

# 遮蔽解析による奥行推定に基づく 2次元画像からの3次元仮想空間生成に関する研究

益本 大樹<sup>†1,a)</sup> 福元 伸也<sup>†1,b)</sup> 鹿嶋 雅之<sup>†1,c)</sup> 渡邊 睦<sup>†1,d)</sup>

**概要：**物体検出を用いて得られた領域に基づく、2次元画像の遮蔽解析と3次元シーンのモデリング手法を提案する。1枚の画像に写っているシーンに対して、より汎用的なモデリング手法の構築を目的とした。物体検出により、画像を要素ごとに分け、各要素に対して異なる処理を行うことが可能であることを利用した。その際に、物体検出の結果で遮蔽関係が存在することが確認された場合は、物体間の前後関係について輪郭画像を用いることで遮蔽解析を行い整合性をとることを目指す。異なる遮蔽関係を持つシーンに対する効果と、結果として得られる3Dモデルについて考察した。

**キーワード：**画像認識・理解, 3次元形状推定, 仮想現実

## A study on 3D virtual space generation from 2D images based on depth estimation by shielding analysis

DAIKI MASUMOTO<sup>†1,a)</sup> SHINYA FUKUMOTO<sup>†1,b)</sup> MASAYUKI KASHIMA<sup>†1,c)</sup> MUTSUMI WATANABE<sup>†1,d)</sup>

**Abstract:** We propose a shielding analysis and 3D scene modeling method for 2D images based on regions obtained using object detection. The objective is to construct a more general modeling method for scenes in a single image. We took advantage of the fact that the image can be divided into elements. In object detection. In the case of confirming the existence of a shielding relationship, a shielding analysis was performed using the image's outline to determine the front-back relationship between the objects, aiming for consistency. We discussed the effects on scenes with different shielding relationships and the resulting 3D models.

**Keywords:** Image recognition and understanding, 3D shape estimation, virtual reality

### 1. はじめに

近年、3DCGは映画やゲーム・アニメーションといった分野をはじめ、建築分野や自動車などに挙げられる設計業といった幅広い分野に活用されており、従来手法と比較して製作コストの削減や作業者の負担軽減に役立っている。また、仮想空間についてはVR機器の発展はもとより、人と人がコミュニケーションをとる新しい活動の場として注目されている。しかし、3DCGの扱う範囲が広がるほど、

これを扱う作業者の負担が増える。これら背景から、より汎用的で簡略的な3D空間生成の手法が必要になると考えられる。また、複数視点から生成される高精度なモデリング手法は多くある。しかし、安定してそれらのデータを集めることができない状況や実際に人が訪れることができない場所でのデータとしての単眼視点画像を用いることや、また、以前訪れた場所をはじめとする、昔の写真から3次元状況を復元するような、アプローチを行うことができる研究は、現状不十分であると言える。

本研究では、1枚の画像に写っているシーンに対してより汎用的なモデリング手法を構築すること。その際に、物体間の前後関係について、できるだけ現実の3次元空間に近いものを再現するため整合性を取ることを目的とする。

<sup>†1</sup> 現在、鹿児島大学大学院 理工学研究科  
Presently with The Graduate School of Science and Engineering, Kagoshima University  
a) k1160252@kadai.jp  
b) fukumoto@ibe.kagoshima-u.ac.jp  
c) kashima@ibe.kagoshima-u.ac.jp  
d) mutuyas256@gmail.com

## 2. 関連研究

3D モデルの生成に関して、斎藤らは1枚の画像から人の3次元形状を復元する手法を考案した [1]. 前身である PIFu は2次元画像から、人のピクセルに対応する3D オブジェクトの3次元位置を推定することにより人の3D モデルが生成される [2]. PIFuHD ではこれに加え、GAN 手法の一つである Pix2PixHD を使用し人の背面を高解像度で生成することで3D モデルの精度やメモリ問題を解決している. これを活用することにより1枚の画像から高精度な人の3D モデルの生成を可能としている. また、人以外に関しても1枚の画像から三角形メッシュの3次元形状を推論する手法がある. 単一 RGB 画像について、GCN (Graph Convolutional Networks) をベースとし、楕円形からの変形を学習する [3]. これらの研究は、1枚の画像から3次元化を行うことができるものの、対象が限定されていることや物体の背景にモデルの精度が影響される. また、山本らの研究では移動ロボットについて環境モデリングを行った [4]. その中で、実空間の知識を付与することでモデリングの精度を向上する空間的制約の考えを提唱した. 本研究では、物体検出を用いることで1枚の画像から複数の要素を対象にシーン全体を3次元復元することを目的とする. 画像中に表れる遮蔽関係において、前後関係について整合性をもった位置関係を再現することを目指す.

