

卒業論文
シーンマッチングによる自動運転ナビゲーション

学籍番号：T190022
氏名：大野 大樹

2023年5月25日

目次

1	序論	2
1.1	研究の背景と目的	2
2	本論	2
2.1	風景画像	2
2.2	自己位置推定	2
2.3	テンプレートマッチング	4
2.4	画像縮小	5
2.5	エッジ検出	5
2.6	LBP 特徴	7
2.7	HOG 特徴	8
2.8	実験の手法	9
3	結論	10
3.1	実験の結果	10
4	考察	14
5	おわりに	14
6	参考文献	15
7	付録	16

1 序論

1.1 研究の背景と目的

1.1.1 背景

近年、人工知能の機械学習法の著しい発展により、自動運転や無人ドローンといった自律走行、自律飛行技術への関心が高まっている。自動運転を搭載した車では、スバルアイサイト[1]のように周囲の情報をステレオカメラのみを使用したものがあるが、テスラ Model3[2] やトヨタ Advanced Drive[3] といった3次元センサー(3DLidar)とステレオカメラを使用して取得しているものもある。3次元センサーを用いることで、道路の路面状況、車間距離、歩行者の有無といった周囲の状況をより緻密に把握することができ、光を使わないと暗闇や雨天など天候の影響を受けにくいという特徴を持っている。しかし、周りに障害物のない開けた土地やデータ処理の負荷が高い、そしてセンサーそのものが高価であるなどのデメリットもある。また、カメラセンサーは天候の変化や暗闇に弱いというデメリットはあるが、画像の解像度の向上や画像から得られる情報の多さ、3次元センサーと比べコストが低いというメリットがある。そこで、本研究では3次元センサーと比べ比較的安価であるカメラと基本的な画像処理を用いた自己位置推定を行っていく。

1.1.2 目的

自動運転車両が目的地まで移動する場合、自動運転車両は現在の位置を知る必要がある。これを自己位置推定という。例えばLiDARなどのセンサーを用いる場合、自己位置推定は環境の3Dデータを持つ地図と、現在の環境の形状を3次元的に照合すること等で自己位置推定を行う。しかし、人間は周囲の風景を見て、自分の位置を知ることができるために、自動運転車両も人間と同様に、カメラで周囲の風景を撮影した画像を照合することで、自己位置推定が行えないか実験する。

2 本論

2.1 風景画像

シーンマッチングでは、シーンと呼ばれる風景画像を使用する。(図1)風景画像には不变的な情報と可変的な情報の2種類の情報が描画されている。不变的な情報とは、建物や遠景に見える山の輪郭、道路の形状など数時間単位等では決して変化することのない情報をさす。可変的な情報とは、天候や時間などの影響による、雲の動きや画像全体の明度の変化等が相当する。(図1a)風景画像はこの天候や時間等の影響を受けやすい画像であるため、こういった変化に強い画像特徴を使用することで影響をなくす必要がある。

そこで、今回風景画像を照合する前にいくつかの画像特徴を使用した前処理を行う事で、どの画像特徴を使用した時が天候や時間の影響を受けずに自己位置推定を行えるかを調べていく。

2.2 自己位置推定

自動運転車両が自動運転中に自己位置推定を行う方法はいくつかある。その中でも今回使用するのは風景画像を使った自己位置推定である。風景画像を使った自己位置推定では、自動運転走行するルートの風景画像を事前に走行して撮影しておく。そして、実際に走行しながら取得した画像と照らし合わせながら自分の位置を



(a) 晴れの時の風景画像

(b) 曇りの時の風景画像

図 1: 風景画像



図 2: 自己位置推定のイメージ

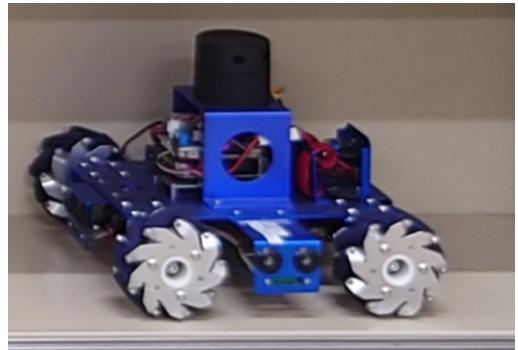
推定していく。例えば、走行しながら取得した画像が A 地点で撮影した画像と照合し同じと判断された場合、その取得した画像によって自動運転車両は自分の位置が A 地点にあると推定することができる。取得した画像と事前走行で得られた風景画像を照合する手法としてテンプレートマッチングを使用する。

また、テンプレートマッチングを行う際の処理として、以下の処理を行う。

1. 0.1 倍のスケールに画像縮小してマッチング
2. サイズを変えずに Canny のエッジ検出をしてマッチング
3. 0.1 倍のスケールに画像縮小し Canny のエッジ検出をしてマッチング



(a) 探索対象画像



(b) テンプレート画像

図 3: テンプレートマッチングに必要な画像

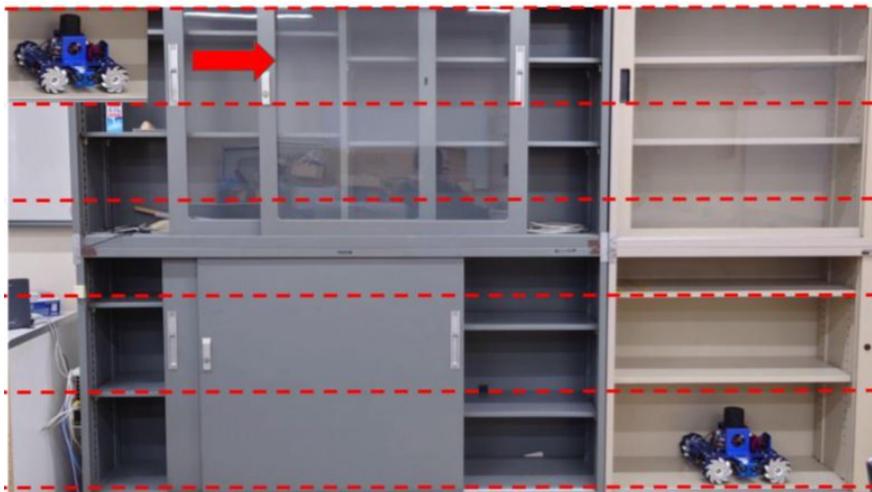


図 4: テンプレートマッチングの動作イメージ

4. サイズを変えずに LBP 特徴を計算してマッチング
5. 0.1 倍のスケールに画像縮小して LBP 特徴を計算してマッチング
6. サイズを変えずに HOG 特徴を計算してマッチング

