

平成 30 年度卒業論文
深層学習で抽出した特徴量を用いた
特徴点マッチング

宮崎大学 工学部 情報システム工学科
梶原 奏美
指導教員 椋木雅之

目次

1. はじめに	1
2. 特徴点マッチング	2
2.1. 特徴点マッチングとは	2
2.2. AKAZE を用いた特徴点マッチング	3
2.3. AKAZE の精度評価	4
2.3.1. 実験方法	4
2.3.2. 実験結果	5
3. 深層学習を用いた特徴点マッチング	7
3.1. 特徴点検出	7
3.2. 特徴量記述	7
3.3. マッチング	8
4. 実験	9
4.1. 実験方法	9
4.2. 実験結果	11
4.3. オリエンテーションを考慮しない実験	17
4.3.1. 実験方法	17
4.3.2. 実験結果	18
5. おわりに	19
謝辞	19
参考文献	20

1. はじめに

同じ物体を違う視点から映した画像から、その物体の同じ点を対応させる特徴点マッチングは、画像からの形状復元や物体認識などの多くのコンピュータビジョンの分野で重要な役割を果たしている。そのため多くの研究が行われており、SIFT[1]やSURF[2]、AKAZE[3]などのさまざまな手法が提案されている。これらの手法は、画像が回転、拡大・縮小されていてもマッチング結果が影響を受けにくい等、優れた特徴を持っているが、マッチング精度には、改善の余地がある。特徴点マッチングの精度向上は、コンピュータビジョンの多くの関連分野にも成果が波及するため、重要な基礎技術である。

一方、近年、機械学習をより発展させた手法である深層学習がコンピュータビジョンの多くの分野で用いられ、大きな成果を挙げている。深層学習を特徴点マッチングに用いた手法の研究も行われている[4][5]。これは、深層学習の特徴量記述能力が優れているためと考えられている。

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。本研究では、この中の特徴量記述に深層学習を取り入れることで、マッチングの精度を向上させることを目指す。

2. 特徴点マッチング

2.1. 特徴点マッチングとは

特徴点マッチングは、異なる画像間で、固有の点を対応付けることである。特徴点検出、特徴量記述、マッチングの3段階からなる。特徴点検出では、画像中から角や線の交わり、エッジなどの他と区別できるような固有の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で表現した値を特徴量として算出する。マッチングでは、対応付ける画像同士の特徴量を比較する。特徴量の距離が近いものを類似度が高いとし、特徴点を対応付ける。以降、対応付いた2つの特徴点を対応点と呼ぶ。

図1に、簡略化した特徴点マッチングの例を示す。この中で、赤丸が特徴点となる。特徴点検出では、他の部分と区別できるような点の座標を検出することが重要である。特徴量記述では、特徴点が他の特徴点と区別できるような特徴量を記述することが重要である。マッチングでは、同じ物体の同じ点同士が対応することが重要である。

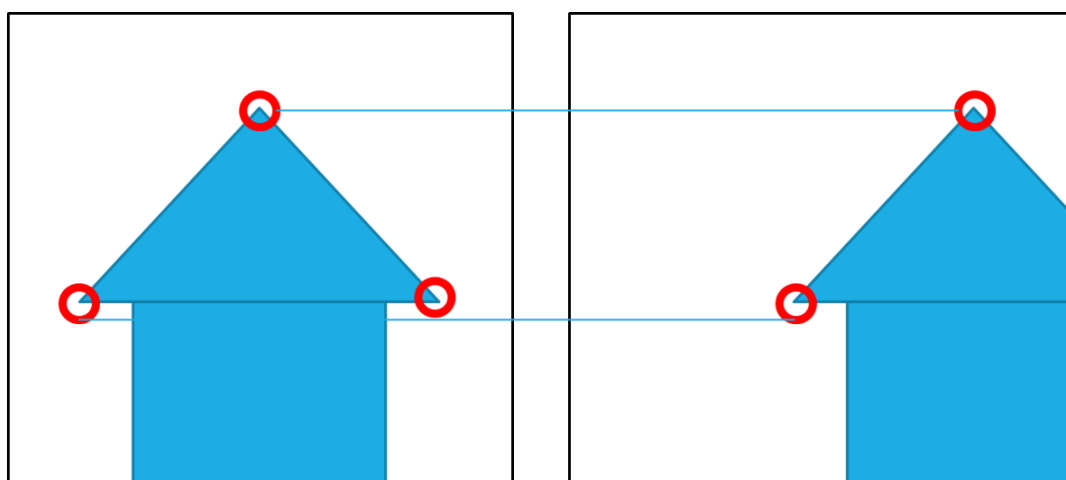


図 1 特徴点マッチングの例

2.2. AKAZE を用いた特徴点マッチング

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。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。

