- half-plane detection, real-plane or facade reconstruction, plane sweeping - Baillard, Zisserman, ISPRS 2000

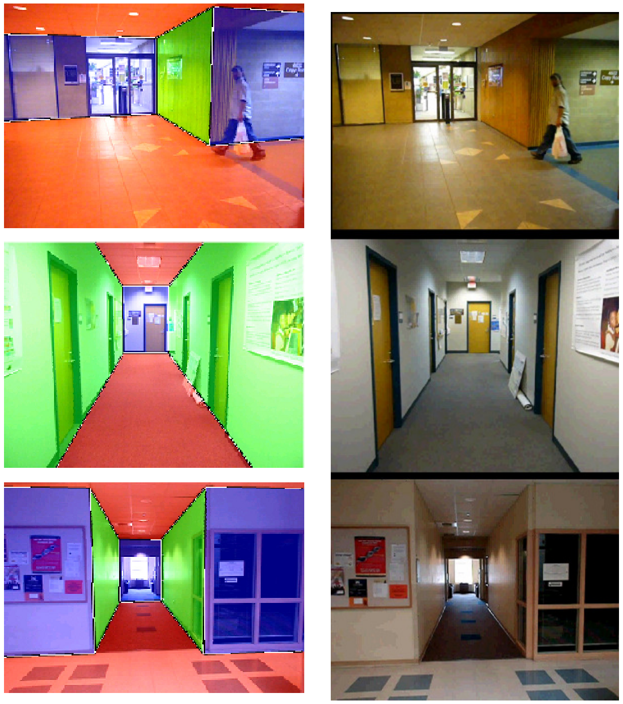


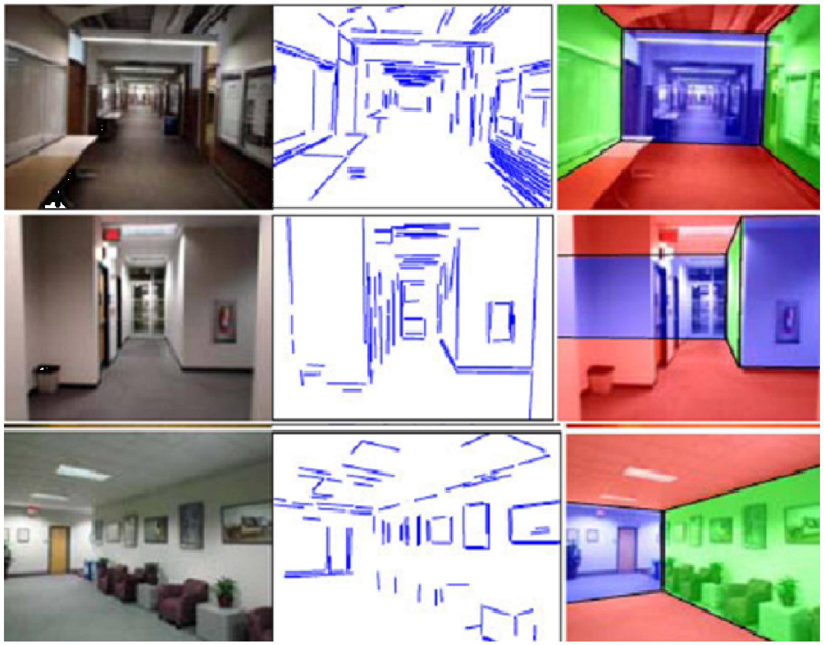
Current Indoor Structural Modeling and Doorway Detection Scenario using EO

- Machine Learning based Door detection - Murillo, Kosecka, RAS 2008
- Facade detection and multi-level regeneration - Lee, Nevatia, ICCV 2003
- Panoramic camera based mapping
- Piecewise planar modeling – Dick, Cipolla, ICCV 2001, Triangulation – Morris, Kanade, CVPR 2000 or Space carving – Kutulakos 2000

Current Indoor Structural Modeling and Doorway Detection Scenario using EO

- Geometric constraints on corners and edges - Lee et al., CVPR 2009





EO Processing - Problems

- Rapid degradation in presence of high amounts of noise under conditions of low illumination and in regions of low-texture or sparse features
- 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
- Lack of adaptive clustering metrics
- Clutter
- Wall and floor reflectance
- Open doors, partial view or case of doorway being structurally similar to an arch, lacking the actual door frame

Range Processing - Problems

- Inference and Message Passing techniques are integral to surface generation
  - Dynamic Programming, Belief Propagation, Relaxation, Diffusion, Graph Cut etc.
- Computational complexity too high to support real-time 3D surface generation esp. on robots
- Excessive smoothing at depth discontinuities resulting in loss of structure (esp. where the 2D image does not provide structural cues)
- Unsuitable for diffusion of extremely sparse depth data (such as - homogenous surfaces)
- Propagation of gross errors in the initial data can significantly affect end reconstruction – traditional scene agnostic filtering schemes are slow and ineffective on surfaces with sparse data points
- Assumption of dependence on co-planarity or curvature metrics of data points

Proposed Solution

- Depth based doorway detection
- Novel framework fusing 2D local and global features such as edges, textures and regions, with geometry data obtained from pixel-wise dense stereo
- Room functionality hypothesis
- Key algorithms
  - Wall detection
  - Real-time depth diffusion
  - Depth segmentation algorithms – identification of depth edges
  - Grouping walls for room reconstruction and doorway detection
  - Room utility labeling

Solution Pipeline

- Three tier process
  - Detection of walls:
    a) Walls and wall-like surfaces detected using 2D edge, texture and region features.
    b) Piecewise depth diffusion
    c) Depth segmentation to identify intra - wall depth orientation changes, discontinuities
  - Modeling of enclosing room:
    - Built by selecting wall surfaces to fit approximate cuboidal constraints
  - Estimation of doorways:
    - Doorways estimated by clustering of the dense stereo data pixels that do not conform to the concave room hypothesis.

