深層学習を用いた局所特徴量による建築画像の位置合わせに関する研究
A Research on Deep Local Feature Description for Geometric Matching

○堀江　周平*1，加戸　啓太*2
Shuhei HORIE*and Keita KADO*2

*1 千葉大学大学院　融合理工学府　博士前期課程
Graduate Student, Graduate School of Sci. and Eng., Chiba University.
*2 千葉大学大学院工学研究院　助教　博士(工学)
Assistant Professor, Graduate School of Engineering, Chiba University.

キーワード：深層学習; 局所特徴量; 画像位置合わせ
Keywords: deep learning; local features; image registration

1. 背景

施工プロセスにおける出来形の確認や、運用プロセスにおける状況の確認において写真記録は様々な役割を持っている。写真記録の管理において、その撮影された位置・向き(以下、単に撮影位置とする)が分かると、検索のキーとしてはもちろん、BIM(Building Information Modeling)と併用すれば、三次元モデル上に撮影位置をプロットし写真に映る部品との紐付けを行う、そこに進捗や不具合情報などを付記するといった、様々な活用が見込まれる。写真記録の活用の観点から、屋内における写真の撮影位置の推定手法は重要な要素技術であるといえる。

画像を用いた位置の推定技術は、SLAM(Simultaneous Localization and Mapping)、SfM(Structure from Motion)、VO(Visual Odometry)やフォトグラメトリなど、その目的や処理の違いからいくつかの用語がある。細かい違いはあると思われるが、共通して連続的に撮影された画像群を入力に、以下の手順により撮影位置の取得が行われる。

1) 画像内の特徴的な箇所を、ディテクタにより特徴点として検出し、ディスクリプタを用い特徴量として記述する。特徴点とその特徴量は、合わせて局所特徴量と呼ばれる。
2) 2 枚の画像について、局所特徴量の特徴量の類似度を比較し、画像 A と B において、画像 A の点 a と画像 B の点 b は同じ箇所を写した可能性が高い、といった具合にマッチングを行う。
3) マッチングの結果およびカメラの内部パラメータ[注1]から、画像間の相対的な回転と並進ベクトルの成分を求める。
4) 複数の画像の組み合わせについて 1)～3)を行い並進ベクトルの成分のノルムを推定することで、それぞれの画像の撮影位置を推定する。同時に映る対象の疎な三次元点群を推定する。

この手順において局所特徴量は極めて重要な役割を果たしている。2)でのマッチングにおいて、誤ってマッチングしないことや、マッチした特徴点が実空間において誤差なく一致していることが正確な撮影位置推定に繋がるからである。3)や 4)で用いられる処理では収束計算が行われることから、適切にマッチングされた局所特徴量が多いことも精度に寄与する。

局所特徴量とその三次元座標を取得し蓄積しておけば、新たに撮影した画像についても、その局所特徴量とのマッチングにより撮影位置を推定することができる。近年、このような局所特徴量を用いた位置取得手法は VPS(Visual Positioning System)として活用が試みられており、その一例として自動運転車両の自己位置推定における試行を挙げることができる[注2]。また、近年ではスマートフォン、タブレットに組み込まれた加速度センサーなどの IMU(Inertial Measurement Unit)を撮影位置の推定に援用する VIO(Visual Inertial Odometry)、なども普及しているが、これにおいても局所特徴量の重要性は同様である。

さて、画像中の局所特徴量を抽出する手法として代表的なものに Lowe[1] が提案した SIFT(Scale Invariant Feature Transform)がある。SIFT はそれまでの局所特徴量抽出において課題であったスケール、および回転に対する不変性を持った特徴量抽出手法であり、それ以降画像マッチングのみならず物体検出など多方面で用いられ、現在に至るまで様々な分野での活用が見られる手法である。

SIFT のような特徴量抽出手法はアルゴリズムベースであるが、近年では深層学習を取り入れることも行われている。ディテクタおよびディスクリプタを所謂学習によって得ようとするものであり、前述した局所特徴量の重要性も相まって様々な手法が報告されている。その汎用的な性能への期待に加え、建築空間あるいは特定の建築での位置合わせに特化するよう学習を行うことができる点も興味深く、SIFT が苦手とされる特徴に乏しい、逆に繰り返しのある壁面においても有効な局所特徴量を抽出できる可能

