Lifting 3D Manhattan Lines from a Single Image (ICCV13)

Manhattan Junction Catalogue for Spatial Reasoning of Indoor Scenes (CVPR13)

Srikumar Ramalingam et al.
Mitsubishi Electric Research Labs
Outline

- Overview
  - Manhattan-world assumption

- Single View Line Reconstruction (SVLR)
  - Lifting 3D Manhattan Lines from a Single Image – ICCV13

- Indoor Scenes
  - Manhattan Junction Catalogue for Spatial Reasoning of Indoor Scenes – CVPR13
  - Recovering the Spatial Layout of Cluttered Rooms – ICCV09

- Summary
Outline

• Overview
  ◦ Manhattan-world assumption

• Single View Line Reconstruction (SVLR)
  ◦ Lifting 3D Manhattan Lines from a Single Image – ICCV13

• Indoor Scenes
  ◦ Manhattan Junction Catalogue for Spatial Reasoning of Indoor Scenes – CVPR13
  ◦ Recovering the Spatial Layout of Cluttered Rooms – ICCV09

• Summary
Overview

• Single View Reconstruction
  ◦ Learning based - MAKE3D;

• Manhattan–world assumption
  ◦ Structured scenes;

• Based on lines and line junctions
  ◦ Instead of textures;
Manhattan-world assumption*

- [Coughlan and Yuille, 1999]
- “The scene is defined by four types of lines: random lines or lines parallel with one of the X, Y or Z axes.”

Applications
- Automatic Camera Calibration from a Single Manhattan Image
- Bayesian algorithms for autonomous vision systems
  - The Manhattan algorithm is used to guide an autonomous robot vehicle, using an artificial retina.
- Extracting 3D information from single images
  - Exploiting the Manhattan assumption to obtain 3D reconstructions from a single image.
  - Extracting ambiguous features such as image depth and identify objects.
  - Generating views of interior rooms
  - Automated model generator of building interiors

* from Wikipedia
Outline

• Overview
  ◦ Manhattan-world assumption

• Single View Line Reconstruction (SVLR)
  ◦ Lifting 3D Manhattan Lines from a Single Image – ICCV13

• Indoor Scenes
  ◦ Manhattan Junction Catalogue for Spatial Reasoning of Indoor Scenes – CVPR13
  ◦ Recovering the Spatial Layout of Cluttered Rooms – ICCV09

• Summary
SVLR - overview

- Single View Reconstruction based on lines;
- 3D intersections informative;
- Challenges:
  - Occluding edges producing several false intersections;
  - Detected lines are often broken and cropped;
SVLR - contributions

- A graph is built that models connectivity constraints between nearby lines in real images;
- Linear programming (LP) is used to lift line segments in 3D space using Manhattan-world assumption;
- Junction features are used to design the penalty terms in the LP;
- Computationally efficient;
• **Constraint graph**
  
  ◦ **Connectivity constraints**
    
    • Shortest distance between two line segments;
    
    • Two types of connectivity:
      
      • Two orthogonal lines – intersection;
      
      • Two collinear lines – incidence;
  
  ◦ **Graph:**
    
    • Vertices: image lines;
    
    • Edge: possible intersections or incidences;
SVLR - step II

- **Linear Programming**
  - One unknown parameter for each vertex - depth of one endpoint;
  - Slack variable $S_{ij}$: the distance between the two 3D lines.

\[
\begin{align*}
\min_{\lambda_i} \quad & \sum_{(i,j) \in \mathcal{E}} (||s_{ij}||_0) \\
\text{s.t.} \quad & |\lambda_id_{ia} - \lambda_jd_{ja}| \leq s_{ij}, a \in \{x, y, z\} \setminus \{\mathcal{D}_i, \mathcal{D}_j\} \\
& \lambda_i \geq 1, \ i \in \mathcal{V}. \\
& \lambda_i \geq 1, \ i \in \mathcal{V}.
\end{align*}
\]
Junctions

- Junction features for penalty terms;
- Different junctions:
  - L and X junctions occur on planar surfaces.
  - T junctions occur on both planar surfaces and occluding boundaries.
  - Y and W junctions are common on convex and concave edges.

Figure 7. We show the prominent junctions detected in two outdoor images. First column displays L junctions. The second displays T junctions.
SVLR - experiments
SVLR - problems
Outline

• Overview
  ◦ Manhattan-world assumption

• Single View Line Reconstruction (SVLR)
  ◦ Lifting 3D Manhattan Lines from a Single Image – ICCV13

• Indoor Scenes
  ◦ Manhattan Junction Catalogue for Spatial Reasoning of Indoor Scenes – CVPR13
  ◦ Recovering the Spatial Layout of Cluttered Rooms – ICCV09

• Summary
Indoor Scenes - overview

- Recovering spatial layouts for indoor scenes;
  - Manhattan world;
  - 3D box + surface labels of pixels (ICCV09);
- Junction detection;
  - Challenging due to missing and spurious lines;
  - A voting scheme to detect and classify junctions in real images;
Indoor Scenes - contributions

- Exploit Manhattan junctions for spatial understanding of indoor scenes;
- Present an efficient voting-based method to detect the junctions;
- Show a CRF formulation to incorporate junction features for the layout estimation problem;
- Demonstrate state-of-the-art performance for the layout estimation problem on standard datasets;
Indoor Scenes - review

- Get candidate box layouts
  - Evenly spaced sampling;
- Select the best layout
  - Learning to rank box layouts with structured outputs;
Indoor Scenes - learning

