

東京電機大学

視覚障がい者のための  
日常生活支援システムの開発

Daily Living Activities Support System  
for the Visually Impaired

平成 30 年(2018 年) 3 月

淵田 正隆



## 概要

視覚障がい者の日常生活における不便の一部を解決することを目標として、本研究では「地面性状識別」および「指先指定文字抽出」それぞれに関する手法を提案し、システムを開発する。

視覚障がい者のための支援器具として、白杖と点字が思い浮かぶ。白杖は外出時の移動支援に利用され、点字は文字情報の取得支援に利用されている。しかし、白杖に関して、周囲の障害物情報を知覚するには便利であるが、打突しただけでは分かりにくい情報の取得には不都合である。例えば地面の性状を識別することは難しく、すべりやつまづきにつながる。点字に関しては、そもそも点訳情報が少なく、また、視覚障がい者のうち約一割しか判読することができないといわれている。それに対し、商品や書籍の表面に書かれている文字をすべて読み上げた場合、情報過多による混乱につながる。

以上の事情に鑑み、本研究の目的を、「地面性状識別」すなわち前方の近距離地面の性状を識別する手法と、「指先指定文字抽出」すなわち指先で指定された文字列を抽出し読み上げる手法、それぞれに関して提案・実装することと定める。

本論文は、7章より構成される。第1章は序論であり、背景、先行研究一部紹介、目的を述べる。第2章では先行研究紹介および問題設定を概説する。第3章および第4章は「地面性状識別」、第5章および第6章は「指先指定文字抽出」に関して述べる。第3章と第5章ではそれぞれの手法に関して詳述し、第4章と第6章は提案手法の有効性を検証するとともにその結果に関して考究を加える。以上を踏まえたうえ、第7章で本研究を総括する。

第1章では、視覚障がい者支援の現状から「地面性状識別」および「指先指定文字抽出」を取り扱う意義を述べ、先行研究の一部を概説した後、上述した本研究の目的を詳述する。

第2章では、第1章で触れなかった先行研究の多くを紹介・評価し、これまでに解決されていない問題点を俯瞰する。「地面性状識別」および「指先指定文字抽出」それぞれに対して、問題設定を行うとともに、解決のためのアプローチを決定して概説する。「地面性状識別」に関しては、教師あり学習手法の一種である **Support Vector Machine (SVM)** を用いて、地面性状の識別器を作成し、それを用いて前方の地面性状を識別する。「指先指定文字抽出」に関しては、赤外線照射機器と指先検出用指輪型マーカを用いて照明変化に対して頑健に指先を検出する。検出した指先で指定した領域内の文字のみ抽出する。

第3章では「地面性状識別」に関して詳述する。前方地面が単一の性状ではなく、複数の性状が混在している可能性を考慮して、撮影画像下部を分割して画像を収集し、

データベースを作成する。画像からは形状を表す Speeded Up Robust Features (SURF) と色を表す色ヒストグラムを抽出し、Bag of Features (BoF) にてベクトル量子化する。得られた特徴を利用して SVM にて識別器を作成し、一対他分類法で複数性状を識別する。

第4章では「地面性状識別」手法の有効性を検証するとともにその結果に関して考究を加える。4種類の地面（芝生、砂利、コンクリート、ウッドデッキ）を識別した結果、再現率、適合率ともに90%以上であることを確認した。

第5章は「指先指定文字抽出」に関して詳述する。照明の変化やカメラの揺動の影響を防ぐために、赤外線を利用した指先検出手法を提案した。赤外線を前方に照射し、指先検出用指輪型マーカ上の3個の赤外線反射材の位置関係から指先位置を算出する。また、文字が混在する環境において、指先で指定した領域内から、色の分布と文字候補の形状に基づき、傾きやノイズに対処した上で文字を抽出する手法を提案する。

第6章では第5章で提案した「指先指定文字抽出」手法の有効性を検証するとともにその結果に対して考究を加える。様々な照明条件下で指先を検出した結果、十分な赤外線が照射されている範囲では90%以上で検出できた。また、様々な照明条件下で10種類の物品上の文字を抽出した結果、細文字を除くと、70%以上で抽出できた。

第7章で本研究を総括する。視覚障がい者の日常生活支援システムとして、「地面性状識別」および「指先指定文字抽出」それぞれに関する画像処理手法を提案した。実験を通して、それぞれの手法の有効性を示した。「地面性状識別」に関しては、代表的な4種類の地面性状を識別できることを示した。「指先指定抽出手法」に関しては、赤外線照射機器と指先検出用指輪型マーカを用いた指先検出手法が照明の変化に頑健なことを示した。また、色や形状の情報を利用して指先指定領域内の文字を限定して抽出する手法が複数文字の混在環境下からの文字抽出に有効であることを示した。

本研究は、視覚障がい者が日常生活で抱える些細な不満・問題を取り上げ、工学的なアプローチで解決を図ったものであり、今後、現実に即した視覚障がい者支援システムを開発するための基礎的な知見を得ることができたと考える。



## 目次

1. 序論.....	1
1.1 背景.....	2
1.2 先行研究.....	5
1.2.1 地面性状識別.....	5
1.2.2 文字情報取得.....	6
1.3 目的.....	8
1.4 論文構成.....	9
2. 提案システムの概要.....	11
2.1 はじめに.....	12
2.2 視覚障がい者の定義.....	14
2.3 従来研究.....	16
2.3.1 誘導に関する研究.....	16
2.3.2 障害物検知に関する研究.....	18
2.3.3 地面性状の識別に関する研究.....	20
2.3.4 文字情報認識に関する従来研究.....	25
2.3.5 文字抽出に関する従来研究.....	29
2.3.6 領域の指示に関する従来研究.....	33
2.4 従来研究のまとめ.....	37
2.5 提案する視覚障がい者のための地面性状識別システムの概要.....	41
2.6 提案する視覚障がい者のための文字取得支援システムの概要.....	43
2.7 おわりに.....	45
3. 地面性状識別.....	47
3.1 はじめに.....	48
3.2 地面性状識別手法の流れ.....	49
3.3 データベースの構築.....	50
3.4 SVMによる地面性状の識別.....	65
3.5 おわりに.....	70
4. 地面性状識別システムの評価.....	71
4.1 はじめに.....	72
4.2 実験条件.....	73
4.3 実験結果.....	75
4.4 おわりに.....	77
5. 墨字抽出.....	79
5.1 はじめに.....	80

5.2 指先検出 .....	81
5.2.1 素手のまま利用可能な指先検出 .....	82
5.2.2 マーカを用いた指先検出 .....	86
5.2.3 指先検出のまとめ .....	91
5.3 指先文字抽出 .....	93
5.3.1 指先文字抽出の概要 .....	93
5.3.2 指先文字抽出手法 .....	96
5.3.3 文字形状に基づく処理 .....	101
5.3.4 文字列傾き補正 .....	102
5.3.5 文字候補抽出およびノイズ除去 .....	104
5.3.6 文字の高さに基づく処理 .....	105
5.3.7 文字の隙間に基づく文字列合成処理 .....	106
5.4 おわりに .....	109
6. 墨字認識システムの評価 .....	111
6.1 はじめに .....	112
6.2 指先位置検出実験 .....	113
6.3 文字抽出実験 .....	117
6.4 指先指定文字抽出実験 .....	120
6.5 おわりに .....	124
7. 結論 .....	125
謝辞 .....	129
参考文献 .....	133
業績リスト .....	143