日本建築学会情報システム技術委員会
第45回情報・システム・利用・技術シンポジウム論文集，64-67，2022年12月，東京
Proceedings of the 45th Symposium on Computer Technology of Information, Systems and Applications，AIJ，64-67，Dec.，2022，Tokyo

性もある。

以上を背景に本研究では深層学習を用いた局所特徴量抽出手法である SuperPoint[2]、SuperGlue[3]を用いた建築画像の位置合わせを行い、SIFT との比較を通してその効果について考察する。

2. 研究概要

本研究は建築画像の位置合わせにおける深層局所特徴量の効果について検証するものである。深層局所特徴量の抽出手法として、その代表的な手法である SuperPoint と SuperGlue を試行する。詳しくは後述するが、SuperPoint はディテクタ及びディスクリプタを学習する、SuperGlue はディスクリプタのマッチングを学習するものである。汎用データセットで学習したパターンと、特定の建築物の画像データで学習を追加したパターンを比較に含める。

よって比較検討のパターンは以下のように整理できる。なお、本研究では汎用データセットとして MS-COCO(以下 COCO と表記）を、具体的な建物として千葉大学の墨田サテライトキャンパスを設定し、墨田サテライトキャンパスにて撮影した専用データセットを SSC と表記することにする。また比較の上で重要度が低いと考えられる一部の組み合わせについては除外している。

A) SIFT : SIFT を特徴量抽出に用いるパターン
B) SPCOCO : COCO で学習した SuperPoint を特徴量抽出に用いるパターン
C) SPSSC : SPCOCO をさらに SSC で学習した SuperPoint を特徴量抽出に用いるパターン
D) SG-SIFT-COCO:SIFT にて抽出した局所特徴量のマッチングを COCO で学習した SuperGlue をマッチングに用いるパターン
E) SG-SIFT-SSC : SG-SIFT-COCO をさらに SSC で学習した SuperGlue をマッチングに用いるパターン
F) SG-SPCOCO-COCO : SPCOCO(B)にて抽出した局所特徴量のマッチングを COCO で学習した SuperGlue を用いるパターン
G) SG-SPCOCO-SSC : SPCOCO(B)にて抽出した局所特徴量のマッチングを SSC で学習した SuperGlue を用いるパターン

ところで、特定の建築物で学習を行うことは長期的な維持管理を想定したものである。本研究では具体的な建物として千葉大学の墨田サテライトキャンパスを設定したが、本研究においてここに特化した局所特徴量抽出が学習できれば、今後の施設運用における写真記録管理に継続的に利用できる。

2.1. SuperPoint

SuperPoint はディテクタおよびディスクリプタを CNN(Convolutional neural network)を用いて行うネットワークである。人工的に作った教師ありデータを用いた教師あり学習であり、以下 3 ステップで行われる。

1) 人工の図形データで正解データを自動生成し特徴量を学習
2) 学習したモデルを使って複雑な画像の特徴点を擬似正解データとして作成
3) 作成したデータから特徴点・記述子を学習

論文の著者らにより、デモンストレーション用のスクリプトが公式に公開されている[注3]が、学習環境は非公開となっている。本研究ではこれに学習環境を加えた eric-yyjau によるリポジトリ[注4]に手を加えたものを用いた。図 1 は SIFT、SuperPoint それぞれで局所特徴量を抽出したものである。

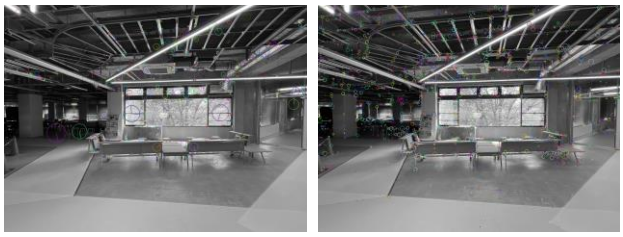

図 1 SIFT（左）と SuperPoint（右）による局所特徴量

2.2. SuperGlue

SuperGlue は、GNN(Graph Neural Network)と Attention 機構を用い、画像内の局所特徴量同士の位置や特徴量の関係に加え、もう一方の画像内の局所特徴量との関係からマッチングを学習するネットワークである。簡単には、マッチングの対象とするペアの画像における局所特徴量の傾向からマッチングを行うものであり、人間が画像の位置合わせを行う際に、局所に注目しつつも大域との整合にも配慮するのに似たアプローチが採られている。

