

pix2pix を用いた影画像からの 3D モデリングのための照明数の検討

菊地悠李[†] 田村仁[†]

日本工業大学先進工学部ロボティクス学科[†]

1. はじめに

3D スキャナは十数年前から存在し始めた(1)がそのいずれもが専門性が高く、知識を持たない者が使用することができなかつた。近年になって手ごろな価格で手に入るようになったとはいえ ToF など特殊なセンサを必要とする。そこで本研究では、単眼カメラの画像を pix2pix で学習させることによる 3D スキャンの手法を提案する。本手法は単眼カメラの画像を使用するため手軽で安価に扱うことができる。しかし単眼カメラの画像一枚では十分な精度が出ない、そこで入力データは複数枚の画像データを組み合わせたものを使用する。具体的な手法の検討としては対象に多方向から一度ずつ光を照射し、そのたび撮影する。この時発生する影を用い情報を増やすことで精度を上げる見込みである。その複数枚撮影した画像を一枚の画像に合成する。本研究の目的はこの時照射する適切な数の照明数を分析し、提案手法の有効性を検討することである。

2. 関連研究

影そのものを 3D データ生成の情報として使用する研究は見受けられなかつた。2 次元の影を用い、物体の表面形状の形を取得する研究(4)などは存在するが地面に落ちた影のみを情報とする手法のため例えば影の形状に変化を与えない位置にある凹型構造がある場合など、原理的に再現できない形状が存在する。

画像を pix2pix に学習させ距離画像を生成する研究(5)があるが推定精度の問題があり、データセットの情報量が向上する余地はない。本研究では地面に落ちた二次元的な影のみ使用するのではなく、物体そのものに落ちた影を含めて利用することで凹構造が存在する場合もデータを生成可能であるとする。

3. 実験方法

Blender を使用し、光の入らない箱状の環境

を用意する。対象となる物体を中央に置き、この物体に光を照射する。照明は天井に格子状に配置している。

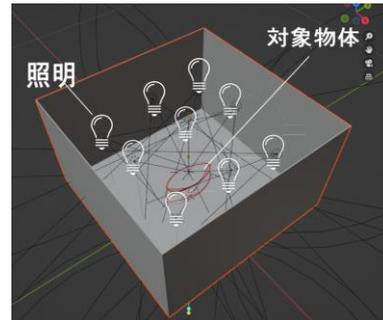


図1 シミュレーションを Blender で行った際の環境。この照明の数を変更し、それぞれ学習を行う。今回はこの照明数を 1,4,9,16,25 と設定した。画像の大きさは 256px×256px とする。この学習を行い生成した画像の PSNR と SSIM を出力し、その結果で比較、検討を行う。SSIM は Ground truth 画像と出力画像の類似度を示しているものであり、その値が 1 に近いほど類似しているといえる。PSNR は画像の劣化度を表すものだが画像がどの程度一致しているかの指標として扱う。

解像度が結果に影響を与えている可能性が浮上したため先の実験で最も良い結果となった照明数の解像度に変更し、もう一度実験を行った。

この実験では比較検討を目的とするため学習に使用する画像の枚数は、テスト用画像 100 枚、学習用画像 200 枚、評価用画像 100 枚とする。

影を生成するための対象物は Blender を用い、生成を行った。今回は手法の有効性を確認することが目的のため対象の形状は円筒形で一貫性のあるものとなっている。

今回は入力画像を照明数に応じて 4 分割、9 分割といったようにまとめて一枚の画像にする。左側の画像が入力画像、右側の画像が Ground truth 画像である。

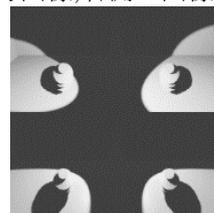


図3 pix2pix において使用するペア画像例

画像生成には pix2pix を用いる。pix2pix とは

[†]Investigation of the number of lights for 3D modeling from shadow information using pix2pix

[†]Yuri Kikuchi

[†]Hitoshi Tamura,

[†]Nippon Institute of Technology Department of Faculty of Advanced Engineering Department of Robotics

敵対的生成学習(GAN)を利用した画像合成アルゴリズムの一種であり入力画像と Ground truth 画像の 2 つをペアとして学習させることにより二つの画像の関係性を学び、入力画像から Ground truth 画像に近づけたものを出力する手法である。