それぞれの前処理を行うことで画像に特徴を持たせることができ、似ている画像同士の類似度を高めることができると予想した。使用した画像処理技術については次節にて説明する。

2.3 テンプレートマッチング

テンプレートマッチングとは、画像同士の類似度を調べる画像処理である。テンプレートマッチングを行うにあたってまずテンプレート画像と探索対象画像と呼ばれる二つの画像を用意する必要がある。テンプレート画像とは、探索対象画像中で探し出したい画像情報が大きく描かれた画像のことである。(図 3b) また、テンプレート画像の画像サイズは探索対象画像よりも小さいことが必要である。探索対象画像とは、テンプレート画像に描かれた画像情報とどれだけ類似しているかを調べたい画像のことである。(図 3a) 探索対象画像内に

はテンプレート画像と同じ画像情報が必ずしも含まれている必要性はない。

テンプレートマッチングでは画像同士の類似度を調べる手法がいくつか存在するが、ここでは実験でも使用したZNCC(零平均正規化相互相関)について説明する。 $I(i, j)$ が探索対象画像の座標 (i, j) における画素値、 $T(i, j)$ がテンプレート画像の座標 (i, j) における画素値とする。また、テンプレート画像の解像度を $M \times N$ とする。このとき、ZNCC は次の式で求めることができる。

$$ZNCC = \frac{\sum_{j=0}^{N-1} \sum_{i=0}^{M-1} (I(i, j) - \bar{I})(T(i, j) - \bar{T})}{\sqrt{\sum_{j=0}^{N-1} \sum_{i=0}^{M-1} (I(i, j) - \bar{I})^2 \times \sum_{j=0}^{N-1} \sum_{i=0}^{M-1} (T(i, j) - \bar{T})^2}} \quad (1)$$

ここで \bar{I} と \bar{T} はそれぞれ探索対象画像とテンプレート画像の画素値の平均値となる。

ZNCC のもととなった NCC(正規化相互相関) があり、NCC はベクトルの内積の式と同一なので ZNCC は計算結果が 1 に近づくほど類似度が高くなる。NCC との違いは画像全体から全画素の平均値を引くことでノイズや明るさの影響を受けにくくという利点がある。

これらの画像と類似度計算を使ってマッチングを行う。テンプレートマッチングでは探索対象画像の左上の画素にテンプレート画像を合わせ、その画素からテンプレート画像と同じサイズの領域とテンプレート画像で類似度を調べる。その結果の値をその画素の類似度とする。これを探索対象画像上を走査しながら行っていく。(図 4) その後、一番大きい類似度を持った画素がマッチングされた部分となる。

2.4 画像縮小

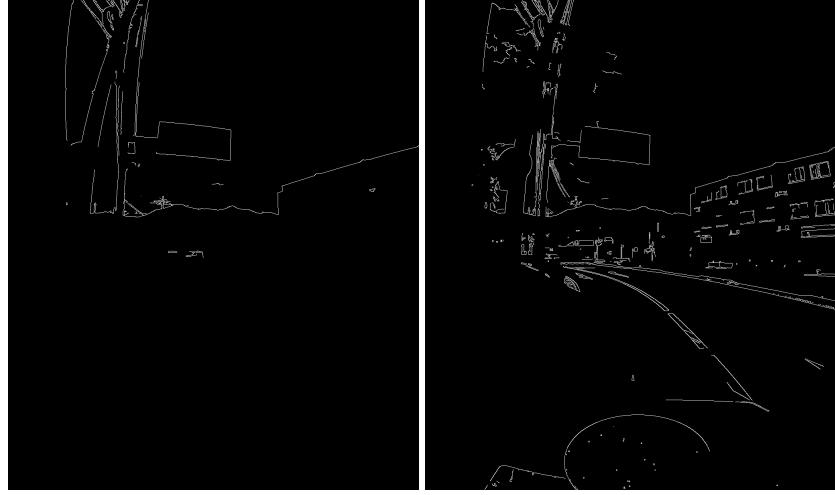
0.1 倍スケールに縮小した画像を使ってマッチングを行った理由としては、画像のスケールを下げることで細かい部分のノイズを削減し大まかな形状を得ることができるためである。解像度が高い画像では、木々の揺れや葉の色、タイルの網目などの細かい情報が散乱している。その状態でテンプレートマッチングを行った場合、そういった細かい情報まで一致しなければ類似の画像として扱うことができない。風景画像では、遠方の山や建物の輪郭といった大まかな物体の形が分かればよいので、細かい情報をなくす簡単な処理としてスケールの縮小を行った。また、画像のスケールを縮小することでテンプレートマッチングの走査領域が減るために、実行時間を削減することができる効果もある。

2.5 エッジ検出

エッジ検出とは、画像内の物体の輪郭や輝度差が閾値以上の部分のみを抽出する画像処理のことである。基本的にエッジは輝度の不連続な部分であり、明るいところから極端に暗いところ、暗いところから極端に明るくなるところにエッジができやすい。ここでは、いくつかあるエッジ検出のひとつである Canny 法について説明する。

Canny 法ではいくつかの処理に分けてエッジ検出を行っていく。まず、入力画像にグレースケール化を行い、その入力画像にガウシアンフィルタをかける。この時、ガウシアンフィルタをかけた画像を G とする。ガウシアンフィルタとは加重平均フィルタと呼ばれ、注目画素からの距離に応じて重み付けを行いその合計を画素値とするフィルタの事である。ガウシアンフィルタをかけて画像を平滑化することで、画像内のノイズを消去、軽減する目的がある。