## 図番

Fig. 1 本論文の構成 .....	9
Fig. 2 ランドルト環 .....	15
Fig. 3 バーコード .....	26
Fig. 4 QR コード .....	27
Fig. 5 OCR の結果 .....	28
Fig. 6 画像の加工と認識結果の変化 .....	29
Fig. 7 環境中に存在する複数の文字 .....	33
Fig. 8 Google 翻訳(Google, Inc.) .....	36
Fig. 9 複数の地面性状が撮影された画像の例 .....	38
Fig. 10 対象とする文字 .....	43
Fig. 11 地面性状識別手法のシステムフロー .....	49

Fig. 12 複数枚の平滑化画像.....	50
Fig. 13 DoG.....	51
Fig. 14 特徴点の検出.....	51
Fig. 15 特徴点の周りの輝度勾配.....	52
Fig. 16 輝度勾配のヒストグラム.....	52
Fig. 17 オリエンテーション.....	53
Fig. 18 基準方向への回転.....	53
Fig. 19 領域ごとの輝度勾配.....	54
Fig. 20 色ヒストグラムの例.....	55
Fig. 21 減色処理の例 (RGB).....	55
Fig. 22 色相と彩度の分割(減色処理).....	56
Fig. 23 彩度と明度の分割(減色処理).....	56
Fig. 24 4種の地面性状の例.....	58
Fig. 25 入力画像の分割.....	59
Fig. 26 グリッドサンプリング.....	60
Fig. 27 特徴点の転写.....	60
Fig. 28 k-means 法が適用できない例.....	61
Fig. 29 k-means 法(1).....	61
Fig. 30 k-means 法(2).....	62
Fig. 31 k-means 法(3).....	62
Fig. 32 VW による特徴ベクトルの置き換え.....	63
Fig. 33 空間の分割.....	63
Fig. 34 VW による置き換え.....	64
Fig. 35 画像を表すベクトル.....	64
Fig. 36 線形識別関数を用いた2クラス分離問題の分離結果(線形).....	66
Fig. 37 線形識別関数を用いた2クラス分離問題の分離結果(線形以外).....	66
Fig. 38 マージンの最大化.....	67
Fig. 39 カーネルトリック.....	68
Fig. 40 クラスが入り混じる際の分割.....	68
Fig. 41 使用した地面性状の一部.....	73
Fig. 42 Ai-ball(Trek).....	74
Fig. 43 識別結果.....	76
Fig. 44 指先以外の指示方法.....	82
Fig. 45 指先による指示方法.....	82
Fig. 46 画像ピラミッド.....	84
Fig. 47 オプティカルフローを用いた指先検出.....	85

Fig. 48	オプティカルフロー(GF 法)	85
Fig. 49	高速で動く領域の抽出	85
Fig. 50	凸法	86
Fig. 51	指先の検出 (オプティカルフロー使った手法)	86
Fig. 52	マーカの装着位置	88
Fig. 53	カメラユニットと指輪型マーカ	89
Fig. 54	再帰性反射	89
Fig. 55	赤外線反射光の検出	89
Fig. 56	指先位置の計算と外れ値除去	91
Fig. 57	指先検出	91
Fig. 58	将来的なシステムの概要図	93
Fig. 59	RGB 表色系のモデル	97
Fig. 60	L*a*b*表色系のモデル	97
Fig. 61	指先周りの領域	97
Fig. 62	指先の色分布 (L*a*b*表色系)	98
Fig. 63	ボクセル(L*a*b*表色系)	99
Fig. 64	文字列候補画像	99
Fig. 65	大きさによる同一文字列の抽出	100
Fig. 66	細線化	101
Fig. 67	傾いた文字の中心(白黒反転)	102
Fig. 68	傾き補正	103
Fig. 69	文字線の太さ測定	104
Fig. 70	文字線の太さを基準にしたマスク	105
Fig. 71	膨張させたマスク	105
Fig. 72	指先周辺のヒストグラム	106
Fig. 73	抽出される文字列	106
Fig. 74	ノイズを含んだ文字列	107
Fig. 75	文字列の縦幅と文字間距離	107
Fig. 76	文字の拡大	108
Fig. 77	指先指定文字の抽出	108
Fig. 78	指先検出実験 : 実験環境	114
Fig. 79	文字抽出実験 : 実験環境	117
Fig. 80	実験対象	118
Fig. 81	文字の欠け	119
Fig. 82	文字の欠け(250mm)	119
Fig. 83	文字の欠け(450mm)	119

Fig. 84 指先指定文字抽出実験：実験環境 .....	121
Fig. 85 成功例 .....	121
Fig. 86 指先検出の失敗 .....	122
Fig. 87 文字色推定の失敗 .....	122
Fig. 88 文字列生成の失敗 .....	122
Fig. 89 傾き補正の失敗 .....	122

#### 表番

Table 1 同じような外形で区別しづらい包装用紙 .....	4
Table 2 視覚障害等級([厚生労働省 2009]) .....	14
Table 3 地面性状識別に関する従来研究まとめ .....	38
Table 4 文字の抽出に関する従来研究まとめ .....	40
Table 5 指先検出に関する従来研究まとめ .....	40
Table 6 地面性状識別システムの結果 .....	76
Table 7 指先検出実験の実験環境 .....	113
Table 8 指先検出結果(実験環境 1) .....	114
Table 9 指先検出結果(実験環境 2) .....	115
Table 10 指先検出結果(実験環境 3) .....	115
Table 11 指先検出結果(実験環境 4) .....	115
Table 12 指先検出結果(実験環境 5) .....	115
Table 13 指先検出結果(実験環境 6) .....	116
Table 14 それぞれの実験環境の平均指先検出結果 .....	116
Table 15 文字抽出実験の実験環境 .....	118
Table 16 文字抽出実験結果 .....	118
Table 17 指先指定文字抽出実験：結果 .....	121
Table 18 読取革命：認識結果 .....	123
Table 19 Google 翻訳：認識結果 .....	123



# 1. 序論

---

1.1 背景 .....	2
1.2 先行研究 .....	5
1.2.1 地面性状識別 .....	5
1.2.2 文字情報取得 .....	6
1.3 目的 .....	8
1.4 論文構成 .....	9

---

## 1.1 背景

2018年現在、IoT技術が著しい発展を遂げているが、一般の人には、その技術がどのように発展し、どのようなことが実現できているのかが正しく伝わっていない。視覚障がい者も同様で、彼らとの会話や支援機器の展示会を見学した際、画像処理技術としては教科書に記載されているような基礎的な技術を用いた支援機器が高額で提供されており、愕然とした。視覚障がい者の日常の些細な不満の中には、工学的に解決ができるにもかかわらず、それを視覚障がい者自身が知らないため、我慢してしまうことや情報発信もなされていないことがある。すなわち、工学的なシーズと視覚障がい者のニーズが乖離している。本博士論文の提案により、視覚障がい者の些細な不満を無視せず、工学的なアプローチで多少なりとも解消すべきである。

人間が環境から得る情報のほとんどは視覚を通してであり、視覚を失うことは情報の大部分の取得手段を失うことである、と考えがちである。しかし、視覚障がい者は、自分の発する足音およびその反響音を聞き取る聴覚、道路の勾配を感じる足裏の触覚、周囲の食べ物屋の匂いを感じる嗅覚を利用する。すなわち、聴覚、触覚、嗅覚といった、健常者も特に意識することなく識別できる視覚以外の他の感覚器官を利用して、環境から行動に役立つ情報を得ている。