この手法は、拡大縮小不変、回転不変となっている。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。

本研究では OpenCV3.1 に実装されている AKAZE を用いる。

2.3. AKAZE の精度評価

本研究を行うにあたり、AKAZE による特徴点マッチングの精度評価の予備実験を行った。

2.3.1. 実験方法

画像の中心を原点とし、反時計まわりに 45° 回転させた画像（以下、回転画像）を用意する。回転画像と元の画像（以下、元画像）に対して AKAZE を用いて特徴点検出、特徴量記述を行う。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。総当たりマッチングでは、2 枚の画像が与えられた時、画像 1 から抽出したすべての特徴点について画像 2 中で最も特徴点間距離が小さい特徴点と対応付ける。そのため、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。クロスチェックでは、元画像→回転画像の総当たりマッチングと回転画像→元画像の総当たりマッチングの双方の結果で一致するもののみをマッチング結果とすることでこの問題を解消し、信頼性の高いマッチング結果を得る。

式 (1) を用いてマッチング率を求める。

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

元画像の i 番目の特徴点座標 Q_i と回転画像の j 番目の特徴点座標 P_j が対応点になっているとする。 P_j を時計回りに 45° 回転させた画像座標を P'_j とする。この時、許容座標誤差 T に対して $|Q_i - P'_j| < T$ を満たす対応点を「正しい対応点」とする。

2.3.2.実験結果

実験結果を図 2 に示す。横軸は対応点の許容座標誤差 T 、縦軸は T におけるマッチング率である。

図 2 より、 $T < 10$ では、マッチング率の変化が大きい。これは、図 3 のようにほぼ同じ場所であるが特徴点検出時の誤差が原因で、位置がずれた点同士を対応付けたと考えられる。また $T \geq 10$ では変化が小さく、約 93%のマッチング率でグラフが横ばいとなっている。図 4 のように全く違う場所の特徴点と対応する対応点が存在することで、マッチング率が 100%にならないと考えられる。

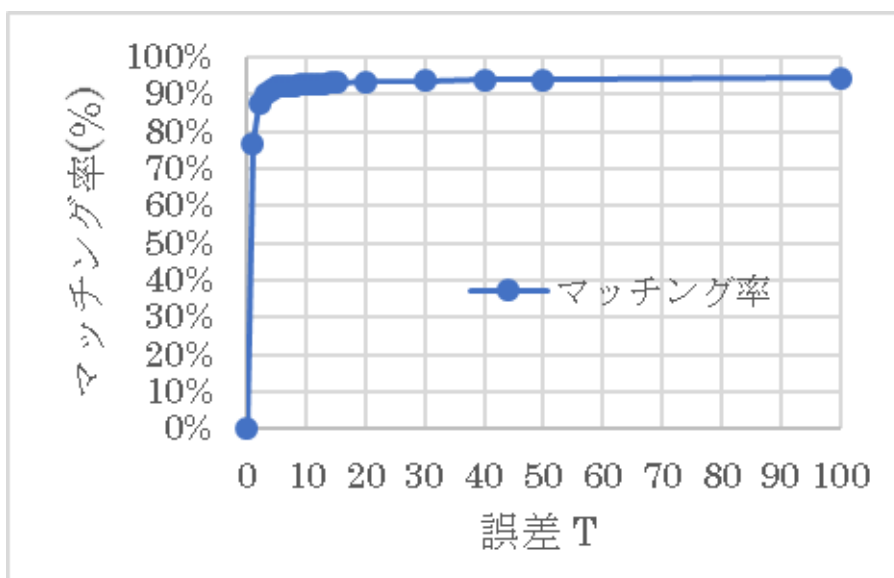


図 2 マッチングした点間の座標誤差とマッチング率

本研究では、特徴点検出は既存手法を使い、特徴量記述によるマッチングの性能の評価をしたい。そのため、検出時の誤差は正解であると考え。これを踏まえ、 $T < 10$ をマッチング成功とみなすと、本実験のマッチング成功率は約 93%といえる。