次に、画像 G に Sobel フィルタをかけて x 方向と y 方向の一次微分画像 G_x, G_y を求める。Sobel フィルタもエッジ検出法の一つで、入力画像に対し x 方向か y 方向に一次微分を行い x 方向であれば縦線、 y 方向であれば横線のエッジを検出する。



(a) 曇りの時のエッジ特徴

(b) 晴れの時のエッジ検出

図 5: 風景画像に対してエッジ検出を適用した場合

先ほど用いた G_x, G_y を使って、エッジ強度と勾配方向を求めていく。エッジ強度 I と勾配方向 θ は次の式で求めることができる。

$$I(i, j) = \sqrt{G_x(i, j)^2 + G_y(i, j)^2} \quad (2)$$

$$\theta = \arctan \frac{G_y}{G_x} \quad (3)$$

次に、先ほど求めたエッジの勾配を使って非極大抑制と呼ばれる処理を行う。非極大抑制とは注目している画素値がその画素のエッジ勾配方向に隣接している 2 つの画素値と比べて極大値であるかどうかを調べる。極大値であればそのまま残し、非極大値であればその画素はエッジでないため画素値を 0 とする。

最後に先ほどの非極大抑制処理で得られたエッジが正しいかどうか調べるために、ヒステリシス閾値処理を行う。ヒステリシス閾値処理では、まず最大閾値と最小閾値の二つの閾値を決定する。そして、非極大抑制で残ったエッジを最大閾値を以上のエッジ、最小閾値以上最大閾値未満のエッジ、最小閾値に満たないエッジの 3 つに分ける。最小閾値に満たない場合はそのエッジをエッジとみなさず消去する。最小閾値以上最大閾値未満のエッジの場合は、そのエッジをエッジの断片または始点と考える。その後、そのエッジに隣接するエッジを全て調べる。調べた中に最大閾値以上のエッジが存在するのであれば、始点から終点までのエッジの集合をエッジとする。最大閾値以上のエッジが存在しない場合、そのエッジの集合はエッジとみなさず消去する。

以上の処理を繰り返すことで画像内のエッジを検出するのが Canny 法である。[4]

今回の実験で Canny のエッジ検出を前処理で用いる理由としては、Canny のエッジ検出では導出でも説明したがヒステリシス閾値処理のおかげで弱いエッジを消すことができ、雲の模様や電線などの細い線等をある程度消すことができる。そのため、建物や山の輪郭線のような強いエッジのみ残すことができテンプレートマッチングを行った際に一致している時と一致していない時での類似度に差が出ると予想した。

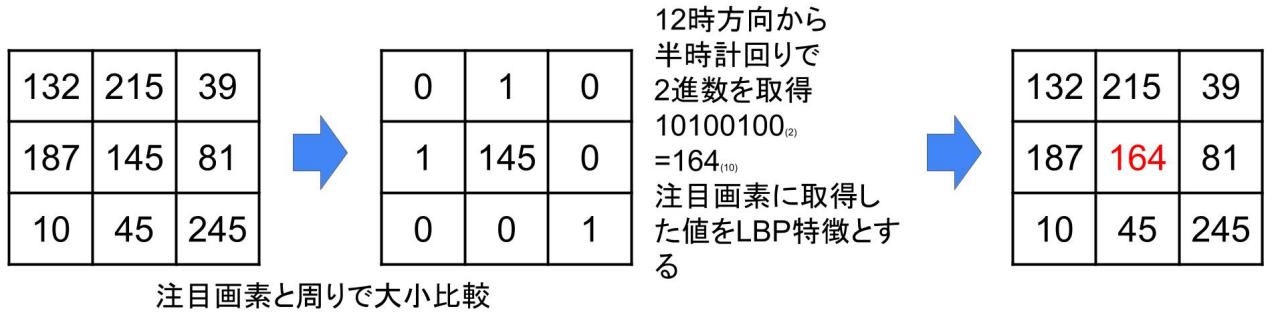


図 6: LBP 特徴の求め方

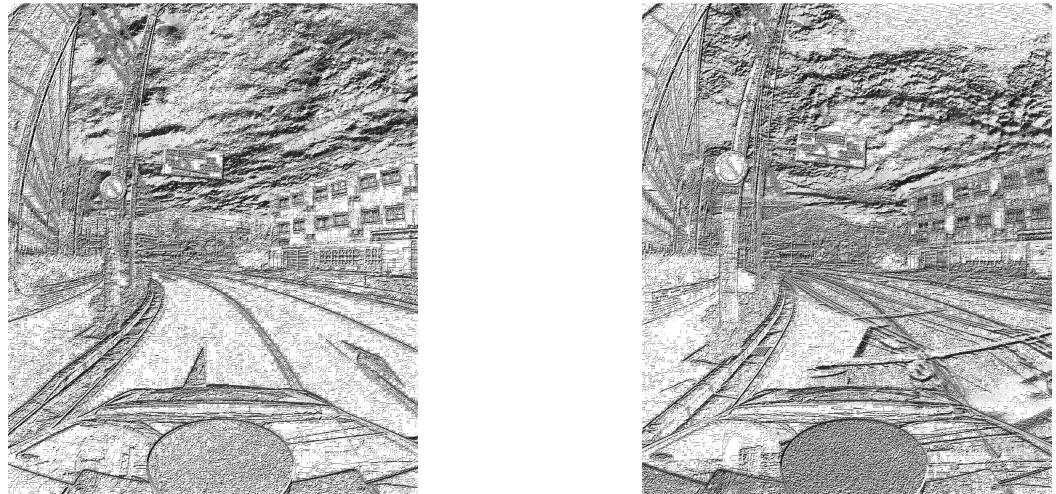


図 7: LBP 特徴を適用した画像

2.6 LBP 特徴

LBP 特徴 (Local Binary Patterns) とは、画像特徴量の一つで注目画素とその周りの輝度のパターンを特徴とするものである。LBP 特徴は、まず入力画像にグレースケール化を行う。その後、注目画素とその周り八近傍に存在する画素の画素値を比較する。比較した結果注目画素よりも近傍に存在する画素の方が画素値が大きいのであれば 1、小さければ 0 とする。その後、注目画素に対して 12 時方向に存在する画素から反時計回りに 1 週 0,1 の値を取得する。取得したその値を 2 進数に置き換えて得られた値が、注目画素の LBP 特徴となる。例えば、図 6 のように中心の画素値が 144 の場合、八近傍に存在する画素値をそれぞれ 144 よりも大きければ 1、小さければ 0 とする。こうして 0 と 1 に振りなおされた八近傍の中から注目画素の 12 時方向から反時計回りに値を取得していく。今回の例でいえば、12 時方向の値は 1 でありここから反時計回りに取っていくと、10100100 が得られる。10100100 を 10 進数に直すと 164 となり、これが注目画素の LBP 特徴値となる。[5]

