AKAZE 特徴点を学習した SuperPoint による特徴点マッチング

小板弦ノ介 [1]　椋木雅之 [1]

概要：本研究では、特徴点抽出手法である SuperPoint において、学習の初期に使用する正解データが、特徴点マッチングにどのような影響を与えるか調査する。SuperPoint は学習データに自動生成した幾何図形の画像（幾何画像）を用いて学習を行なっている。しかし、実際に特徴点検出を適用する対象は、自然画像が多い。幾何画像と自然画像では、特徴点の位置や数が異なり、特徴点マッチングの結果に影響が出ると考えた。そこで、本研究では SuperPoint の学習に幾何画像を用いるのではなく、自然画像を用いた。正解データには AKAZE を適用し自然画像から検出された特徴点を用いた。自然画像を用いることで全体的に平均マッチング成功率を落とすことなく、平均マッチング成功数が増加した。

キーワード：特徴点マッチング、AKAZE、SuperPoint

1. はじめに

2 枚の画像間で、固有の点を対応づける特徴点マッチングは画像からの 3 次元形状復元、類似画像の検索、移動体の自己位置推定と環境地図作成などコンピュータビジョンの多くの分野で重要な役割を担っている。そのため多くの研究がされており、SIFT[1]や SURF[2]、KAZE[3]、AKAZE[4]など多くの手法が提案されている。これらの手法は、画像の回転や拡大・縮小の変化、照明の変化に対して強い耐性あるなど優れた特徴がある。しかし未だに、特徴点マッチングには研究の余地がある。

一方、近年、機械学習の発展手法である深層学習は、コンピュータビジョンの様々な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。深層学習による特徴点マッチングでは、正解データをどのように与えるのかが大きな問題となる。従来、深層学習が適用されていた物体認識や物体検出の問題では、画像中の対象物体の種類や位置、範囲を正解データとして与え、それと同様の出力が得られるように学習が行われる。一方、特徴点マッチングの問題では、画像中のどのような点を特徴点として検出することが正解であるか自明ではなく、人手で正解データを与えることが難しい。LIFT[5]では、SIFT で検出した特徴点の内、正しくマッチングできた対応点のみを選別して正解データとすることで、この問題に対処している。このように得られた正解データを深層学習のネットワークで学習することで SIFT より多くの特徴点を検出し、それらを正しくマッチングできている。

深層学習を用いた特徴点マッチング手法の一つに SuperPoint[6]がある。SuperPoint では、正解データを自動生成しながら 3 段階に学習している。第 1 段階では、自動生成可能な幾何図形の画像（幾何画像）を用いて学習を行う。正解データとなる特徴点としては、幾何図形の端点や交点を用いる。特徴点の位置は、幾何画像の生成時に計算可能であり、正解データを自動生成できる。第 2 段階では、第 1 段階で学習した特徴点検出器を自然画像に適用し、正解データを作成する。まず、与えられた自然画像にホモグラフィ変換という幾何変換を適用した画像を生成する。それらに対して第 1 段階で学習した特徴点検出器を適用し、特徴点を得る。それらの特徴点座標を元画像に逆投影したものを全て合わせて正解データとする。これにより、多くの特徴点を正解データとして自動生成できる上、画像の回転や拡大・縮小といった幾何変換にも対応できる。第 3 段階では、与えられた自然画像とその画像をホモグラフィ変換した画像から正解データを作成する。第 2 段階で学習した特徴点検出器をそれぞれの画像に適用して、特徴点を検出する。適用したホモグラフィ変換は既知なので、逆変換することで 2 つの画像間での特徴点の正しい対応関係がわかる。第 3 段階では、この正しい対応関係を正解データとして、特徴点位置だけではなく対応する特徴点同士が類似した特徴量を持つよう、特徴量記述も同時に学習する。SuperPoint は、このように正解データを段階的に自動生成しながら学習することで、多くの特徴点を検出し、正しくマッチングができる高性能な手法である。そのため、その後の特徴点マッチングに関する研究でも多く参照される手法となっている。

SuperPoint は特徴点マッチングにおいて高い性能を示しているが、このような正解データの与え方が最良であるとは限らない。本研究では、特に学習の第 1 段階に注目した。SuperPoint では、第 1 段階で幾何画像を用いて学習を行っている。しかし、実際に特徴点検出を適用する対象は、自然画像が多い。幾何画像中の幾何図形の端点や交点の位置は明確であるが、そのような点は自然画像中にはあまり現れない。このように、幾何画像と自然画像では、画像や特徴点の性質、数等が異なる。このことが、特徴点マッチングの結果に影響を与えると考えられる。

1 宮崎大学大学院　工学研究科

そこで、本研究では、SuperPoint において、学習の第 1 段階で使用する正解データが、特徴点マッチングにどのような影響を与えるか調査する。具体的には 3 つの観点から実験を行った。1 つ目は、第 1 段階で使用する画像を自然画像とすることの影響調査である。2 つ目は、正解データを評価値の高い特徴点に絞った場合の影響調査である。3 つ目は、LIFT と同様の考え方で、正解データを正しくマッチングできた特徴点に絞った場合の影響調査である。自然画像からの正解データの作成には、AKAZE を使用する。

2. 特徴点マッチング

2.1 特徴点マッチングとは

特徴点マッチングとは、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させることである。特徴点マッチングは特徴点検出、特徴量記述、マッチングの 3 段階からなる。