Key Contributions

- Innovative framework for functionality based room boundary detection
- Scheme for room functionality determination and place learning based on structural semantics of the room
- Integrated framework for complete functional room modeling from stereo range data

Color Image Processing

- Color Pre-processing
  - Range Registration
  - Bilateral Filtering
- Intrinsic Reflectance Gradients Extraction
  - Gradient Classification – Reflectance and Shading components
  - Extension of Weiss – Tappen scheme

$$I(x,y) = S(x,y) \text{ x } R(x,y)$$

$$c_{x+1} = \alpha c_x$$

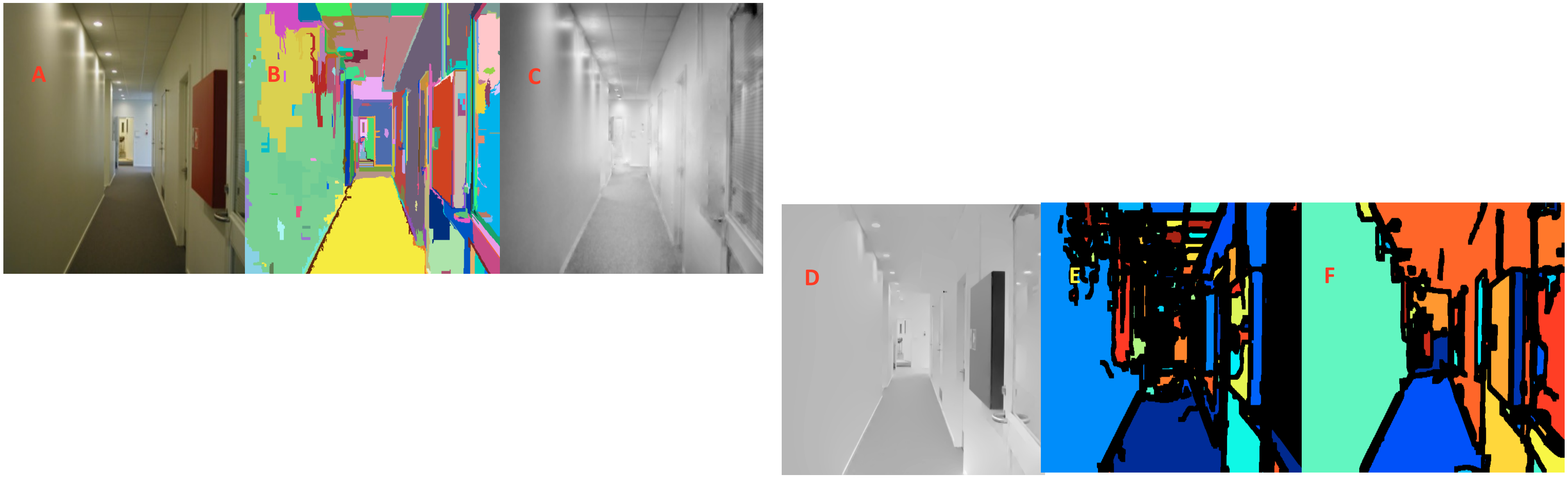
$$c_{y+1} = \alpha c_y$$

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

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

2D Reflectance Gradient based Segmentation

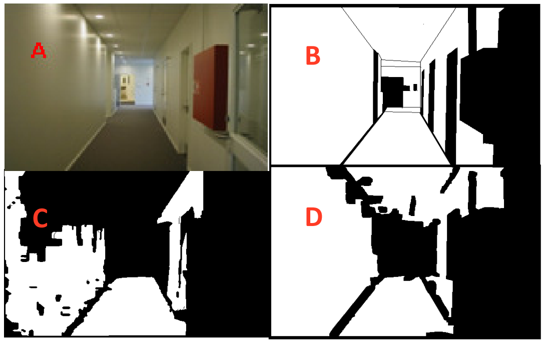
- Low complexity multi-scale edge analysis scheme
- Based on need for real-time operation



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.

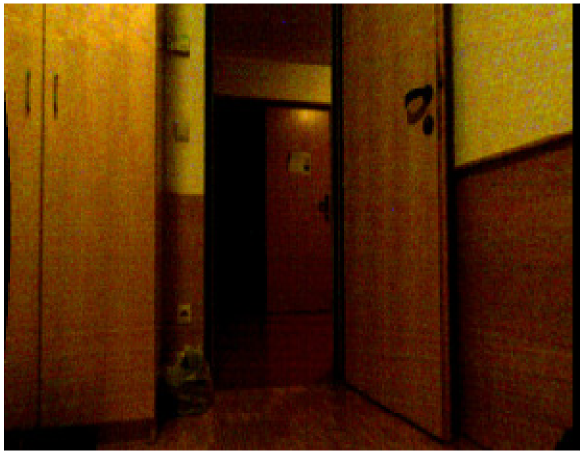
Region Isolation using Texture Analysis

- Characteristics of walls: Low texture, High homogeneity, Large pixel spans (> $I_w * I_h/15$) and representations using high gray-scale intensity values (> 100/255)
- Entropy (E), homogeneity (H), uniformity energy (U), correlation (R), contrast (C)
  - Soft Thresholds:
    - H>0.99 (1.0)
    - C<0.0275 (1.0)
    - R>0.9 (0.9, also undefined/ negative) reducing for walls with rough texture
    - U>0.6 (0.3) – unreliable, varies with lighting changes
    - E<5.5 (0.8)
  - Hard Thresholds:
    - 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)
  - Positive Classification for > 3.0/4.0 : 95% Accuracy
  - Machine Learning for deployment scene

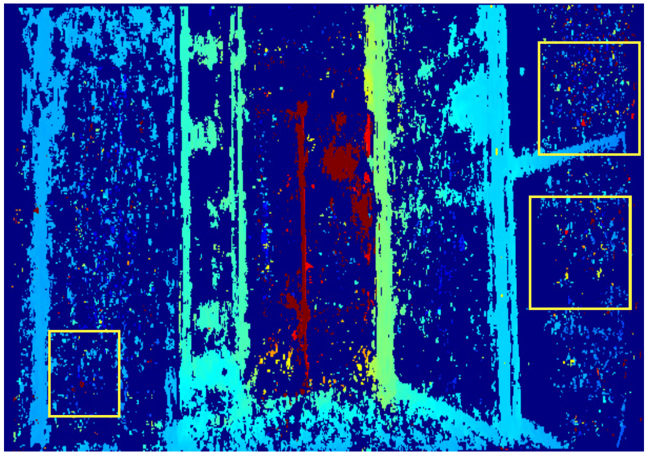


Segmentation and Regions of Interest Selection (A) Input color image (B) Ground Truth (Manual Segmentation) for walls and wall-like regions (floors, ceiling etc.) (C) Results using FH (mislabeled pixels: 70292) (D) Results using our framework (mislabeled pixels: 32069) – note the correspondence to Fig.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

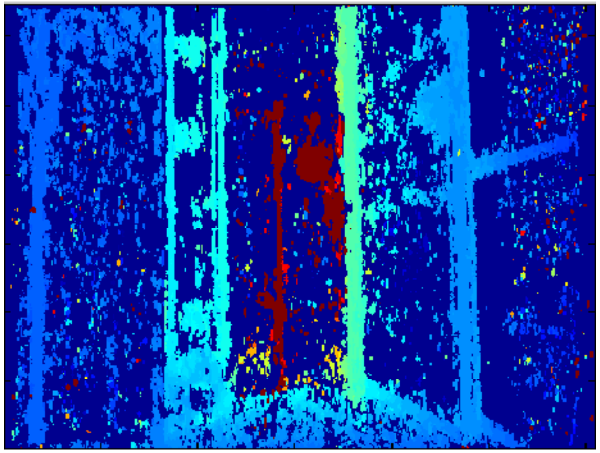
Iterative Hysteresis Filtering and Morphological Reconstruction