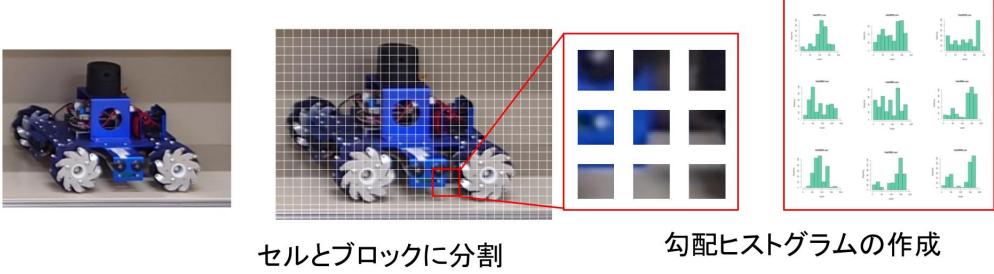


図 8: HOG 特徴の導出



図 10: HOG 特徴を適用した結果

LBP 特徴は、注目画素とその周りの八近傍の画素値との相対的な輝度差によって得られるため、画像全体の明るさやコントラストが変化したとしても LBP 特徴は変化しにくく、画像に対して回転などの線形変換が行われたとしても 2 進数のパターンがシフトするだけで全体的な LBP 特徴の変化は少ない。

今回の実験で LBP 特徴を使った理由としては、風景画像には目印となる建物や山といった変化の少ない情報のほかに、天候や時間帯による明度の変化などの常に変化する周りの環境の情報も含まれる。前処理なしで晴れの日の画像と曇りの日の画像でテンプレートマッチングを行った場合、快晴の空と曇りの空で差が出てしまいマッチングが上手くいかないことがある。そこで LBP 特徴を使うことで快晴と曇りのコントラストの差をある程度取ることができ、さらに画像のスケールを縮小することでさらに差を取ることができる。

2.7 HOG 特徴

HOG 特徴とは、LBP 特徴と同じ画像特徴の一つで入力画像の輝度勾配ヒストグラムを特徴するものである。HOG 特徴の求め方は、まず入力画像にグレースケール化を行った後、いくつかのピクセルが集まってできたセルとセルがいくつか集まってできたブロックに分ける。今回の実験では 1 セル 20×20 ピクセルとし、1 ブロック 3×3 のセルの集合とした。次に 1 ブロックに含まれるセルごとの強度と輝度勾配を計算する。今回、

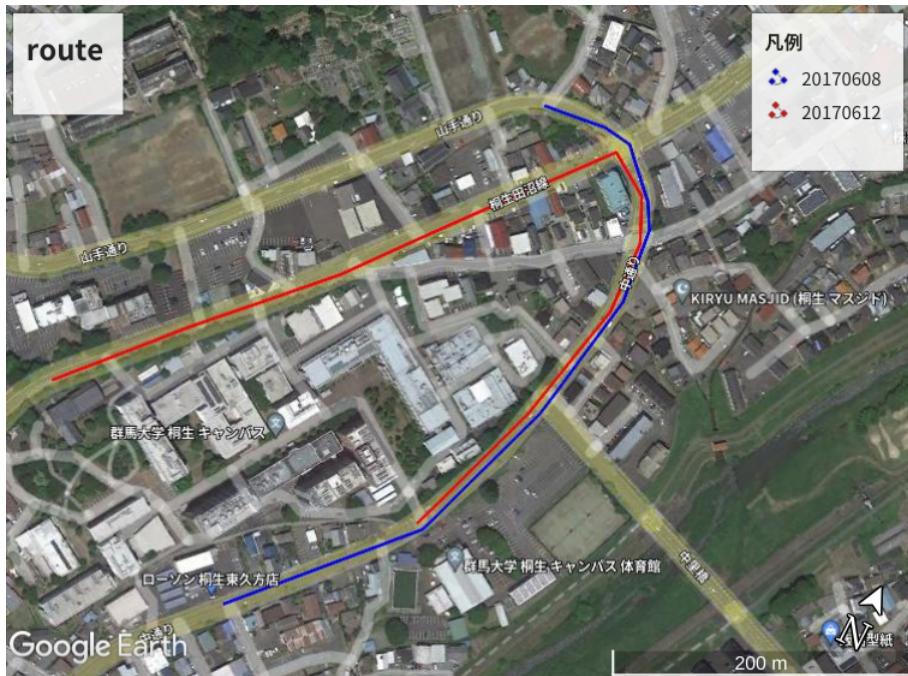


図 11: 群馬大学周りの走行ルート

プログラムの実行時間を少なくするために積分ヒストグラムを使用した。[6] 強度と輝度勾配を求めるために Canny のエッジ検出と同様に画像に対して x 方向と y 方向に対して Sobel フィルタをかけていく。強度と輝度勾配の式は (2)、(3) 式を使用する。その後、セル毎にセル内の輝度勾配のヒストグラムを作成(図 8)し、それをブロック単位で連結し HOG 特徴とする。[7]

HOG 特徴は、画像の特定領域内の輝度勾配を用いるので物体の形状を表現することができ、SVM と使うことで人物検出や車両の検出などに用いられている。

今回の実験で HOG 特徴を使用した理由としては、風景画像には建物や山のような目印となるものほかに道路の白線等の直線といった形状情報を含んでいる。そういった、情報をより強調することができるが HOG 特徴である。前処理で HOG 特徴を使用することでテンプレートマッチングを行った際に一致している時と一致していない時での類似度に差が出ると予想した。

