

修士論文

インスタンスセグメンテーションと
3次元回転文字認識を用いた
車番認識の高精度化

令和5年度修了

三重大学大学院 工学研究科 情報工学専攻
ヒューマンコンピュータインタラクション研究室

佐々木鉄郎

はじめに

現代人にとって、自動車は移動手段として欠かせない存在である。また、自動車にはナンバープレートの取り付けが義務付けられており、駐車場の出入庫管理システムや手配車両の追跡システムなど、様々な場面でカメラで撮影された画像に対する車番認識システムが用いられている。しかし、これまでのシステムではカメラを固定設置する方法がとられているため、認識対象は一定の角度からの撮影画像に限定されている。したがって本研究では様々な角度から撮影された入力画像に対応する車番認識システムの構築を目指す。

車番認識システムは一般に、ナンバープレートの領域検出、領域内の文字検出、文字認識の3つのフェーズで構成されている [1]。プレート検出は、後のステージに影響を与える重要な処理である。従来の研究ではエッジ特徴と文字サイズによるプレート検出手法 [2] が提案されているが、車体の色とプレートの背景色が同じ場合や、多数の辺が検出される複雑な背景を持つ画像の場合には、誤検出する可能性がある。従って、中村らはナンバープレートの領域検出を行うことなく、入力画像から直接文字を検出し、ナンバープレートの文字認識を行う手法を提案した [3]。また、3次元回転文字認識 [4] [5] を用いることで、斜めから撮影された車番の認識を可能にした。この手法には、下記の三つの課題が存在する。

- 実行時間の長さ (入力画像 1 枚あたりの平均実行時間が 13 秒)
- 入力画像内においてナンバープレート領域が占める割合が閾値の範囲内である必要があること
- 入力画像内に複数のナンバープレートが含まれている場合、そのうちの 1 枚しか認識できないこと

そこで、本研究ではこれらの課題を解決すべく、深層学習による物体検知によりナンバープレート領域を検出した後に、その領域内でのみ文字検出を行い、3次元回転文字認識を行う新たな手法を提案する。プレート領域を検出するための深層学習には精度と速度を併せ持つ YOLOv5 [6] を用いる。YOLOv5 により抽出したプレート領域に対して、輝度の差が大きい部分を境界として領域を分割する Maximally Stable Extremal

Regions (MSER) [7] を適用し、文字候補を検出する。続いて、文字サイズやアスペクト比を用いて明らかに車番文字でない文字候補の棄却を行う。残った文字候補に3次元回転文字認識を適用し、文字の認識とその文字の3次元の回転角度推定を行う。残った文字候補にはまだ車番文字でないものが含まれている可能性がある。そこで、予め作成した車番テンプレートと文字候補の間で、文字サイズの比と回転角度を特徴量としたグラフマッチングを行うことにより、車番文字だけを出力する。文字検出や文字認識、その後の棄却方法について上記以外の様々な手法を検証した。それらについては本編で論ずる。

本研究では、Hsu ら [8] により公開された台湾のデータセット AOLP と、中村の研究のために撮影された画像群に本研究で新たに収集した画像を加えた独自データセット JLP を用いる。提案手法の実行時間は中村らの研究に比べて、データセット AOLP では 4.38 秒から 1.88 秒へ、データセット JLP では 27.2 秒から 2.13 秒へ短縮された。また、車番認識成功率は、データセット AOLP では 80.1% から 84.5% に、データセット JLP では 68.9% から 81.7% に向上した。プレートを検出した後、その領域内でのみ文字検出・認識を行うことで、大幅な処理速度の向上を実現した。また、インスタンスセグメンテーションにより画素単位でプレート領域を抽出することで、文字候補の誤検出が減少し認識精度向上に繋がった。

目次

はじめに	i
第 1 章 緒言	1
1.1 研究背景	1
1.2 先行研究	1
1.3 研究目的	2
1.4 データセット	4
第 2 章 提案手法	6
2.1 前処理 (プレート検出)	6
2.2 グラフマッチング手法	6
2.3 台形補正手法	8
2.4 辞書探索手法	9
第 3 章 関連技術	11
3.1 YOLOv5	11
3.2 Difference of Gaussian(DoG) フィルタ	11
3.3 Maximally Stable Extreamal Regions(MSER)	12
3.4 3次元回転文字認識	14
3.5 グラフマッチング	14
第 4 章 実験	20
4.1 YOLOv5 の学習	20
4.2 評価値	20
4.3 実験結果	21
4.4 考察	22
第 5 章 結言	24

目次		iv
5.1	まとめ	24
5.2	今後の課題	24
第 6 章	付録	26
謝辞		27

第 1 章

緒言

1.1 研究背景

現代人にとって自動車は必要不可欠な移動手段であり，そのすべての車両に車番(ナンバープレート)の取り付けが義務付けられている．そして，この車番を対象とするカメラベースの自動車番認識システムを用いて，様々な場面で自動車の照合を自動で行うべく研究がなされている．代表的な応用例として，オービスなどの道路交通法違反の取り締まりや駐車場での出入庫管理，パトロール車両による路上駐車中の手配車両の発見などが挙げられる．これらの場面を想定すると，車番は必ずしも真正面から撮影されているとは限らない．中村ら [3] は，斜めから撮られた車番を認識可能にするべく，3次元回転文字認識 [4] を用いた手法を提案している．

1.2 先行研究

従来の研究では，エッジ特徴と文字サイズによるプレート検出を行った後，文字検出，文字認識を行う手法が主流であった [2]．しかし，この手法は車番が長方形であるという特徴を利用しているので，斜めから撮影された画像の場合，プレート検出に失敗するという問題がある．そこで中村ら [3] は，プレート検出の過程を省き，入力画像全体に文字検出・文字認識を適用して車番を読み取る手法を提案した (図 1.1)．入力画像全体に対して，輝度の差が大きい箇所を境界として領域を切り取る MSER を適用し，文字候補を出力する．続いて探索ウィンドウをスライドさせながら，そのウィンドウ内に含まれる文字候補に 3次元回転文字認識を適用することで，文字の認識と 3次元の回転角度推定を行う．最後に，予め作成した車番テンプレートと文字候補の間で文字サイズの比と回転角度を特徴としたグラフマッチングを行うことで最終的な車番文字のみを出力する．この手法により，斜めから撮影された車番文字の認識を可能

にしている。評価実験では入力画像に対する車番サイズの比率が一定の条件を満たしたデータセットであれば車番認識率 96.3%が得られることを示している。この条件は探索ウィンドウのサイズが実際の車番サイズと同等でなければならないというものである。この制限に対して、入力画像に対する探索ウィンドウのサイズを 1%~100%に可変とすれば全サイズ的車番に対応できると考えられる。しかし、実際には車番ではない文字候補を車番文字として出力するリスクに加えて、実行時間が非常に長くなるという問題がある。(最低値を 5%, 最大値を 25%に設定し 640 × 480 の入力画像を処理する場合、平均で 13 秒を要する。) 中村らの研究における課題を以下に挙げる。

