Free your Camera: 3D Indoor Scene Understanding from Arbitrary Camera Motion Supplementary Material

Axel Furlan[1]
furlan@disco.unimib.it
Stephen David Miller[2]
sdmiller@stanford.edu
Domenico G. Sorrenti[1]
sorrenti@disco.unimib.it
Li Fei-Fei[2]
feifeili@stanford.edu
Silvio Savarese[3]
silvio@umich.edu

[1] Computer Science Department
University of Milano - Bicocca
Milano, Italy

[2] Computer Science Department
Stanford University
Stanford, CA, USA

[3] EECS Building
University of Michigan
Ann Arbor, MI, USA

1 Failure and success cases analysis

Here we briefly discuss the failure and success cases on the proposed dataset. In Table 1 we present the classification accuracy for each sequence in the dataset achieved by the compared methods, while in Figures 1, 2 and 3 we show the corresponding visual results. Figures are organized so that columns represents sequences (see the captions of the Figures), while rows are encoded with the following order:

- First frame of the sequence.
- Reconstruction result by [2].
- Reconstruction result by [1].
- Reconstruction result by the proposed method fed with VSLAM (projection).
- Reconstruction result by the proposed method fed with VSLAM (3D model).
- Reconstruction result by the proposed method fed with VSfM (projection).
- Reconstruction result by the proposed method fed with VSfM (3D model).

There are two main types of scenarios where the proposed method achieves a poorer reconstruction result:

- Scenarios in which the observer do not explore enough the scene (*e.g.* Room 1, Room 5 and Lounge 1, where the right walls are framed for a very short period)
- Scenarios in which the ceiling is very dark and featureless. For example, in the sequence Room 1, the proposed method is perfectly able to recover the non Manhattan

Sequence/Method	Baseline	[2]	[1]	Our + VSLAM	Our + VSfM
Corridor	66.92	61.17	73.69	**81.96**	68.68
Entrance 1	74.36	88.04	89.75	89.43	**93.72**
Entrance 2	69.15	67.55	64.95	**97.24**	94.06
Lounge 1	71.21	51.76	**86.28**	81.14	83.31
Lounge 2	57.56	25.11	68.74	**88.82**	31.28
Room 1	65.32	60.23	**76.17**	70.08	63.75
Room 2	83.27	74.54	82.69	**97.03**	91.88
Room 3	71.48	64.68	55.13	93.66	**95.52**
Room 4	67.26	33.20	54.98	**84.79**	59.34
Room 5	79.83	66.57	**83.52**	78.25	77.88
Average	70.64	59.29	73.59	**86.24**	75.94

Table 1: Classification accuracy on the proposed dataset. The scores of all the compared methods are shown for each sequence in the dataset.

geometry of the left side of the room (unlike [2] and [1]), but fails in estimating the correct ceiling height, resulting in a poorer classification accuracy wrt to [1].

On the positive edge, there are also two main scenarios where the proposed method significantly outperforms state-of-the-art methods:

- Scenarios in which only a small portion of the scene falls within the field of view of the camera due to the small dimensions of the scene itself (the observer is pretty close to the walls) or to long focal length (pretty common on smartphones). A good example of this scenario is represented by the sequence Room 3, where [2] and [1] totally fail in reconstructing the correct scene layout, while the proposed method achieves very good performances.
- Scenes that cannot be represented with a box layout (*e.g.* Room 4) or that violate the Manhattan world assumption (*e.g.* Room 1).

Finally, a short note on the failure of the proposed method fed with the VSfM reconstruction in the sequence Lounge 2. Repetitive patterns like chessboards or some modern art drawings can lead the VSfM approach to generate a confuse cloud of 3D points around the patterns (see the 3D reconstruction in Figure 2 left column). In Lounge 2, a modern art drawing is present in the second half of the sequence and, in this case, led to the failure of the whole reconstruction. Please note that, on the same sequence, the proposed method fed with the VSLAM reconstruction achieved good results.

References

[1] Varsha Hedau, Derek Hoiem, and David Forsyth. Recovering the spatial layout of cluttered rooms. In *ICCV*, 2009.

[2] Derek Hoiem, Alexei A. Efros, and Martial Hebert. Recovering surface layout from an image. *IJCV*, 75(1), 2007.