SuperGlue についても公式としてはデモンストレーション用のスクリプト[注5]のみ公開されている。これをベースに学習環境を加えた HeatherJiaZG によるリポジトリ[注6]に手を加えたものを用いた。

2.3. 評価方法

特徴量抽出パターンの比較については、ホモグラフィ変換を行った画像間マッチングの結果を可視化しつつ、ホモグラフィ変換の推定精度を指標に評価を行う。具体的には、元画像とそれにホモグラフィ変換をかけたペアについて、局所特徴量の抽出、マッチングを行い、その結果からホモグラフィ変換を推定することを行う。推定したホモグラフィ変換の正しさは再投影誤差(reprojection error)で評価する。これは、画像内の点について、真のホモグラフィ変換を行った場合と推定したホモグラフィ変換を行った際とのズレの平均であり、小さければ精度が良いものである。

また、ホモグラフィ推定を行う過程で RANSAC[注7]により外れ値(outlier、外れマッチング)が排除されるので、マッチングした局所特徴量に対する外れマッチングでない

もの(inlier)の割合を局所特徴量の精度(precision)、抽出された局所特徴量に対する inlier の割合をスコア(matching score)として評価する。なお、RANSAC は多数決的な振る舞いをするため、誤マッチングが多数であれば局所特徴量の精度は高く評価されてしまうが、この場合、再投影誤差が大きくなる。よって、第一に再投影誤差が小さいこと、第二に局所特徴量の精度、スコアが高いことが重要であると整理できる。この指標をもとに各手法についての評価を行う。図 2 は評価を行った例である。図中左上に評価指標となる数値が記載されている。n_matched はマッチングの総数であり、図中緑の線が inlier を、赤の線が outlier を表している。n_inlier の値の右括弧内の数値は前述の局所特徴量の精度(precision)であり、全体のマッチング数に対する inlier の割合を示している。

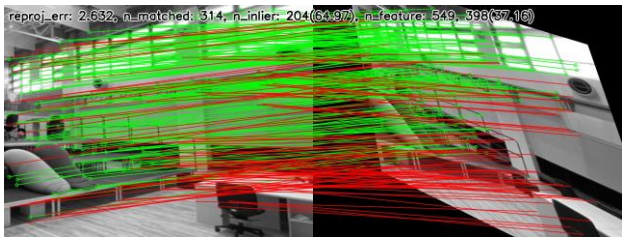

図 2 マッチングの例

2.4. 墨田サテライトキャンパスのデータセット(SSC)

墨田サテライトキャンパスにて撮影した画像は 200 枚であり、維持管理を想定した際に設備等の不具合を検知することを踏まえ、なるべく全体が移るような引き気味の画像を撮影した上で、その一部をトリミングしてデータの拡張処理を行った。なおデータ強化のため、輝度調整とぼかしの 2 種類の処理をランダムに加えつつ、画像 1 枚につき 50 枚データの水増しを行うことで、計 10000 枚のデータセットを作成した。

図 3 データの例

3. 評価

専用データセット(SSC)のテスト用データ 50 件について、各パターンでの再投影誤差の平均値、ホモグラフィ推定が上手く行えなかった数、および特徴量抽出方法の違いが顕著に現れた二つの画像の例を表 1 にまとめた。図 4 はこのうち無作為に選んだ 25 件について、それぞれのパターンの再投影誤差をプロットしたものである。

3.1. 考察

表 1 を概観する。局所特徴量の抽出方法の観点から A~B を比較すると、SIFT による A が SuperPoint による B、C よりも良い結果を示していることがわかる。図 4 の picture_id(以下、id)3、9、20 などでこの傾向が確認できた。これは DeTone らの報告 [2]においても確認することができ、一因として SIFT がアルゴリズム内においてサブピクセル推定を行っていることが挙げられている。学習に