特徴点検出は画像中から角、線の交わり等の他と異なる固有の点の座標を検出する。特徴量記述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などで表現した値を特徴量として算出する。

マッチングは、特徴点の特徴量を比較し、類似度の高い特徴点同士を対応付けることである。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例を示す。総当たりマッチングでは、画像 1 から検出された各特徴点と画像 2 で検出された全特徴点の特徴量同士の距離を計算し、類似度の高い特徴点同士を対応付ける。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い、双方の結果が一致するものをマッチングしたとみなす。この 2 つを組み合わせることで総当たりマッチングの問題である一つの特徴点に対して複数個の特徴点が対応付くことはなくなり、信頼性の高いマッチング結果を得ることが可能である。

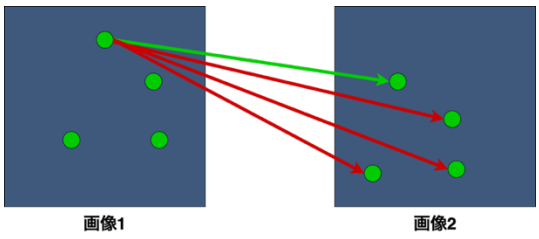

図 1 総当たりマッチング

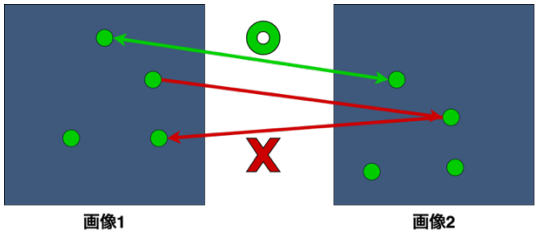

図 2 クロスチェック

2.2 AKAZE

AKAZE は、特徴点検出と特徴量記述を行う手法である。SIFT や SURF の欠点を改善した KAZE をもとにしている。そのため、SIFT や SURF 等の他手法と比較してロバスト性や処理速度の面で優れている。また、AKAZE は回転、拡大・縮小に強い性能がある。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。特徴点検出では、非線形スケールスペースでヘッセ行列を適用し、その行列式の値が極大となる点を求める。これにより、様々なスケールを考慮した上で、他と異なる特徴を持つ点を特徴点として検出できる。また、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の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。

2.3 SuperPoint

SuperPoint[6]は、Tone らによって提案された特徴点検出器である。特徴点の正解データを 3 段階に学習することで、特徴点検出器を作成する。SuperPoint は、特徴点マッチングの 3 つの処理の内、特徴点検出と特徴量記述までを行う。Tone ら[6]の論文内で、マッチング精度に関して、既存の SIFT[1]、LIFT[5]、ORB[7]より高い精度を示している。以下、SuperPoint の詳細について述べる。

2.3.1 SuperPoint のネットワーク構造

SuperPoint は完全な畳み込みニューラルネットワークである。画像が与えられると、画像中の特徴点位置とその特徴点のもつ特徴量を出力する。SuperPoint のネットワーク構造を図 3 に示す。

SuperPoint はエンコーダ・デコーダ構造を持つ。エンコーダは、VGG ネットワーク[8]に似た構造を持つ。このエンコーダにより、入力画像の縦横 1/8 のサイズの特徴マップを出力する。デコーダは、特徴点の位置を出力する Interest Point Decoder と特徴点の特徴量を出力する Descriptor Decoder の 2 つで構成される。エンコーダで得られた特徴マップは、これら 2 つのデコーダに入力され、それぞれ処理される。Interest Point Decoder では、Softmax と reshape 処理で元の画像サイズでの特徴点位置が出力される。Descriptor Decoder では共 3 次補完と L2-正規化で特徴点位置での特徴量が出力される。

SuperPoint のネットワーク構造から、Descriptor Decoder を削除したものを MagicPoint と呼ぶ。MagicPoint では、特徴量記述は行わず、特徴点検出のみが行える。SuperPoint の学習では、まず、特徴点検出を行う MagicPoint を学習した後、それを利用して段階的に特徴点検出の学習を行い、最後に、特徴量記述も含めた SuperPoint の学習を行う。

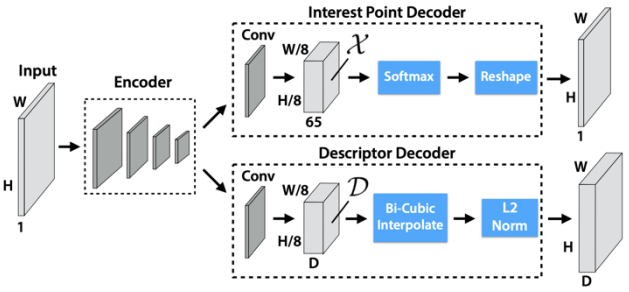

図 3　SuperPoint のネットワーク構造[6]