2.8 実験の手法

今回の実験では、群馬大学周りのコースを走行したときの風景画像を使用して自己位置推定を行い画像毎の類似度のピークの調査と精度検証を行った。風景画像は 2 種類あり、2017 年 6 月 8 日に撮影されたもの 671 枚(以後青ルートと呼称する)と 2017 年 6 月 12 日に撮影されたもの 490 枚(以後赤ルートと呼称する)がある。今回、青いルートの画像を探索対象画像とし赤いルートの画像をテンプレート画像とした。これらの画像を使って節で説明した前処理を行いテンプレートマッチングを行う。まず、青ルートの 0 フレーム目の画像に対して赤ルートの 0 フレーム目から 489 フレーム目までの画像を使ってテンプレートマッチングをおこないそれぞれの類似度を求める。この処理を青ルートの 670 フレーム目までの画像に対して同様の処理を行う。これによって青ルートの n フレーム目は赤ルートの m フレーム目に類似度のピークが来るといったことがわ

かる。

精度検証で用いる式は次の式を用いた。

$$accuracy = \frac{\text{ピーク時の類似度の値}}{\text{ピーク時の類似度の値} + 2 \text{ 番目に高い類似度の値}} \quad (4)$$

この式により accuracy が 1 に近づくほどピークが突出して高く、0.5 に近づくほど他に似たようなピークが出現しているということがわかる。この検証式を用いる理由としては、自己位置推定を行う際にピークの類似度と似ている値が多く出現した場合、誤った風景画像が結果として出力されてしまう。そのため、ピーク時の値がほかの値よりも突出しているのであれば突出している程よく、誤った風景画像が出力されるのを防ぐことができる。