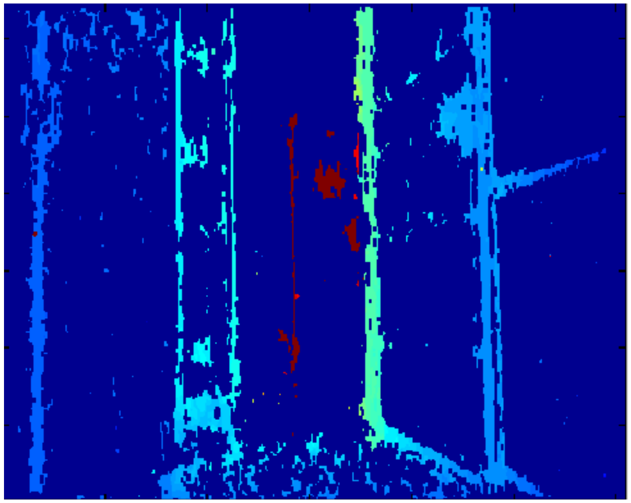
Input Image



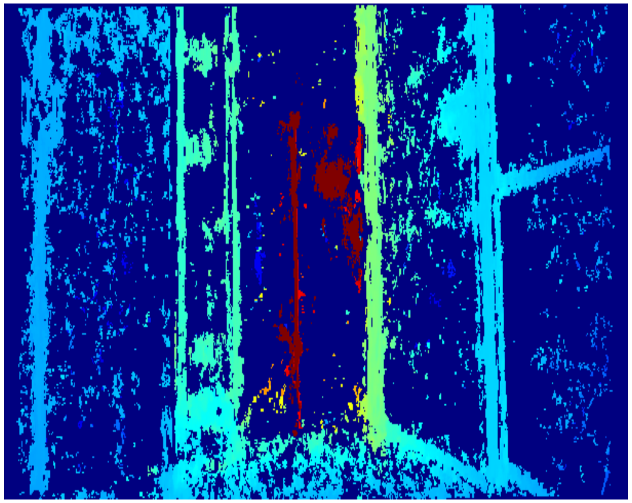
Input Depth Map



Mode Filter Output



Median Filter Output



Proposed Filter Output

Piecewise Isotropic Depth Diffusion

- Heat Diffusion

$$\frac{\partial u(\mathrm{r},t)}{\partial t} = c\left(\frac{\partial^2 u(\mathrm{r},t)}{\partial x^2} + \frac{\partial^2 u(\mathrm{r},t)}{\partial y^2}\right)$$

$$\frac{\partial u(\mathrm{r},t)}{\partial t} = c\nabla^2 u(\mathrm{r},t)$$

- Formulation for Depth Diffusion

  c is the binary image mask for piecewise isotropic formulation

$$\begin{aligned}\frac{\partial u(r_{(i,j)},t)}{\partial t} &= u\left(r_{(i,j)},t+1\right) - u\left(r_{(i,j)},t\right)\\ &= \varphi\big[\, c_{(i-1,j)}.\nabla u\left(r_{(i-1,j)},t\right)\\ &+ c_{(i,j-1)}.\nabla u\left(r_{(i,j-1)},t\right)\\ &+ c_{(i+1,j)}.\nabla u\left(r_{(i+1,j)},t\right)\\ &+ c_{(i,j+1)}.\nabla u\left(r_{(i,j+1)},t\right)\big]\end{aligned}$$

Piecewise Isotropic Depth Diffusion

- where, the constant ≤ 0.25 controls the overall rate of diffusion. In the steady state

$$\begin{aligned}(1/\varphi).u\left(r_{(i,j)},t_{ss}\right)\\ &- \left[c_{(i-1,j)}.u\left(r_{(i-1,j)},t_{ss}\right)\right.\\ &+ c_{(i,j-1)}.u\left(r_{(i,j-1)},t_{ss}\right)\\ &+ c_{(i+1,j)}.u\left(r_{(i+1,j)},t_{ss}\right)\\ &+ \left.c_{(i,j+1)}.u\left(r_{(i,j+1)},t_{ss}\right)\right] = 0\end{aligned}$$

- Representing (1/Φ) as λ and linearizing the tuple indices, a matrix system is obtained. Sample matrix for a 3x3 depth image

$$\begin{bmatrix} \lambda & c_{12} & 0 & c_{21} & 0 & 0 & 0 & 0 & 0 \\ c_{11} & \lambda & c_{13} & 0 & c_{22} & 0 & 0 & 0 & 0 \\ 0 & c_{12} & \lambda & 0 & 0 & c_{23} & 0 & 0 & 0 \\ c_{11} & 0 & 0 & \lambda & c_{22} & 0 & c_{31} & 0 & 0 \\ 0 & c_{12} & 0 & c_{21} & \lambda & c_{23} & 0 & c_{32} & 0 \\ 0 & 0 & c_{13} & 0 & c_{22} & \lambda & 0 & 0 & c_{33} \\ 0 & 0 & 0 & c_{21} & 0 & 0 & \lambda & c_{32} & 0 \\ 0 & 0 & 0 & 0 & c_{22} & 0 & c_{31} & \lambda & c_{33} \\ 0 & 0 & 0 & 0 & 0 & c_{23} & 0 & c_{32} & \lambda \end{bmatrix} \begin{bmatrix} u_{11} \\ u_{12} \\ u_{13} \\ u_{21} \\ u_{22} \\ u_{23} \\ u_{31} \\ u_{32} \\ u_{33} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$Ax = B$$

Piecewise Isotropic Depth Diffusion

- This results in a block-tridiagonal matrix system with fringes (blocks are red, tri-diagonals are blue and violet along with the main diagonal, upper fringe in green and lower fringe in orange)
- $a_1(i,j)$ – the lower diagonal elements (blue), $b_1(i,j)$ – the middle diagonal elements, $c_1(i,j)$ – the upper diagonal elements (violet), $a_2(i,j)$ – the lower fringe elements (orange), $c_2(i,j)$ – the upper fringe elements (green)
- Adaptation of Del'Osso IBS method: Pseudo-code

```
FringeTriDiagSolver := {
InitializeSolution,
    InitializeMatrixComputation,              $i_{iter}$ -> 0,
While[{CurrEps > EpsTol && $i_{iter}$ < MaxItr && AbsErr > AbsErrTol},{
        $i_{iter}$ -> $i_{iter}$ + 1,
        StorePreviousResult,
        ForwardSubstitution,BackwardSubstitution,
        ComputeMaximumResidual}]          }
```