ただ、だからといって全ての視覚障がい者が自由に環境を動き回れるわけではない。厚生労働者によると、現在日本には31万人の視覚障がい者がおり、そのほとんどが中途失明者である[厚生労働省 2008]。中途失明者の大部分は視覚以外の感覚器官を使っても日常行動に不自由ないほど周囲の環境を識別できるとは限らない。学習・経験を積み、熟練すればよいのだろうが非常に時間がかかる。

まず、私たち健常者と視覚障がい者自身が考える「必要な支援」には齟齬があると考え、視覚障がい者の方のニーズを調査した。共用品推進機構が2010年に実施したアンケート調査では、「もしもこの世の中にあったら、便利な器具」という質問について調査しており、そのうち上位3位が「安全歩行」に関するもの、「情報取得」に関するもの、「移動」に関するものであった[共用品推進機構 2011]。安全歩行と移動は同じ種類と見なすと、視覚障がい者の方には「安全な移動」と「情報取得」を支援する機器のニーズが高く、それを解消することが視覚障がい者の方の支援として最も効果的であると推察した。

「視覚障がい者」のための「安全な移動」と「文字情報の取得」と聞いて思いつく支援は白杖と点字である。

白杖は視覚障がい者に携帯が義務付けられており、それを持つことで周囲に視覚障がい者であることを伝えると同時に、地面を打突することで、反響音や振動により聴覚や触覚を通して周囲の情報を得るための重要かつ有効な手段である。しかし、白杖では検知が難しい情報も存在する。



そこで、白杖では取得が困難な情報を得るための手法について検討する。白杖で取得が困難な情報には、白杖が届かない範囲の情報や細かい凹凸が挙げられる。

白杖が届かない範囲の情報に関しては、距離センサを用いたシステムが効果的であり、ガイドロボットの障害物検知などに利用されている。一方、細かい凹凸を検知することは容易ではない。先に挙げた距離センサは 10mm 程度の凹凸は検知が難しく、白杖で探ることも難しい。

10mm 程度の凹凸では視覚障がい者の安全に影響を及ぼさないと考えがちである。しかし、僅か 10mm 程度の段差が致命的な事故につながる可能性がある。高齢者の転倒しやすい場所について、内閣府がまとめた資料によると、高齢者のうち約 1 割の方が年 1 回以上の転倒を経験している[内閣府 2005]。転倒した場所に関しては、庭や居間、浴室といった大きな段差がほとんどない場所でもそれぞれ 26.5%、17.0%、9%の方が転倒していた。これは、わずかな段差での躓きや滑りやすい床面での滑りが転倒の原因の 1 つであることを示している。

しかも、我が国における視覚障がい者は、後天的に障害を受けた中途障害者の割合が先天的に障害を受けた方よりも多く、この傾向は昨今の少子高齢化の影響を受け、ますます顕著になっていくと考えられる。ただでさえ危険な転倒だが高齢者の場合足腰に障害が残ることもあり、ときには寝たきり状態になってしまうこともある。

このような転倒を防ぐため、小さな凹凸や危険な地面（地面性状）を識別することが必要である。つまり、地面性状を識別することができればつまずきや滑りの防止につながる。

点字は触覚を介して文字情報を伝達することができ、商品パッケージやエレベータ、券売機など様々な場所に設置されている。点字を習得している方にとっては、非常に有効な支援になる。しかし、中途失明者は点字を習得していない方が多く、視覚障がい者全体を見ても約一割の方しか習得していない[厚生労働省 2010]。

そこで、点字を習得していない方のために、点字の翻訳システムや 2 次元コードの認識システムが提案されている。しかし、我々の周囲にあるすべての物品や書籍に点字や 2 次元コードが付属しているわけではない。例えば、食品の包装用紙には点字が打刻していないものも多く、打刻してあってもその内容は簡易な情報であることが多い。なぜならば、点字は小さな範囲に大量の情報を与えることが難しいためである。QR コードなどに代表される 2 次元コードは小さな範囲に点字と比べて大量の情報を与えることが可能である。例えば、商品名のみならず賞味期限や値段なども得ることができ、しかも、スマートフォンを用いれば利用も容易である。このように、有用なシステムではあるが、すべての物品や書籍には対応しきれない。そこで、文字自体を認識するシステムについて検討する。

さて、一概に文字の認識と言ってもその応用例は広い。そこで、視覚障がい者の方に役に立つ文字認識システムに関して検討する。そのため、視覚障がい者が文字認識に関して感じる不便さについて調査した。

視覚障がい者の方と話す機会があり、普段の生活でストレスを感じる場面についてお話を伺った。その際、味の異なるポテトチップスがあったときに、好みの味を選ぶことができないことがストレスであると聞いた。ポテトチップスは味毎に手触りの差が全くなく、重さの差も、揺らした際の音も、温度にも違いがない。このような物品の外装を見分けることは視覚障がい者にとっては困難である。

これは共用品推進機構が行った「同じような外形で区別しづらい包装用紙」に関するアンケート調査の結果からも明らかである[共用品推進機構 2011]。

Table 1 同じような外形で区別しづらい包装用紙

物品	全盲	弱視
箱・袋・パック入り食品・冷凍食品	62.7%	54.4%
缶詰め食品	60.4%	50.8%
缶・瓶・紙パック飲料	55.7%	49.7%
詰め替え用のもの	54.9%	53.8%
調味料	52.4%	48.7%

この問題に対して、文字自体を認識し、読み上げを考える。しかし、商品や書籍の表面には多数の文字が書かれているため、書かれている文字をすべて読み上げても情報過多による混乱につながる。

以上、本研究では視覚障がい者の日常生活支援システムとして、地面性状の識別と文字情報の取得手法について検討を加える。

## 1.2 先行研究

背景で述べたように視覚障がい者の日常生活支援として、地面性状の識別と文字情報の取得が重要である。そこで、これらの先行研究について調査し、研究方針を検討する。

### 1.2.1 地面性状識別

地面性状識別に関しては、地面の感触を振動や圧力情報として捉え識別する手法や投影した光の反射パターンの違いを捉えて識別する手法、画像中のパターンの違いを捉え識別する手法と、3種類の識別手法が提案されている。

第1の手法として、ロボットに加速度計を設置しておき、地面の上を走行したときの振動の違いを識別することで、地面性状を識別する手法が提案されている[Weiss 2006]。しかし、このように地面の感触を測定する手法は、精度が高いという利点があるが、対象の地面に接近、進入する必要がある。

第2の手法として、地面へ赤外光を照射して、その反射光をカメラで撮影し、反射光の違いから地面の種類を識別する手法が提案されている[Lu 2009]。この手法は対象の手前から地面を識別できるだけでなく、地面の濡れや照明条件の変化にも頑健であるという利点がある。しかし、対象と離れた位置からカメラに捉えられる程度の強い赤外線を照射する必要があり、機器の大型化する可能性が高い。システムを身体に装着することを考慮すると、機器の大型化は操作者に与える負担が大きくなることが懸念される。

第3の手法として、ロボットに装着したカメラで撮影した画像から画像のパターンを学習し、地面性状を識別する手法が提案されている[Zenker 2009]。この手法は対象の地面に進入することなく、その手前から認識できる。しかし、1フレームの画像内には1種の地面しか存在しないと仮定しているため、複数の地面が混在した場合にも対応した手法が必要である。しかも、地面を形状情報として捉えているため、なだらかな平面のように特徴のない地面を見分けることが難しい。

これらの先行研究から地面性状の識別手法として、画像処理による手法が多くの特長を持っていることが分かる。画像処理を用いた手法では、処理フレーム内に多数の地面性状が混在する可能性が高く、その対策が必要である。また、画像処理により識別するためには、地面性状ごとの特徴を捉える必要があり、そのときに捉える特徴についても検討する。