- 探索ウィンドウのサイズと大幅に異なるサイズ的車番を認識できないこと
- 実行時間の長さ
- 入力画像に複数の車番が写っていても一つしか認識できないこと

上記のような課題を解消すべく、本研究では新たな手法を提案する。

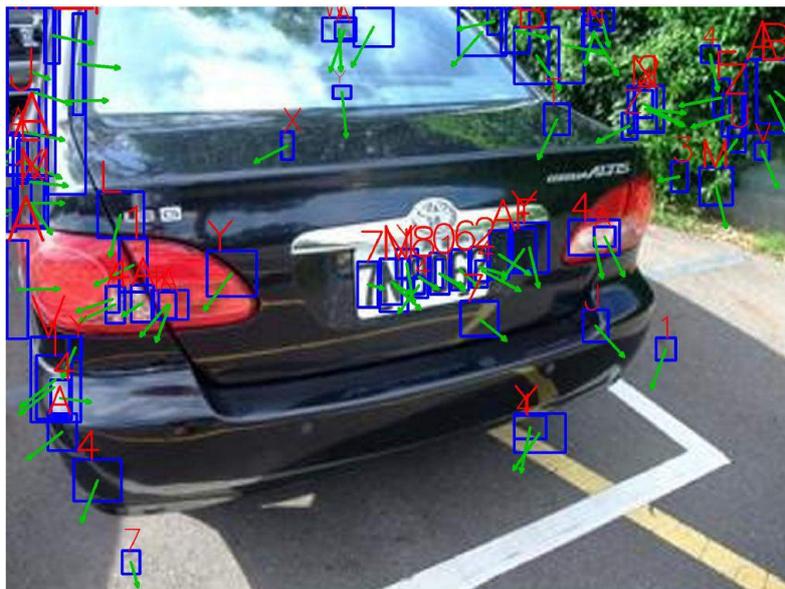


図 1.1: 文字認識・回転角度推定後に残った文字候補の例

1.3 研究目的

本研究では深層学習によりナンバープレート領域検出を行った後に、その領域内のみで文字検出を行い、3次元回転文字認識による文字認識を行う新たな車番認識手法を提案する。深層学習を用いることにより、様々なサイズ・撮影角度の車番に対応できると考えた。また、先にナンバープレート領域を検出することで、実行時間の短縮

を図る。複数の入力画像が含まれている入力画像に対しても、それぞれのナンバープレート領域で文字検出・文字認識を行うことができるという強みも存在する。

1.4 データセット

本研究では日本国内の車番を含むデータセットと台湾での車番を含むデータセットの2つを使用する. 表 1.1 に詳細を記す.

表 1.1: データセットの詳細

	地域	形式	サイズ	枚数
AOLP	台湾	jpg	640 × 480	582
JLP	日本	jpg	2,778 × 1,284	393

1.4.1 Application-oriented license plate (AOLP)

Hsu ら [8] により公開されたデータセットで, 台湾の車両を道路脇に設置された固定カメラにより撮影した低解像度画像群である. 画像例を図 1.2 に示す.

1.4.2 JLP

先行研究 [3] で用いていた独自データセットに本研究で新たに収集した画像を加えたものである. 所有者の許可を得た上で, スマートフォンのカメラで日本の車両を様々な角度から撮影している. 画像例を図 1.3 に示す.



図 1.2: データセット AOLP



図 1.3: データセット JLP

第2章

提案手法

本研究では高速化と複数のナンバープレート同時認識を実現するため、文字検出を行う前に、深層学習によるナンバープレート領域検出を行う手法を提案する。プレート検出後のプロセスでは、様々な比較・検討した。ここでは最も高い精度を示すグラフマッチングを用いた手法に加えて、台形補正によって画像を長方形に変換する手法、回転テンプレートとの重複度を用いる手法の3つを示す。

2.1 前処理 (プレート検出)

全ての手法に共通する処理として、初めに入力画像のナンバープレート領域を検出する。深層学習のモデルには、精度と速度を併せ持つ YOLO v5 [6] を用いる。プレート検出を行った出力画像例を図 2.1 に示す。ナンバープレートの領域を画素単位で検出し、外接枠がバウンディングボックスで表示されている。バウンディングボックスの左上に表示されている数値 (図 2.1 では 0.95) は検出されたオブジェクトの信頼度 (confidence score) である。信頼度は 0 から 1 の範囲で表現され、1 に近いほど信頼性が高いことを意味する。プレートの検出に失敗した場合、その時点で車番認識失敗となるため、プレート検出フェーズは最重要項目であるといえる。ここでプレートでないものをプレートとして認識 (誤検出) してしまうことがある。その場合、YOLOv5 が出力したプレート領域の外接枠のアスペクト比と面積を用いて、それらを棄却している。

2.2 グラフマッチング手法

この手法は、前処理で得られたプレート領域内の文字候補を検出・認識し、最後にグラフマッチングを行う手法である。プレート領域内は一定の輝度分布であるとは限らないため、ガウスフィルタでぼかした画像と原画像の差分画像を求めることにより、



図 2.1: YOLOv5 によるプレート領域検出例

画像の低周波成分を除去し、背景輝度を均一化する DoG (Difference of Gaussian) フィルタを適用する。中村らの先行研究 [3] では DoG フィルタを用いることによる文字検出精度の向上が示されている。

文字候補検出には、Matas らが提案した領域分割手法である Maximally Stable Extremal Regions (MSER) [7] を用いる。プレート領域に対して文字候補検出を行った例を図 2.2(a) に示す。この文字候補検出では、図 2.2(a) 左端のように、車番文字以外を文字として誤検出する可能性がある。この誤検出した文字候補は後のグラフマッチング工程により取り除かれる。

検出した文字候補に対して 3 次元回転文字認識 [4] により文字認識を行う。3 次元回転文字認識では文字候補に対して特徴ベクトル (加重方向指数ヒストグラム) を求め、疑似ベイズ識別関数で認識を行っている。3 次元回転文字認識を適用した結果の画像を図 2.2(b) に示す。図中の矢印は 3 次元回転文字認識で推定される文字印字面の法線ベクトルである。右斜め上から見下ろす方向で撮影した数字は左斜め下から見上げる方向で撮影した数字とほぼ同じ変形をするため、図 2.2(b) では後者の法線ベクトルが出力されている。

文字候補検出では車番文字でないものを文字として検出することがある。そのため、3 次元回転文字認識で算出した識別関数値や回転角度を用いて候補を絞るとともに、ナンバープレートに対する文字サイズの割合や文字の配置、文字同士の位置関係などを規定したテンプレートを様々な角度で回転させたものとの間でグラフマッチングを行うことで、文字候補から最も車番文字らしい組み合わせを選択する。詳細は 3.5 で論ずる。グラフマッチングを行った後の最終的な出力画像を図 2.2(c) に示す。

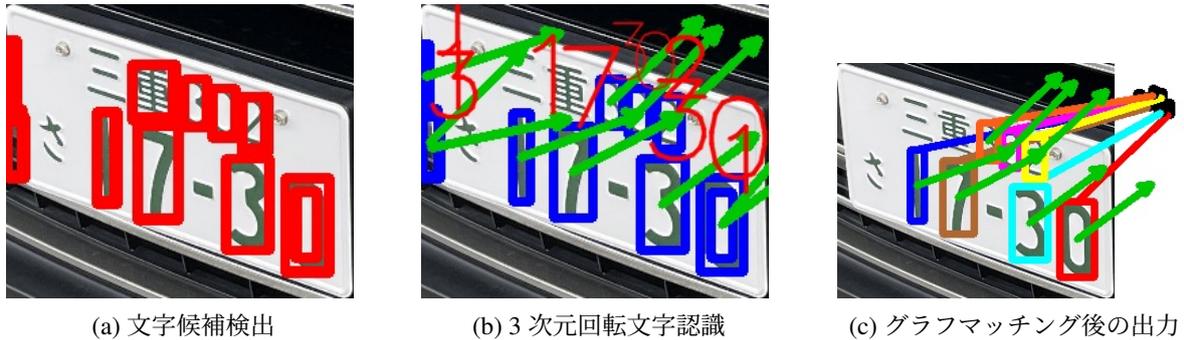


図 2.2: プレート検出後の流れ

2.3 台形補正手法

この手法はナンバープレート領域をあたかも正面から撮影したように変形した後、文字の検出・認識を行う手法である。前処理での YOLOv5 の出力はプレート領域を囲む多角形の頂点の集合である (頂点群を線で繋いでプロットしたものが図 2.3 である。) その頂点群に折れ線近似アルゴリズム (Douglas-Peucker アルゴリズム [10]) を適用することで実際のプレートの辺に沿った四辺形を予測する。Douglas-Peucker アルゴリズムは多角形を単純化するアルゴリズムの一つである。そのプロセスを下記に示す。