## 3. 遮蔽解析

1枚の画像から情報を取得する際に発生する遮蔽問題について、物体検出を行ったときに得られる情報から遮蔽解析を行う手法について述べる. 物体検出を行うと、検出結果は対象の物体を囲う矩形情報 (バウンディングボックス) として出力される. 本研究では、矩形領域が重なっていた際に、2つの物体は遮蔽関係にあるとし、どちらが前か後ろかを判定する. 重なっている領域に対してエッジ検出手法である Canny edge detector を用いて遮蔽部分を求める. 同時に、それぞれの対象についてどの位置が他の検出物体と重なっているのかを求める. 遮蔽部分について、チェーンコード化を行い0から8の数値で表現する. このエッジの連結成分について、全てのピクセルの連結する向きの合計を特徴量とする. また、本研究では物体検出の検出対象と対象の遮蔽のない輪郭画像を事前に与え手法となっている. 物体検出対象の輪郭画像における、重なっている位置の特徴量を同様に求め、遮蔽部分のものと比較を行うことでどちらが前にあるのかを判定する. 遮蔽解析手法を図1に示す.

### 3.1 物体検出と遮蔽領域取得

物体検出と遮蔽領域取得について述べる. 本研究では物

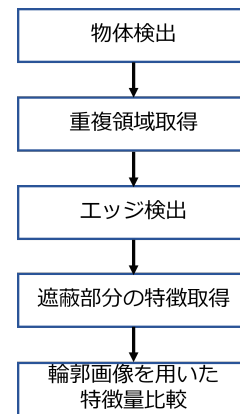


図1 遮蔽解析手法

体検出手法に YOLOv8 を使用する. 物体検出の検出結果から得られた領域について、検出物体の名称と左上の座標と右下の座標を保存する. その後、バウンディングボックスが重なっている部分が存在する場合はその領域の座標も取得する. また、遮蔽領域が各検出物体のバウンディングボックスに対してどの位置にあるのかを求める.

### 3.2 遮蔽部分の特徴取得

遮蔽解析には画像のエッジ情報を用いる. エッジ検出には Canny 法を用いた. また、本研究では遮蔽領域として得られたバウンディングボックスから繋がるエッジのみを抽出する. 物体間の遮蔽に関係のあるエッジのみを検出することで、物体のテクスチャによるエッジをはじめとする物体の輪郭に関係のないエッジを除外することができる. 本研究では、ここまでの処理で得られたエッジのうち、最長のエッジを2つの物体間の遮蔽部分として扱うこととする. その後、得られたエッジについて8近傍チェーンコードを作成する. 得られた符号をもとに、各方向の角度を合成していくことで  $-90 < x < 90$  の範囲の特徴量  $x$  を求める.

### 3.3 輪郭画像を用いた特徴量比較

本研究では、対象の物体が遮蔽されていない状態の輪郭画像を用いて、入力画像中の物体が遮蔽されているかどうかを判別する. 物体検出で得られた領域の内、遮蔽関係のある位置をもとに輪郭画像から同様の位置を求める. 求めた位置について向きの特徴量を同様に計算する. 2つの検出された物体の遮蔽されていない状態の特徴量と遮蔽領域における特徴量の、求められた3つの特徴量を比較する. 遮蔽領域の特徴量により近い物体が遮蔽されていないものとし、より前方にあるものと判断する.

## 4. 3D モデルの生成

入力画像から実空間での前後関係を反映した3次元仮想空間モデルの生成手法を述べる. 入力画像に対して物体検出を行い、検出結果ごとにバウンディングボックスで別画

像として切り出す。このとき、バウンディングボックスの遮蔽関係から、3次元空間での前後関係を求めた。物体検出の結果について、人と人以外に分け、人の場合はPIFuHDで3Dモデルの生成を行う。人以外は、検出結果に対応する3Dモデルを呼び出す。また、切り出された画像から得られた色をもとにモデルの配色を行う。3Dモデルを用意した後、画像から得られた前後関係と空間的制約をもとに配置を行う。3Dモデルの生成手法を図2に示す。