本研究では、対象となる物体の影を生成し、距離画像をペアとして学習させることで、距離画像を生成する。

4. 結果

表 1 に各照明数の PSNR と SSIM を示す。すると照明数は 9 個の値が一番良いと示された。

表 1 各照明数の PSNR と SSIM

照明	1	4	9	16	25
PSNR	19.149	18.974	19.390	19.077	19.254
SSIM	0.854	0.851	0.861	0.851	0.859

図 4 として照明数 9 個の生成画像を示す。

左から SSIM が一番高い値を出した生成物、この生成物の目標、SSIM が一番低かった生成物、同じく目標である。

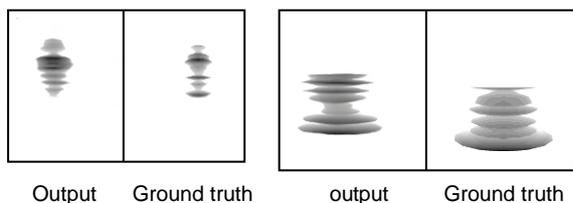


図 4 照明数 9 個での画像生成の例

次に照明 9 個での一枚当たりの解像度である、85px×85px をほかの照明数に当てはめた結果を表 2 に示す。照明数 1 及び 4 は解像度下げることになるため割愛する。結果照明 9 個の値が最も精度が良かった。

表 2 一枚当たりの解像度を 85px×85px とした実験結果

照明数	9	16	25
PSNR	19.390	18.576	18.865
SSIM	0.861	0.843	0.853

5. 考察

実験の結果から入力画像のサイズ上限を 256px×256p とした時の最適な照明数は 9 個であるとわかった。照明数が一個であると純粋に情報量が不足し、精度は悪い。そして照明 4 個では 4 分割にまとめた際の画像の切れ目などがノイズになり SN 比が低下していることが考えられる。照明数 9 以降ではノイズで劣化してしまうことより増えた情報量が大きくその分精度が向上したことによって照明 1 個及び 4 個の学習結果より良い精度が示されたと考える。そしてその SN 比バランスが最も良いものが 9 分割であると示唆される。加えて照明数 9 が最も良い精度を出した理由として画像ごとの変化の差が考えられる。照明

を格子状に配置しているため照明を増やせば増やすほど一つ一つの差が減るため情報量が増えづらくなっていく。そのためノイズでの劣化を情報量で超えることができなくなり照明数 16 及び 25 は 9 に比べて精度が低いものと考えられる。

ここで画像の数が多いほど一枚に対する解像度が下がるということ踏まえて一番精度の良かった照明数 9 個の解像度ほかの照明数の解像度も変更し実験を行うことにした。そうすると同じく照明数 9 個が最も良い結果となった。これは先の実験と同じく SN 比が最も良いのが照明 9 個であったのであると考えられる。加えて照明数 16 及び 25 は解像度が上がっているにも関わらず精度が低下している。これは解像度が向上した分ノイズは増えたが情報量は向上しなかったということが予想でき、解像度は SN 比を向上させる要因にはならないことが示唆される。

6. おわりに

本研究では単眼カメラを用い、それぞれの照明数で撮影し、画像を pix2pix に学習させることによる 3D スキャンの手法の提案とその有効性の評価を行った。その結果複数の照明を用い入力画像を分割した時画像一枚で学習させるより、高い精度を出すことができることがわかった。加えて照明数 9 個が最適な照明数であることも分かった。一枚当たりの解像度が一律の場合は照明数個がより高い精度であることが示された。これにより照明数と解像度により精度が変化することがわかった。今後は結果を参照し、使用枚数を増やすことでさらに精度を高めるため実験を行い、新たな提案手法を検討している。

参考文献

- (1) 井口 征士：3次元形状計測の最近の動向、計測と制御、計測自動制御学会、Vol.34, No.6, pp.429-434 (1995)
- (2) 大谷 幸利：3次元計測技術、日本画像学会誌、画像学会、Vol.53, No.2, pp.128-135 (2014)
- (3) Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, Alexei A. Efros.: Image-to-Image Translation with Conditional Adversarial Networks. In CVPR, 2018
- (4) 森村 周平, 間 博人, 今林 仁広, 松井 健人, 三木 光範 “影情報を用いた物体表面の形状取得手法の検討” 研究報告高齢社会デザイン (ASD), 2016-ASD-5, 10号, pp1-8
- (5) 佐藤 颯人, 田村 仁 “機械学習による RGB 画像からの距離画像の生成”, 2017 年度日本工業大学創造システム工学科卒業論文