2.3.2 SuperPoint の学習

SuperPoint では、正解データを自動生成しながら、以下の 3 段階で学習を行っている。

Step1 幾何画像を用いた MagicPoint の初期学習

Step2 自然画像から生成した正解データによる MagicPoint の改良学習

Step3 自然画像の正解対応点による SuperPoint の特徴点検出と特徴量記述の学習

Step1 では、単純な幾何図形の写った幾何画像を自動生成し、その中の幾何図形の端点や交点を正解データとして特徴点検出の学習を行う。この際、特徴量記述の学習を行わないので、ネットワーク構造としては SuperPoint から Descriptor Decoder を削除した MagicPoint を用いる。Step2 では、Step1 で学習済みの MagicPoint を自然画像に適用し、正解データを作成する。与えられた自然画像を様々にホモグラフィ変換した画像を生成し、それらの画像から MagicPoint で得られた特徴点を合成して正解データとする。この処理を Homographic Adaptation と呼ぶ。Homographic Adaptation で得られた正解データを MagicPoint で学習する。この処理を 2 回繰り返すことで、MagicPoint の特徴点検出性能を向上させる。Step3 では、自然画像とその画像をホモグラフィ変換した画像に Step2 で学習した MagicPoint を適用し、正しく対応する特徴点の組を抽出して正解データとする。この正解データを用いて SuperPoint を学習することで、特徴点検出と特徴量記述を同時に学習する。

特徴点検出の学習では、正解データの特徴点座標と周囲の座標での出力値のクロスエントロピーを損失関数とすることで、正解データの特徴点座標のみで出力値が高くなるよう学習する。特徴量記述の学習では、対応する特徴点の特徴量同士は同じ値をもち、対応しない特徴点同士では異なる値（特徴量ベクトルが直交する方向）になるよう学習する。以下の節で、学習の各段階についての詳細を述べる。

2.3.3 幾何画像を用いた MagicPoint の初期学習

学習の第 1 段階では、特徴点検出の学習を行う。特徴点検出の学習は、SuperPoint から特徴量記述のためのデコーダである Descriptor Decoder を削除したネットワーク構造（MagicPoint）に対して行う。

学習データには、自動生成した幾何図形が写った幾何画像と幾何図形の端点や交点を特徴点とした座標（正解データ）の組を用いる。幾何画像の例を図 4 に示す。幾何図形には、四角形、三角形、線、立方体、チェッカーボードなどがある。図中の緑の点が正解データとなる特徴点である。幾何画像と特徴点座標からなる学習データは、様々な見え方のものを自動で無数に生成することができる。幾何画像が入力された時、正解データの特徴点座標が出力されるように、MagicPoint の学習が行われる。

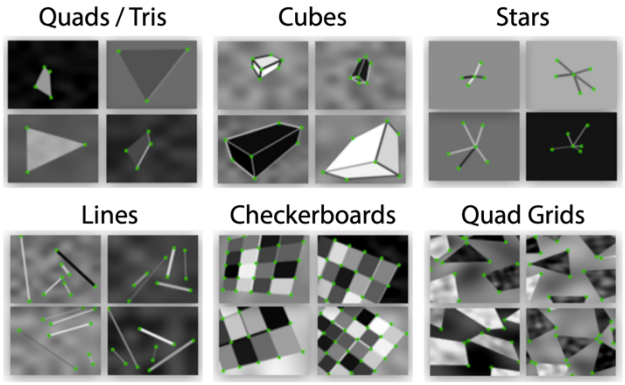

図 4 幾何画像の例[6]

2.3.4 自然画像から生成した正解データによる MagicPoint の改良学習

学習の第 2 段階では、自然画像を用いて特徴点検出の性能改善を行う。そのために、自然画像をホモグラフィ変換した画像を利用する。ホモグラフィ変換は 8 自由度の幾何変換で、3 次元空間上の 2 次元平面同士の間の見え方の変換を記述できる。回転や拡大・縮小、平行移動といった 6 自由度のアフィン変換に加え、奥行き方向の縮小も表現できる。

ホモグラフィ変換した自然画像に対して第 1 段階で学習済みの MagicPoint を適用して、特徴点の正解データを作成する。この正解データの作成処理のことを Homographic Adaptation と呼ぶ。Homographic Adaptation の処理の流れを図 5 に示す。まず、自然画像の大規模なデータセット内の画像に対し複数パターンのランダムなホモグラフィ変換を行う。全てのパターンに対し MagicPoint を適用して特徴点検出を行う。検出した特徴点に対して、逆変換をかけ、元画像上での特徴点座標を求める。全てのパターンで検出された特徴点の元画像上での座標を合成したものを正解データとして MagicPoint で学習をする。これにより、元の学習済み MagicPoint よりも多くの特徴点を学習でき、結果としてより多くの特徴点を検出できるようになる。また、種々のホモグラフィ変換に対してロバストに安定して特徴点を検出できるようになる。この処理を 2 回行い、MagicPoint の特徴点検出性能を向上させる。

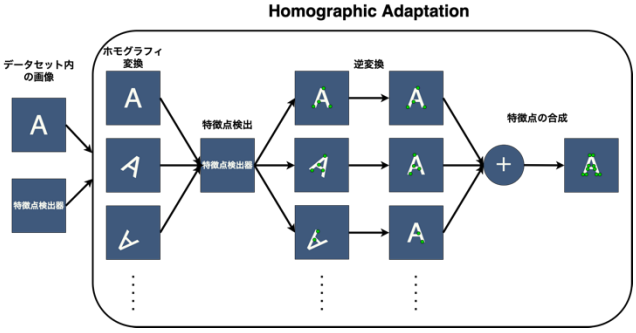

図 5 Homographic Adaptation

2.3.5 自然画像の正解対応点による SuperPoint の特徴点検出と特徴量記述の同時学習

学習の第 3 段階では、3 画像間の特徴点の対応関係を正解データとして特徴点検出と特徴量記述の学習を行う。この学習では、SuperPoint の 2 つのデコーダを用いる。第 3 段階の学習の流れを図 6 に示す。