表 1 各パターンの再投影誤差の平均値、ホモグラフィ推定が上手く行えなかった数

	A SIFT	B SPCOCO	C SPSSC	D SG-SIFT-COCO	E SG-SIFT-SSC	F SG-SPCOCO-COCO	G SG-SPCOCO-SSC
Mean Reproj. err	75.52	30.31	225.92	29.76	22.64	27.49	**19.20**
Num Failed	5	19	20	7	**4**	8	8
Exp1(img_ID: 3) Reproj. err	0.79	1.77	1.58	**0.28**	0.69	2.14	2.35
Exp2(img_ID: 16) Reproj. err	348.86	failed	failed	30.47	225.47	36.48	**8.77**

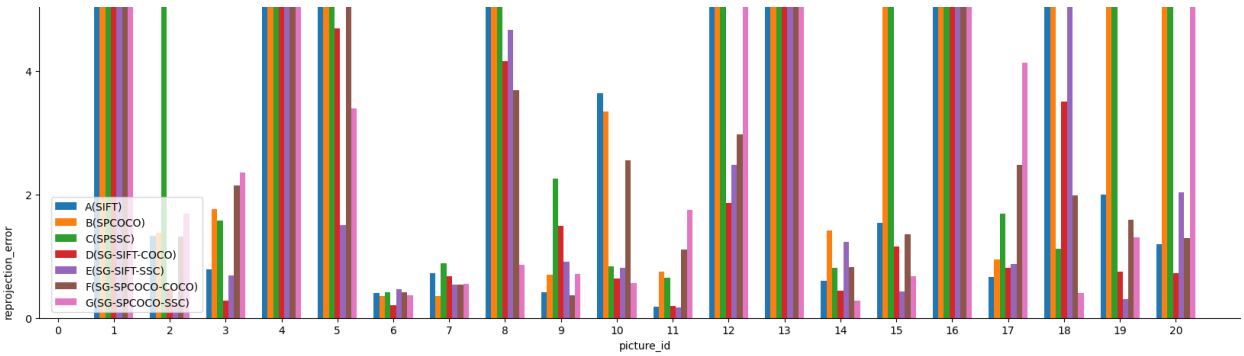

図 4 各パターンにおける画像毎の再投影誤差

SSC を用いた C を確認すると、id18 のように C が A の数値を大幅に上回るようなものも見られた一方で、再投影誤差の平均値や推定に失敗した数は最も悪い結果となった。過学習となった可能性もあり、改善の余地があると考えている。

SuperGlue によりマッチングを学習したパターンを見ると、E、G をはじめ A～C と比較して良い結果が得られていることがわかる。図 4 においては id7、15、19 にこの傾向が見られる。また、図 5 は表 1 に示したマッチングが難しいケースの一つであった id16 について、B、G の結果を示したものである。B、G は局所特徴量の抽出は同じ抽出器(B のもの)である。id16 では抽出された局所特徴量は 400 個と少なくはないが、似たものが多いのか B ではマッチングに失敗している一方、G では再投影誤差が 8.77 とまずまずの結果が得られている。前述のように SuperGlue の適用により局所特徴量単体でなく画像内の局所特徴量との関係からマッチングが行われることが寄与していると考えられる。専用データセットの効果については再投影誤差の平均が最も小さい G、ホモグラフィ推定が上手く行えなかった数が最も小さい E は共に SSC での学習を行ったものである。少なからず効果が現れたと考えている。

一方で比較的マッチングがしやすいケースでは、SIFT を用いる A、D、E の結果が良い傾向にあった。図 6 は表 1 に示した id3 について、A、D、G の結果を示したものである。SIFT による A、D は再投影誤差が 0.79、0.28 と良好な結果が得られた一方、G では 2.35 とやや誤差のある結果となった。G ではマッチングした特徴量 146 点のうち 46 点が outlier であり適切でないマッチングが増えたことが誤差につながったと考えられる。このあたりは学習時において、正解のマッチングと誤ったマッチングの評価バランスに調整の余地があると思われる。

