

多視点照明合成を活用した複数枚画像入力 pix2pix モデル改良による 3D モデリング精度向上

菊地悠李[†] 田村仁[†]

日本工業大学工学研究科機械システム工学専攻

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った。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、多視点照明を使用した単眼カメラの画像を機械学習で学習させることにより距離画像を生成する 3D スキャンの手法を提案した。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。具体的な手法には対象に多視点から一度ずつ光を照射し、そのたび撮影し、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手法を提案した[6]。この先行研究[6]では、Blender を使用し、シミュレーションを行い、画像を生成、学習を行った。

画像の生成環境は図 1 のように光の入らない箱状の環境を用意した。対象となる物体を中央に置き、この物体に白色の光を照射した。照明は天井に格子状に配置した。

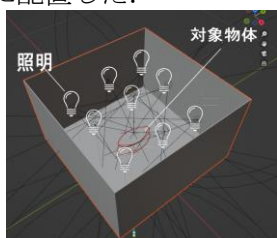


図 1 シミュレーションを Blender で行った際の環境
この照明の数を変更し、影響を確かめた。

形状はランダムな円筒を積み重ねた構造とした。

画像生成には pix2pix を用いる。pix2pix とは敵対的生成学習(GAN)を利用した画像合成アルゴリズムの一種であり図 2 のように入力画像と Ground truth 画像の 2 つをペアとして学習させる。

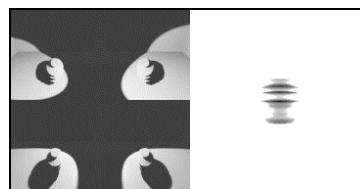


図 2 pix2pix において使用するペア画像例

先行研究[6]では対象となる物体の影を生成し、それを入力画像として距離画像とペアで学習させることで、距離画像を生成した。

入力画像は照明数に応じてまとめて一枚の画像にする。図 2 の左側の画像が入力画像、右側の画像が Ground truth 画像である。

先行研究[6]では照明数を 1, 4, 9, 16, 25 でそれぞれ学習を行い、評価、検討を行った。

学習に使用した画像の枚数は、テスト用画像 100 枚、学習用画像 200 枚、評価用画像 100 枚とした。

画像の評価には SSIM, PSNR を使用した。PSNR は本来、画像の劣化度を表すものだが本稿では画像がどの程度一致しているかの指標として扱う。

先行研究[6]では、情報量が向上するため照明数 1 よりも照明の数が多いそのほかの照明数により良い精度が出るのと考えた。結果は照明数 1 に比べて、ほかのいくつかの照明数で評価値が下がったこと以外に有意な変化はなかった。しかしその一部を除いて評価値が減少しないということはその減少しなかった分評価値が上昇したといえる。また評価値が減少する要因としては画像に切れ目ができることで生じるノイズやそのノイズと照明を増やすことの影響の差によって照明数ごとに評価値にばらつきが出ていると考えた。しかし今回は学習に使用できた画像の枚数が少ないという問題もあり、十分に結果が得られたといづらい。

これらの結果から確認の必要がある以下のような不明点が発生していることがわかった。

- (1) 照明の配置よっての精度への影響
- (2) 分割形式の画像を使用した時の影響
- (3) ライトの色による精度への影響
- (4) 学習枚数が少なかったことからの精度への影響

本稿では以上の不明点を検討することを目的

Improvement of pix2pix model with multiple image inputs using multi-viewpoint lighting synthesis for increased accuracy in 3D modeling

[†]Yuri Kikuchi

[†]Hitoshi Tamura,

[†]Nippon institute of technology Graduate school Mechanical Systems Engineering Major

にそれぞれ実験を行う。

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った。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なく、物体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。

