

博士論文

機械学習を用いたナンバープレート
複合劣化画像の画質改善と文字認識
に関する研究

**(Study on Image Quality Improvement and
Character Recognition for Composite Degradation
of License Plate Images Using Machine Learning)**

2018年9月

立命館大学大学院理工学研究科
電子システム専攻博士課程後期課程

辻 広生

立命館大学審査博士論文

機械学習を用いたナンバープレート複合劣化画像の
画質改善と文字認識に関する研究

**(Study on Image Quality Improvement and
Character Recognition for Composite Degradation
of License Plate Images Using Machine Learning)**

2018年9月

September 2018

立命館大学大学院理工学研究科

電子システム専攻博士課程後期課程

Doctoral Program in Advanced Electrical,

Electronic and Computer Systems

Graduate School of Science and Engineering

Ritsumeikan University

辻 広生

TSUJI Hiroo

研究指導教員：道関 隆国 教授

Supervisor : Professor DOUSEKI Takakuni

要旨

本論文は監視カメラにより撮影されたナンバープレート複合劣化画像を対象とし、ナンバープレート文字読み取りシステムに対するロバスト性及び効率性を高める画質改善法、並びにシステムの適応性を高める文字認識法について提案したものである。

画質改善法としては、第一にシステムのロバスト性を高めるため、ナンバー灯による環境光の外乱がある複合劣化画像を対象とする **Retinex** 処理を提案した。この手法では安定性が高い機械学習法であるサポートベクター回帰を用いて画像中の環境光成分の推定精度を高めることにより外乱除去効果を向上させた。提案の **Retinex** 処理と従来の **Retinex** 処理との比較実験の結果、提案の **Retinex** 処理は、**Retinex** 処理の共通の副作用である **Halo** 作用を低減させ、従来の **Retinex** 処理に比べて文字認識の正答率を 25 パーセント以上、向上させた。

第二にシステムの効率性を高めるため、文字情報が少ない複合劣化画像を対象とするマルチフレーム超解像処理を提案した。この手法では信頼性が高い機械学習法であるサポートベクターマシンを用いて適切な画素値を選択することにより文字情報の再構成効果を向上させた。提案の超解像処理と従来の超解像処理との比較実験の結果、提案の超解像処理は、エッジ復元効果を向上させ、従来の超解像処理に比べて文字認識の正答率を 32 パーセント以上、向上させた。

文字認識法としては、システムの適応性を高めるため、複合劣化画像に映った形状が複雑で種類が多い文字を対象とする畳み込みニューラルネットワーク (CNN) による文字認識法を提案した。この手法では柔軟性が高い機械学習モデルである CNN を多重構造として CNN に入力画像の解像度調整機能と候補数の適応的制御機能を与えることにより複合劣化に対する適応性能を改善した。提案の文字認識法と従来の文字認識法との平仮名認識における比較実験の結果、提案の文字認識法を用いた場合の正答率は、従来の文字認識法を用いた場合の正答率の約 1.17 倍となり、提案の文字認識法でのみ第 2 候補以内で 90 パーセ

ント以上の正答率が得られ，実用レベルの認識性能が達成された．

最後に本論文で提案した全手法を統合することで，文字認識の正答率が従来手法に比べて，粗い見積もりではあるものの約 2 倍に向上し，さらに実用レベルの認識性能でナンバープレートの平仮名を認識可能とすることで従来手法を適用した場合に比べて車両の絞り込み台数を 2.4 パーセント以下に低減できることを実証し，これによって車両の検索に要する人的負担と費用が劇的に軽減されることを明らかにした．

目次

1. 序論

1.1 監視カメラの歴史	1
1.2 監視カメラの設置台数の推移	2
1.3 ナンバープレート文字読み取りシステム歴史	3
1.4 ナンバープレート画像の画質改善法の歴史	5
1.5 ナンバープレート文字認識法の歴史	5
1.6 ナンバープレート画像の画質改善法と文字認識法の開発経緯	6
1.7 本研究の目的および課題	10
1.8 本論文の構成	10

2. サポートベクター回帰を用いた Retinex 処理による輝度値補正法

2.1 まえがき	17
2.2 一般的な輝度値補正法	18
2.2.1 トーンカーブによる輝度値補正法	18
2.2.2 ヒストグラム平坦化による輝度値補正法	19
2.3 従来の Retinex 処理	21
2.3.1 Retinex モデル	21
2.3.2 従来の Retinex 処理	22
2.4 サポートベクター回帰を用いた Retinex 処理	25
2.4.1 全体的な処理手順	26
2.4.2 サポートベクター回帰を用いた Retinex 処理の利点	26
2.4.3 サポートベクター回帰の実装方法	28
2.5 Halo 作用低減効果の評価	30
2.5.1 評価の条件	30
2.5.2 Halo 作用低減効果	31
2.5.3 考察	31

2.6	画質の定量的評価実験	32
2.6.1	実験条件	32
2.6.2	実験に用いた畳み込みニューラルネットワークの構成	35
2.6.3	画質改善効果	37
2.6.4	考察	38
2.7	まとめ	38
3.	サポートベクターマシンを用いた画素値選択型マルチフレーム超解像処理による高解像度化法	
3.1	まえがき	43
3.2	従来の超解像処理	44
3.2.1	マルチフレーム超解像処理の基本概念	44
3.2.2	サブピクセル位置合わせの方法	45
3.2.3	マルチフレーム超解像画像の推定方法	47
3.3	サポートベクターマシンを用いた画素値選択型マルチフレーム超解像処理	48
3.3.1	全体的な処理手順	50
3.3.2	画素値の選択方法	51
3.3.3	画素値選択型マルチフレーム超解像処理の利点	52
3.3.4	サポートベクターマシンを用いたサブピクセル位置合わせの方法	53
3.3.5	サポートベクターマシンを用いたテンプレート画像の選択方法	55
3.4	エッジ復元効果の評価	57
3.4.1	評価の条件	57
3.4.2	エッジ復元効果	58
3.4.3	考察	59
3.5	画素値選択効果の評価実験	59
3.5.1	実験条件	59
3.5.2	画素値選択効果	60
3.5.3	考察	61
3.6	まとめ	62

4. 多重構造の畳み込みニューラルネットワークによる文字認識法	
4.1 まえがき	65
4.2 一般的な畳み込みニューラルネットワークの構成	66
4.3 一般的な畳み込みニューラルネットワークによる文字認識法の問題 ...	67
4.4 従来の複数候補選択法	70
4.5 多重構造の畳み込みニューラルネットワークによる文字認識法	70
4.5.1 多重構造の畳み込みニューラルネットワークの全体構成	71
4.5.2 多重構造の畳み込みニューラルネットワークによる文字認識法の 利点	73
4.5.3 畳み込みニューラルネットワークの実装方法	74
4.6 認識性能改善効果の評価実験	76
4.6.1 実験条件	76
4.6.2 認識性能改善効果	78
4.6.3 考察	80
4.7 まとめ	81
5. 評価	
5.1 まえがき	85
5.2 ロバスト性の評価実験	85
5.2.1 実験条件	85
5.2.2 評価方法	91
5.2.3 外乱の除去効果	92
5.2.4 考察	95
5.3 効率性の評価実験	97
5.3.1 実験条件	97
5.3.2 評価方法	98
5.3.3 文字情報の再構成効果	98
5.3.4 考察	100

5.4 適応性の評価実験	100
5.4.1 実験条件	100
5.4.2 評価方法	101
5.4.3 複合劣化に対する適応性能	101
5.4.4 考察	103
5.5 実環境での検証実験	103
5.5.1 実験条件	103
5.5.2 評価方法	107
5.5.3 実環境での性能	107
5.5.4 考察	109
5.6 全実験についての考察	111
5.6.1 正答率に対する効果についての考察	111
5.6.2 車両の絞り込み台数に対する効果についての考察	112
5.7 まとめ	113
6. 結論	117
7. 今後の課題	
7.1 まえがき	119
7.2 ナンバープレート領域の抽出性能の向上	120
7.3 ナンバープレートの傾き補正性能の向上	121
7.4 ナンバープレート文字領域の抽出性能の向上	122
7.5 実環境におけるナンバープレート文字読み取りシステムの性能確認	123
7.6 動いている車両のナンバープレートを対象とする場合の対策	124
7.7 まとめ	126
謝辞	129
本研究に関する発表文献リスト	131

目次

図 1.1	日本における防犯設備機器の推定市場売上高の推移	2
図 1.2	ナンバープレート文字読み取りシステムの一般的な処理手順	3
図 1.3	本論文の構成	11
図 2.1	折れ線型トーンカーブによる輝度値補正例 (a)入力画像, (b)出力画像	18
図 2.2	補正に用いたトーンカーブ	18
図 2.3	ヒストグラム平坦化例 (a)入力画像, (b)出力画像	20
図 2.4	入出力画像の輝度値ヒストグラム (a)入力画像, (b)出力画像	21
図 2.5	Retinex モデルの模式図	22
図 2.6	SSR 処理の手順	24
図 2.7	提案 Retinex の全体的な処理手順	26
図 2.8	サポートベクター回帰を用いた Retinex 処理の利点の説明図 (a)観測画像, (b1)C/S 画像, (b2)SVR 画像, (c1) (b1)の 3 次元表示図, (c2) (b2)の 3 次元表示図	27
図 2.9	Retinex 処理画像 (a)補正前の画像, (b)SSR 処理後の画像, (c) 2MSR 処理後の画像, (d) 提案 Retinex を行った後の画像	31
図 2.10	テンプレート画像	34
図 2.11	学習画像の例 (a)例 1, (b)例 2, (c)例 3	35
図 2.12	CNN のネットワーク構造	36
図 2.13	各 Retinex 処理画像の正答率	37
図 3.1	マルチフレーム超解像処理の基本概念を表す模式図	45
図 3.2	フィッティング関数によって求まるサブピクセル位置 (a)等角直線 フィッティング, (b)パラボラフィッティング	47
図 3.3	提案超解像における分類モデルの模式図	50
図 3.4	提案超解像の全体的な処理手順	51

図 3.5	画素値選択型マルチフレーム超解像処理の利点の説明図 (a)3次元表示図, (b)2次元表示図, (c)注目線上の画素値変化を表す模式図	53
図 3.6	離散的なサブピクセル位置の定義の例 (a)ピクセル1の擬似的高解像度化, (b)ピクセル2の擬似的高解像度化, (c)互いにサブピクセル(1/6ピクセル)の位置ずれがあるピクセル1とピクセル2	55
図 3.7	画素値を選択した超解像画像の生成例 (a)低解像度画像, (b)テンプレート画像 (c)画素値を選択した超解像画像	56
図 3.8	対象画像	58
図 3.9	学習画像の例	58
図 3.10	超解像画像の例 (a)従来超解像の画像, (b)提案超解像の画像	59
図 3.11	生成した高解像画像 (a)画素値を選択していない高解像度画像, (b)画素値を選択した高解像度画像	60
図 3.12	各高解像度画像の振幅スペクトル	61
図 4.1	CNNの構成を表す模式図	66
図 4.2	解像度と正答率との関係	67
図 4.3	テスト画像の劣化強度と正答率との関係	69
図 4.4	提案の文字認識法における認識モデルの模式図	71
図 4.5	全体構成を示す模式図 (a)学習時, (b)テスト時	72
図 4.6	多重解像度画像の具体的な階層	73
図 4.7	テスト画像の劣化強度と平均候補数との関係	74
図 4.8	平均候補数と正答率との関係 (a)劣化強度9の場合 (b)劣化強度11の場合, (c)劣化強度13の場合	79
図 4.9	平均候補数と正答率の比との関係	80
図 5.1	ナンバー灯の設置状況	87
図 5.2	ナンバー灯設置位置の説明図	87
図 5.3	実写画像1と実環境画像 (a)実写画像1, (b)実環境画像	88
図 5.4	一般的な数字領域の切り出し手順	91
図 5.5	テンプレート画像と縮小した画像の例 (a)テンプレート画像, (b)縮小した画像	92

図 5.6	対象画像 1 における正答率	93
図 5.7	対象画像 2 における正答率	93
図 5.8	対象画像 3 における正答率	94
図 5.9	対象画像 4 における正答率	94
図 5.10	数字認識の正答率	99
図 5.11	サブピクセル位置合わせの誤差の平均値	99
図 5.12	テスト画像の劣化強度と平均候補数との関係	102
図 5.13	環境光の外乱が加わった実環境画像 (a)実環境画像 1, (b)実環境画像 2	104
図 5.14	ヒストグラム拡張後の実環境画像 (a)実環境画像 1, (b)実環境画像 2	104
図 5.15	多様な劣化を受けた, ナンバープレート平仮名の実環境画像 (a)実環境画像 3, (b)実環境画像 4, (c)実環境画像 5, (d)実環境画像 6, (e)実環境画像 7, (f)実環境画像 8, (g)実環境画像 9, (h)実環境画像 10, (i)実環境画像 11, (j)実環境画像 12, (k)実環境画像 13, (l)実環境画像 14	105
図 5.16	距離と角度の定義 (a)平面図, (b)立面図	106
図 5.17	正答を得るのに必要な候補数	109
図 7.1	ナンバープレート領域抽出性能の向上法	120
図 7.2	ナンバープレートの傾き補正性能の向上法	121
図 7.3	ナンバープレート文字領域抽出性能の向上法	122
図 7.4	複合劣化画像に対するシステム全体の実行アルゴリズム	123
図 7.5	動きに特化した多重構造の畳み込みニューラルネットワークの模式図 (a)学習時, (b)テスト時	125

表目次

表 1.1	ナンバープレート画像の画質改善法と文字認識法の開発経緯と、 本研究の位置づけ	9
表 2.1	対象画像と学習画像の各劣化強度範囲	34
表 2.2	ネットワーク構造の諸元	36
表 2.3	正答率の比	38
表 3.1	実験用画像の劣化要因と劣化強度	58
表 4.1	テスト画像と学習画像の劣化要因と劣化強度	69
表 4.2	ネットワーク構造の諸元	75
表 4.3	テスト画像の例	77
表 4.4	学習画像の例	77
表 5.1	ナンバープレートの諸元	86
表 5.2	ナンバー灯の設置位置	87
表 5.3	輝度値の標準偏差	88
表 5.4	パラメータ k_3 で表される各劣化強度の範囲	89
表 5.5	グリッドサーチのパラメータ探索範囲	90
表 5.6	正答率の比	95
表 5.7	数字認識の正答率と、サブピクセル位置合わせの誤差の平均値	99
表 5.8	単一候補数当たりの正答率	102
表 5.9	各実環境画像の諸元	106
表 5.10	実環境画像 1 における実験結果	108
表 5.11	実環境画像 2 における実験結果	108

第1章 序論

1.1 監視カメラの歴史

世界初の監視カメラ装置は、1942年にナチスドイツにおいて、長距離弾道ミサイル V-2 の打ち上げを監視することを目的として、シーメンス社により開発されたものである⁽¹⁾。商用の監視カメラ装置としては1949年にアメリカで発売された Vericon と呼ばれるものが最初であり、このカメラは危険な産業工程を監視することなどに用いられた⁽²⁾。日本では1954年に三菱電機社が国産初の監視カメラ装置を開発し、このカメラは火力発電所の炉内の監視に利用された⁽³⁾。ただし、これらの初期の監視カメラ装置には録画機能が付いておらず、コントロールセンターなどに送られてモニターに映し出された映像を監視員などが常時監視しなければならなかった。

オープンリール式録画装置が開発されると、監視映像を録画することが可能となった。ただしこれらの装置では、オペレーターが磁気テープを手作業で交換する必要があった。この作業は非常に煩雑なものであり、装置の信頼性は低く、運用には多額の費用を要した。これらの欠点のため、オープンリール式録画装置を用いた監視カメラ装置が一般に普及することはなかった。

1970年代にビデオテープレコーダー技術が利用可能になると、簡単に情報を記録、消去できるようになり、監視カメラ装置が一般に普及し始めた⁽⁴⁾。

1990年代にデジタル多重化技術が開発されると、複数のカメラ映像を同時に録画することが可能になった。タイムラプスや動体検知録画もこの時代に出現した。これらの技術は、監視カメラ装置の維持に要する時間や費用を大幅に削減するものであったため、監視カメラ装置をより一層普及させることになった⁽⁵⁾。

近年には、インターネットベースの装置や顔認証などのバイオメトリクス技術を搭載した装置が出現している⁽⁶⁾。

1.2 監視カメラの設置台数の推移

監視カメラの設置台数は各国で増加している。イギリス全土の設置台数は、2014年時点で590万台（人口11人に1台）に達している。アメリカでも、2016年時点で3,000万台（人口10.8人に1台）以上が設置されているといわれている。アジアにおいても設置台数の増加が進んでいる。中国では設置台数が2017年12月時点で1億7,000万台（人口8.1人に1台）に達している。韓国では2015年12月時点で800万台（人口6.4人に1台）が設置されていると推測されている。日本では、2016年時点で500万台（人口25.4人に1台）以上が設置されているといわれている⁽⁷⁾。

日本における防犯設備機器の市場推定売上高の推移を図1.1に示す⁽³⁾。2008年ころ、売上高に一時的な減少が見られるが2012年ころから再び増加に転じており、全体として1990年代以降、市場規模が拡大傾向にあることを確認できる。上記の一時的な減少は、リーマンショックによる景気低迷などが原因として考えられ、市場規模の拡大傾向は今後も続く予測される。

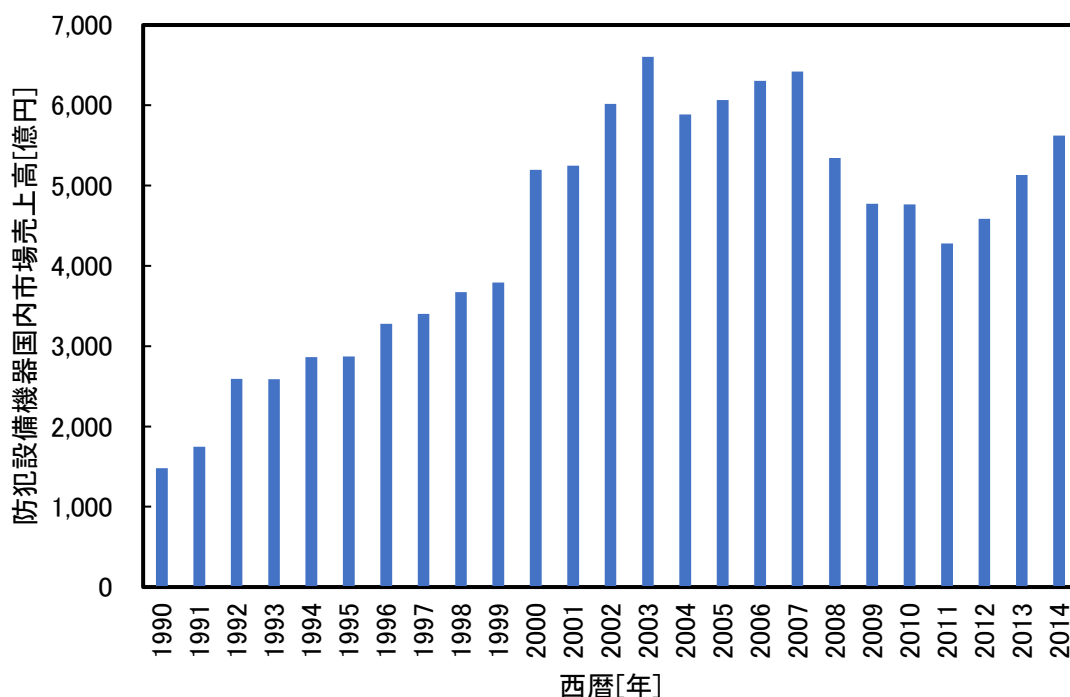


図 1.1 日本における防犯設備機器の推定市場売上高の推移

1.3 ナンバープレート文字読み取りシステムの歴史

1976年にイギリスの警察科学開発部が、ナンバープレート文字読み取りシステムを発明し、1979年にはその試作機が稼働した⁽⁸⁾。ナンバープレート文字読み取りシステムとは、監視カメラ装置の1つであり、犯罪捜査を目的としている。日本では、1981年に警察庁が自動車犯罪の激増に対処することを目的としてナンバープレート文字読み取りシステムの開発を始め、1986年にこのシステムが全国配備された。このシステムはナンバーの頭文字を取ってNシステムと呼ばれる。ナンバープレート文字読み取りシステムの一般的な処理手順を図1.2に示す。

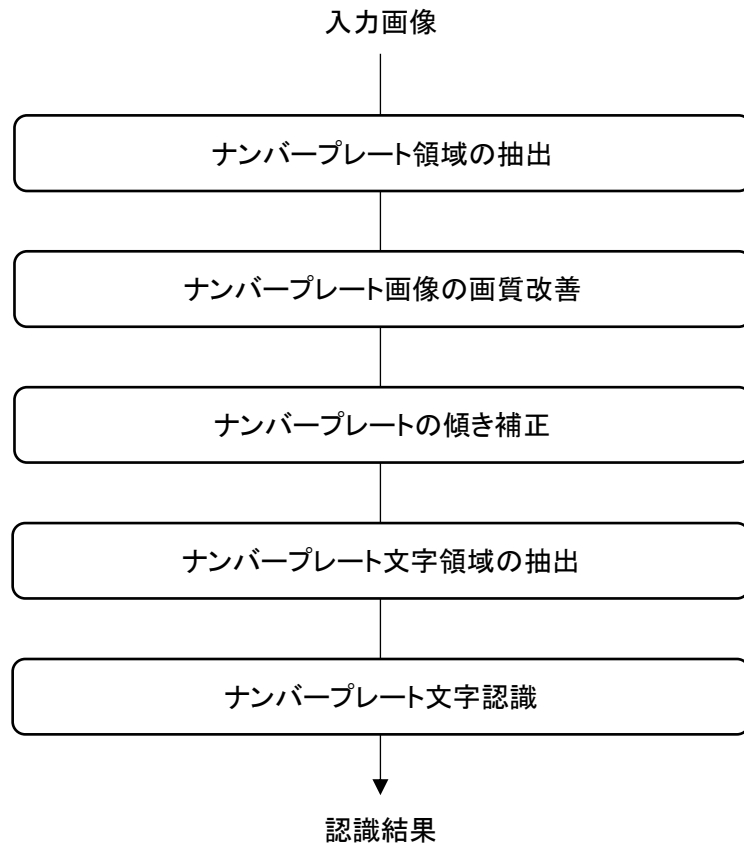


図 1.2 ナンバープレート文字読み取りシステムの一般的な処理手順

初期のナンバープレート文字読み取りシステムでは撮影角度や撮影距離などの撮影条件が固定されていた。また劣化が弱く人が容易に文字を視認できる画像のみが対象とされた。この頃は認識の正答率を高めることが課題であった。

1990年代も、固定された条件で撮影された劣化が弱いナンバープレート画像が対象とされ、認識の正答率を高めるための研究が継続された⁽⁹⁾。この時代にデジタル画像処理がナンバープレート文字読み取りシステムに導入された⁽¹⁰⁾。また明るさの変化に強い撮像方式を採用したカメラの開発など、ハードウェア技術も改良された⁽¹¹⁾。

2000年代以降にはハードウェア技術とソフトウェア技術の発展によりシステムにより多くのデジタル画像処理を適用することができるようになった。このため撮影角度や撮影距離などの撮影条件が変動する画像⁽¹²⁾や劣化が比較的強い画像⁽¹³⁾といった認識が難しい画像も研究対象とされるようになった。さらに街頭や店頭に設置された一般的な監視カメラもこれに録画された画像を回収して画像解析を行うことで文字情報を得ることができることから、ナンバープレート文字読み取りシステムの1つとして利用されるようになった⁽¹⁴⁾。この監視カメラのナンバープレート画像は偶然撮影されていたものであるため撮影条件が劣悪であることが多い。よってこれらの画像は、多くの要因により複合的に劣化しており、かつ各劣化強度も非常に強いため、多くの場合、人が容易にこの文字を視認できない。さらに、これらの画像における劣化の要因と強度は画像ごとに異なる。したがって劣化要因が多く、かつ劣化要因と強度が広範囲にわたる画像を研究対象とする必要が出てきた。これらの画像では画質が低いことが文字読み取りを難しくする根本原因であるため、図 1.2 の処理手順の中で画質改善が重要である。また劣化要因と強度が広範囲にわたる画像では、画像の状態に応じて適応的に文字認識を行う必要があるため、文字認識も重要である。このため、近年には、劣化要因が多く、かつ劣化要因と強度が広範囲にわたる画像を対象とした画質改善法と文字認識法に対する需要が高まっている。

1.4 ナンバープレート画像の画質改善法の歴史

ナンバープレート画像の画質改善は、図 1.2 に示したとおり、ナンバープレート文字読み取りシステムの間接処理の 1 つとして位置づけられる。1990 年代には明るさ調整などの一般的なデジタル画像処理を用いた画質改善法がナンバープレート文字読み取りシステムに導入された⁽¹⁰⁾。この時代には、劣化が弱く人が容易に文字を視認できる画像が対象とされていたため、ナンバープレート文字読み取りシステムに特化した画質改善法が研究されることはなかった。

2000 年代以降に劣化が強い画像もナンバープレート文字読み取りシステムの対象とされるようになると、ナンバープレート文字読み取りシステムに特化した画質改善法が研究され始めた。具体的には、主要な劣化要因の 1 つである解像度低下を解決するための研究が多く行われた⁽¹⁵⁾。近年には、街頭や店頭に設置された一般的な監視カメラもナンバープレート文字読み取りシステムの 1 つとして利用されるようになったことに対応するため、劣化要因が多く、かつ劣化要因と強度が広範囲にわたる画像を対象とした画質改善法に対する需要が高まっている。

1.5 ナンバープレート文字認識法の歴史

ナンバープレート文字認識は、図 1.2 に示したとおり、ナンバープレート文字読み取りシステムの最終処理として位置づけられる。1990 年代にはテンプレートマッチング法などの一般的なデジタル画像処理を用いた文字認識法がナンバープレート文字読み取りシステムに用いられていた⁽¹⁶⁾。この時代には、劣化が弱く人が容易に文字を視認できる画像が対象とされていたため、ナンバープレート文字読み取りシステムに特化した文字認識法が研究されることはなかった。1990 年代の中頃になると、明るさの変化に強いテンプレートマッチング法を用いたシステムが開発されるなど、一般的なデジタル画像処理の中で、よりシステムの応用に適合した手法が研究され、用いられるようになった⁽¹⁰⁾。

2000 年代になって劣化が強い画像が研究対象とされるようになると、劣化画像に特化した文字認識法に対する需要が高まった。この需要の高まりに伴い、

まず劣化要因と劣化強度が狭い範囲に限定された画像を対象とした研究が行われた。しかし街頭や店頭に設置された一般的な監視カメラ画像では、撮影条件が劣悪であることが多く、しかもその条件を限定できない場合が多い。このため、近年には、劣化要因が多く、かつ劣化要因と強度が広範囲にわたる画像に対応できる文字認識法に対する需要が高まっている。

1.6 ナンバープレート画像の画質改善法と文字認識法の開発経緯

本研究は、2000年代以降に街頭や店頭に設置された一般的な監視カメラもナンバープレート文字読み取りシステムの1つとして利用されるようになったことに対応して行ったものである。このナンバープレート文字読み取りシステムでは、システムの性能低下の主要因となる外乱光の影響の低減、システムの性能の向上に大きく寄与する解像度の改善、システムの劣化に対する適応性の向上に寄与する劣化強度推定が主要課題となる。また対応可能な劣化要因はシステムの達成度を測る指標となる。以下に監視カメラを用いたナンバープレート文字読み取りシステムにおいて重要な、画質改善法と文字認識法の開発経緯を述べるとともに、開発された各手法の、主要課題に対する達成状況と対応可能な劣化要因について説明する。

1976年にイギリスの警察科学開発部が世界で初めて光学的文字認識技術を用いてナンバープレート文字認識を行った⁽⁸⁾。1995年にはIBM社がデジタル画像処理をナンバープレート文字読み取りシステムに導入した⁽¹⁰⁾。具体的にはトーンカーブによる明るさ調整が画質改善法として用いられ、また明るさの変化に強い正規化相互相関が文字認識法に用いられた。これらの手法では劣化が弱い画像が対象とされていて本研究の主要課題に対する重要性が低かったため、いずれの主要課題も検討されなかった。1998年にはトヨタ社がカメラの撮像方式を改良し、システムの、主要課題の1つである外乱光に対する対応能力を高めた⁽¹¹⁾。これはハードウェア技術の面での進展であったといえる。2004年には甲南大学がファジィルールを用いた文字認識法を開発した⁽¹⁴⁾。この手法は

システムを初めて複合劣化に対応させた。2010年には大分県警が擬似劣化画像を用いた文字認識法を開発し、システムを解像度低下、ぼけ、画像圧縮の劣化に対応させた⁽¹⁷⁾。2012年には、Southeast Universityがナンバープレート文字読み取りシステムのためのマルチフレーム超解像処理による画質改善法を開発した⁽¹⁵⁾。この手法で初めてシステムが主要課題の1つである解像度改善に対応できるようになった。2013年にはNational Chiao Tung Universityが機械学習モデルの1つであるサポートベクターマシンを文字認識に導入した⁽¹³⁾。この手法は、文字認識法として初めて主要課題の1つである外乱光の影響の低減を実現した。この手法では、99パーセントという高い文字認識の正答率が達成されており正答率の面での進展もあった。また同じ年に兵庫県警が機械学習モデルの1つであるベイズ識別を用いた文字認識法を開発した⁽¹⁸⁾。この手法では、極めて低い解像度の画像が対象とされており、顕著な解像度低下の劣化に対する対応能力の面での進展があった。2015年にはNational University of Defense Technologyが、畳み込みニューラルネットワークを用いた文字認識法を開発した⁽¹⁹⁾。この手法は適応性が高いという畳み込みニューラルネットワークの特徴を活かし、初めてナンバープレート文字読み取りシステムを主要な劣化要因の全てに対応させた。2017年にはVellore Institute of Technologyがぼけ除去処理による画質改善法を開発した⁽²⁰⁾。この手法では、初めてナンバープレート文字読み取りシステムの画像のぼけ除去が行われた。2017年に著者らは、サポートベクターマシンを用いた画素値選択型マルチフレーム超解像処理による画質改善法を開発した⁽²¹⁾。この手法はシステムが主要課題の1つである解像度改善に対応できるようにしたものである。この手法により、初めて主要な劣化要因の全てに対応したナンバープレート文字読み取りシステムの画質改善法が実現した。さらにこの手法は画素値選択という新しい方法を提案しシステムの文字認識の正答率を大幅に向上させた。同じ年に著者らはサポートベクター回帰を用いたRetinex処理による画質改善法を開発した⁽²²⁾。この手法はナンバープレート文字読み取りシステムのソフトウェア技術として初めて主要課題の1つである外乱光の影響の低減を実現したものである。さらにこの手法はサポートベクター回帰を用いた新しい外乱光推定方法を提案しシステムの文字認識の正答率を大幅に向上さ

せた．この手法は主要な劣化要因の全てに対応することができ，劣化要因の面でも高い性能を達成した．さらに 2018 年に著者らは多重構造の畳み込みニューラルネットワークを用いた文字認識法を開発した⁽²³⁾．この手法は，ナンバープレート文字読み取りシステムとして初めて劣化強度推定に成功した．また多重解像度画像を処理することでシステムの文字認識の正答率を大幅に向上させた．この手法は主要な劣化要因の全てに対応することができ，劣化要因の面でも高い性能を達成した．

ナンバープレート画像の画質改善法と文字認識法の開発経緯と，本研究の位置づけを表 1.1 にまとめる．表 1.1 では劣化要因と主要課題の対応状況を併記した．著者らが提案した画質改善法と文字認識法は，近年有効性が明らかになっている機械学習を活用することにより，主要課題についての性能を高めるとともにシステムが主要な劣化要因の全てに対応できるようにしたことに特徴がある．著者らが提案した手法を用いて街頭や店頭に設置された一般的な監視カメラを，ナンバープレート文字読み取りシステムとして有効活用することで，これまでは強い複合劣化を受けているために利用が制限されていたカメラ映像がより広範に利用されるようになる可能性が高い．

表 1.1 ナンバープレート画像の画質改善法と文字認識法の開発経緯と、本研究の位置づけ

年次	画質改善法	文字認識法	劣化要因				主要課題			報告者
			解像度低下	ぼけ	ノイズ	画像圧縮	外乱光の影響低減	解像度改善	劣化強度推定	
1976		世界初の光学的文字認識技術によるナンバープレート文字認識	×	×	×	×	×	×	×	警察科学開発部 (イギリス)
1995	デジタル画像処理により明るさを調整	正規化相互相関で明るさの変化に対応	×	×	×	×	×	×	×	IBM, P. Comelli 他 ⁽¹⁰⁾
1998	外乱光の影響を低減できる撮像方式の開発		×	×	×	×	○	×	×	トヨタ, 内藤 貴志 他 ⁽¹¹⁾
2004		ファジィルールで初めて複合劣化に対応	○	×	○	×	×	×	×	甲南大学, 吉川 歩 他 ⁽¹⁴⁾
2010		疑似劣化画像を用いて解像度低下, ぼけ, 画像圧縮に対応	○	○	×	○	×	×	×	大分県警, 藤田 佳久 他 ⁽¹⁷⁾
2012	ぼけ, ノイズがある条件での解像度改善法の開発		○	○	○	×	×	○	×	Southeast University, Weili Zeng 他 ⁽¹⁵⁾
2013		サポートベクターマシンで外乱光の影響を低減	○	×	×	×	○	×	×	National Chiao Tung University, Bing-Fei Wu 他 ⁽¹³⁾
2013		ベイズ識別で顕著な解像度低下に対応	○	○	×	×	×	×	×	兵庫県警, 四宮康治 他 ⁽¹⁸⁾
2015		畳み込みニューラルネットワークで全劣化に対応	○	○	○	○	×	×	×	National University of Defense Technology, Liu Pan 他 ⁽¹⁹⁾
2017	ノイズがある条件でのぼけ除去処理の開発		×	○	○	×	×	×	×	Vellore Institute of Technology, P. S. Prashanth Rao 他 ⁽²⁰⁾
2017	全劣化がある条件での解像度改善法の提案		○	○	○	○	×	○	×	立命館大学, 辻 広生 他 ⁽²¹⁾
2017	全劣化がある条件での外乱光の影響の低減法を提案		○	○	○	○	○	×	×	立命館大学, 辻 広生 他 ⁽²²⁾
2018		多重構造の畳み込みニューラルネットワークで劣化強度推定に成功	○	○	○	○	×	×	○	立命館大学, 辻 広生 他 ⁽²³⁾

