Estimating Generic 3D Room Structures from 2D Annotations

Denys Rozumnyi Stefan Popov Kevis-Kokitsi Maninis Matthias Nießner Vittorio Ferrari

ETH zürich Google Research TUM NEURAL INFORMATION PROCESSING SYSTEMS

Introduction

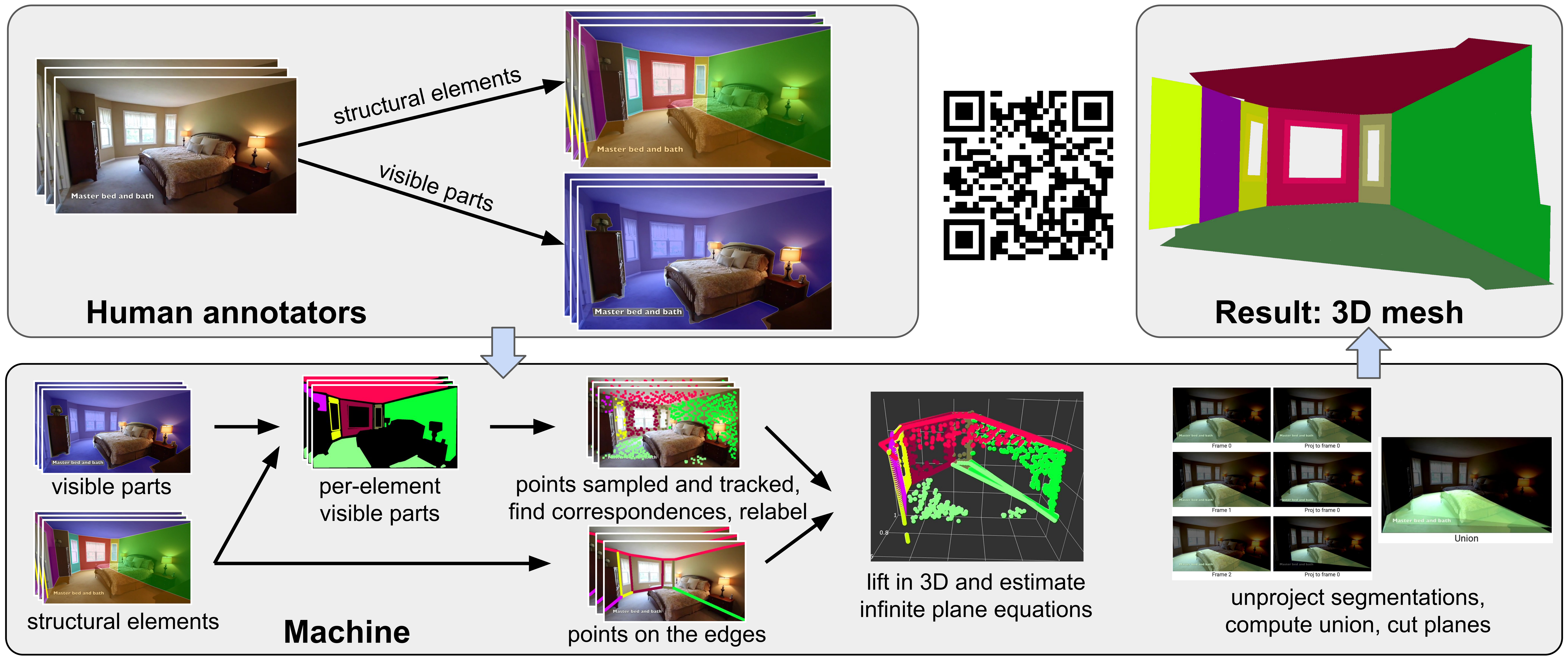
Goal: Create 3D room layouts from RGB video (no depth) → commonly available

Method: create 3D room layouts only *from 2D annotations* → easy for humans!

- Few (real) prior datasets, all requiring special acquisition devices (RGB-D, pano)
- The dataset is released here: https://github.com/google-research/cad-estate