学習データの生成には、自然画像の大規模な画像データセットを用いる。まず、元画像となる自然画像に対して、ランダムなホモグラフィ変換を適用し、Warp 画像を作成する。次に、元画像と Warp 画像に第 2 段階で学習済みの MagicPoint を適用し、それぞれの画像から特徴点を検出する。この際、適用したホモグラフィ変換は既知なので、Warp 画像の特徴点座標を逆変換することで 2 画像間での特徴点の対応関係がわかる。このようにして得られた元画像と Warp 画像、それぞれの画像の特徴点座標、2 画像間の特徴点の対応関係を学習データとして用いる。特徴量記述の学習は、2 画像間で特徴点が対応関係にある場合、それらの特徴点の特徴量同士が近い値になるよう行う。

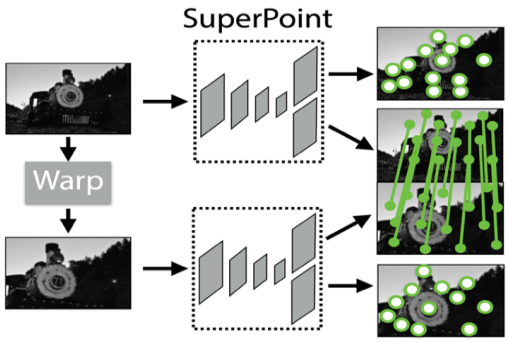

図 6 第 3 段階の学習[6]

3. AKAZE 特徴点を学習した SuperPoint の評価

3.1 自然画像による SuperPoint の学習

SuperPoint では、学習の第 1 段階で幾何画像を用いて MagicPoint を学習している。幾何画像を用いる利点は、大きく 2 つある。1 つ目は、自動で無数の幾何画像を生成できること、2 つ目は正解データである特徴点座標も自動生成できることである。これらによって、MagicPoint を自動で学習することが可能である。しかし、このような正解データの与え方が最も良いとは限らない。実際に特徴点検出を適用する対象は自然画像が多い。自然画像中には幾何図形の端点や交点のような明確な特徴点はあまり現れない。逆に、自然画像では不規則な濃度変化が多く存在する。このような濃度変化が他の点との違いとなり得るため、それをヒントに特徴点検出が行える。このように、幾何画像と自然画像では、画像や特徴点の性質、数等が異なる。自然画像を対象に特徴点マッチングを行う場合、SuperPoint の学習の第 1 段階から自然画像を用いた方が、自然画像の性質に合った特徴点をより多く検出でき、結果的にマッチングの性能も向上できる可能性がある。

そこで本研究では、SuperPoint において、学習の第 1 段階で使用する正解データが、特徴点マッチングにどのような影響を与えるのか調査する。具体的には 3 つの観点から実験を行う。1 つ目は第 1 段階で使用する画像を自然画像とすることの影響調査である。2 つ目は、正解データを評価値の高い特徴点に絞った場合の影響調査である。3 つ目は、LIFT と同様の考え方で、正解データを正しくマッチングできる特徴点に絞った場合の影響調査である。以下、調査手順の詳細を述べる。

3.2 自然画像による SuperPoint の調査手順

本研究では、以下の手順で SuperPoint の学習を行う。

Step1 自然画像を用いた MagicPoint の初期学習
Step2 自然画像から生成した正解データによる MagicPoint の改良学習
Step3 自然画像の正解対応点による SuperPoint の特徴点検出と特徴量記述の同時学習

SuperPoint では Step1 で幾何画像を用いていたが、本研究では自然画像を用いて学習を行う。この自然画像に AKAZE を適用して特徴点を検出する。この自然画像のデータセットと正解データである AKAZE 特徴点を MagicPoint の学習データとして用いる。

幾何画像を学習に用いる利点として、無数の画像を自動生成できる点が挙げられる。自然画像を自動生成することは困難であるが、近年では大規模な自然画像のデータセットが提供されており、学習に用いるのに十分な大量の画像が利用できる。幾何画像を学習に用いるもう一つの利点は、正解データとなる特徴点座標も自動生成できる点である。これに対しては、既存の特徴点検出手法である AKAZE を利用することで自動化する。AKAZE は、深層学習を用いない画像処理ベースの特徴点検出、特徴量記述手法である。画像の回転や拡大・縮小等の影響を受けにくく、自然画像に対して安定して多数の特徴点を検出できる。特徴点マッチングの性能は、近年の深層学習を用いた手法に及ばない面もあるが、MagicPoint の初期学習のために、自然画像から多数の特徴点を抽出する目的には適している。

Step2、Step3 は SuperPoint と同じ処理を行う。これらの処理では Step1 と同じ自然画像のデータセットを用いる。

学習に用いるデータの影響を調査するために、データセットや正解データを変えながら評価実験を行う。

3.3 学習データの違いによる SuperPoint の検出性能の評価

本研究では、SuperPoint において、学習に用いる正解データが特徴点マッチングにどのような影響を与えるのか調査する。具体的には、3 つの観点から調査を行う。それぞれの調査について述べる。

3.3.1 学習データを自然画像とすることの影響調査

この実験では、SuperPoint の学習の第 1 段階で使用する画像を自然画像にすることによってどのような影響があるのか調査する。そのために、自然画像のデータセットとして、大規模で多様な自然画像のデータセットを用いる。正解データの作成には、AKAZE を用いる。比較対象は幾何画像を学習した SuperPoint 及び AKAZE とする。