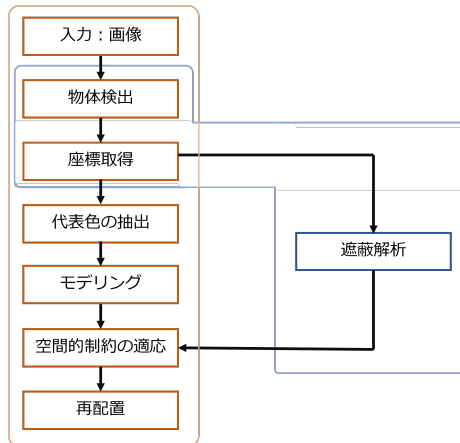


図2 3Dモデル生成手法

#### 4.1 物体検出と対象の画像切り出し

入力画像に対して、はじめに物体検出を行う。検出の結果に応じてモデリングするかどうかの判定、また、モデリングの前の画像処理及び遮蔽関係の有無を確認する。その後、検出結果に従って座標を取得する。この際、遮蔽解析で使用する検出対象の座標と重複している領域の取得も行う。また、対象の得られた座標について、3Dモデル色を決定するために使用するため画像の切り出しを行う。

#### 4.2 モデリングとモデル色選択

物体検出の結果の内、人に対してはPIFuHDを使用し、3Dモデリングを行う。また、人以外の対象については検出結果に対応する3Dモデルを呼び出す。このとき、代表色の選択で得られた色をもとに3Dモデルの色を変更する。代表色の選択はk-means法で行った。k-means法により各ピクセルの色をクラスタリングする。最大クラスタの中心部分を代表色として扱う。本研究では、入力する画像は物体領域で切り抜かれているため、対象物がある前景と背景に二分することができる。そのため、クラスタの数は2とした。2つのクラスタのうち、より要素数が多い最大クラスタの中心の色を3Dモデルの色とする。

#### 4.3 再配置

作成した3Dモデルの配置・表示は3DモデリングソフトウェアであるBlenderを使用する。出力結果について、

3D空間に配置した後、既に取得した物体間の前後関係及び、空間的制約をもとに移動を行う。本研究で適応した制約の例としては、モニターが宙に浮くことは無いためモニターの下に机を補完するといったものである。

### 5. 評価実験

実験には、本研究グループの研究室及び実験室で撮影された画像を使用した。また、モデリングの対象として人・椅子・モニター・机を対象とした。

#### 5.1 使用画像

本実験で使用した画像の例を図3と図4に示す。



図3 画像1



図4 画像2

#### 5.2 実験結果

4枚の画像について、遮蔽解析で前方にあるとされた物体の各関係を表1に示す。また、実際に存在する遮蔽と物体検出で検出できた遮蔽、遮蔽解析で正答できた遮蔽の数を表2に示す。また、画像1と画像2についてモデリング後の結果を図5～図10に示す。

表1 前方にあると判断された対象

対象画像	人と椅子	人とモニター	モニターとモニター
画像1	椅子	人	モニター (左側)
画像2	人	人	モニター (左側)
画像3	人	モニター	—
画像4	人	人	モニター (右側)

表2 遮蔽解析による検出結果

対象画像	遮蔽関係の総数	遮蔽関係の検出数	正答数
画像1	3	3	3
画像2	3	3	1
画像3	2	2	1
画像4	3	3	3

### 6. 考察

複数の遮蔽関係を持つ画像について遮蔽解析及び、1枚の画像からシーンのモデリングを行った。遮蔽解析においては、11の遮蔽関係のうち、8つの前後関係を判別することができた。また、人や椅子などのモデルが互いに干渉せずに行えていることが確認できる。画像2の前後判定の誤

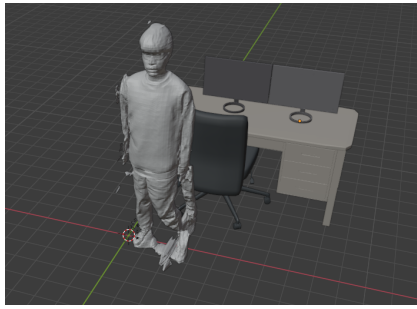


図 5 画像 1 の出力結果

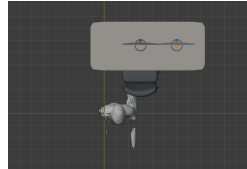


図 6 画像 1 の出力結果（左側面） 図 7 画像 1 の出力結果（上面）

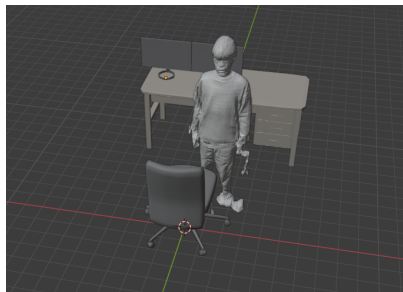


図 8 画像 2 の出力結果

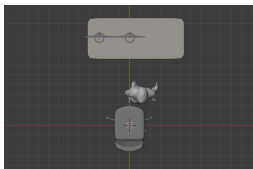
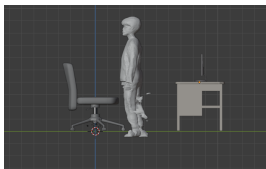


