7P-03

直方体を用いた単眼画像からの室内シーン再構成に関する研究

仙田　朋也†　上田　芳弘†　坂本　一磨†
公立小松大学　　生産システム科学部†

１　はじめに

近年，深層学習の 3D 分野への利用に多く関心が寄せられている．また，自動運転や VR，AR などの 3D を扱った技術の発展は著しく，普及に向けて多くの研究が行われている．さらに毎年開催されるコンピュータビジョンに関する学会である CVPR(Conference on Computer Vision and Pattern Reconstruct)でも 3D を扱った研究が多くみられ，ボクセル生成，点群生成，メッシュ生成，3D 物体認識など多岐に渡る．その中でも本研究では，画像を用いた 3D 物体認識に注目する．画像を用いる際，単眼画像か複眼画像を用いる方法があるが，今回は単眼画像から室内シーン理解と再構成に取り組む．室内シーンの理解と再構成とは，対象となる室内シーンのレイアウトすなわち，どこに何があるかを理解し，仮想空間上で現実の室内シーンを再現することである．この技術の活用例として期待されているが，不動産のアプリケーションであり，内見の際に AR で仮想家具を設置し，部屋選択の質を高めることがあげられる．しかし，単眼画像から 3D シーンを理解することは，人間にとっては容易であるが，コンピュータにとっては未だ困難な課題がある．本研究では，室内シーンを扱った既存研究をベースとして，様々な室内シーンへ対応するために直方体を用いた再構成の研究に取り組んだ．

２　既存研究

室内シーンの理解と再構成を扱った研究[1]では，図 1 に示すとおり，1 枚の画像から 3D シーンの再構成のために，屋内のレイアウト推定，3D 物体検出，並びにメッシュ生成を End-To-End で扱う学習手法を提案している．各タスクは，深層学習をベースとしており，２つのデータセット SUN RGB-D[2]と Pix3D[3]を学習に用いている．3 つのタスクを共同で学習させることによって，各タスクにおいて先行研究を上回る結果を出した. これによりシーン再構成において各タスクが影響を及ぼし合っていることを実証した.

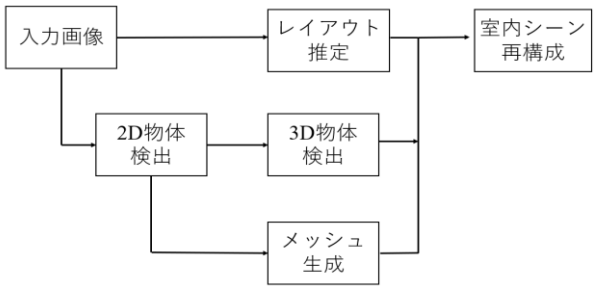

図 1　手法概要

２．１　レイアウト推定

1 枚の RGB 画像を入力として，カメラポーズとレイアウトのバウンディングボックスの向き，中心位置，大きさを推定する．レイアウトのバウンディングボックスとは部屋の形を直方体で簡易的に表現した物である.

２．２　3D 物体検出

2D 物体検出の結果を入力として，3D 空間上の物体の位置，向き，大きさを推定するタスクである．このタスクにより，2D のバウンディングボックスから 3D のバウンディングボックスを推定することが可能である．本研究での 3D のバウンディングボックスの座標は，既存研究[1]と同様にカメラ座標系上で扱う.

２．３　メッシュ生成

3D 物体検出と同様に 2D 物体検出の結果を入力として，メッシュを用いて物体の形状を再現するタスクである．本研究においても既存研究[1]と同様にあらかじめテンプレートとなる球を用意し，これを変形させることで物体形状の再現を行う.