自然画像を学習することで、幾何画像を学習した SuperPoint より自然画像に合ったより多くの特徴点を検出できると考えられる。また、AKAZE と比較して、より適切な特徴量記述を学習できるため、特徴点検出の精度が向上すると考えられる。

3.3.2 正解データを評価値の高い特徴点に絞った場合の影響調査

この実験では、3.3.1 節の実験において、正解データとなる特徴点を評価値の高いもののみに絞った場合の影響を調査する。AKAZE により自然画像から正解データを作成する際に閾値を定め、特徴点の評価値が閾値よりも高いもののみを正解データに加える。評価値としては、AKAZE による特徴点検出時の反応値（ヘッセ行列の行列式の値）を用いる。反応値の高い点ほど、AKAZE の特徴点検出に強く反応しており、AKAZE による評価観点でより明確な特徴点と言える。評価値が閾値未満の点を除くので、特徴点の数は減るが、明確な特徴点のみを正解データとすることができる。使用するデータセットは、3.3.1 節の実験と同じである。

3.3.1 節の実験と比較して、正解データの数が減るため検出される特徴点数は減少するが、特徴点マッチングの精度は向上すると考えられる。

3.3.3 正解データを正しくマッチングできる特徴点に絞った場合の影響調査

この実験では、対象とする自然画像をある程度限定して、その種の自然画像において特徴点マッチングがうまく行えた特徴点を正解データとすることの影響を調査する。データセットとしては、「道路シーン」のような類似した自然画像を用いる。正解データの作成では、元画像と元画像をアフィン変換した画像間で AKAZE を用いて特徴点マッチングを行い、正しくマッチングが行えた特徴点のみを使用する。

AKAZE で検出した全ての特徴点を正解データとして使用する場合に比べ、自然画像の特定のシーンにより適応した特徴点検出が行えることで、特徴点マッチングの精度が特定シーンに対して向上すると考えられる。

4. 実験

4.1 実験 1

この実験では、SuperPoint の学習の第 1 段階で使用する画像を自然画像にすることでどのような影響があるのか調査する。

4.1.1 実験内容

まず、学習用画像と検証用画像を用意する。本実験では、COCO dataset[9]から 42000 枚の学習用画像と 2000 枚の検証用画像を 640×480 画素の大きさに変更して用いる。これらの画像に対して、AKAZE を適用して特徴点検出を行う。次に、画像を 160×120 画素の大きさに変更する。それに合わせて、特徴点の座標も 160×120 画素の大きさに合わせて変更する。座標を変更する方法は以下の式(1)(2)を用いた。

$$\text{変更後の}x\text{座標} = \frac{\text{変更前の}x\text{座標}}{\text{元画像の幅}} \times \text{変更後の画像の幅} \quad (1)$$

$$\text{変更後の}y\text{座標} = \frac{\text{変更前の}y\text{座標}}{\text{元画像の高さ}} \times \text{変更後の画像の高さ} \quad (2)$$

この学習用画像と学習用特徴点の組と検証用画像と検証用特徴点の組が SuperPoint の第 1 段階の学習である MagicPoint の学習用データと検証用データとなる。MagicPoint では学習回数を 200000 回、学習率は 0.001 とした。

SuperPoint の第 2 段階の学習では第 1 段階の学習でも用いた学習用画像と検証用画像を 320×240 画素の大きさに変更して使用する。Homographic Adaptation では、100 個のランダムなホモグラフィ変換を行い特徴点の検出を行う。検出された特徴点を正解データとして MagicPoint を学習する。この処理を 2 回行う。

SuperPoint の第 3 段階の学習では、第 2 段階と同じ学習用画像と検証用画像（320×240 画素）を用いる。学習回数を 200000 回、学習率は 0.0001 とした。以降、この手法を「提案手法」と呼ぶ。

比較手法として、SuperPoint の第 1 段階の学習で幾何画像を用いた手法（以下、「従来手法」と呼ぶ）と AKAZE を用いた手法（以下、「AKAZE 手法」と呼ぶ）を用いる。従来手法では、第 1 段階の学習で、学習用画像として 90000 枚、検証用画像として 1800 枚の幾何画像を生成して利用する。第 2 段階、第 3 段階の学習では、COCO dataset から 80000 枚の学習用画像と 40000 枚の検証用画像を利用する。その他の設定は、提案手法と同じである。AKAZE 手法は、OpenCV4.5.5 を利用して実装した。各種設定は、デフォルトのままにした。

次に、評価方法について説明する。まずテスト用画像を

50 枚用意する。学習データにも使用した COCO dataset から学習用に使用していない画像を選び、画像サイズを 640×480 画素の大きさにして用いる。次に、テスト用画像に対して、アフィン変換で回転や拡大縮小変化させたテスト用変換画像を作成する。テスト用画像を 10°、30°、45°、60°、90°に回転させ、それぞれ 50%、80%、100%、120%、150%の大きさに変換した 25 通りをテスト用変換画像とする。