注 ○ : 対応可 × : 対応不可

1.7 本研究の目的および課題

本研究は街頭や店頭に設置された一般的な監視カメラを利用したナンバープレート文字読み取りシステムの性能を向上させることを目的とする。具体的にはこのカメラが撮影したナンバープレート複合劣化画像の文字を視覚的に見やすくし、正確に読み取れるようにすることを目指す。研究を進めるにあたり、近年有効性が明らかとなった機械学習を活用することに主眼を置き、以下の点を課題として検討を進めることとした。

(1) ロバスト性

ナンバー灯による環境光の外乱が加わる状況下でも、安定性が高い機械学習法であるサポートベクター回帰を活用して外乱を正確に除去することにより、視覚的に見やすい画像を復元して正確に文字を読み取ることができる。

(2) 効率性

ナンバープレートが小さく映っているため文字情報が少ない場合でも、信頼性が高い機械学習法であるサポートベクターマシンを活用して少ない文字情報を確実に利用することにより、視覚的に見やすい文字画像を効率的に再構成し、正確に文字を読み取ることができる。

(3) 適応性

形状が複雑で種類が多い文字を認識対象とし、また、多様な複合劣化に対応しなければならない場合でも、柔軟性の高い機械学習法である畳み込みニューラルネットワークを活用して文字の特徴や劣化の状態に柔軟に適応することにより、正確に文字を読み取ることができる。

1.8 本論文の構成

本論文の構成を図 1.3 に示す。

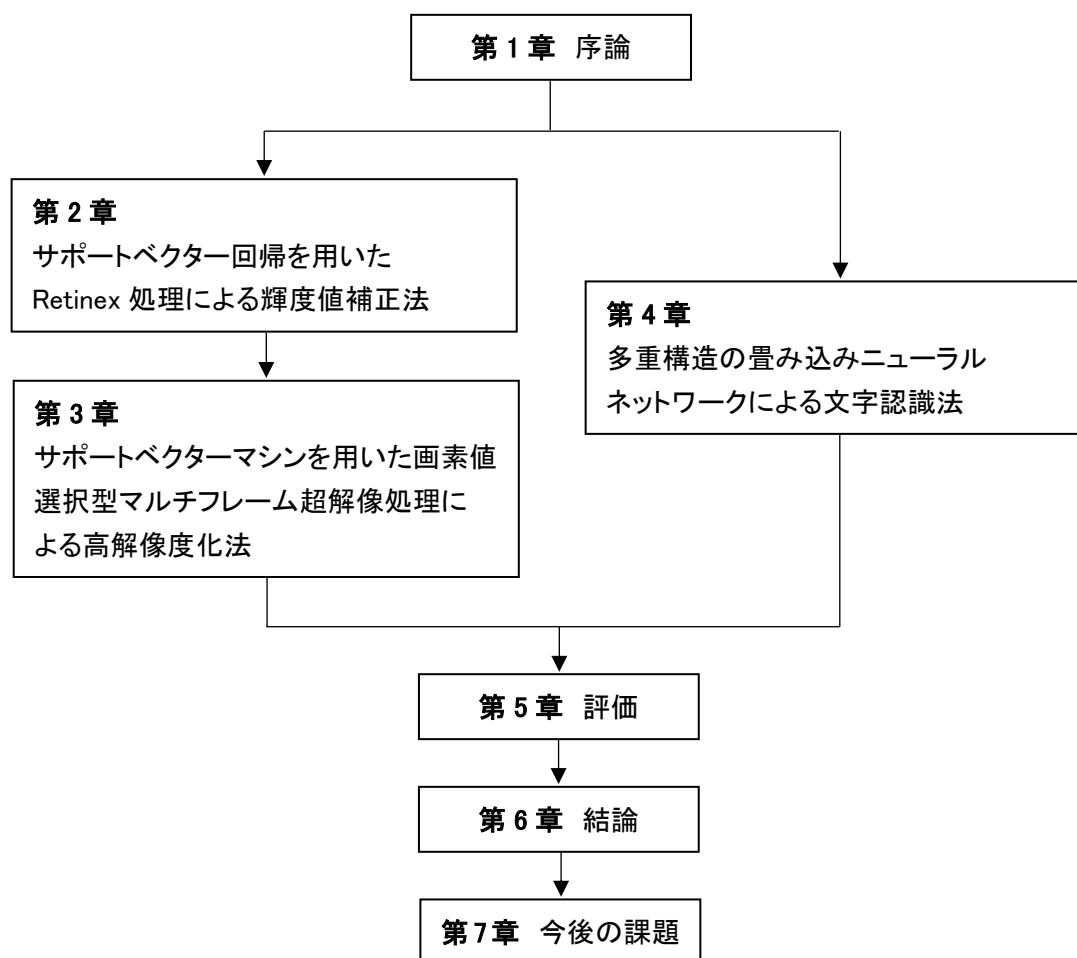


図 1.3 本論文の構成

第 2 章でナンバー灯による環境光の外乱に対するロバスト性を高めるための画質改善法について述べる。第 3 章では少ない文字情報の利用効率を高めるための画質改善法について述べる。第 4 章では複雑な文字形状や多様な複合劣化に柔軟に適應するための文字認識法について述べる。第 5 章では第 2 章と第 3 章で提案した画質改善法と第 4 章で提案した文字認識法の評価について述べる。

各章は以下のように構成されている。

第 2 章ではまず一般的な輝度値補正法について述べた後、環境光の外乱を除いて視覚的に見やすい画像を復元することに適している Retinex 処理について述べる。次に従来の Retinex 処理では環境光成分の推定精度が低いという問題があることを述べ、この問題を低減するためサポートベクター回帰で環境光成

分を推定する Retinex 処理を提案する。そして Retinex 処理の副作用である Halo 作用低減効果の評価と画質の定量的評価実験について述べる。

第 3 章では、まず画像情報が少ない場合に対する画質改善法である超解像処理について述べた後、超解像処理のうちマルチフレーム超解像処理が犯罪捜査の目的に適合することを述べる。次にこのマルチフレーム超解像処理には開口問題と呼ばれる問題があることを述べ、この問題を低減するためサポートベクターマシンを用いて開口部で適切な画素値を選択するマルチフレーム超解像処理を提案する。そして画質改善効果を明らかにするため、エッジ復元効果の評価と画素値選択効果の評価実験について述べる。

第 4 章では、まず文字特徴に柔軟に適応できる文字認識手法である畳み込みニューラルネットワークについて述べる。次にこの畳み込みニューラルネットワークには入力画像の解像度を調整して認識性能を高める機能がないという問題があることを述べ、この問題を低減するため、多重構造の畳み込みニューラルネットワークで多重解像度画像を処理する文字認識法を提案する。そして認識性能改善効果の評価実験について述べる。

第 5 章では第 2 章と第 3 章で提案した画質改善法と第 4 章で提案した文字認識法の各評価について述べた後、これらの評価結果を総括し、提案した全手法を統合したときの性能について述べる。

参考文献

- (1) Walter Dornberger: “V-2”, Ballantine, p. 14, 1954.
- (2) Edward L. Youmans: “Popular Science”, Bonnier, p. 179, 1949.
- (3) 賀来泉: “社会を変える防犯カメラ”, 幻冬舎, p. 275, 2017.
- (4) Herman Kruegle: “CCTV Surveillance: Video Practices and Technology”, Elsevier, p. 276, 2013.
- (5) CVRL Industries 社 HP: 「History of Video Surveillance and CCTV」
<http://www.wecusurveillance.com/cctvhistory>
- (6) Fennoturvapalvelut 社 HP: 「Internet Based CCTV on Cloud Services」
<http://fennoturvapalvelut.com/>
- (7) 英国放送協会 (BBC) HP: 「中国の監視網がたちまち人を特定 AI 付き監視カメラ全国に」
<http://www.bbc.com/japanese/video-42304882>
- (8) S. Rasheed, A. Naeem and Omer Ishaq: “Automated Number Plate Recognition Using Hough Lines and Template Matching”, Proceedings of The World Congress on Engineering and Computer Science, pp. 199-203, 2012.
- (9) P. Davies, N. Emmott and N. Ayland: “License Plate Recognition Technology for Toll Violation Enforcement”, IEE Colloquium on Image Analysis for Transport Applications, pp. 1-5, 1990.
- (10) P. Comelli, P. Ferragina and M.N. Granieri: “Optical Recognition of Motor Vehicle License Plates”, IEEE Transactions on Vehicular Technology, Vol. 44, No. 4, pp. 790-799, 1995.
- (11) 内藤貴志, 塚田 敏彦, 山田 啓一, 山本 新: “環境の明るさの変化にロバストな走行車両のナンバープレート認識”, 電子情報通信学会論文誌, Vol. J81-D- II, No. 9, pp. 2019-2026, 1998.

- (12) H. J. Lee, S. Y. Chen and S. Z. Wang: "Extraction and Recognition of License Plates of Motorcycles and Vehicles on Highways", Proceedings of International Conference on Pattern Recognition(ICPR), pp. 356-359, 2004.
- (13) B. F. Wu, H. Y. Huang and Y. L. Chen: "Degraded License Plate Recognition System for Town Buses on Highway", Proceedings of 10th IEEE International Conference on Networking, Sensing and Control (ICNSC), pp. 446-450, 2013.
- (14) 吉川歩, 吉川奈緒子, 藤田和弘: "記号識別装置, 記号識別方法, およびコンピュータプログラム", 特許第 4243941 号, 出願日 2002 年 10 月 17 日.
- (15) W. Zeng and X. Lu, "A Generalized DAMRF Image Modeling for Superresolution of License Plates" : IEEE Transactions on Intelligent Transportation Systems, Vol. 13, No. 2, pp. 828-837, 2012.
- (16) K. Miyamoto, K. Nagano, M. Tamagawa, I. Fujita and M. Yamato: "Vehicle License-plate Recognition by Image Analysis", IEEE International Conference on Industrial Electronics, Control and Instrumentation(IECON), pp. 1734-1738, 1991.
- (17) 藤田佳久, 佐藤和彦, 渡邊慎太郎, 岩津陽一郎: "低解像度ナンバー推定プログラムの試作と精度向上のための検討", 日本法科学技術学会誌, 第 15 巻別冊号, no.D-09, p. 83, 2010.
- (18) K. Shinomiya, N. Takamura, K. Fujita and N. Nakamori: "Discriminating Car License Plate Numbers on Low Resolution Using Moment Characteristics", IEEE Transactions on Media Technology and Applications, Vol. 1, No. 4, pp. 271-277, 2013.
- (19) P. Liu, G. H. Li and T. Dan: "Low-quality License Plate Character Recognition Based on CNN", Proceedings of 8th International Symposium on Computational Intelligence and Design(ISCID), Vol. 2, pp. 53-58, 2015.
- (20) P. S. P. Rao and R. K. Muthu: "A New De-blurring Technique for License Plate Images with Robust Length Estimation", IEEE Proceedings of International Conference on Intelligent Computing and Control (I2C2), pp. 1-6, 2017.

- (21) H. Tsuji, Y. Fukumizu, T. Douseki and H. Yamauchi: “Super-resolution on Pixel Values Selection for Degraded Image Recognition by Support Vector Machine”, Proceedings of 21st Triennial Meeting of the International Association of Forensic Sciences (IAFS), p. 248, 2017.
- (22) 辻広生, 福水洋平, 道関隆国, 山内寛紀: “Retinex モデルにおけるサポートベクター回帰を用いた劣化ナンバープレート画像の輝度値補正手法と文字認識への応用”, 画像電子学会誌, Vol. 46, No. 4, pp. 510-521, 2017.
- (23) 辻広生, 福水洋平, 道関隆国, 山内寛紀: “多重構造とした畳み込みニューラルネットワークによる劣化したナンバープレート画像の平仮名認識”, 信号処理学会誌, Vol. 22, No. 3, pp. 121-134, 2018.

第2章 サポートベクター回帰を用いた Retinex 処理による輝度値補正法

2.1 まえがき

街頭や店頭に設置された一般的な監視カメラ画像のナンバープレートは偶然映っていたものであり、明るさや画角などの撮影条件がナンバープレートに合わせられていない。このためナンバープレート画像は複合的な劣化を受ける。犯罪捜査ではこの複合劣化画像の文字を視覚的に見やすくし、正確に読み取れるようにすることが求められる。

特にナンバー灯による環境光の外乱はナンバープレート画像の画素値を大きく変化させてしまうため、安定した文字読み取りを難しくする。よって本研究では、環境光の外乱を正確に除去する手法を開発して外乱光に対するロバスト性を高めることを課題とした。

環境光の外乱を除いて視覚的に見やすい画像を復元することに適した従来の輝度値補正処理に Retinex 処理⁽¹⁾がある。Retinex 処理では、処理過程で環境光成分を推定しなければならないが、従来の Retinex 処理ではこの推定精度が低いという問題があり、上記の課題を克服することができない。そこで本研究では環境光成分の推定が回帰問題として一般化できることに着目した。回帰問題には機械学習が有効なことが近年明らかになっており、本論文では安定性が高い機械学習法であるサポートベクター回帰⁽²⁾を用いて環境光成分を推定する Retinex 処理⁽³⁾を提案する。

まず2節で一般的な輝度値補正法について述べ、3節で従来の Retinex 処理について述べる。次に4節でサポートベクター回帰を用いた Retinex 処理を提案する。5節で Halo 作用低減効果について述べ、6節で画質の定量的評価実験について述べる。

2.2 一般的な輝度値補正法

2.2.1 トーンカーブによる輝度値補正法

画像を構成する各画素の輝度値を変化させるためには，入力画像の各輝度値に対し出力画像の各輝度値をどのように対応づけるかを指定すればよい．この対応関係を与える関数を階調変換関数，それをグラフで表したものをトーンカーブと呼ぶ⁽⁴⁾

最も代表的なトーンカーブは折れ線型のものである．図 2.1 に折れ線型トーンカーブによる輝度値補正例，図 2.2 に補正に用いたトーンカーブを示す．



図 2.1 折れ線型トーンカーブによる輝度値補正例

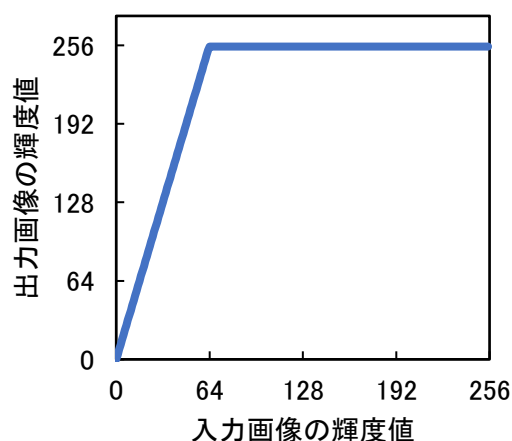


図 2.2 補正に用いたトーンカーブ

折れ線型トーンカーブが水平な部分では出力画像の輝度値が一定になるため，入力画像のその範囲の明暗変化は除去されそれ以外の範囲の明暗変化が強調される．図 2.1 の例では入力画像上部が明るくなっていることから，上部でナンバー灯による環境光の外乱が大きい．画質改善の対象となる数字部分は入力画

像下部の暗い部分にあるため、図 2.2 では入力画像の明るい部分（輝度値が高い部分）を水平にして入力画像の暗い部分（輝度値が低い部分）の明暗変化が強調されるようにしてある。この補正により出力画像の文字の背景部分は上部も下部も全て明るくなっており、上部にあった外乱光成分は小さくなっている。しかし外乱光成分は上部にだけあるわけではなく少ないながら下部の文字部分にも存在し、この文字部分の外乱光成分についてはトーンカーブによる輝度値補正では取り除くことができない。出力画像の文字部分を見ると上方から下方へ徐々に暗くなっており、文字部分に外乱光成分が残っていることがわかる。

2.2.2 ヒストグラム平坦化による輝度値補正法

輝度値のヒストグラム分布は画像によりまちまちで、一般的に偏りがある。ヒストグラム平坦化は出力画像のヒストグラムが輝度値の全域にわたって均等に分布するように変換するものであり、一般的にはコントラストを改善することができる⁽⁴⁾

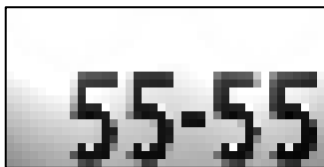
画像サイズを $M \times N$ 画素、出力画像の明暗レベル数を L とすると、出力画像のヒストグラムが完全に平坦なれば各輝度値の頻度は MN/L となる。そこで例えば輝度値の小さい方から頻度を積算し、その数が MN/L に達するまでの輝度値をひとまとめにして、出力画像の輝度値に割り当てるという処理を行えばよい。

しかし実際には、入力画像のある輝度値に対する積算を行ったときにちょうど MN/L に等しくなるということはまれである。その場合、近くの輝度値にまとめて割り当ててしまう方法が一般的である。近くの輝度値にまとめて割り当てると出力画像のヒストグラムは完全には平坦にならずでこぼこが生じたり頻度が高い輝度値の部分では間が空いてまばらに分布したりするが、それでも平均的にみればヒストグラムが均等化される。

図 2.3 にヒストグラム平坦化例、図 2.4 に入出力画像の輝度値ヒストグラムを示す。

図 2.3 よりヒストグラム平坦化はナンバー灯による環境光の外乱を取り除く効果がないことがわかる。図 2.4 の入力画像の輝度値ヒストグラムをみると、

環境光の影響で輝度値が高い画素の頻度が高い。よってヒストグラムを平坦化すると入力画像の輝度値が高い画素に対して出力画像では広い範囲の輝度値が割り当てられることになる。その結果、出力画像では入力画像で明るい部分のコントラストが高くなり、暗い部分のコントラストが低くなる。しかし入力画像で明るい部分は環境光の外乱が大きい部分であり暗い部分に画質改善の対象となる数字部分が含まれることから、外乱が強調され数字部分が見えにくくなっている。すなわちヒストグラム平坦化は外乱に対するロバスト性を高めるといふ課題に対しては有効でない。

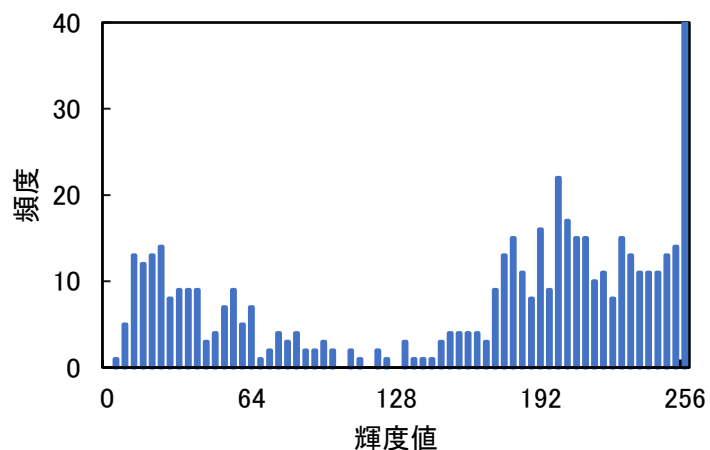


(a) 入力画像

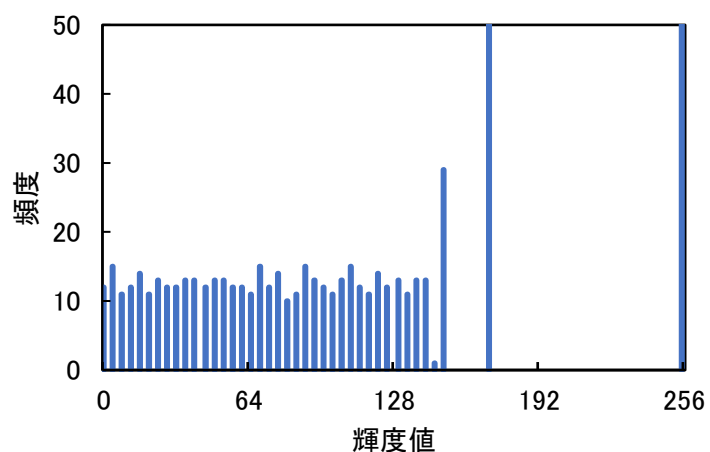


(b) 出力画像

図 2.3 ヒストグラム平坦化例



(a) 入力画像



(b) 出力画像

図 2.4 入出力画像の輝度値ヒストグラム

2.3 従来の Retinex 処理

2.3.1 Retinexモデル

人が知覚する明るさは入射光の強度ではなく物体固有の反射率に基づくことが知られている。たとえば白い衣服と黒い衣服を外光の直射日光の下で観察しても室内光の下で観察しても、白い衣服は白く黒い衣服は黒く見える。物理的には網膜に到達している光強度は直射日光下の黒い衣服の方が室内光下の白い

衣服よりも圧倒的に高いが、白と黒の関係が逆転することはない。すなわち人が知覚する明るさは入射光から環境光成分を除去した物体固有の反射率に基づく⁽⁵⁾。

Retinex モデルは上述した人が知覚する明るさ、観測画像、物体固有の反射率、環境光成分の間に成り立つ関係をモデル化したものである。Retinex モデルの模式図を図 2.5 に示す。

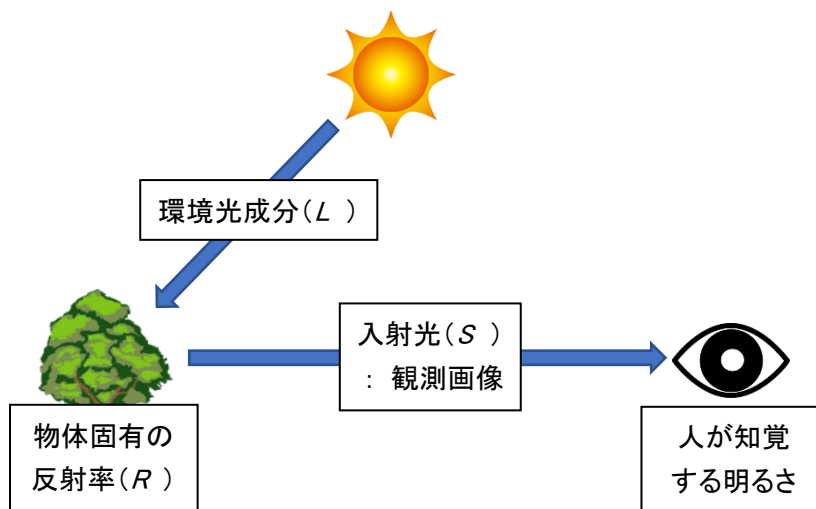


図 2.5 Retinex モデルの模式図

Retinex モデルでは観測画像 S は式(2.1)で定式化される⁽¹⁾。

$$S(x, y) = L(x, y) \cdot R(x, y) \quad (2.1)$$

式(2.1)において L は環境光成分、 R は環境光成分に依存しない物体固有の反射率、 x は画像横方向の座標、 y は画像縦方向の座標である。Retinex モデルでは人の知覚は式(2.1)の R に基づいていると考える。本研究が対象とするナンバープレート画像の場合、 L は主としてナンバー灯による照明光成分となる。

2.3.2 従来のRetinex処理

人が知覚する明るさは入射光から環境光成分を除去した物体固有の反射率に基づくのに対し、観測画像の輝度値は基本的に入射光の物理的な光強度に基づく。よって人の知覚と観測画像との間には違いが生じ、人が見やすい画像を復

元するためには画像の階調特性を人が知覚する明るさに近づける必要が生じる。Retinex モデルは画像の階調特性を人が知覚する明るさに近づくように変換する目的に用いることができる。この変換は、観測画像 S から環境光成分 L を推定した後、式(2.1)により物体固有の反射率 R を導出することで実現される。この Retinex モデルに基づいて R を導出する処理を Retinex 処理と呼ぶ。

一般的に L の推定には center/surround (以下, C/S) アルゴリズム⁽⁶⁾が用いられる。本論文では C/S アルゴリズムによる L の推定結果の画像を C/S 画像と呼ぶこととする。本研究の目的の 1 つは文字を正確に読み取れるようにすることであるが、この目的にも C/S アルゴリズムによる Retinex 処理が用いられる。例としてデジタルカメラで撮影された複合劣化画像の文字を対象とした研究⁽⁷⁾がある。

C/S アルゴリズムでは環境光は空間的になめらかであるという仮定をおき、 S にガウス関数を畳み込むことによって L を推定する。この推定において単一の周辺領域で定義されるガウス関数のみを用いる方法は single scale retinex (以下, SSR) 処理と呼ばれる。

SSR 処理の手順を図 2.6 に示す。SSR 処理では対数強調後の観測画像と C/S 画像との差分をとり、この差分画像にヒストグラム拡張を行ったものを補正後の画像とする。

なおヒストグラム拡張とは 2.2.1 節で述べた折れ線型トーンカーブによる輝度値補正を行って画質改善対象領域のコントラストを改善することを意味する。

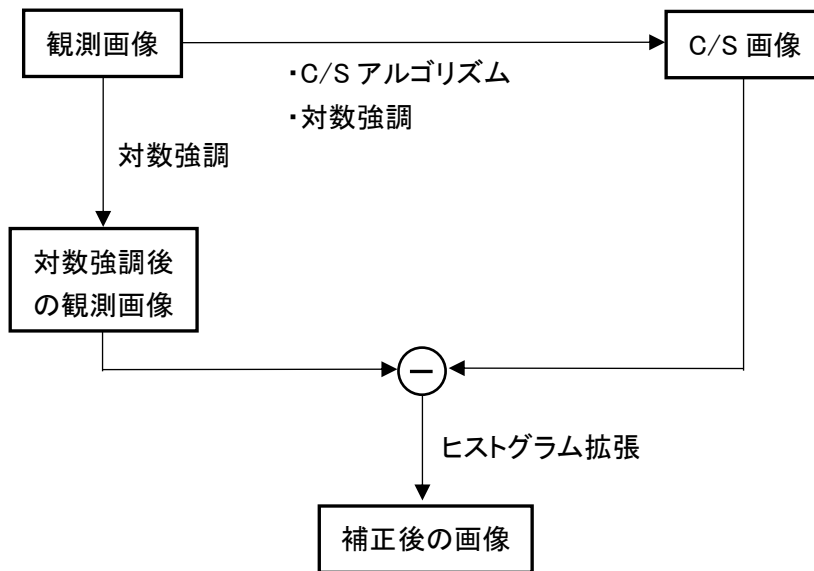


図 2.6 SSR 処理の手順

SSR 処理について具体的に説明すると，上記の差分画像（以下，SSR 出力）は式(2.2)で与えられる．

$$R_{SSR}(x, y) = \log S(x, y) - \log\{F(x, y) * S(x, y)\} \quad (2.2)$$

式(2.2)において $F(x, y)$ はガウス関数であり

$$F(x, y) = K \cdot e^{-(x^2+y^2)/2\sigma^2} \quad (2.3)$$

と表され，また $*$ は畳み込み積分を表す．式(2.3)の K は式(2.4)を満たすように定める．

$$\iint F(x, y) dx dy = 1 \quad (2.4)$$

SSR 出力では観測画像を単に推定照明光成分で割るだけでなく対数強調を行っている．対数強調を行うことで特に黒潰れ領域におけるコントラストが改善される．SSR 処理は輝度値の勾配が緩やかな領域では有効であるが，勾配の大きい領域の境界付近でアーティファクト（画像処理の副作用として出現する偽構造）の一種である帯状妨害（以下，Halo）が発生する．

Halo を改善するため周辺領域が異なるガウス関数による SSR 出力を合成する手法である multi-scale retinex（以下，MSR）処理が提案されている⁽⁸⁾．MSR 処理でもヒストグラム拡張を行って補正後の画像を得るが，このヒストグラム拡張を行う前の画像（以下，MSR 出力）は式(2.5)で与えられる．

$$R_{\text{MSR}}(x, y) = \sum_{n=1}^N w_n \cdot R_n(x, y) \quad (2.5)$$

式(2.5)において $R_n(x, y)$ は周辺領域が σ_n のガウス関数を用いた SSR 出力であり、 N は SSR 出力の総数を表す。また w_n は各 SSR 出力に対する重みであり、

$$0 < w_n < 1 \quad (2.6)$$

$$\sum_{n=1}^N w_n = 1 \quad (2.7)$$

となるように設定する。

2.4 サポートベクター回帰を用いた Retinex 処理

従来の Retinex 処理（以下、従来 Retinex）には以下の 2 つの問題がある。

第 1 の問題は、従来 Retinex では任意の点の L を推定する際にその近傍の点における L の影響しか考慮できないことである。従来 Retinex では S にガウス関数を畳み込んで L を推定するので任意の点の L はその近傍の点の L から強い影響を受け、距離が離れるに従いその影響は幾何級数的に低下する。しかし実際にはナンバー灯の光に照らされたナンバープレートの場合、いずれの点の L も主としてナンバー灯の光に起因しており、ナンバー灯から離れた点の L もナンバー灯近傍の L から強い影響を受けている。

第 2 の問題は、従来 Retinex では L の方向性を考慮できないことである。ガウス関数は等方的であるので従来 Retinex では各点の L がその周りに及ぼす影響も等方的である。しかし実際には上記ナンバープレートの場合、 L はナンバー灯近傍の点を中心として放射状に拡がるため、ナンバープレート内では L は方向性をもつ。

上記の 2 つの問題を低減するため、提案の Retinex 処理（以下、提案 Retinex）では L を推定する問題を回帰問題として一般化する。すなわち提案 Retinex ではナンバープレート内の座標と L との間に成り立つ関数を導き出すことで L を推定する。回帰問題には機械学習が有効であり、街頭や店頭に設置された一般的な監視カメラ画像のように多様な劣化が存在し得る不安定な環境では安定性

が高い機械学習法を用いる必要がある。サポートベクター回帰（以下、SVR）は大域的最適解を数学的に導き出せ⁽²⁾、かつ多重共線性問題が発生しないため⁽⁹⁾安定性が高く、かつ現在知られている機械学習法の中で性能が優れている。よって本研究ではサポートベクター回帰を用いて L を推定した。

2.4.1 全体的な処理手順

提案 Retinex の全体的な処理手順を図 2.7 に示す。図 2.7 で比較のため従来 Retinex の経路を点線で示した。

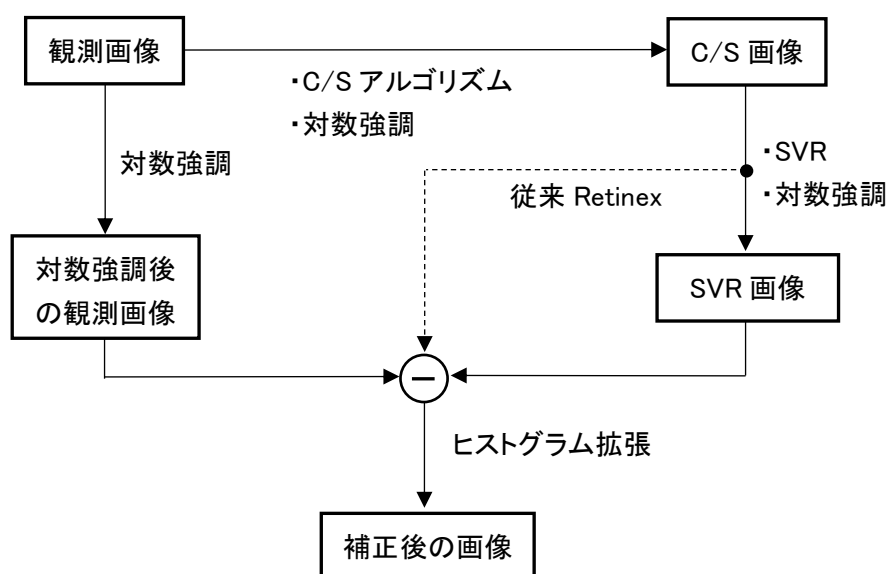


図 2.7 提案 Retinex の全体的な処理手順

提案 Retinex では最初に従来 Retinex と同じ方法で C/S 画像を生成する。次に C/S 画像において SVR を行い C/S 画像よりも高い精度で L を推定した画像（以下、SVR 画像）を生成する。そして対数強調後の観測画像と SVR 画像との差分をとり、この差分画像にヒストグラム拡張を行ったものを補正後の画像とする。

2.4.2 サポートベクター回帰を用いた Retinex 処理の利点

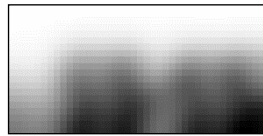
サポートベクター回帰を用いた Retinex 処理の利点の説明図を図 2.8 に示す。

図 2.8 の(b1)と(b2)はそれぞれ(a)の観測画像に対する C/S 画像と SVR 画像であり (c1)と(c2)はそれぞれ(b1)と(b2)の 3 次元表示図である. (c1)と(c2)における座標軸の説明を下記に示す.

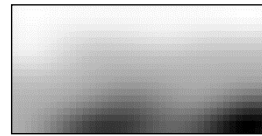
- x 軸：画像の横方向[画素]
- y 軸：画像の縦方向[画素]
- z 軸：輝度値



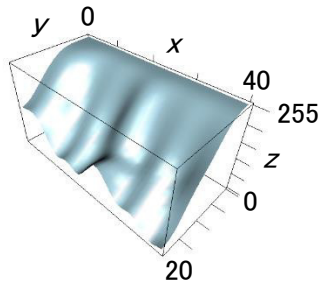
(a) 観測画像



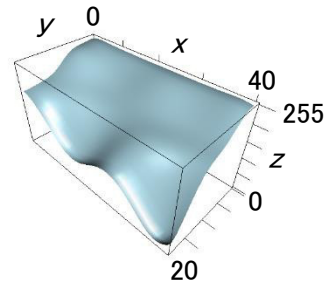
(b1) C/S 画像



(b2) SVR 画像



(c1) (b1)の 3 次元表示図



(c2) (b2)の 3 次元表示図

図 2.8 サポートベクター回帰を用いた Retinex 処理の利点の説明図

図 2.8 では環境光成分 L はナンバープレート上方のナンバー灯に起因し, 本来 L は光源の光学的性質から上方から下方へ向かって大きさが滑らかに小さくなっていく.