Piecewise Isotropic Depth Diffusion

- Intermediate matrices are computed as

$$G(i,j) = 1/(-a_1(i,j) * Q_1(i-1,j) - a_2(i,j) * Q_2(i,-1) - b_1(i,j))$$

$$Q_1(i,j) = G(i,j) * (a_2(i,j) * Q_1(i,j-1) * Q_2(i+1,j-1) + c_1(i,j))$$

$$P_1(i,j) = Q_1(i,j) * X(i+1,j)$$
$$P_2(i,j) = Q_2(i,j) * X(i,j+1)$$

$$Q_2(i,j) = G(i,j) * c_2(i,j)$$

- *ForwardSubstitution* and *BackwardSubstitution* modules are iterated until convergence of X estimated as

$$M(i,j) = G(i,j) * (a_1(i,j) * (M(i-1,j) + P_2(i-1,j) + P_3(i-1,j)) + a_2(i,j) * (Q_1(i,j-1) * (M(i+1,j-1) + P_1(i+1,j-1)) + M(i,j-1) - S(i,j))$$

$$P_1(i,j) = Q_1(i,j) * X(i+1,j)$$

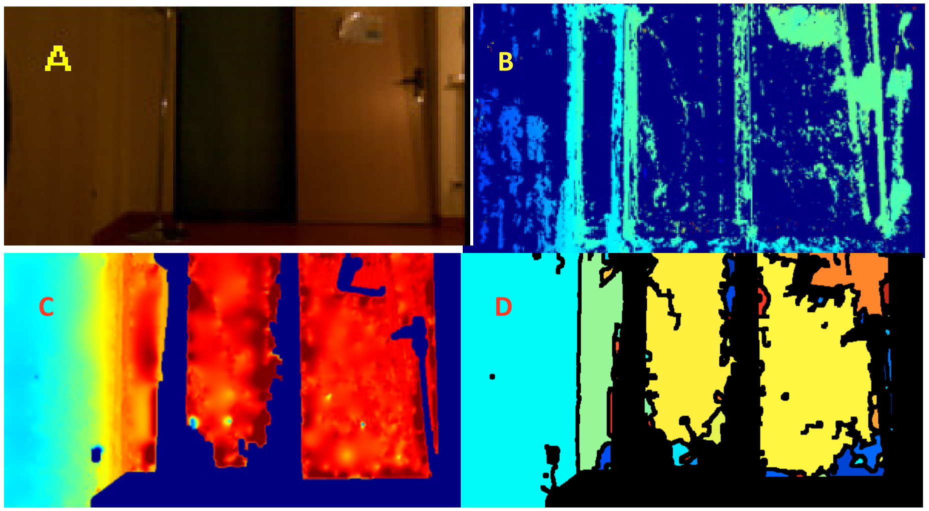
$$P_2(i,j) = Q_2(i,j) * X(i,j+1)$$

$$X(i,j) = M(i,j) + P_1(i,j) + P_2(i,j) + P_3(i,j)$$

- Multi-grid to enhance convergence

Depth Segmentation

- Low-complexity multi-scale edge detection and linking approach
- Identifies wall junctions, boundaries, columns etc.



Depth Diffusion and Segmentation (A) Input image – please note that the depth edge formed by the junction of wall plane parallel to the observer (in front) and perpendicular to the observer (to the left) is hardly visible and hence color image segmentation produces one single surface (B) Input depth map (C) Diffused depth map (after filtering & using mask from step C) – here the depth edge is clearly visible (D) Segmentation in depth identifies the depth edge (between the blue and green segments)

Functional 3D Room Detection

- Surface Fitting
  - Plane fitting using Iteratively Re-weighted Least Squares Robust Linear Regression
  - Plane Selection

$$\frac{1}{Z}=\left(\frac{A}{f_x D}\right)x+\left(-\frac{B}{f_y D}\right)y+\left(\frac{C}{D}-\frac{A(1-C_x)}{f_x D}+\frac{B(1-C_y)}{f_y D}\right)$$

- Functional Room Boundaries Detection
  - PCA grouping
  - Cuboidal and Manhattan constraints
  - Convexity assumptions critical
  - No visibility tests required