図 9 画像 2 の出力結果（左側面） 図 10 画像 2 の出力結果（上面）

りについて、図 11 のようにモニターとモニターにおいては、左側のモニターのベゼル部分のエッジが強く検出されており、エッジが曲がっている。そのため、左側にあるモニターが前方にあると判別された。また、人と椅子の遮蔽関係については図 12 に示すように、人と椅子で囲まれたエッジが遮蔽部分のエッジとされたため、正しく判定されなかったと考えられる。本手法は、遮蔽部分が適切にエッジとして検出でき、検出物体の輪郭の特徴を十分に反映している場合に有効であると考えられる。また、物体の回転や同じ対象の形の違いによって結果が左右されるため、一つの対象に対して複数の見方をすることや、輪郭の情報を増やすなど手法の改善が必要であると考えられる。今回の実験では、屋内環境を対象とした。汎用的なシステムを目指した際に、屋外環境では、光源の不確定さ、スケール感の違いなどが新たな問題となると考えられる。



図 11 遮蔽部分のエッジ 1

図 12 遮蔽部分のエッジ 2

## 7. おわりに

本研究では、単一画像からのモデリングとその配置による 3次元シーンの再現を最終的な目標とした。物体検出を活用した遮蔽解析を行うことで物体間の前後関係を判定して 3D モデルに反映することを試みた。人と椅子、モニターを対象とした実験では、あらかじめ物体の輪郭情報を与えることで、これらの対象の重なりについて前後の判別をさせることができた。物体間の重なり方の違いから、判別しやすい遮蔽関係としにくいものについて考察を進めることができた。また、出力された前後関係の結果を反映することで整合性の取れたシーンのモデリングを行うことが可能であると確認することができた。

## 参考文献

- [1] Shunsuke Saito, Tomas Simon, Jason Saragih, Hanbyul Joo, "PIFuHD: Multi-Level Pixel-Aligned Implicit Function for High-Resolution 3D Human Digitization", CVPR2020(Oral Presentation), 2020.
- [2] Shunsuke Saito, Zeng Huang, Ryota Natsume, Shigeo Morishima, Angjoo Kanazawa, Hao Li, "PIFu: Pixel-Aligned Implicit Function for High-Resolution Clothed Human Digitization", ICCV2019, pp.2304-2314, 2019.
- [3] Nanyang Wang, Yinda Zhang, Zhuwen Li, Yanwei Fu, Wei Liu, Yu-Gang Jiang, "Pixel2Mesh: Generating 3D Mesh Models from Single RGB Images", ECCV2018, 2018.
- [4] 山本泰平, 福元伸也, 鹿嶋雅之, 渡邊睦, "物体認識に基づく経時的環境モデリングに関する研究", ロボティクス・メカトロニクス講演会講演概要集 2020, pp.2P1-K13, 2020.
- [5] 加藤 大晴, 牛久 祥孝, 原田 達也, "Neural 3D Mesh Renderer", 2019 年度人工知能学会全国大会 (第 33 回), 2019.
- [6] Tinghui Zhou, Matthew Brown, Noah Snavely, David Lowe, "Unsupervised Learning of Depth and Ego-Motion from Video", CVPR2017 (Oral Presentation), 2017.
- [7] Vincent Casser, Soeren Pirk, Reza Mahjourian, Anelia Angelova, "Depth Prediction Without the Sensors: Leveraging Structure for Unsupervised Learning from Monocular Videos", AAAI2019, 2019.
- [8] René Ranftl, Katrin Lasinger, David Hafner, Konrad Schindler, Vladlen Koltun, "Towards Robust Monocular Depth Estimation: Mixing Datasets for Zero-shot Cross-dataset Transfer", IEEE 2020, 2020.
- [9] Derek Hoiem Andrew N. Stein Alexei A. Efros Martial Hebert, "Recovering Occlusion Boundaries from a Single Image", 2007 IEEE 11th International Conference on Computer Vision, 2007.
- [10] YOLOv8, "https://ultralytics.com/yolov8".