C/S 画像は観測画像から生成されるが, 観測画像は物体 (ナンバープレート)

の局所的な反射率の影響を強く受けている。また C/S 画像はこの観測画像に等方的なガウス関数を畳み込むことで生成されるため、C/S 画像では観測画像の近傍点同士のみが等方的に影響し合うことになる。よって C/S 画像では物体の局所的な反射率の影響のみが強まり、また本来の L には存在する方向性も反映されない。その結果、物体の反射率が低い領域の L が、反射率が高い領域の L よりも低い値に推定される。また反射率が異なる領域の境界で推定結果の値に高低差が生じ、本来の L がもつ大きさが滑らかに変化するという性質が反映されにくくなる。図 2.8(a)のナンバープレートの場合 4 桁の数字部分の反射率が低い。この場合の C/S 画像では数字部分が暗くなることを図 2.8 の(b1)から確認でき、また数字部分と背景部分の境界で輝度値の高低差が大きくなることを図 2.8 の(c1)から確認できる。

一方、提案 Retinex では C/S 画像において SVR を行って反射率が異なる領域を含む物体全体（ナンバープレート全体）の輝度値に基づいて L を推定する。よって SVR 画像では物体の局所的な反射率の違いによる影響が低減され、また L は 2 次元座標に対する関数として導き出されるため L の方向性も反映される。ナンバープレートの場合、SVR 画像では数字部分が暗くならないことを図 2.8 の(b2)から確認でき、また数字部分と背景部分の境界における輝度値の高低差が小さいことを図 2.8 の(c2)から確認できる。

以上より SVR を用いた Retinex 処理には、物体の局所的な反射率の違いが環境光成分の推定精度に及ぼす影響が低減され、また環境光成分がもつ大きさが滑らかに変化するという性質が推定結果に反映されやすくなる。すなわち提案 Retinex では従来 Retinex に比べて環境光成分の推定精度が向上するという利点がある。

2.4.3 サポートベクター回帰の実装方法

SVR による環境光成分の推定は、C/S 画像における i 番目の画素が有する座標で構成する 2 次元入力ベクトルを \mathbf{x}_i 、 i 番目の画素が有する輝度値を z_i 、 M を画素数とすると、

$$(\mathbf{x}_i, z_i) \quad (i = 1, 2, \dots, M) \quad (2.8)$$

で表されるデータ対を入出力教師データとして行う。

環境光成分の推定は、入出力教師データを用いて式(2.9)の関数を決定することに相当する。

$$f(\mathbf{x}) = \mathbf{w}^T \phi(\mathbf{x}) + \beta \quad (2.9)$$

ただし、 $\phi(\mathbf{x})$ は \mathbf{x} を特徴空間に写像する写像関数、 \mathbf{w} は係数ベクトル、 β はバイアス項である。

SVR を定式化するため⁽²⁾、出力 z と z の推定値との残差を、

$$r = z - f(\mathbf{x}) \quad (2.10)$$

で定義し、誤差関数に式(2.11)のロバスト関数近似を用いる。

$$E(r) = \begin{cases} 0 & (|r| \leq \varepsilon) \\ |r| - \varepsilon & (|r| > \varepsilon) \end{cases} \quad (2.11)$$

ただし、 ε は正の数である。

$$r = D(\mathbf{x}, z) \quad (2.12)$$

で定義し、 $f(\mathbf{x}) \pm \varepsilon$ により表される ε チューブの外にも教師データが存在することを許容するため、 \mathbf{x}_i にスラック変数 ξ_i, ξ_i^* を式(2.13)と式(2.14)のように導入する。

$$\xi_i(r) = \begin{cases} 0 & (D(\mathbf{x}_i, z_i) - \varepsilon \leq 0) \\ D(\mathbf{x}_i, z_i) - \varepsilon & (D(\mathbf{x}_i, z_i) - \varepsilon > 0) \end{cases} \quad (2.13)$$

$$\xi_i(r)^* = \begin{cases} 0 & (\varepsilon + D(\mathbf{x}_i, z_i) \geq 0) \\ -\varepsilon - D(\mathbf{x}_i, z_i) & (\varepsilon + D(\mathbf{x}_i, z_i) < 0) \end{cases} \quad (2.14)$$

以上より、環境光成分についての SVR は式(2.15)として定式化される。

$$\min Q(\mathbf{w}, \beta, \xi, \xi^*) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^M (\xi_i + \xi_i^*) \quad (2.15)$$

subject to

$$z_i - \mathbf{w}^T \phi(\mathbf{x}_i) - \beta = \varepsilon + \xi_i \quad (i = 1, 2, \dots, M) \quad (2.16)$$

$$\mathbf{w}^T \phi(\mathbf{x}_i) + \beta - z_i = \varepsilon + \xi_i^* \quad (i = 1, 2, \dots, M) \quad (2.17)$$

$$\xi_i \geq 0, \xi_i^* \geq 0 \quad (i = 1, 2, \dots, M) \quad (2.18)$$

ただし、 C は超平面からの距離と教師データに対する近似誤差との間におけるトレードオフを調整するマージンパラメータである。

ラグランジュ乗数を導入すると、式(2.15)の最適化問題を解くことができる。

この最適化問題を解いた結果として得られる SVR 画像 $f(\mathbf{x})$ は、ラグランジュ乗数 α_i, α_i^* により式(2.19)で表現される。

$$f(\mathbf{x}) = \sum_{i=1}^M (\alpha_i - \alpha_i^*) \phi^T(\mathbf{x}_i) \phi(\mathbf{x}) \quad (2.19)$$

なお式(2.19)において、

$$K(\mathbf{x}, \mathbf{x}_i) = \phi^T(\mathbf{x}_i) \phi(\mathbf{x}) \quad (2.20)$$

と定義することで、カーネルトリックを適用できる。

2.5 Halo 作用低減効果の評価

Halo は原画像にはない画像構造であるため、犯罪捜査の応用では証拠価値を低下させてしまう一因となり、この評価は犯罪捜査の応用上重要である。Halo 作用は見た目に明らかであるためその強度を人が評価する場合でも主観が評価結果に及ぼす影響は少ないと考えられる。よって Halo 作用については定性的に評価することとした。

2.5.1 評価の条件

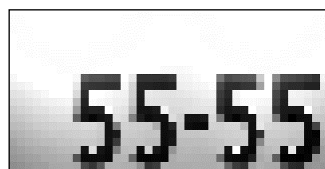
Halo 作用は輝度値の勾配が大きい領域で帯状に出現するアーティファクトであり、ぼけ、ノイズ、JPEG 圧縮の劣化を受けるとこれを視認しにくくなる。よって解像度低下の劣化のみを付与した画像を対象画像とした。本実験ではナンバー灯による環境光の外乱については実環境で与え、解像度低下については計算機上で与えた。具体的にはまず劣化ができるだけ少なくなるよう、暗室でデジタルカメラ (Nikon 製 D200) により近距離 (1.7 メートル) で真正面からナンバー灯 (小糸製作所製自動車用電球 12V 5W) に照らされた新品のナンバープレートを高解像度 (横 3,872[画素], 縦 2,592[画素]) で撮影し無圧縮で保存した。以降においてこの無圧縮で保存した画像を実写画像 1 と呼ぶこととする。実写画像 1 のナンバープレート領域の解像度は横 483[画素]で縦 227[画素]であった。そしてこの実写画像 1 を計算機で縮小し、これを評価対象画像とした。

従来 Retinex のパラメータと提案 Retinex のパラメータについては、いずれも

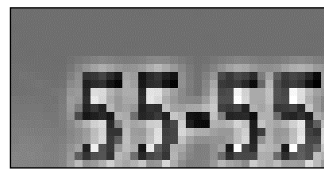
正答の数字に対する正規化相互相関の値が最大となる値をグリッドサーチで求めることにより設定した。

2.5.2 Halo作用低減効果

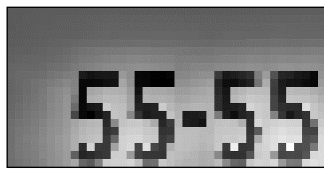
輝度値補正前の対象画像（以下，補正前の画像），従来 Retinex（SSR 処理，2スケールの MSR（以下，2MSR）処理）を行った画像，提案 Retinex を行った画像を図 2.9 に示す．ここに 2 スケールとは，式(2.5)において $N=2$ であることを表す．



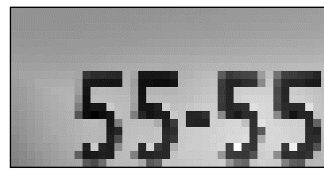
(a) 補正前の画像



(b) SSR 処理後の画像



(c) 2MSR 処理後の画像



(d) 提案 Retinex を行った後の画像

図 2.9 Retinex 処理画像

2.5.3 考察

図 2.9 より SSR 処理後の画像においては，顕著な Halo 作用が見られ，2MSR 処理後の画像においては，SSR 処理後の画像より，Halo 作用が低減している．提案手法を行った後の画像においては，2MSR 処理後の画像よりもさらに Halo 作用が低減している．

以上より，提案 Retinex は従来 Retinex よりも Halo 作用低減効果の観点で性能が高いといえる．

2.6 画質の定量的評価実験

Retinex 処理は観測画像から環境光成分を除いて人の視覚による見え方に近づけた画像を復元する処理である。したがって Retinex 処理の評価方法の 1 つとして復元画像の見やすさを人に評価してもらう方法が考えられるがこの評価結果は人の主観の影響を受けてしまう。大勢の人に評価してもらえば主観の影響を排除できる可能性があるが、どの程度の人数の人に評価してもらえばよいのか不明である。したがってこの方法で正確に性能を評価することは難しい。そこで本研究では人の視覚システムを模擬した構造をもつ畳み込みニューラルネットワーク（以下、CNN）⁽¹⁰⁾を用いて性能を評価することとした。具体的には CNN にナンバープレート数字の形状を学習させ、学習済み CNN に従来 Retinex で復元した数字画像と提案 Retinex で復元した数字画像を認識させ、認識の正答率に基づいて各性能を評価した。

2.6.1 実験条件

ナンバー灯による環境光の外乱と、解像度低下、ぼけ、ノイズ、JPEG 圧縮による複合劣化画像を各 Retinex 処理の対象画像とした。本実験ではナンバー灯による環境光の外乱については実環境で与え、その他の劣化については計算機上で与えた。具体的には 2.5.1 節で述べた実写画像 1 に、画像劣化モデル⁽¹¹⁾に基づき計算機で各劣化を付与した画像を生成しこれを対象画像とした。本実験の画像劣化モデルを表す数式を式(2.21)に示す。

$$g(x, y) = J(s\downarrow(h(u, v)*f(x, y)) + \eta) \quad (2.21)$$

f : 原画像

g : 劣化画像

h : 点拡がり関数

$s\downarrow$: 解像度低下

η : ノイズ

J : JPEG 符号化, 復号化

x, u : 画像横方向の座標

y, v : 画像縦方向の座標

ただし, 式(2.21)の*は畳み込み積分を表す. また点拡がり関数 h とノイズ η の確率密度関数には, それぞれ式(2.22)と式(2.23)の関数を用いた. 式(2.22)はぼけを近似するガウス関数であり, 式(2.23)は, ガウスノイズの確率密度関数である. 式(2.22)の σ_b と式(2.23)の σ_n は, いずれも標準偏差を表す.

$$h(u, v) = \frac{1}{2\pi\sigma_b^2} \exp\left(-\frac{x^2 + y^2}{2\sigma_b^2}\right) \quad (2.22)$$

$$\rho(\eta) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma_n^2}\eta^2\right) \quad (2.23)$$

Retinex 処理の対象画像における, ガウスフィルタによるぼけの標準偏差 σ_{b1} , ガウスノイズの標準偏差 σ_{n1} , JPEG 圧縮の品質 Q_1 についてはそれぞれ式(2.24), から式(2.26)で表される値を用いた.

$$\sigma_{b1} = 0.95 + 0.05k_1 \quad (2.24)$$

$$\sigma_{n1} = 0.13 + 0.01k_1 \quad (2.25)$$

$$Q_1 = 84 - 2k_1 \quad (2.26)$$

ここに劣化強度は k_1 で定義される. k_1 は 1 から 5 までの 5 段階あり値が大きいほど劣化が強い.

学習画像は解像度低下, ぼけ, ノイズ, JPEG 圧縮による複合劣化画像とした. 具体的には図 2.10 の 0 から 9 までの数字のテンプレート画像に式(2.21)に基づき計算機で各劣化を付与し, 学習画像を生成した.

0 1 2 3 4 5 6 7 8 9

図 2.10 テンプレート画像

パラメータ k_1 で表される対象画像の各劣化強度範囲と学習画像の各劣化強度範囲を表 2.1 に示す.

表 2.1 対象画像と学習画像の各劣化強度範囲

	対象画像	学習画像
解像度低下[画素]	横 6×縦 11	横 6×縦 11
ガウシアンフィルタによるぼけ (対象画像の標準偏差 σ_{b1} [画素]) (学習画像の標準偏差 σ_{b2} [画素])	$1.00 \leq \sigma_{b1} \leq 1.20$	$0.90 \leq \sigma_{b2} \leq 1.30$
ガウスノイズ (対象画像の標準偏差 σ_{n1}) (学習画像の標準偏差 σ_{n2})	$0.14 \leq \sigma_{n1} \leq 0.18$	$0.05 \leq \sigma_{n2} \leq 0.27$
JPEG 圧縮 (対象画像の品質 Q_1) (学習画像の品質 Q_2)	$74 \leq Q_1 \leq 82$	$64 \leq Q_2 \leq 92$

なおガウスノイズについては、不規則性があるため、同じ劣化強度でもシードごとに認識精度が異なる。よって対象画像と学習画像のいずれにおいても劣化強度ごとにシードを変えて 10 枚生成し、不規則性が認識精度に与える影響を軽減した。

対象画像から切り出した数字画像と学習画像の数字画像の総数は、それぞれ 1,950 枚, 49,500 枚とした。

学習画像の例を図 2.11 に示す。

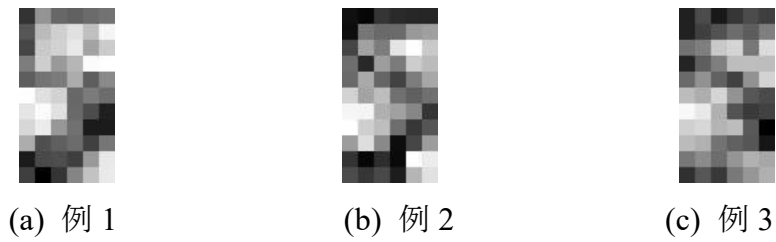


図 2.11 学習画像の例

従来 Retinex のパラメータと提案 Retinex のパラメータについては、いずれも数字認識の正答率が最大となる値をグリッドサーチで求めることにより設定した。

2.6.2 実験に用いた畳み込みニューラルネットワークの構成

実験に用いた CNN のネットワーク構造を図 2.12, ネットワーク構造の諸元を表 2.2 に示す。この CNN では出力計算と勾配計算を高速化するため活性化関数に ReLU⁽¹²⁾を用いた。また汎化能力を高めるためにドロップアウト⁽¹³⁾を導入した。

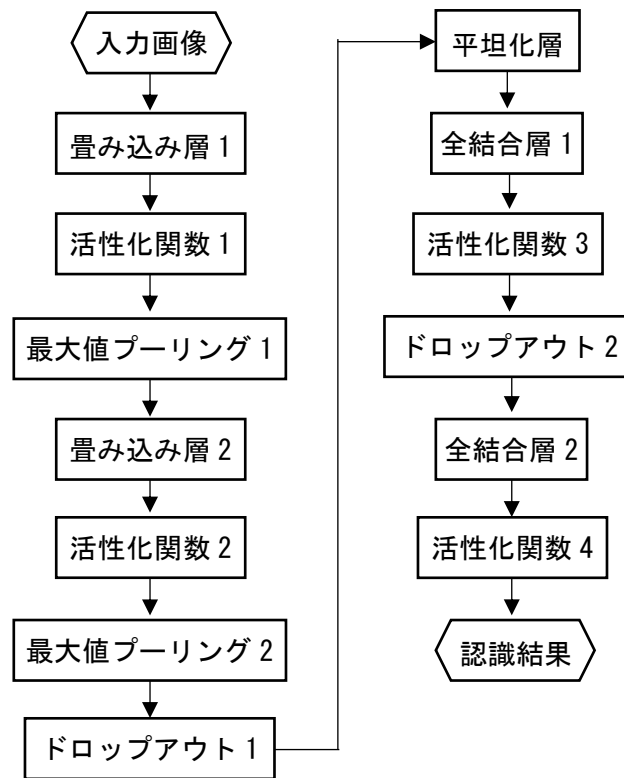


図 2.12 CNN のネットワーク構造

表 2.2 ネットワーク構造の諸元

畳み込み層 1	<ul style="list-style-type: none"> ・フィルタサイズ：横 5×縦 5[画素] ・フィルタ数：20
畳み込み層 2	<ul style="list-style-type: none"> ・フィルタサイズ：横 5×縦 5[画素] ・フィルタ数：50
活性化関数 1 から 3	種類：ReLU
活性化関数 4	種類：ソフトマックス関数
最大値プーリング 1, 2	注目領域：横 2×縦 2[画素]
ドロップアウト 1, 2	ドロップアウト率：0.5
全結合層 1	ノード数：500
全結合層 2	ノード数：10

2.6.3 画質改善効果

対象画像と各 Retinex 処理画像の正答率を図 2.13 に示す.

なお, 図 2.13 における凡例の説明を下記に示す.

- ・ 補正前 : 輝度値補正前の対象画像
- ・ SSR : single scale retinex 処理後の対象画像
- ・ 2MSR : 2 スケールの MSR 処理後の対象画像
- ・ 3MSR : 3 スケールの MSR 処理後の対象画像
- ・ 提案 Retinex : 提案 Retinex を行った後の対象画像

ここに 2 スケール, 3 スケールとは, それぞれ 2.3.2 節の式(2.5)において $N=2, N=3$ であることを表す.

提案 Retinex における正答率の SSR における正答率に対する比について, 対象画像の全劣化強度について最大値, 平均値, 最小値を算出したものを表 2.3 に示す.

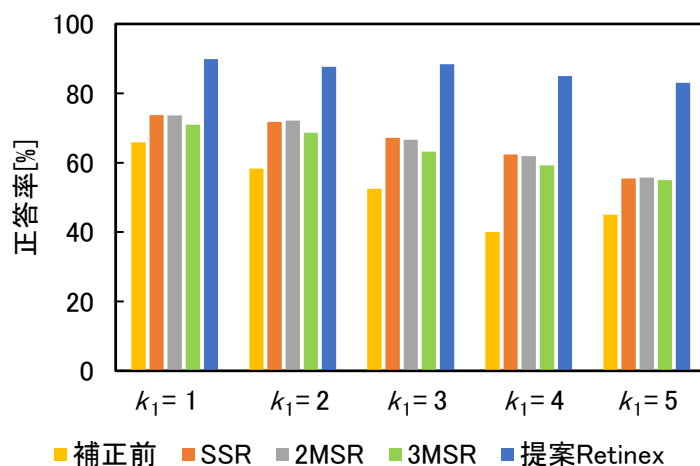


図 2.13 各 Retinex 処理画像の正答率

表 2.3 正答率の比

	テスト画像
最大値	1.50
平均値	1.32
最小値	1.22

なお、図 2.13 において 3 つの従来 Retinex (SSR, 2MSR, 3MSR) における正答率を比較すると、SSR が僅かに優位と認められたため、表 2.3 における提案 Retinex の比較対象として SSR を選んだ。

2.6.4 考察

図 2.13 より、全劣化強度において、提案 Retinex の正答率は、SSR、2MSR、3MSR の正答率よりも高い。対象画像は全劣化が付与された画像であることから、実際の監視カメラ画像と同様の複合劣化画像において、提案 Retinex の性能は、SSR、MSR、3MSR の性能よりも高いといえる。

表 2.3 より提案 Retinex の正答率は、SSR の正答率に比べて約 32 パーセント向上した。また図 2.13 において従来 Retinex のうち SSR が最も優位であることから、本実験では従来 Retinex においてガウス関数の周辺領域を複数設定しても性能は向上しないといえる。よって周辺領域の数を 3 よりも大きい数にしても、性能は向上しないと考えられ、提案 Retinex の性能は、いずれの従来 Retinex の性能よりも高いと考えられる。

すなわち実際の監視カメラ画像と同様の複合劣化画像を対象画像とする場合に、提案 Retinex は、いずれの従来 Retinex よりも高性能であると考えられる。

2.7 まとめ

本章ではナンバー灯による環境光の外乱に対するロバスト性を高めるため、サポートベクター回帰を用いて環境光成分を推定する Retinex 処理を提案する

とともに、提案した手法の利点を述べた。またサポートベクター回帰の実装方法を述べた。そして Halo 作用の評価実験と画質の評価実験について述べた。以下に得られた結果を要約する。

- (1) Retinex 処理は環境光の外乱を除いて視覚的に見やすい画像を復元することに適しているが、従来の Retinex 処理では環境光成分の推定精度が低くナンバー灯による環境光の外乱に対するロバスト性を高めることができない。そこで、このロバスト性を高めるためサポートベクター回帰を用いて環境光成分を推定する Retinex 処理を提案した。
- (2) Retinex 処理においてサポートベクター回帰を用いて環境光成分を推定した場合、物体の局所的な反射率の違いが推定結果に及ぼす影響が低減され、また環境光成分がもつ光学的性質が推定結果に反映されやすくなるため、従来よりも高い精度の推定結果が得られることを述べた。
- (3) Halo 作用低減効果の観点で提案する Retinex 処理の性能を定性的に評価した結果、提案の Retinex 処理の性能は従来の Retinex 処理の性能よりも高かった。また人の視覚システムを模擬した構造をもつ畳み込みニューラルネットワークを用いて提案の Retinex 処理の性能を評価した結果、提案の Retinex 処理画像に対する文字認識の正答率は従来の Retinex 処理画像に対する文字認識の正答率に比べて約 32 パーセント向上した。

参考文献

- (1) E. H. Land, J. J. McCann: “Lightness and Retinex Theory”, Journal of the Optical Society of America, Vol. 61, No. 1, pp. 1-11, 1971.
- (2) V. Vapnik: “The Nature of Statistical Learning Theory”, Springer, New York, pp.181-224, 1995.
- (3) 辻広生, 福水洋平, 道関隆国, 山内寛紀: “Retinex モデルにおけるサポートベクター回帰を用いた劣化ナンバープレート画像の輝度値補正手法と文字認識への応用”, 画像電子学会誌, Vol. 46, No. 4, pp. 510-521, 2017.
- (4) 奥富正敏: “デジタル画像処理”, 公益財団法人画像情報教育振興協会, 東京, pp. 82-87, 2015.
- (5) 田口亮, “視覚システムとその特性に基づく画像処理”, Journal of Signal Processing, Vol. 19, No. 1, pp3-14, 2015.
- (6) D. J. Jobson, Z. Rahman and A. wooddell: “Properties and Performance of a Center/Surround Retinex”, IEEE Transactions on Image Processing, Vol. 6, No.3, pp.451-462, 1997.
- (7) H. Kiragu and E. Mwangi: “An Improved Enhancement of Degraded Binary Text Document Images Using Morphological and Single Scale Retinex Operations”, Proceedings of IET Conference on Image Processing (IPR), pp. 1-6, 2012.
- (8) Z. Rahman, D. J. Jobson and G. A. Woodell: “A Multiscale Retinex for Color Rendition and Dynamic Range Compression”, Proceedings of the international Society for Optical Engineering (SPIE), Vol. 2847, pp. 183-191(1996).
- (9) 田辺和俊, 鈴木孝弘; “サポートベクターマシンを用いた所得格差の決定要因の実証分析”, 情報知識学会誌, Vol. 25, No. 3, pp. 223-242, 2015.
- (10) Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R.E. Howard, W. Hubbard and L. D. Jackel: “Backpropagation applied to handwritten zip code recognition”, Neural Computation, Vol. 1, No. 4, pp. 541-551, 1989.
- (11) D. Capel: “Image Mosaicing and Super-resolution”, Springer, pp. 86-87, 2004.

- (12) V. Nair and G. E. Hinton: “Rectified linear units improve restricted boltzmann machines”, Proceedings of the 27th International Conference on International Conference on Machine Learning(ICML), pp. 807-814, 2010.
- (13) N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov: “A simple way to prevent neural networks from overfitting”, Journal of Machine Learning Research, Vol. 15, No. 1, pp.1929-1958, 2014.

第3章 サポートベクターマシンを用いた画素値選択型マルチフレーム超解像処理による高解像度化法

3.1 まえがき

街頭や店頭に設置された一般的な監視カメラ画像のナンバープレートは偶然映っていたものであり、明るさや画角などの撮影条件がナンバープレートに合わせられていない。このためナンバープレート画像は複合的な劣化を受ける。犯罪捜査ではこの複合劣化画像の文字を視覚的に見やすくし、正確に読み取れるようにすることが求められる。

画角が合っておらずナンバープレートが映っている領域が小さいと、文字情報が少なくなる。文字情報が少なくなると文字は見えにくくなり、これを正確に読み取ることが難しくなる。よって本研究では、少ない文字情報を確実に利用する効率性の高い画質改善法を開発することを課題とした。

少ない文字情報を効率的に利用できる従来の画像処理には複数の低解像度画像のサブピクセル位置を合わせた後、各低解像度画像を統合して1枚の高解像度画像を再構成するマルチフレーム超解像処理⁽¹⁾がある。しかしマルチフレーム超解像処理には複数の低解像度画像の位置を合わせたときに、いずれの低解像度画像の画素位置も当てはまらない高解像度画像の位置（開口部）が出現してしまう問題（開口問題）⁽²⁾がある。開口問題が発生すると少ない文字情報を確実に利用できなくなる。すなわち、開口問題は上記課題解決の障害となる。

開口問題を低減して上記課題を解決するにあたり、本研究ではナンバープレートの文字を対象としており目標とする高解像度画像は規格により字体が定まった高解像度文字画像であることに着目した。すなわち目標とすべき高解像度文字画像を正確に選択し、選択した高解像度文字画像と複数の低解像度文字画像の各位置を正確に合わせることができれば、高解像度文字画像を参照して低解像度文字画像の中から適切な画素値を選択できる。画像の選択や位置合わせ

は分類問題として一般化でき，分類問題には機械学習が有効である．そこで本研究では信頼性が高い機械学習法であるサポートベクターマシン（以下，SVM）を用いて上記の画素値選択を行うマルチフレーム超解像処理を提案する⁽³⁾．

まず 2 節で従来の超解像処理⁽¹⁾について述べる．そして 3 節でサポートベクターマシン⁽⁴⁾を用いた画素値選択型マルチフレーム超解像処理を提案する．4 節でエッジ復元効果の評価について述べ，最後に 5 節で画素値選択効果の評価実験について述べる．

3.2 従来の超解像処理

超解像処理とは低解像度画像から高解像度画像を復元する処理である．超解像処理は画像データベースを用意してこれを学習することで 1 枚の低解像度画像から高解像度画像を復元するシングルフレーム超解像処理と，複数の低解像度画像を再構成して高解像度画像を復元するマルチフレーム超解像処理とに大別される⁽¹⁾．

シングルフレーム超解像処理では高解像度画像と低解像度画像の対応関係を学習して高解像度画像を復元するので低解像度画像に元々含まれていない情報が生成される．したがって，ナンバープレートの損傷，汚損などの特異な特徴は学習画像に通常含まれないので，対象画像がこのような特異な特徴をもつ場合，推定後の画像ではこれらが消失してしまう恐れがある．よって犯罪捜査の応用では一般的にこのような偶発的な情報消失が起こるリスクが低いマルチフレーム超解像処理が用いられる⁽⁵⁾．以降では本研究の目的に適合するマルチフレーム超解像処理について述べる．

3.2.1 マルチフレーム超解像処理の基本概念

マルチフレーム超解像処理の基本概念を表す模式図を図 3.1 に示す⁽¹⁾．動画の撮像過程では互いにサブピクセルの位置ずれがある複数の低解像度画像が取得される．マルチフレーム超解像処理はこの撮像過程の逆の過程に相当し，複数の低解像度画像のサブピクセル位置を合わせ 1 枚の高解像度画像を再構成する画質改善法である．

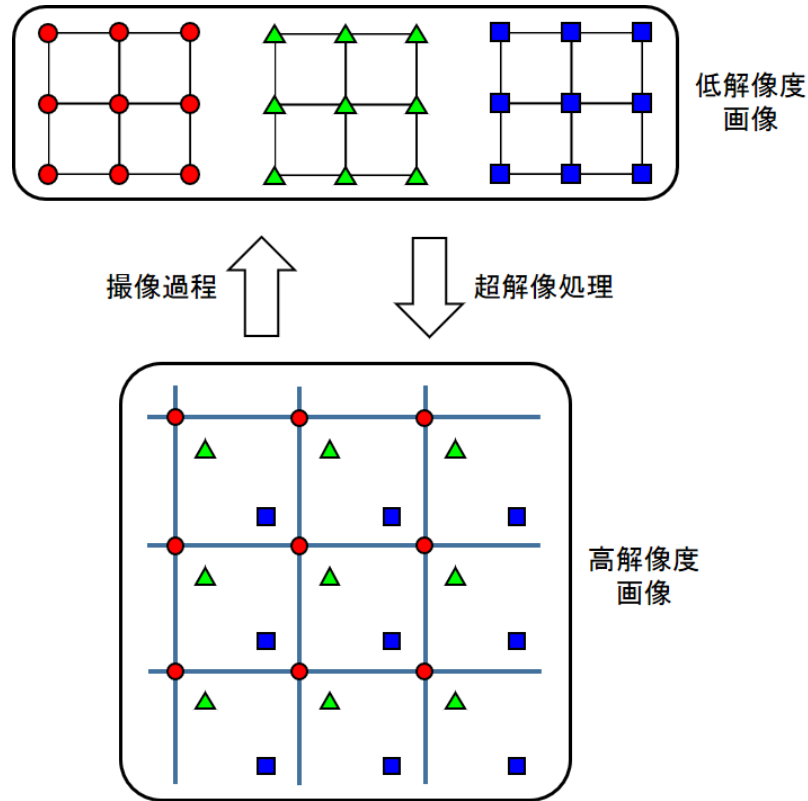


図 3.1 マルチフレーム超解像処理の基本概念を表す模式図

3.2.2 サブピクセル位置合わせの方法

サブピクセル位置合わせは一致度評価値を用いたピクセル単位の位置合わせの後に行う。この一致度評価値には差の二乗和（以下，SSD），差の絶対値和（以下，SAD），正規化相互相関（以下，ZNCC）が用いられる。SSD と ZNCC は画像の類似度であり，それぞれ式(3.1)，式(3.2)で表される。

$$R_{\text{SSD}} = \sum_{j=0}^{N-1} \sum_{i=0}^{M-1} (I(i,j) - T(i,j))^2 \quad (3.1)$$

$$R_{\text{ZNCC}} = \frac{\sum_{j=0}^{N-1} \sum_{i=0}^{M-1} (I(i,j) - \bar{I})(T(i,j) - \bar{T})}{\sqrt{\sum_{j=0}^{N-1} \sum_{i=0}^{M-1} (I(i,j) - \bar{I})^2 \times \sum_{j=0}^{N-1} \sum_{i=0}^{M-1} (T(i,j) - \bar{T})^2}} \quad (3.2)$$

ここに位置合わせの対象となる 2 枚の画像の画素数はいずれも横 M [画素]，縦

N [画素]であるとする. $I(i,j)$ と $T(i,j)$ はこれらの2枚の画像の座標 (i,j) における画素値を表す. \bar{I} と \bar{T} はいずれも領域内における画素値の平均値であり, それぞれ式(3.3), 式(3.4)で表される.

$$\bar{I} = \frac{1}{MN} \sum_{j=0}^{N-1} \sum_{i=0}^{M-1} I(i,j) \quad (3.3)$$

$$\bar{T} = \frac{1}{MN} \sum_{j=0}^{N-1} \sum_{i=0}^{M-1} T(i,j) \quad (3.4)$$

SAD は画像の相違度であり, 式(3.5)で表される.

$$R_{\text{SAD}} = \sum_{j=0}^{N-1} \sum_{i=0}^{M-1} |I(i,j) - T(i,j)| \quad (3.5)$$

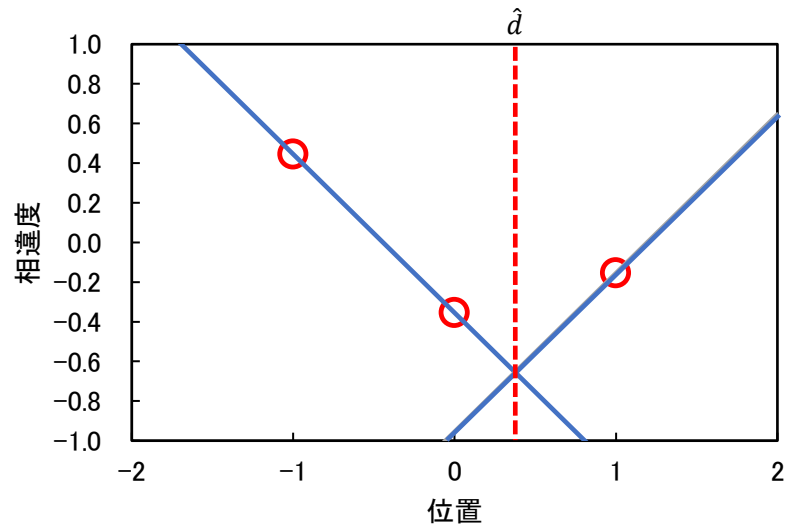
次に, より精密な位置合わせを行うためピクセルで得られた一致度評価値を連続なフィッティング関数で補間しフィッティング関数の最大値または最小値を与えるサブピクセル位置を求める. 通常は画像の横方向と縦方向を独立と考えて各方向に対してサブピクセル位置を求める. 補間関数には等角直線フィッティング, パラボラフィッティングがあり, 各サブピクセル位置 \hat{d} は, それぞれ式(3.6), 式(3.7)で表される.

$$\hat{d} = \begin{cases} \frac{1}{2} \frac{R(1) - R(-1)}{R(0) - R(-1)} & R(1) < R(-1) \text{ のとき} \\ \frac{1}{2} \frac{R(1) - R(-1)}{R(0) - R(1)} & \text{その他} \end{cases} \quad (3.6)$$

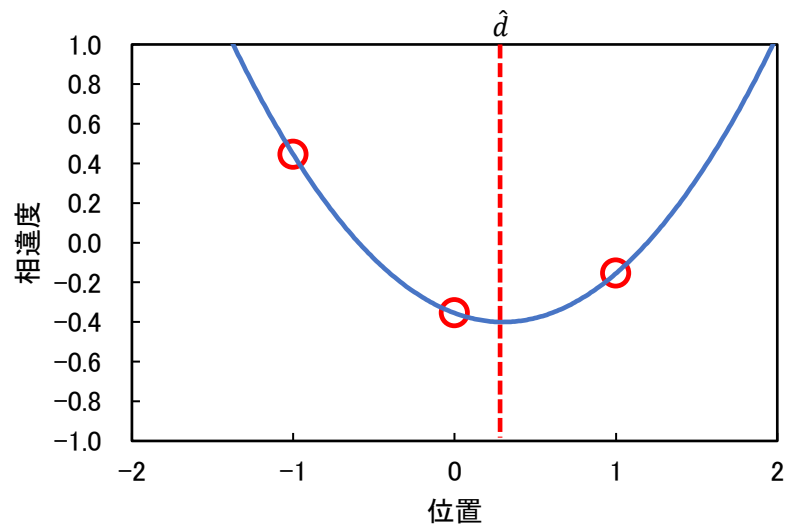
$$\hat{d} = \frac{R(-1) - R(1)}{2R(-1) - 4R(0) + 2R(1)} \quad (3.7)$$

ただし $R(0)$ は相違度 (類似度) が最小 (最大) の位置における相違度 (類似度) の値, $R(-1), R(1)$ はその隣接位置における相違度 (類似度) の値を表す. これらのフィッティング関数によって求まるサブピクセル位置を図 3.2 に示す.