1. YOLOv5 の出力データから任意の折れ線の始点と終点をプロットの対象として決定する。
2. 視点と終点を通る直線 (セグメント) と、その折れ線に含まれる全ての頂点との距離をそれぞれ算出する。
3. その最大距離が許容距離 (ϵ) を超える場合、その点を新たなセグメントの頂点として追加する。
4. これを繰り返し、 ϵ 以上離れた点が存在しなくなれば終了する。

YOLOv5 のインスタンスセグメンテーションで出力された画素領域の輪郭にこの処理を適用することで 4 辺形へと変換することができる (図 2.4)。許容距離 ϵ はプレートのサイズの 1/20 に設定している。

次に、台形補正と呼ばれる画像処理技術を用いて斜めに写っているプレートを正面から撮影したように補正する。この時に用いる変換行列は、算出した四変形の 4 点とナンバープレートのアスペクト比から求められる。この変換行列を用いて画像全体を変形することで、あたかもほぼ正面から撮影されたような画像が得られる (図 2.5)。

最後にナンバープレートのレイアウトに基づき、文字候補検出と文字認識を適用することで車番を出力する。



図 2.3: YOLOv5 の出力



図 2.4: 折れ線近似適用後の四変形

2.4 辞書探索手法

この手法は予め様々な角度に回転させたナンバープレートの辞書を作成し、前処理で検出したプレート領域と照合することで車番文字の位置を推測する手法である。回転テンプレート辞書を下記の通り作成する。



図 2.5: 台形補正適用後のプレート画像

1. 真正面から見たナンバープレートの角の頂点 4 つと重心, そして車番文字の位置を定義する.
2. -50° から $+50^{\circ}$ の範囲内で, x 軸, y 軸, z 軸方向に 2° ずつ回転させた時の回転行列を算出する.
3. 1 のそれぞれの座標に, 算出した回転行列を適用し, 回転後のナンバープレートの頂点・重心・車番文字の重心座標を得る.
4. 回転後の 4 つの頂点, 重心, 車番文字の座標を csv ファイルに記録する.

ここから実際のシステムの処理を説明する. 前処理で得られたプレート領域と, 準備しておいた csv ファイルの回転後のプレート領域との重複度を計算する. まず, YOLOv5 が実際に検出したプレート領域と辞書のプレート領域のバイナリマスクを定義する. それらに対して IoU(Intersection over Union) を計算する.

1. 2 つのバイナリマスクの論理積 (AND 演算) を計算する. これにより, 2 つのマスクの交差領域 (重なり領域) を算出する.
2. 2 つのバイナリマスクの論理和 (OR 演算) を計算する. これにより, 2 つのマスクの合併領域 (どちらかのマスクに含まれる領域) を算出する.
3. 交差領域のピクセル数を合併領域のピクセル数で割ることで IoU を算出する.
4. 1.~3. を辞書内の全ての回転後プレートで実行し, 最も IoU が大きい回転テンプレートを選択する

選択した回転テンプレートに記録されている車番文字座標に基づき, 文字候補検出と文字認識を適用することで車番を出力する.

第 3 章

関連技術

3.1 YOLOv5

物体検出とは、与えられた入力画像から特定のオブジェクトを検出する技術である。そのオブジェクトは本研究ではナンバープレートにあたる。近年の物体検出は YOLO に加えて、Faster-RCNN, SSD, RetinaNet, CenterNet など深層学習を用いた手法が主流である。YOLOv5 [6] は 2020 年 6 月に Ultralytics [9] が公開した新しい物体検出手法である。"You Only Look Once" 「見るのは一度きり」が名前の由来である。予め画像全体をグリッド分割し、各領域ごとに物体のクラスとバウンディングボックスを求めため、YOLO v3 や他の深層学習モデルより処理が高速であるという特徴がある。本手法では扱っていないが、リアルタイムでの実行も可能である。YOLOv5 では 4 種類の大きさのモデル (小さい方から YOLOv5 s, YOLOv5 m, YOLOv5 l, YOLOv5 x) が提供されている。モデルサイズが大きい方が精度が高く、速度は遅い。公式ドキュメントに記されている。バッチサイズ 32 で 5000 枚の COCO val 2017 画像を推論した平均値によりモデルを比較したグラフを図 3.1 に示す。本研究では YOLOv5 m を用いる。

3.2 Difference of Gaussian(DoG) フィルタ

本研究ではプレート検出フェーズの後、文字候補検出を行う前に DoG フィルタを用いて輝度値の調整をする。用いる DoG フィルタは、元のグレースケール画像とそれに対してガウシアンフィルタをかけた画像との輝度値の差分を取り、低周波部分を取り除くフィルタである。図 3.2 に 1 次元画素列に対する DoG のイメージグラフを示す。一般的に DoG は、画像処理分野においてエッジ検出手法として用いられ、 σ の異なる二つのガウシアンフィルタを適用した画像の差分を取る。この際、輝度の変化領域のみを抽出することを目的としているため、算出する差分に対し絶対値を取ることが一

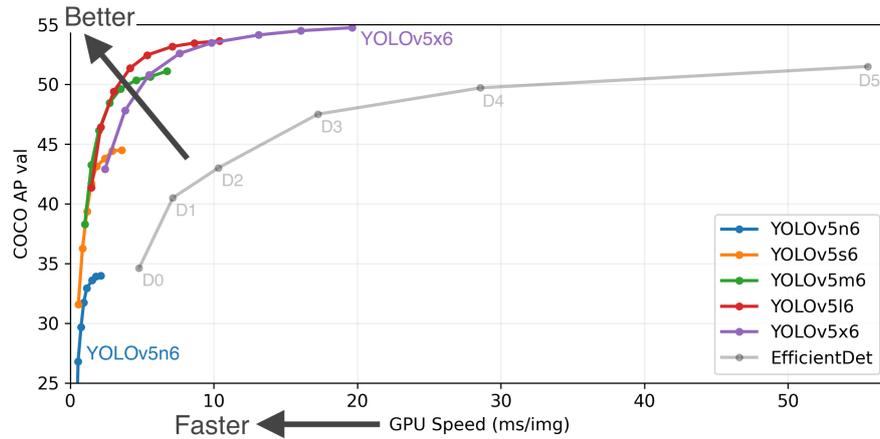


図 3.1: YOLOv5 のモデル比較

一般的である．図 3.2 においては，元画像とガウシアン画像との交点付近である．しかし，提案手法中の DoG フィルタは絶対値をとらず $-255 \sim 255$ の輝度値を $0 \sim 255$ に正規化して画像全体のなだらかな輝度変化を除去して均一にするフィルタとなっている．図 3.3 に DoG フィルタによる背景輝度の均一化の例を示す．図 3.3(a) の元画像に対し，ガウシアンフィルタをかけた画像 3.3(b) との差分を取り，輝度正規化を行うことで，図 3.3(c) のように背景輝度が均一化される．この後 2 値化することで，文字連結成分を安定して抽出することが可能となる．