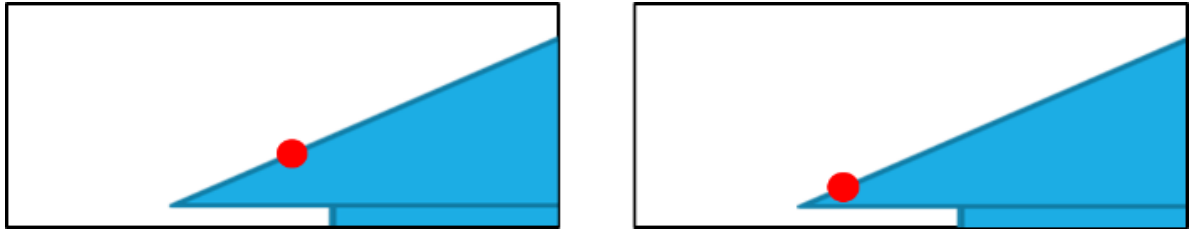


図 3 特徴点の検出エラー

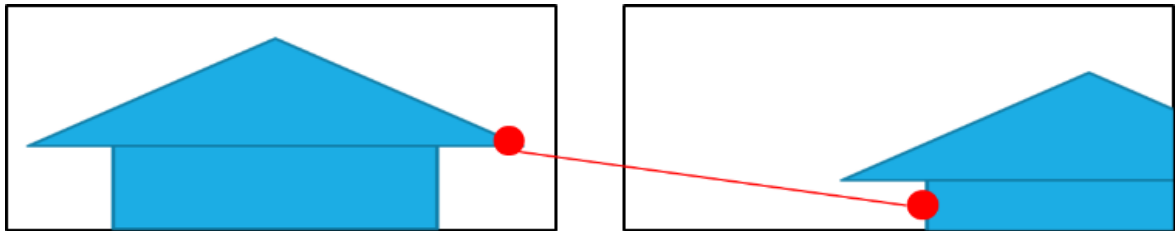


図 4 特徴点のマッチングエラー

3. 深層学習を用いた特徴点マッチング

本研究では、特徴点検出には AKAZE をそのまま使用する。特徴量記述では、学習済みの深層学習を用い、特徴量を抽出する。これを使い、マッチングを行う。以下、各処理について詳細を述べる。

3.1. 特徴点検出

AKAZE によりマッチングさせたい画像からそれぞれ特徴点を検出する。このとき、オリエンテーションも同時に求める。

3.2. 特徴量記述

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[6]と呼ばれる構造を持った畳み込みニューラルネットワーク (CNN) に ILSVRC2012 の 1000 カテゴリ、120 万画像のデータを学習させた学習済みの公開モデル (bvlc_reference_caffenet[7]) を用いる。このモデル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。

3.3. マッチング

2.3 節の予備実験同様、総当たりマッチングとクロスチェックによりマッチングを行う。特徴量間の距離計算には、cos 距離を用いる。 \vec{q}_i をマッチングさせる1つ目の画像の特徴点 i の特徴量、 \vec{p}_j を2つ目の画像の特徴点 j の特徴量とする。この時、cos 距離を式(2)で求める。

$$\frac{\vec{q}_i \cdot \vec{p}_j}{|\vec{q}_i| |\vec{p}_j|} \quad (2)$$

cos 距離は、ベクトル同士の成す角度の近さを表現するため、1に近ければ類似しており、0に近ければ似ていないことになる。

4. 実験

4.1. 実験方法

提案手法と AKAZE でのマッチング成功率の比較実験を行う。回転、拡大・縮小の変換に対して、マッチング成功率がどのように変化するか評価する。

画像の中心を原点とし、反時計まわりに回転させ、拡大縮小させた画像を用いる。5, 15, 35, 45, 90 度に回転させた画像を用意する。これらの画像をそれぞれ 50%, 60%, 70%, 80%, 90%, 100%, 110%, 120%, 130%, 140%, 150% の大きさに変換した画像（以下、変換画像）で実験を行う。図 5 は元画像、図 6 は 15° 回転させたスケール 100% の変換画像である。元画像の大きさは 1747×1310 である。

特徴点には、AKAZE で検出した同じ点を用いる。提案手法の特徴量記述では、元画像と変換画像の特徴点の周辺を 16×16 画素と 32×32 画素の範囲で切り取り、特徴量を求める。図 7 に 16×16 で切り取った画像、図 8 に 32×32 で切り取った画像の例を示す。マッチングには、どちらも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。マッチング結果の評価では、2.3 節より、対応点の座標の誤差が $T = 10$ 未満のものをマッチング成功とする。



(左) 図 5 元画像

(右) 図 6 15° 回転させた変換画像



図 7 16×16 で切り取った画像



図 8 32×32 で切り取った画像

4.2. 実験結果

表 1 に提案手法で特徴点の周辺 16×16 を切り取った際の実験結果、表 2 に特徴点の周辺 32×32 を切り取った際の実験結果を示す。さらに、表 3 に AKAZE でのマッチングの実験結果を示す。また、スケール 100%のときのマッチング成功率を図 9 に、画像を各角度で回転させたときのそれぞれのマッチング成功率を図 10 から図 15 に示す。

表 1 16×16 で切り取った際の実験結果

	0°	5°	15°	35°	45°	90°
50%	25.23%	21.10%	22.60%	20.32%	17.81%	22.65%
60%	39.59%	37.54%	34.41%	33.81%	31.55%	37.64%
70%	66.13%	65.40%	63.87%	61.20%	61.45%	62.21%
80%	88.72%	86.91%	86.75%	84.85%	83.33%	83.63%
90%	97.11%	95.53%	94.57%	94.28%	93.60%	92.71%
100%	100.00%	99.24%	98.56%	97.62%	97.20%	96.60%
110%	95.86%	90.52%	88.27%	86.27%	84.44%	85.98%
120%	85.16%	74.76%	72.90%	70.77%	70.02%	73.82%
130%	67.35%	55.76%	52.77%	50.97%	50.15%	55.04%
140%	48.87%	38.45%	37.04%	35.90%	35.01%	39.58%
150%	35.61%	25.37%	24.03%	20.71%	22.17%	28.67%