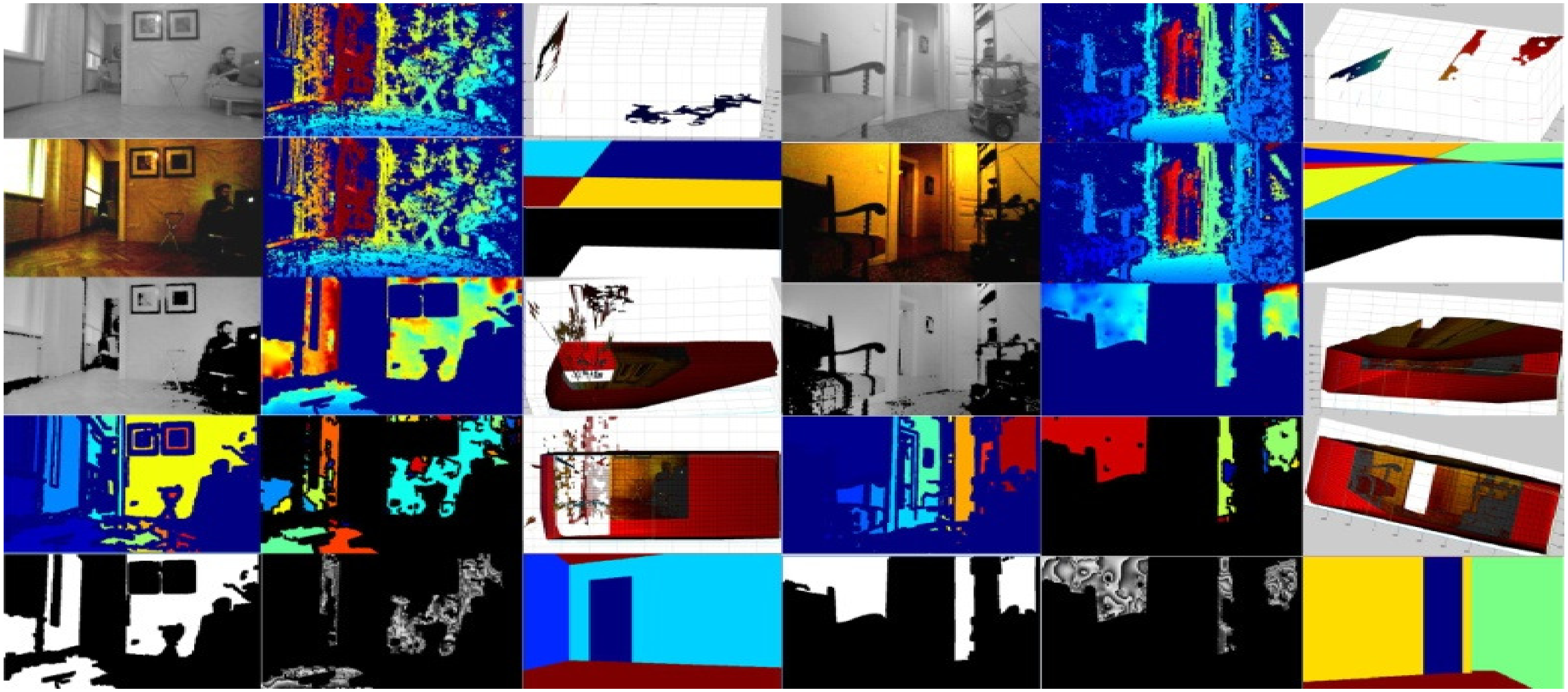
Functional 3D Room Detection

- Functional Room Boundary Detection
  - In domestic environments, barriers represent boundaries of functional separation
    - American kitchen - usually separated from the hall by a simple barrier and not a doorway
    - Dining halls - typically sections of the hall separated from the functional hall area by an open boundary, i.e. while the functional hall area forms a convex shape such as a cuboid, the dining area forms a separate cuboid adjoining it
  - Planes ranked based on consistency, span, texture content and Manhattan constraints
- 3D Room Reconstruction and Doorway Detection
  - Clustering of depth pixels at jump discontinuity

Functional 3D Room Detection

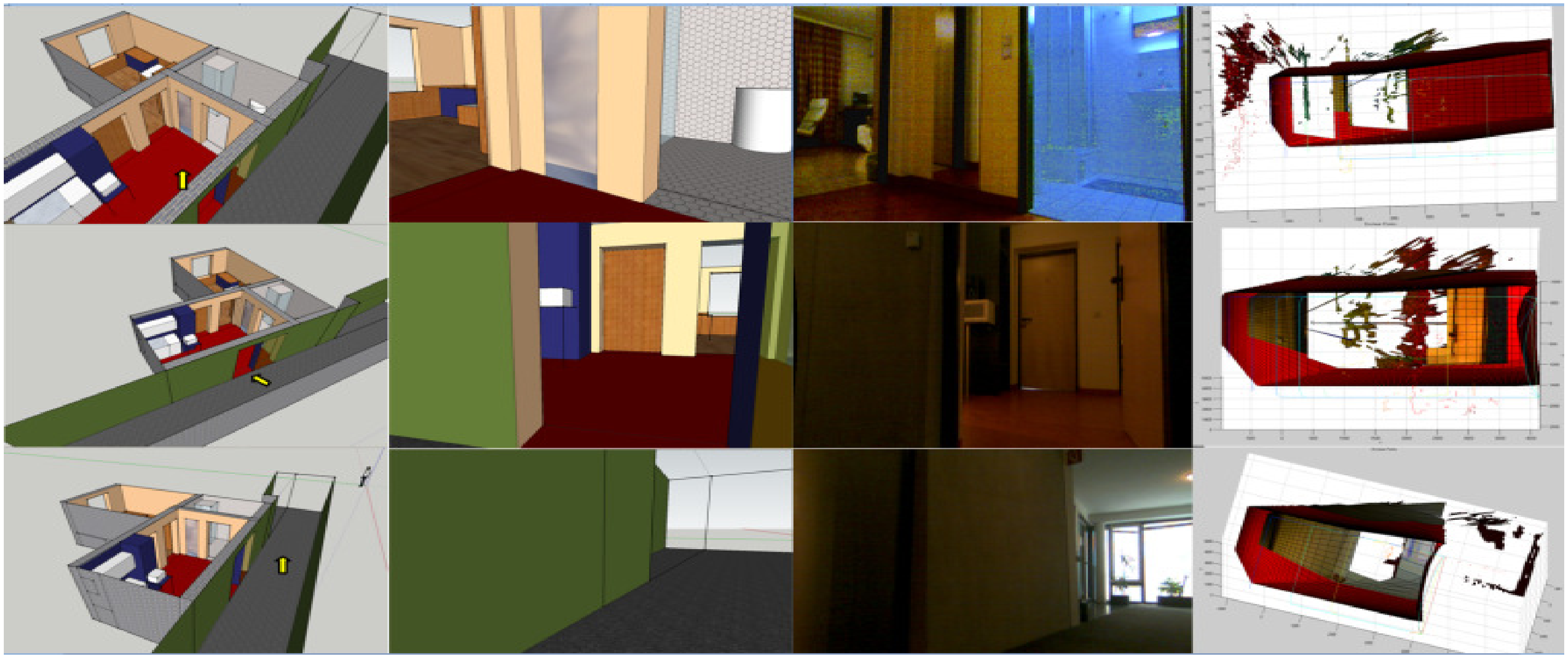
- Room Functionality Hypothesis
  - Based on area, number of doorways, number and size of open boundaries and topology of the scene
  - Semantic Rules
    - Large areas with more than one door – halls/ living area
    - Parallel walls with low numerical values for the distance between the walls and open boundaries in orthogonal directions – Corridors
    - Areas adjoining halls with open boundaries leading to the hall – Kitchen or Dining – based on size of open boundary
    - Regions with a single doorway - bedrooms or bath – based on area and depth
  - Label assignment to areas beyond doorways in the field of view
  - Single-shot image processing, extensible with metric SLAM maps or topological maps
  - Multi-frame processing, full 3D processing