一般的に SAD では等角直線フィッティング, SSD と ZNCC ではパラボラフィッティングを用いると, サブピクセル位置の推定精度が高まることが示されている⁽⁶⁾.



(a) 等角直線フィッティング



(b) パラボラフィッティング

図 3.2 フィッティング関数によって求まるサブピクセル位置

3.2.3 マルチフレーム超解像画像の推定方法

マルチフレーム超解像画像の推定は低解像度画像のサブピクセル位置合わせの後に行われる。この推定はコスト関数を最小化する処理として式(3.8)のよう

に定式化できる⁽⁷⁾.

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \left\{ \sum_{k=1}^N \|\mathbf{Y}_k - M_k \mathbf{X}\|^2 + \lambda A(\mathbf{X}) \right\} \quad (3.8)$$

ここで k は低解像度画像の番号, N は低解像度画像の枚数, \mathbf{Y}_k は k 番目の低解像度画像の画素値をラスタスキャン順に並べたベクトル, \mathbf{X} は高解像度画像の画素値をラスタスキャン順に並べたベクトル, $\hat{\mathbf{X}}$ は推定結果の \mathbf{X} , M_k は k 番目の低解像度画像の劣化を表す行列, λ は拘束の強さを表すパラメータ, $A(\mathbf{X})$ は拘束を表す関数, $\|\cdot\|^2$ は L2 ノルムをそれぞれ表す.

式(3.8)を満たす超解像画像 $\hat{\mathbf{X}}$ を推定する最も代表的な方法は Irani と Peleg の方法と呼ばれる方法⁽⁸⁾であり, 反復法を用いる. Irani と Peleg の方法では, まず M_k が既知であると仮定することで, 式(3.8)を式(3.9)で表される最尤推定の形式に単純化する.

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \left\{ \sum_{k=1}^N \|\mathbf{Y}_k - M_k \mathbf{X}\| \right\} \quad (3.9)$$

次に適当な初期高解像度画像 $\hat{\mathbf{X}}(0)$ を設定し, 式(3.10)で $\hat{\mathbf{X}}(0)$ に対する k 番目の低解像度画像の推定結果 $\hat{\mathbf{Y}}_k(0)$ を計算する.

$$\hat{\mathbf{Y}}_k(0) = M_k \hat{\mathbf{X}}(0) \quad (3.10)$$

そして $\hat{\mathbf{Y}}_k(0)$ と実際に観測された画像 \mathbf{Y}_k との誤差を計算し, その誤差が小さくなるように $\hat{\mathbf{X}}$ を更新していく. すなわち Irani と Peleg の方法は式(3.11)で定式化される.

$$\hat{\mathbf{X}}(i+1) = \hat{\mathbf{X}}(i) + c \sum_{k=1}^N M_k^T (\mathbf{Y}_k - M_k \hat{\mathbf{X}}(i)) \quad (3.11)$$

ここに c は解を収束させるための定数を表す.

3.3 サポートベクターマシンを用いた画素値選択型マルチフレーム超解像処理

従来のマルチフレーム超解像処理 (以下, 従来超解像) には開口問題と呼ばれる, 低解像度画像と高解像度画像の画素位置を一意に対応付けられない問題がある⁽²⁾. 一意に対応付けられない部分 (開口部) では本来の位置とは異なる位

置の画素値が高解像度画像に反映されることになるのでこの部分で画像が劣化してしまう。エッジなどの画素値の変化が大きい箇所ではわずかに位置がずれるだけで画素値が大きく変わってしまうため、開口問題はこれらの箇所で顕著になる。具体的にはエッジ付近で開口問題が発生するとエッジが鈍ってしまう。本研究が対象とする文字画像では、多くのエッジがあるため開口問題を解決することの重要性が高い。

開口問題を低減するため、提案のマルチフレーム超解像処理（以下、提案超解像）では目標とする高解像度文字画像（以下、テンプレート画像）と複数の低解像度文字画像の各サブピクセル位置を合わせた後、テンプレート画像を参照して低解像度文字画像の中から適切な画素値を選択する。

提案超解像では機械学習を用いてテンプレート画像の選択と位置合わせを行うが、画像の選択や位置合わせを間違えると開口部に全く異なる画素値を当てはめることになってしまう。よってこの選択や位置合わせには高い信頼性が求められる。信頼性を高めるためには学習を確実に行う必要がある。このため提案超解像では対象画像の劣化の要因と強度を限定する。これらを限定できる場合とは、対象となる文字画像が撮影された条件と同様の条件でナンバープレートを撮影して学習画像を取得して劣化要因と劣化強度を推定できる場合（処理対象画像を再現して劣化要因と劣化強度を推定できる場合）である。この限定された複合劣化画像では特徴量を柔軟に抽出する必要性が低下するため、提案超解像では特徴量抽出以外の面で信頼性が高い機械学習法であるサポートベクターマシン⁽⁴⁾を用いることとした。提案超解像で画像の選択と位置合わせに使用した分類モデルの模式図を図 3.3 に示す。ここに理想的な環境で撮影したナンバープレート文字画像とは、できるだけ劣化を受けないように新品のナンバープレートを真正面から近距離でデジタルカメラを用いて高画質で撮影した画像から切り出した文字画像を意味する。このモデルでは学習画像を画像劣化モデル⁽⁹⁾に基づいて計算機上で生成するので均衡が取れた劣化要因や劣化強度を設定でき、かつ設定した劣化強度を正確に与えることができる。図 3.3 では比較のため実環境から得た劣化画像データベースを用いる一般的な分類モデルを点線で併記した。

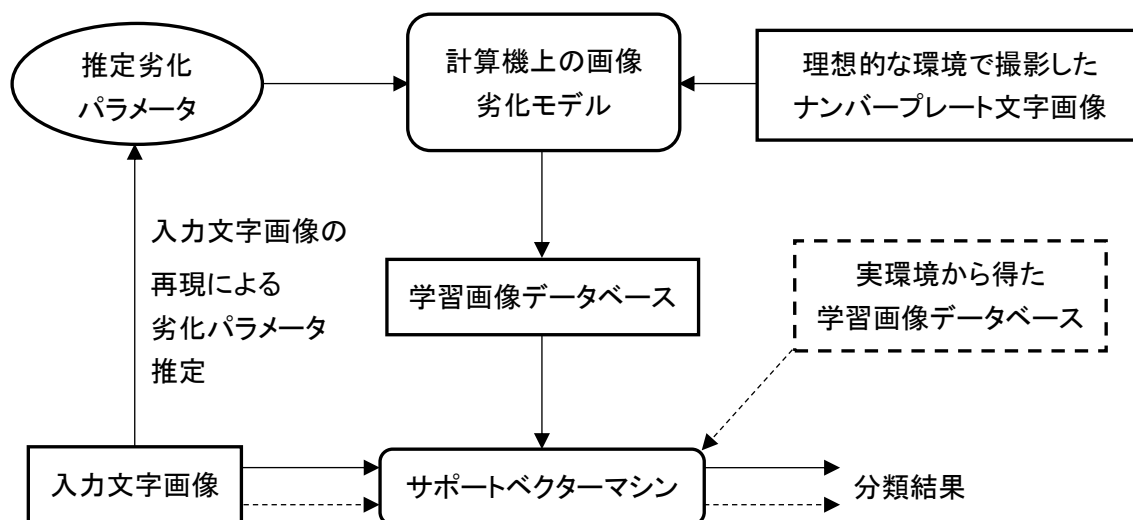
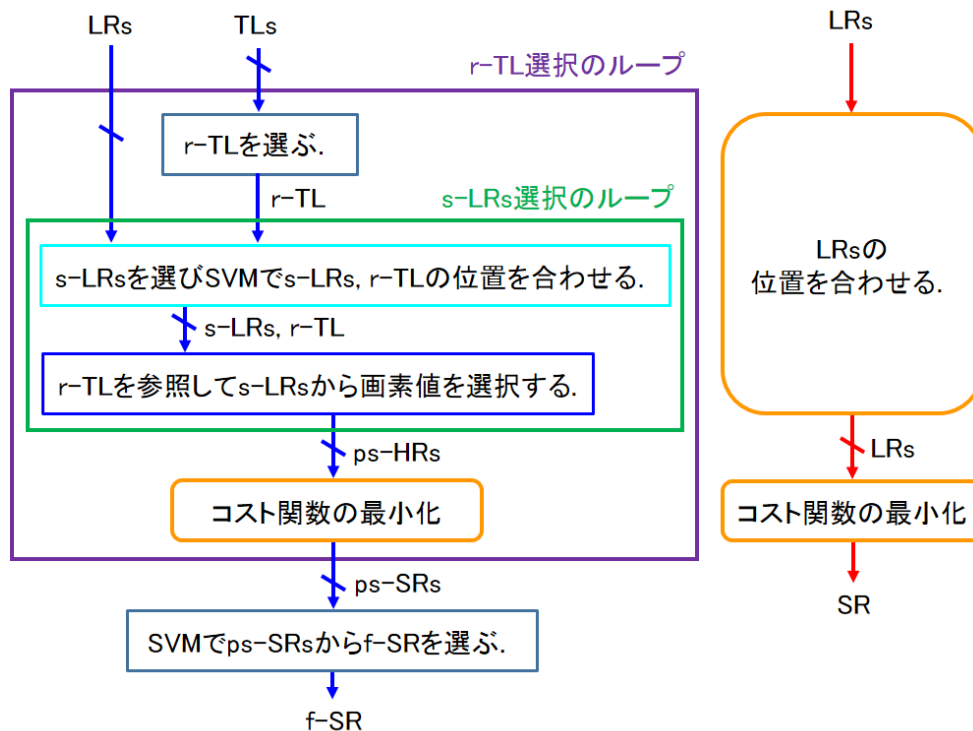


図 3.3 提案超解像における分類モデルの模式図

3.3.1 全体的な処理手順

提案超解像の全体的な処理手順を図 3.4 に示す. 図 3.4 では比較のため従来超解像の全体的な処理手順を併記した. 最初に LR_s から超解像処理に用いる低解像度画像を複数枚選び, これらを s-LR_s とし, また TL_s から 1 枚選び, これを r-TL とする. 次に, サポートベクターマシンを用いて s-LR_s と r-TL のサブピクセル位置を合わせる. サポートベクターマシンを用いたサブピクセル位置合わせの方法については 3.3.4 節で述べる. 次に r-TL を参照して s-LR_s から画素値を選択して ps-HR_s (画素値選択後の高解像度画像) を生成する. 画素値の選択で画質が改善する原理については 3.3.3 節で述べる. s-LR_s の選択方法は複数通りあるため, ps-HR_s は複数の画像から成る画像群となる. この ps-HR_s に対して最尤推定を行うことで ps-SR_s を生成する. r-TL は複数あるため, ps-SR_s も複数の画像から成る画像群となる. 最後にこの ps-SR_s からサポートベクターマシンで f-SR (最終的な超解像画像) を選ぶ.



⇓⇓: 提案超解像の処理の流れ

⇓⇓: 従来超解像の処理の流れ

LRs: 低解像度画像(観測画像)

s-LRs: LRs から選んだ超解像処理に用いる低解像度画像

TLs: テンプレート画像(数字を対象とする場合、「0」から「9」までの劣化がない高解像度画像)

r-TL: TLs から選んだ, 画素値選択の際に参照するテンプレート画像

ps-HRs: 画素値選択後の高解像度画像

ps-SRs: 画素値選択後の超解像画像(適合しない r-TL を参照した画像を含む.)

f-SR: 最終的な超解像画像(適合する r-TL を参照した画像)

図 3.4 提案超解像の全体的な処理手順

3.3.2 画素値の選択方法

開口部では低解像度画像の開口部に対応する画素値の中から, テンプレート画像の開口部に対応する画素値との差の絶対値が最小となる画素値を選択する. すなわち選択する画素値が含まれる低解像度画像の番号の推定結果 \hat{k} は式

(3.12)で表される.

$$\hat{k} = \arg \min_k |V_T - V_{LR}(k)| \quad (3.12)$$

ここで k は低解像度画像の番号, $V_{LR}(k)$ は k 番目の低解像度画像の対応する画素値, V_T はテンプレート画像の対応する画素値を表す.

3.3.3 画素値選択型マルチフレーム超解像処理の利点

画素値選択型マルチフレーム超解像処理の利点の説明図を図 3.5 に示す. ここに PVS の表記は画素値選択を意味する.

図 3.5(a)の最上層にはテンプレート画像が配置されており, 中間層には低解像度画像が配置されている. これらの画像はいずれもサブピクセル位置を合わせた後の画像である. そして最下層にはテンプレート画像を参照して複数の低解像度画像から画素値を選択して得た高解像度画像が配置されている. すなわち図 3.5(a)の 3 次元表示図は各画素の対応関係を示している.

図 3.5 (b)は図 3.5 (a)の最上層と中間層を重ね合わせた 2 次元表示図である. この 2 次元表示図では, 各開口部に対して青色の画素と緑色の画素が等距離に位置するため, 開口部において一意に画素を対応付けられない.

図 3.5(b)の注目線上における画素値変化の一例を示した図が図 3.5(c)である. 図 3.5(c)において, 提案超解像の場合, 青色の画素値と緑色の画素値のうち, 橙色で示したテンプレート画像の画素値との差が小さい画素値を選択する. 図 3.5(c)の真中のグラフに示すとおり提案超解像では画素値を選択することで低解像度画像の開口部に対応する画素値のうち劣化した画素値の影響が除去されるため高解像度画像の劣化が低減される. さらにエッジ付近ではこの選択で本来の画素値に近い画素値が高解像度画像の画素値となるため高解像度画像では急峻なエッジが復元される. 一方, 従来超解像の場合, 最尤推定により開口部の画素値を決定するので開口部の画素値は青色の画素値と緑色の画素値の影響を等しく受け図 3.5(c)の下側のグラフに示すとおり両者の平均値となる. よって従来超解像の場合, 劣化した画素値の影響を受け, またエッジも正確に復元できない.

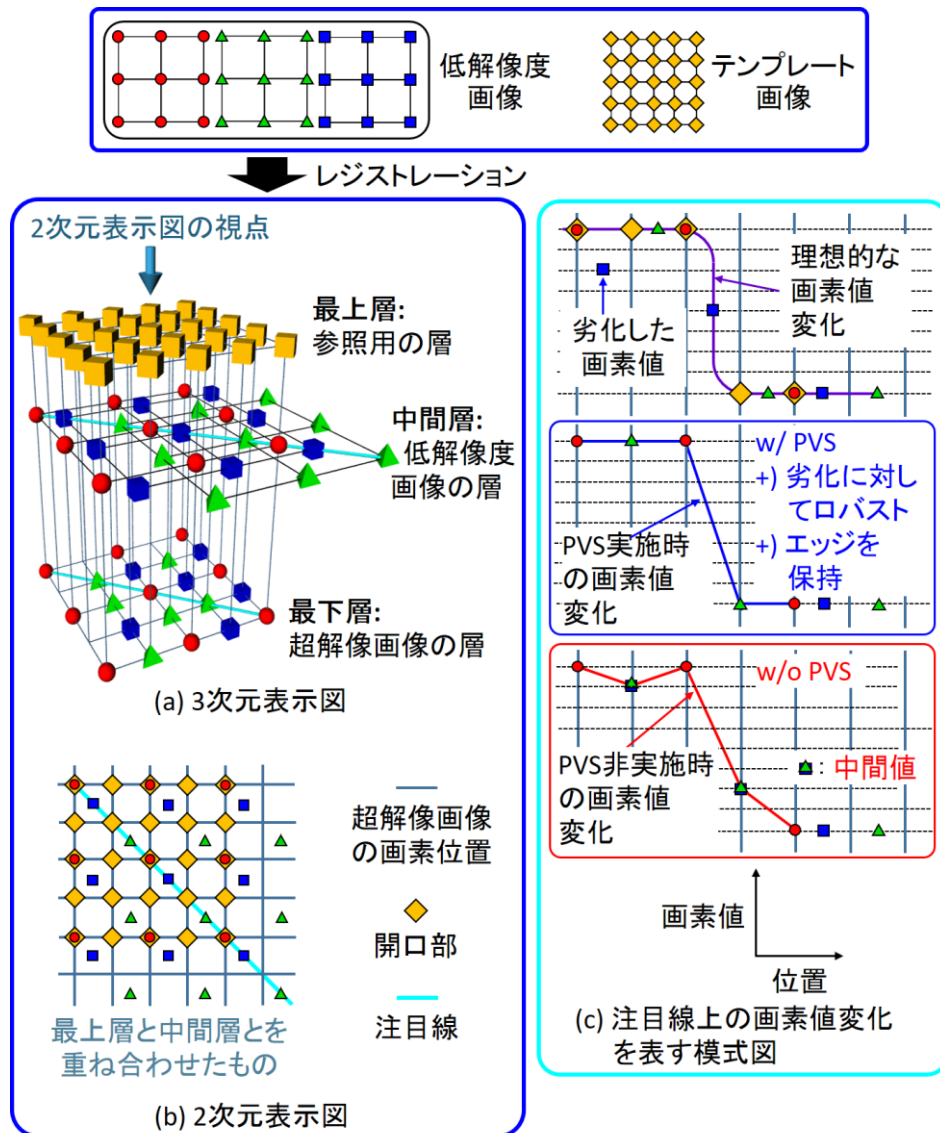


図 3.5 画素値選択型マルチフレーム超解像処理の利点の説明図

3.3.4 サポートベクターマシンを用いたサブピクセル位置合わせの方法

従来のサブピクセル位置合わせにおける誤差は、一致度評価関数とフィッティング関数との不一致により生じる。ここに、一致度評価関数とは、画像の位置と SSD, ZNCC, SAD のうちのいずれかの一致度評価値との間に成立する関数を意味する。SAD に対しては等角直線フィッティング, SSD と ZNCC に対してはパラボラフィッティングを用いると、この不一致が小さくなるが、一致度

評価関数は画像の特徴にも依存することが知られており⁽⁶⁾，複合劣化画像の場合は劣化要因や強度に依存することになる．したがって従来のサブピクセル位置合わせでは複合劣化画像を対象とする場合，一致度評価関数とフィッティング関数が適合せず位置合わせ誤差が大きくなってしまう．

よって本研究ではフィッティング関数を使用せず，各ピクセルをコピーして疑似的に高解像度化することによりサブピクセル位置を離散的に定義し，この離散的なサブピクセル位置に対象画像を分類することで位置を合わせる方法を用いた．

離散的なサブピクセル位置の定義の例を図 3.6 に示す．図 3.6 の赤色の丸と青色の四角は位置合わせの対象となる 2 枚の画像の対応するピクセルを表しており，赤色の丸をピクセル 1，青色の四角をピクセル 2 と呼ぶこととする．図 3.6 の例では横方向と縦方向にそれぞれ 5 通り，合計 25 通りのサブピクセル位置が定義される．この方法ではサブピクセル位置の離散化に伴う誤差が生じるが，図 3.6 のように離散化の幅を小さくすればこの誤差を実用上問題ない大きさに抑えることができる．分類問題には機械学習が有効であり，提案超解像では信頼性が高い多値サポートベクターマシン⁽⁴⁾によりこの分類を行う．

多値サポートベクターマシンを用いたサブピクセル位置合わせの手順を以下に示す．

(i) 対象とする監視カメラを用いて対象画像の撮影状況を再現し，主要な劣化要因（ぼけ，ノイズ，画像圧縮）について対象画像の劣化強度を推定する．そして理想的な環境で撮影したナンバープレート画像を用いて図 3.6 で定義した全サブピクセル位置の文字画像を生成し，生成した画像に画像劣化モデル⁽⁹⁾に基づき複合劣化を与え，この複合劣化画像を学習画像とする．

なお理想的な環境で撮影したナンバープレート画像とは，できるだけ劣化を受けないように新品のナンバープレートを真正面から近距離でデジタルカメラを用いて高画質で撮影した画像を意味する．また上記画像劣化モデル⁽⁹⁾については，2.6.1 節で述べたとおりである．

- (ii) (i)の学習画像を用いて、多値サポートベクターマシンによりサブピクセル位置を学習し、分類器を作成する。
- (iii) (ii)で作成した分類器を用いて対象画像とテンプレート画像のサブピクセル位置を合わせる。

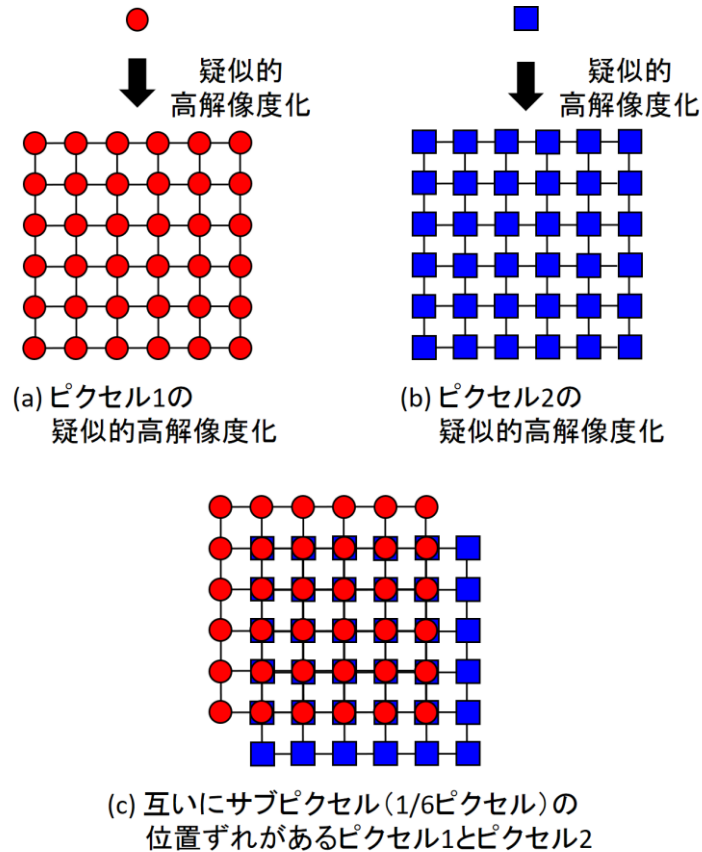


図 3.6 離散的なサブピクセル位置の定義の例

3.3.5 サポートベクターマシンを用いたテンプレート画像の選択方法

テンプレート画像の選択は提案超解像の最終処理として行われる。例として数字画像に提案超解像を適用する場合について説明する。テンプレート画像は0から9までの劣化のない10枚の高解像度数字画像であり、どの数字画像を参照すべきか画素値を選択する段階ではわからない。よって全てのテンプレート画像を順に参照する。したがって、この過程で10枚の超解像画像が生成される。これらの10枚の超解像画像の生成例を図3.7に示す。図3.7の(a)は4枚の

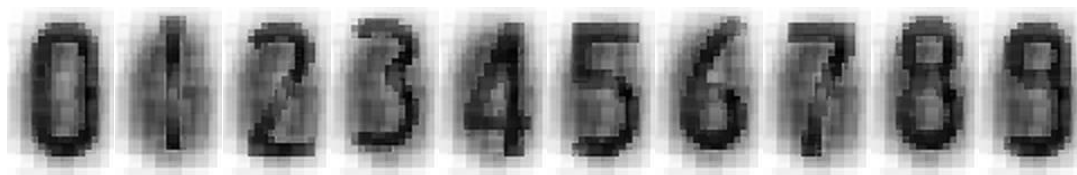
低解像度画像, (b)は左から順に 0 から 9 までのテンプレート画像, (c)は左から順に 0 から 9 までのテンプレート画像を参照して画素値を選択した超解像画像であり図 3.4 の ps-SRs に相当する. これらの 10 枚の画像から適切なテンプレート画像が参照された 1 枚の画像を選択しなければならないが, この 1 枚の画像の選択を間違えると本来とは異なる画像が生成されてしまう. よって適切なテンプレート画像が参照された超解像画像を確実に選択する必要があるが, 図 3.7 (c)に示すように不適切なテンプレート画像を参照した場合には本来の数字と本来とは異なる数字とが混合した特徴的な画像が生成されるため, これらを除外することは比較的容易である. 提案超解像では適切なテンプレート画像が参照された超解像画像と不適切なテンプレート画像が参照された超解像画像についての学習画像を用意し, 信頼性が高い機械学習法の多値サポートベクターマシン⁽⁴⁾によりこの分類を行って, 不適切なテンプレート画像が参照された超解像画像を除外する.



(a) 低解像度画像



(b) テンプレート画像



(c) 画素値を選択した超解像画像

図 3.7 画素値を選択した超解像画像の生成例

3.4 エッジ復元効果の評価

エッジは画像の高周波数成分が失われるとなだらかになる。したがってエッジ復元効果の評価は画像が細かい構造まで表現できているかどうかを判断する指標の1つとなる。エッジは見た目に明らかであるためその特性を人が評価する場合でも主観が評価結果に及ぼす影響は少ないと考えられる。よってエッジ復元効果については定性的に評価することとした。

3.4.1 評価の条件

解像度低下、ぼけ、ノイズ、JPEG 圧縮による複合劣化画像を各高解像度画像生成処理の対象画像とした。本実験では上記の各劣化を計算機上で与えた。具体的には最初に新品のナンバープレートをスキャナ（エプソン製 LP-M5000A）により高解像度（横 4,960[画素]，縦 3,507[画素]）で撮影し無圧縮で保存した。以降においてこの無圧縮で保存した画像を実写画像 2 と呼ぶこととする。実写画像 2 のナンバープレート領域の解像度は横 3,895[画素]で縦 1,935[画素]であった。そしてこの実写画像 2 から切り出した数字画像に計算機で図 3.6 と同じ定義の 25 通りのサブピクセル位置を与えた画像を生成した。最後に計算機で各サブピクセル位置の画像に画像劣化モデル⁽⁹⁾に基づき各劣化を付与し、これを評価用画像とした。この画像劣化モデルについては 2.6.1 節で述べたとおりである。これらの評価用画像のうち、超解像処理の対象となる画像を対象画像、サブピクセル位置合わせのためサポートベクターマシンにより学習する画像を学習画像と呼ぶこととする。学習画像と対象画像の劣化要因と劣化強度を表 3.1 に示す。また対象画像と学習画像の例をそれぞれ図 3.8、図 3.9 に示す。

表 3.1 実験用画像の劣化要因と劣化強度

	対象画像	学習画像
解像度低下[画素]	横 6×縦 11	横 6×縦 11
ガウシアンフィルタによるぼけ (対象画像の標準偏差 σ_{b3} [画素]) (学習画像の標準偏差 σ_{b4} [画素])	$\sigma_{b3} = 0.91$	$\sigma_{b4} = 0.90$
ガウスノイズ (対象画像の標準偏差 σ_{n3}) (学習画像の標準偏差 σ_{n4})	$\sigma_{n3} = 0.091$	$\sigma_{n4} = 0.090$
JPEG 圧縮 (対象画像の品質 Q_3) (学習画像の品質 Q_4)	$Q_3 = 30$	$Q_4 = 30$

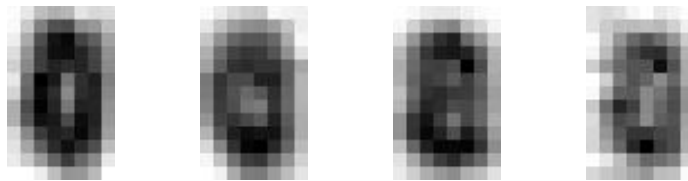


図 3.8 対象画像

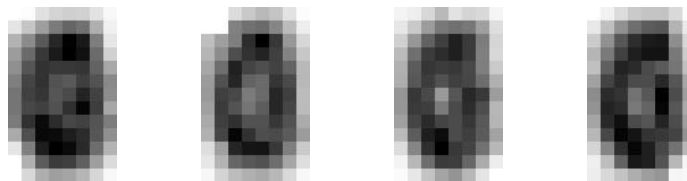
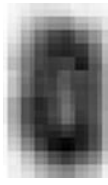


図 3.9 学習画像の例

3.4.2 エッジ復元効果

従来超解像の画像と提案超解像の画像の例を図 3.10 に示す. 図 3.10 の(a)と(b)はいずれも図 3.8 の 4 枚の画像に対して超解像処理を行ったものである.



(a) 従来超解像の画像



(b) 提案超解像の画像

図 3.10 超解像画像の例

3.4.3 考察

図 3.10 より提案超解像を適用した画像は，従来超解像を適用した画像よりもエッジが鮮鋭であることがわかる．本来の数字画像は鮮鋭なエッジをもつことから，提案超解像はエッジ復元効果が高いといえる．

3.5 画素値選択効果の評価実験

サポートベクターマシンを用いた画素値選択型マルチフレーム超解像処理における画質改善は画素値選択に起因する．そこで画素値選択の効果を確かめるため，低解像度画像のサブピクセル位置合わせを行った後，画素値を選択せずに対応する複数の画素値の平均値を用いて生成した高解像度画像と，画素値を選択して生成した高解像度画像の振幅スペクトルを求め，これらの振幅スペクトルに基づいて画素値選択効果の評価した．

3.5.1 実験条件

対象画像と学習画像については 3.4.1 節と同じものを用いた．

高解像度画像の生成方法について，本実験は画素値選択の効果を確かめることが目的であるので，比較する 2 枚の画像において画素値選択以外の処理は同じにし，テンプレート画像は目標とすべき適切な画像のみを用いた．

まず，画素値を選択しない場合の高解像度画像生成方法を以下に示す．

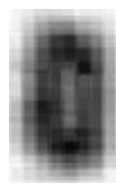
- (i) サポートベクターマシンにより対象画像のサブピクセル位置を合わせる.
- (ii) 対象画像のサブピクセル位置を合わせた結果として出現した開口部で, 対象画像の対応する全画素値の平均値を求め, この平均値を高解像度画像の開口部の画素値とする.

画素値を選択する場合の高解像度画像生成方法を以下に示す. ここに画素値を選択する場合の高解像度画像は図 3.4 の ps-HRs のうち適切なテンプレート画像を用いた画像に相当する.

- (i) サポートベクターマシンにより対象画像と目標とすべき適切なテンプレート画像のサブピクセル位置を合わせる.
- (ii) 対象画像のサブピクセル位置を合わせた結果として出現した高解像度画像の開口部で, テンプレート画像の対応する画素値との差が最も小さい低解像度画像の画素値を選択し, 選択した画素値を高解像度画像の開口部の画素値とする.

3.5.2 画素値選択効果

画素値を選択していない高解像度画像と画素値を選択した高解像度画像をそれぞれ図 3.11 の(a)と(b)に示す.



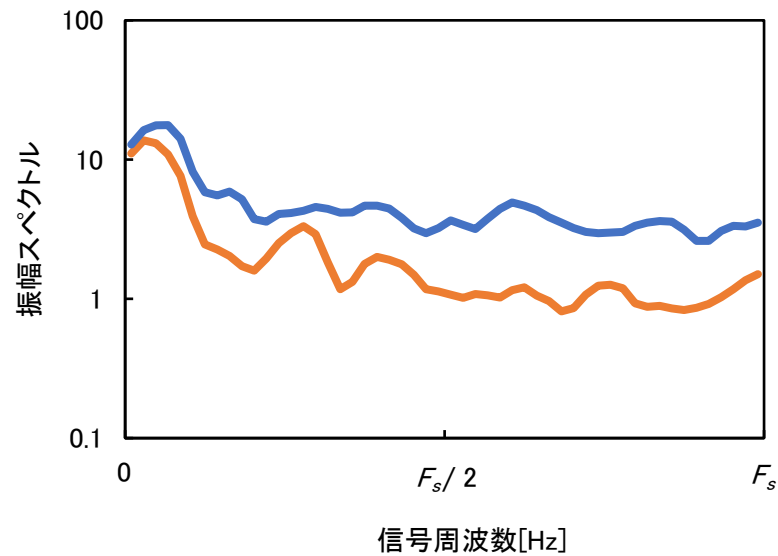
(a) 画素値を選択していない高解像度画像



(b) 画素値を選択した高解像度画像

図 3.11 生成した高解像画像

図 3.11 に示した各高解像度画像の振幅スペクトルを図 3.12 に示す.