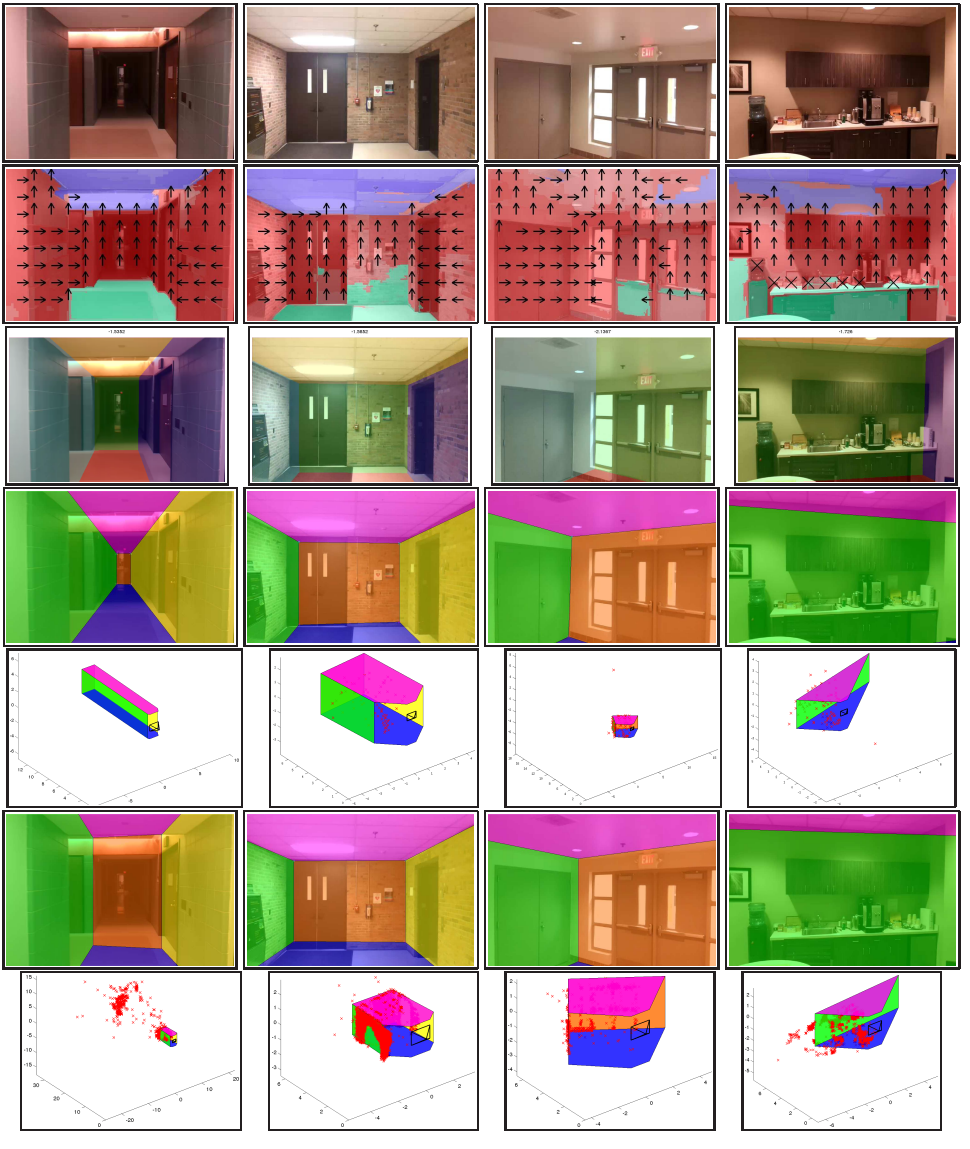

Figure 1: Sequence names: Corridor, Entrance 1, Entrance 2, Lounge 1.