Results



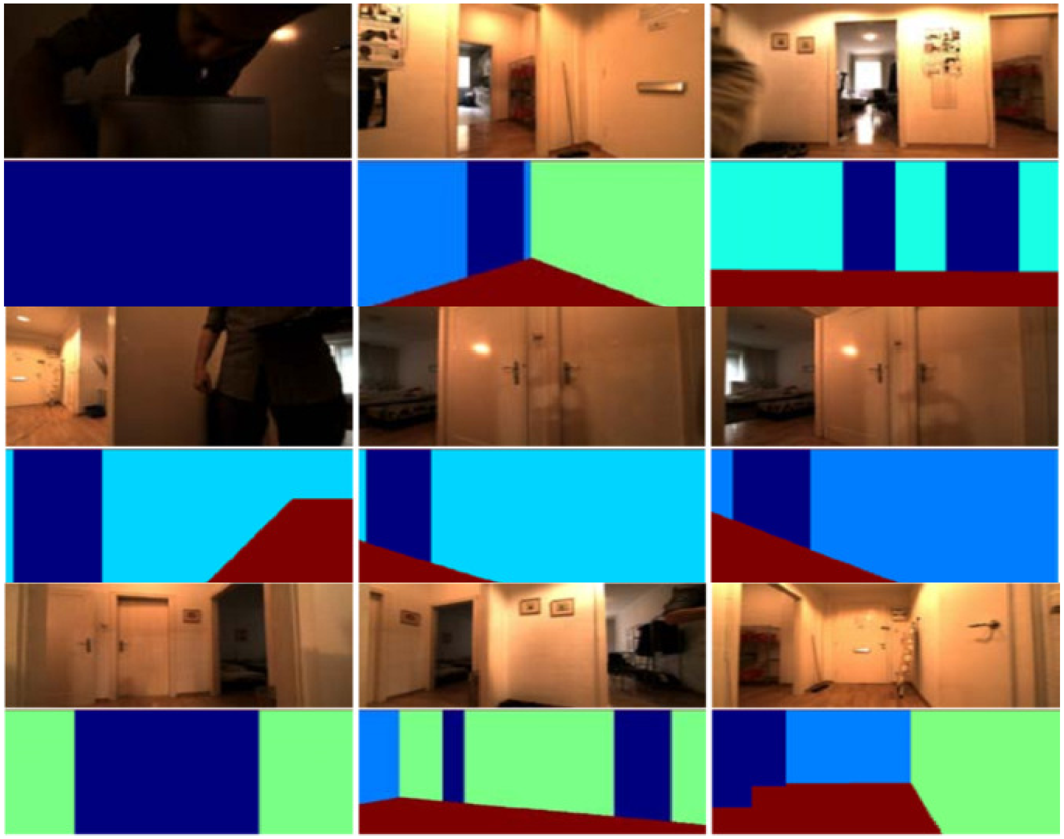
Complete Algorithmic Pipeline for 2 input scenes (columns 1-3 and 4-6) (Top to bottom) Left Row – Color image processing: (A) 2D input monochrome left image (B) 2D color image (C) Reflectance image (D) Image segmentation output (E) Region selection output (of wall-like structures) Middle Row - (A) Input depth map (B) Filtered depth map (C) Depth diffusion output (D) Depth segmentation output (E) Masked depth map Right Row - (A) Surface fitting results (near top view) (B) Surface categorization & PCA based boundary detection (C) 3D Reconstruction (top view (D) 3D Reconstruction (front view) (E) 3D map reprojected to camera plane

Results



3D Room reconstructions for 3 scenes (row-wise) (Left to right) (A) 3D ground truth model of room along with camera viewpoint (B) Synthetic image from camera viewpoint (C) Input scene (D) Generated 3D model of the room with doorways denoted by cavities in the model leading to 3D data points from the room observed through the doorway. For the 3 cases, the functional labels output by the system are (1) hall and adjoining rooms as bedroom and bath (2) unknown (3) corridor

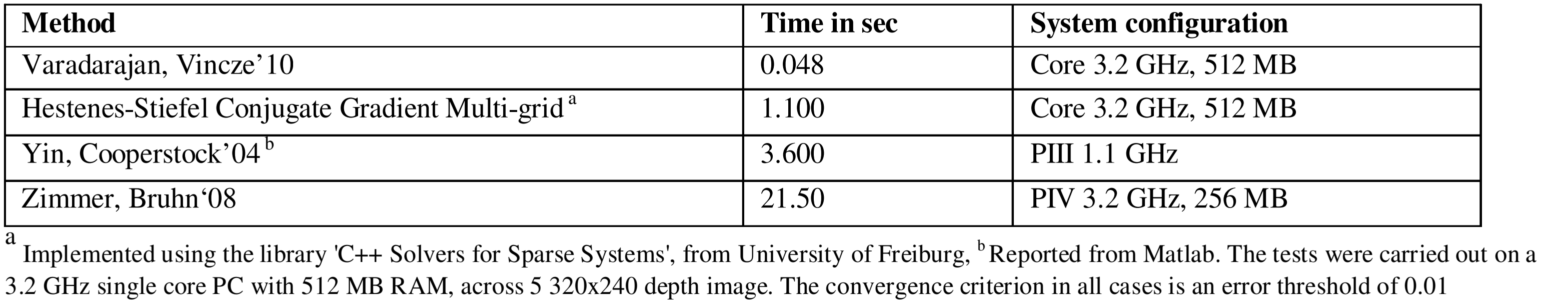
Results



Results from test environment (Top rows) Input scenes (Bottom rows) 3D model reprojected on to the image plane. The percentage of mislabeled pixels (w.r.t to manual segmentation) was 5%.

Results

- Run-time comparison of Depth Diffusion solvers



Method	Time in sec	System configuration
Varadarajan, Vincze'10	0.048	Core 3.2 GHz, 512 MB
Hestenes-Stiefel Conjugate Gradient Multi-grid[a]	1.100	Core 3.2 GHz, 512 MB
Yin, Cooperstock'04[b]	3.600	PIII 1.1 GHz
Zimmer, Bruhn'08	21.50	PIV 3.2 GHz, 256 MB

[a] Implemented using the library 'C++ Solvers for Sparse Systems', from University of Freiburg, [b] Reported from Matlab. The tests were carried out on a 3.2 GHz single core PC with 512 MB RAM, across 5 320x240 depth image. The convergence criterion in all cases is an error threshold of 0.01

- Pixel mislabeling error is 5 times higher for RANSAC (our framework: ~8713, RANSAC: ~40125)
- Reprojection error is about 5% in typical scenes
- Robust detection of room boundaries and doorways, even in conditions of heavy clutter , specular highlights and non-salient EO edges

Conclusion

- Framework for functionality based room boundary detection and place learning in domestic environments
- Integrated system for functional room boundary detection from stereo
  - Highly efficient with extremely sparse range data
  - Preserves and detects depth edges in regions where there are no visible edges in color data
  - Handles shadows and specular highlights effectively
- Simple semantic features based on doorways and area used for room functional analysis
- Future work: Improvements to model based on other semantic features, Full SLAM integration

Thank you !!!

Extra Slides

Iterative Hysteresis Filtering and Morphological Reconstruction

- Algorithm

1. Divide the input depth map into core-blocks (sqrt. $I_w/10$) and macro-blocks ($I_w/10$) based on expected spans of surfaces at mean ranges from camera (for given FOV) - typical block sizes of 50x50 and 7x7

2. σ_c and $\sigma_{m.}$ estimated using values of depth pixels with high confidence measures (in the confidence map)

3. Logical maps of valid pixels (falling within a pre-determined threshold times σ_c and σ_m) are estimated. Rough rule of the thumb calculation for the macro-block and the core-block thresholds are 0.25/P and 0.175/P (P is the percentage pixel density, typical threshold values being 1.0 and 0.7 for 25% pixel density)