表 2 32×32 で切り取った際の実験結果

	0°	5°	15°	35°	45°	90°
50%	35.05%	35.35%	31.15%	33.49%	30.65%	37.44%
60%	60.75%	61.65%	57.25%	53.75%	51.72%	59.86%
70%	81.13%	78.88%	80.51%	76.91%	77.64%	80.32%
80%	91.83%	91.40%	91.32%	91.21%	90.89%	91.48%
90%	98.61%	97.57%	97.21%	96.69%	97.12%	98.34%
100%	100.00%	99.37%	99.02%	98.89%	98.62%	99.60%
110%	95.71%	91.19%	91.26%	90.37%	89.69%	93.69%
120%	88.90%	83.04%	83.21%	81.03%	81.47%	86.11%
130%	76.23%	66.04%	66.38%	65.97%	66.28%	74.48%
140%	62.47%	52.62%	51.13%	48.91%	48.92%	59.07%
150%	0.00%	34.66%	36.91%	33.89%	35.03%	45.16%

表 3 AKAZE での実験結果

	0°	5°	15°	35°	45°	90°
50%	86.80%	72.96%	72.96%	64.25%	64.85%	86.97%
60%	80.58%	68.66%	63.92%	63.87%	63.39%	79.95%
70%	88.52%	79.59%	76.20%	77.29%	75.99%	88.26%
80%	82.38%	75.79%	72.67%	73.06%	72.62%	83.23%
90%	88.39%	83.01%	81.04%	81.34%	82.15%	87.46%
100%	100.00%	95.25%	94.03%	93.45%	92.73%	98.70%
110%	88.95%	86.62%	85.32%	84.95%	84.76%	89.10%
120%	76.85%	73.84%	72.16%	72.35%	72.98%	77.07%
130%	83.91%	82.14%	80.92%	80.53%	81.42%	82.69%
140%	86.40%	85.71%	84.46%	84.34%	84.18%	85.94%
150%	84.96%	83.36%	83.07%	81.76%	82.05%	84.10%