3 結論

3.1 実験の結果

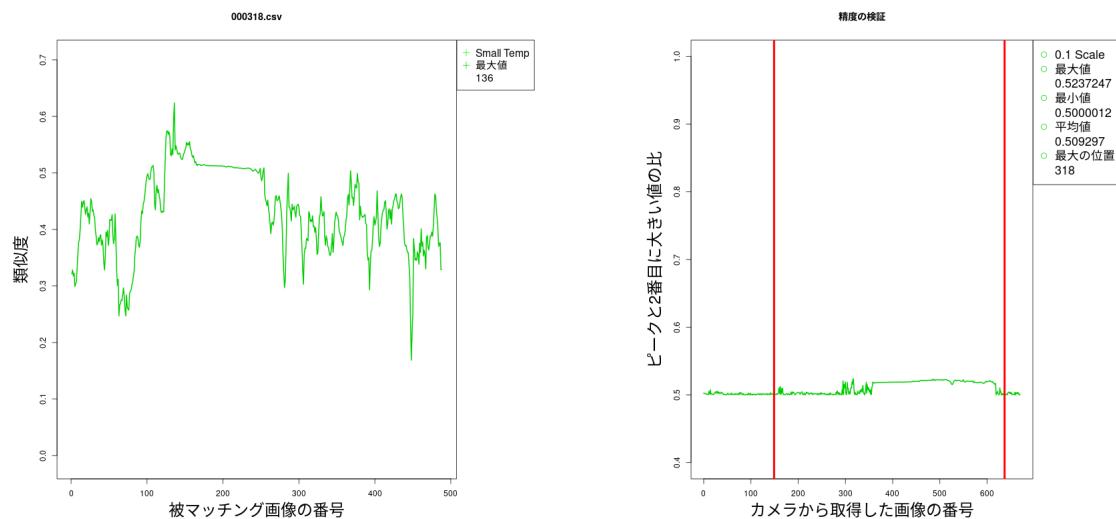


図 12: 画像縮小のみを行った時の自己位置推定の結果と精度検証

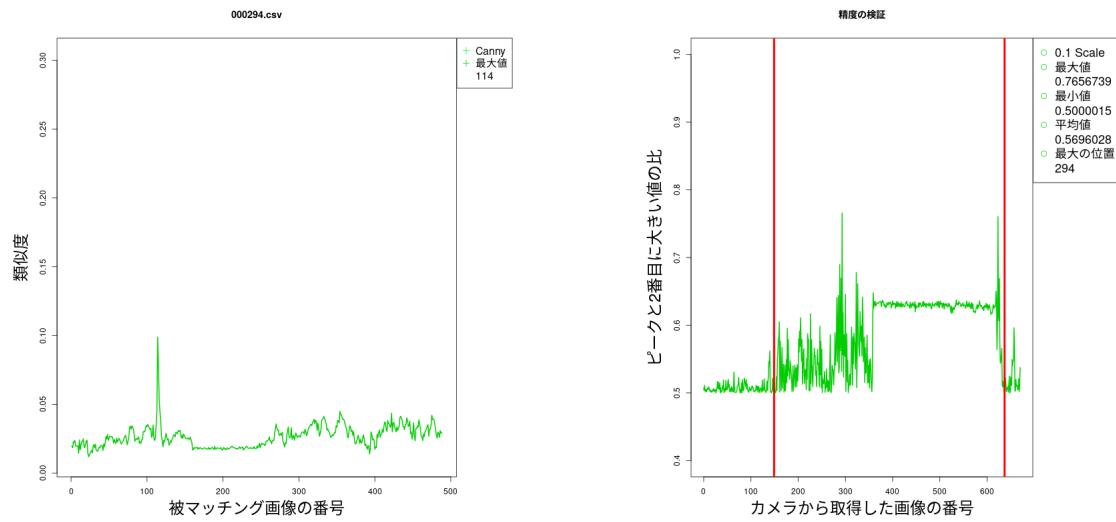


図 13: エッジ検出を行った時の自己位置推定の結果と精度検証

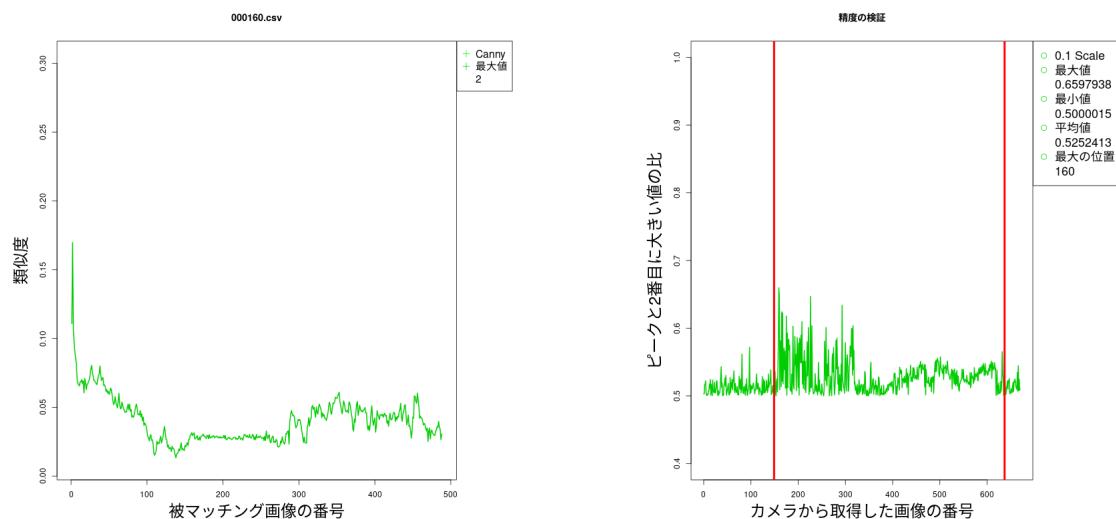


図 14: 画像縮小とエッジ検出を行った時の自己位置推定の結果と精度検証

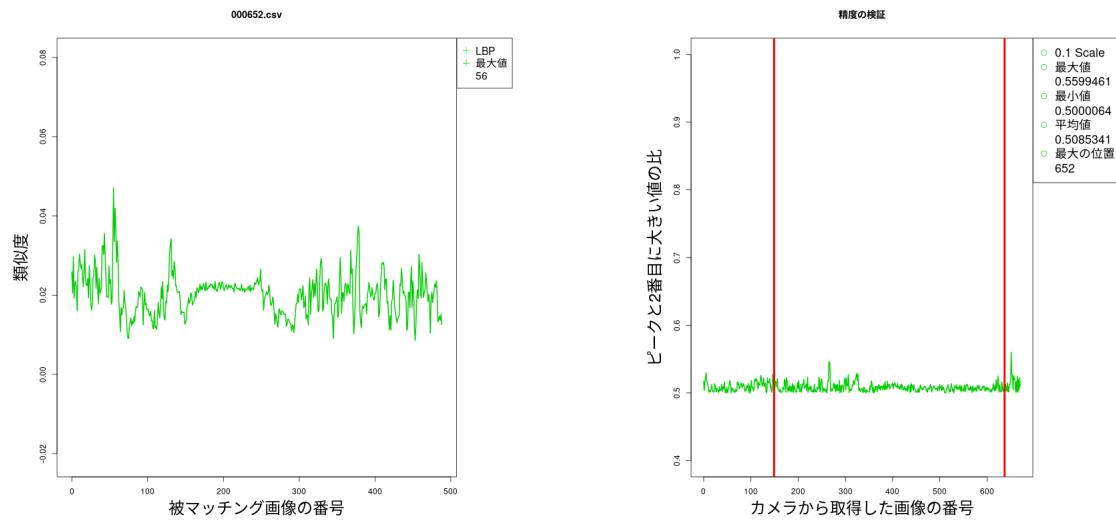


図 15: LBP 特徴のみを適用時の自己位置推定の結果と精度検証

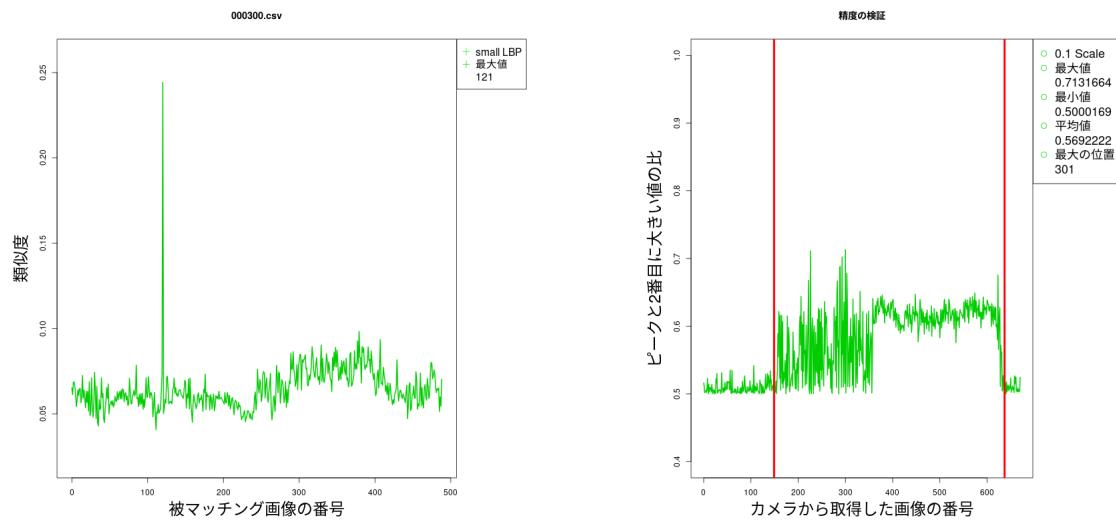


図 16: 画像縮小と LBP 特徴を適用時の自己位置推定の結果と精度検証

表 1: accuracy の最大最小と平均値のまとめ

自己位置推定の手法	最大値	最小値	平均値
画像縮小	0.524	0.500	0.509
Canny	0.766	0.500	0.570
画像縮小 +Canny	0.660	0.500	0.525
LBP	0.56	0.500	0.509
画像縮小 +LBP	0.713	0.500	0.569
HOG	0.575	0.500	0.523