## 1.2.2 文字情報取得

視覚障がい者の文字の取得支援には、前述したとおり、点字の認識や2次元コードの認識が提案されており、一部普及している。しかし、実際にはほとんどの文字情報は活字でのみ表記されている。そのため、文字自体を認識し、読み上げる手法が障害の程度によらず、最も有効であると考えた。

文字の認識手法としては、大量の文字画像を用意し、それを学習することで画像を認識する手法が提案されている[Shi 2016]。用意する画像の種類や多様さによっては、様々な背景や照明条件下の文字も認識でき、特徴量を工夫すれば文字の変形にも頑健になる。しかし、学習用の画像は数万枚や数10万枚と大量に必要であり、学習画像が少ない場合は汎用性が失われる過学習という現象を起こしやすい。

このように、文字の認識手法は様々な手法が提案されており、実用化されているシステムもある。そのシステムの1種として、Optical Character Recognition(OCR)システムが開発・販売されている。OCRはスキャナで読み込んだ書類等に印刷された文字を高い精度で認識することができる。しかし、OCRはスキャナで読み込んだ照明条件などの整った画像の入力を前提としており、また、読み込まれた背景を文字と誤認識することも多い。本研究では、認識についてはOCRを用いることを考え、OCRで認識するために複雑な背景や様々な照明条件から文字を抽出する手法について考える。

Gotoらは、環境中の文字を抽出し、読み上げるウェアラブル型支援システムを提案した[Goto 2009]。視認性を高めるために、文字と背景の輝度値に差が出やすい色が文字には使用される。そのため、文字と背景の境では急激に輝度値が変化することになり、高周波成分が強く検出される。この高周波成分はDigital Cosign Transform(DCT)特徴によってその強さを測定することができる。したがって、撮影映像内のDCT特徴が強い領域を抽出することで文字を抽出することができる。このシステムは高周波成分が文字以外に発生しないことを前提としている。すなわち、背景が単色の壁や紙の場合を想定している。しかし、実環境では文字以外の高周波成分が強い領域が抽出されることが想定されたため、それに対応する必要がある。

平岩らは、視覚障がい者の支援システムとして、環境中の文字情報を利用した道案内システムを提案した[平岩 2002]。前述の通り、背景との境において輝度値の差が大きくなるように文字は書かれている。すなわち、画像中から輝度値が急激に変化する領域「エッジ」を抽出しやすい。そこで、エッジが検出される領域を文字領域として抽出する。しかし、エッジが文字領域以外にも出現することがある。そのため、実環境での使用を考慮して、抽出した文字候補に対して、エッジの密度やバランス、文字の大きさなどに閾値を設定し、それらの文字領域以外のエッジ、すなわちノイズを除去している。

このように、文字が持つ特徴に着目することで環境中から文字を抽出することができる。ここで、文字が正確に抽出できた際に発生する問題として、多量の文字列（単語）の抽出が挙げられる。つまり、乱雑に存在する文字に対して文字を抽出した場合、OCRは複数の認識することになる。すなわち、システムは文字を連続で読み上げてしまう。この連続読み上げが発生すると、視覚障がい者へ正しく情報が伝わらず、混乱を招くことが懸念される。例えば、スーパーの食品棚には多量の商品が並んでいる。もしも棚の上に1つの商品しかなく、「ポテトチップス」と伝えれば、ポテトチップスの存在をはっきりと認識することができるだろう。しかし、「せんべい」、「チョコレート」、「ポテトチップス」…等々と延々と読み上げた後、そこに何があったのか判別することは容易ではない。

このように、視覚障がい者が大量の文字の読み上げによって混乱しないためには、読み上げる文字の量を限定することが重要である。しかし、単純に量を絞ってもうまくいかない。なぜならば、視覚障がい者にはどこの文字が読み上げられるか全くの未知であるためである。また、文字は点字と異なり、書かれている場所を判断することができない。

そこで、読み上げる文字を限定するために、指先を使用することを考える。晴眼者でも暗闇で手先の感覚だけで物品を探し当てることは可能であり、その際に物体の形状や材質、温度などを把握することができる。視覚障がい者もペットボトルや缶、本、手紙を触れたとき、自身がどのような物品に触れているか把握することができる。この際、視覚障がい者は触れている物品のどこに文字が書かれているか推測することができるので、推測した場所の文字を読み上げれば、その物品の詳細を知ることができる。

例えば、ある視覚障がい者が「塩味のポテトチップス」を探している場面を想像する。その視覚障がい者は手の感触から自宅に置いてあった「ポテトチップス」の袋を発見した。しかし、袋の手触りやにおい、音、重さでは味まで調べることはできない。このような場面では袋の文字を読み上げることが視覚障がい者への助けになることが考えられる。ここで、注意点として視覚障がい者は周りの様子が分からないので、袋と一緒に壁や他のパッケージなどが映り込んでしまうことが考えられる。そのような場合に、袋上の文字が書かれていそうな場所を指定することができれば、より文字情報を正確に伝達することができると考えられる。

以上、本論文では「地面性状識別」および「指先指定文字抽出」に関する視覚障がい者のための日常生活支援システムについて提案する。

### 1.3 目的

1.1 節で挙げた視覚障がい者のための歩行支援, 文字情報取得支援として解決すべき事項をまとめる.

これまでに挙げた視覚障がい者のための安全な歩行支援のために, 障害物検知や誘導システムが提案・研究されてきた. しかし, 安全な歩行のためには白杖や距離センサでは識別が難しい地面の性状の情報も必要である. 本論文では, 視覚障がい者の安全な歩行支援のための地面性状識別システムについて考える. その実現のためには, 以下のような機能が求められる.

- 対象の地面に進入することなく識別
- 複数の地面性状が入り混じった場合の対処

また, これまでに挙げた視覚障がい者のための文字情報取得支援の問題点を解消するためには, 複数の文字が混在する環境から特定の範囲の文字を抽出することが必要である. そこで, 指先を用いて文字を抽出する範囲を限定するシステムについて考える. その実現のためには, 以下のような機能が求められる.

- 様々な照明環境下で安定した文字の抽出
- 複数の文字がある環境下で限定された領域内の文字認識
- ぶれることが考慮されるカメラで撮影した画像を用いた安定した指先検出

本研究では, これらの要素を満足させた視覚障がい者のための地面性状識別システムと指先指定文字抽出システムについて検討を加える.

そこで, 本研究の目的は,

- 画像処理を用いた地面性状識別手法の提案
- 様々な照明環境下で多数の文字が混在する環境における, 物体表面上の文字抽出・音読手法の提案

とする.

## 1.4 論文構成

本論文の構成は全7章で構成される(Fig. 1).