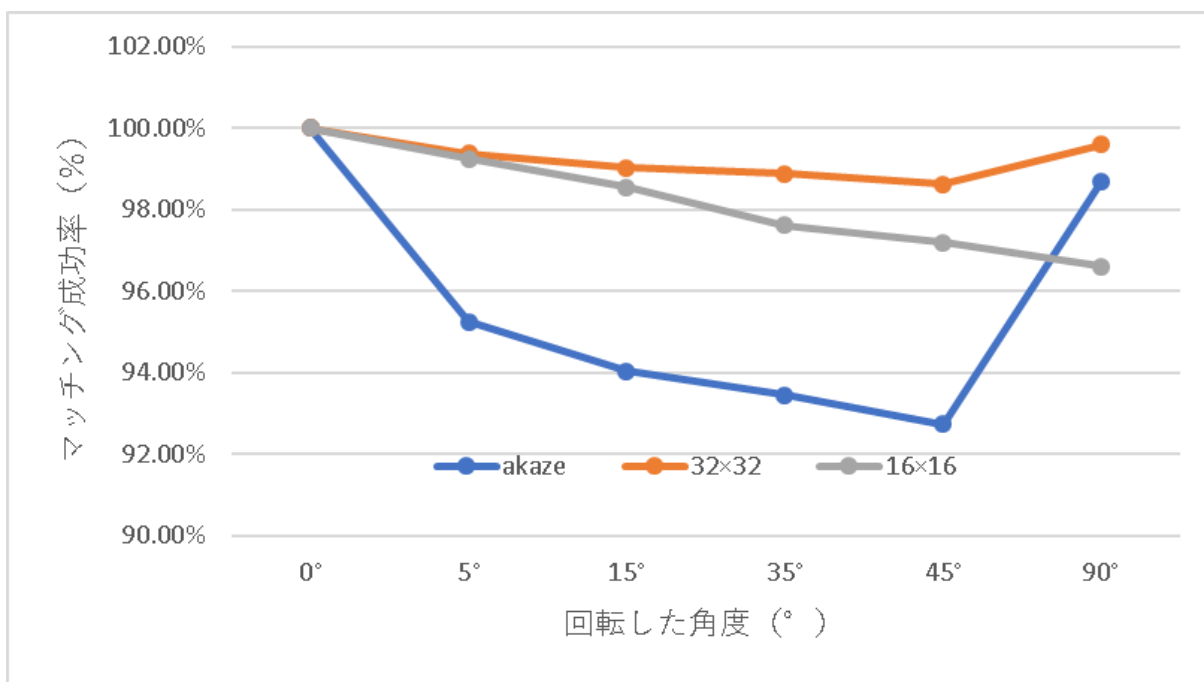


図 9 スケール 100%のときのマッチング成功率

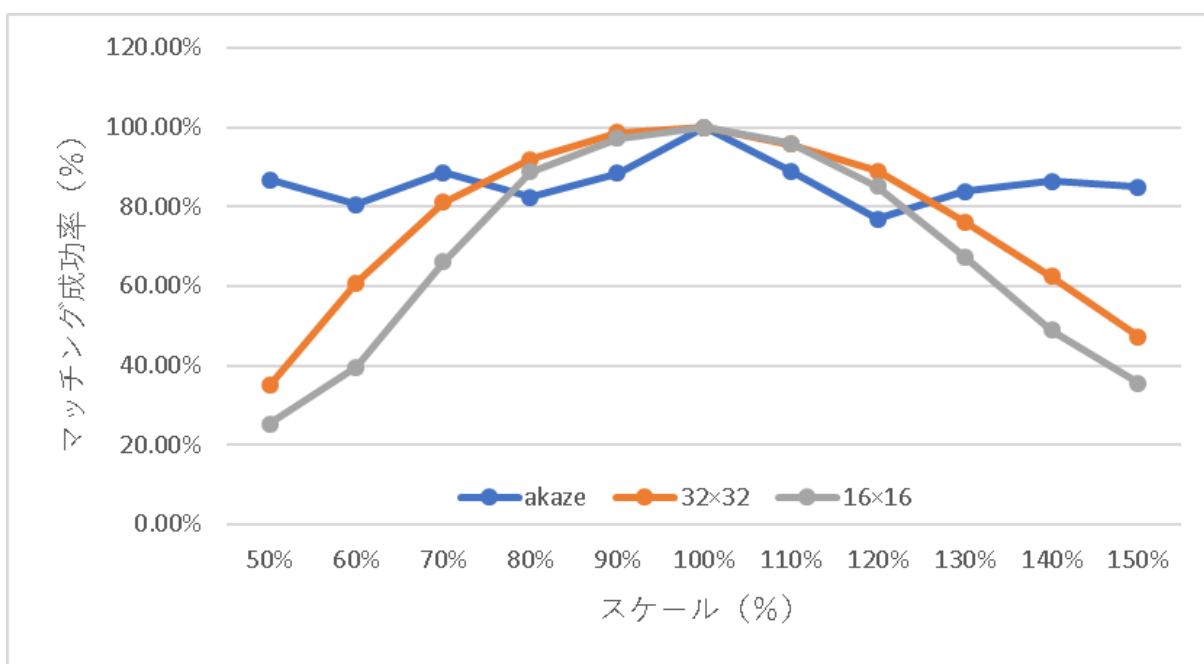


図 10 回転していないときのマッチング成功率

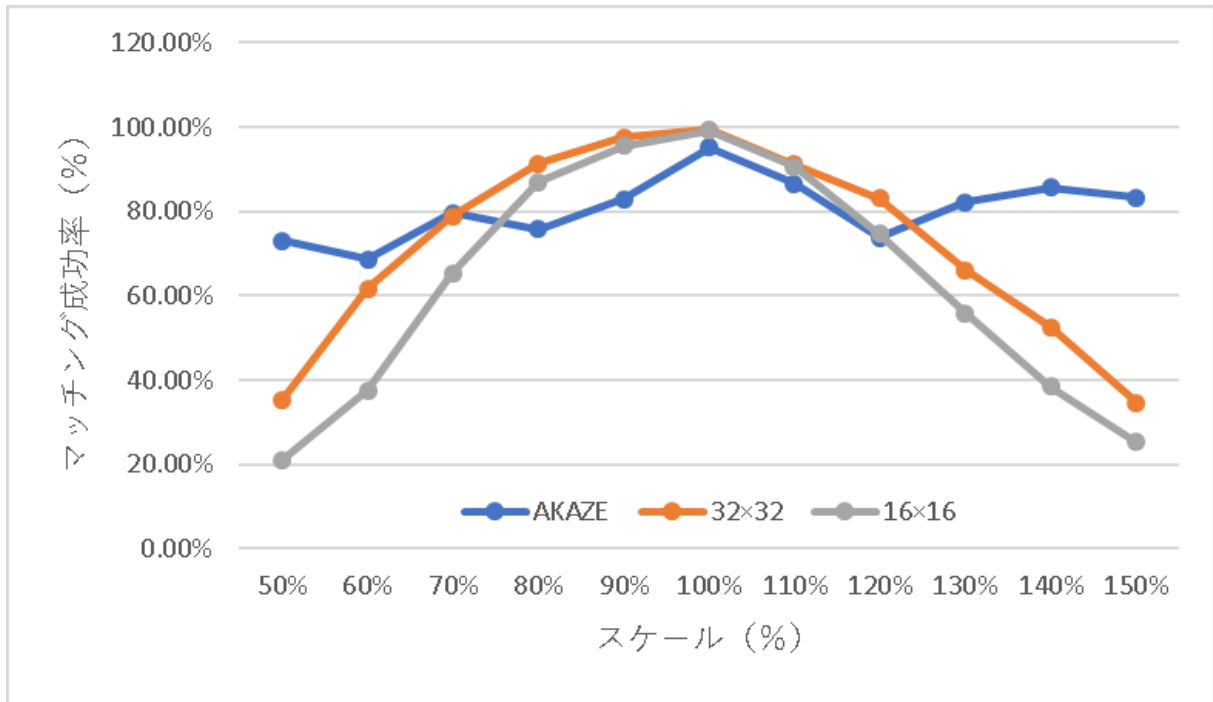


図 1 1 5° 回転させたときのマッチング成功率

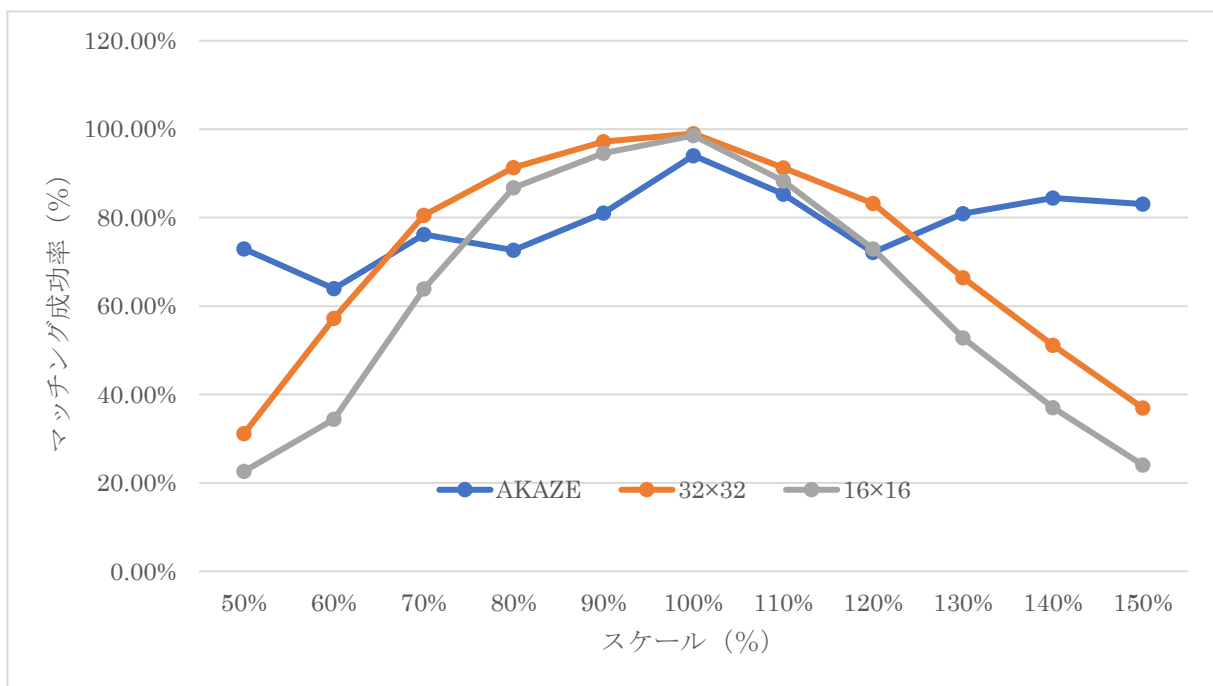


図 1 2 15° 回転させたときのマッチング成功率

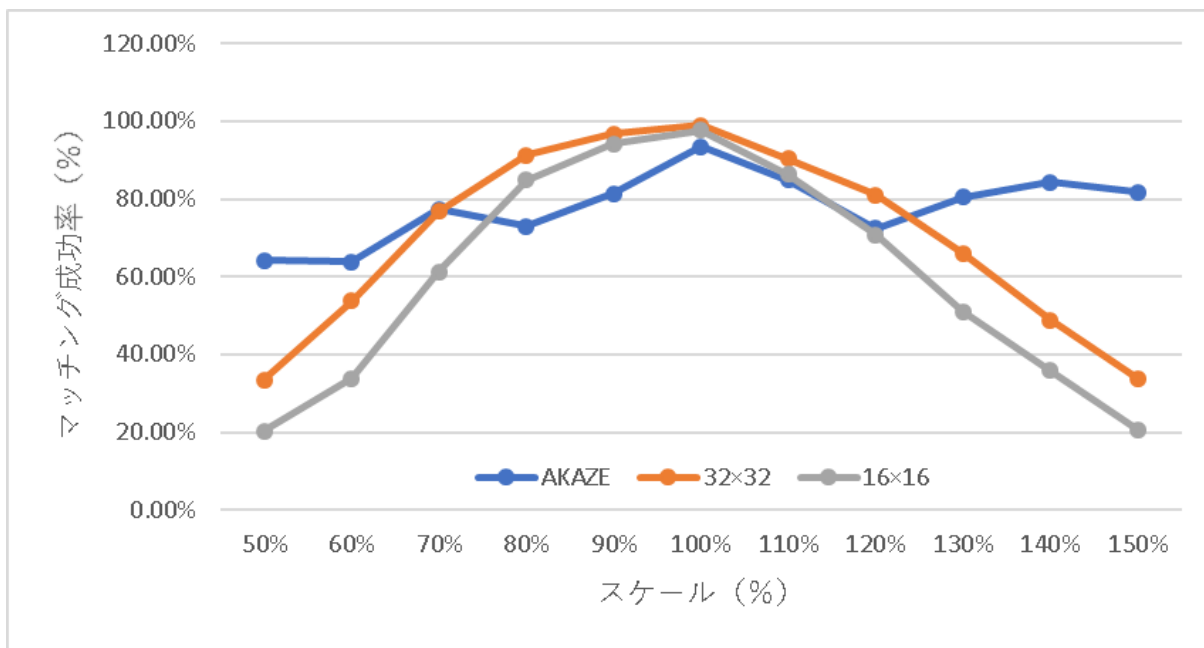


図 1 3 35° 回転させたときのマッチング成功率

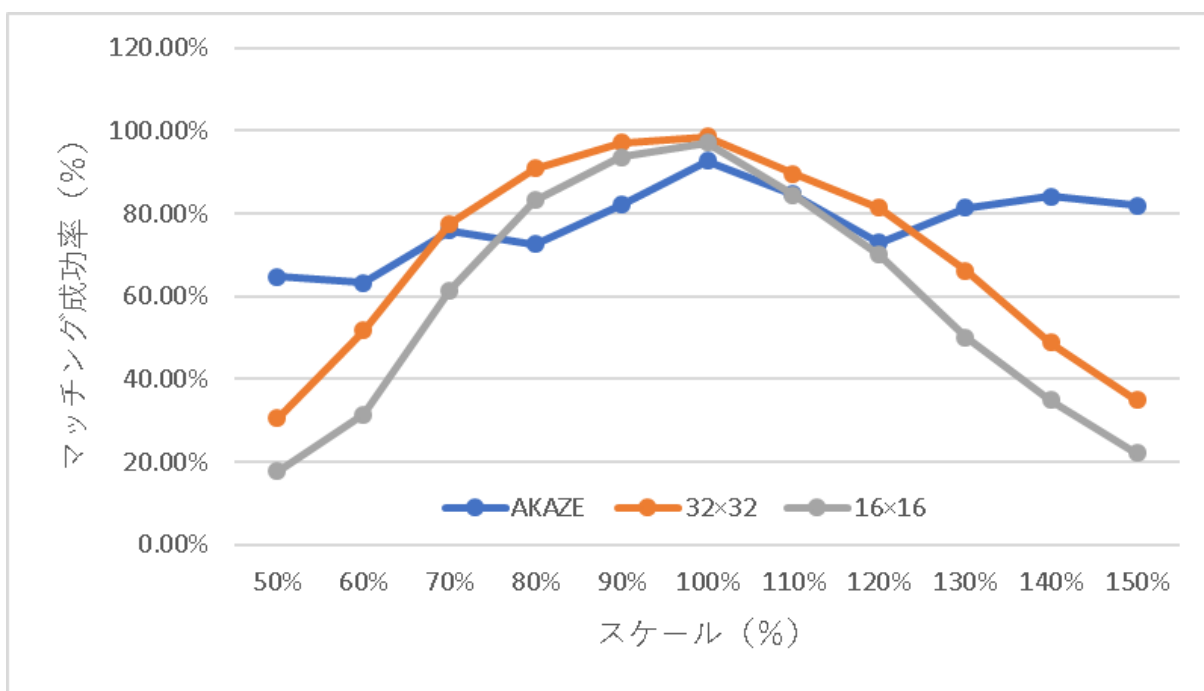


図 1 4 45° 回転させたときのマッチング成功率

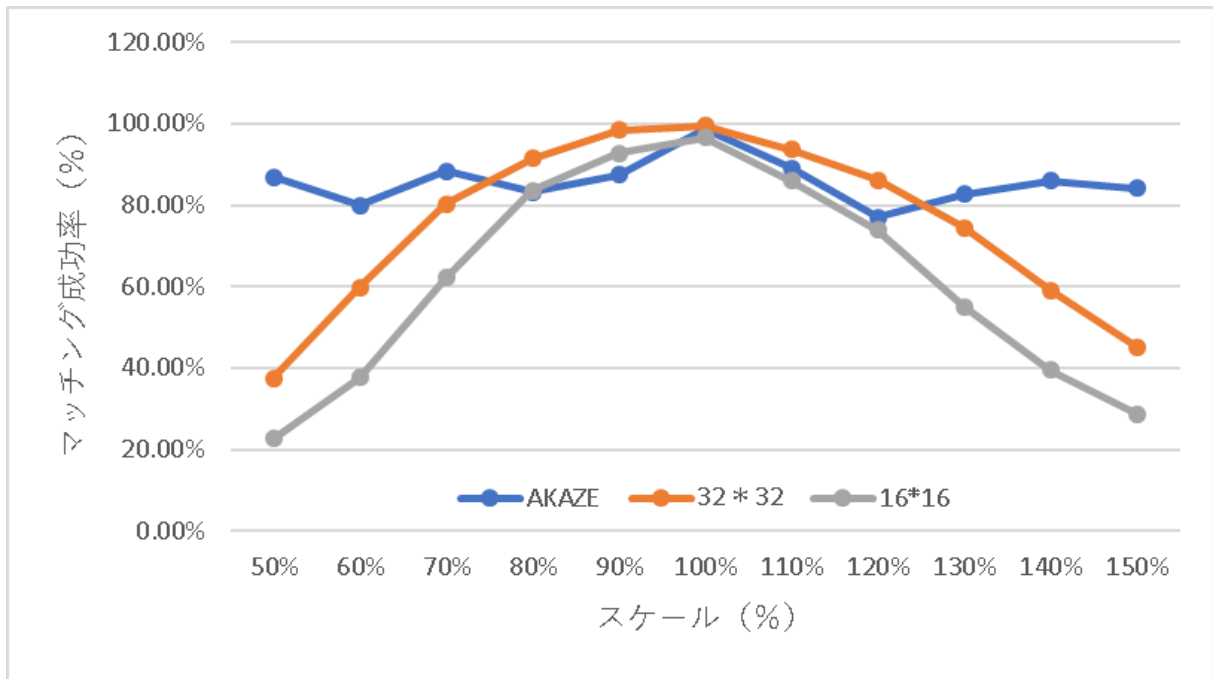


図 15 90° 回転させたときのマッチング成功率

提案手法において、切り取り範囲を変えた際の結果を比較すると 32×32 での結果が 16×16 での結果を上回っている。これは、本研究で用いた深層学習の学習済みモデルが物体認識用であるため、小さな画像を利用することに向いていないことが原因と考えられる。