- Training images $X$ and their layouts $Y$
  \[ \{x_1, x_2, \ldots, x_n\} \in X \quad \{y_1, y_2, \ldots, y_n\} \in Y \quad y = \{F_1, F_2, F_3, F_4, F_5\}. \]

- For test image $x$, choose the correct layout
  \[ y^* = \arg \max_y f(x, y; w) \]

\[ f(x, y) = w^T \psi(x, y), \text{ where } \psi(x, y) \text{ is a vector of features.} \]
The mapping $f$ is learned discriminatively by solving
\[
\min_w, \xi \left( \frac{1}{2} \|w\|^2 + C \sum_i \xi_i \right) \\
\text{s.t. } \xi_i \geq 0 \ \forall i, \quad \text{and} \\
w^T \psi(x, y_i) - w^T \psi(x, y) \geq \Delta(y_i, y) - \xi_i, \\
\forall i, \forall y \in Y / y_i
\]
Indoor Scenes - functions

- \( \Delta \): loss functions quantifying the deviation between two layouts;

\[
\begin{align*}
\Delta(y_i, y) &= \Delta_t(y_i, y) + \Delta_c(y_i, y) + \Delta_p(y_i, y) \\
\Delta_t(y_i, y) &= \sum_{k \in [1,5]} \delta(F_{ik}, F_k) \\
\Delta_c(y_i, y) &= \sum_{k \in [1,5]} \| c_{ik} - c_k \|^2 \\
\Delta_p(y_i, y) &= \sum_{k \in [1,5]} (1 - \frac{\text{Area}(F_{ik} \cap F_k)}{\text{Area}(F_{ik} \cup F_k)})
\end{align*}
\]

- \( \Psi \): the set of features extracted for image layout pair \((x_i, y)\);
  - For each face \( F_k \), the unweighted line membership feature \( f_l \) is:

\[
f_l(F_k) = \frac{\sum_{l_j \in C_k} |l_j|}{\sum_{l_j \in L_k} |l_j|}
\]

  - \( C_k \): the set of lines belonging to the two vanishing points for face \( F_k \);
  - \( L_k \): the set of lines in \( F_k \);

- To summarize: \( F_k \) is characterized by two VPs. Lines inside \( F_k \) should have the same VPs as \( F_k \).
  - Occlusion?
Indoor Scenes - review

- Get candidate box layouts
  - Evenly spaced sampling;
  - Junction-based sampling;

- Select the best layout
  - Learning to rank box layouts with structured outputs;
  - Energy function based on a CRF model;

- Estimate surface labels
  - Superpixel segmentation;
  - Boosted decision tree classifier;

- Re-estimate
Indoor Scenes - step I

- Voting-based junction detection
  - Two-stage algorithm
    - Vote for 6 accumulator arrays using line segments along the vanishing points (weighted by length);
    - Detect different types of junctions by applying a product operation to the contents of the 6 accumulator arrays;
      - S: 6 directions; A: subset of S, i.e. a junction cfg.;

\[ f(p, A) = \prod_{i \in A} V_i(p) \prod_{j \in S \setminus A} \delta(V_j(p)) \]

Figure 4. (a) There are 6 possible directions at a point p based on the three vanishing points. (b) By choosing different combinations of these directions, we can generate junctions of types L, T, Y, W, X and K.
Indoor Scenes - step II

- Junction-based sampling

Figure 5. Illustration of junction-based sampling of layouts. (a) By sampling two horizontal rays passing through \( v_{px} \) and two vertical rays passing through \( v_{py} \), we can generate a box layout where the five faces correspond to left wall, ceiling, middle wall, floor and right wall. (b) The image is divided into four quadrants based on the positions of the three vanishing points. Each corner of the room, if visible in the image, can only appear in one of the quadrants and it should belong to a specific subtype of \( Y \) junctions as shown. In each of these quadrants we store the detected \( Y \) junctions. (c) Given a regular sampling, we identify top scoring \( Y \) junctions in the cone spanned by two consecutive rays. These high scoring \( Y \) junctions are used to generate a new set of rays to sample layouts.
Indoor Scenes - step III

- CRF model

Figure 6. Our CRF model. (a) Given the rays used for sampling the layouts, we divide the image space into a set of polygons. Each polygon corresponds to a node $x_i$ in the CRF. Each node can take 5 labels \{Left, Middle, Right, Floor, Ceiling\} that correspond to the five faces of the room. (b) Two nodes $x_i$ and $x_j$ are adjacent if they share a line segment $pq$. Our pairwise potentials in the CRF can be computed based on the presence of a line segment in the image that coincides with $pq$ or specific junctions detected at a point $r$ on the line segment $pq$. (c) Corners of a true layout usually coincides with specific Y junctions. If $t$ is detected as a Y junction, we would like to incorporate this prior in the form of a triple clique involving the incident nodes $x_i$, $x_j$ and $x_k$. 
Indoor Scenes - experiments
Outline

• Overview
  ◦ Manhattan-world assumption

• Single View Line Reconstruction (SVLR)
  ◦ Lifting 3D Manhattan Lines from a Single Image – ICCV13

• Indoor Scenes
  ◦ Manhattan Junction Catalogue for Spatial Reasoning of Indoor Scenes – CVPR13
  ◦ Recovering the Spatial Layout of Cluttered Rooms – ICCV09

• Summary
Summary

- Line-based method
  - Structures;
  - Hidden structures;

- Junctions
  - Connectivity, junction types

- Single view vs multiple view