

令和 3 年度卒業論文

特徴点抽出法 AKAZE、AKAZE-ALEX、
SuperPoint の比較

宮崎大学 工学部 情報システム工学科

熊田陸

指導教員 椋木雅之

概要

本研究では、特徴点マッチング手法のうち、アルゴリズムベースの手法である AKAZE、深層学習ベースの手法である SuperPoint、梶原が提案する AKAZE-ALEX を比較する。異なる画像間で、固有の点を対応付ける特徴点マッチングはカメラの移動量推定、類似画像の検索、画像内の物体検出など多くのコンピュータビジョンの分野で重要な役割を担っており多くの手法が提案されている。例えば、アルゴリズムベースの手法では AKAZE が提案されており、拡大縮小、回転に強い性能となっている。一方、機械学習の発展手法である深層学習は、近年コンピュータビジョンの分野で大きな成果を挙げており、深層学習を特徴点マッチングに取り入れた研究も行われている。例えば、SuperPoint は特徴点抽出をするモデルを疑似正解特徴点データを学習することにより作成しており高いマッチング精度を示している。また、梶原が提案する AKAZE-ALEX は、AKAZE で検出した特徴点に対し、特徴点まわりの画像の切り抜きを行い、学習済みの深層学習モデルを用いて切り抜いた画像の特徴量を与えマッチングする。これにより、アルゴリズムベースの手法である AKAZE よりも高い精度を出しているが、他の深層学習ベースの手法とは比較がされていない。そこで本研究では、AKAZE、AKAZE-ALEX、SuperPoint の 3 つの手法について、画像列からの (Structure from Motion ; SfM) および 3 次元形状復元、移動体の自己位置推定と環境地図の作成 (Simultaneous Localization and Mapping ; SLAM) の利用場面を想定し、それぞれの場面で必要とされる性質を検討した上で、評価基準を定め比較を行った。実験結果より、マッチング成功率が高く、成功マッチング数が多く、対象物体撮影時に照明の変化があってもマッチングに影響がないことから、3 次元形状復元においては SuperPoint が適していることが分かった。また、撮影対象の大きさの変化や、ノイズが入ってもマッチング成功率が大きく下がらず、かつ処理時間が短いことから、SLAM においては、AKAZE が適していることが分かった。AKAZE-ALEX は、処理速度が他手法と比較すると極端に遅いこと、拡大縮小の度合いが大きいとマッチング成功率が大きく下がるといった課題点が見つかった。比較実験と考察により、それぞれの手法の特性を示すことができた。

目次

1. はじめに.....	1
2. 特徴点抽出法の比較.....	2
2.1 特徴点マッチング.....	2
2.2 特徴点マッチングの利用場面.....	2
2.3 特徴点抽出手法の評価基準.....	3
2.3.1 回転、拡大縮小不変.....	2
2.3.2 照明変化不変.....	3
2.3.3 ノイズ変化不変.....	4
2.3.4 処理時間.....	4
2.4 マッチング処理.....	5
3. 比較する特徴点抽出法.....	6
3.1 AKAZE.....	6
3.2 AKAZE-ALEX.....	6
3.3 SuperPoint.....	8
4. 実験.....	10
4.1 特徴点抽出手法.....	10
4.2 回転拡大縮小変化.....	11
4.2.1 実験方法.....	11
4.2.2 実験結果.....	12
4.3 照明変化.....	17
4.3.1 実験方法.....	17
4.3.2 実験結果.....	18
4.4 ノイズ変化.....	21
4.4.1 実験方法.....	21
4.4.2 実験結果.....	21
4.5 処理時間.....	22
4.5.1 実験方法.....	22
4.5.2 実験結果.....	23
4.6 考察.....	25
5. おわりに.....	27
謝辞.....	28
参考文献.....	29

1. はじめに

異なる画像間で、固有の点を対応付ける特徴点マッチングはカメラの移動量推定、類似画像の検索、画像内の物体検出など多くのコンピュータビジョンの分野で重要な役割を担っている。そのために多くの研究がされており、アルゴリズムベースの手法では SIFT[1],SURF[2],KAZE[3],AKAZE[4]などの多くの手法が提案されている。これらの手法は、画像に回転拡大縮小などの変化があっても特徴点マッチングの結果が影響を受けにくいなどの優れた特徴がある。しかし未だ、特徴点マッチングの精度には改善の余地がある。特徴点マッチングの精度向上が達成できれば、特徴点マッチングを用いたコンピュータビジョン分野での様々な手法で高精度な成果が期待できる。

一方、機械学習の発展手法である深層学習は、近年コンピュータビジョンの分野で大きな成果を挙げており、深層学習を特徴点マッチングに取り入れた研究も行われている。例えば、SuperPoint[5]は特徴点抽出をするモデルを疑似正解特徴点データを学習することにより作成している。この手法は従来の手法よりもマッチング精度が高く、コンピュータビジョンに関する世界トップレベルの国際会議 CVPR (Computer Vision and Pattern Recognition) で行われたマッチングコンペなどで上位の成績を挙げている。梶原[6]、黒河[7]も、深層学習を特徴点マッチングに取り入れた手法を提案している。AKAZE で検出した特徴点に対し、特徴点まわりの画像の切り抜きを行い、学習済みの深層学習モデルを用いて切り抜いた画像の特徴量を与えマッチングする。これにより、アルゴリズムベースの手法である AKAZE よりも高い精度を出している。この内、梶原[6]が提案している AKAZE-ALEX はマッチング精度に関して従来手法である AKAZE を上回っているが、深層学習ベースの他の特徴点マッチング法との比較は行われてはいない。

そこで本研究では、アルゴリズムベースの手法である AKAZE、深層学習ベースの手法である Super Point、梶原が提案する AKAZE-ALEX を比較する。特徴点マッチングの利用場面を分析して、それぞれの場面で必要とされる性質を検討した上で、評価基準を定め比較を行う。また、比較実験の結果に基づき、それぞれの手法の特性について考察する。

以下 2 章では、特徴点抽出法を比較する際の評価基準について検討する。3 章では比較する特徴点抽出法の詳細について述べる。4 章では比較実験を行い、5 章で本論文のまとめを行う。