第1章では、視覚障がい者のための安全な歩行の支援と文字認識支援における問題点と既存の支援手法をまとめ、本論文の目的を明らかにした。

第2章では、視覚障がい者のための安全な歩行の支援手法と文字取得支援手法について提案されている従来研究を調査し、それぞれの従来手法では取得が困難な地面の情報があることと、複数の文字が存在する環境では読み上げる文字を減らす必要があることを明らかにする。その結果、明らかになった各問題点に対して、それを解決するためのアプローチ「地面性状識別」と「指先指定抽出手法」を提案し、概説する。それぞれのアプローチの詳細については、第3,4,5章で詳述する。

第3章では、2章で提案した歩行支援のためのアプローチである「地面性状識別」について詳述する。第4章では提案した地面性状識別手法の性能を、実験を通して確認し、得られた知見を述べた。

第5章では、2章で提案した情報取得支援のためのアプローチである「指先指定抽出手法」について詳述する。第6章では提案した指先指定抽出手法の性能を、実験を通して確認し、得られた知見を述べた。

第7章では、視覚障がい者のための日常生活支援システムについて、本研究の総括を行い、今後の展望を示す。

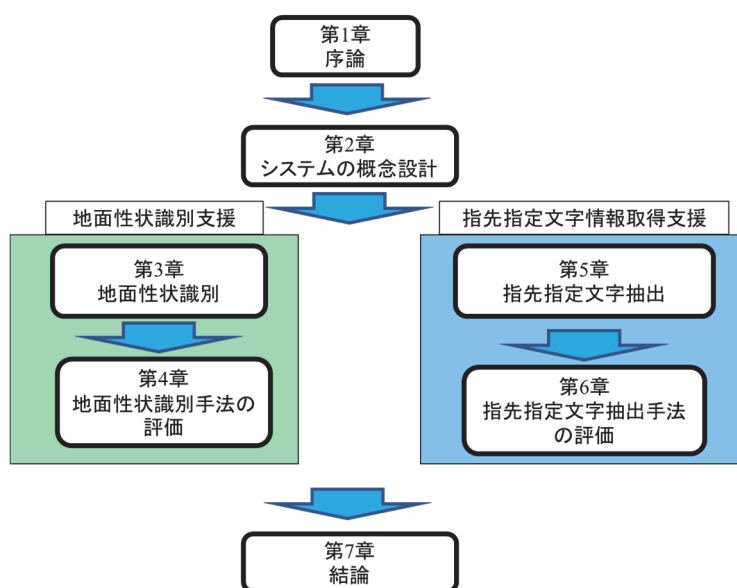


Fig. 1 本論文の構成





## 2. 提案システムの概要

---

2.1 はじめに .....	12
2.2 視覚障がい者の定義 .....	14
2.3 従来研究 .....	16
2.3.1 誘導に関する研究 .....	16
2.3.2 障害物検知に関する研究 .....	18
2.3.3 地面性状の識別に関する研究 .....	20
2.3.4 文字情報認識に関する従来研究 .....	25
2.3.5 文字抽出に関する従来研究 .....	29
2.3.6 領域の指示に関する従来研究 .....	33
2.4 従来研究のまとめ .....	37
2.5 提案する視覚障がい者のための地面性状識別システムの概要 .....	41
2.6 提案する視覚障がい者のための文字取得支援システムの概要 .....	43
2.7 おわりに .....	45

---

## 2.1 はじめに

第2章では、1章で明らかにした目的を達成するためのアプローチについて従来研究の調査を通じて、改めて検討する。本説の構成は、2.2.1-2.2.3項で、視覚障がい者の安全な移動を支援するシステムに関して紹介、評価し、2.2.4-2.2.6項で、視覚障がい者の文字取得を支援するシステムに関して紹介、評価する。

### 2.2.1 項：誘導に関する従来研究

視覚障がい者の安全な移動を支援するためには、道迷いを防ぐ必要があり、そのための視覚障がい者の誘導手法について調査する。誘導とは、視覚障がい者の自己位置と目的地の関係から進行方向を与えることである。この調査を通じて、道迷いを防ぐための支援システムがすでに提案されていることを示す。

### 2.2.2 項：障害物検知に関する従来研究

視覚障がい者の安全な移動を支援するためには、障害物の回避が必要であり、そのための周囲の障害物検知手法について調査する。障害物の検知とは、距離センサ等により障害物までの距離や周囲の形状を検知し、その情報を与えることである。この調査を通じて、障害物との衝突や転落を防ぐための支援システムがすでに提案されていることを示す。

### 2.2.3 項：地面性状識別に関する従来研究

視覚障がい者の安全な移動を支援するためには、地面性状の識別が必要である。そこで、地面性状識別手法について調査する。地面性状の識別とは、距離センサでは取得することが難しい微細な地面のパターンを識別することである。視覚障がい者の支援のためには、この微細な情報を対象の地面の手前から取得する支援が必要であることを示す。

### 2.2.4 項：文字情報認識に関する従来研究

文字情報の取得を支援するためには、環境中の文字情報を認識する必要がある。文字情報は、点字や2次元コード、文字の形式で表現されており、それぞれの形式ごとに認識手法が提案されている。3種の形式のうち、点字や2次元コードは環境中に少なく、文字から直接情報を得るべきことを示す。文字は様々な学習手法を用いて認識されるが、十分な能力を発揮させるためには大量の画像の収集が必要である。文字の認識用アプリケーションとして、OCRが開発・販売されており、本研究目的においては、文字の認識よりも文字の抽出が重要であることを示す。

### 2.2.5 項：文字情報抽出に関する従来研究

前節で紹介したOCRを用いて文字を認識するためには、文字の抽出が必要である。OCRは文字の認識に有効であるが、スキャナによる文章の読み取りが必要なことや、認識対象が書籍に限られることへの対策が必要である。そのため、撮影した身の回りの文字を直接OCRに認識させても文字の認識精度は低下してしまう。そ

ここで、撮影した画像から OCR で認識しやすいように文字を修正して抽出することが必要であることを示す。さらに環境中には、複数の文字が混在していることもあり、商品や書籍の表面上の文字をすべて抽出し読み上げた場合、情報過多による混乱を招くことを示す。

#### 2.2.6 項：文字抽出領域の限定に関する従来研究

複数の文字が混在する環境から文字情報を明快に伝達するためには、抽出する文字の領域を限定することが必要である。本節では、文字抽出領域の限定方法について検討する。指定方法として、特定の文字を用いた手法やマーカを用いた手法、指先位置を用いた手法が提案されているが、指先位置を用いた手法が有効であることを示す。

## 2.2 視覚障がい者の定義

本研究で取り扱う視覚障がい者に関して定義する。視覚障害の大きな分け方として、全く視力を失っている完全失明（盲）と、多少は見える弱視に分けられるが、本研究ではどちらも利用が可能な支援システムを提案する。すなわち、最も条件の厳しい完全失明者を対象としたシステムであれば、どちらの視覚障がい者にも利用できるシステムとなる。

法律的には、視覚障害はメガネやコンタクトレンズを使用した左右の矯正視力の合計と視野の広さから、その等級が次のように定義されている(Table 2)。健康診断などの視力検査によく用いられるランドルト環(Fig. 2)では、最も大きい環が 5m先から 0.1 の視力で確認できるものになっている。つまり、矯正視力が 0.1 未満となると、5m先の信号機や駅の案内など周囲の環境を確認することが難しくなり、日常生活に大きな障害となる。また、視野に問題がある場合も日常生活における障害になる。

Table 2 視覚障害等級([厚生労働省 2009])

1 級	両眼の視力（万国式試視力表によって測ったものをいい、屈折異常がある者については、きょう正視力について測ったものをいう。以下同じ。）の和が 0.01 以下のもの
2 級	1. 両眼の視力の和が 0.02 以上 0.04 以下のもの 2. 両眼の視野がそれぞれ 10 度以内でかつ両眼による視野についての視能率による損失率が 95 パーセント以上のもの
3 級	1. 両眼の視力の和が 0.05 以上 0.08 以下のもの 2. 両眼の視野がそれぞれ 10 度以内でかつ両眼による視野について視能率による損失率が 90 パーセント以上のもの
4 級	1. 両眼の視力の和が 0.09 以上 0.12 以下のもの 2. 両眼の視野がそれぞれ 10 度以内のもの
5 級	1. 両眼の視力の和が 0.13 以上 0.2 以下のもの 2. 両眼による視野の 2 分の 1 以上が欠けているもの
6 級	一眼の視力が 0.02 以下、他眼の視力が 0.6 以下のもので、両眼の視力の和が 0.2 を超えるもの