２．４　室内シーン再構成

列挙した 3 つのタスクの出力を組み合わせていくことで室内シーン再構成を行う．生成されたメッシュオブジェクトに 3D 物体検出の結果を用いて大きさをあわせ，3D バウンディングボックス内に設置し，推定されたカメラポーズを用いてカメラ座標系からワールド座標系へ変換する．メッシュオブジェクトに加え，レイアウトバウ

Research for Indoor Scene Reconstruction from Monocular Camera Image using Cuboids

† Tomoya Senda，Yoshihiro Ueda and Kazuma Sakamoto
Faculty of Production Systems Engineering and Sciences, Komatsu University

ンディングボックスも 3D 空間内に設置する．以上の手順で 3 つのタスクの出力をもとに，室内シーンの理解と再構成を実現している．

3　課題

3D を扱った研究における課題として既存研究[1]だけでなく，学習に用いるデータの取得が困難である．そのため[1]ではメッシュ生成可能なオブジェクトが椅子やテーブル，本棚などデータセットに用意されている数個のカテゴリに限られている．また，同じカテゴリのオブジェクトにおいてもその形は多岐にわたるため，そのすべてをメッシュで完全に再現することは難しい．そのため生成可能なカテゴリに含まれていないオブジェクトは再構成することは不可能であり，生成可能なカテゴリにおいても現実ではその形状は異なるため元画像の形状を再現できない場合がある．以上の理由により，再構成できる室内シーンが限定されてしまうという課題がある．

4　提案手法

既存研究[1]においてメッシュ生成が難しい物体に対応するため，学習していないカテゴリに属する物体が検出された場合，直方体として 3D 空間上に適切に配置することを提案する．本手法では，図 2 に示す室内シーン画像を対象としている．以下では図 2 を例として提案手法の解説を行う．なお，図 1 中における 2D 物体検出の処理には YOLO v4[4]を用いる．YOLO を用いて得られた物体検出の結果を 3 D 物体検出とメッシュ生成のタスクの入力とする．

図 2 中にはテーブル，椅子，モニター，ハンガーラックなどが存在することがわかる．テーブルや椅子，モニターはメッシュ生成可能なカテゴリであるが，ハンガーラックは学習していないため，メッシュ生成が困難である．そこで，ハンガーラックを図 3 のような直方体で表現することでシーン再構成を改善する．方法として直方体を画像内のシーンにあわせて配置するために他 2 つのタスクの出力を用いる．そして，メッシュ生成以外のレイアウト検出と 3D 物体検出の出力をもとにハンガーラックの大きさ，位置，向きを求め図 3 中のように直方体を配置する．

5　今後の展望

メッシュ生成が難しい物体に対して直方体を適応することによる室内シーン表現に取り組んだ．提案する手法により，様々な室内シーンの再構成が可能だと考えられる．しかし，直方体を用いた場合の各タスクの学習を考慮していないため，既存手法において定義された共同学習時に用いる損失関数を適用することは困難である．そのため直方体を用いた上で共同学習を可能とする手法の提案が必要である．また，課題として前述した学習データの取得が難しいことに関しては，データの人間によるラベリングを必要としない自己教師あり学習が行える仕組みを作ることによって解決可能であると考える．

図 2　室内シーン画像例

図 3　直方体適用例

参考文献

[1]Yinyu, N. Xiaoguang, H. Shihui, G. Yujian, Z.Jian, C. and Jian, J. Z.. Total3Dunderstanging: Joint Layout, Object Pose and Mesh Reconstruction for Indoor Scenes from a Single Image. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, p. 55-64.

[2]Shuran, S. Samuel, P. L. and Jianxiong, X.. SUN RGB-D: A RGB-D Scene Understanding Benchmark Suite. I*n Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, p. 567–576.

[3]Xingyuan, S. Jiajun, W. Xiuming, Z. Zhoutong, Z. Chengkai, Z. Tianfan, X. Joshue, B. T. and Wiliam, T F.. Pix3D: Dataset and Methods for Single-Image 3D Shape Modeling. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, p. 2974–2983.

[4]Alexey, B. Chien, Y. and Hong-Yuan, M. L.. YOLOv4: Optimal Speed and Accuracy of Object Detection. *Computer Vision and Pattern Recognition*, 2020, p.1-17.