2. 特徴点抽出法の比較

2.1 特徴点マッチング

特徴点マッチングは、同じ物体が写る 2 枚の画像間で、同じ物体上の同じ点を対応付ける手法であり、特徴点検出、特徴量記述、マッチングという三段階からなる。特徴点検出では、画像中から角や線の交わり等の他と異なる固有の点(特徴点)の座標を検出する。特徴量記述では、検出した特徴点の固有性を、ベクトルやバイナリコード表現した値(特徴量)で算出する。マッチングでは、対応付ける特徴点の特徴量を比較し、類似度が高い特徴点同士を対応付ける。本研究では、特徴点検出と特徴量記述の二つのステップを合わせた工程のことを「特徴点抽出」と呼ぶ。この 2 画像間で対応付けた点が同じ物体上の同じ点であることが重要である。

2.2 特徴点マッチングの利用場面

特徴点マッチングは Structure from Motion (SfM) という技術を用いた物体の 3 次元形状復元、自己位置推定と環境地図の作成 (Simultaneous Localization and Mapping 以下 SLAM)、物体追跡などに用いられる。

3 次元形状復元とは、復元対象物体が写る 2 次元の画像を異なる位置で撮影をし、対象物体の固有の各特徴点を対応付けることにより、2 次元の画像から 3 次元の物体を復元する技術である。3 次元形状復元では、前述の対象物体の固有の各特徴点の対応付けに特徴点マッチングを用いる。

SLAM は、移動体の自己位置推定と環境地図の作成を同時に行う技術であり、構築した地図情報を用いて障害物などを回避しつつ特定のタスクを遂行するお掃除ロボットや自動運転などに利用される。SLAM では移動体が撮影した画像から得られた特徴点同士をマッチングすることで、撮影対象の 3 次元形状およびカメラの外部パラメータの推定を行う。

物体追跡とは、与えられた動画から、指定した対象がどのように移動したかを推定する技術である。物体追跡では、動画における各フレーム間で追跡対象物体の固有の特徴点を対応付けることにより物体を追跡する。この対応付けのために特徴点マッチングを行う。

2.3 特徴点抽出手法の評価基準

2.3.1 回転拡大縮小不変

本研究でいう回転・拡大縮小不変とは、特徴点マッチングを行う 2 枚の画像間で、対象物体が異なる大きさや角度で写っていても、特徴点マッチングに影響しない性質のことである。同じ物体が写っている 2 枚の画像があるとき、必ずしも同じ場所、角度で撮影されているとは限らない。そのため同じ物体でも画像によって大きさや角度が異なる。そのような変化が、特徴点マッチングに影響しないことが重要である。

物体の 3 次元形状復元においては、回転に不変であることが重要である。3 次元形状復元では、復元対象物体を様々な角度から撮影するので、復元対象物体の見え方が異なる。回転などにより、復元対象物体の見え方が異なっても、復元対象物体の固有の点を正しく対応づけできれば、適切な 3 次元形状復元が行える。

一方、SLAM においては、特に拡大縮小不変であることが重要である。SLAM では移動体に搭載したカメラから得られる時系列の画像間に特徴点マッチングを適用する。この時、進行方向を向いたカメラで撮影した画像では、移動体の進行により物体までの距離が近づき、画像上での大きさが変化する。画像上での大きさ変化しても特徴点マッチングを正しく行えば、移動体の自己位置推定や環境地図の作成が正しく行える。

2.3.2 照明変化不変

本研究でいう照明変化不変とは、定点カメラで撮った画像間で、時間変化による自然な明るさの変化、物体自体の明るさの変化があっても特徴点マッチングに影響しない性質のことである。同じ物体でも時間によって照明の変化度合いは異なる。そのような変化が特徴点マッチングに影響しないことが重要である。

特徴点マッチングは画像からの 3 次元形状復元で利用される。特徴点マッチングを行う異なる画像間においては同じような物体を撮影しても撮影時間帯や状況に応じて光源の位置が違うなどして物体の見え方が異なる場合があるという問題がある。それに対応するためには照明変化不変が必要である。

2.3.3 ノイズ変化不変

本研究でいうノイズ変化とは、定点カメラで撮った画像間において、撮影者が予想できなかったノイズが発生する変化のことをいう。同じ物体でも撮影器具の不調などによってノイズの変化度合いは異なる。そのような変化が特徴点マッチングに影響しないことが重要である。

特徴点マッチングは物体追跡技術に使用されている。例えば、定点カメラで撮影した画像内の人物の追跡などに用いられる。定点カメラはコスト面の制約などで、一般にそれほど画質が高くなく、ノイズの多い画像となる。このようなノイズを含んだ画像間においても正しく特徴点マッチングを行えば、実場面での物体追跡に有効である。

2.3.4 処理時間

処理時間が短いことは、特徴点マッチングを利用する技術全般において、処理の効率化が期待できる。例えば、物体追跡においては、追跡対象物体を時間的に連続して捉え続ける必要があるが、処理時間が長いと画像間の違いが大きくなり、物体を途中で見失ってしまう可能性がある。このように処理時間が長いと特徴点マッチングを利用する技術全般でデメリットが発生するケースが多く、処理時間は早ければ早いほど良いとされる。

2.4 マッチング処理

特徴点マッチングの3段階目のマッチングでは、対応付ける特徴点の特徴量を比較し、類似度が高い特徴点同士を対応付ける。本論文では、総当たりマッチングとクロスチェックを組み合わせた手法を用いる。図1に総当たりマッチング、図2にクロスチェックの簡略図を示す。総当たりマッチングは、画像1から検出された各特徴点に対して、画像2のすべての特徴点と特徴点類似度を計算し、類似度が最も高い特徴点と対応付ける。この際、画像2の特徴点が画像1の複数の特徴点と対応することがあるのでクロスチェックを行う。クロスチェックでは画像1と画像2の立場を入れ替えて同様に総当たりマッチングを行う。双方の結果で一致するものをマッチング結果とする。この2つを組み合わせて行うことで総当たりマッチングの問題である1つの特徴点に対して複数個対応付くことがなくなり、より信頼性の高いマッチング結果を得ることが出来る。