また光の偏光に基づき形状復元を達成する研究[5]では、物体表面の偏光状態からの法線方向を計算に加えて法線を密に求めるために陰に基づく法線推定手法を提案している。この研究と本稿の研究の相違点は、厳密な計算を用いない点である。本稿では形状復元を機械学習することで達成しようとしている。この研究に比べ密な推定は難しいがより簡易的な装置、手法で 3D 形状の復元を目指している。

3. 実験手法

3.1 実験 1

先行研究[6]において照明の配置による影響を調べる必要が出たためその影響を調べた。これを実験 1 とする。先行研究[6]では照明を格子状に配置し、照明の数を変えることで実験を行ったがこの実験では照明を図 3 のように以下の形状、格子状、内接円上、直線状に配置することで比較を行った。なお照明の数はこれらの形状をとることができる最小の数である 9 個を採択した。



図 3 照明の各配置方法

実験 1 では学習に使用する画像をそれぞれテスト用画像 1000 枚、学習用画像 2000 枚、評価用画像 1000 枚、学習回数 200 回とし学習を行った。

3.2 実験 2

次に実験 2 として分割形式の画像を使用した時の影響を検証するため RGB 照明を使用し、実験を行った。対象のオブジェクトに同時に RGB それぞれの照明を図 4 のように 3 点照射し、撮影

を行いその画像で学習を行う。

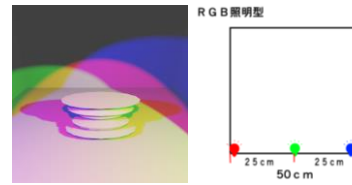


図 4 RGB 照明型の照射例と配置図

他の実験では単色のライトを使用し影を発生させていたがライト自体に色がついていた場合どういった影響が出るのかも加えて確認する。

実験 3 では、実験 2 と同様の画像枚数、学習回数で学習を行った。

3.3 実験 3

実験 3 では、先行研究[6]では少なかった学習に使用する画像の枚数を大幅に増やし、テスト用画像 4000 枚、学習用画像 72000 枚、評価用画像 4000 枚、学習回数 200 回とし学習を行った。

枚数を増やす以外の条件は先行研究[6]の実験と同じである。これは、画像の枚数を増やすことでの精度への影響を確認するとともに、先行研究[6]の実験結果が枚数を増やした時と同様の傾向を示すかどうかを見ることを目的としている。

3.4 実験 4

実験 4 では pix2pix のモデルを改良していくことを予定している。今まで行った実験では図 2 のように画像を分割形式で学習させてきた。しかし画像に切れ目ができること、正解画像とのシルエットがかけ離れてしまうことによる悪影響が予想される。ここで 2 種類の手法で解決を図る。一つ目の手法は pix2pix の損失関数を改良するという方法である。通常、以下の図 5 のように、デプス画像を合成するときは入力画像、Ground truth 画像ともに外形が大まかに一致しているものである。



図 5 分割形式でない画像

しかし、図 6 のように今までの手法では入力画像と共通のシルエットである右の画像とかけ離れてしまう。これが原因で生成結果では外形がうまく再現できていないことが考えられる。



図 6 Ground truth 画像との類似性

そのため Discriminator 側または, Generator 側の損失関数に新たな数値を与えることで今回使用する分割形式の画像において最適な形に改良を行う。そのため, 今回損失関数に加え入れる関数は生成結果の外形を整えられるように設定する。具体的には図 6 の右の画像のようなデプス画像, そのシルエットの画像を以下の図 7 のように用意する。



図 7 Ground truth 画像のシルエット

この図 7 の画像とエポックごとに合成されたデプス画像で差をとる。その値を損失関数に加え入れることで損失関数を調整する。損失関数を loss とし, 改良された損失関数を LOSS, シルエット画像を S, 合成されたデプス画像を D とした計算式を以下に示す。

$$LOSS = loss + (S - D)$$

二つ目の手法としては学習の入力画像に使用する複数枚の画像を 32bit, 64bit の画像一つにまとめその画像を使用し, 学習を行う手法である。この手法は図 8 のように複数枚の画像の同一の場所のピクセルからデータを取り出し, 32bit, 64bit の画像の同一場所のピクセルにまとめ, これをすべてのピクセルで行い画像を生成し, 学習を行うものである。