提案手法で 32×32 で切り取った結果と AKAZE での結果との比較では、スケール 100% の時、いずれの回転角度でも提案手法が同等以上である。提案手法の回転への対処は AKAZE のオリエンテーションを使っているため、この差は、特徴量記述の差といえる。このことから、提案手法の特徴量記述は、AKAZE より優れているといえる。一方、スケール変化に対しては、スケールが 70%～120% では、提案手法が上回っているが、スケール変化が大きくなると、AKAZE の方が良い結果であった。提案手法はスケール変化について考慮しておらず、改善が必要であることが分かった。

4.3. オリエンテーションを考慮しない実験

4.3.1. 実験方法

オリエンテーションを考慮しない場合の評価実験を行う。画像の中心を原点とし、反時計まわりに回転させた画像を用いる。5, 15, 35, 45, 90 度に回転させた画像（以下、回転画像）で実験を行う。

特徴点には、AKAZE で検出した同じ点を用いる。提案手法の特徴量記述では、元画像と回転画像の特徴点の周辺を 32×32 画素の範囲で切り取り、特徴量を求める。切り取った画像の例を図 16 に示す。マッチングには、どちらも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。マッチング結果の評価では、2.3 節より、対応点の座標の誤差が $T = 10$ 未満のものをマッチング成功とする。



図 16 オリエンテーションを考慮していない場合の 32×32 で切り取った画像

4.3.2.実験結果

表 4 に、本実験とオリエンテーションを考慮した場合のマッチング成功率を示す。

表 4 本実験とオリエンテーションを考慮した場合のマッチング成功率

	オリエンテーション なし	オリエンテーション あり