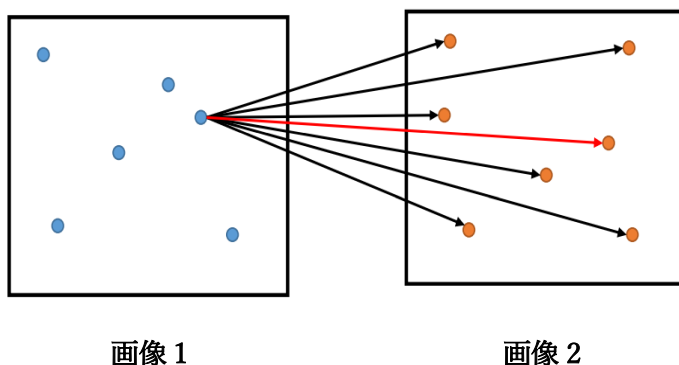


図1：総当たりマッチング

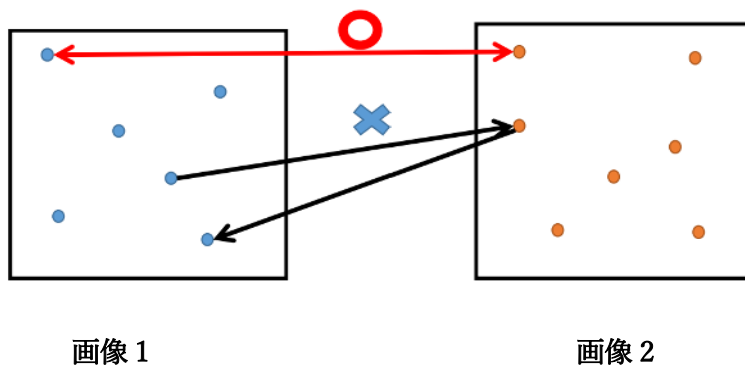


図2：クロスチェック

3. 比較する特徴点抽出法

3.1 AKAZE

AKAZE は、特徴点検出と特徴量記述を行う手法である。他の従来手法である SIFT や SURF の欠点を改善した KAZE をもとにしている。SIFT や SURF などの他の手法と比べロバスト性や処理速度の面で優れている。AKAZE は、拡大縮小、回転に強い性能となっている。特徴点検出では、特徴点と特徴点のスケールを検出する。スケールを正規化することで拡大縮小不変となる。特徴量記述では、オリエンテーションにより向きを正規化を行うことで、回転不変な特徴量を得ることができる。オリエンテーションとは、特徴点における方向のことであり、勾配の方向、強度により求める。特徴量はバイナリコードで表される。そのためマッチングでは、特徴量間のハミング距離を計算し、その距離を用いて対応付けを行う。

3.2 AKAZE-ALEX

AKAZE-ALEX は特徴点検出により検出された点周りの画像を切り取り、その画像に対し学習済みの深層学習モデルから特徴量を与えることによりマッチングする手法である。図 3 に AKAZE-ALEX の流れを示す。AKAZE-ALEX では特徴点検出には AKAZE を使用する。特徴量記述には事前学習済みの深層ニューラルネットワーク AlexNet を用いる。特徴量記述においての特徴量算出の簡略図を図 4 に示す。検出した特徴点の周辺の $D \times D$ 画素の範囲を画像から切り取る。切り取った画像から特徴量を求める。ここで、画像の外側の特徴点においては、 $D \times D$ 画素の範囲を切り取れない場合がある。そのような特徴点には特徴量を与えず除外する。この際、画像を AKAZE で求めたオリエンテーション分回転させることで、回転の影響に対処する。

特徴量の算出には、AlexNet[8]と呼ばれる深さが 8 層の事前学習済みの深層ニューラルネットワークを用いる。このネットワークは 100 万枚を超える画像で学習されている。このネットワークに上記の切り取った画像を与えて得られる fc7 層の出力を特徴量とする。特徴量の次元数は 4096 次元である。