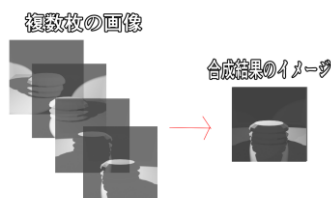


図 8 複数枚の画像の画像の合成イメージ

この手法は結果的に学習させる画像が分割形式ではなく一枚ため画像に切れ目が入ることなくノイズを増やさず得られる情報の量を増やすことができるうえ外形を再現しやすい利点があると予想される。

4. 結果

4.1 先行研究

表 1 に比較のため先行研究[6]の実験結果である各照明数の PSNR と SSIM の平均を示す。

表 1 各照明数の PSNR と SSIM の平均

照明数	1	4	9	16	25
PSNR	20 . 397	20 . 139	20 . 106	19 . 913	20 . 186
SSIM	0 . 864	0. 868	0. 863	0. 854	0. 867

4.2 実験 1

以下に表 2 として実験 2 の PSNR と SSIM の平均の数値を示す。表の結果から格子型が最も良い結果を出した。

表 2 実験 2 の各配置方法の PSNR と SSIM の平均

照明形式	格子型	内接円型	直線型
PSNR	21. 297	20. 726	20. 435
SSIM	0. 875	0. 869	0. 867

4.3 実験 2

以下に表 3 として実験 3 の PSNR と SSIM の平均の数値を示す。格子型に比べて PSNR は減少し, SSIM は微増した。

表 3 実験 3 の結果と格子型照明数 9 の平均

照明形式	格子型	RGB 照明
PSNR	21. 297	21. 121
SSIM	0. 875	0. 880

4.4 実験 3

以下に表 4 として実験 4 の PSNR と SSIM の平均の数値を示す。照明数 1, 9 は特筆して上昇した。なお学習が間に合わなかったため照明数 25 のデータが取れていない。

表 4 実験 4 の各照明数の PSNR と SSIM の平均

照明数	1	4	9	16	25
PSNR	23 . 998	22 . 426	22 . 971	22 . 281	データなし
SSIM	0. 928	0. 908	0. 920	0. 897	データなし

5. 考察

実験 1 では, 図 4 のように照明を配置した。内接円型では外周に照明が多くなるよう, 直線型では, 一方向からの影に多様性を持たせるためこの配置になった。結果は格子型がより良い結果となった。しかし, 今回は照明数 9 を採択したことによって各配置方法に差が少なくなってしまったように見える。特に格子型と内接円型ではオブジェクトの頭上に照明があるか, 外周に一つ照明が増えたかの違いしかなくこの結果だけでは判断しにくい結果となってしまった。

次に行った実験 2 では先行研究[6]で判明した画像の切れ目がノイズになったことを踏まえ入力画像を一枚で, なおかつ情報量を増やすことを目的に RGB 照明を同時に照射し, 撮影, 学習を行うことにした。結果は, 格子型に比べ PSNR は減少し, SSIM は微増することになった。その際生成された画像を以下に図 9 として示す。

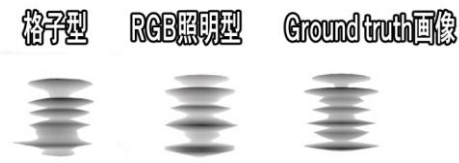


図9 実験3にて生成した画像比較例

図9を見ると格子型は突起の数などの特徴をうまく再現しており、RGB照明型では、照明数1よりも精度が高いためRGB照明を使用し、色を付けたこと、複数照明を一度に照射したことが良い影響を及ぼしたことがわかる。外形に関しては入力画像とGround truth画像のプロポーションが同じため、生成した画像の外形が近くなっていると考えられる。格子型では分割形式の画像であることからシルエットが正解からかけ離れ、生成した画像のプロポーションが崩れてしまっていることが考えられRGB照明型では、格子型に比べ特徴を捉えられていないことがわかった。今後はこの点を改善する必要がある。このことから分割形式の画像では分割なしの画像に比べ外形がとりにくい。しかし、分割形式の画像においてもRGB照明よりも特徴を抜き出しやすいという優位性が確認できた。