— 画素値を選択していない高解像度画像 — 画素値を選択した高解像度画像

F_s : サンプル周波数

図 3.12 各高解像度画像の振幅スペクトル

図 3.12 において、画素値を選択した高解像度画像の振幅スペクトルの総和は、画素値を選択していない高解像度画像の振幅スペクトルの総和に比べて約 208 パーセント向上した。

3.5.3 考察

図 3.11 より画素値を選択することにより、本来の数字がもつ急峻なエッジを復元できていることを確認できる。

また図 3.12 より画素値を選択することにより全体的に振幅スペクトルが増加していることがわかる。劣化強度が強ければ強いほどエッジが鈍りノイズが混入するため画像のコントラストが低下する。よって劣化強度が強い画像ほど振幅スペクトルが低下する。したがって画素値を選択することにより画像のコントラストが向上し、画質が改善しているといえる。さらに図 3.12 をより詳しくみると、画素値を選択することにより特に高い周波数において振幅スペクトルが増加していることがわかる。高い周波数における振幅スペクトルの増加は特

にエッジが急峻になることによってもたらされる。したがって図 3.11 の視覚的なエッジの特徴が、図 3.12 では定量的な形で表れたといえる。

3.6 まとめ

本章では少ない文字情報を確実に利用して視覚的に見やすい画像を効率的に再構成するため、サポートベクターマシンを用いた画素値選択型マルチフレーム超解像処理を提案するとともに、提案した手法の利点を述べた。またサポートベクターマシンを用いた、サブピクセル位置合わせの方法とテンプレート画像の選択方法について述べた。そしてエッジ復元効果の評価と画素値選択効果の評価実験について述べた。以下に得られた結果を要約する。

- (1) 少ない文字情報を効率的に利用する従来手法にマルチフレーム超解像処理があるが、従来のマルチフレーム超解像処理では開口問題により文字情報を確実に利用できなくなり効率性が低下する。そこで、この効率性を向上させるためサポートベクターマシンを用いた画素値選択型マルチフレーム超解像処理を提案した。
- (2) 画素値を選択した高解像度画像では、開口部で低解像度画像の劣化した画素値の影響が除去され、さらにエッジ付近で本来の画素値に近い画素値が選択されることにより急峻なエッジが復元されるため、画質が改善することを述べた。
- (3) 画素値を選択した高解像度画像と画素値を選択していない高解像度画像とを定性的に評価した結果、画素値を選択することにより急峻なエッジが復元されることがわかった。さらにこれらの画像を振幅スペクトルにより定量的に評価した結果、画素値を選択した場合、画素値を選択していない場合に比べて振幅スペクトルが約 208 パーセント向上することを示し、画素値の選択が画質を改善させる効果をもつことを明らかにした。

参考文献

- (1) P. Milanfar: “Super-resolution Imaging”, CRC Press, 2010.
- (2) H. Takeshima, T. Kaneko: “Image Registration Using Subpixel-shifted Images for Super-resolution”, 15th IEEE International Conference on Image Processing (ICIP), pp. 2404-2407, 2008.
- (3) H. Tsuji, Y. Fukumizu, T. Douseki and H. Yamauchi: “Super-resolution on Pixel Values Selection for Degraded Image Recognition by Support Vector Machine”, Proceedings of 21st Triennial Meeting of the International Association of Forensic Sciences (IAFS), p. 248, 2017.
- (4) V. Vapnik: “The Nature of Statistical Learning Theory”, Springer, New York, pp.181-224, 1995.
- (5) J. Sapiro, K. Nasrollahi, P. Correia, T. B. Moeslund: “Super-resolution of Facial Images in Forensics Scenarios”, IEEE 5th International Conference on Image Processing Theory, Tools and Applications (IPTA), pp.55-60, 2015.
- (6) 清水雅夫, 奥富正敏: “画像のマッチングにおけるサブピクセル推定の意味と性質”, 電子情報通信学会論文誌 D, Vol.J85-D2, No.12, pp.1791-1800, 2002.
- (7) S. C. Park, M. K. Park and M. G. Kang, “Super-resolution Image Reconstruction: A Technical Overview”, IEEE Signal Processing Magazine, Vol. 20, No. 3, pp. 21-36, 2003.
- (8) M. Irani and S. Peleg: “Motion Analysis for Image Enhancement Resolution Occlusion and Transparency”, Journal of Visual Communication and Image Representation, Vol. 4, pp. 324-335, 1993.
- (9) D. Capel: “Image Mosaicing and Super-resolution”, Springer, pp. 86-87, 2004.

第4章 多重構造の畳み込みニューラルネットワークによる文字認識法

4.1 まえがき

街頭や店頭に設置された一般的な監視カメラ画像のナンバープレートは偶然映っていたものであり、明るさや画角などの撮影条件がナンバープレートに合わせられていない。このためナンバープレート画像は複合的な劣化を受ける。犯罪捜査ではこの複合劣化画像の文字を正確に読み取れるようにすることが求められる。

監視カメラ画像の文字認識では、複合劣化のため単一の候補を選ぶだけでは正答の文字を実用上必要とされる確率で得ることができない場合が多い。よって監視カメラ画像の文字認識では、通常、複数候補を選ぶ⁽¹⁾。しかし可能な限り候補に含まれる文字の数（以下、候補数）を少なくしないと、車両を絞り込めないだけでなく、認識結果の信用性も損なわれる。一般的に、犯罪捜査を目的とした監視カメラ画像のナンバープレート文字認識では、第2候補まで選んで90パーセント以上の正答率を達成できるかどうかを実用レベルの認識性能の基準となる。上記の正答率は正答の文字が候補に含まれる確率を意味する。また上記の基準における正答率と候補数は、多数回の認識における平均値である。

形状が複雑で種類が多い文字を認識対象とし、また多様な複合劣化に対応しなければならぬ場合には文字認識法に高い適応性が要求され、実用レベルの認識性能を達成することが難しくなる。よって本研究では、文字の特徴や劣化の状態に柔軟に適応できる文字認識法を開発することを課題とした。

文字の特徴や劣化の状態に柔軟に適応することに適した従来の文字認識法に畳み込みニューラルネットワーク（以下、CNN）による文字認識法⁽²⁾がある。しかしCNNには入力画像の解像度を調整する機能がなく、十分に上記の適応性を高めることができないため、実用レベルの認識性能で平仮名を認識できない。そこで本論文では、多重構造のCNNにより多重解像度画像⁽³⁾を処理する文字認識法⁽⁴⁾を提案する。

まず 2 節で一般的な CNN の構成について述べ、3 節で一般的な CNN による文字認識法の問題について述べる。4 節で従来の複数候補選択法について述べ、5 節で提案の文字認識法について述べる。最後に 6 節で認識性能改善効果の評価実験について述べる。

4.2 一般的な畳み込みニューラルネットワークの構成

CNN は人の視覚系の受容野をモデル化した構成となっている。CNN の構成を表す模式図を図 4.1 に示す。

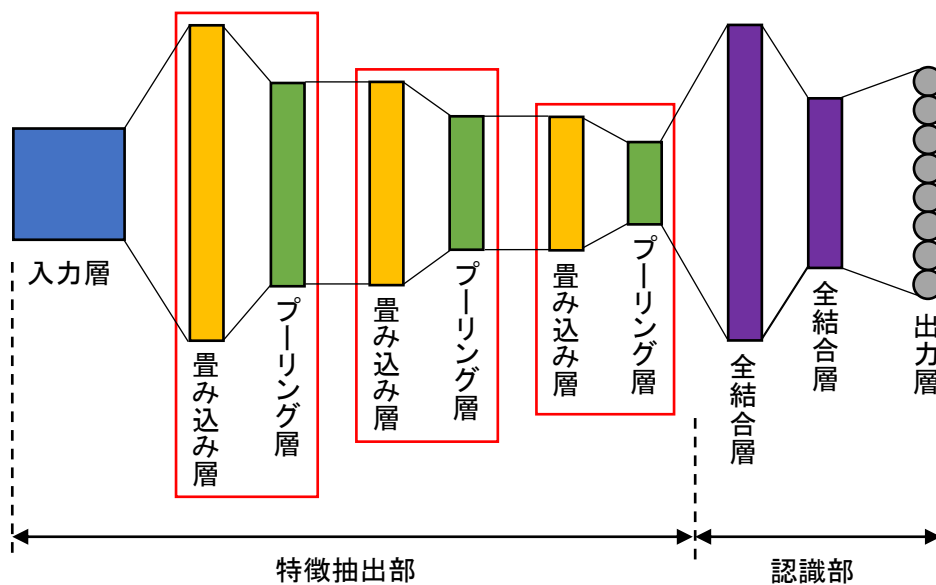


図 4.1 CNN の構成を表す模式図

CNN は図 4.1 に示すように入力層、畳み込み層、プーリング層、全結合層、出力層から構成されている。畳み込み層とプーリング層が複数回繰り返されて深い層が形成され、その後の全結合層も同様に複数回繰り返される。

CNN などの深層学習が出現する前は、人手による試行錯誤で特徴を抽出していたが、CNN では画像の画素値を入力とし、学習により特徴を獲得して認識する。CNN は各階層が全て結合しておらず、結合が疎になっているため誤差の勾配が拡散しづらく過学習を防ぐことができる。

CNN ではまず入力画像に対して重みフィルタにより畳み込み処理を行う。畳み込み処理の結果は特徴マップと呼ばれ、次の層ではこの特徴マップに対してプーリング処理を行う。プーリング処理によって認識に有効な新たな特徴マップが得られる。特徴抽出部ではこの畳み込みとプーリング処理を繰り返して特徴を抽出する。認識部では特徴抽出部の最後の層の特徴マップを入力とし、通常の階層型ニューラルネットワークと同じ処理を行う。

CNN の学習は誤差逆伝播法⁽⁵⁾を用いた最急降下法により重みフィルタと結合係数を更新することにより行う。

4.3 一般的な畳み込みニューラルネットワークによる文字認識法の問題

CNN で手書き文字認識を行った論文の成果 (LeNet⁽²⁾) を劣化文字認識に応用し、多様な劣化に対応した認識を実現した例がある⁽⁶⁾。しかしながらこの一般的な CNN では、90 パーセントの正答率を達成するのに 18 もの候補を選ばなければならない。以下の 2 つの問題を解決して認識性能を改善することが望まれている。

第 1 の問題は、一般的な CNN は入力画像の解像度を調整して正答率を最大化する機能を持たないことである。その例を図 4.2 に示す。

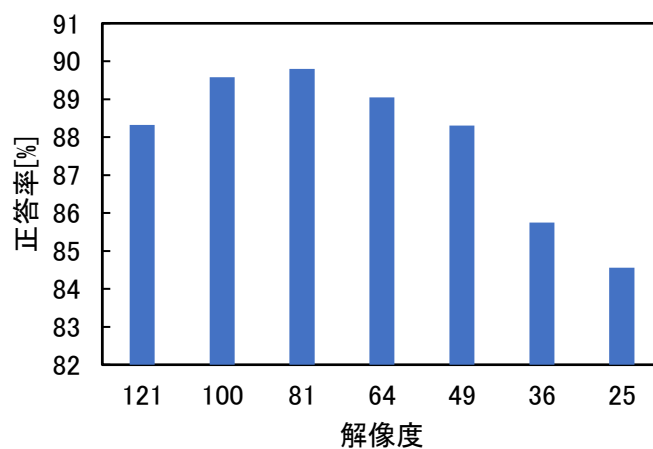


図 4.2 解像度と正答率との関係

これは CNN により劣化文字認識を行った例であり，劣化文字の解像度と正答率との関係を表している．解像度の指標には画素数を用いた．この実験で用いたテスト画像と学習画像は，実写した画像に計算機上で複合劣化を与えたものである．この複合劣化画像を縮小して7段階の多重解像度画像を生成した．

なお以降の3節，5節，6節の画像は全て，上記同様，計算機上で生成した．

理想的には，認識に最適な解像度の画像が自動生成されるので，正答率に極大値が現れない．すなわち，解像度が下がると正答率は単調減少する．しかしながら，実験結果はそのようになっていない．これは CNN に解像度を自動調整する機能がないためであると考えられる．

第2の問題は，CNN には入力画像の解像度を調整して候補数を適応的に制御し正答率（例えば90パーセント）を保証する機能がないことである．その例を図4.3に示す．

これは，テスト画像の劣化の強さ（以下，劣化強度）と正答率との関係を表している．この実験で用いたテスト画像と学習画像の，劣化要因と劣化強度を表4.1に示す．テスト画像の劣化強度を学習画像の劣化強度より強くしたのは，テスト画像の劣化強度を正確に算出することはできないからである．解像度は11に固定した．テスト画像の劣化強度をパラメータ k_2 で表すこととする． k_2 は1から13までの13段階あり値が大きいほど劣化が強い．パラメータ k_2 で表されるガウスフィルタによるぼけの標準偏差 σ_{b5} ，ガウスノイズの標準偏差 σ_{n5} ，JPEG圧縮の品質 Q_5 をそれぞれ式(4.1)から式(4.3)に示す．

$$\sigma_{b5} = 2.35 + 0.05 k_2 \quad (4.1)$$

$$\sigma_{n5} = 0.065 + 0.015 k_2 \quad (4.2)$$

$$Q_5 = 54 - 2k_2 \quad (4.3)$$

ガウスノイズについては，不規則性があるため，同じ劣化強度でもシードごとに認識精度が異なる．よって劣化強度ごとにシードを変えて10枚生成し，不規則性が認識精度に与える影響を軽減した．劣化強度が強い場合，正答率が90パーセント以下（赤色の線以下）となることが確認できる．

なお，3節，5節，6節のテスト画像と学習画像は，いずれも表4.1に示す諸元の画像である．

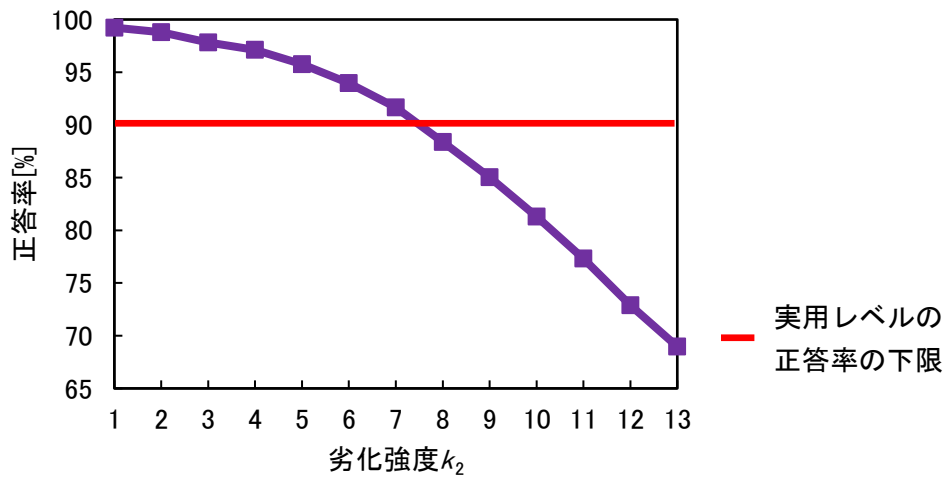


図 4.3 テスト画像の劣化強度と正答率との関係

表 4.1 テスト画像と学習画像の劣化要因と劣化強度

		テスト画像	学習画像
劣化強度固定	解像度低下[画素]	横 11×縦 11	横 11×縦 11
	切り出し誤差[画素]	上下左右に各 1	上下左右に各 1
劣化強度可変	ガウシアンフィルタによるぼけ (テスト画像の標準偏差 σ_{b5} [画素]) (学習画像の標準偏差 σ_{b6} [画素])	$2.40 \leq \sigma_{b5} \leq 3.00$	$1.00 \leq \sigma_{b6} \leq 2.50$
	ガウスノイズ (テスト画像の標準偏差 σ_{n5}) (学習画像の標準偏差 σ_{n6})	$0.080 \leq \sigma_{n5} \leq 0.260$	$0.050 \leq \sigma_{n6} \leq 0.110$
	JPEG 圧縮 (テスト画像の品質 Q_5) (学習画像の品質 Q_6)	$28 \leq Q_5 \leq 52$	$48 \leq Q_6 \leq 90$

4.4 従来の複数候補選択法

従来の複数候補選択手法⁽⁷⁾は、信頼度（所属確率）が高い順にクラスを順位付けした後、候補数 m を信頼度の分布に基づいて決定し、1位から m 位のクラスを候補とする。クラスは本研究の場合、平仮名の種類となる。 m を決定するため、5種類のルールが提案されている。各ルールを定義する数式を下記に示す。

なお、下記の各数式では、各クラスの信頼度を降順にランク付けしたときにおける j 番目の信頼度を $Conf_j$ 、 $Conf_j$ に対して設定するしきい値を T で表す。

- Confidence Accumulation (CA)

$$\min_m \left\{ \sum_{j=1}^m Conf_j \geq T \right\} \quad (4.4)$$

- Confidence Thresholding (CT)

$$\max_m \{ Conf_m \geq T \} \quad (4.5)$$

- Difference Thresholding 1 (DT1)

$$\min_m \{ Conf_m - Conf_{m+1} > T \} \quad (4.6)$$

- Difference Thresholding 2 (DT2)

$$\max_m \{ Conf_1 - Conf_m \leq T \} \quad (4.7)$$

- Ratio Thresholding (RT)

$$\max_m \left\{ \frac{Conf_m}{Conf_1} \geq T \right\} \quad (4.8)$$

4.5 多重構造の畳み込みニューラルネットワークによる文字認識法

一般的な CNN では入力画像の解像度を調整する機能がないことが問題である。そこでこの問題を解決して認識性能を改善するため、多重構造の CNN で多重解像度画像を処理する文字認識法を提案する。

多重構造の CNN による文字認識法（以下、提案の文字認識法）における認識

モデルの模式図を図 4.4 に示す．ここに理想的な環境で撮影したナンバープレート文字画像とは，できるだけ劣化を受けないように，新品のナンバープレートを真正面から近距離でデジタルカメラを用いて高画質で撮影した画像から切り出した文字画像を意味する．また処理対象画像の劣化強度は，パワースペクトル解析などに基づいて推定する．このモデルでは学習画像を画像劣化モデル⁽⁸⁾に基づいて計算機上で生成するので広範囲で均衡が取れた劣化要因や劣化強度を設定でき，かつ設定した劣化強度を正確に与えることができる．図 4.4 では比較のため実環境から得た劣化画像データベースを用いる一般的な認識モデルを点線で併記した．

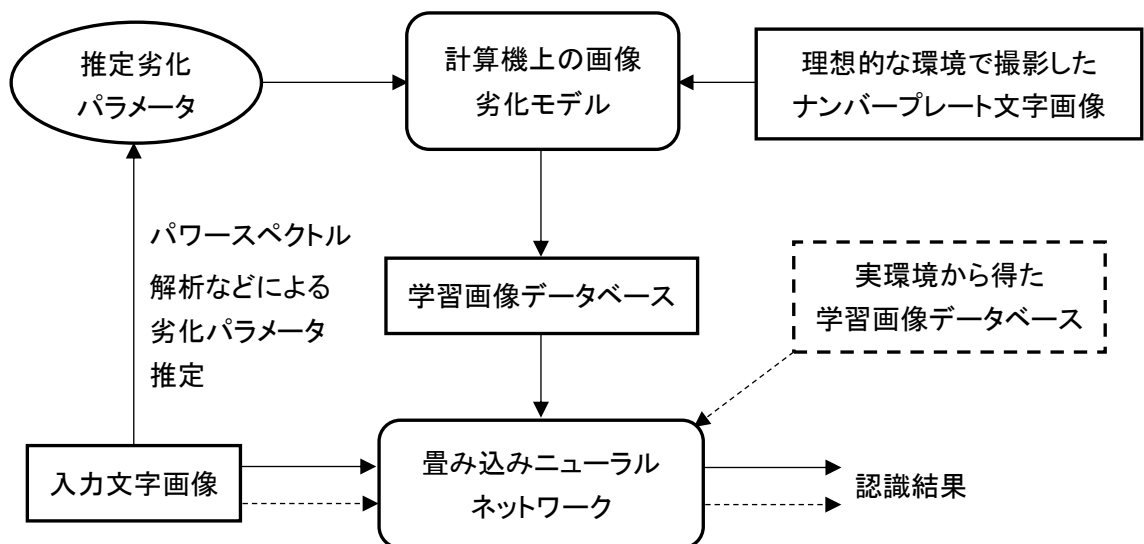


図 4.4 提案の文字認識法における認識モデルの模式図

4.5.1 多重構造の畳み込みニューラルネットワークの全体構成

全体構成を示す模式図を図 4.5 に示す．図 4.5 における記号の説明を下記に示す．

- L1 から L7：多重解像度画像の 7 段階の階層を表し，L1 から L7 の順に解像度が低くなる．
- CL：正解ラベル

なお多重解像度画像の具体的な階層については4.5.2節で述べる。

図4.5の実線部分と点線部分は、それぞれ学習時とテスト時における、使用部分と不使用部分を表す。ただし、図4.5の破線部分については、どの階層の特徴量を抽出するかにより、複数通りの構成が考えられる。本研究では t を2以上7以下の整数とし、1番目の階層(L1)から t 番目の階層(L t)までを使用する構成とした。候補数は t により調節することができ、 t を大きくすればするほど候補数が増える。図4.5で各認識結果の重複を除去と記載したブロックの処理は、各階層で得られる平仮名の全種類(1種類以上、7種類以下)を求めめるために行う。上記の処理により求められる平仮名の全種類を元(要素)とする集合 P は、式(4.9)で表される。

$$P = \bigcup_{s=1}^t C_s \quad (4.9)$$

ここに C_s は、 s 番目の階層において求められる平仮名の種類を元とする単集合である。 P における元の数 $|P|$ で定義すると、候補数は $|P|$ となる。

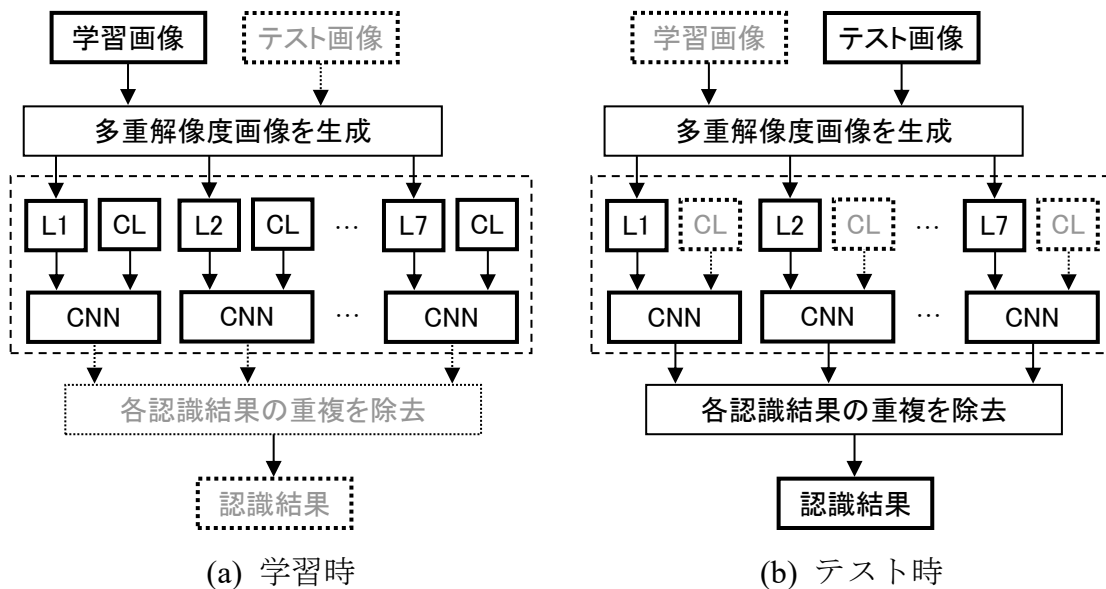


図4.5 全体構成を示す模式図

4.5.2 多重構造の畳み込みニューラルネットワークによる 文字認識法の利点

多重構造の畳み込みニューラルネットワークでは 4.3 節で述べた一般的な畳み込みニューラルネットワークの 2 つの問題が低減されるため適応性が向上する。

まず多重構造の畳み込みニューラルネットワークが処理する多重解像度画像の中には最も認識精度が高い解像度の画像が含まれているので、第 1 の問題が低減される。多重解像度画像の具体的な階層を図 4.6 に示す。

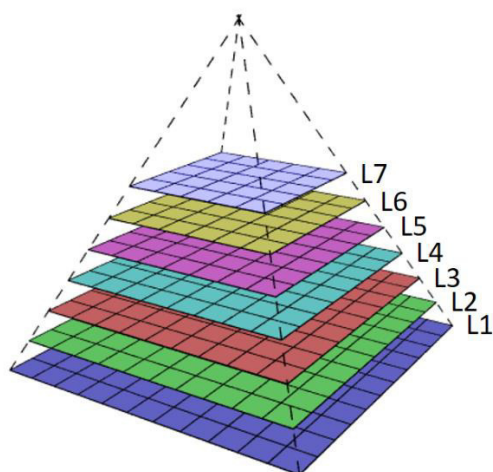


図 4.6 多重解像度画像の具体的な階層

L1 から L7 は画像の階層を表す。L1 の画像は、一般的な CNN の、テスト画像と学習画像に相当し、L2 から L7 の画像は、いずれも L1 の画像を縮小した画像である。図 4.6 に示すように、隣接する階層の大きさが縦横に 1 画素ずつ異なるようにしたのは、オクターブピラミッド⁽³⁾よりも拡大縮小比率を細かく変化させるためである。これによって、多くの解像度特性を認識結果に反映させることができる。提案の CNN では全階層を用いて認識結果を判断する。

さらに、多重解像度画像の各認識結果が統合されることで解像度を調整して候補数を適応的に制御する機能が得られるので第 2 の問題が低減される。具体的には、式(4.9)により各階層からの出力を統合して複数候補を出すことで図 4.3

の劣化強度が強い領域の正答率が向上し、正答率（例えば 90 パーセント）が保証されるようになる。上記の統合を行った場合の、テスト画像の劣化強度と平均候補数との関係を図 4.7 に示す。

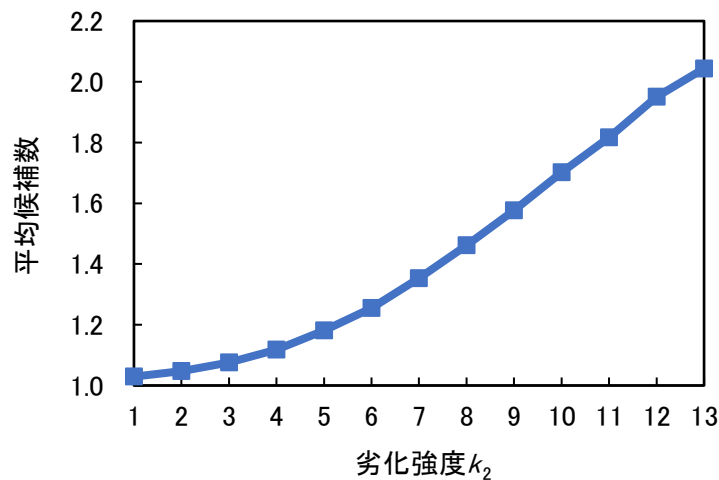


図 4.7 テスト画像の劣化強度と平均候補数との関係

劣化強度が強くなるほど、平均候補数が増加していることを確認できる。平均候補数の増加は、正答率を増加させる効果があるため、劣化強度が強い場合の正答率が向上する。すなわち、劣化強度が強くなって正答率が低下すれば、自動的に候補数が増え、システムとしての信頼性が維持される。

4.5.3 畳み込みニューラルネットワークの実装方法

使用した各 CNN（図 4.5 で CNN と記載したブロック）のネットワーク構造は第 2 章の図 2.12 で示される構造と同じである。各 CNN の諸元を表 4.2 に示す。具体的には、以下の 2 点で LeNet⁽²⁾に改良を行っている。

- 出力計算と勾配計算を高速化するため、活性化関数に ReLU⁽⁹⁾を用いた。
- 汎化能力を高めるために、ドロップアウト⁽¹⁰⁾を導入した。

なおこの CNN は比較的小規模な構造となっていて実用性が高い。

表 4.2 ネットワーク構造の諸元

畳み込み層 1	<ul style="list-style-type: none"> ・フィルタサイズ：横 5×縦 5[画素] ・フィルタ数：32
畳み込み層 2	<ul style="list-style-type: none"> ・フィルタサイズ：横 5×縦 5[画素] ・フィルタ数：80
活性化関数 1 から 3	種類：ReLU
活性化関数 4	種類：ソフトマックス関数
最大値プーリング 1, 2	注目領域：横 2×縦 2[画素]
ドロップアウト 1, 2	ドロップアウト率：0.25
全結合層 1	ノード数：800
全結合層 2	ノード数：42

各 CNN における認識手順を以下に示す。

- (i) 監視カメラ画像の劣化強度を大まかに分類する。実務的には分類数を 3 程度とするのが適当である。そして理想的な環境で撮影したナンバープレート画像に対し、画像劣化モデル⁽⁸⁾を用いて、分類ごとに学習画像を用意する。
 なお、理想的な環境で撮影したナンバープレート画像とは、できるだけ劣化を受けないように、新品のナンバープレートを真正面から近距離でデジタルカメラを用いて高画質で撮影した画像を意味する。
- (ii) テスト画像の劣化強度を大まかに推定する。実務的には劣化要因を区別せず、全体的な劣化強度を、テスト画像のパワースペクトルに基づいて推定する方法が有効である。劣化要因ごとに、より詳細に推定する場合には、ぼけ、ノイズ、画像圧縮について、それぞれ R. Yan らの方法⁽¹¹⁾、C. Tang らの方法⁽¹²⁾、X. Zhang らの方法⁽¹³⁾を用いる。解像度低下による劣化については、学習画像の解像度を、テスト画像の解像度と同じにすればよい。
- (iii) (i)で用意した学習画像のうち、(ii)で推定した劣化強度に最も近い学習画像を選び、CNN を学習する。
- (iv) 学習した CNN でテスト画像を認識する。

4.6 認識性能改善効果の評価実験

定量的評価実験を行って従来の文字認識法（4.2 節の一般的な CNN において 4.4 節の従来の複数候補選択法を用いる認識法）と提案する多重構造の畳み込みニューラルネットワークによる文字認識法⁽⁴⁾を比較した。

文字認識法の指標について、複数候補を選ぶ場合、候補数、正答率、処理時間がトレードオフの関係になるが、本研究の応用では数日間の処理時間を許容できるため処理時間については実務上、考慮する必要がない。よって候補数と正答率の両指標を用いた。

本実験における測定項目を下記に示す。

- (1) 平均候補数と正答率との関係
- (2) 平均候補数と正答率の比との関係

なお測定項目（2）の正答率の比とは、提案の文字認識法の正答率と従来の文字認識法の正答率の最大値との比を意味する。

4.6.1 実験条件

解像度低下、ぼけ、ノイズ、JPEG 圧縮による複合劣化画像に切り出し誤差を与えた画像を各文字認識法の実験用画像（テスト画像と学習画像）とした。本実験では各劣化を計算機上で与えた。具体的にはデジタルカメラ（Nikon 製 D200）により近距離（1.4 メートル）で真正面から新品のナンバープレートを高解像度（横 3,872[画素]、縦 2,592[画素]）で撮影し無圧縮で保存した。以降においてこの無圧縮で保存した画像を実写画像 3 と呼ぶこととする。実写画像 3 の平仮名領域の解像度は平仮名の種類により異なるが、最小の場合でも、横 120 画素、縦 180 画素であった。環境光については、蛍光灯の光のみとすることで、光源環境を固定した。この実写画像 3 に画像劣化モデル⁽⁸⁾に基づき計算機で各劣化を付与した画像を生成し、さらにこの画像に切り出し誤差を与えた画像を実験用画像とした。

なお本実験の画像劣化モデルは 2.6.1 節で述べたとおりである。

テスト画像と学習画像の劣化要因と劣化強度は表 4.1 に示したとおりである。またテスト画像の劣化強度は 4.3 節と同じパラメータ k_2 で表され、その劣化強度の詳細は式(4.1)から式(4.3)に示したとおりである。本実験では劣化強度 k_2 が 4 以上となるものに限定した。4 以上とした理由は、3 以下の場合、提案手法を適用しなくても 90 パーセントの正答率を十分に確保できるからである。

対象とした文字は、数字よりも複雑な形状をもち、種類が多い平仮名である。本実験ではナンバープレートに使用されている 42 種類の平仮名を対象とした。テスト画像の劣化強度ごとの総数と学習画像の総数は、それぞれ 18,900 枚、10,752 枚とした。