視覚障害には、上記の視力や視野に関する障害の他にも色を見分けることが難しい色覚障害や明るさを調整することが難しい光覚障害もあるが、本研究では視力や視野に関する視覚障害を取り扱う。

また、先に述べたように視覚障がい者は大きく盲と弱視者の2種に分けられる。この際、視覚障がい者を区別するとき、実際の障害の度合いで区別するのではなく、受けてきた教育、すなわち、視覚情報を他の感覚器官から主に得ているか、自身の視力を様々な支援機器によって矯正した視覚から得ているのかで区別することが適当である[青柳 2015]。中途障がい者は視覚情報を他の感覚器官(触覚や聴覚など)からの情報として受け取り、補間することが難しいとされている。例えば、振動や音階などで周囲の環境を伝達することも可能であるが、中途障がい者には視覚で得ていた情報を効率よく他の器官で伝えることが重要である。

以上をまとめると、本研究における「視覚障がい者」とは、盲や弱視者など障害の程度によらず、障害を受けた時期も区別しない、視野や視力に問題がある方とする。

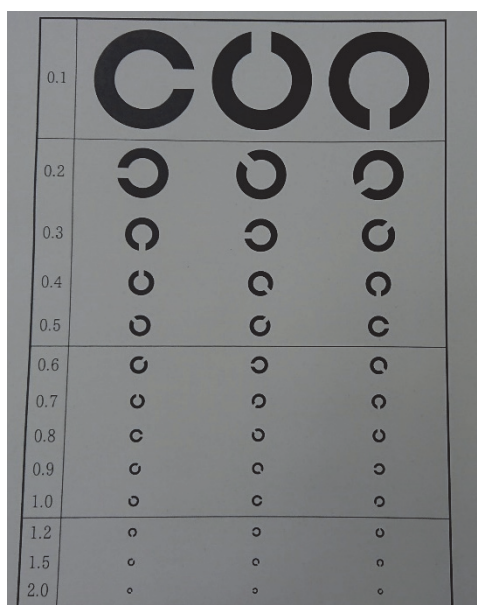


Fig. 2 ランドルト環

## 2.3 従来研究

従来研究として、下記に挙げる 6 種の項目について調査する。

### 歩行支援

- 誘導に関する従来研究
- 障害物検知に関する研究
- 地面性状の識別に関する研究

### 文字情報取得支援

- 文字情報認識に関する従来研究
- 文字抽出に関する従来研究
- 領域の指示に関する従来研究

### 2.3.1 誘導に関する研究

視覚障がい者の安全な移動を支援するために、道迷いの防止が挙げられる。そこで、視覚障がい者の誘導システムを紹介する。

視覚障がい者の道迷いを防ぐため、GPS による視覚障がい者の案内システムが提案されている[Soeda 2004]。視覚障がい者は外出する際に、頭の中に地図を思い浮かべながら移動する。しかし、初めての目的地に向かう際には、地図を思い浮かべることが難しいので視覚障がい者は多大なストレスを受けるだけでなく、道迷いの危険がある。2017 年現在、GPS を用いたナビゲーションシステムが普及しており、初めての場所を訪れる際に大いに役立っている。このシステムを転用した場合、視覚障がい者にも有効であると考えられる。

屋外の誘導システムでは、すでに GPS を利用した有効な手法が提案されている。しかし、GPS は屋内で使用した場合、精度の低下が発生することがある。

GPS の使用が難しい環境の場合、案内装置を設置し、視覚障がい者を誘導することが考えられる。Kim らは、環境に設定した音声案内装置による視覚障がい者ガイドシステムを提案した[Kim 2009]。予めレストランやトイレ、非常口などに音声案内装置を設置しておき、受信デバイスを装着している視覚障がい者が近づいたとき、その音声案内機ごとに持っている情報を伝える。そのうえ、デバイスを装着した人にしか案内が行われないので、晴眼者に対する騒音の問題も解決している。

また、Liao らは、視覚障がい者の安全な歩行のために、地面に埋め込んだ RFID タ

グとそれを読み込む改造白杖を用いた誘導システムを提案した [Liao 2013]. 周辺の建物や信号などの情報を埋め込むことで、誘導だけでなく周囲の情報も視覚障がい者へ伝達することができる.

これらの手法の利点としては、屋内でも使用できる他、GPS よりも誘導の精度が高く、屋外・屋内問わずに 1m 以下の精度で誘導が可能という利点がある. このように案内装置や RFID を用いた手法は案内のために非常に有効ではある. しかし、あらかじめインフラの整備が必要なので、設備が整っている場所でないと使用できない.

インフラの整備なしの誘導システムとして、ロボットによるガイドが考えられる. 例えば、商業施設に視覚障がい者の方が訪問したとき、ガイドが目的地まで誘導することが考えられる. このとき、人件費や労働時間の増加を考慮すると、ロボットが代わりに誘導できれば有用である. 日本精工は視覚障がい者のために、障害物自動回避機能を持った誘導ロボットを開発した[日本精工 2015]. このロボットは視覚障がい者が取手に手を乗せた状態で走行し、目的地まで牽引しながら誘導する. また、このロボットは、誘導だけではなく障害物回避機能も有する. 例えば、誘導の際には人や想定していない障害物などが通路上に存在することが考えられる. そこで、ロボットに障害物を回避するための距離センサを内蔵し、誘導ルート内に障害物が現れた場合も回避しながら誘導を続行する. ロボットによる誘導システムは視覚障がい者の誘導として有効であるが、その構造上段差の上り下りが難しく、購入資金や使用方法の熟知、メンテナンス等を含めた導入コストも高いため一般的な誘導手法としてはまだまだ敷居が高い.

このように、誘導ロボットの導入には問題があるが、その問題の大部分はロボットという大型のモジュールを用いていることが原因である. そこで、誘導時に情報だけを与えることを考える. Sekiguchi らは相対的な位置情報の伝達により、視覚障がい者を誘導するシステムを提案した[Sekiguchi 2014]. 屋内では、出発地点から目的地までに存在する分岐路ごとに進行方向を指示することで、十分誘導が可能である. 測域センサによって、ソナーのように周囲の壁の位置を把握し、分岐路を発見する. 分岐路を発見したとき、予め所持している経路情報に従い目的地への方向へ誘導する. このシステムでは、通路の形状を時系列データとして確認することで、自己位置を推定している. しかし、分岐路の誤検出や指示以外の方向に移動を行ったとき、自己位置の推定に失敗することがある.