実験3では学習枚数を大幅に増やし、先行研究[6]と同じ条件で学習を行った。これは学習枚数による影響を調べるためである。結果は全体的に精度が向上し、特に照明数1, 9が大幅に上昇したように見えた。しかしこの実験では学習時間の長さを鑑みてバッチサイズを途中で2から50に変更した。実際に変更され学習を行ったのが照明数4と16である。このことからバッチサイズを変更して学習を行ったことが精度に影響を与えた可能性が考えられる。そのためもう一度バッチサイズをそろえ学習を行う必要がある。

ここで画像の生成結果とSSIM, PSNRの評価を見たところオブジェクトの面積が大きいほど評価値が低く、面積が小さいほど評価値が高い傾向にあることに気づいた。これは評価の折、白い空白部分を類似点として評価されてしまっているため、空白部分が大きいほど評価値が高くなってしまったことがわかった。これでは一つ一つで比較した時の結果が正しくなくなってしまっている。今回導いた結果は全体で統一はされているため一つの結果としては見れるが具体的な傾向を知ることができない。今後の実験では生成した画像のデプス画像の面積を一定にすることによって正しく評価できるように評価用のプログラムを改善する。その点を踏まえても結果として出た評価値は現状の時点で学習枚

数を増やす以外に大きな差が出たものはない。このことから細部まで確認できていないものもあるが照明の数を変更すること、照明の配置を変えること、色を変えるなどの画像自体を調整することに大きな有効性がないように見える。

そのため、画像自体に工夫を施すのではなく本研究のアイデアである影画像や分割形式の画像に適するよう生成アルゴリズム本体の方を改良調整する必要があると考える。具体的な悪影響として画像に切れ目ができること、正解画像とのシルエットがかけ離れてしまうことで生成画像の外形が安定しないことが考えられる。これを改善するために2種類の手法で解決を図る。一つ目の手法はpix2pixの損失関数を改良する方法である。これは損失関数を分割形式の入力画像に対応できるように調整するものである。

二つ目の手法としては学習の入力画像に使用する複数枚の画像を32bit, 64bitの画像一つにまとめその画像を使用し、学習を行う手法である。この手法の目的は分割形式の画像を使用せず、情報量を増やし、学習を行うものである。今後は画像の評価方法を見直し、こちらの手法を最優先として実験を進めていき、発表時に結果を報告する。

6. おわりに

本稿では先行研究[6]で判明した不明点を確認するため複数の実験を行った。しかし、評価方法に問題があったこと、照明の配置を変更する、照明の色を変更するなどの方法に大きな優位性はなく、根本的な精度向上にはつながらなかったことから、今後はアプローチを変え、生成アルゴリズムそのものの損失関数や学習形式を改良することで精度の向上を図っていく。

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] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, Alexei A. Efros. : Image-to-Image Translation with Conditional Adversarial Networks. In CVPR, 2018
- [3] 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
- [4] 佐藤 颯人, 田村 仁, 檜山 正樹, 入江 俊, 仲田 仁, 機械学習によるRGB画像からの距離画像の生成, 第80回全国大会講演論文集(2), 2018, p229-230
- [5] 吉田 百花, 川原 僚 岡部, 孝弘, 偏光と陰に基づくワンショット法線推定, 第85回全国大会講演論文集(4), 2023, p99-100
- [6] 菊地 悠李, 田村 仁, pix2pixを用いた影情報からの3Dモデリングのための照明数の検討, 第85回全国大会講演論文集(4), 2023, p107-108