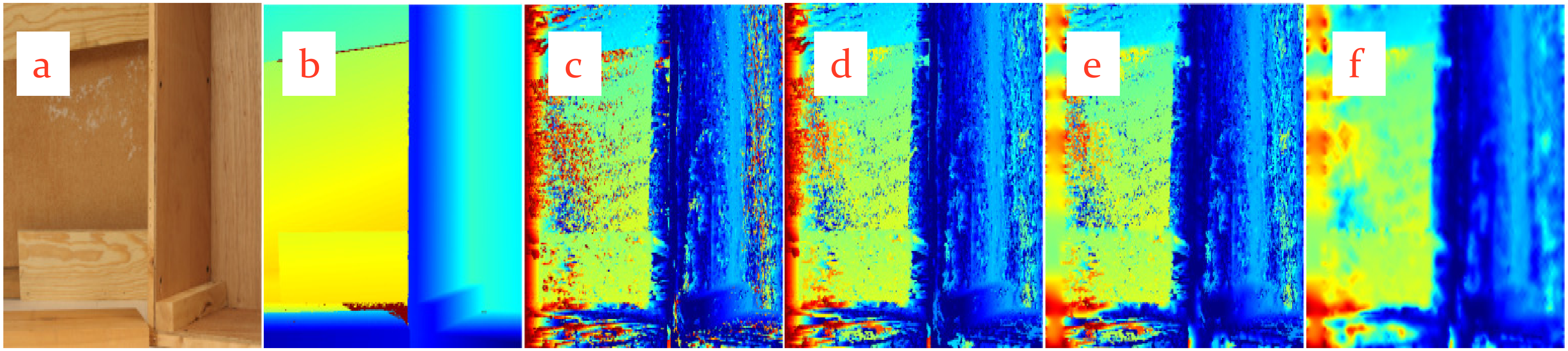
4. Valid pixels in the core-block flagged true in both maps retain original values. These pixels are well-behaved; satisfy topological smoothness constraints and are likely to belong to the same surface.

Iterative Hysteresis Filtering and Morphological Reconstruction

5. Pixels flagged true in only one map are categorized as hysteresis pixels - might belong to surfaces at a depth discontinuity with respect to the most prominent surface in core-block. Neighborhood pixels are ascertained (based on limits on depth values - set at 300 for the 16-bit depth pixel range and connected component analysis) in the macro-block map.
6. Hysteresis pixel along with the neighborhood pixels are added to the filtered map if neighborhood region is significant.
7. Above steps are iterated for the entire depth map until the number of pixels classified as noise pixels between iterations falls below threshold.
8. The final noise filtered depth map obtained by morphological reconstruction of the marker under the mask map, where the iterative hysteresis filtered depth map obtained in the previous step is used as the marker and the original depth map as the mask.

Depth Refinement

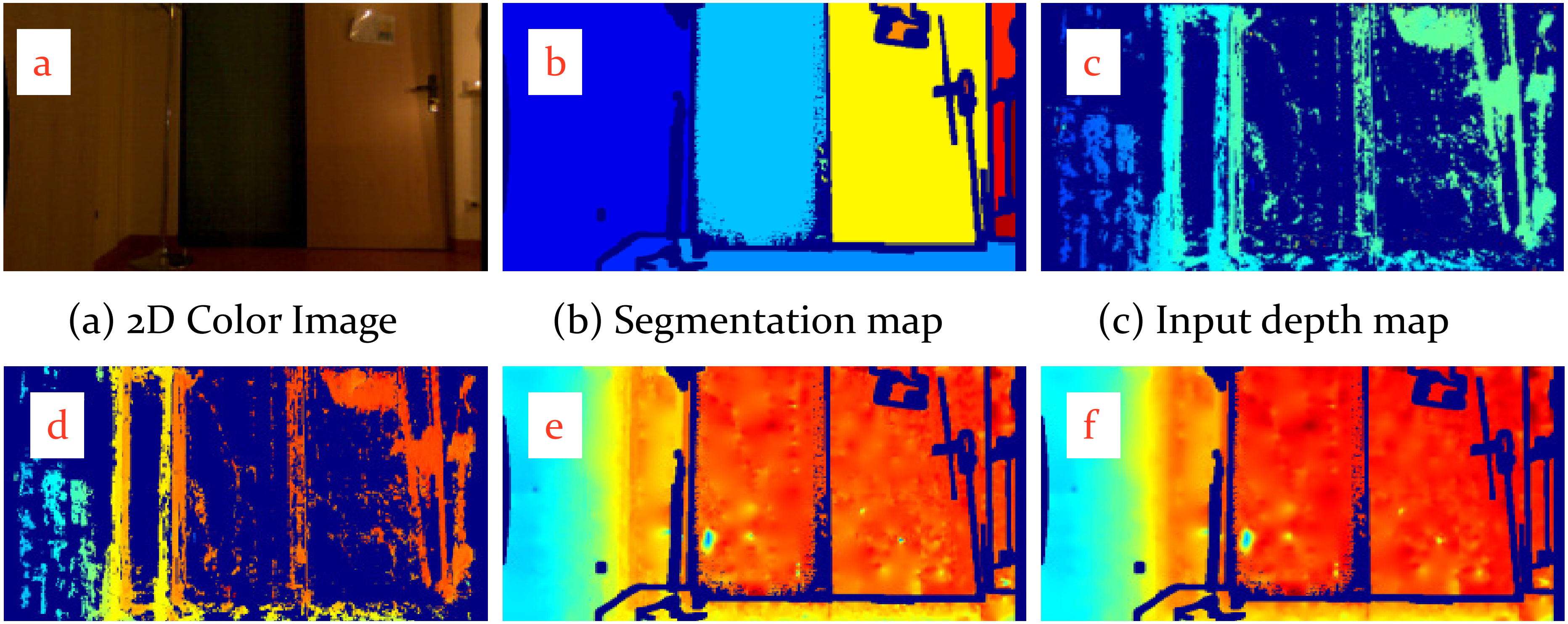
Image gradients guided anisotropic depth diffusion



(a) 2D Color image from Middlebury dataset (b) Ground truth (c) Noisy depth map from stereo-system (MSE = 2307) (d) De-noised depth map (MSE = 1752) (e) Depth map obtained by IBS based diffusion in estimation mode – note that sharp edges are preserved (MSE = 1507). (f) Depth map obtained by IBS based diffusion in filtering mode (MSE = 1184) - note that in the filtering mode noise removal is superior, though at the cost of lost edges

