

# 1枚のイラスト画像からの3Dモデル再構成

理学専攻 情報科学コース 2140675 SHEN QIAN (指導教員: 伊藤 貴之)

## 1 はじめに

画像からの3Dモデル復元は、コンピュータビジョン、医用画像処理、バーチャルリアリティなど多くの領域での課題となる。画像からの3Dモデル再構成は、コンピュータビジョンにおける主要なタスクの一つであり、単一画像または複数画像から対応する3次元構造の生成を研究することを目的としている[1]。

人間は明暗や影などの光学的特性から物体の奥行きを推察する能力を有する。一方で、1枚の画像から計算によって深さ情報を推定するのは容易ではなく、相応しい結果が得られない事例が多数生じている。この問題を解決するために多くの研究がなされている。例として、写真撮影に用いたカメラのパラメータを解析することで、深さ情報の復元を補助する手法がある。一方で、写真に限らずより多様な入力画像からの3次元再構成の可能性も興味深い課題である。その一環として本研究では、手描き画像を入力として、明暗や影などの光学特性、投影や遠近感が現実に即していないようなイラストからも、物体を認識して3Dモデルを復元することが可能か否かを検討する。

## 2 先行研究

単一画像からの3Dモデルの再構成に関して多くの研究が発表されている。例としてWangら[2]は、単一カラー画像から三角形の3D形状を生成するエンドツーエンドの深層学習アーキテクチャを提案した。このネットワークは、グラフベースの畳み込みニューラルネットワークで3Dメッシュを表現し、入力画像から抽出した知覚的特徴を用いて楕円体を徐々に変形させることで、適切な形状を生成する。

## 3 提案手法

### 3.1 データセット構造

#### 3.1.1 3Dモデル可視化

イラスト画像にはカメラ情報もなく、光と影の正確な描画も保証できない。そこでイラスト画像から3Dモデルを再構成するには、既存の3Dモデルデータベースの参照が必要であると考えた。本研究では、プリンストン大学の椅子点群889個を収録した「ModelNet10」をデータベースの適用事例として選択した。さらに本研究では、椅子だけで6778個の3Dモデルを有するShapeNetCoreというデータセットを併用した。この2つのデータベースを組み合わせると、7667個のモデルからなる3Dモデルデータベースを作成した。

3Dモデルとスケッチとの類似性比較を実装するた

めに、まず3Dモデルをレンダリングする必要がある。本研究ではBlenderを適用してレンダリング画像を生成した。ここで、本研究で適用した機械学習モデルでは、10枚程度の画像の集合を同一オブジェクトとして学習できることから、高さ方向の角度を75°に固定し、30°間隔で視点を回転移動して1枚のスクリーンショットを生成した。図1に例を示す。



図1: 3Dモデルの可視化

#### 3.1.2 Canny演算子によるエッジ抽出

本研究では、比較する際に入力画像をできるだけデータベース画像に近づけるために、画像加工ソフトウェアのスケッチ風フィルタを用いたスクリーンショットの加工を試みている。しかし結果として、同じ画像でも処理前と処理後では同じ物体として認識することはできなかった。代替手段として本研究では、入力画像とデータセットの全画像に対して前処理として、Canny演算子によるエッジ抽出を適用している。



図2: 入力画像とデータセットのエッジ抽出例

### 3.2 機械学習からモデルの選択

比較画像の類似性を考慮したツインニューラルネットワークによって画像の類似性を比較する手法[3]が公開されている。この手法では2つのニューラルネットワークで別々に画像から特徴を抽出して比較する。しかし、この2つの画像に対して、異なるニューラルネットワークを独立に使用して特徴を抽出すると、異なる空間に特徴を抽出することになってしまう。この問題を回避する手段として、単一のメインニューラルネットワークで同時に特徴を抽出することも考えられるが、本研究では、ツインニューラルネットワークにおける一方の入力情報を他方のニューラルネットワークでの空間にマッピングするという手段をとる。ここで2つの入力の特徴性を Loss の計算によって評

価する。

この過程で特徴を抽出するためのネットワークとしては、本研究では Simonyan らが提案した VGG16[4] を適用する。この方法は非常に小さな(3×3)畳み込みフィルタを持つアーキテクチャを使用して深さを 16～19 層程度に増やすことで、先行技術と比べて大きな改善を達成している。

### 3.3 モデルの分割・合併

本研究では、半教師あり手法による 3 次元形状セグメンテーションを適用する[5]。機械学習ベースの 3 次元分割技術の問題点として、ラベル付き 3 次元データセットが少ないという点がある。そこでこの手法は、少数のラベル付き 3 次元形状と大量のラベルなし 3 次元データから、3 次元分割のための効果的な半教師あり学習を実現する。

## 4 実行結果・考察

本研究では、前述した 7667 個の椅子モデル画像を含むデータセットを学習のための訓練データセットとして用いた。訓練過程は十分に収束しており、学習結果はある程度信用できると考えられる。学習終了後、入力画像とデータセット内の全画像の類似度を比較し、類似度が最も高い画像がモデルに対応していることを確認した。

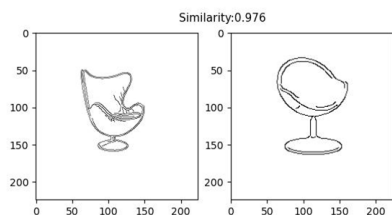


図 3：入力画像とデータセット内の画像を比較した類似度結果



図 4：異なる入力画像と、それに対応する最も類似度の高いモデル

続いて、ある程度類似していると判断された 3D モデルを、次のステップの学習セットとして選択する。モデルの 3 次元分割が完了したら、各モデルの中から類似度の高いパーツを選んで合体することで、3D モデルの再構成を完成する。図 5 にその結果を示す。



図 5：3Dモデル分割・合体の結果

## 5 まとめと今後の課題

本報告では、3D モデルデータセットのレンダリング画像を学習し、入力画像と比較して類似度の高い 3 次元モデルを抽出し、それを分割・合体することにより、1 枚の手書き風画像から 3 次元モデルを再構成する手法を提案した。今後の課題としては、3 次元モデルの変形手法を適用することで、より精度の高い再構成モデルの構築を検討する。

## 参考文献

- [1] Pollefeys M., Koch R., Vergauwen M., Gool L.V., Automated reconstruction of 3D scenes from sequences of images. ISPRS J Photogramm Remote Sens 55(4), 251–267, 2000.
- [2] Wang, Nanyang and Zhang, Yinda and Li, Zhuwen and Fu, Yanwei and Liu, Wei and Jiang, Yu-Gang. Pixel2Mesh: Generating 3D Mesh Models from Single RGB Images. arXiv, 2018.
- [3] One Shot Learning and Siamese Networks in Keras. <https://sorenbouma.github.io/blog/oneshot/>.
- [4] Simonyan K., Zisserman A., Very deep convolutional networks for large-scale image recognition (arXiv:1409.1556), 2014.
- [5] Sun C., Yang Y., Guo H., Wang P., Tong X., Liu Y., Shum H.-Y., Semi-Supervised 3D Shape Segmentation with Multilevel Consistency and Part Substitution. Computational Visual Media, 2022.