テスト画像の例を表 4.3、学習画像の例を表 4.4 に示す。

表 4.3 テスト画像の例

平仮名の種類	さ	す	せ
劣化強度 4 の画像例			
劣化強度 7 の画像例			
劣化強度 10 の画像例			
劣化強度 13 の画像例			

表 4.4 学習画像の例

平仮名の種類	さ	す	せ
平仮名画像の例			

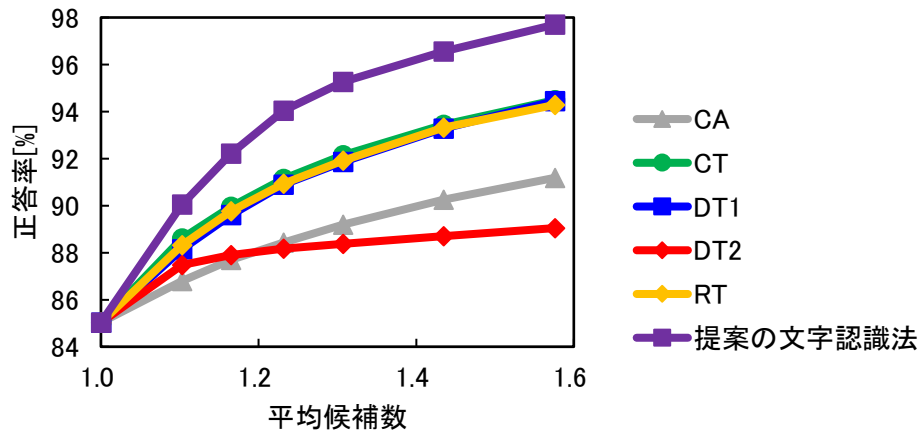
CNN のネットワーク構造については，従来の文字認識法と提案する多重構造の畳み込みニューラルネットワークによる文字認識法の両方において表 4.2 に示す構造とした．多重構造の畳み込みニューラルネットワークによる文字認識法における，全体構成は図 4.5 と同じ構成とし，多重解像度画像の具体的階層は図 4.6 と同じにした．

4.6.2 認識性能改善効果

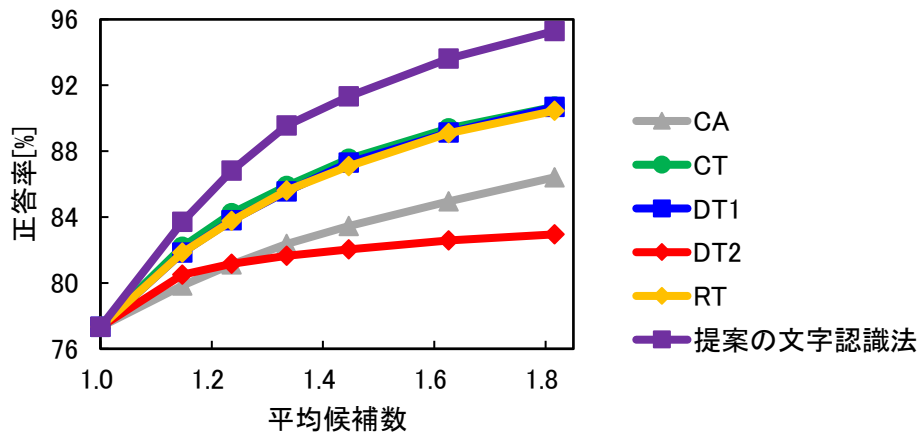
測定項目（1）の測定結果を図 4.8，測定項目（2）の測定結果を図 4.9 に示す．

図 4.8 の(a), (b), (c)は，それぞれ劣化強度 9，劣化強度 11，劣化強度 13 の場合である．図 4.8 の CA, CT, DT1, DT2, RT の表記は，4.4 節で説明した従来の複数候補選択手法⁽⁷⁾を表す．

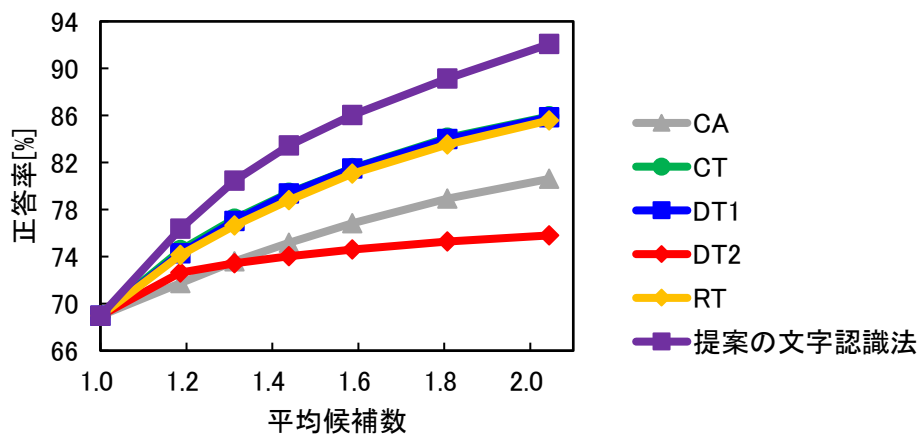
図 4.9 で正答率の比が最大となるのは劣化強度 k_2 が 13 のときで，その比は 1.072 であった．



(a) 劣化強度 9 の場合



(b) 劣化強度 11 の場合



(c) 劣化強度 13 の場合

図 4.8 平均候補数と正答率との関係

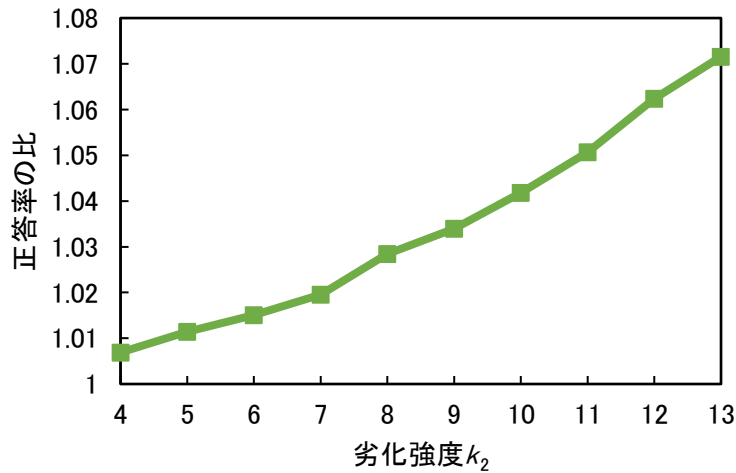


図 4.9 平均候補数と正答率の比との関係

4.6.3 考察

図 4.8 より，いずれの劣化強度においても提案の文字認識法の正答率が従来の文字認識法の正答率よりも高いことがわかる．提案の文字認識法の正答率が高くなるのは，一般的な単一解像度の画像を処理する従来の CNN は，解像度を調整する機能がないため，必ずしも最適ではない解像度の画像を用いて認識を行うのに対し，提案の文字認識法は多重解像度画像を処理することで，最適な解像度の画像から得られる認識結果を，最終的な認識結果に反映できるためであると考えられる．また，いずれの手法を用いた場合でも，平均候補数が大きくなるに従い正答率の傾きが小さくなるのは，平均候補数を増やしていくと，比較的認識しやすいサンプルから順に検出されていくためであると考えられる．すなわち，平均候補数が増えるに従い，認識が難しいサンプルが増えるため，傾きが小さくなったと考えられる．

図 4.9 より，劣化強度が強くなるにつれて正答率の比が大きくなっていくことを確認できる．これは劣化強度が強くなると単一候補での正答率が低下するため候補数を増やした場合における正答率の増分が大きくなり，各手法の差が出やすくなることに起因すると考えられる．

4.7 まとめ

本章では文字の特徴や劣化の状態に柔軟に適応した文字認識を行うため、多重構造の畳み込みニューラルネットワークで多重解像度画像を処理する文字認識法を提案した。一般的な畳み込みニューラルネットワークによる文字認識法の問題を述べ、多重構造の畳み込みニューラルネットワークによる文字認識による文字認識法の利点を述べた。畳み込みニューラルネットワークの実装方法を述べ、最後に認識性能改善効果の評価実験について述べた。以下に得られた結果を要約する。

- (1) 畳み込みニューラルネットワークは文字の特徴や劣化の状態に柔軟に適応して文字認識を行うことに適しているが、一般的な畳み込みニューラルネットワークでは入力画像の解像度を調整する機能がないためこの適応性が不十分である。そこでこの適応性を向上させるため多重構造の畳み込みニューラルネットワークで多重解像度画像を処理する文字認識法を提案した。
- (2) 一般的な畳み込みニューラルネットワークによる文字認識法には 2 つの問題があり、入力画像の解像度を調整して認識精度を最大化する機能がないことが第 1 の問題であり、入力画像の解像度を調整して候補数を適応的に制御し正答率を保証する機能がないことが第 2 の問題であることを明らかにした。
- (3) 提案の文字認識法では、多重解像度画像の中に最も認識精度が高い解像度の画像が含まれているので上記の第 1 の問題が低減され、多重解像度画像の各認識結果が統合されることで解像度を調整して候補数を適応的に制御する機能が得られるので上記の第 2 の問題が低減されることを述べた。そしてこの 2 つの問題が低減される結果として提案の文字認識法の適応性が高まることを述べた。

(4) 提案の文字認識法の有用性を確かめるため複合劣化画像を認識対象とする定量的評価実験を行った。その結果、いずれの劣化強度においても提案の文字認識法の正答率は従来の文字認識法の正答率に比べて高く、提案の文字認識法の正答率は従来の文字認識法の正答率に対して最大で約 7.2 パーセント高くなった。

参考文献

- (1) K. Shinomiya, N. Takamura, K. Fujita and N. Nakamori: “Discriminating car license plate numbers on low resolution using moment characteristics”, *ITE Transactions on Media Technology and Applications*, Vol. 1, No. 4, pp. 271-277, 2013.
- (2) Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R.E. Howard, W. Hubbard and L. D. Jackel: “Backpropagation applied to handwritten zip code recognition”, *Neural Computation*, Vol. 1, No. 4, pp. 541-551, 1989.
- (3) P. Burt and E. Adelson: “The Laplacian Pyramid as a Compact Image Code”, *IEEE Transactions on Communications*, Vol. 31, No. 4, pp. 532-540, 1983.
- (4) 辻広生, 福水洋平, 道関隆国, 山内寛紀: “多重構造とした畳み込みニューラルネットワークによる劣化したナンバープレート画像の平仮名認識”, *信号処理学会誌*, Vol. 22, No. 3, pp. 121-134, 2018.
- (5) D. E. Rumelhart, G. E. Hinton and R. J. Williams: “Learning Representations by Back-propagating Errors”, *Nature*, Vol. 323, No. 9, pp. 533-536, 1986.
- (6) P. Liu, G. H. Li and T. Dan: “Low-quality License Plate Character Recognition Based on CNN”, *Proceedings of 8th International Symposium on Computational Intelligence and Design (ISCID)*, Vol. 2, pp. 53-58, 2015.
- (7) C.-L. Liu and M. Nakagawa: “Precise Candidate Selection for Large Character Set Recognition by Confidence Evaluation”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 6, pp. 636-641, 2000.
- (8) D. Capel: “Image Mosaicing and Super-resolution”, Springer, pp. 86-87, 2004.
- (9) V. Nair and G. E. Hinton: “Rectified Linear Units Improve Restricted Boltzmann Machines”, *Proceedings of the 27th International Conference on International Conference on Machine Learning (ICML)*, pp. 807-814, 2010.
- (10) N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov: “A Simple Way to Prevent Neural Networks From Overfitting”, *Journal of Machine Learning Research*, Vol. 15, No. 1, pp. 1929-1958, 2014.

- (11)R. Yan and L. Shao: “Blind Image Blur Estimation via Deep Learning”, IEEE Transactions on Image Processing, Vol. 25, No. 4, pp. 1910-1921, 2016.
- (12)C. Tang, X. Yang and G. Zhai: “Noise Estimation of Natural Images via Statistical Analysis and Noise Injection”, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 25, No. 8, pp. 1283-1294, 2014.
- (13)X. Zhang, W. Lin and J. Liu: “Compression Noise Estimation and Reduction via Patch Clustering”, Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA), pp. 715-718, 2015.

第5章 評価

5.1 まえがき

街頭や店頭に設置された一般的な監視カメラが撮影したナンバープレート複合劣化画像の文字を視覚的に見やすくし、正確に読み取れるようにするためには、ロバスト性、効率性、適応性が高い文字読み取りシステムを構築する必要がある。そこで本論文で提案した手法の、ロバスト性、効率性、適応性に対する評価実験を行った。そして本論文で提案した手法の実用性を確認するため、実環境での検証実験を行った。

2節でロバスト性の評価実験、3節で効率性の評価実験、4節で適応性の評価実験について述べる。次に5節で実環境での検証実験について述べる。最後に6節で全実験を総括し、考察を述べる。

5.2 ロバスト性の評価実験

第2章で提案した Retinex モデルにおけるサポートベクター回帰を用いた輝度値補正法⁽¹⁾を適用することで、文字読み取りシステムの、環境光の外乱に対するロバスト性が向上するかどうかを定量的に評価した。

5.2.1 実験条件

実験用画像は環境光の外乱と、解像度低下、ぼけ、ノイズ、JPEG 圧縮の各劣化による複合劣化画像とした。実環境画像において環境光の外乱はナンバー灯に起因し画像の違いによる変動は比較的小さいため、条件を固定した。一方、上記の劣化の中には劣化強度の変動が大きいものがあり、これらについては変動させた。本実験は定量的評価実験であるため劣化強度を正確に定量化する必要がある。しかし、実環境画像を使用した場合、劣化強度を定量化することができない。

第1に、実環境のナンバープレートには汚損があり、汚損が画像の劣化と混合した場合の影響が不明である。よって、実環境のナンバープレートに対して設定した劣化強度は正確性を欠く。

第2に、実環境画像は、複数の劣化要因により複合的に劣化しているが、一般的に各劣化強度の推定は不良設定問題⁽²⁾に該当するため、各劣化強度を正確に推定することはできない。

以上を考慮し、理想的な環境でナンバー灯に照らされた新品の実験用ナンバープレートをデジタルカメラ（Nikon 製 D200）で撮影し、無圧縮で保存した画像（第2章の実写画像1）に対し、正確な定量化が可能な画像劣化モデル⁽³⁾に基づき計算機上で各劣化を付与して実験用画像を生成することとした。

なお本実験の画像劣化モデルについては2.6.1節で述べたとおりである。

以降で実験用画像の生成方法について具体的に説明する。

暗室において新品のナンバープレートにナンバー灯（小糸製作所製自動車用電球 12V 5W）を設置し、ナンバー灯を点灯してナンバープレートを撮影した。ナンバー灯は左右に1つずつ、ナンバープレートに対して対称な位置に設置した。使用したナンバープレートの諸元を表5.1、ナンバー灯の設置状況を図5.1、ナンバー灯設置位置の説明図を図5.2に示す。左側ナンバー灯の設置箇所を図5.2の×印 d とし、×印 d の位置をナンバープレート左端からの距離（図5.2 距離 ab ）、ナンバープレートからの水平距離（図5.2 距離 bc ）、ナンバープレートからの垂直距離（図5.2 距離 cd ）で定義した。定量的評価と定性的評価における左側ナンバー灯の設置位置を表5.2に示す。

表 5.1 ナンバープレートの諸元

大きさ[cm]	横 33.0, 縦 16.5 (中型番号標規格)
枚数[枚]	10 (各数字 1)
一連指定番号	・000, 1111, 2222, 3333, 4444, 5555, 6666, 7777, 8888, 9999



図 5.1 ナンバー灯の設置状況

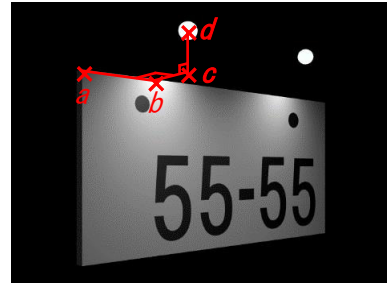


図 5.2 ナンバー灯設置位置の説明図

表 5.2 ナンバー灯の設置位置

左側 ナンバー灯	距離 ab [cm]	8.5
	距離 bc [cm]	4.5
	距離 cd [cm]	1.0

上記撮影時におけるナンバープレートの中心から監視カメラのレンズ中心までの距離は、図 5.2 の bc 方向に 1.7 メートルであり、レンズ中心から見てナンバープレート中心は俯角 0 度に位置し、レンズ中心とナンバープレート中心を結ぶ直線と図 5.2 の bc 方向の方位線との間の方位角は 0 度である。以上の条件で撮影し、無圧縮で保存した画像（実写画像 1）の全体の画素数は横 1,024[画素]、縦 685[画素]であり、ナンバープレート領域の画素数は、横 483[画素]、縦 227[画素]であった。さらに実写画像 1 における照明条件が現実世界における照明条件を模倣できていることを確認するため、実写画像 1、実環境の監視カメラ画像（以下、実環境画像）、ナンバー灯が点灯していない状態の監視カメラ画像（以下、参照画像）について、ナンバープレート領域における輝度値の標準偏差を計算した。計算結果を表 5.3 に示す。

表 5.3 輝度値の標準偏差

	標準偏差
実写画像 1	0.15
実環境画像	0.14
参照画像	0.071

参照画像は、実環境画像と同一ナンバープレートを、昼間においてナンバープレート全体がほぼ均一の明るさであると考えられる条件下で、実環境画像と同一のカメラにより撮影したものである。また、上記 3 枚の画像における比較条件を同一とするため、3 枚の画像はいずれも横 40[画素]、縦 20[画素]となるように拡大縮小した。ナンバープレートの文字領域は照明条件の影響が低下するため、標準偏差は文字領域を除外して計算した。表 5.3 より実写画像 1 の標準偏差と、実環境画像の標準偏差の差は比較的小さく、実写画像 1 は実環境画像に近い照明条件を再現できていることを確認できる。さらに実写画像 1 における照明条件が現実世界の監視カメラ画像における照明条件を模倣できていることを視覚的に確認できるよう画像の大きさを調整した後の両画像を図 5.3 に並べて示す。



(a) 実写画像 1



(b) 実環境画像

図 5.3 実写画像 1 と実環境画像

実写画像 1 をグレースケール画像に変換した後，変換後の画像に計算機で解像度低下，ぼけ，ノイズ，JPEG 圧縮の劣化を付与した。

まずグレースケール画像を横 40[画素]，縦 20[画素]に縮小することにより解像度を低下させた。この画素数は，標準画質（SD 画質）の監視カメラ画像において，頻度が高い値に相当し，この値は画角が約 60 度のカメラで，カメラから 5 メートル離れた位置にあるナンバープレートに SD 画質で撮影したときの値に相当する。

ガウスフィルタによるぼけの標準偏差 σ_{b7} ，ガウスノイズの標準偏差 σ_{n7} ，JPEG 圧縮の品質 Q_7 についてはそれぞれ式(5.1)から式(5.3)で表される値を用いた。

$$\sigma_{b7} = 0.80 + 0.08 k_3 \quad (5.1)$$

$$\sigma_{n7} = 0.10 + 0.09 k_3 \quad (5.2)$$

$$Q_7 = 100 - 10 k_3 \quad (5.3)$$

ここに劣化強度は k_3 で定義される。 k_3 は 0 から 10 まで 11 段階あり，値が大きいほど劣化が強い。パラメータ k_3 で表される各劣化強度の範囲を表 5.4 に示す。

表 5.4 パラメータ k_3 で表される各劣化強度の範囲

ガウスフィルタによるぼけ (標準偏差 σ_{b7} [画素])	$0.80 \leq \sigma_{b7} \leq 1.60$
ガウスノイズ (標準偏差 σ_{n7})	$0.10 \leq \sigma_{n7} \leq 1.00$
JPEG 圧縮 (品質 Q_7)	$0 \leq Q_7 \leq 100$

ガウスノイズについては，不規則性があるため，同じ劣化強度でもシードごとに認識精度が異なる。よって劣化強度ごとにシードを変えて 10 枚生成し，不規則性が認識精度に与える影響を軽減した。

以上の方法で下記の認識対象画像をそれぞれ 1,100 枚生成した。

- ・対象画像 1：ガウスフィルタによるぼけ，ガウスノイズ，JPEG 圧縮が重畳した画像（全劣化が重畳した画像）
- ・対象画像 2：ガウスフィルタによるぼけのみの画像
- ・対象画像 3：ガウスノイズのみの画像
- ・対象画像 4：JPEG 圧縮のみの画像

従来 Retinex については，SSR 処理と 2 スケールの MSR（以下，2MSR）処理を行った．ここに 2 スケールとは，2.3.2 節の式(2.5)において $N=2$ であることを表す．

提案 Retinex については，式(2.20)において式(5.4)で表されるガウスカーネルを用いた．

$$K(\mathbf{x}, \mathbf{x}_i) = \exp(-\gamma \|\mathbf{x} - \mathbf{x}_i\|^2) \quad (5.4)$$

2MSR における式(2.5)の重み w_n は，式(5.5)により定義した．

$$w_n = \frac{1}{N} \quad (n = 1, 2, \dots, N) \quad (5.5)$$

従来 Retinex における w_n 以外のパラメータと提案 Retinex におけるパラメータについては，数字認識の正答率が最大となる値をグリッドサーチで求めることにより設定した．グリッドサーチのパラメータ探索範囲を表 5.5 に示す．

表 5.5 グリッドサーチのパラメータ探索範囲

従来 Retinex	式(2.3)における σ もしくは式(2.5)の R_n における σ_n の範囲[画素]	[2, 4, 6, 8, 10]
提案 Retinex	式(2.3)における σ の範囲[画素]	[2, 4, 6, 8, 10]
	式(5.4)における γ の範囲[画素]	[0.2, 3.2, 6.2, 9.2, 12.2]
	式(2.11)における ε の範囲	[0.002, 0.275, 0.547, 0.820]
	式(2.15)における C の範囲	[0.003, 0.019, 0.034, 0.050]

5.2.2 評価方法

従来の Retinex 処理⁽⁴⁾の画像に対する数字認識の正答率とサポートベクター回帰を用いた Retinex 処理⁽¹⁾の画像に対する数字認識の正答率を比較した。数字認識は認識対象画像から切り出した数字画像に対して行った。一般的な数字領域の切り出し手順を図 5.4 に示す。この手順ではまず視認性が高く検出しやすいナンバープレート 4 頂点を検出する。そして射影歪みがあるときは縦横比が 2 であるというナンバープレートの規格に基づきナンバープレート画像を正面から見た画像に射影変換する。この射影変換されたナンバープレートではナンバープレートの規格により 4 頂点と数字領域の位置関係は定まっているため、この規格に基づいて数字領域を切り出す。射影歪みがない場合には正規化する必要がないので射影変換を行わずに 4 頂点と規格により定まっている上記の位置関係に基づき数字領域を切り出す。本実験の対象画像はいずれも射影歪みがないため、図 5.4 の分岐では NO の経路を辿る。

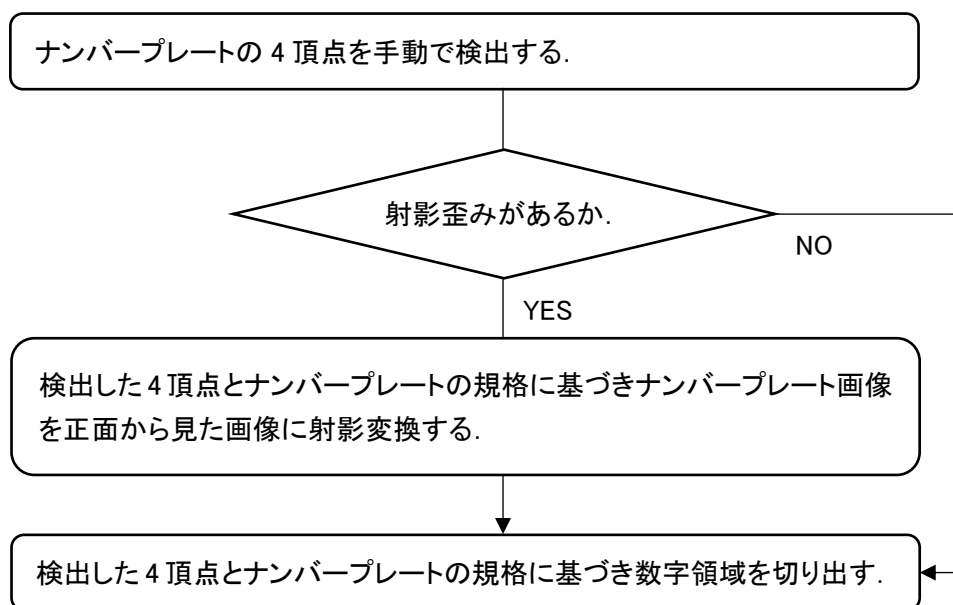


図 5.4 一般的な数字領域の切り出し手順

数字認識には定量的評価に適する正規化相互相関（以下、ZNCC）を用いた。ZNCC による数字認識の手順を以下に示す。

- (i) 対象とする数字画像の，縦横の画素数（横 U [画素]，縦 V [画素]）を計算する．
- (ii) 大きさが横 20 [画素]，縦 40 [画素]であるナンバープレート数字のテンプレート画像を，横 U [画素]，縦 V [画素]に縮小する．図 5.5 にテンプレート画像と縮小した画像の例を示す．
- (iii) 対象とする数字画像と縮小したテンプレート画像に対して，式(5.6)で定義される正規化相互相関係数 R_{ZNCC} を算出し， R_{ZNCC} が最も大きくなる数字を認識結果とする．



(a) テンプレート画像

(b) 縮小した画像

図 5.5 テンプレート画像と縮小した画像の例

$$R_{ZNCC} = \frac{\sum_{j=0}^{V-1} \sum_{i=0}^{U-1} (I(i, j) - \bar{I})(T(i, j) - \bar{T})}{\sqrt{\sum_{j=0}^{V-1} \sum_{i=0}^{U-1} (I(i, j) - \bar{I})^2 \sum_{j=0}^{V-1} \sum_{i=0}^{U-1} (T(i, j) - \bar{T})^2}} \quad (5.6)$$

ただし，式(5.6)において， \bar{I} と \bar{T} は，領域内における画素値の平均値である． \bar{I} と \bar{T} をそれぞれ式(5.7)と式(5.8)に示す．

$$\bar{I} = \frac{1}{UV} \sum_{j=0}^{V-1} \sum_{i=0}^{U-1} I(i, j) \quad (5.7)$$

$$\bar{T} = \frac{1}{UV} \sum_{j=0}^{V-1} \sum_{i=0}^{U-1} T(i, j) \quad (5.8)$$

5.2.3 外乱の除去効果

全劣化が重畳した画像（対象画像 1）における正答率を図 5.6，ガウスフィル

タによるぼけのみの画像（対象画像 2）における正答率を図 5.7，ガウスノイズのみの画像（対象画像 3）における正答率を図 5.8，JPEG 圧縮のみの画像（対象画像 4）における正答率を図 5.9 に示す。

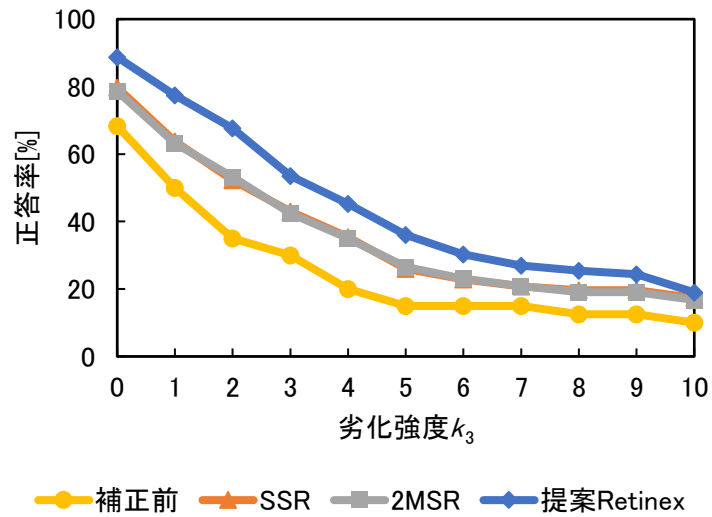


図 5.6 対象画像 1 における正答率

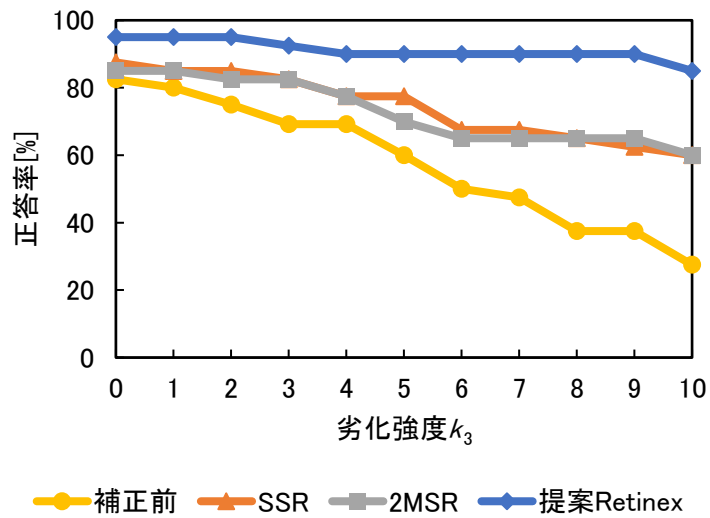


図 5.7 対象画像 2 における正答率

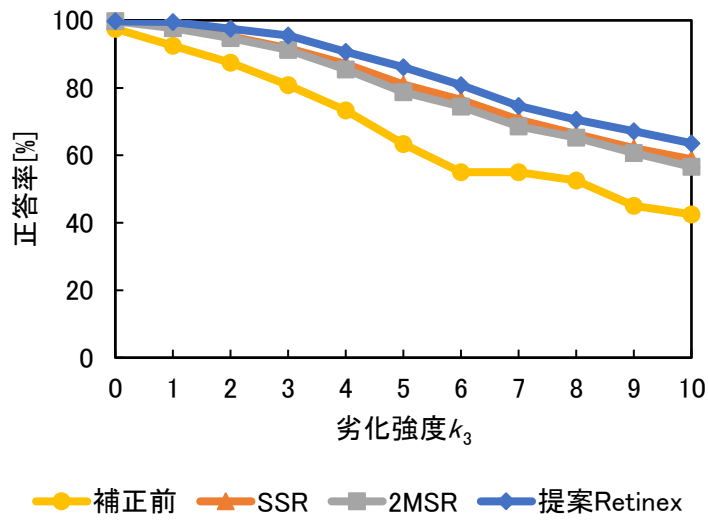


図 5.8 対象画像 3 における正答率

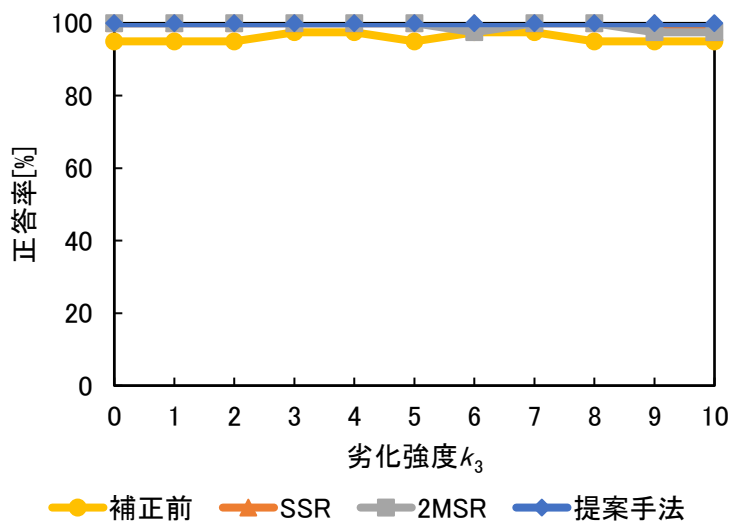


図 5.9 対象画像 4 における正答率

なお、図 5.6 から図 5.9 における凡例の説明を下記に示す。

- 補正前：輝度値補正前の対象画像
- SSR：single scale retinex 処理後の対象画像
- 2MSR：2 スケールの MSR 処理後の対象画像
- 提案 Retinex：提案 Retinex を行った後の対象画像

ここに2スケールとは、2.3.2節の式(2.5)において $N = 2$ であることを表す。

提案 Retinex における正答率の SSR における正答率に対する比について、対象画像ごとに全劣化強度の最大値、平均値、最小値を算出したものを表 5.6 に示す。

表 5.6 正答率の比

	対象画像 1	対象画像 2	対象画像 3	対象画像 4
最大値	1.39	1.44	1.08	1.03
平均値	1.25	1.24	1.05	1.00
最小値	1.10	1.09	1.00	1.00

なお、図 5.6 から図 5.9 において2つの従来 Retinex (SSR と 2MSR) における正答率を比較すると、両者にはほぼ差がないか、SSR が僅かに優位と認められたため、表 5.6 における提案 Retinex の比較対象として SSR を選んだ。

5.2.4 考察

図 5.6 より、対象画像 1 の全劣化強度において提案 Retinex の正答率は SSR と 2MSR の正答率よりも高い。対象画像 1 は全劣化が重畳した画像であることから、実環境画像と同様の複合劣化画像において、提案 Retinex の外乱除去効果は、SSR と 2MSR の外乱除去効果よりも高いといえる。表 5.6 より、提案 Retinex の正答率は、SSR の正答率に比べて約 25 パーセント改善した。また、図 5.6 の全劣化強度において2つの従来 Retinex (SSR と 2MSR) における正答率はほぼ同じ値であることから、ZNCC を精度の指標に用いる場合には、従来 Retinex においてガウス関数の周辺領域を複数設定しても外乱除去効果は向上しないといえ、全劣化が重畳した画像において、提案 Retinex の外乱除去効果は、いずれの従来 Retinex の外乱除去効果よりも高いと考えられる。

図 5.6 において、SSR における正答率と 2MSR における正答率がほぼ同じ値

になった原因には Halo 作用 が関わっていると考えられる。SSR において顕著に出現する Halo 作用は本来好ましくない効果であるが、Halo 作用が数字と背景の境界付近の画像特徴量を強調する場合がある。一方で一般的に SSR は 2MSR よりも外乱除去効果が低い。SSR における Halo 作用による画像特徴量の強調がもたらす SSR の 2MSR に対する優位が SSR は 2MSR よりも外乱除去効果が低いという SSR の 2MSR に対する劣位と相殺した結果、SSR における正答率と 2MSR における正答率がほぼ同じ値になったと考えられる。