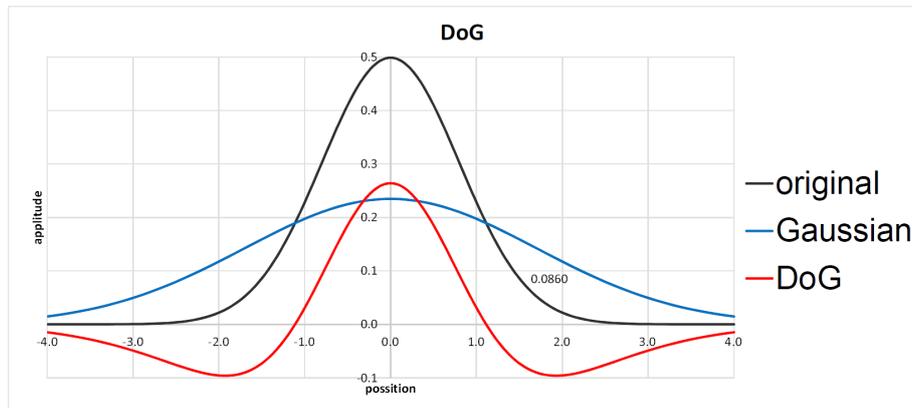


図 3.2: DoG のイメージグラフ

3.3 Maximally Stable Extremal Regions(MSER)

Maximally Stable Extremal Regions (MSER) [7] は，Matas らが提案した領域分割手法である．MSER の処理の流れは以下の通りである．

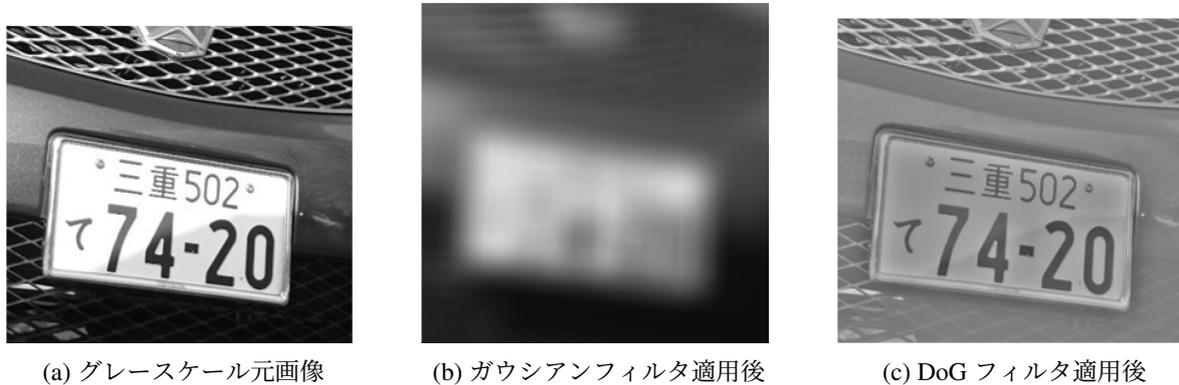


図 3.3: DOG フィルタによる背景輝度の均一化

1. 濃淡画像において、閾値を推移させながら、閾値毎に 2 値画像を生成する。
2. 各 2 値画像の連結成分を求める。
3. 任意の輝度閾値 t において、 $(t - \Delta t) \sim (t + \Delta t)$ における面積の変化が最も緩やかな連結成分を特徴領域とする。 Δt はパラメータとなっており、本研究では $\Delta t = 1$ としている。MSER においてこの値は最小の閾値範囲となっており、より多くの候補を抽出することを目的としている。

MSER は車番認識における文字検出に留まらず、一般的な文字検出において広く用いられている。文献 [8] 中では、車番文字の検出に MSER を用いるメリットとして以下のことが述べられている。

- 領域は、画像の歪みに関わらず、アフィン不変であることを示す連続的な(つまり射影的な)変換で閉じられる。これは、回転した車番文字の検出に適している。
- 輝度変化に頑健である。
- 検出には平滑化を必要とせず複数のスケールに渡り実行可能、すなわち、小領域、大領域の両方を検出可能である。

MSER では膨大な数の文字候補が検出されるため、文献 [3] の手法では車番文字の大きさとアスペクト比による候補棄却を行っている。また、全文字候補ペアに対し、重なり率 10%以上で認識結果の文字クラスが等しかった場合、識別関数値が大きい(より文字らしくない)文字候補を棄却する。さらに、2つの文字候補の重なり率が 80%以上の場合、外接矩形の面積が小さい文字候補を棄却する。

3.4 3次元回転文字認識

成田 [4], 蔵元ら [5] の3次元回転不変文字認識と回転角度推定について記す. 図 3.4, 図 3.5 に提案手法の処理の流れを示す. 3次元回転不変文字認識は, 大きく分けると学習と認識の2つのステージから成る. さらに, 学習は回転文字生成, 特徴抽出, 辞書作成から成り, 認識は検出, 文字認識, 回転角度推定から成る. はじめに, 標準文字に対して計算機上で3次元の回転処理を施し, 2次元平面に平行投影した回転文字を生成する. そして得られた回転文字から特徴抽出を行い, 文字のクラスごとに学習辞書を作成する. 一方で, 認識対象の文字に対しては回転などの補正は行わずに特徴抽出を行い, 得られた特徴ベクトルを回転文字を学習した辞書と照合することで認識を行う. ここで, 標準文字とは図 3.6(a) に示すような正面から撮影した回転や歪みのない文字のことを言い, 回転文字とは図 3.6(b)(c)(d) に示すような回転のある文字のことを言う.

< Learning Stage >

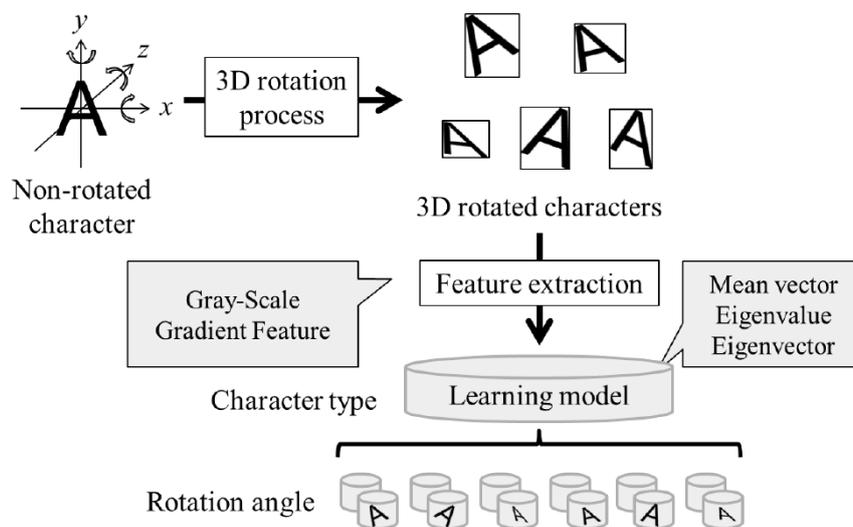


図 3.4: 学習

3.5 グラフマッチング

先行研究 [3] では, 予め車番テンプレートを様々な角度で3次元回転させたものと画像中を移動するスライドウィンドウ内の文字候補群との間で, グラフマッチングを行うことで, 3次元回転不変な車番認識手法を提案している. グラフマッチングとは,