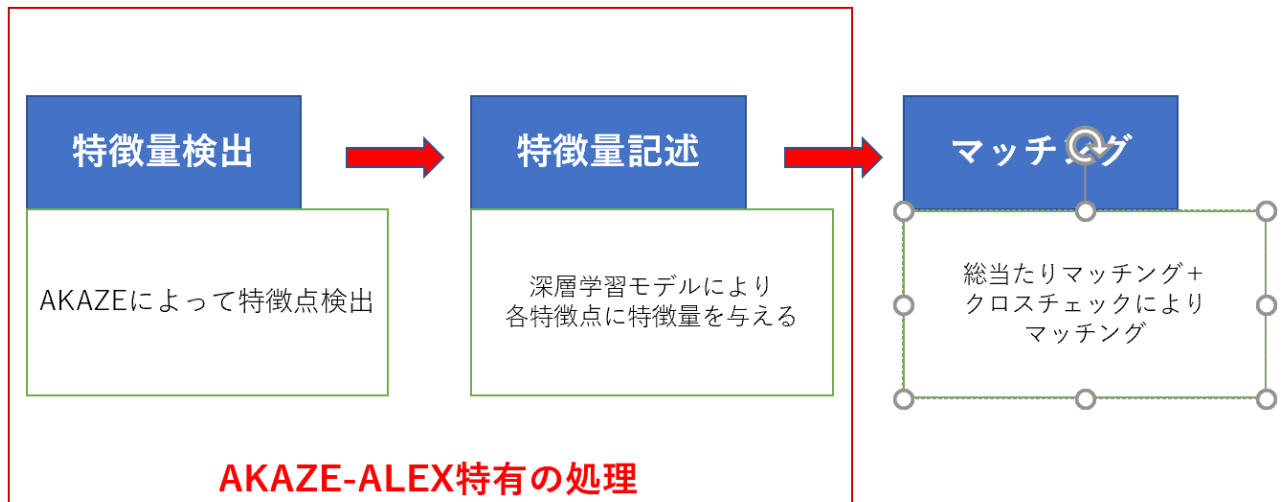


図3：AKAZE-ALEXの流れ

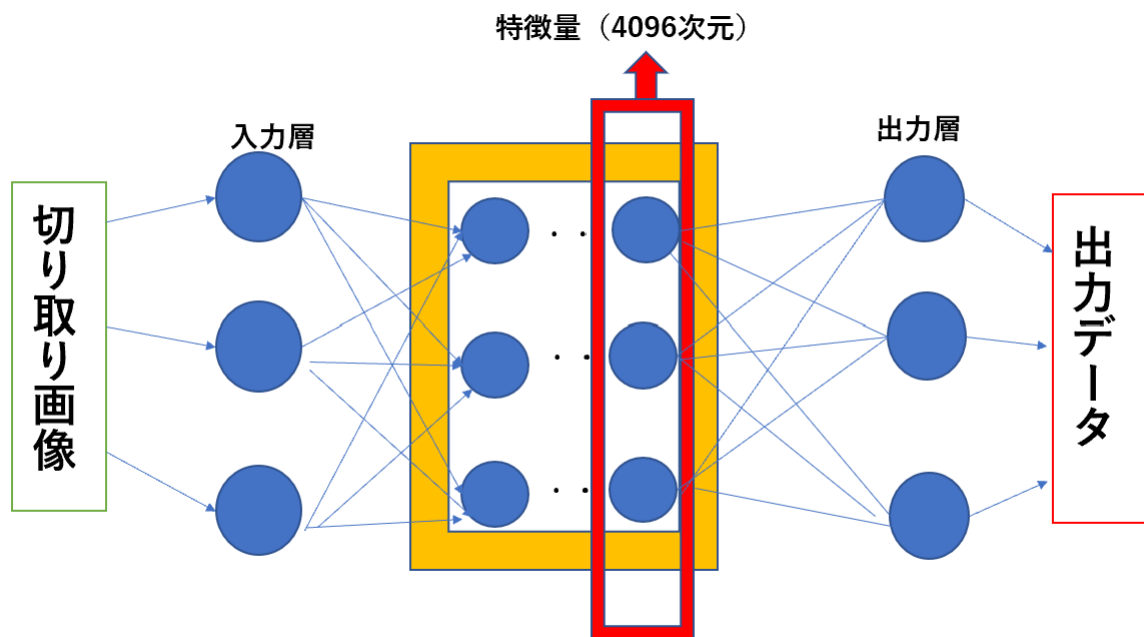


図4：事前学習済み深層ニューラルネットワークによる特徴量算出

3.3. SuperPoint

Super Point[5]は、Tone らによって提案された特徴点抽出器である。疑似特徴点正解データを学習することにより特徴点抽出器を作成する。SuperPoint は

(a)初期の特徴点検出器”Magicpoint”の作成 (図 5)

(b)疑似特徴点正解データの作成 (図 6)

(c)疑似特徴点正解データの学習による SuperPoint 特徴点抽出器の作成 (図 7)

の3ステップで構成される。(a)のステップでは、単純な図形の端点を特徴点としたデータセットを学習することにより初期の特徴点検出器”MagicPoint”を作成する。(b)のステップでは疑似特徴点正解データの作成を行う。データセットにある画像に対し、複数パターンのホモグラフィ変換を行い、(a)で作成した MagicPoint でそれぞれの変換した画像に対し特徴点検出を行う。変換し、特徴点を検出した画像に対し、逆変換をかけ特徴点の位置の対応付けを行う。この工程を Homographic Adaptation という。そのようにして疑似正解特徴点データを作成する。(c)では(b)で作成された疑似正解特徴点データを学習することにより、特徴点抽出器を作成する。SuperPoint は特徴点マッチングの3ステップの内、特徴点検出と特徴量記述を行う。このようにして作成された SuperPoint は、Tone らの論文[5]内でマッチング精度に関し、既存手法の LIFT[9]、SIFT[1]、ORB[10]と比較して高い精度を示している。

単純な図形の端点を
特徴点としたデータ
セット

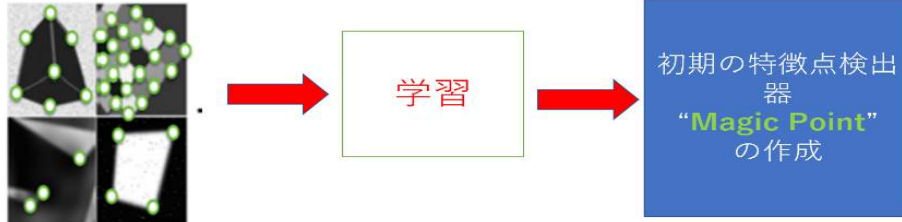


図 5：初期の特徴点検出器“Magic Point”の作成

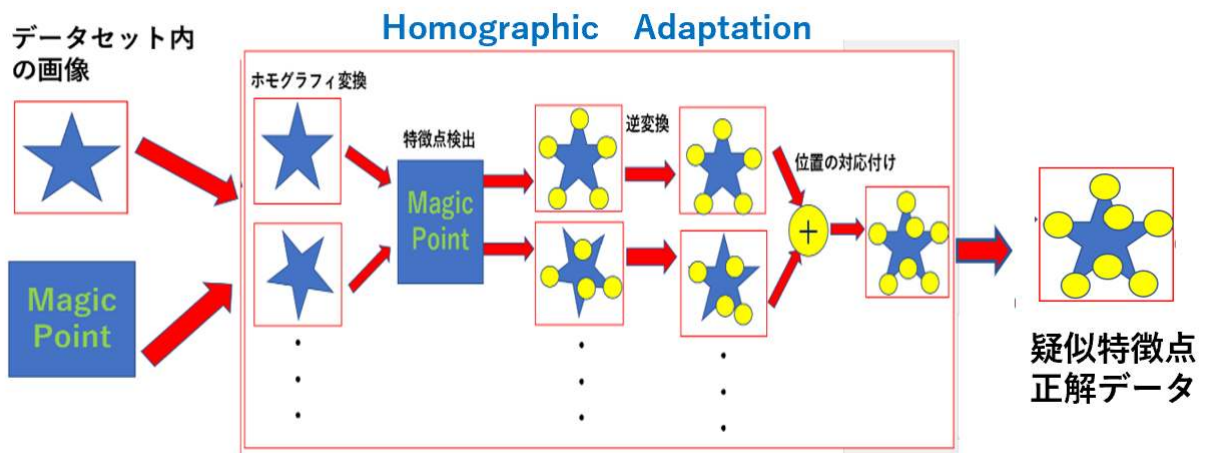


図 6：疑似特徴点正解データの生成

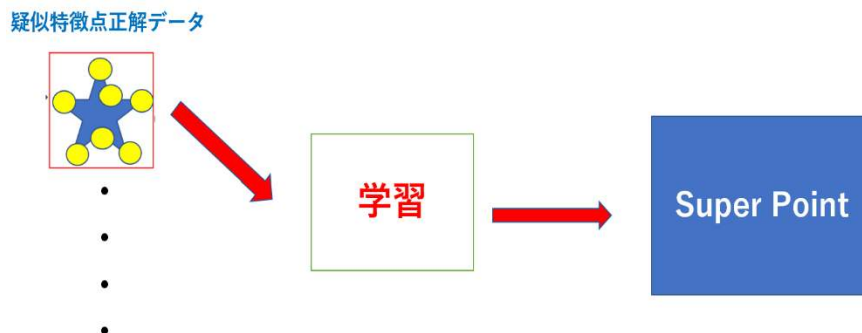


図 7：疑似特徴点正解データの学習による SuperPoint 特徴点抽出器の作成

4. 実験

4.1 特徴点抽出手法

本実験では、AKAZE、AKAZE-ALEX、SuperPoint の3つの手法を比較する。特徴点マッチングの利用場面を分析し、それぞれの場面で必要とされる性質を検討したうえで、評価基準を定め比較実験を行う。AKAZEはOpenCV4.5.5.62に実装されているもの、AKAZE-ALEXはPytorchを使って自身で実装したプログラム、SuperPointはGithub上で公開されているプログラム[11]を利用する。表1に特徴点マッチングの三段階のステップの内それぞれの手法が何に依存するかを示す。AKAZE-ALEXにおいては、切り取り範囲は 33×33 画素とする。SuperPointに関してはモデル作成者によって最高のモデルと述べられている疑似正解特徴点データを170000回学習したモデル[11]を使用する。