そこで、Bai らは、ランドマークを利用した視覚障害者のための誘導手法を提案した[Bai 2014]. ランドマークとの相対位置により、自己位置の推定ができる. また、認

識したランドマークの情報として、ランドマークの大きさから距離や予め登録しておいた階数や部屋名の情報も得ることができる。予め撮影したランドマーク画像と撮影した画像からそれぞれ SIFT 特徴量を抽出し、一致する場所を画像中から探すことでランドマークを検出している。このように、屋内において効果的に視覚障がい者の道迷いを防ぐことができる。

このように、視覚障がい者の安全な移動を支援するための様々な誘導システムが提案されており、屋外、屋内の視覚障がい者の誘導において成果を挙げている。

### 2.3.2 障害物検知に関する研究

2.2.1 項で紹介したように、視覚障がい者の安全な移動のための1手段である誘導に関しては、成果が上がっている。しかし、安全な移動のためには道迷いを防ぐだけではなく、障害物との衝突や段差からの転落も防がなくてはならない。

神籾らは、視覚障がい者のホームからの転落を防止するために、音声による注意喚起システムを提案した[神籾 2001]。駅のホームの縁に警告領域を設置しておき、その範囲に利用客が進出したとき、指向性の高いスピーカから警告を発する。デバイスが環境中に設置されているので、利用者の白杖等の使用を阻害することがない。また、装置が環境内に固定されていることで警告領域内の検出時に環境の変化が少なく、検出精度の低下も発生しにくい。しかし、このように環境中にセンサを設置した場合、危険な場所を網羅するよういくつものセンサを準備しなくてはならないため、設置の手間がある。また、このシステムは障害物自体に取り付けているとも考えられるが、屋外、屋内のすべての危険な場所にこのようなシステムを取り付けることは現実的ではない。そこで、ユーザの周囲の危険を検知するシステムが求められる。

センサを視覚障がい者の周りに取り付け、周囲の危険を検知するシステムとして、Ulrich は、前方の広い範囲の距離を調べられるように超音波センサを並べたロボットを白杖の先に装着する Guide Cane を提案した[Ulrich 2001]。このシステムは、視覚障がい者と障害物の距離を常に測定しておき、障害物に接近した場合に、検知したセンサの位置と進行方向からそれを避けるように視覚障がい者を牽引する。

また、Takizawa らは視覚障がい者の段差や障害物を識別するデバイスとして、白杖を改造した Kinect Cane を提案した[Takizawa 2015]。白杖に距離センサを装着し、任意の物体までの距離を3次元計測する。視覚障がい者には物体までの距離に応じた周波数でデバイスを振動させて注意を喚起する。また、前方の3次元情報から空間中の大まかな物体の位置や形状を調べ、何もない床面や椅子、上り段差や下り段差を識別す



ることができる。

しかし、これらのシステムは白杖自体に装着するので、白杖の重量の増加することや、白杖の使い勝手が変わることによって、普段と同じような使い方が困難になる恐れもある。また、後者のシステムは、ユーザが空間内の物体をある程度把握できることを前提としているため、中途視覚障がい者が利用することは困難であると考えられる。

そこで、白杖ではなく身体に支援装置を装着することを考える。Uedaらは視覚障がい者の階段からの転落を防止するために、測域センサを用いた段差検出手法を提案した[Ueda 2006]。視覚障がい者の身体に2次元測域センサを装着し、そのセンサを回転させることで3次元計測する。このような測域センサは赤外線を用いて空間一面に照射し、それが反射して戻ってくるまでの時間を計測することで、前方の障害物までの距離を計測することができる。照射した赤外線ごとに照射した方向と距離が分かるので、3次元空間中に周りの障害物を点群として表現することができる。この空間をブロック状に分割して考え、ブロック内に点群が一定数以上存在する場合、そこに障害物があると判断する。そして、障害物と判定されたブロックの並び方から段差の有無を判別する。身体に装着することで、視覚障がい者の白杖の併用を阻害することが少なく、また、広範囲の障害物の有無を調べることが可能である。しかし、身体に装着する場合、センサが大きく動作した場合、視覚障がい者の他の動作を阻害してしまうことが懸念される。そこで、視覚障がい者の行動の邪魔になりにくいセンサが提案されている。

Ishiwataらは、2次元距離センサを用いた段差検出手法を提案した[Ishiwata 2013]。視覚障がい者の胸部に2次元距離センサを装着し、視覚障がい者の立っている床との高さの差を調べ、その変化によって下り段差や上り段差を識別する。センサが胸部にあるおかげで白杖と共用することができ、また白杖を自由に扱うことができる。

このようなウェアラブル型の障害物検知デバイスは多数提案されており、その形状も服型、メガネ型、胸部装着型と多様である。

例えば、Bahadirらは、視覚障がい者が障害物を認知するために、全面に距離センサを装着したウェアラブル型デバイスを提案した[Bahadir 2012]。このウェアラブル型デバイスは、服型であり、着衣のように装着でき、視覚障がい者の行動を阻害することがほとんどない。障害物までの距離に応じて装置内に配置されたモータが強さを変化させながら振動することで、障害物の有無を感じ取れるようになっている。

Setiawan らは、メガネ上部に設置した超音波センサと振動モータを組み合わせた誘導眼鏡を開発した[Setiawan 2014]. 超音波センサが測定した障害物までの距離に応じて振動数を変化させる。これにより、視覚障がい者は障害物への接近を認識することができる。

Cardin らは、身体に取り付けた複数の超音波センサの出力値より振動の強さを変化させる手法を提案した[Cardin 2007]. 発見した障害物の位置に近い超音波センサの出力値に大きい重みを設定し、計算した出力値の合計から振動の強さを決定する。

このように、視覚障がい者の安全な移動を支援するための様々な障害物検知システムが提案されている。障害物の種類も様々であるが、ここで紹介した全てのシステムは測距センサを用いている。衝突や転落が発生する障害物の検知には測距センサが測定範囲、測定精度ともに優れていることを示している。

### 2.3.3 地面性状の識別に関する研究

2.2.1 項, 2.2.2 項で紹介したように、様々な誘導システムや障害物検知システムが提案されている。しかし、視覚障がい者の安全な移動のための支援としては、白杖では分かりにくい情報を得ることも重要である。例えば、地面の性状を白杖で識別することは難しく、すべりやつまずきにつながる。その結果、転倒や転落など大きな事故が起こる可能性もある。

地面性状識別に関する研究には、測距センサを用いた手法や地面を移動した際の振動や圧力のパターンを調べる手法、画像を用いた手法が挙げられる。しかし、測距センサでは測定精度の限界から表面の凹凸のような微細な変化を捉えることが難しい。振動や圧力のパターンは、微細な変化も捉えることが可能であるが、今まさに歩いている地面の性状しか識別ができない。すなわち、危険な性状を持つ地面であっても進入する必要がある。そして、画像を用いた手法は危険な地面に進入する必要はないが、精度を保つことが難しい。ここでは、上記の3種の手法について調査した。

#### (1) 測距センサを用いた地面性状識別手法

今井らは、3D 測距センサを用いて前方の深度情報を取得し、足元から続く平面を安全な歩行領域として検出する手法を提案した[今井 2017]. 50 mm以上の凹凸を駅のホームや階段など、転落の恐れがある段差として認識し、危険領域として視覚障がい者の方の伝達することを想定している。

また、Leung らはサングラス型のディスプレイに 2 個のカメラと 1 個の IMU(Inertial Measurement Unit)を装着し、ステレオビジョンによって深度情報を得る、視覚障がい者の歩行支援システムを提案した[Leung 2014]. 視覚障がい者が立っている場所は安全な平面と見なし、深度情報の時系列データから時間ごとに変化していく地面性状のうち、連続して続いている安全な地面性状を捉えることで、危険な場所を識別して視覚障がい者に伝達することができる。

しかし、これらの測距センサやステレオビジョンを用いた手法では測定精度の限界から表面の凹凸のような微細な変化を捉えることが難しく、そのため地面の性状を識別することは困難である。先に述べたように、人は数十 mm の高さで躓くこともあるため、その凹凸に対応できる手法が必要である。