< Recognition Stage >

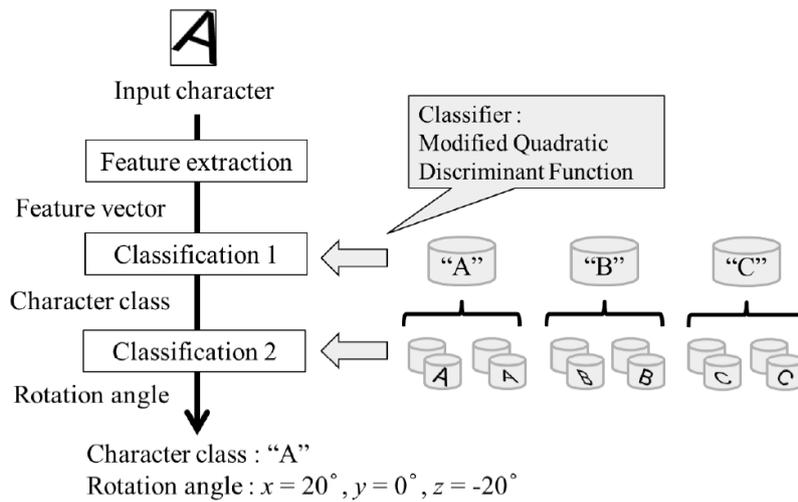


図 3.5: 文字認識・回転角度推定

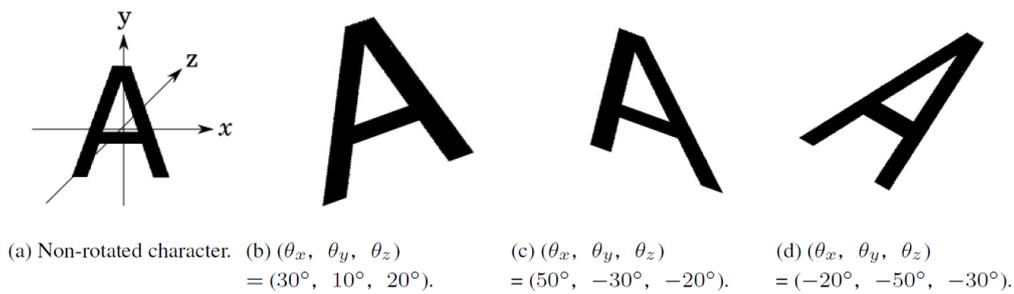


図 3.6: 3次元回転文字の例

グラフに関わる構造を反映させたグラフ頂点や辺の間の対応を確立することによって2つのグラフを比較する処理である。以下、予め参照グラフを生成する学習ステージと、車番認識時に参照グラフとのマッチングを行う車番認識ステージについて述べる。図 3.7 に提案手法の処理の流れを示す。

3.5.1 マッチンググラフ構成

本節では、グラフマッチングで用いるグラフの構成を述べる。有向グラフ G は、文字候補 (参照グラフにおいては車番文字) から成る頂点と、頂点を結ぶ辺によって構成され、それぞれ特徴量を持つ。頂点数を k とし、頂点特徴量の集合を $V = \{v_1, \dots, v_k\}^T$ 、辺特徴量の集合を $E = \{e_{12}, \dots, e_{(k-1)k}\}^T$ で表す。

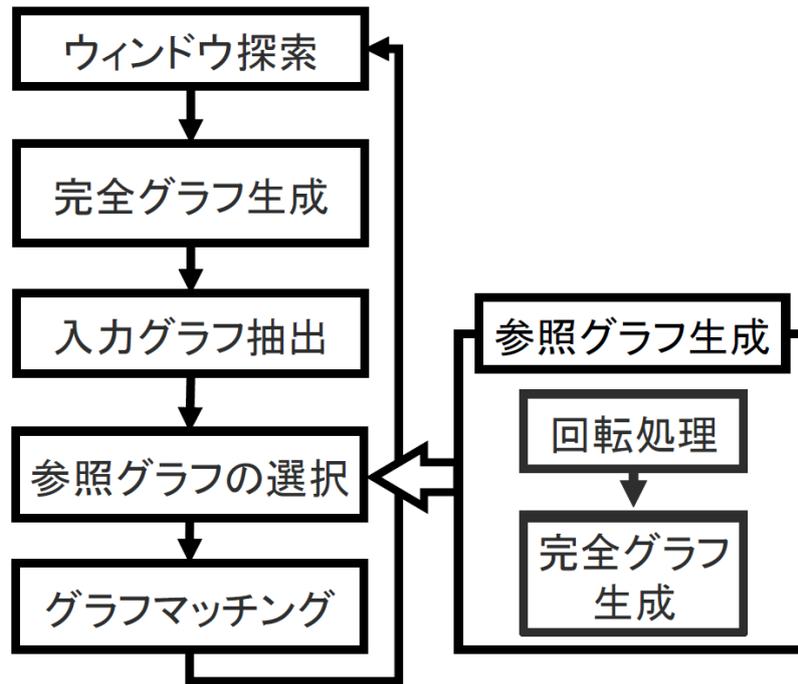


図 3.7: 提案手法の流れ

また, i 番目の頂点特徴量を以下で定義する.

$$\mathbf{v}_i = \begin{pmatrix} N_i \\ \theta_z \end{pmatrix}$$

ここで, N_i と θ_z はそれぞれ, 法線ベクトル, z 軸周りの回転角度である. ただし, 参照グラフ中の頂点は全て同一平面上の車番文字であるため, 全頂点の頂点特徴量が全て等しい. そこで, 参照グラフでは頂点特徴量の代わりに, グラフ全体の特徴量として, 車番プレートの法線ベクトルと z 軸周りの回転角度を持つものとする.

一方で, i 番目と j 番目の文字候補ペア ($i \neq j$) に対する辺特徴量を以下で定義する.

$$\mathbf{e}_{ij} = \begin{pmatrix} \mathbf{W}_{ij} \\ h_{ij} \\ d_{ij} \end{pmatrix}$$

$$h_{ij} = \frac{h_i}{h_j}$$

$$d_{ij} = \frac{\|\mathbf{p}_{ji}\|}{h_i}$$

$$\mathbf{w}_{ij} = \frac{\mathbf{p}_{ij}}{\|\mathbf{p}_{ji}\|}$$

ここで、 h_i , h_j はそれぞれ各文字候補の外接矩形の高さ、 p_{ij} は i 番目の文字候補の中心から j 番目の候補を結ぶベクトルである。すなわち、 h_{ij} , d_{ij} , w_{ij} はそれぞれ、接続された文字候補ペアの高さの比、中心間ユークリッド距離と高さの比、中心間を結ぶベクトルの単位ベクトルとなる。これら3つを辺特徴量として記述し、各頂点の位置関係を表現する。辺特徴量としてこれらを用いた理由を以下に示す。

1. 多くの車番プレート上の所定の位置の文字群は高さが一定で、クラス不変であるため。英語における小文字が利用される例は確認されていない。一方で文字の高さの代わりに面積を利用した予備実験も行ったが、面積は回転不変ではあるがクラス不変でないため、精度は悪化した。
2. 車番文字は隣接しあっており、距離情報が有効であるため。ただし、次節の通りに本手法は3次元回転不変ではあるがサイズ不変ではないため、高さで正規化している。高さを用いる理由は上記同様である。
3. 車番文字間を結ぶ線分の方向が有効であるため。ただし、距離情報は2で評価されるため単位ベクトルとした。

3.5.2 3次元回転車番テンプレートの生成

標準車番テンプレート (回転していない車番テンプレート) に対して計算機上で3次元の回転処理を施し、2次元平面に平行投影して斜めから見たテンプレートを生成する。標準車番テンプレートの例を図3.8(a)、生成した回転車番テンプレートの例を図3.8(b)~(d)に示す。 θ_x , θ_y , θ_z はそれぞれ x 軸, y 軸, z 軸まわりの回転角度である。色付きの長方形は車番文字の外接矩形である。次に、生成した各3次元回転車番テンプレートに対して前節のマッチンググラフ構成の完全グラフを生成することで、参照グラフ群が生成される。入力グラフの車番文字群から推定された車番プレート平面の3次元回転角度が最も近い参照グラフを、生成した参照グラフ群から選択してグラフマッチングを行う。これにより、様々な回転角度の車番レイアウトとマッチングを行うことが可能となり、3次元回転不変な車番認識を実現している。