表1：比較する手法

手法	特徴点検出	特徴量記述	マッチング
AKAZE	AKAZE	AKAZE	総当たりマッチング＋クロスチェック
AKAZE-ALEX	AKAZE	AlexNet	総当たりマッチング＋クロスチェック
SuperPoint	SuperPoint	SuperPoint	総当たりマッチング＋クロスチェック

実験におけるマッチング成功率は式(1)で求める。

$$\text{マッチング成功率} = \frac{\text{正しい対応点の数}}{\text{得られた対応点の数}} \dots \dots (1)$$

マッチングに使用する元画像の*i*番目の特徴点座標 Q_i と比較画像の*j*番目の特徴点座標 P_j が対応しているとする。また、比較画像の座標 P_j に対し本来対応すべき正しい対応点の座標を P'_j とする。この時、許容座標誤差 T に対して $|P_j - P'_j| < T$ を満たす対応点を「正しい対応点」とする。本実験では、許容座標誤差 $T = 10$ 画素とする。

4.2 回転拡大縮小変化

4.2.1 実験方法

本実験では、回転拡大縮小変化の度合いに対するマッチング成功率の変化を評価する。そのために、元画像をアフィン変換により回転、拡大縮小変化させた画像との特徴点マッチングを行う。

本実験での特徴点マッチングの利用場面として、画像からの3次元形状復元とSLAMを想定し、2つの設定で実験を行う。

実験1では3次元形状復元を想定する。3次元形状復元においては、対象物体の見え方が異なっても撮影した画像間の復元対象物体の固有の点に対応づることが望ましい。よって回転角度は、0、15、30、45、60、75、90度の広い範囲で、拡大縮小率は、復元対象物体撮影時にカメラのブレで多少拡大縮小率が変化すると考慮し、それぞれの角度に対し、90%、100%、110%に変換した21通りのアフィン変換画像を用いる。元画像に関してはImageNet(ILSVRC2012)データセット[12]の評価用画像から復元対象物体が中心に写り、復元対象物体の周りの背景が複雑でないものを3枚使用する。画像サイズはそれぞれ407×500、500×375、500×354とする。図8に実験1の元画像と変換画像の一例を示す。

実験2ではSLAMを想定する。SLAMについては、前方に写る物体の位置が遠くても近くても正しく物体の固有の点に対応付くことが望ましい。よって回転角度はカメラの外からの衝撃による少々はずれなどを考慮し、0、10、20度の狭い範囲で、拡大縮小率は、それぞれの角度に対し、50%、70%、90%、110%、130%、150%の広い範囲に変換した18通りのアフィン変換画像を用いる。また元画像に関しては、SLAMにCityscapesデータセット[13]から前方に複数の物体が写った、道路の様子を撮影した画像を3枚使用する。画像サイズはすべて2048×1024である。図9に実験2の元画像と変換画像の一部を示す。

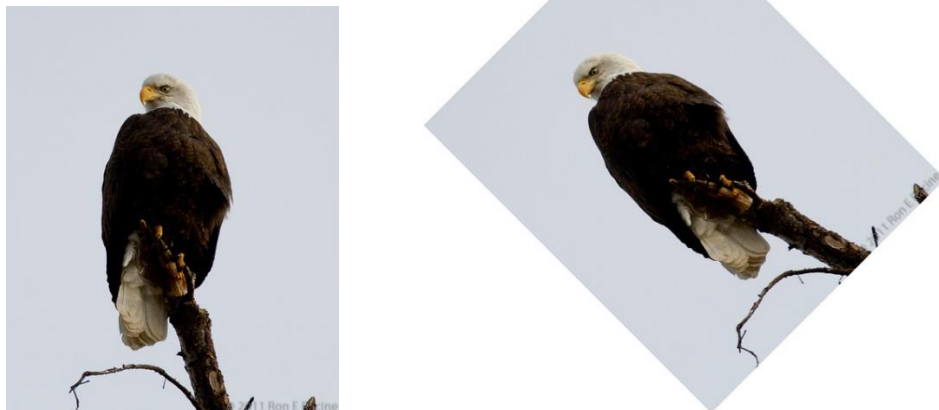


図 8：実験 1 画像例（左：元画像 右：45 度回転画像）



(右) 図 9：実験 2 画像例（左：元画像 右 20 度回転画像）

4.2.2 実験結果

表 2～表 4 に実験 1 での AKAZE、AKAZE-ALEX、SuperPoint それぞれのマッチング率の平均を、表 5～7 にマッチング数の平均を示す。表 8～表 10 に実験 2 での AKAZE、AKAZE-ALEX、Super Point それぞれのマッチング率の平均を、表 11～13 にマッチング数の平均を示す。実験結果において、マッチング率に関しては 90%以上の結果を高精度とし赤字で、マッチング成功数に関しては、3 手法で比較し 1 番数が多いものを青字で記述する。

表 2 : AKAZE のマッチング成功率の平均 (%) (実験 1)

	0°	15°	30°	45°	60°	75°	90°
90%	99.64	98.98	99.37	98.13	99.03	98.68	99.76
100%	100.00	99.48	99.39	99.33	99.23	99.35	100.00
110%	99.32	98.79	99.14	98.36	98.79	99.34	99.26

表 3 : AKAZE-ALEX のマッチング成功率の平均(%) (実験 1)

	0°	15°	30°	45°	60°	75°	90°
90%	100.00	99.42	99.58	99.88	99.70	99.20	99.87
100%	100.00	99.56	99.86	99.90	99.60	99.62	99.91
110%	99.30	99.30	99.10	99.02	99.30	99.45	99.85