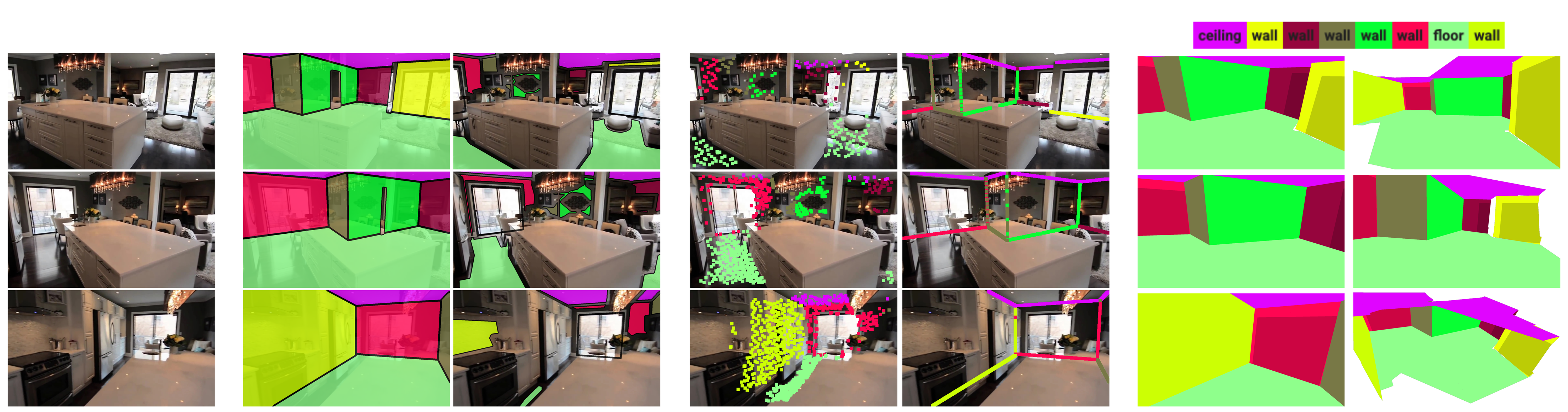
Pipeline

- Inputs are familiar 2D segmentation
- Each frame is annotated *independently,* without any correspondences



Method

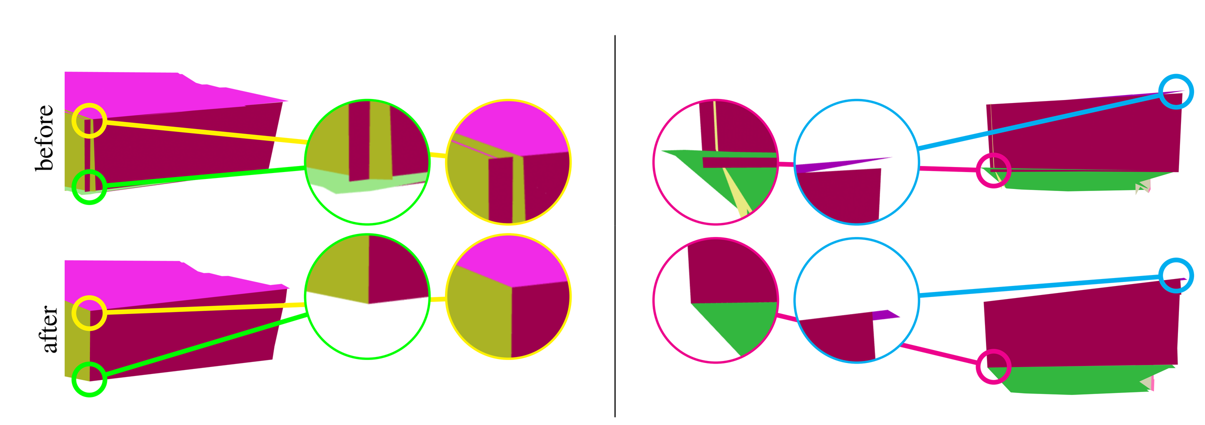
Given an input video and manual 2D annotations of structural elements and their visible parts, we combine **point tracks fitting**, **edge matching**, and **perpendicularity** constraints to generate a 3D room layout.



Inputs Structure Visible Tracked points Edges Final layout

Spatial extent refinement, before (top) and after (bottom).

We cut hanging walls extending outside the room boundary, and fill in the holes between neighboring planes (blue).



Evaluation

- Low depth errors *and* very high IoU values → high quality reconstructions
- Automatic quality control: reject reconstructions with IoU < 0.8 → IoU and depth worse when turned off → works well

	Runs	RE10k	ScanNet [11]	
		IoU↑	IoU↑	$\epsilon\downarrow$
Ours (full method)	100	**0.89**	**0.90**	**0.22**
Ours (no quality control)	100	0.83	0.85	0.30
Ours (no quality control)	30	0.81	0.84	0.33
Ours (no quality control)	1	0.72	0.79	0.36

- Run method many times and select automatically based on IoU → indirectly minimize depth error → good to have many runs
- We train and evaluate a baseline method [31] that performs at the state-of-the-art on the existing datasets, with a low error around 6% − 7%.
- Instead, it performs much worse on our dataset (26%), demonstrating it offers a harder challenge than the previous ones

	LSUN dataset [66]	Hedau dataset [19]	Our dataset
	Pixel Error (%)↓	Pixel Error (%)↓	Pixel Error (%)↓
Hedau *et al.* [19]	24.23	21.20	-
Mallya *et al.* [35]	16.71	12.83	-
DeLay [12]	10.63	9.73	-
CFILE [45]	7.57	8.67	-
Zhang *et al.* [68]	6.58	12.70	-
ST-PIO [71]	5.29	6.60	-
Lin *et al.* [31] (baseline)	6.25	7.41	26.3

Input annotations and final reconstructions