3.5.3 完全グラフ生成・入力グラフ抽出

MSER で検出した文字候補から文字認識における識別関数値が小さい (文字らしい) 順に15個を取り出し、それらを頂点とする完全グラフ (入力完全グラフ) を生成する。次に、その完全グラフから車番文字数個の頂点を持つ全ての部分グラフを入力グラフとして抽出する。以下に続く処理を全ての入力グラフに対し行う。図3.9, 図3.10にそれぞれ完全グラフ生成, 入力グラフ生成例を示す。

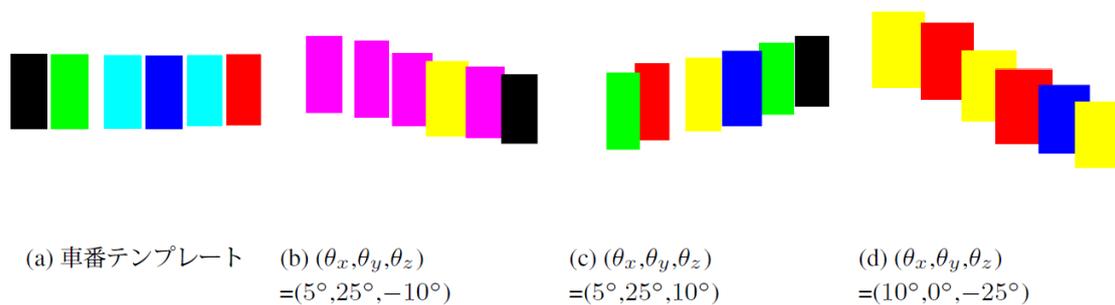


図 3.8: 3次元回転車番テンプレートの例

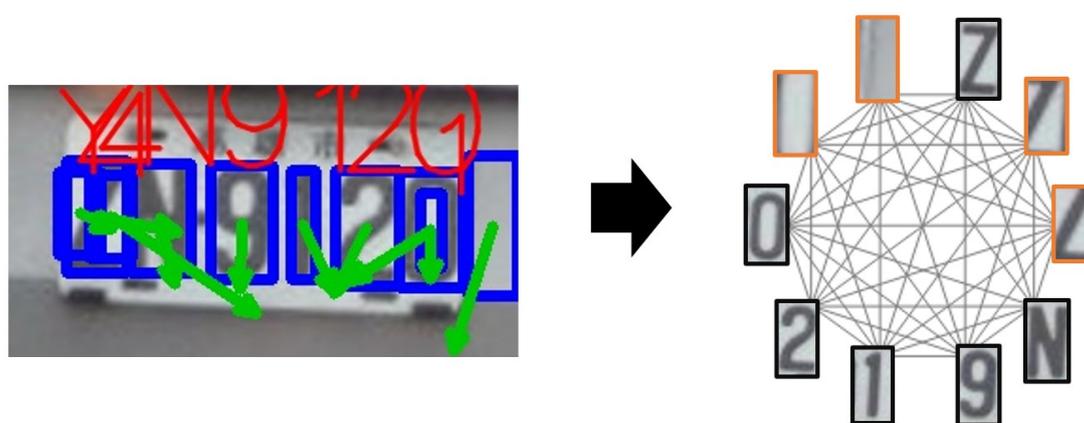
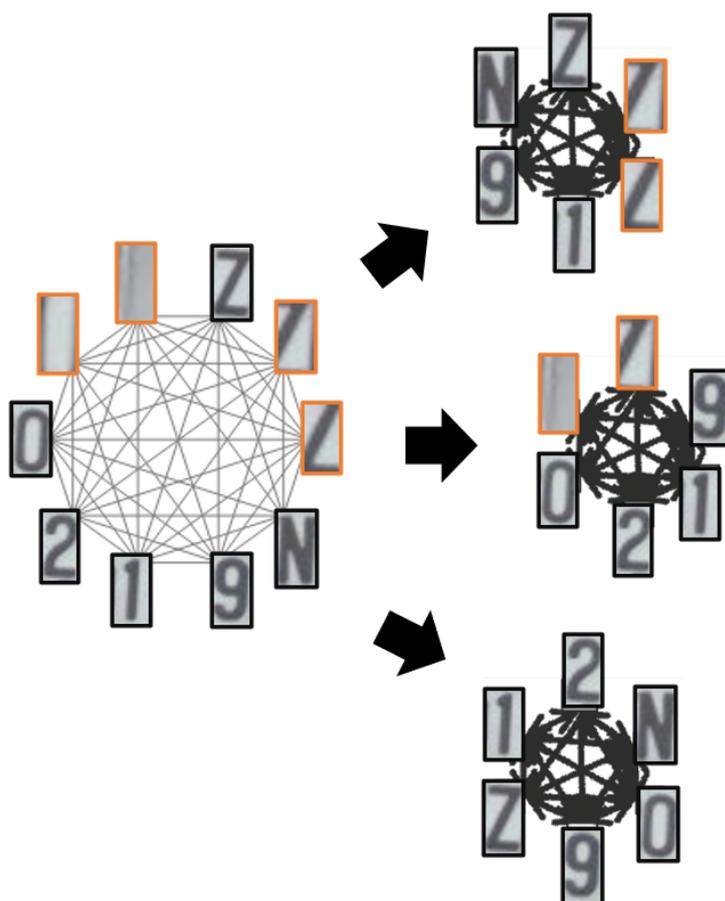


図 3.9: 完全グラフ生成



3.5.4 参照グラフの選択

入力グラフの頂点特徴量 (法線ベクトルと z 軸周りの回転角度) の平均から推定車番プレート回転角度を求め、参照グラフ群の中から回転角度が最も近い参照グラフを選択する。入力グラフの車番プレート候補の推定法線ベクトルと参照グラフの車番プレートの法線ベクトルがなす角 $\Delta\theta_n$ 、入力グラフの車番プレートの z 軸周りの推定回転角度と参照グラフの車番プレートの z 軸周りの回転角度との差 $\Delta\theta_z$ は、それぞれ

$$\Delta\theta_n = \cos^{-1}\left(\frac{\sum_{i=1}^{k_l} \mathbf{N}_i}{k_l} \cdot \mathbf{N}_t\right)$$

$$\Delta\theta_z = \left| \frac{\sum_{i=1}^{k_l} \theta_{zi}}{k_l} - \theta_{zt} \right|$$

となる。

ここで、 k_l は入力グラフの頂点数 (車番文字数)、 \mathbf{N}_i 、 θ_{zi} はそれぞれ、入力グラフにおける i 番目の文字候補の推定済み法線ベクトル、 z 軸周りの回転角度、 \mathbf{N}_t 、 θ_{zt} は参照グラフにおける車番プレートの法線ベクトル、 z 軸周りの回転角度である。

$\Delta\theta_n$ が最小となる参照グラフの中から、さらに $\Delta\theta_z$ が最小となる参照グラフを選択する。

3.5.5 グラフマッチング

参照グラフ T と入力グラフ X 間のマッチングコストを以下の式より算出する。

$$\text{cost}(T, X) = \sum_{i=1}^{k_l} \sum_{j=1}^{k_l} [\|h_{ij}^t - h_{ij}^x\| + \alpha \|d_{ij}^t - d_{ij}^x\| + \beta \{ \cos^{-1}(\mathbf{w}_{ij}^t \cdot \mathbf{w}_{ij}^x) \}]$$