テスト用画像とテスト用変換画像を用いて、提案手法と従来手法、AKAZE 手法を適用し、特徴点検出、特徴量記述を行う。その後、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によりマッチング結果を得る。テスト用変換画像を生成した際のアフィン変換は既知なので、マッチング結果で正しい対応点が得られたかは、逆変換を適用することで判定できる。具体的には、マッチング結果において、テスト用画像のi番目の特徴点Q_iとテスト用変換画像のj番目の特徴点P_jが対応している時、P_jに逆変換を適用して得られる点P_j'が式(3)を満たすとき、正しい対応点とする。

$$\left|Q_i - P_j'\right| < T \quad (3)$$

ここでTは許容座標誤差である。本研究では、$T = 10$画素とする。

評価には、平均マッチング成功率と平均マッチング成功数を用いる。マッチング成功率は、テスト用画像とテスト用変換画像の組毎に、式(4)で求める。

$$\text{マッチング成功率} = \frac{\text{正しい対応点の数}}{\text{得られた対応点の数}} \quad (4)$$

平均マッチング成功率と平均マッチング成功数は式(5)(6)で求める。

$$\text{平均マッチング成功率} = \frac{\text{マッチング成功率の総和}}{\text{テスト画像の数}} \quad (5)$$

$$\text{平均マッチング成功数} = \frac{\text{マッチング成功数の総和}}{\text{テスト画像の数}} \quad (6)$$

4.1.2 実験結果

実験結果を図 7、図 8 に示す。図 7 は横軸を拡大縮小率、奥行きを回転角度、縦軸を平均マッチング成功率とする。図 8 は横軸を拡大縮小率、奥行きを回転角度、縦軸を平均マッチング成功数とする。

提案手法と従来手法の平均マッチング成功率を比べると、拡大縮小率が 80%から 120%の間では、いずれの回転角度でも提案手法の方が高かった。しかし、その差は 1 ポイント程度で、性能として大きな違いは見られなかった。一方、平均マッチング成功数では、拡大縮小率が 80%から 150%の間で、提案手法の方が多かった。従来手法に比べ、およそ 2 倍程度の成功数となっている。提案手法では、学習の第 1 段階で自然画像を用い、AKAZE を適用して特徴点抽出を行っている。これにより、幾何画像を使う従来手法より多くの特徴点を利用できる。この影響により、最終的な学習済みの SuperPoint でも、提案手法の方が自然画像から多くの特徴点を抽出できたと考えられる。

提案手法と AKAZE 手法を比べると、平均マッチング成功率、平均マッチング成功数とも、50%縮小の際は、AKAZE 手法の方が良い結果を示している。SuperPoint の第 3 段階の学習では、適用するホモグラフィ変換が極端な変形とならないよう制限している。50%縮小は、極端な変形となるため、対応していないことが原因と考えられる。

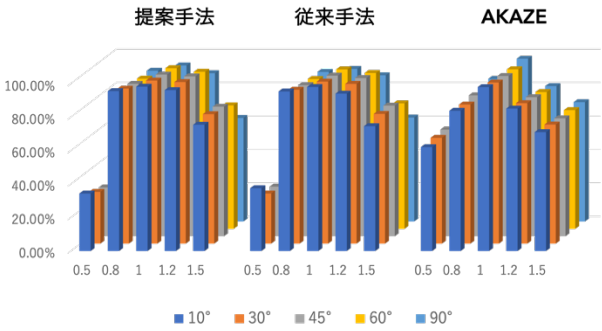

図 7 平均マッチング成功率(%)

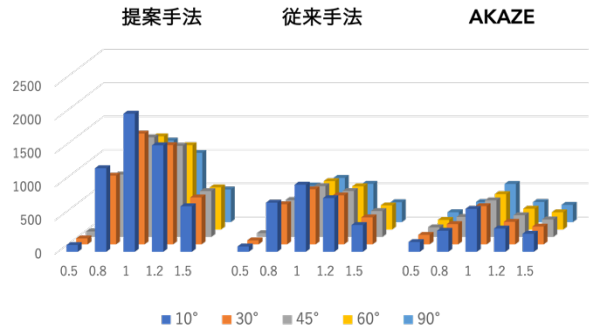

図 8 平均マッチング成功数(個)

4.2 実験 2

この実験では、提案手法の学習の第 1 段階において、正解データとなる特徴点を評価値の高いもののみに絞った場合の影響を調査する。

4.2.1 実験内容

学習用画像と検証用画像は実験 1 と同じものを用いる。これらの画像に対して、AKAZE を適用して特徴点検出を行う。次に、閾値を設定して学習用画像から検出された各特徴点の評価値と比較し閾値よりも高い特徴点を正解データとして使用する。評価値には、AKAZE の反応値（OpenCV で得られる response の値）を用いる。閾値は 0.005、0.01 の 2 通り行なった。得られた画像あたりの正解データの平均個数を表 1 に示す。

表 1 得られた正解データの平均個数

	学習用データ	検証用データ
閾値なし	907.36	915.72
閾値 0.005	197.76	197.86
閾値 0.01	65.79	64.10

その他の設定については実験 1 と同様である。評価方法については、4.1.1 節と同様の評価を行った。

4.2.2 実験結果

実験結果を図 9、図 10 に示す。図 9 の座標軸は図 7 と同様である。図 10 の座標軸は図 8 と同様である。

平均マッチング成功率に注目すると、閾値 0.005 で最良の場合が多いが、閾値 0.01 や閾値なしと比較して大きくは向上していない。閾値により AKAZE の評価観点でより明確な特徴点のみを学習に用いることで、マッチングがしやすい特徴点抽出が行え、マッチング成功率が向上すると予想していたが、そのような傾向は見られなかった。一方、平均マッチング成功数に注目すると、閾値なしが最も多く、次に閾値 0.005、最少が閾値 0.01 となっている。これは、学習に用いた特徴点の数に対応している。学習の第 1 段階でより多くの特徴点を用いることで、最終的に得られる学習済みの SuperPoint でより多くの特徴点を検出できるようになると考えられる。しかし、学習の第 1 段階で使用した特徴点の数の差に比べて、平均マッチング成功数の差は小さい。これは、学習の第 2 段階で Homographic Adaptation により検出できる特徴点数を増やしている効果と考えられる。