表 4 : Super Point のマッチング成功率の平均(%) (実験 1)

	0°	15°	30°	45°	60°	75°	90°
90%	99.51	98.90	98.99	98.90	97.31	98.17	98.47
100%	100.00	99.53	98.61	98.36	98.69	97.67	98.00
110%	99.74	98.41	97.90	97.68	97.11	97.35	98.14

表 5 : AKAZE のマッチング成功数の平均(個) (実験 1)

	0°	15°	30°	45°	60°	75°	90°
90%	220	222	221	217	220	222	219
100%	312	266	270	267	270	271	296
110%	267	255	258	252	257	257	265

表 6 : AKAZE-ALEX のマッチング成功数の平均 (個) (実験 1)

	0°	15°	30°	45°	60°	75°	90°
90%	196	199	195	197	193	195	191
100%	311	239	239	232	232	236	257
110%	222	197	198	193	198	197	218

表 7 : SuperPoint のマッチング成功数の平均 (個) (実験 1)

	0°	15°	30°	45°	60°	75°	90°
90%	394	376	351	329	320	328	320
100%	553	434	398	379	371	360	364
110%	467	440	392	392	369	370	385

表 8 : AKAZE のマッチング成功率の平均 (%) (実験 2)

	0°	10°	20°
50%	95.38	86.47	83.83
70%	97.07	93.07	91.03
90%	98.45	96.70	96.71
110%	98.50	97.45	97.09
130%	96.58	95.69	95.10
150%	95.94	94.50	94.40

表 9 : AKAZE-ALEX のマッチング成功率の平均 (%) (実験 2)

	0°	10°	20°
50%	52.05	50.53	46.81
70%	89.96	88.60	86.97
90%	99.34	99.06	99.09
110%	99.31	98.48	98.35
130%	88.81	85.79	85.54
150%	61.29	53.27	53.25

表 10 : SuperPoint のマッチング成功率の平均 (%) (実験 2)

	0°	10°	20°
50%	48.24	38.32	33.07
70%	92.75	90.54	86.31
90%	98.88	98.20	96.90
110%	98.68	97.77	97.30
130%	94.83	93.17	92.63
150%	79.30	80.69	79.16

表 11 : AKAZE のマッチング成功数の平均 (実験 2)

	0°	10°	20°
50%	746	718	711
70%	1250	1224	1203
90%	1741	1682	1680
110%	2038	1973	1939
130%	1881	1856	1841
150%	2041	1998	1983

表 12 : AKAZE-ALEX のマッチング成功数の平均 (実験 2)

	0°	10°	20°
50%	95	116	106
70%	637	676	666
90%	1566	1539	1522
110%	1724	1557	1532
130%	898	729	720
150%	316	223	219

表 13 : SuperPoint のマッチング成功数の平均 (実験 2)

	0°	10°	20°
50%	401	323	242
70%	2991	2867	2215
90%	6012	5443	4604
110%	7026	5840	5486
130%	6510	5554	5268
150%	4442	3943	3634

結果として実験 1 においては、3 手法すべてにおいて、マッチング成功率は 97~100%と高精度であった。マッチング数に関しては、3 手法の 21 個のマッチングパターンすべてにおいて SuperPoint が 1 番多くのマッチング数を示した。また、3 手法それぞれのマッチング数において、21 パターンの平均を示すと、AKAZE が約 252 個、SuperPoint が約 385 個、AKAZE-ALEX が約 215 個となり SuperPoint が最も多かった。3 次元形状復元においては、復元対象物体の固有の特徴点が、見え方が違っても正しく対応付き、正しいマッチング数が多いことが重要であるので、3 次元形状復元においては、SuperPoint が特に優れているといえる。

実験 2 においては、3 手法において 50%、150%のマッチング成功率の平均をとった際、AKAZE が 91.75%、SuperPoint が 59.80%、AKAZE-ALEX が 52.87%となり、拡大縮小率の変化の度合いが大きいところ（拡大縮小率 50%,150%など）に関しては、AKAZE がマッチング成功率に関して特に高精度であった。SLAM については、前方に写る物体の位置が遠くても近くても正しく物体の固有の点が対応付くことが望ましい。このことから、SLAM においては拡大縮小に強い AKAZE が優れているといえる。

4.3 照明変化

4.3.1 実験方法

本実験では、定点カメラで撮った元画像に対し自然に明るさを変化させた画像セット (Hpatches[14]) の中から 3 パターンの画像(図 10~図 12)を用いる。それぞれ図中の左上の 1 の画像を元画像とし、1 と 2、1 と 3 のように 5 つのマッチングを 3 パターン分行う。おおまかな傾向として比較画像の数字が大きいくほど照明変化が大きくなっている。画像サイズはそれぞれパターン 1 が 640×480、パターン 2 が 1280×700、パターン 3 が 800×600 となっている。



図 10：使用画像（パターン 1）



図 11：使用画像（パターン 2）



図 12：使用画像（パターン 3）

4.3.2 実験結果

表 14～表 16 にマッチング成功率、表 17～表 19 にマッチング数を示す。それぞれの実験パターンにおいてマッチング成功率に関しては、3 手法の中で 1 番良いものを赤字で、マッチング数に関しては 1 番多いものを青字で示す。

表 14：照明条件変化におけるマッピング成功率（％）（パターン 1）

	1and2	1and3	1and4	1and5	1and6	平均
AKAZE	96.94	53.85	15.87	20.00	15.94	40.52
AKAZE-ALEX	92.59	40.91	13.64	37.04	25.00	41.84
SuperPoint	96.00	80.32	57.59	64.56	58.80	71.45

表 15：照明条件変化におけるマッピング成功率（％）（パターン 2）

	1and2	1and3	1and4	1and5	1and6	平均
AKAZE	80.05	68.51	69.23	57.39	25.00	60.04
AKAZE-ALEX	88.98	85.50	68.95	53.42	0.00	59.37
SuperPoint	94.55	90.95	91.87	82.12	43.34	80.57

表 16：照明条件変化におけるマッピング成功率（％）（パターン 3）

	1and2	1and3	1and4	1and5	1and6	平均
AKAZE	84.87	84.35	82.95	76.87	53.03	76.41
AKAZE-ALEX	92.33	90.67	91.71	88.41	81.08	88.84
SuperPoint	96.58	97.22	96.85	96.40	93.85	96.18