ただし、 $i \neq j$ であり、予備実験の結果より、 $\alpha = 12$ 、 $\beta = 5$ とした。このとき、 T と X の頂点の対応は未知であるため、 T と X の頂点の全組み合わせについてコストを算出し、最小値を T と X のマッチングコストとする。

$$\text{Cost}(T, X) = \min\{\text{cost}(T, X_i)\}$$

以上の処理を、YOLOv5 で検出したプレート領域内の文字候補から生成した入力完全グラフから抽出された入力グラフに対して行い、さらに最小値を求める。最小値となったグラフの頂点を車番文字として出力する。ただし、マッチングコスト計算については、コストがその時点の最小コストより上回った時点で打ち切り、次の入力グラフに対する計算に移行することで高速化を図る。

第 4 章

実験

4.1 YOLOv5 の学習

提案手法のプレート検出フェーズでは共通して YOLOv5 m を用いる。また、2 種類のデータセット (AOLP, JLP) それぞれを学習用, 検証用, 評価用に分割する (表 4.1)。学習用に割り振った画像群にアノテーションを施し, YOLOv5 に学習させる。学習時のパラメータを表 4.2 に示す。学習済みモデルのナンバープレートの検出率 (Recall) は 0.96 となった。

表 4.1: データセットの分割

	train	valid	test
AOLP	382	100	100
JLP	170	100	123

表 4.2: YOLOv5 学習時のパラメータ

バッチサイズ	8
エポック数	300
リサイズ解像度	384 × 640

4.2 評価値

車番認識において、車番中の 1 文字でも認識に失敗した場合、同一プレート中の他の文字を認識していても、1 つの車番を認識できたとはいえない。従って 1 つのプレート内の全ての文字を認識できたプレート単位の認識率「車番認識率」(式 4.1) を評

価値の一つとする。ただし、本研究で使用しているデータセット AOLP の公開元である Hsu ら [8] の実験では、単純な文字単位での車番文字認識率を採用している。比較を行うため、文字単位の認識率「車番文字認識率」(式 4.2) を二つめの評価値とする。また、従来手法の課題であった実行速度を比べるために平均実行時間 (式 4.3) も評価値として用いる。なお、車番文字とはナンバープレート中の英数字とする。

$$\text{車番認識率 (\%)} = \frac{\text{全ての車番文字を認識成功したプレート数}}{\text{全入力画像に含まれるプレートの総数}} \times 100 \quad (4.1)$$

$$\text{車番文字認識率 (\%)} = \frac{\text{全ての車番文字数}}{\text{車番文字の総数}} \times 100 \quad (4.2)$$

$$\text{平均実行時間 (s)} = \frac{\text{全入力画像の実行時間}}{\text{入力画像枚数}} \quad (4.3)$$

4.3 実験結果

台湾のデータセット AOLP 及び日本のデータセット JLP に対する実行結果を表 4.3, 表 4.4 に示す。

表 4.3: データセット AOLP に対する実行結果

	車番認識率	車番文字認識率	平均実行時間
Hsu らの手法 [8]	-	86.6 %	-
中村らの手法 [3]	80.1 %	88.9 %	4.38 s
台形補正手法	83.4 %	86.1 %	0.95 s
辞書探索手法	83.9 %	88.3 %	2.02 s
グラフマッチング手法	84.5 %	89.3 %	1.88 s

表 4.4: データセット JLP に対する実行結果

	車番認識率	車番文字認識率	平均実行時間
中村らの手法 [3]	68.9 %	75.3 %	27.23 s
台形補正手法	76.4 %	82.2 %	1.10 s
辞書探索手法	78.6 %	88.0 %	3.06 s
グラフマッチング手法	81.7 %	91.9 %	2.13 s

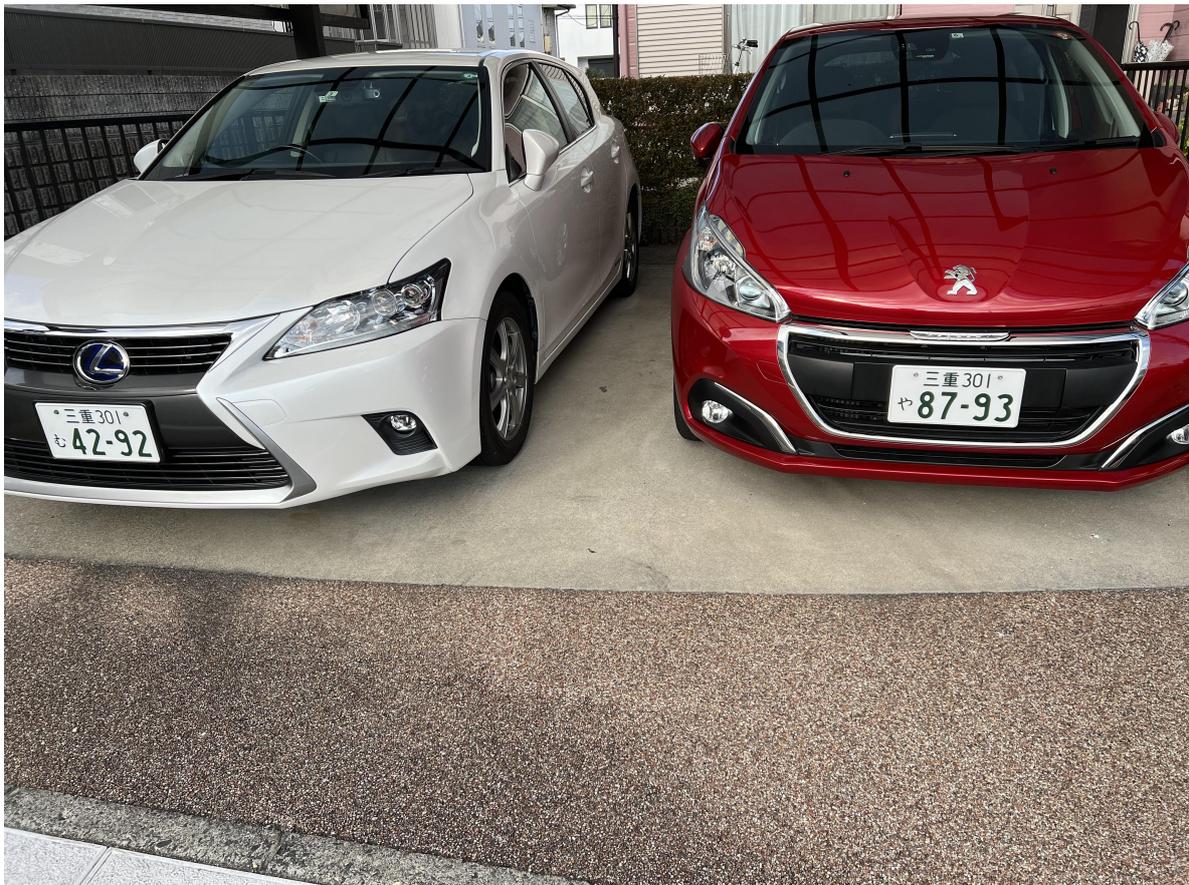
データセット AOLP において、車番認識率は中村らの手法 [3] の 80.1% に比べて、提案手法全てにおいて向上した。中でもグラフマッチング手法が最も高く、84.5% で

あった。車番文字認識率は Hsu らの手法 [8] の 86.6% に対して、中村らの手法 88.9%、グラフマッチング手法が 89.3% となっている。車番文字認識率において提案手法の中で中村らの手法からの向上が見られたのはグラフマッチング手法のみである。平均実行時間は、全ての提案手法で中村らの手法よりも短縮された。中でも台形補正手法が最も高速であった。