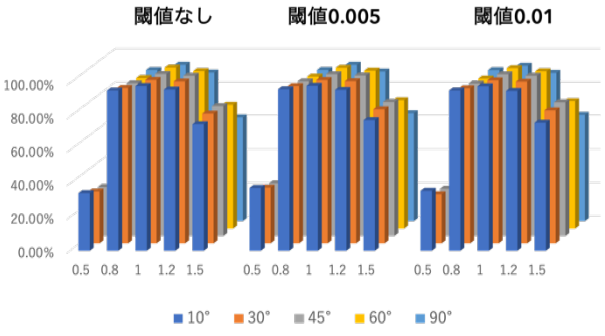

図 9 平均マッチング成功率(%)

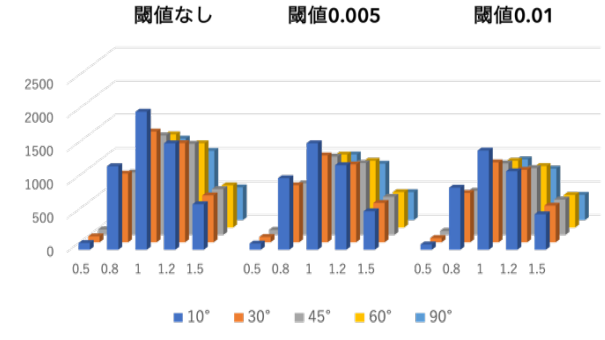

図 10 平均マッチング成功数(個)

4.3 実験 3

この実験では、学習データを特定のシーンに限定することで、そのようなシーンに適用した学習が行えるか調査する。

4.3.1 実験内容

本実験では、特定のシーンの画像データセットとして、Cityscapes Dataset[10]を用いる。このデータセットは、車に搭載されたカメラから進行方向を撮影した動画を数フレーム毎に抽出して作成したもので、様々な道路シーンが含まれている。一方で、道路の路面上で道路の方向に沿って撮影されている、撮影時の高さはフロントガラス程度、街中の道路を撮影しており周囲は主にビルや街路樹であるといった共通した性質を持っている。本実験では、Cityscapes Dataset の計 5000 枚から 4400 枚の学習用画像と 500 枚の検証用画像を 1024×512 画素の大きさに変更して用いる。次に、学習用画像に対して、アフィン変換で回転させた学習用変換画像を作成する。学習用画像を 30°、45°、60°、90°に回転させた 4 通りのデータセットを作成する。学習用画像と学習用変換画像に対して、AKAZE を適用して特徴点検出、特徴量記述及びマッチングを行う。マッチングに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。学習用変換画像は、適用した変換が既知であるため、各特徴点は逆変換により学習用画像での位置を計算できる。そのため、得られたマッチング結果が正しいか判断できる。各学習用画像において、マッチングが正しく行えた AKAZE 特徴点の座標を求める。検証用画像も学習用画像同様に行う。4 通りの回転それぞれのデータセットについて、得られた特徴点を正解データとして SuperPoint の第 1 段階の学習を行う。第 2 段階、第 3 段階の学習でも、同じ学習用画像と検証用画像を用いる。その他の学習時の設定は、実験 1 と同様である。

評価時にも、Cityscapes Dataset を用いる。Cityscapes Dataset から学習時に使用していない画像 50 枚をテスト用画像とする。評価方法に関するその他の設定は、実験 1 と同様である。

4 通りの回転それぞれのデータセットを用いた学習結果をそれぞれ「提案手法 30°」「提案手法 45°」「提案手法 60°」「提案手法 90°」と呼ぶ。比較手法として、第 1 段階の学習に Cityscapes Dataset を用いて実験 1 と同様に学習した結果（以下、「提案手法選別なし」と呼ぶ）、実験 1 と同じ従来手法、AKAZE 手法を用いる。従来手法は、実験 1 と同じ COCO dataset で学習したものである。

4.3.2 実験結果

本稿では、実験結果として、提案手法選別なし、提案手法 45°、提案手法 60°、従来手法の 4 通りを示し、拡大縮小率 100%の特徴点マッチングの結果を示す。実験結果を図 11、図 12 に示す。図 11 は横軸を回転角度、縦軸を平均マッチング成功率とする。図 12 は横軸を回転角度、縦軸を平均マッチング成功数とする。

平均マッチング成功率に注目すると、提案手法 60°が最良となっているが、他の手法と比較して、大きな向上は見られない。また、例えば提案手法 60°が、60°回転の画像に対してより平均マッチング成功率が向上するといった傾向も見られない。さらに、COCO dataset で学習した従来手法

と比較しても、大きな向上は見られない。本実験では、第 1 段階で使用する学習データを特定のシーンに絞ることで、その特定シーンに適応して平均マッチング成功率が向上すると想定して実施したが、実験結果からはそのような傾向は見られなかった。