0°	100.00%	100.00%
5°	99.71%	99.37%
15°	97.22%	99.02%
35°	54.79%	98.89%
45°	37.54%	98.62%
90°	30.70%	99.60%

回転角度が大きくなるにつれてマッチング成功率が小さくなっている。この結果から、オリエンテーションの考慮が特徴量記述の際に重要であることがわかる。

5. おわりに

本研究では特徴点マッチング中の特徴量記述に深層学習を用いた手法を提案し、マッチングの精度の向上を行った。特徴量記述能力が高く、AKAZE と比べ精度の高い特徴点マッチングを行うことができた。

今後の課題として、提案手法ではスケール変化を考慮していないため、スケール変化に対応させることが挙げられる。

謝辞

本研究を作成するにあたり、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感謝致します。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、研究の相談や助言を快く受けて頂いた椋木研究室の皆様にも感謝致します。

参考文献

- [1] David G. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints”, IJCV, (2004), pp.91-110.
- [2] Herbert Bay¹, Tinne Tuytelaars², Luc Van Gool^{1,2}, “SURF: Speeded Up Robust Features”, CVPR, (2008), pp.346-359.
- [3] Pablo F. Alcantarilla, Jesús Nuevo, Adrien Bartoli, “Fast Explicit Diffusion for Accelerated Features in Nonlinear Scale Spaces”, BMVC, (2013), pp.13.1-13.11.
- [4] Kwang Moo Yi, Eduard Trulls, Vincent Lepetit, Pascal Fua, “LIFT: Learned Invariant Feature Transform”, ECCV, (2016), pp.467-483.
- [5] Mohammad Ali Bagheri Orumi, Mahmoud Famouri, Zohreh Azimifar, Azadeh Nazemi, “Deep Learning-Based Corresponding Points Fast Matching”, ICPRAI, (2018), pp.256-260.
- [6] A. Krizhevsky, I. Sutskever, G.E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks”, NIPS, (2012), pp.1091-1105.
- [7] AlexNet 学習済みモデル
https://github.com/BVLC/caffe/tree/master/models/bvlc_reference_caffenet