表 17：照明条件変化におけるマッチング数（パターン 1）

	1and2	1and3	1and4	1and5	1and6
AKAZE	317	70	10	14	11
AKAZE-ALEX	225	18	3	10	7
SuperPoint	816	253	110	133	127

表 18：照明条件変化におけるマッチング数（パターン 2）

	1and2	1and3	1and4	1and5	1and6
AKAZE	1930	1314	702	233	9
AKAZE-ALEX	1801	1362	322	78	0
SuperPoint	2342	1849	1819	1144	218

表 19：照明条件変化におけるマッチング数（パターン 3）

	1and2	1and3	1and4	1and5	1and6
AKAZE	746	733	715	525	254
AKAZE-ALEX	590	622	653	496	240
SuperPoint	1298	1257	1201	1043	855

結果より、パターン 1 の 1and2 以外の全てのパターンで SuperPoint が他の 2 手法より高精度を出した。マッチング数においてもすべてのパターンで SuperPoint が特に多かった。照明変化においては、光源の位置や、時間帯によつての照明変化により、物体の見え方が異なつてもマッチングに影響しないことが望ましい。よつて、照明変化においては、SuperPoint が特に優れた手法だといえる。

4.4 ノイズ変化

4.4.1 実験方法

本実験では、元画像として Cityscapes データセット [13] から 3 枚を使用する。加えるノイズは「ごま塩ノイズ」とする。ごま塩ノイズは画像上に 2×2 画素の黒色と白色の点をそれぞれ定めた数描画したものである、元画像に対し、ノイズを白と黒合わせて、10000、20000、30000 個発生させたノイズ変化画像をマッチングさせる。使用画像例を図 13、図 14 に示す。



図 13：ノイズ変化画像例（ノイズ 10000 個）

図 14：ノイズ変化画像例（ノイズ 30000 個）

4.4.2 実験結果

表 20 にマッチング成功率、表 21 にマッチング数の平均を示す。実験結果においてマッチング成功率に関しては、3 手法の中で 1 番良いものを赤字で、マッチング数に関しては 1 番多いものを青字で示す。

表 20：ノイズ変化におけるマッチング成功率 (%)

	10000	20000	30000
AKAZE	90.84	78.94	69.04
AKAZE-ALEX	93.55	77.29	55.4
SuperPoint	75.63	58.81	47.2

表 21：ノイズ変化におけるマッチング成功数 (個)

	10000	20000	30000
AKAZE	1803	1445	1203
AKAZE-ALEX	279	60	20
SuperPoint	2068	1314	900

マッチング成功率に関しては AKAZE-ALEX がノイズ総数 10000 個の時には特に高精度であり、30000 個に関しては AKAZE が特に高精度だった。また、AKAZE と AKAZE-ALEX を比較すると、AKAZE が全パターンでマッチング数が多い結果となった。ノイズ変化では、ノイズがかかってもマッチングに影響しないことが望ましい。よって、ノイズ変化においては、AKAZE が優れているといえる。また、SuperPoint はノイズに弱いことも分かった。

4.5 処理時間

4.5.1 実験方法

本実験では 3 手法の特徴点マッチングにおける処理時間を計測する。実験に使用した計算機の仕様を表 22 に示す。AKAZE-ALEX、SuperPoint については、深層学習に関する計算に GPU を用いている。処理時間は、特徴点検出+特徴量記述の時間、マッチング処理の時間それぞれについてすべての場合で 5 回ずつ計測し、その平均処理時間を示している。合計処理時間は特徴点検出+特徴量記述の時間とマッチング処理の時間を足し合わせたものである。また、比較画像の特徴点数も 3 手法それぞれについて示す。4.2.1 節の図 9 のものを元画像とし、実験対象マッチング画像は、元画像と同じ画像とする。

表 22：計算機の仕様

計算機の種類	計算機の仕様
CPU	Intel Corei7 6700K(40GHz)
メモリ	24GBytes
GPU	NVIDIA Geforce 1080(メモリ 8Gbytes)

4.5.2 実験結果

表 23 に 3 手法の各処理時間の平均と比較画像の特徴点検出数を示す。

表 23：処理時間結果、特徴点検出数

	特徴点検出+ 特徴量記述処理 時間 (s)	マッチング処理 時間 (s)	合計処理時間 (s)	特徴点検出(個)
AKAZE	0.459	0.067	0.526	2945
AKAZE-ALEX	23.175	2.83	26.005	2920
SuperPoint	0.619	0.814	1.433	8723

特徴点検出+特徴量記述処理時間においては、AKAZE、SuperPoint、AKAZE-ALEX の順番で処理時間が短かった。3 手法の比較画像の特徴点検出数の結果から、SuperPoint は AKAZE の約 3 倍の特徴点を検出していることが示されている。SuperPoint は、AKAZE よりも検出数が多いことと、検出数の増加に伴う特徴量記述の増加により、AKAZE よりも SuperPoint が特徴点検出+特徴量記述処理時間が長いといえる。また、AKAZE、SuperPoint が検出された点に対し特徴量を与えるのに対し、AKAZE-ALEX は検出された点周りの切り取り画像に対して特徴量を与える。この違いにより AKAZE-ALEX が AKAZE、SuperPoint よりも処理時間が長いといえる。

マッチング処理時間も、AKAZE、SuperPoint、AKAZE-ALEX の順番で処理時間が短かった。マッチング処理時間に違いが出た理由として、マッチングする 2 画像における 2 画像それぞれの特徴点検出数

の違いと、検出された特徴点に与える各手法の特徴量の次元数の違いが挙げられる。マッチングする 2 画像間でそれぞれの画像の特徴点検出数が多ければ多いほど、マッチング処理もその 2 乗に比例し増加する。また、マッチング処理では各特徴点の特徴量から特徴点類似度を計算するが、この計算において、各特徴点に与えられた特徴量の次元数が多いほど計算時間は長くなる。特徴量の次元数は、AKAZE が 61 次元、SuperPoint が 256 次元、AKAZE-ALEX が 4096 次元となっている。このような特徴点検出数の違いと次元数の大小関係がマッチング処理時間に影響したと考える。合計処理時間も、AKAZE、SuperPoint、AKAZE-ALEX の順に短かった。以上の結果より、AKAZE は高速であることが示せた。さらに AKAZE は GPU 処理を必要としない。これより、AKAZE は実時間処理に向くことが示せた。また、AKAZE-ALEX は処理時間は遅く、精度も SuperPoint に劣る場合が多かったため、それほど良い手法とは言えない。