平均マッチング成功数に注目すると、提案手法 45°が最多となっている。4 通りの回転で作成したデータセットでは、マッチングに成功した特徴点のみを利用している。そのため、提案手法選別なしに比べて学習時の特徴点の数が少なくなっている。それにも関わらず、提案手法 45°で最多となっている。学習時の特徴点数と学習済みの SuperPoint で検出できる特徴点数の関係について、より深く検討する必要がある。一方、提案手法と従来手法を比較すると全般に提案手法の方が、平均マッチング成功数が多くなっている。こちらは、学習時に特定シーンの画像を使ったことの効果が現れたと考えられる。

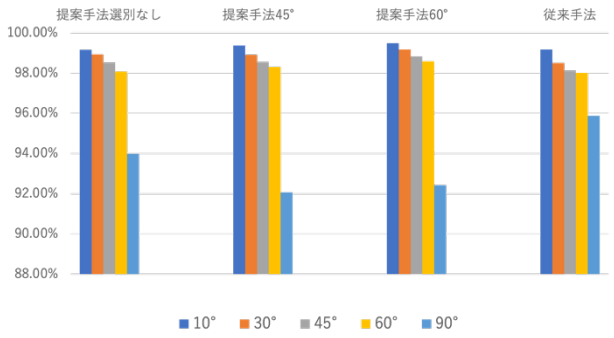

図 11 平均マッチング成功率(%)

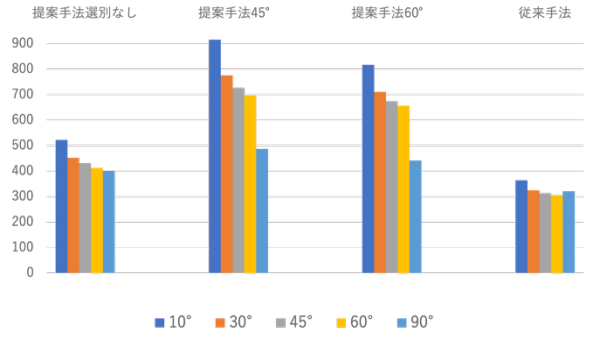

図 12 平均マッチング成功数(個)

5. おわりに

本研究では、SuperPoint において、学習に用いる正解データが特徴点マッチングにどのように影響するのかについて 3 つの観点から調査した。1 つ目は、SuperPoint の学習の第 1 段階で使用する画像を自然画像にすることの影響調査、2 つ目は、提案手法の第 1 段階において、正解データとなる特徴点を評価値の高いもののみに絞った場合の影響調査、3 つ目は、学習データを特定のシーンに限定することで、そのようなシーンに適用した学習が行えるか調査である。

学習データを自然画像にすることで、自然画像に合ったより多くの特徴点を正解データとして SuperPoint に適用できるため、平均マッチング成功率を大きく落とすことなく、平均マッチング成功数が大幅に増加した。正解データを評価値の高い特徴点に絞った場合は、はっきりとした特徴点のみを正解データとして SuperPoint に適用できるため、平均マッチング成功数は減少し、平均マッチング成功率は増加すると想定していた。しかし、平均マッチング成功数は減少したものの、平均マッチング成功率は従来手法の SuperPoint と大差はなかった。正解データを正しくマッチングできる特徴点に絞った場合は、平均マッチング成功率は従来手法の SuperPoint と大差はなかったが、平均マッチング成功数は増加した。全体を通じて、学習に用いる特徴点の数が増えると、平均マッチング成功数が増える傾向が見られたが、平均マッチング成功率には大きな変化がなかった。

今後の課題として、学習データの入力画像のサイズを変更することの影響調査が挙げられる。本研究では、学習データの入力画像のサイズは 160×120 としていたが、入力画像のサイズをより大きなサイズにすることでより多くの特徴抽出が可能となり、特徴点マッチングの結果に影響すると考えられる。第 2 段階の学習や第 3 段階の学習においても入力画像のサイズをより大きなサイズに変更することで特徴点マッチングの結果にどのような影響があるか検討する必要がある。また、本研究では AKAZE を用いた学習データの特徴点検出しか行えていない。他の特徴点検出手法を適用して SuperPoint の学習データとして学習することで、特徴点マッチングの結果に影響すると考えられる。

参考文献

[1] David G. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints”, IJCV, (2004), vol60, pp.91-110.
[2] Herbert Bay, Tinne Tuytelaars, Luc Van Gool, “SURF: Speeded Up Robust Features”, CVPR, (2008), pp.346-359.
[3] Pablo F. Alcantarilla, Adrien Bartoli, Andrew J. Davison, “KAZE features”, ECCV, (2012), pp.214-227.
[4] Pablo F. Alcantarilla, Jesús Nuevo, Adrien Bartoli, “Fast Explicit Diffusion for Accelerated Features in Nonlinear Scale Spaces”, BMVC, (2013), pp.13.1-13.11.
[5] YI Kwang Moo, et.al., “Lift: Learned invariant feature transform”, ECCV, (2016), pp. 467-483.
[6] Daniel DeTone, Tomasz Malisiewicz, Andrew Rabinovich, “SuperPoint: Self-Supervised Interest Point Detection and Description”, CVPR, (2018), pp.337-349.
[7] RUBLEE Ethan, et.al., “ORB: An efficient alternative to SIFT or SURF”, ICCV, (2011), pp. 2564-2571.
[8] Karen SIMONYAN, Andrew ZISSERMAN, “Very deep convolutional networks for large-scale image recognition”, arXiv:1409.1556, (2014).
[9] LIN Tsung-Yi, et.al., “Microsoft coco: Common objects in context”, ECCV, (2014), pp.740-755.
[10] Cityscapes Dataset: https://www.cityscapes-dataset.com