図 5.7 と図 5.8 より、対象画像 2 と対象画像 3 の全劣化強度において、提案 Retinex の正答率は SSR と 2MSR の正答率よりも高い。したがって、ぼけのみの画像とノイズのみの画像においても、提案 Retinex の外乱除去効果は SSR と 2MSR の外乱除去効果よりも高いといえる。対象画像 2 と対象画像 3 の従来 Retinex におけるガウス関数の周辺領域数についても、対象画像 1 の場合と同様の考察が成り立ち、ぼけのみの画像とノイズのみの画像において、提案 Retinex の外乱除去効果は、いずれの従来 Retinex の外乱除去効果よりも高いと考えられる。

表 5.6 において、対象画像 2 における正答率の比と対象画像 3 における正答率の比とを比較すると、最大値、平均値、最小値のいずれにおいても対象画像 2 における正答率の比が対象画像 3 における正答率の比よりも高い。また、図 5.7 より、ぼけの劣化強度が強くなるに従い、提案 Retinex における正答率と従来 Retinex における正答率の差が大きくなっている。したがって、提案 Retinex は劣化要因としてノイズが支配的である画像よりもぼけが支配的である画像において効果的であると考えられる。ぼけが支配的である画像において、従来 Retinex の効果が比較的低い原因としては、従来 Retinex の場合、対象画像をぼかすことにより照明画像を生成するため、対象画像においてぼけが支配的である場合、照明画像と対象画像の差が小さくなり、輝度値補正後の画像において数字の特徴量が不明瞭になることが考えられる。

図 5.9 より、対象画像 4 の全劣化強度において、補正前の画像でも 90 パーセントを超える正答率が得られており、JPEG 圧縮が正答率に与える影響は比較的小さいといえる。また図 5.9 を用いて、提案 Retinex もしくは従来 Retinex に

よる補正後の画像における正答率と補正前の画像における正答率とを比較すると、補正後の画像における正答率は、補正前の画像における正答率よりも向上しており、輝度値補正の効果を確認できる。しかし、図 5.9 において提案 Retinex の外乱除去効果が従来 Retinex の外乱除去効果よりも高いかどうかについては、両者の正答率がいずれも 100 パーセント近くの上限值に達しているため不明である。

以上より、実環境画像と同様の複合劣化画像を対象として ZNCC による文字認識の正答率を指標とする定量的評価実験を行った結果、提案 Retinex における外乱除去効果は、いずれの従来 Retinex における外乱除去効果よりも高く、また提案 Retinex は特にぼけが支配的である画像において効果的であると考えられる。

5.3 効率性の評価実験

第 3 章で提案したサポートベクターマシンを用いた画素値選択型マルチフレーム超解像処理⁽⁵⁾を適用することで、文字読み取りシステムの、文字情報が少ない場合に対する効率性が高まるかどうかを定量的に評価した。

5.3.1 実験条件

3.4.1 節の実験用画像と同じ画像を用いた。具体的には、これらの画像は解像度低下、ぼけ、ノイズ、JPEG 圧縮の各劣化による複合劣化画像である。本実験では定量的な評価を行うため劣化強度を正確に定量化する必要があることから、5.2.1 節で述べた理由により理想的な環境で撮影したナンバープレート画像に対し正確な定量化が可能な画像劣化モデル⁽³⁾に基づき計算機上で各劣化を付与して実験用画像を生成することとした。本実験では、2 節の実験とは異なりナンバー灯による環境光の外乱を考慮する必要がないため、上記の理想的な環境で撮影したナンバープレート画像として、新品のナンバープレートをスキャナで撮影し無圧縮で保存した画像を用いた。定量的評価実験における超解像処理の対象画像の枚数と、サポートベクターマシンの学習画像の枚数は、いずれも 400 枚とした。

従来超解像と提案超解像の各処理は 3.3.1 節の図 3.4 と同じ手順で行った。

従来超解像での位置合わせは、3.3.4 節の図 3.6 に示した擬似的高解像度画像を用いる方法により行った。なぜなら 3.3.4 節で述べたとおり、複合劣化画像では、一致度評価関数とフィッティング関数との不一致が大きいため、フィッティング関数を用いた方法では位置合わせ精度が低下するからである。一致度評価値には、照明変動に強いことから⁽⁷⁾、本研究の応用に適合する ZNCC を用いた。

図 3.4 の従来超解像と提案超解像でのコスト関数の最小化はいずれも 3.2.3 節で述べた方法と同じ方法で行った。すなわちコスト関数を最尤推定の形で定式化し、これを反復法で最小化する方法 (Irani と Peleg の方法⁽⁸⁾) を用いた。

5.3.2 評価方法

以下の 2 つの評価項目について従来超解像⁽⁶⁾の画像と提案超解像の画像を比較した。

- (1) 数字認識の正答率
- (2) サブピクセル位置合わせの平均誤差

文字認識には正規化相互相関 (以下、ZNCC) を用いた。ZNCC による数字認識の手順は 5.2.2 節に示したとおりである。サブピクセル位置合わせの指標については正しい位置からのユークリッド距離を用いた。

5.3.3 文字情報の再構成効果

数字認識の正答率 (評価項目 (1)), サブピクセル位置合わせの平均誤差 (評価項目 (2)) の結果をそれぞれ図 5.10 と図 5.11 に示す。さらに評価項目 (1), (2) の結果を表 5.7 にまとめる。

なお、図 5.10 と図 5.11 における提案超解像 (画素値選択なし) の表記は、サポートベクターマシンによるサブピクセル位置合わせのみを行い、画素値選択は行っていない超解像画像を意味する。

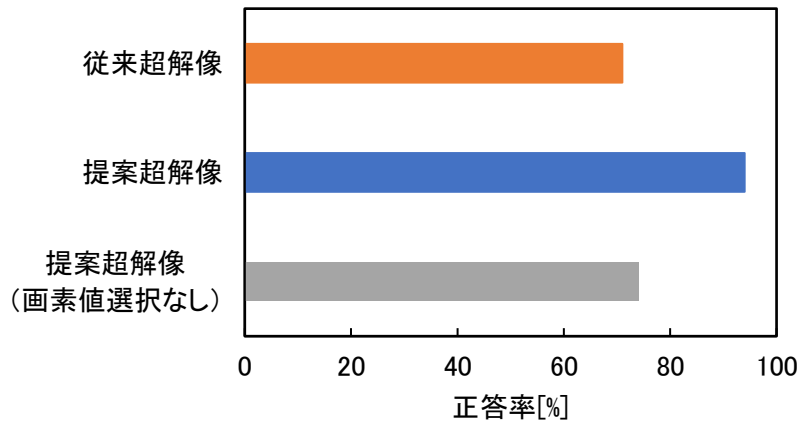


図 5.10 数字認識の正答率

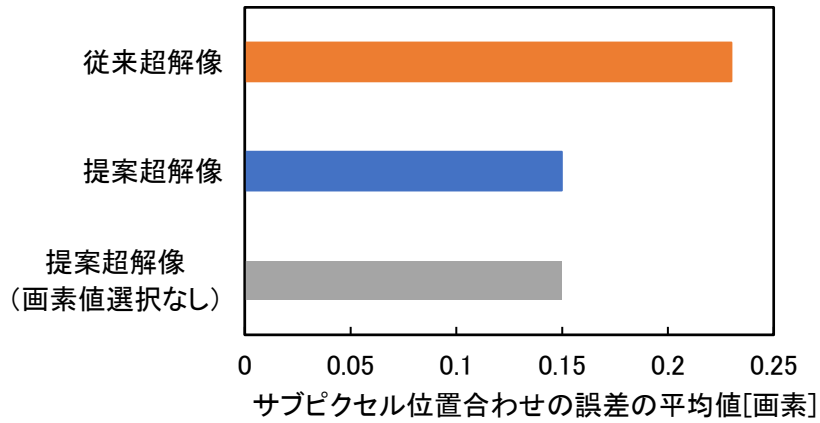


図 5.11 サブピクセル位置合わせの誤差の平均値

表 5.7 数字認識の正答率と、サブピクセル位置合わせの誤差の平均値

	正答率[%]	サブピクセル位置合わせの誤差の平均値[画素]
従来超解像 ⁽⁶⁾	71 (1.00)	0.23 (1.00)
提案超解像	94 (1.32)	0.15 (0.65)
提案超解像 (画素値選択なし)	74 (1.04)	0.15 (0.65)

注：() 内の数値は従来超解像に対する比を示す。

5.3.4 考察

図 5.10 と表 5.7 より提案超解像の正答率は従来超解像の正答率に比べて約 32 パーセント向上している。また、提案超解像（画素値選択なし）の正答率は、従来超解像の正答率に比べて向上しているが、その比率は 4 パーセントであり、比較的小さい。

図 5.11, 表 5.7 より提案超解像のサブピクセル位置合わせの平均誤差は、従来超解像のサブピクセル位置合わせの平均誤差の約 65 パーセントに低減している。

以上より、以下の 3 点がわかる。

- ・ サポートベクターマシンによるサブピクセル位置合わせと画素値選択の両方を行うと高い文字情報の再構成効果が得られる。
- ・ サポートベクターマシンによるサブピクセル位置合わせで大幅に位置合わせ誤差が低減する。
- ・ 提案超解像（画素値選択なし）では比較的低い文字情報の再構成効果しか得られない。

これらの 3 点を総合すると、提案超解像における文字情報の再構成効果は、サポートベクターマシンによるサブピクセル位置合わせと画素値選択の相乗効果により高まっていると考えられる。

5.4 適応性の評価実験

第 4 章で提案した多重構造の畳み込みニューラルネットワークによる文字認識法⁽⁹⁾を適用することで、文字読み取りシステムの、文字の特徴や劣化の状態に対する適応性が高まるかどうかを定量的に評価した。

5.4.1 実験条件

4.6.1 節の実験用画像と同じ画像を用いた。具体的には、これらの画像は解像度低下、ぼけ、ノイズ、JPEG 圧縮の各劣化による複合劣化画像に切り出し誤差

を与えた画像である。これらの画像は定量的評価に用いられるため劣化強度を正確に定量化する必要があることから、5.2.1 節で述べた理由により理想的な環境で撮影したナンバープレート画像に対し正確な定量化が可能な画像劣化モデル⁽³⁾に基づき計算機上で各劣化を付与して実験用画像を生成することとした。一般的に、畳み込みニューラルネットワークで抽出される特徴を定量的に明らかにすることは難しい。よって本実験では学習画像の特徴が実環境画像の特徴に近づくよう、撮像過程で画像に付与される特徴も考慮するため、5.3 節の実験のように上記の理想的な環境で撮影したナンバープレート画像として、新品のナンバープレートをスキャナで撮影して無圧縮で保存した画像を用いるのではなく、新品のナンバープレートをデジタルカメラで撮影し無圧縮で保存した画像を用いた。

5.4.2 評価方法

複合劣化画像を対象として従来の文字認識法（4.2 節の一般的な CNN⁽¹⁰⁾において 4.4 節の従来の複数候補選択法⁽¹¹⁾を用いる認識法）と提案の文字認識法（多重構造の畳み込みニューラルネットワークによる文字認識法⁽⁹⁾）を比較した。従来の文字認識法と多重構造の畳み込みニューラルネットワークによる文字認識法の両方において、CNN のネットワーク構造は表 4.2 に示す構造とした。多重構造の畳み込みニューラルネットワークによる文字認識法における、全体構成は図 4.5 と同じ構成とし、多重解像度画像の具体的階層は図 4.6 と同じにした。

本実験でも 4.6 節冒頭で述べた理由により候補数と正答率が評価指標となる。本実験では各手法が実用レベルの認識性能を満たすかどうかを明らかにするため、正答率を 90 パーセント（実用レベルの正答率の下限值）に固定し、各手法の、劣化強度と平均候補数の関係を測定した。

5.4.3 複合劣化に対する適応性能

正答率を 90 パーセント（実用レベルの正答率の下限值）に固定し、各手法

の、劣化強度と平均候補数の関係を測定した結果を図 5.12 に示す。図 5.12 の CA, CT, DT1, DT2, RT は、4.4 節で述べた従来の複数候補選択手法を表す。

図 5.12 でピンク色の線より下側の領域が実用レベルの認識性能を満たす領域（平均候補数が 2 以下で正答率が 90 パーセントの領域）である。

さらに図 5.12 の結果に基づき、単一候補数当たりの正答率を算出した結果を表 5.8 にまとめる。

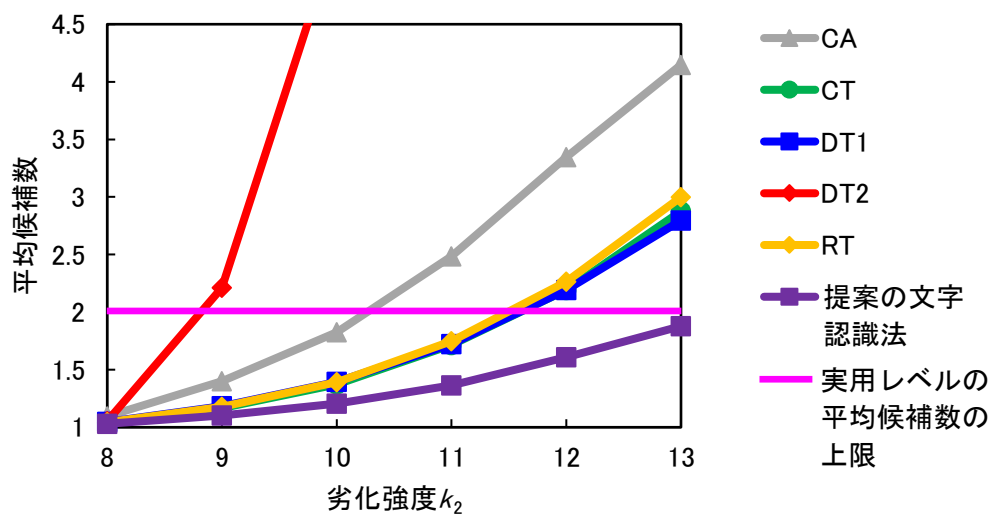


図 5.12 テスト画像の劣化強度と平均候補数との関係

表 5.8 単一候補数当たりの正答率

	$k_2 = 8$	$k_2 = 9$	$k_2 = 10$	$k_2 = 11$	$k_2 = 12$	$k_2 = 13$	平均値
CA	82.2	64.3	49.3	36.2	26.9	21.7	46.8
CT	86.6	77.2	65.6	52.5	41.1	31.3	59.0
DT1	86.0	76.1	64.6	52.3	41.1	32.2	58.7
DT2	85.5	40.7	17.0	10.1	7.4	6.0	27.8
RT	86.2	76.4	64.7	51.5	39.7	30.0	58.1
提案の文字認識法	87.4	81.7	74.7	66.0	55.9	47.9	68.9

5.4.4 考察

図 5.12 より，提案の文字認識法を用いた場合のみ，全劣化強度において実用レベルの認識性能を満足できており，提案の文字認識法は複合劣化に対する適応性能が高いことが分かる．一方，従来の文字認識法の場合には，劣化強度が強い場合に，実用レベルの認識性能を満足できておらず，複合劣化に対する適応性能が低いといえる．

5.5 実環境での検証実験

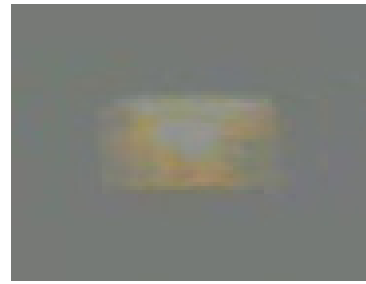
本研究ではロバスト性，効率性，適応性を課題としている．これらの課題のうち，効率性の課題では限定された複合劣化画像を対象としたのに対し，ロバスト性と適応性の課題では多様な複合劣化画像を対象としたので，ロバスト性と適応性の課題については実用性の観点から実環境での検証実験が必要不可欠である．そこでこれらの課題については実環境での検証実験を行った．

5.5.1 実験条件

ロバスト性について検証するためナンバー灯による環境光の外乱が加わった実環境のナンバープレート画像を取得した．具体的には実際の車両に装着され，ナンバー灯に照らされた 2 枚のナンバープレートを，ウェブカメラ Logicool 製 c270(以下，カメラ番号 1 のカメラ)とウェブカメラ ELECOM 製 UCAM-C0113FE(以下，カメラ番号 2 のカメラ)を用いて撮影することにより 2 枚の実環境画像を取得した．以降においてカメラ番号 1 のカメラをカメラ 1，カメラ番号 2 のカメラをカメラ 2 と表記する．これらの実環境画像を図 5.13 に示す．以降において図 5.13(a)の実環境画像の画像番号を 1，図 5.13(b)の実環境画像の画像番号を 2 とし，画像番号 1 の実環境画像を実環境画像 1，画像番号 2 の実環境画像を実環境画像 2 と表記する．実環境画像 1 はナンバー灯が 2 個装着されたナンバープレートの画像であり，実環境画像 2 はナンバー灯が 1 個装着されたナンバープレートの画像である．実環境画像 2 では撮影時に露光が適切に調整されなかったため，コントラストが低下している．



(a) 実環境画像 1



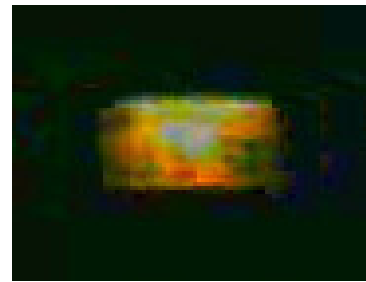
(b) 実環境画像 2

図 5.13 環境光の外乱が加わった実環境画像

実環境画像 1 と実環境画像 2 において環境光に外乱が生じていることを確認するため、各画像のナンバープレート領域における輝度値の最小値が 0，最大値が 255 となるようにヒストグラムを拡張した。ヒストグラム拡張後の画像を図 5.14 に示す。



(a) 実環境画像 1



(b) 実環境画像 2

図 5.14 ヒストグラム拡張後の実環境画像

ヒストグラム拡張後の、実環境画像 1 と実環境画像 2 について数字領域における輝度値の標準偏差の平均値を計算し結果、実環境画像 1 の場合、0.15 であり、実環境画像 2 の場合、0.12 であった。表 5.3 よりナンバー灯が点灯していない状態の監視カメラ画像の場合の標準偏差が 0.071 であることを考慮すると、実環境画像 1 と実環境画像 2 はいずれもナンバー灯により環境光に外乱が生じているといえる。

適応性について検証するため、多様な劣化を受けた、ナンバープレート平仮名の実環境画像を取得した。具体的には実際の車両に装着された 3 枚のナンバ

ナンバープレートを、カメラ1とカメラ2を用いて2通りの撮影角度から撮影することにより12枚の実環境画像を取得した。これらの実環境画像を図5.15に示す。

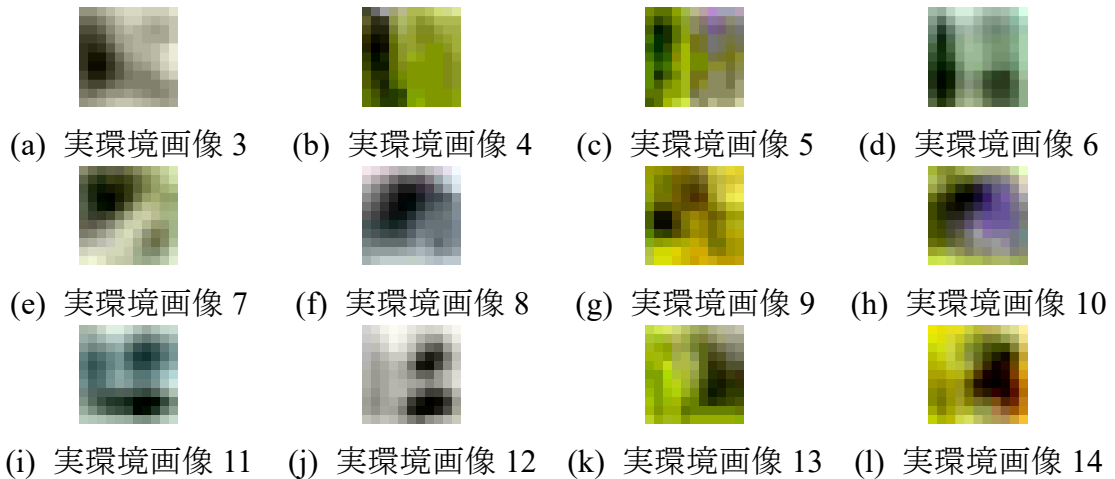


図 5.15 多様な劣化を受けた，ナンバープレート平仮名の実環境画像

以降において図5.15の(a)から(l)の実環境画像の画像番号を順に3から14とし、またこれらの画像を順に実環境画像3から実環境画像14と表記する。図5.15の全画像はナンバープレートの実環境画像から平仮名領域を手動で切り出すことにより取得した画像であり、解像度についてはいずれも横11画素、縦11画素である。実環境画像3から実環境画像14は夜間において、街灯と、道路を走行する車両のライトとに照らされたナンバープレートを撮影したものであり、カメラのホワイトバランス調整状態により本来の色を再現できていない画像がある。図5.15のカメラ2で撮影された画像は、図5.13のカメラ2で撮影された画像とは異なり、撮影時に露光が適切に調整されたため、コントラストが高い。

なお、ウェブカメラも監視カメラとして使用されることがあるため、ウェブカメラを使用することは本実験目的に適合する⁽¹⁾。

上記の実環境画像撮影時における、ナンバープレートとカメラの位置関係を表す距離と角度の定義を図5.16、各実環境画像の諸元を表5.9に示す。図5.16(a)はナンバープレートを真上方向から見た図（平面図）であり、図5.16(b)はナンバープレートを左方向から見た図（立面図）である。

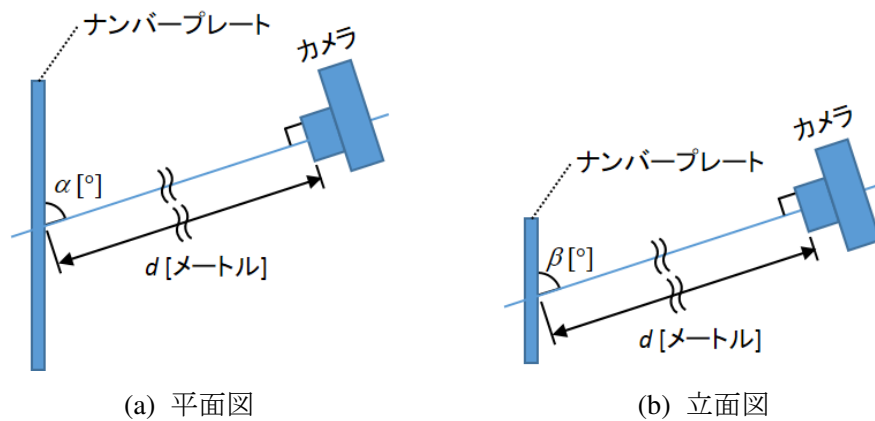


図 5.16 距離と角度の定義

表 5.9 各実環境画像の諸元

画像番号	対象文字	カメラ番号	d [メートル]	α [°]	β [°]
1	数字 (2, 5, 8)	1	6	0	5
2	数字 (0, 1, 5)	2	6	0	9
3	平仮名 (こ)	1	5	72	82
4	平仮名 (こ)	1	5	99	82
5	平仮名 (こ)	2	5	72	82
6	平仮名 (こ)	2	5	99	82
7	平仮名 (め)	1	5	72	82
8	平仮名 (め)	1	5	99	82
9	平仮名 (め)	2	5	72	82
10	平仮名 (め)	2	5	99	82
11	平仮名 (は)	1	5	72	82
12	平仮名 (は)	1	5	99	82
13	平仮名 (は)	2	5	72	82
14	平仮名 (は)	2	5	99	82

Retinex 処理の実験条件は 5.2.1 節と同じである。具体的には従来 Retinex については、SSR 処理と 2 スケールの MSR (以下, 2MSR) 処理を行った。ここに 2 スケールとは、2.3.2 節の式(2.5)において $N = 2$ であることを表す。提案 Retinex ではガウスクアーネルを用い、2MSR における式(2.5)の重み w_n は $1/N$ とした。そして従来 Retinex における w_n 以外のパラメータと提案 Retinex におけるパラメータを、グリッドサーチで求めた数字認識の正答率が最大となる値に設定した。

5.5.2 評価方法

ロバスト性の検証における評価方法は 5.2.2 節と同じである。具体的には従来の Retinex 処理画像⁽⁴⁾に対する数字認識の正答率とサポートベクター回帰を用いた Retinex 処理画像⁽¹⁾に対する数字認識の正答率を比較した。数字認識には正規化相互相関 (以下, ZNCC) を用いた。ZNCC による文字認識は、認識対象画像から切り出した数字画像に対して行った。数字領域の切り出し方法は図 5.4 に示したとおりである。本実験の対象画像はいずれも射影歪みがあるため、図 5.4 の分岐では YES の経路を辿る。ZNCC による数字認識の手順は 5.2.2 節に示したとおりである。

適応性の検証実験では 4.6 節冒頭で述べた理由により候補数と正答率が評価指標となる。本実験では実験サンプル数が少ないため正答率を 100 パーセントに固定し、候補数を評価指標とした。本実験では一般的な畳み込みニューラルネットワーク⁽¹⁰⁾を用いて正答率が高い順に正答が得られるまで候補を抽出する手法 (以下, 従来法) と、多重構造の畳み込みニューラルネットワークによる文字認識法⁽⁹⁾で図 4.6 の階層の番号が小さいものから順に正答が得られるまで候補を抽出する手法 (以下, 提案法) とを比較した。

5.5.3 実環境での性能

ロバスト性についての実験結果を表 5.10 と表 5.11 に示す。表 5.10 は実環境画像 1 の実験結果であり、表 5.11 は実環境画像 2 の実験結果である。表 5.10 と表 5.11 において、○印は正解を表し、×印は不正解を表す。

表 5.10 実環境画像 1 における実験結果

	補正前	SSR	2MSR	提案 Retinex
1 桁目	×	×	×	○
2 桁目	×	×	×	○
3 桁目	×	×	×	×
4 桁目	×	○	○	×
正答率[%]	0	25	25	50

表 5.11 実環境画像 2 における実験結果

	補正前	SSR	2MSR	提案 Retinex
1 桁目	×	○	○	○
2 桁目	×	×	×	×
3 桁目	×	○	○	○
4 桁目	×	×	×	○
正答率[%]	0	50	50	75

適応性についての実験結果を図 5.17 に示す。図 5.17 の横軸は画像番号であり、縦軸は正答を得るのに必要な候補数である。

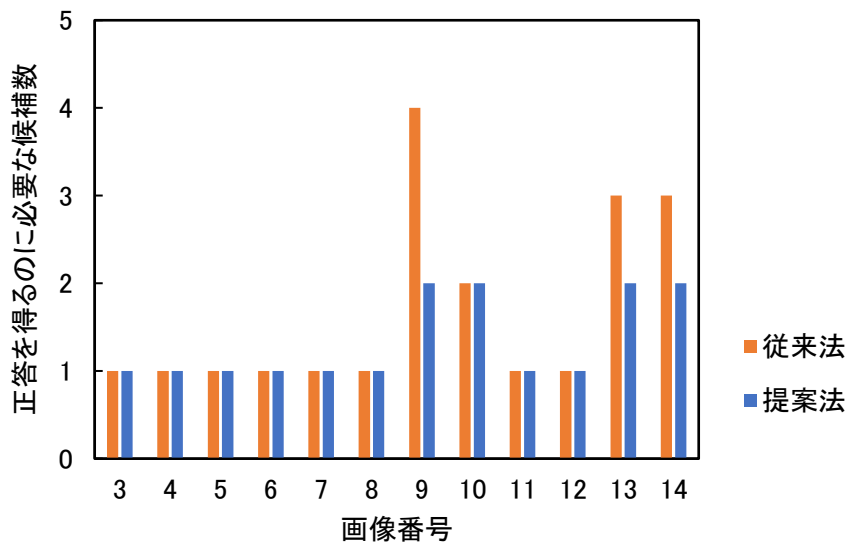


図 5.17 正答を得るのに必要な候補数

図 5.17 の全画像について、認識法ごとに平均候補数を求めると、従来法の平均候補数は 1.67、提案法の平均候補数は 1.33 となった。したがって、提案法の平均候補数は、従来法の平均候補数の 79.6 パーセントになっており、提案法に優位性がある。

提案法に優位性が認められた画像は、画像番号 9, 13, 14 の画像である。これらの画像は、比較的認識しにくい「め」か「は」を対象とし、かつ、低画質の画像が撮影されるカメラ番号 2 のカメラを用いたときの画像である。すなわち、認識が困難な条件の画像において、提案法が優位であった。

5.5.4 考察

最初にロバスト性について考察する。

表 5.10 と表 5.11 より、実環境画像 1 と実環境画像 2 のいずれにおいても、提案 Retinex の正答率は、SSR と 2MSR の正答率よりも高い。実環境画像 1 はナンバー灯が 2 個装着されたナンバープレートの画像であり、実環境画像 2 はナンバー灯が 1 個装着されたナンバープレートの画像であることから、2 枚の照明条件が異なる画像のいずれにおいても提案 Retinex の外乱除去効果は SSR と

2MSR の外乱除去効果よりも高いことがわかる．表 5.10 における各補正後の正答率と表 5.11 における各補正後の正答率とを比較すると，いずれの手法を適用した場合においても実環境画像 2 における正答率の方が実環境画像 1 における正答率よりも高いことがわかる．5.5.1 節で述べたとおり，実環境画像 2 の数字領域での輝度値における標準偏差の平均値は，実環境画像 1 の数字領域での輝度値における標準偏差の平均値よりも小さい．すなわち実環境画像 2 の方が実環境画像 1 よりも数字領域における照明光成分の偏りが小さいといえ，このことが実環境画像 2 における正答率が高くなった原因として考えられる．

次に適応性について考察する．

図 5.17 より，提案法の候補数は，従来法の候補数より少ないか，同じとなっており，いずれの画像においても提案法は従来法を上回る性能か，従来法と同じ性能を達成できていることを確認できる．

認識が困難な条件の画像において提案法が優位であったのは，上記の画像では，両手法において，第 1 候補で正答を得られないためであると考えられる．すなわち，第 1 候補で正答を得られる場合には，そもそも複数候補を選ぶ必要がなく，提案法の優位性が顕在化しないのに対し，第 1 候補で正答を得られない場合には，複数の候補を選ぶ必要が生じ，提案法の優位性が顕在化したと考えられる．

さらに，ナンバープレートを真正面から撮影した画像を学習画像とし，ナンバープレートを斜めから撮影した画像をテスト画像とした場合においても提案法の有効性を確認できることから，学習画像の撮影角度とテスト画像の撮影角度とが異なる場合でも，提案法を適用できることがわかる．

5.6 全実験についての考察

本論文で提案した、サポートベクター回帰を用いた Retinex 処理⁽¹⁾、サポートベクターマシンを用いたマルチフレーム超解像処理⁽⁵⁾、多重構造の畳み込みニューラルネットワークによる文字認識法⁽⁹⁾を全て統合した場合、文字認識の正答率を向上させる効果に加え、車両の絞り込み台数を低減する効果も得られる。前者の効果は提案の各手法により画質改善効果と文字認識性能が向上することに起因し、後者の効果は提案の文字認識法により数字認識に加え従来は不可能であった平仮名認識も可能となることに起因する。

正答率の向上と絞り込み台数の低減は捜査における人的負担と費用を削減する。

次節以降で各効果について具体的に考察する。

5.6.1 正答率に対する効果についての考察

全提案法を統合したときの正答率に対する効果を定量化する。

最初に画質改善法（提案 Retinex と提案超解像）の効果を定量化する。提案 Retinex の正答率の、従来 Retinex の正答率に対する比は、劣化強度により異なるが、その平均値は表 5.6 より 1.25 である。また提案超解像の正答率の、従来超解像の正答率に対する比は表 5.7 より 1.32 である。提案 Retinex と提案超解像とは対象画像に対して独立に適用できることから、提案 Retinex と提案超解像の両方を適用した場合の正答率と、従来 Retinex と従来超解像の両方を適用した場合の正答率との比を粗く見積もると、この比は上記の比の積になると考えられる。すなわちその積を求めると、およそ 1.65 になる。

次に文字認識法の効果を定量化する。提案の文字認識法の正答率の、最高精度の従来の文字認識法の正答率に対する比は表 5.8 より 1.17 である。

最後に提案の画質改善法と文字認識法を両方適用した場合の効果を定量化する。提案の画質改善法と文字認識法とは独立に適用できることから、提案の画質改善法と文字認識法の両方を適用した場合の正答率と、従来の画質改善法と従来の文字認識法の両方を適用した場合の正答率との比を粗く見積もると、この比は上述した画質改善法の比と文字認識法の比の積になると考えられる。す

なわち、その積を求めるとおよそ 1.93 となり、本論文で提案した全手法を統合した場合の正答率は、従来手法に比べて概算で約 2 倍に向上すると見積もることができる。

5.6.2 車両の絞り込み台数に対する効果についての考察

同一の種類ナンバープレート（同一の番号、平仮名、地域名表示文字をもつナンバープレート）は複数あることから、車両の絞り込み台数を直接的に定量化することは難しい。しかしこの台数は、特定できたナンバープレートの種類数に概ね比例するため、この種類数を定量化することで間接的に車両の絞り込み台数を定量化することができる。具体的には数字認識のみが可能な場合と、数字認識と平仮名認識が可能な場合とを比較し、平仮名を特定できることが、ナンバープレートの種類数に与える効果を定量化する。

なお、比較条件を同じにするため、各認識については第 1 候補で 100 パーセント正答率が得られたと仮定する。

まず数字認識のみが可能な場合、ナンバープレートの番号は 1 通りに特定できるが、42 種類ある平仮名と 116 種類ある地域名表示文字についての情報は得られない。よって特定できるナンバープレートの種類数は、平仮名の種類数と地域名表示文字の種類数の積となり、4,872 となる。

次に数字認識と平仮名認識が可能な場合、ナンバープレートの番号に加え、平仮名も 1 通りに特定できる。しかし 116 種類ある地域名表示文字についての情報は得られない。よって特定できるナンバープレートの種類数は、地域名表示文字の種類数と同じになり、116 となる。

以上をまとめると平仮名認識が可能となることにより、特定できるナンバープレートの種類数が 42 分の 1 になり、車両の絞り込み台数は、概ね 42 分の 1 に低減されるといえる。すなわち、実用レベルの認識性能で平仮名を認識できるようになることで車両の絞り込み台数が従来の手法を適用した場合の約 2 パーセントになるといえる。

5.7 まとめ

本章では本論文で提案した手法の、ロバスト性、効率性、適応性に対する評価実験について述べた。そして本論文で提案した手法の実環境での検証実験について述べた。最後に全実験を総括し、考察を述べた。以下に得られた結果を要約する。

- (1) 環境光の外乱に対するロバスト性の評価実験を、複合劣化した数字画像を実験用画像とし、数字認識の正答率を指標として行った結果、サポートベクター回帰を用いた **Retinex** 処理を適用した場合の正答率は、従来の **Retinex** 処理を適用した場合の正答率に比べて 25 パーセント以上向上し、サポートベクター回帰を用いた **Retinex** 処理によりロバスト性が高まることがわかった。
- (2) 少ない文字情報を再構築する際の効率性の評価実験を、複合劣化した数字画像を実験用画像とし、数字認識の正答率を指標として行った結果、サポートベクターマシンを用いた画素値選択型マルチフレーム超解像処理を適用した場合の正答率は、従来マルチフレーム超解像処理を適用した場合の正答率に比べて 32 パーセント以上向上し、サポートベクターマシンを用いた画素値選択型マルチフレーム超解像処理により効率性が高まることがわかった。
- (3) 文字の特徴や劣化の状態に対する適応性の評価実験を、複合劣化した平仮名画像を実験用画像とし、平仮名認識の正答率を指標として行った結果、多重構造の畳み込みニューラルネットワークによる文字認識法を適用した場合の正答率は、従来の文字認識法を適用した場合の正答率に比べて 16 パーセント以上向上し、多重構造の畳み込みニューラルネットワークによる文字認識法により適応性が高まると同時に、実用レベルの認識性能で平仮名を認識できるようになることがわかった。
- (4) 本論文で提案した手法を全て統合した場合の文字認識の正答率を粗く見積もると、この正答率はロバスト性、効率性、適応性が高まることで従来手法

を適用した場合の約 2 倍となることに加え、平仮名認識が可能となることで車両の絞り込み台数が従来の手法を適用した場合の約 2 パーセントになり、車両の検索に要する人的負担と費用が削減されることがわかった。

参考文献

- (1) 辻広生, 福水洋平, 道関隆国, 山内寛紀: “Retinex モデルにおけるサポートベクター回帰を用いた劣化ナンバープレート画像の輝度値補正手法と文字認識への応用”, 画像電子学会誌, Vol. 46, No. 4, pp. 510-521, 2017.
- (2) B. Gunturk, X. Li: “Image Restoration: Fundamentals and Advances”, CRC Press, 2012.
- (3) D. Capel: “Image Mosaicing and Super-resolution”, Springer, pp. 86-87, 2004.
- (4) E. H. Land, J. J. McCann: “Lightness and Retinex Theory”, Journal of the Optical Society of America, Vol. 61, No. 1, pp. 1-11, 1971.
- (5) H. Tsuji, Y. Fukumizu, T. Douseki and H. Yamauchi: “Super-resolution on Pixel Values Selection for Degraded Image Recognition by Support Vector Machine”, Proceedings of 21st Triennial Meeting of the International Association of Forensic Sciences (IAFS), p. 248, 2017.
- (6) P. Milanfar: “Super-resolution Imaging”, CRC Press, 2010.
- (7) L. Stefano, S. Mattoccia, F. Tombari: “ZNCC-based Template Matching Using Bounded Partial Correlation”, Pattern Recognition Letters, Vol. 26, No. 14, pp. 2129-2134, 2005.
- (8) M. Irani and S. Peleg: “Motion Analysis for Image Enhancement Resolution Occlusion and Transparency”, Journal of Visual Communication and Image Representation, Vol. 4, pp. 324-335, 1993.
- (9) 辻広生, 福水洋平, 道関隆国, 山内寛紀: “多重構造とした畳み込みニューラルネットワークによる劣化したナンバープレート画像の平仮名認識”, 信号処理学会誌, Vol. 22, No. 3, pp. 121-134, 2018.
- (10) Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R.E. Howard, W. Hubbard and L. D. Jackel: “Backpropagation applied to handwritten zip code recognition”, Neural Computation, Vol. 1, No. 4, pp. 541-551, 1989.