Depth Refinement

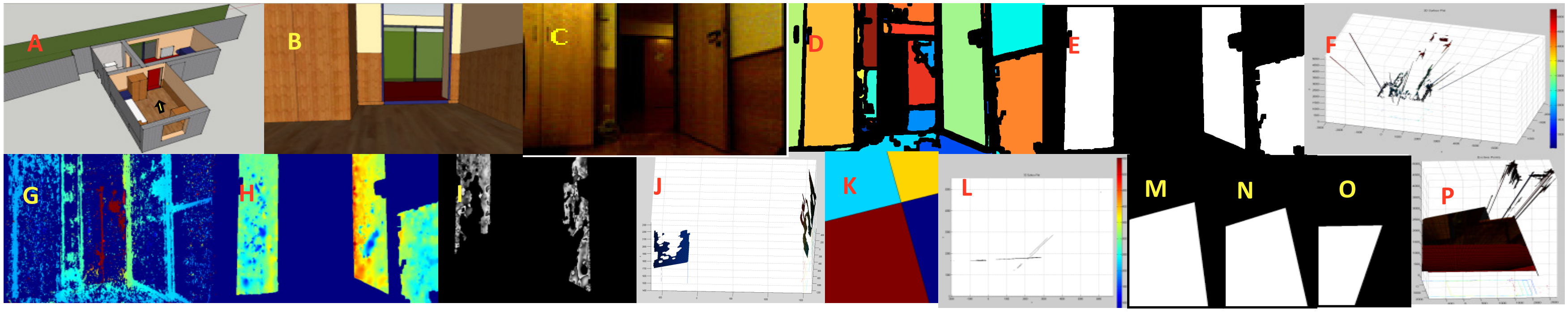
Piecewise isotropic IBS depth diffusion based on segmentation cues



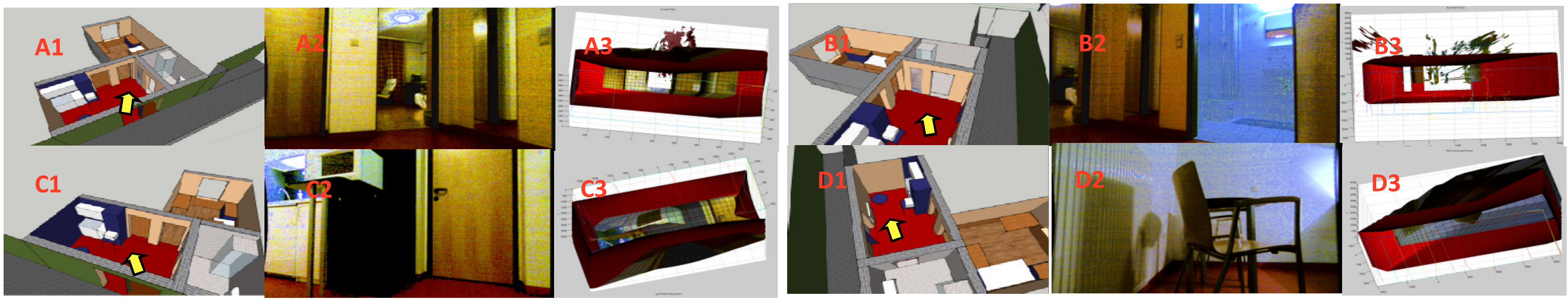
(a) 2D Color Image (b) Segmentation map (c) Input depth map

(d) De-noised depth map – note the change in dynamic range (e) Estimation mode results– note the smooth transition of depth values on the wall to the left and clear depth edge at the intersection of the side and the front wall surfaces. This edge is only visible in the diffused depth map and not in color image (f) Filtering mode results – note that the depth surfaces are smoother.

Results

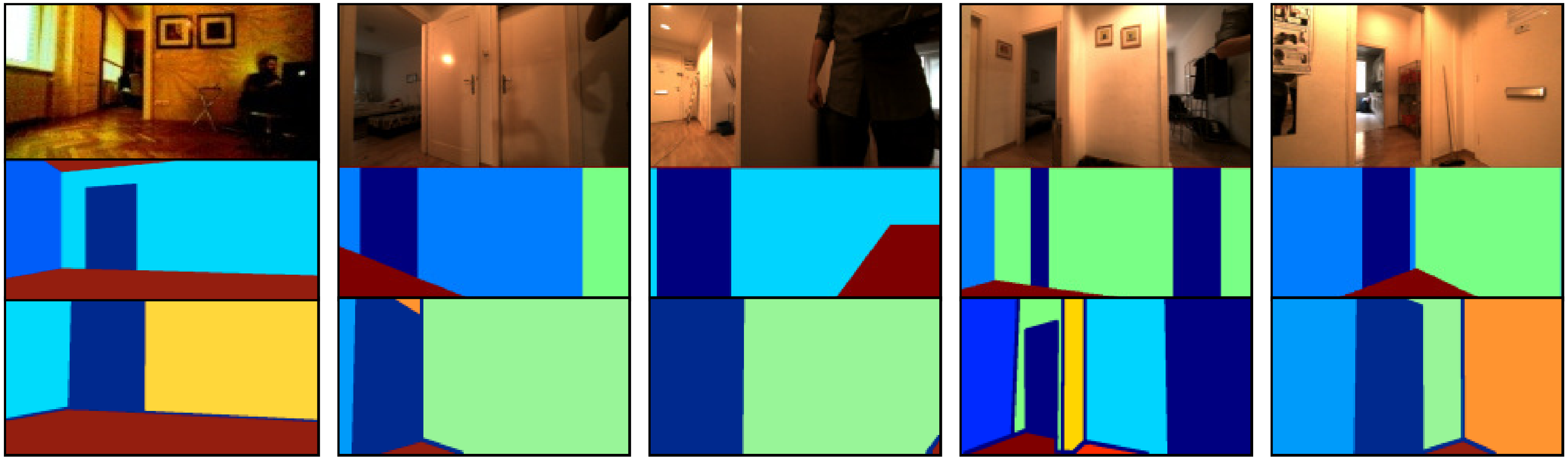


Complete Algorithmic Pipeline (A) 3D ground truth scene (yellow arrow shows position of camera) (B) Synthetic view from camera location – shows corner of a room with two intersecting wall planes and a doorway leading to a second room (C) 2D input image (D) Image segmentation output (E) Region selection output (of wall-like structures) (F) Input 3D point cloud (G) Input depth map (H) Depth diffusion output (I) Depth segmentation output (J) Surface fitting results (near top view) (K) Surface categorization & PCA based boundary detection (L) Comparative results with RANSAC based plane fitting on segmented point clouds (M) Ground truth room sector (N) Room sector results of our framework (mislabeled pixels = 8713) (O) Results for RANSAC (mislabeled pixels = 40125) (P) 3D Reconstruction (top view – note similarity to 5N) with exclave points through the doorway



3D Scene Reconstruction and Doorway Detection - Sets A and B are scenes with true doorways, while sets C an D are cluttered scenes with plenty of negative spaces but no doorways. Images indexed 1 present 3D ground truth of the scene, 2 are 2D input images and 3 are final reconstructions. Note that while true doorways have been estimated in sets A (at the intersection of two perpendicular wall faces) and B (2 doorways - a large one leading to the brighter room directly in front and a small one at the extreme left), the algorithm builds the 3D scene in sets C and D without detecting any doorways, demonstrating the robustness of the scheme to clutter.

Results



Results from test environment (Top to bottom) (a) Input scenes (b) 3D model reprojected on to the image plane (c) Ground truth. The percentage of mislabeled pixels were 5, 4, 17, 12 and 5 respectively.