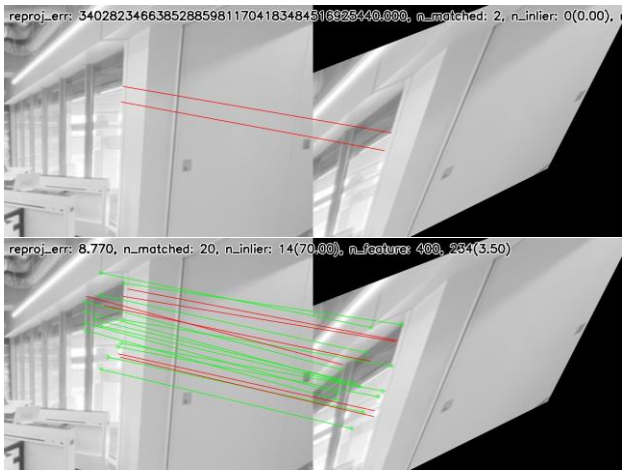

図 5 マッチングが難しいケース（id16、上 B、下 G）

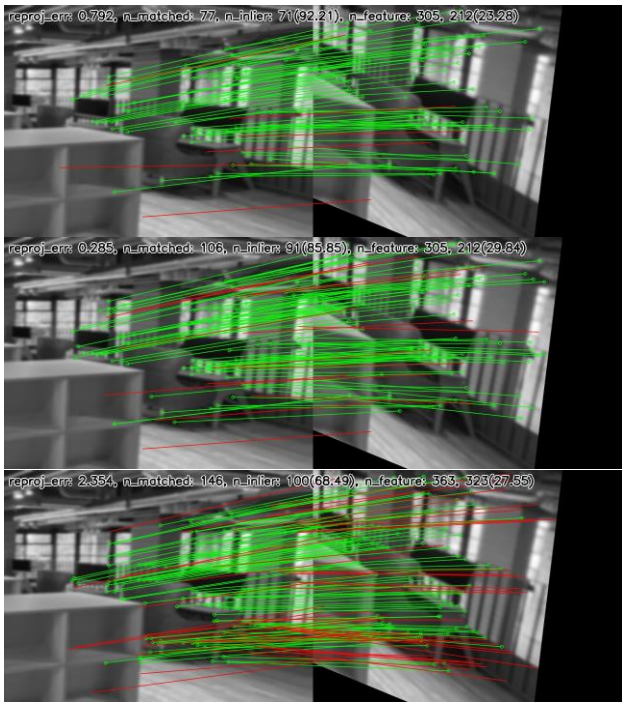

図 6　マッチングしやすいケース（id3、上から A、D、G）

4. まとめ

深層学習を用いた局所特徴量抽出手法によって建築画像の位置合わせを行い、従来手法との比較・検討を行った。

[注釈]

注 1　カメラの光学的中心と焦点距離。カメラ座標を画像座標変換する行列。

注 2　https://www.mlit.go.jp/plateau/use-case/smart-planning3-007/ (2022/08/15 アクセス)

注 3　https://github.com/magicleap/SuperPointPretrained Network (2022/08/15 アクセス)

注 4　https://github.com/eric-yyjau/pytorch-superpoint (2022/08/15 アクセス)

注 5　https://github.com/magicleap/SuperGluePretrained Network(2022/08/15 アクセス)

注 6　https://github.com/HeatherJiaZG/SuperGlue-pytorch (2022/08/15 アクセス)

注 7　Random Sample Consensus

[参考文献]

1) David G. Lowe, Distinctive Image Features from Scale-Invariant Keypoints, International Journal of Computer Vision, 2004.

2) Daniel DeTone, Tomasz Malisiewicz, Andrew Rabinovich, SuperPoint: Self-Supervised Interest Point Detection and Description, Conference on Computer Vision and Pattern Recognition (CVPR), 2017

3) Paul-Edouard Sarlin, Daniel DeTone, Tomasz Malisiewicz, Andrew Rabinovich, SuperGlue: Learning Feature Matching with Graph Neural Networks, onference on Computer Vision and Pattern Recognition (CVPR), 2020

4) 加戸、他、SfM による写真撮影位置推定の建築分野における活用に関する研究、日本建築学会技術報告集、第 24 巻 57 号、pp.873-876、2018