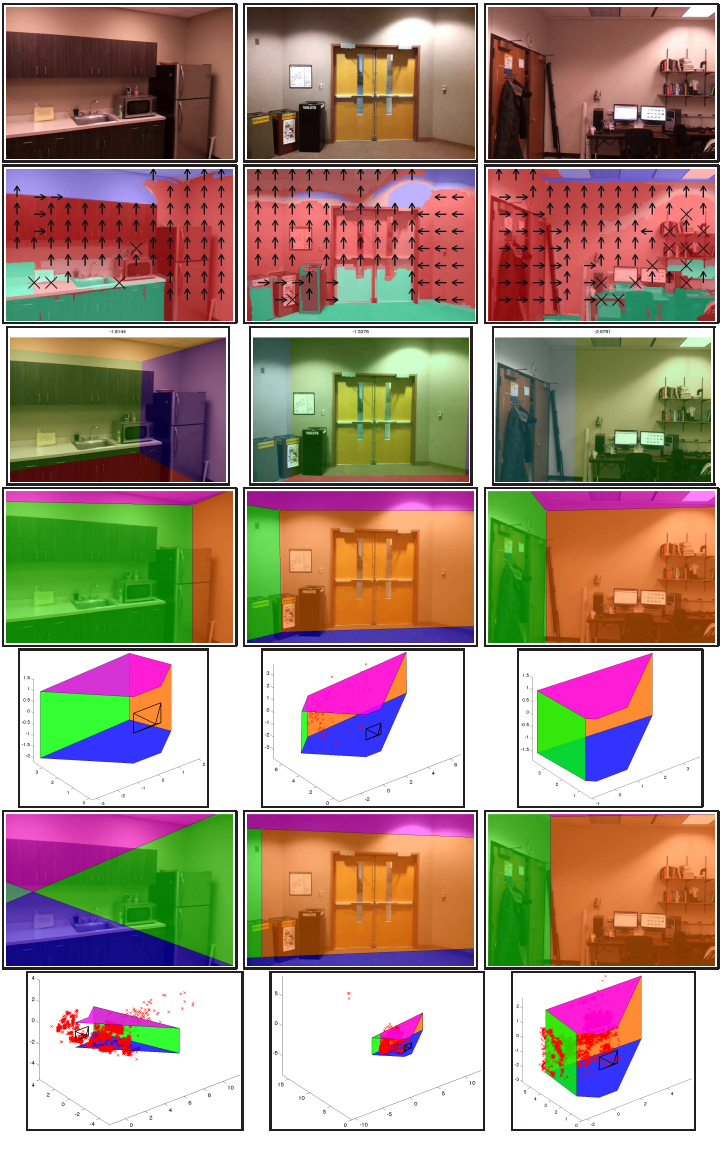

Figure 2: Sequence names: Lounge 2, Room 1, Room 2.

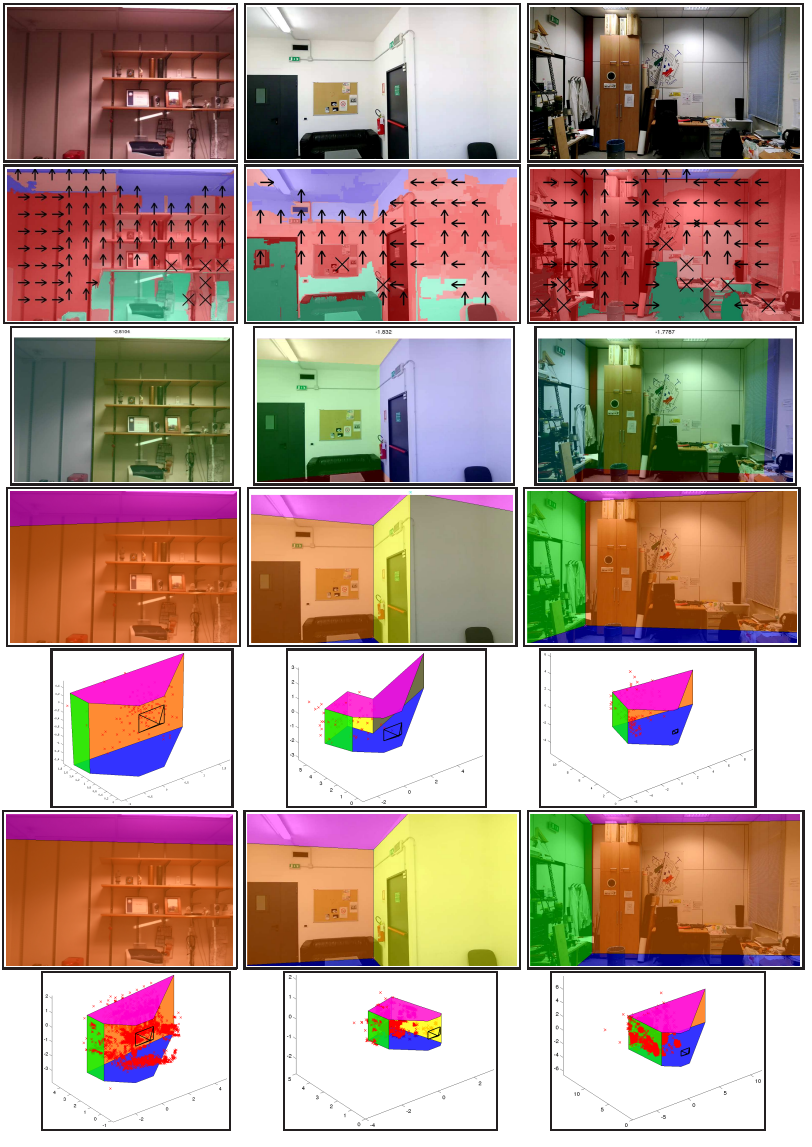

Figure 3: Sequence names: Room 3, Room 4, Room 5.