データセット JLP において、車番認識率は中村らの手法 68.9% に対して、提案手法全てにおいて向上した。中でもグラフマッチング手法が最も高く、81.7% であった。車番文字認識率は中村らの手法の 75.3% から、グラフマッチング手法では 90.7% とこちらも大きく向上した。平均実行時間は、全手法で大幅な高速化を実現した。中でも台形補正手法は 1.10 秒であり、先行研究の 27.23 秒と比べて約 24.8 倍の処理速度である。車番認識率が最も高いグラフマッチング手法は 2.13 秒であり、中村らの手法と比べると 12.8 倍の処理速度である。

4.4 考察

まず、車番テンプレートを変更するだけで、3つの提案手法が外国のナンバープレートにも対応できることを確認できた。先行研究では入力画像全体で文字検出を行った後に車番らしい組み合わせの文字を出力するが、3つの提案手法では YOLO v5 によるプレート検出フェーズの導入により、プレート領域でのみ文字検出を行っている。従って、入力画像 1 枚当たりの実行時間を二つのデータセット双方にて減少させることができた。特に画像サイズが $2,778 \times 1,284$ と大きいデータセット JLP に関しては、大幅な高速化を実現した。これは、先行研究の手法で用いられているウィンドウ探索に時間がかかっていたためである。また、プレートの誤認識が減少し、全ての提案手法で車番認識精度が向上した。最も処理速度が高速であったのは台形補正手法である。その理由として、処理に時間のかかるグラフマッチングや回転テンプレート探索を行っていないことが挙げられる。車番認識率は両方のデータセットにてインスタンスセグメンテーションとグラフマッチングによる提案手法が最も高い結果となった。台形補正手法は YOLOv5 によるインスタンスセグメンテーションが出力する多角形の精度に大きく影響されるため、データセットの補強により精度を高めると予想している。また、従来手法では不可能であったナンバープレートを 2 枚以上含む入力画像の車番認識に対しても車番文字認識を可能にした (図 4.1)。



(a) 入力画像



(b) 出力 1



(c) 出力 2

図 4.1: 複数の車番を含む入力画像の車番認識

第5章

結言

5.1 まとめ

本研究では斜めから撮影されたナンバープレート内の車番文字認識の精度向上を目指して、YOLOv5によるナンバープレート領域のインスタンスセグメンテーションと3次元回転文字認識を用いた車番認識手法を提案し、その有効性を検証した。グラフマッチングを用いる手法、折れ線近似アルゴリズムと台形補正を用いてプレートの方向を正面に直してから文字検出・認識を行う手法、テンプレートを回転させた辞書とプレート領域との重複度を算出し車番文字の位置を予測する手法を比較・検討した。先行研究の手法と比べて全ての手法で車番認識率の向上を実現した。特にグラフマッチングを用いた手法が最も高い精度を誇り、台湾の車番では80.1%から84.5%、日本の車番では68.9%から81.7%に向上した。これにより、グラフマッチングの有効性を示すことができた。また、プレート領域でのみ文字候補検出・文字認識を行うことで実行時間の大幅な短縮を実現した。

5.2 今後の課題

本研究では、主に英数字を対象に車番認識の研究を行ってきた。しかし、国内での実用性を高めるためには、英数字だけではなく、地名や平仮名を含む車番の認識に対応する必要がある。

また、本研究では静止画に基づく車番認識に焦点を当てたが、実際の応用を考えた場合、動的な環境下での車番認識、すなわち動画からの車番認識が求められる。動画を用いた認識では、車両が移動することによる車番の見え方の変化やモーションブラー、ぼけや光の反射など、静止画では考慮しない要素が影響を与える。これらの問題に対処するためには、動的なシーンに適応した認識技術の開発が必要である。YOLOv5ア

ルゴリズムは動画データに対応しており，バウンディングボックスの検出をリアルタイムで行うことができるため，動画からの車番認識への応用が期待される．しかし，動画データを用いた学習には，静止画と比較してさらに多くのデータセットが必要となり，その作成は大きな課題である．特に，実際の道路環境を反映した多様な状況下での動画データの収集と，それに伴うアノテーション作業は大きな労力を必要とすると考えられる．

第 6 章

付録

本研究に関するプログラムは全て以下のディレクトリ下に存在する.

```
/net/nfs2/export/home/sasaki/master_directry
```

実行環境

OpenCV3.0, clang14

※ OpenCV の export が必要な場合あり

例：export PKG_CONFIG_PATH="/opt/homebrew/opt/opencv@3/lib/pkgconfig"

C++コード実行方法 (GraphMatching 手法など)

/Method.Graphmatching/codes/graphMatching で make コマンド実行. (\$ make all)

./bin に移動し (\$ cd bin) 生成された ./lpr3docr を実行. (引数は measure.sh を参考)

./bin 内の measure.sh を実行することで複数の画像を 1 度に入力可能.

Python コード実行方法 (その他の手法)

```
$ python3 main_○○.py 0 1 7
```

第 1 引数：実行したい手法の main コード

第 2 引数：0 に固定

第 3 引数：0 →デバッグ画像なし, 1 →デバッグ画像あり

第 4 引数：車番文字の数

謝辞

本研究を進めるにあたって、様々な助言やご教授を下さった若林哲史教授、盛田健人助教、白井伸宙助教に深く感謝いたします。また、ミーティングなどで研究についてご指導頂いた研究室の先輩方、同期の皆様に感謝します。最後に、ここまでの大学生活を支えてくれた家族に感謝の意を表して、本論文の結びといたします。

参考文献

- [1] Du, S., Ibrahim, M., Shehata, M. and Badawy, W.,” Automatic license plate recognition (ALPR): A state-of-the- art review,” 2013, Circuits and Systems for Video Technology, IEEE Transactions on, 23(2), 311-325.
- [2] S.-L. Chang, L.-S. Chen, Y.-C. Chung, and S.-W. Chen, “Automatic license plate recognition,” IEEE Trans. Intell. Transp. Syst., vol. 5, no. 1, pp. 42-53, Mar. 2004.
- [3] 中村悠暉, “3次元回転文字認識の車番認識への応用に関する研究“, 三重大学大学院工学研究科, 修士論文, 2016.
- [4] 成田了, 大山航, 若林哲史, 木村文隆, “3次元回転不変カメラベース文字認識,” 電気学会論文誌 C, vol.130, no.4, pp.876-882, 2013.
- [5] 蔵元侃太, 大山航, 若林哲史, 木村文隆, “カメラで撮影された文字の3次元回転角度推定,” 電気学会論文誌 C, vol.134, no.12, pp.1817-1823, 2014.
- [6] <https://github.com/ultralytics/yolov5>
- [7] J. Matas, O. Chum, M. Urban, and T. Pajdla, “Robust wide baseline stereo from maximally regions,” Proc. Brit. Mach. Vis. Conf., pp. 384-393, 2002.
- [8] Hsu, Gee-Sern, Jiun-Chang Chen, and Yu-Zu Chung. “Application-oriented license plate recognition,” Vehicular Technology, IEEE Transactions on 62.2 (2013): 552-561.
- [9] <https://ultralytics.com>
- [10] Wu, Shin-Ting, Marquez, Mercedes (2003). “A non-self-intersection Douglas-Peucker algorithm,” 16th Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI 2003), Sao Carlos, Brazil: IEEE. pp. 60–66. CiteSeerX 10.1.1.73.5773. doi:10.1109/SIBGRA.2003.1240992. ISBN 978-0-7695-2032-2. S2CID 10163908.