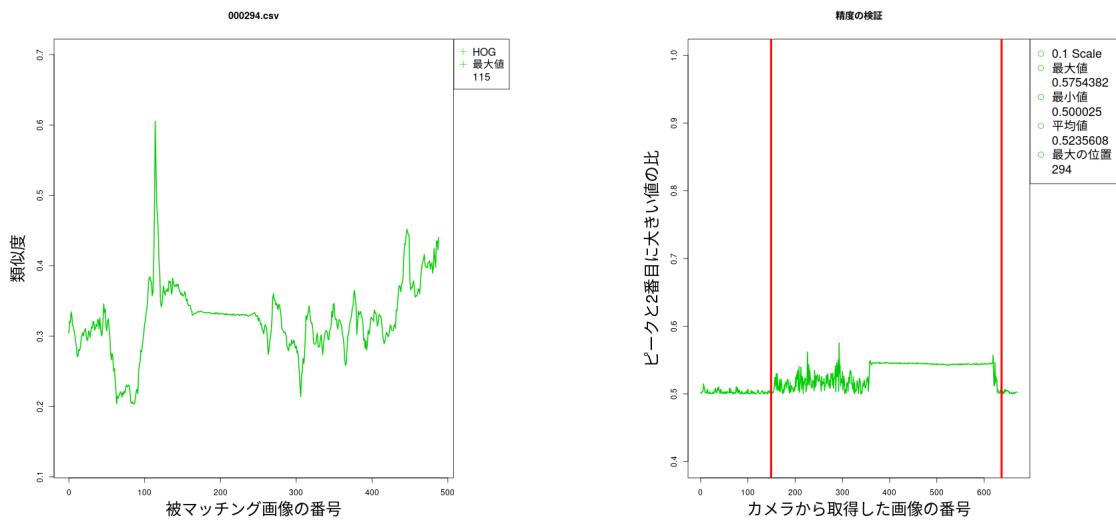


図 17: HOG 特徴のみを適用時の自己位置推定の結果と精度検証

図 12 から図 17 の左の画像はそれぞれの画像特徴を使用した時の自己位置推定の結果の中でも最も顕著なピークが得られた結果となっている。また、左の画像が精度検証によって得られたグラフとなっている。

実験の結果のグラフの中で最もピークがたっている部分がカメラから取得した風景画像と事前走行画像を照合した結果最も類似度が高いと識別された所である。そのため、ピークのところが現在地とみなされている風景画像である。

画像縮小のみと LBP 特徴のみ以外の 4 種類の画像特徴で顕著なピークを得られていることが図 12 から図 17 の左のグラフから分かる自己位置推定の結果のグラフの中で顕著なピークが得られている画像特徴は精度検証のグラフの中でも高い精度を得ることができている。

表 1 はそれぞれの画像特徴に対して行った精度検証の中の最大値、最小値、平均値をまとめたものである。表 1 を見ると精度が最も高い値を持っている画像特徴は Canny のエッジ検出のみを使用した時である。次いで画像縮小して LBP 特徴を使用した時となっている。平均値も同様の順になっている。平均値について、精度検証の最大値が近い値ではないのに画像縮小して Canny のエッジ検出と HOG 特徴のみを使用した時が近い値を持つことがわかった。

4 考察

実験結果から画像特徴ごとの自己位置推定の類似度のグラフに特徴が出ていることが分かった。ピークが顕著に現れている画像特徴は Canny のエッジ検出を使用した時、画像縮小と Canny のエッジ検出を照合した時、画像縮小と LBP 特徴を使用した時と HOG 特徴のみを使用した時である。それぞれの画像特徴でなぜピークが顕著に現れた理由を考えていくと Canny のエッジ検出を行う事で図 5 のような建物や山並みなどの輪郭線のみの画像となっており、輪郭線が一致した時と一致していない時では照合時の結果に差が出るのだと予想することができる。画像縮小と LBP 特徴を使用した時では、画像縮小によって画像内に写る建物などの大まかな形状を取り出すことができ、LBP 特徴による天候の変化による照明の影響を受けにくくするする特徴が上手く働いたのだと予想することができる。HOG 特徴のみを使用した時では、画像内の建物や看板などの形状をより強調することができており、形状の特徴と元の風景の二つの情報で照合を行っているため、一致している部分でピークが高くなっているのだと予想することができる。

次に精度検証によって得られた結果についてみていく。まず、精度検証の結果のグラフに描かれている 2 本の赤い縦線については、縦線と縦線の間の区間が今回使用した群馬大学周りの青い経路と赤い経路の同一区間であることを表している。表 1 より、accuracy の最大値が 0.6 を超えている画像特徴が Canny のエッジ検出のみ、画像縮小して Canny のエッジ検出と画像縮小して LBP 特徴の 3 つである。その中でも、Canny のエッジ検出のみと画像縮小して LBP 特徴を使用した時のグラフは似ており、画像縮小して Canny のエッジ検出をしたときは違うグラフとなっている。これは、同一区間の後半の部分では青い経路の画像軍は信号の前で停車している時の風景画像が映し出されているが、この時の画像照合ができているかどうかの差であると見られる。図 18 は青い経路の画像群で赤信号で停車中の時の自己位置推定の結果である。これを見ると Canny のエッジ検出のみの時と画像縮小して LBP 特徴の時にはピークが立っているが、画像縮小して Canny のエッジ検出をした時ではピークが乱立していることがわかる。実際に Canny のエッジ検出で画像縮小しない場合とした場合の画像照合で使用された画像を確認すると図 19 となっている。図 19 を見ると画像縮小した場合としなかった場合では検出できた輪郭線の密度に違いがみられる。そのため、本来不一致な部分でも輪郭線の密度の差から一致とみなされてしまったのではないかと予想される。

5 おわりに

本研究では自動運転の基幹技術の中のひとつである自己位置推定について、その手法の一つであるシーンマッチングに着目して人間と同じように今見ている風景を事前走行時に作成した風景画像のデータベースと照合することで現在地を推定できるかどうかを調べた。結果としては、画像特徴を使用、組み合わせることで照合の精度を上げ、より現在地に近い地点の画像を得ることができた。今後は実際に自動運転車両に搭載するにあたっての必要な要件満たし、さらに精度の高い自己位置推定の手法を模索していきたい。