- (11) C.-L. Liu and M. Nakagawa: "Precise Candidate Selection for Large Character Set Recognition by Confidence Evaluation", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, No. 6, pp. 636-641, 2000.

第6章 結論

本論文は、複合劣化画像を対象としたナンバープレート文字読み取りシステムのロバスト性、効率性、適応性を高めるための、画質改善法と文字認識法を明らかにしたものである。

以下に本研究で得られた主要な結果を要約する。

- (1) ナンバー灯に起因する環境光の外乱を低減してナンバープレート文字読み取りシステムのロバスト性を高めるため、サポートベクター回帰を用いた Retinex 処理を提案した。提案の Retinex 処理には、物体の局所的な反射率の違いが環境光成分の推定精度に及ぼす影響が低減され、また環境光成分がもつ大きさが滑らかに変化するという性質が推定結果に反映されやすくなるため、従来の Retinex 処理よりも環境光成分の推定精度が高まるという利点があることを述べた。そして実験により提案の Retinex 処理は、従来の Retinex 処理と比べて、副作用の Halo 作用低減効果が大きく、総合的な画質改善効果も向上することを示した (第2章)。
- (2) 複合劣化により少なくなったナンバープレートの文字情報を確実に利用してナンバープレート文字読み取りシステムの効率性を高めるため、サポートベクターマシンを用いた画素値選択型マルチフレーム超解像処理を提案した。提案の超解像処理では低解像度画像の劣化した画素値の影響が除去され、さらにエッジ復元効果が高まるという利点があることを述べた。そして実験により提案の超解像処理は、従来の超解像処理に比べて、エッジ復元効果が大きく、総合的な画質改善効果も向上することを示した (第3章)。
- (3) 文字の特徴や劣化の状態に柔軟に対応してナンバープレート文字読み取りシステムの適応性を向上させるため、多重構造の畳み込みニューラルネットワークによる文字認識法を提案した。提案の文字認識法には一般的な畳み込

みニューラルネットワークの、入力画像の解像度を調整して正答率を最大化する機能を持たないという問題と、解像度を調整して候補数を適応的に制御し正答率を保証する機能がないという問題が低減される利点があることを述べた。そして実験により提案の文字認識法を用いた場合、従来の文字認識法を用いた場合に比べて文字認識の正答率が向上することを示した（第4章）。

(4) 提案の各手法の評価実験について述べ、提案の各手法を用いることで文字読み取りシステムのロバスト性、効率性、適応性が向上することを定量的に示した。そして提案の各手法の実環境での検証実験について述べ、これらの手法の実用性を示した。さらに各評価実験結果を総括し、提案した全手法統合した場合の効果を定量的に示した（第5章）。

第7章 今後の課題

7.1 まえがき

街頭や店頭に設置された一般的な監視カメラが撮影した複合劣化画像を対象としたナンバープレート文字読み取りシステムにおいて重要な画質改善法と文字認識法について、機械学習を活用した手法を提案し、その有用性を示してきた。

本章では一般的な監視カメラが撮影した複合劣化画像を対象としたナンバープレート文字読み取りシステムの今後の課題について考察する。このシステムを完成させるためには、画質改善法と文字認識法に加え、ナンバープレート領域の抽出法、ナンバープレートの傾き補正法、ナンバープレート文字領域の抽出法が必要となる。これらについての一般的な手法では複合劣化画像に対応できないため、現状ではこれらの作業を手動で行ってシステムを補助するしかない。しかしシステムの効率性や認識結果の信頼性、客観性を高めるためには手動による手法よりも高精度で信頼性の高い手法が求められる。よって複合劣化画像に対応できるよう、ナンバープレート領域の抽出性能、ナンバープレートの傾き補正性能、ナンバープレート文字領域の抽出性能を向上させることが今後の課題となる。そして全手法を効率的に実行するアルゴリズムを構築し、全種類のナンバープレートに対してシステム全体が機能するかを実環境で確認することが課題となる。さらには動いている車両の画像は、モーションブラーによる劣化を受ける可能性が高い。このモーションブラーが顕著な場合、認識性能が著しく低下する。したがって、動いている車両を対象とする場合の対策が課題となる。

以下に、上記の各課題を解決する手法について考察する。

7.2 ナンバープレート領域の抽出性能の向上

監視カメラの複合劣化画像ではぼけやノイズの影響でエッジ抽出や2値化の精度が低下し、さらに射影歪みやレンズ歪みによりナンバープレート領域が複合的に歪むため、一般的なナンバープレート領域抽出法が機能しない。よって、この抽出性能を向上させる必要がある。

これを向上させる手法として機械学習システムに複合劣化と、射影歪みやレンズ歪みなどの歪みを受けたナンバープレート画像の特徴を学習させ、この学習させた機械学習システムでナンバープレート領域を抽出する手法が考えられる。この向上法を図7.1に示す。この向上法ではまず計算機上で画像劣化、歪みモデルを構築する。次に入力画像から推定した劣化パラメータ、射影歪みやレンズ歪みについての全歪みパラメータ、理想的な環境で撮影したナンバープレート全体の画像をこのモデルに入力し学習画像データベースを作成する。ここで歪みパラメータは入力画像のみから推定することが困難と考えられるため全パラメータを機械学習システムに入力する。そしてナンバープレート領域抽出用の機械学習システムを構築し、このシステムに学習画像データベースと入力画像を入力することでナンバープレート領域を抽出する。この手法を用いればぼけやノイズなどの複合劣化の影響と、射影歪み、レンズ歪みの影響を、学習を通して考慮できるようになるため領域抽出性能が向上すると考えられる。

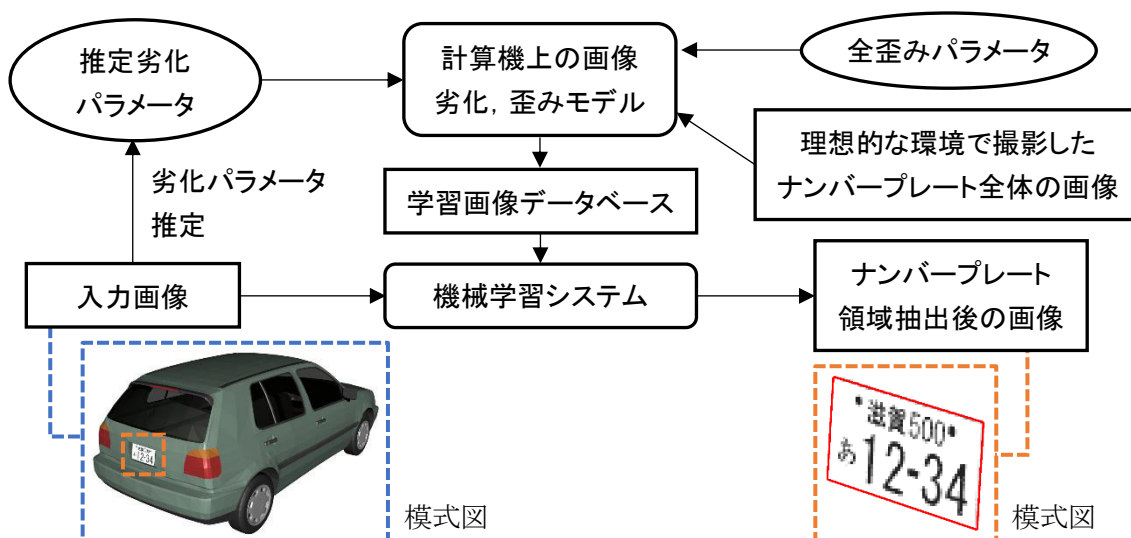


図 7.1 ナンバープレート領域抽出性能の向上法

7.3 ナンバープレートの傾き補正性能の向上

監視カメラの複合劣化画像では上述したナンバープレート領域抽出法の場合と同じ理由により，一般的なナンバープレートの傾き補正法が機能しない．よって，この補正性能を向上させる必要がある．

これを向上させる手法として，機械学習システムに複合劣化と，射影歪みやレンズ歪みなどの歪みを受けたナンバープレート画像の特徴を学習させ，この学習させた機械学習システムで歪みパラメータを推定して傾きを補正する手法が考えられる．この向上法を図 7.2 に示す．この手法ではまず計算機上で画像劣化，歪みモデルを構築し，学習画像データベースを作成する．そして計算機上で構築した歪み推定用の機械学習システムを用いて歪みパラメータを推定する．最後に計算機上で構築した傾き補正モデルに推定歪みパラメータとナンバープレート領域抽出後の画像を入力し，ナンバープレート領域抽出後の画像の傾きを補正する．この手法を用いれば，ぼけやノイズなどの複合劣化の影響と，射影歪み，レンズ歪みの影響を，学習を通して考慮できるようになるため，傾き補正性能が向上すると考えられる．

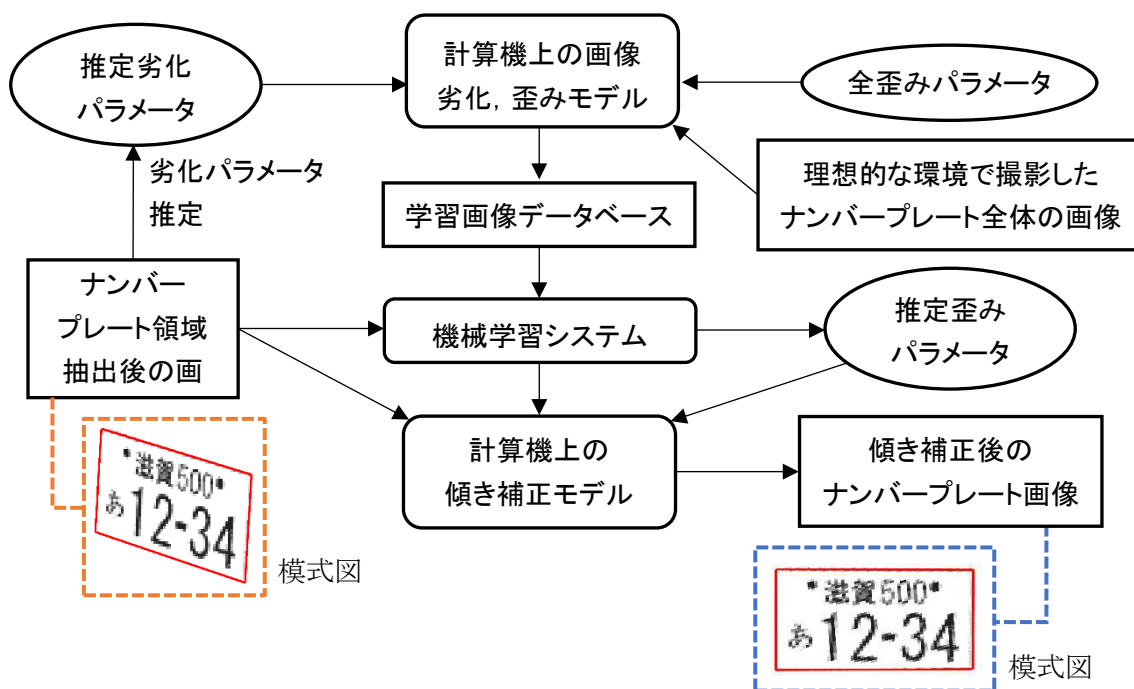


図 7.2 ナンバープレートの傾き補正性能の向上法

7.4 ナンバープレート文字領域の抽出性能の向上

監視カメラの複合劣化画像では、ぼけやノイズの影響でエッジ抽出や2値化の精度が低下するため、一般的なナンバープレート文字領域抽出法が機能しない。よって、この抽出性能を向上させる必要がある。

これを向上させる手法として、機械学習システムに複合劣化を受けたナンバープレート文字画像の特徴を学習させ、この学習させた機械学習システムで文字領域を抽出する手法が考えられる。この向上法を図 7.3 に示す。この手法では入力画像から推定した劣化パラメータと理想的な環境で撮影したナンバープレート文字画像を計算機上の画像劣化モデルに入力し学習画像データベースを作成する。そしてナンバープレート文字領域抽出用の機械学習システムを構築し、このシステムに学習画像データベースと傾き補正後のナンバープレート画像を入力することでナンバープレート文字領域を抽出する。この文字領域は複数抽出されるが、各抽出領域の相対位置はナンバープレートの規格により定まっている。よって最後にこの規格も考慮して各抽出位置の誤差が最小になるように各位置を調整する。この手法を用いれば、ぼけやノイズなどの複合劣化の影響も学習を通して考慮できるようになるため、領域抽出性能が向上すると考えられる。

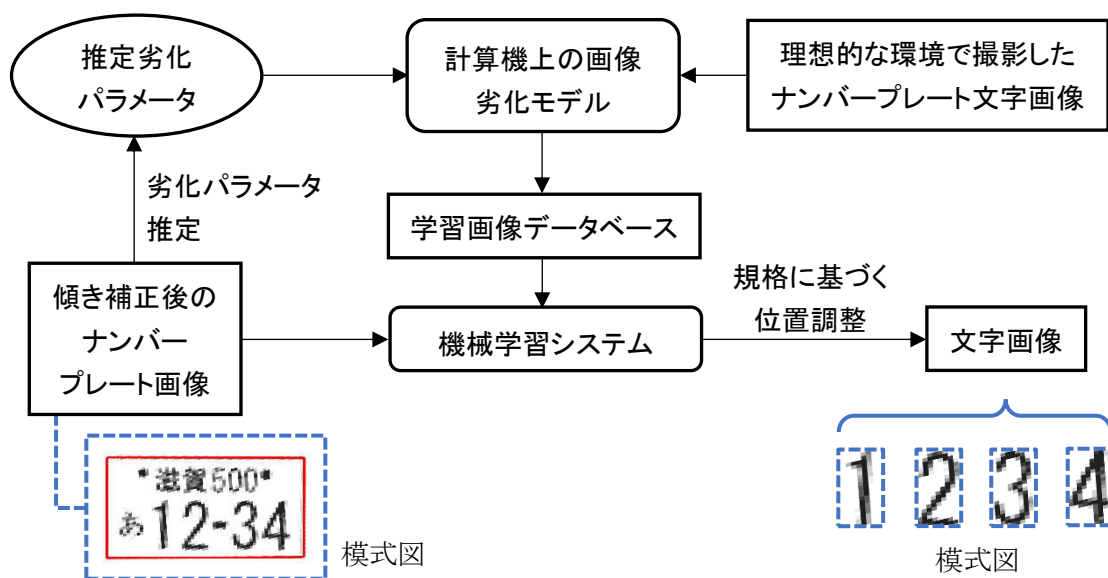


図 7.3 ナンバープレート文字領域抽出性能の向上法

7.5 実環境におけるナンバープレート文字読み取りシステムの性能確認

監視カメラが撮影した複合劣化画像を対象としたナンバープレート文字読み取りシステムでは、ナンバープレート領域抽出法、ナンバープレート画像の画質改善法、ナンバープレートの傾き補正法、ナンバープレート文字領域の抽出法、ナンバープレート文字認識法が構成要素となる。

ナンバープレート文字読み取りシステム全体を稼働させるためには第1に上記の全手法を効果的に実行するアルゴリズムを構築し、システム全体として機能するかを実環境で確認する必要がある。複合劣化画像に対応するためには画質改善法を他の各手法の前処理として用いるアルゴリズムが有効であると考えられる。このアルゴリズムを図7.4に示す。画質改善法のパラメータは直後の手法の性能が最大限に引き出されるよう前処理ごとに最適化する必要がある。

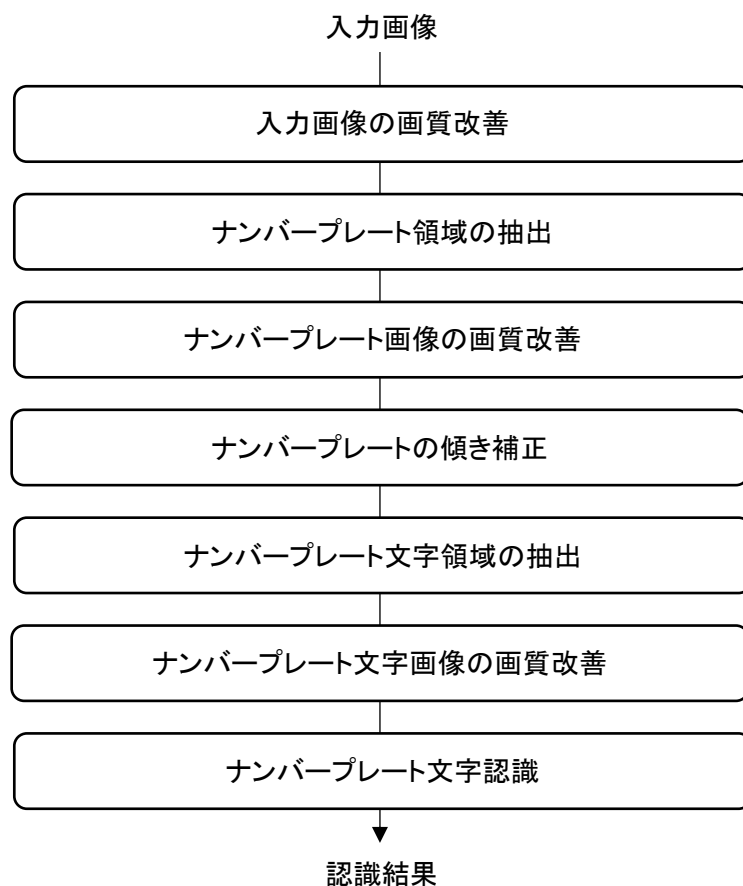


図 7.4 複合劣化画像に対するシステム全体の実行アルゴリズム

第2に開発したシステム全体の、全種類のナンバープレートに対する、文字読み取りの性能を実環境で確認する必要がある。本論文の実験はいずれも白色のプレートに深緑色の文字が記載されたナンバープレートか黄色のプレートに黒色の文字が記載されたナンバープレートを用いて行った。その結果、本論文の実験の範囲内では両者に差は認められなかった。本論文の提案手法はいずれも対象画像をグレースケール画像に変換してから処理を行うので上述した塗色の違いが処理結果に及ぼす影響は小さいと考えられる。しかし字光式ナンバーについては文字から発せられる光とナンバー灯の光とがどのように影響し合うかなど、不明点があり、さらなる検討が必要となる可能性がある。

7.6 動いている車両のナンバープレートを対象とする場合の対策

街頭や店頭に設置された一般的な監視カメラの映像では、対象とする車両が停止している場合と、動いている場合がある。対象車両が動いている場合、一般的な監視カメラのシャッター速度は、通常、この動きに対して十分な速さとなっていない。このため対象車両が動いている場合、一般的に、ナンバープレート画像は顕著なモーションブラーを受ける。このモーションブラーにより認識性能が大きく低下するという問題が生じるため、対策が必要となる。

本研究で提案した多重構造の畳み込みニューラルネットワークによる文字認識法は、適応性が高い文字認識法であり、この畳み込みニューラルネットワークにモーションブラーを学習させることで、上記の問題を低減できると考えられる。しかし、この文字認識法は、顕著なモーションブラーの劣化を受けた画像に対しては、実用レベルの認識性能を達成できない可能性がある。なぜなら、この手法は多重解像度画像を処理することで認識性能を高めるものであるが、解像度を低下させた各画像では、原画像の局所的な劣化は除去されるが、モーションブラーのように画像の広域にわたる劣化は除去されないからである。したがって、動きに特化した多重構造の畳み込みニューラルネットワークを開発するといった対策が求められる。この畳み込みニューラルネットワークの模式

図を図 7.5 に示す。図 7.5 における記号の説明を下記に示す。

- M1 から M_n : モーションブラーの影響を低減させた n 段階の画像
- CL : 正解ラベル

図 7.5 の実線部分と点線部分とは、それぞれ学習時とテスト時における、使用部分と不使用部分を表す。この畳み込みニューラルネットワークでは、まず学習時に、学習画像からモーションブラーの影響を低減させた画像を複数生成する。そして、生成した画像を正解ラベルとともに各畳み込みニューラルネットワークに入力し、入力画像と正解ラベルの関係を学習する。テスト時には、学習時と同様に、対象画像からモーションブラーの影響を低減させた画像を複数生成し、これらの画像を上述した学習済み畳み込みニューラルネットワークに入力する。最後に、各畳み込みニューラルネットワークの認識結果を統合し、統合結果を最終的な認識結果とする。この方法を用いることで、モーションブラーの影響が低減される可能性が高まり、認識性能が改善すると考えられる。

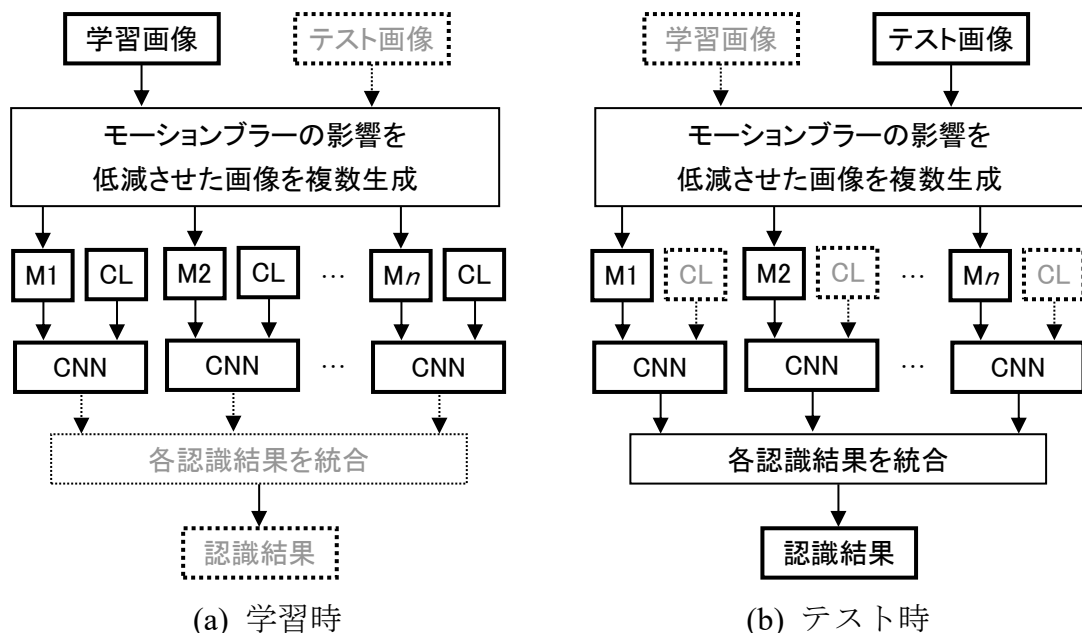


図 7.5 動きに特化した多重構造の畳み込みニューラルネットワークの模式図

7.7 まとめ

本章では監視カメラが撮影した複合劣化画像を対象としたナンバープレート文字読み取りシステムの今後の課題について考察した。以下に考察した内容をまとめる。

- (1) 監視カメラの複合劣化画像ではぼけやノイズの影響でエッジ抽出や 2 値化の精度が低下し，さらに射影歪みやレンズ歪みによりナンバープレート領域が複合的に歪むため，一般的なナンバープレート領域抽出法が機能しない。この抽出性能を向上させる方法として，機械学習システムに複合劣化と，射影歪みやレンズ歪みなどの歪みを受けたナンバープレート画像の特徴を学習させ，この学習させた機械学習システムでナンバープレート領域を抽出する方法が考えられる。この方法を用いれば，ぼけやノイズなどの複合劣化の影響と，射影歪みやレンズ歪みなどの歪みの影響を，学習を通して考慮できるようになるため，領域抽出性能が向上すると考えられる。
- (2) 監視カメラの複合劣化画像ではぼけやノイズの影響でエッジ抽出や 2 値化の精度が低下し，さらに射影歪みやレンズ歪みによりナンバープレート領域が複合的に歪むため，一般的なナンバープレートの傾き補正法が機能しない。この補正性能を向上させる方法として，機械学習システムに複合劣化と，射影歪みやレンズ歪みなどの歪みを受けたナンバープレート画像の特徴を学習させ，この学習させた機械学習システムで歪みパラメータを推定した後，計算機上で構築した傾き補正モデルでこの傾きを補正する方法が考えられる。この方法を用いれば，ぼけやノイズなどの複合劣化の影響と，射影歪み，レンズ歪みなどの歪みの影響を，学習を通して考慮できるようになるため，傾き補正性能が向上すると考えられる。
- (3) 監視カメラの複合劣化画像では，ぼけやノイズの影響でエッジ抽出や 2 値化の精度が低下するため，一般的なナンバープレート文字領域抽出法が機能しない。この抽出性能を向上させる方法として，機械学習システムに複合劣

化を受けたナンバープレート文字画像の特徴を学習させ、この学習させた機械学習システムで文字領域を抽出する向上法が考えられる。この手法を用いれば、ぼけやノイズなどの複合劣化の影響も学習を通して考慮できるようになるため、領域抽出性能が向上すると考えられる。

(4) ナンバープレート文字読み取りシステムを稼働させるためには、ナンバープレート領域抽出法、ナンバープレート画像の画質改善法、ナンバープレートの傾き補正法、ナンバープレート文字領域の抽出法、ナンバープレート文字認識法の全てを効果的に実行するアルゴリズムを構築し、システム全体として機能するかを実環境で確認する必要がある。複合劣化画像に対応するためには画質改善法を他の各手法の前処理として用いるアルゴリズムが有効であると考えられる。そして開発したシステム全体の、全種類のナンバープレートに対する、文字読み取りの性能を実環境で確認する必要があると考えられる。

(5) 動いている車両のナンバープレートを対象とする場合、一般的な監視カメラのナンバープレート画像は、通常、顕著なモーションブラーを受け、システムの認識性能が大きく低下するという問題が生じる。この問題を解決する方法として、モーションブラーの影響を低減させた画像を複数生成し、これらの画像を動きに特化した多重構造の畳み込みニューラルネットワークで処理する方法が考えられる。この方法を用いれば、入力画像のモーションブラーの影響を低減できる可能性が高まり、認識性能が改善すると考えられる。

謝辞

本論文は、筆者が立命館大学大学院理工学研究科電子システム専攻博士課程後期課程在学中に行った研究をまとめたものです。

本研究をまとめるにあたり、ご指導とご助言ならびにご高配を賜った立命館大学理工学部電子情報工学科 道関隆国教授に深甚なる感謝の意を表します。

本論文の審査にあたり、ご指導とご助言を賜りました立命館大学理工学部電子情報工学科 小倉武教授，同学科 泉知論教授，同学部電気電子工学科 福水洋平准教授に深く感謝いたします。特に，福水洋平准教授には，本研究全般にわたり，多大なご指導とご助言を賜りました。重ねてお礼申し上げます。

本研究の全般にわたり多大なご指導とご助言ならびにご高配を賜りました，立命館大学理工学部電子情報工学科 山内寛紀名誉教授に深甚なる感謝の意を表します。

本研究の遂行にあたり，ご理解いただきました滋賀県警察本部科学捜査研究所の歴代所長である，山田直司氏，中村準一氏，岡本勝弘氏，兼正晃氏，山元光信氏，ならびに同研究所の所員の皆様に感謝いたします。

最後に，本研究を行うにあたり応援いただきました，父，母，友人に心より感謝いたします。

本研究に関する発表文献リスト

1. 学術論文誌

- (1) 辻広生, 福水洋平, 道関隆国, 山内寛紀: “Retinex モデルにおけるサポートベクター回帰を用いた劣化ナンバープレート画像の輝度値補正手法と文字認識への応用”, 画像電子学会誌, Vol. 46, No. 4, pp. 510-521, 2017.
- (2) 辻広生, 福水洋平, 道関隆国, 山内寛紀: “多重構造とした畳み込みニューラルネットワークによる劣化したナンバープレート画像の平仮名認識”, 信号処理学会誌, Vol. 22, No. 3, pp. 121-134, 2018.

2. 査読付き国際会議

- (1) H. Tsuji, Y. Fukumizu, T. Douseki and H. Yamauchi: “Super-resolution on Pixel Values Selection for Degraded Image Recognition by Support Vector Machine”, Proceedings of 21st Triennial Meeting of the International Association of Forensic Sciences (IAFS), p. 248, 2017.

3. 研究会

- (1) 辻広生, 福水洋平, 道関隆国, 山内寛紀, 吉川歩: “画像認識のための回帰分析を用いた光量補正”, 電子情報通信学会技術研究報告, Vol. 115, No. 208, pp. 63-68, 2015.
- (2) 辻広生, 福水洋平, 道関隆国, 山崎史紘, 山内寛紀, 吉川歩: “機械学習によるナンバープレート数字画像認識精度の向上”, 情報処理学会研究報告, Vol. 2016, No. CVIM-202, pp. 1-6, 2016.

4. 講演会

- (1) 辻広生, 吉川歩: “Haar-like 特徴を用いた低解像度ナンバープレート平仮名の識別”, 日本法科学技術学会第 19 回学術集会, D-16, 2013.
- (2) 辻広生, 福水洋平, 道関隆国, 中瀬裕太, 山内寛紀, 吉川歩: “ナンバープ

プレート数字画像認識に適合した画素値選択型超解像”, 2016 年電子情報通信学会総合大会, D-11-62, 2016.

- (3) 辻広生, 福水洋平, 道関隆国, 横山智之, 山内寛紀, 吉川歩: “機械学習による劣化ナンバープレート平仮名画像の認識”, 日本法科学技術学会第22回学術集会, D-20, 2016.
- (4) 辻広生, 福水洋平, 道関隆国, 横山智之, 山内寛紀, 吉川歩: “多重解像度画像を用いた劣化ナンバープレート平仮名画像認識”, 2017 年電子情報通信学会総合大会, D-12-1, 2017.
- (5) 辻広生, 福水洋平, 道関隆国, 瀬尾賢哉, 山内寛紀: “劣化したナンバープレート文字画像に対する超解像のための深層学習によるサブピクセル位置推定”, 2018 年電子情報通信学会総合大会, D-12-20, 2018.