4.6 考察

本実験では AKAZE、AKAZE-ALEX、SuperPoint の 3 つの特徴点マッチング法に対し、マッチング成功率、成功マッチング数、処理時間の値を比較することで、AKAZE-ALEX の課題点の発見、状況にあったマッチング法の使い分けの分析を目的とした。

実験結果を踏まえて、AKAZE-ALEX の課題点としては処理速度が他手法と比較すると極端に遅いこと、拡大縮小に対し、拡大縮小率の度合いが大きいとマッチング成功率が大きく下がることが挙げられる。前者の処理時間に関しては、AKAZE、SuperPoint が検出された固有の特徴点に特徴量を与えるのに対して、AKAZE-ALEX は切り取った画像に対して AlexNet から特徴量を与えるといった手法の違いの影響が大きい。後者のマッチング成功率に関しては比較画像の画像サイズの拡大縮小率が大きく変化すると、元画像の切り取り画像と比較して比較画像の切り取り画像は大きく違ったものになってしまう。AKAZE-ALEX は切り取った画像に対し特徴量を与えその特徴量を比較しマッチングを行うので切り取り画像の見た目に大きく差があると、マッチング率は大きく下がってしまう。

SuperPoint は、3次元形状復元への適用において優れている。3次元形状復元を想定した比較においては、マッチング成功率は 97~100% の高精度でありなおかつ、マッチング数の平均が他の 2 手法と比較して、特に多かった。また、照明変化を想定した比較においては、マッチング成功率は他の 2 手法よりも 1 つを除き全てのパターンで高精度であり、成功マッチング数も他の 2 手法と比較して多かった。3次元形状復元においては、マッチング成功率が高く、成功マッチング数が多く、対象物体撮影時に光の当たり方や時間による照明の変化があってもマッチング成功率が高く成功マッチング数が多いことが望ましい。よって、3次元形状復元においては、SuperPoint が特に優れた手法だといえる。

AKAZE は、SLAM において特に優れている。これは、実験結果より、拡大縮小率の度合いが大きく変化しても他の 2 手法と比較してマッチング成功率が大きく下がらなかったからである。また、実験結果より他 2 手法と比較して AKAZE は処理時間が短い。他にも、ノイズ変化において、ノイズ数が変化しても他の 2 手法と比較してマッチング成功率、成功マッチング数ともに大きく下がることがなかった。SLAM においては、拡大縮小率が変化してもマッチング成功率に影響せず、カメラで撮影した映像にノイズが入っても対象物体を見失わずとらえることができ、かつ処理時間が短いことが望ましい。よって

AKAZE が特に優れた手法だといえる。

5 おわりに

本研究では、AKAZE、AKAZE-ALEX、SuperPointの3つの特徴点抽出法の利用場面を分析して、それぞれの場面で必要とされる性質を検討した上で、評価基準を定め比較を行い、比較実験の結果に基づき、それぞれの手法の特性について考察をした。実験結果より、マッチング成功率が高く、成功マッチング数が多く、対象物体撮影時に光の当たり方や時間による照明の変化があってもマッチング成功率が高く成功マッチング数が多いことから、3次元形状復元においてはSuperPointが適していることが分かった。また、拡大縮小率が変化してもマッチング成功率に影響しづらく、カメラで撮影した映像にノイズが入ってもマッチング成功率が大きく下がらず、かつ処理時間が短いことから、SLAMにおいては、AKAZEが適していることが分かった。前述のような比較実験と考察により、それぞれの手法の特性を示すことができた。

今後の課題として、今回の比較実験の考察を生かし、実際の利用場面に適用した場合どのような結果が出るかを確認することが挙げられる。今回は比較実験による考察のみで、実際の利用場面においてはどのような結果が出るかは確認してない。例えば、3次元形状復元において、3手法を適用し実際に正しく3次元形状復元ができているか、SLAMができているかなどを確かめることができれば、より実際の利用場面においてどのようなマッチング手法が適しているかが分かる。

謝辞

本論文の作成にあたり、丁寧で熱心なご指導を頂いた卒業論文指導教員の椋木雅之教授に深く感謝いたします。指導教員である椋木雅之教授には、実験を行う上でのアドバイスや論文に関する助言やご指導を沢山いただきました。また、研究の相談や議論を通して、多くの知識や示唆を頂いた椋木研究室の皆様にお礼申し上げます。特に、研究に必要なプログラムの作成において多くのアドバイスをいただいた児玉光平さん、森慎太郎さん、小坂弦ノ助さんには深くお礼申し上げます。

参考文献

- [1] D. G. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints”, IJCV, (2004), pp.91-110.
- [2] H. Bay, T. Tuytelaars, L. V. Gool, “SURF: Speeded Up Robust Features”, CVPR, (2008), pp.346-359.
- [3] P. F. Alcantarilla, A. Bartoli, and A. J. Davison. “KAZE features”, ECCV, (2012), pp.214–227.
- [4] P. F, Alcantarilla, J. Nuevo, A.Bartoli, “Fast Explicit Diffusion for Accelerated Features in Nonlinear Scale Spaces”, BMVC, (2013), pp.13.1-13.11.
- [5] Daniel DeTone, Tomasz Malisiewicz, Andrew Rabinovich,” SuperPoint: Self-Supervised Interest Point Detection and Description”(2018), arXiv:1712.07629 (cs).
- [6] 梶原奏美,“深層学習で抽出した特徴量を用いた特徴点マッチング”,H30 年度宮崎大学工学部情報システム工学科卒業論文, (2019).
- [7] 黒河英理子,“深層学習を用いた特徴点マッチングにおける回転拡大縮小不変の実現”,R1 年度宮崎大学工学部情報システム工学科卒業論文,(2020).
- [8] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems 25 (2012): 1097-1105.
- [9] Yi,K.M.and Trulls, E., Lepetit, V.et al,, LIFT: Learned Invariant Feature Transform ECCV(2016), pp. 467–483.
- [10] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, “Orb: An efficient alternative to sift or surf,” in Proc. of ICCV, (2011), pp. 2564-2571.
- [11]pytorch-super point 学習済みモデル
<https://github.com/eric-yyjau/pytorch-superpoint>

[12] ImageNet(ILSVRC2012)データセット

<https://image-net.org/challenges/LSVRC/2012/browse-synsets>

[13] Cityscapes データセット

<https://www.cityscapes-dataset.com/>

[14] Hpatches データセット

<http://icvl.ee.ic.ac.uk/vbalnt/hpatches/hpatches-sequences-release.tar.gz>