6 参考文献

参考文献

- [1] SUBARU の総合安全, 株式会社 SUBARU, 2023-02-10 閲覧 <https://www.subaru.jp/safety/eyesight/>
- [2] Tesla MODEL3, Tesla Mortors, 2023-02-10 閲覧 https://www.tesla.com/ja_jp/model3
- [3] TOYOTA Advanced Drive 株式会社 TOYOTA, 2023-02-10 閲覧 https://toyota.jp/mirai/feature/?padid=from_mirai_top_topics_advanceddrive#ft-dspt
- [4] Canny, J., A Computational Approach To Edge Detection, IEEE Transactions on Pattern Analysis and Machine Intelligence, December1986 pp.679–698.
- [5] T. Ojala, M. Pietikainen, D. Harwood, "Performance evaluation of texture measures with classification based on Kullback discrimination of distributions", Proceedings of 12th International Conference on Pattern Recognition, Octorber 1994, vol. 1, pp. 582-586.
- [6] Faith Porikli,"Integral Histogram: A Fast Way to Extract Histograms in Cartesian Spaces" , IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2005, vol. 1, pp. 829-836.
- [7] N. Dalal, B.B Triggs, "Histograms of oriented gradients for human detection", 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), June 2005, vol. 1, pp. 886-893.

7 付録

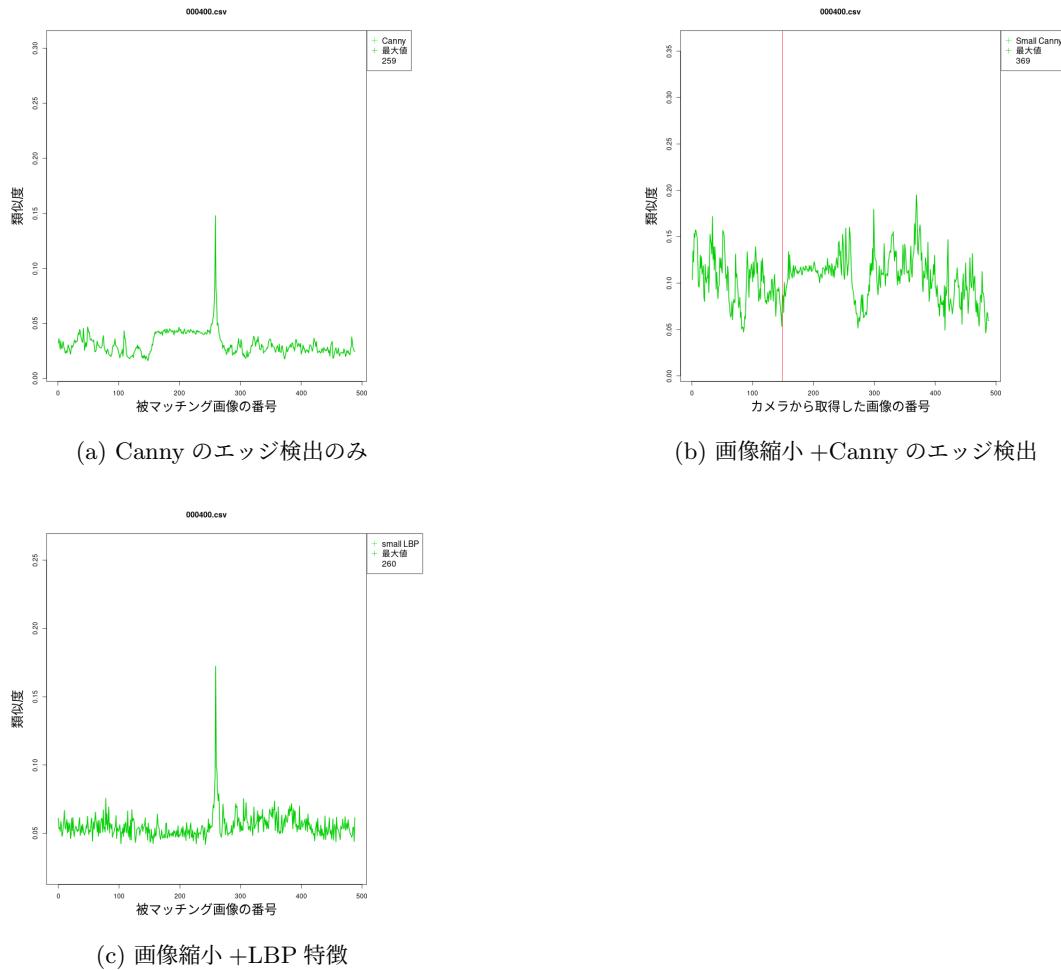
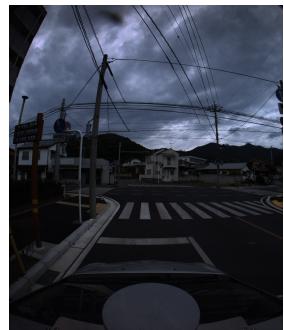


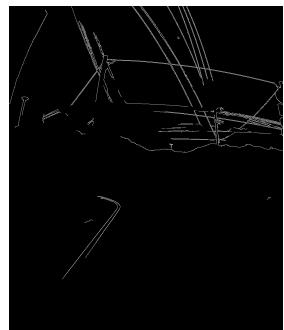
図 18: それぞれの画像特徴の 400 フレーム目の自己位置推定の結果



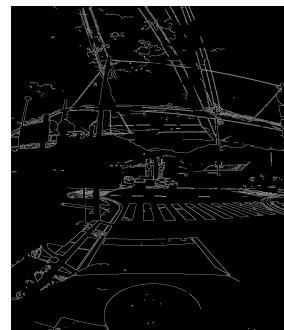
(a) 曇り



(b) 晴れ



(c) 曇り Canny のエッジ検出のみ



(d) 晴れ Canny のエッジ検出のみ



(e) 曇り 画像縮小 +Canny のエッジ検出



(f) 晴れ 画像縮小 +Canny のエッジ検出

図 19: 赤信号で停車している時の風景画像に画像特徴を適用した時の結果画像