## (2) 振動や圧力を用いた地面性状識別手法

Weiss らは、車両型ロボットが安定して走行するために、地面性状を識別する手法を提案した[Weiss 2006]. 屋外を走行することを考えたとき、地面性状ごとに滑りやすさは異なる。そこで、地面性状に合わせて速度や走り方を調整することができれば、走行は安定し、効率的になる。提案されている手法は、予め、ロボットの車体に取り付けられた加速度計から車体の振動を調べ、その振動パターンを地面性状ごとに記録し、データベースを構築する。加えて、サポートベクターマシンを用いて、振動のパターンデータベースから識別器を構築する。そして、構築した識別器を用いて、7 つの地面性状(ポリ塩化ビニルの廊下、アスファルト、砂利、芝生、石畳、土、振動ナシ)を識別した。砂利こそ 66.83%の精度であったが、その他の地面性状に関しては、90%以上の精度で識別が可能であった。

車輪型だけでなく、歩行型のロボットについても地面性状識別手法が提案されている。Best らは、6 脚のクモ型ロボットの安定した歩行のために、地面性状識別手法を提案した[Best 2013]. 提案された手法は、ロボットが地面性状を歩いた際の情報に基づいて、地面性状を識別する。6 本の脚にはそれぞれサーボモータが取り付けられており、その指令した位置と実際の位置は地面性状ごとに異なる。そのため、この差を学習することで地面性状を識別することが可能になる。Best らは、識別器の構築にサポートベクターマシンを用いている。

構築した識別器を用いて、5 つの地面性状(砂利、芝生、小枝、芝生、コンクリート、振動ナシ)を識別した。砂利と小枝の識別が難しいという問題があるが、その他の地面性状に関しては、95%以上の精度で識別が可能である。

振動を用いる手法以外にも例えば、圧力を用いた手法も提案されている。Ascari らは、昆虫型の6脚ロボットの安定した歩行のために、ロボットの足裏に取り付けられた圧力センサとロボット上部に取り付けられたカメラを用いた地面性状識別手法を提案した[Ascari 2006]。歩行したときの足裏にかかる圧力の違いを学習し地面性状を識別することが可能である。

このように、車両やロボット移動時の振動や圧力のパターンは安定しているので、振動や圧力のみで地面性状を調べることも可能である。しかし、特に障がい者の場合、歩行時の振動パターンに個人差が大きいことや、体調の変化によってもパターンが変化することが考えられる。そのうえ、これらの手法は対象の地面に進入して始めて識別が可能になるので、人に適用することを考慮すると、安全面から問題がある。

そこで、前方を撮影し、画像のテクスチャ特徴を解析することで識別することが考えられる。

## (2)画像を用いた地面性状識別手法

先に挙げた Ascari らが提案した識別手法では、圧力以外に画像を用いた手法も併用する。撮影した画像からテクスチャ特徴を抽出し、データベースを構築し、SVM(Support Vector Machine)で識別器を作成した。しかし、テクスチャ特徴のみでは精度に不安があるため、2つの認識手法を併用することで精度の改善を試みているが、圧力センサを用いた認識はやはり地面に進入する必要がある。

画像を用いた手法では地面性状を識別するための特徴量の選定が重要となる。そこで、画像内から抽出が容易な特徴を選定し、危険を知らせる目印にすることが考えられる。

例えば、転落の危険がある場所には点字ブロックがあることを前提とし、森らは画像による点字ブロック認識手法を提案した[森 2004]。点字ブロックは視覚障がい者の安全のために危険な場所や重要な場所へ誘導するために設置されている。例えば、駅ホームにおいてはその縁に点字ブロックが並べてあり、視覚障がい者の転落を防止することに役立っている。すなわち、危険な場所を見分ける目印として活用することができる。

点字ブロックの色は周りの地面と比べて目立つ配色になっていることが多く、点字ブロックの色を抽出することで点字ブロックを検出できる。誘導の際には、点字ブロックを目印(ゴール)として、自己位置と方向を万歩計と地磁気センサを組み合わせる

ことで推定し、点字ブロックの位置を伝えることも可能である。しかし、色のみを特徴として点字ブロックを検出した場合、似た色が多数のノイズとして抽出されることが懸念される。

また、抽出が容易な特徴的なテクスチャを目印とする場合もある。

Wang らは、RGB-D センサを用いて、階段や横断歩道が持つ特徴(平行な直線が並んでいること)と距離情報を併用して、階段や横断歩道を検出する手法を提案した[Wang 2014]。画像情報と距離情報を併用することによって、階段だけでなく、視覚障がい者に必要な情報の1つである横断歩道の検出も可能である。

このように、特徴的な画像が取得できれば、測距センサでは難しい地面の状態を調べることが可能である。しかし、環境中のすべての地面が特徴的な画像として撮影できるとは限らないため、その対策が必要となる。

カラー画像にはノイズが入りやすいため、赤外線画像を用いた手法も提案されている。

Lu らは自動車の自動操縦のための地面性状識別手法を提案した[Lu 2009]。地面性状情報を画像から取得した場合、天候や照明条件の変化による地面の色の変化を考慮しなくてはならない。例えば、濡れた地面と日陰の地面、日向の地面でまったく同一の場所であっても画像の特徴は変化してしまう。そこで、赤外線を照射し、その反射光をカメラで撮影し、地面性状の識別に用いた。撮影した画像から空間周波数応答とテクスチャを抽出し、それを特徴として、データベースを構築した。その結果、アスファルト、芝生、砂利、砂の4つ地面性状を90%以上の精度で識別することができた。

しかし、赤外線を用いると、装置が大型化してしまう。ロボット用のデバイスとしては問題ないが、視覚障がい者への装着を考慮すると、装置はなるべく小型化する必要がある。そのため、画像のみを用いた手法が使いやすさの面では最も優れている。しかし、安定した識別のためには特徴量の工夫が必要である。

Zenker らは、ロボットの安定した歩行のために、画像を用いた地面性状識別手法を提案した[Zenker 2009]。撮影した画像を用いてデータベースを構築し、そのデータベースを学習して識別器を作成する。画像をデータベースを構築する際、画像をベクトルとして表現できると、サポートベクターマシンによる学習が簡便になる。そこで、撮影した画像を Bag of Features(BoF)を用いて、ベクトルで表現する。

BoF で表現するための特徴には Scale Invariant Feature Transform (SIFT)を選択した。識別した地面性状は芝生，砂利，舗道，砂，アスファルト，床，泥，小さな砂利の 8 つの地面性状であり，最大 90%の精度で識別が可能である。しかし，画像全体を 1 つのベクトルとして表現しているので，1 枚の画像内に複数の地面性状が入り込むと識別が難しい。また，形状の特徴である SIFT のみを使用しているため，似た模様を持つ，もしくは何の模様も持たない地面性状を識別することは難しい。

Bajracharya は，複数の特徴から色特徴を選択した[Bajracharya 2008]。提案された地面性状識別手法は，遠距離の地面性状を識別するためのものであり，離れた位置の形状特徴は劣化してしまうので取得することは困難である。一方，色情報は多少離れてもほとんど変化しない。色情報のこの特性は，わざわざ遠距離の地面性状を学習するための画像収集を必要とせず，近距離の地面の画像を学習データとして用いることを可能としている。

そこで，画像を小領域に分割し，それぞれの領域を色ヒストグラムで表現した。色ヒストグラムはベクトルと見なせるため，これを基にデータベースを構築し，SVM を用いて学習した。しかし，色は照明条件の変化によって様々に変化してしまうため，色特徴のみで地面性状を識別することは難しい。

2017 年現在，深層学習による画像認識が強力なツールとして認知されている。Ros らは，自動車に搭載したカメラで撮影した画像を収集し，走行時の semantic segmentation 用のデータセットを作成した[Ros 2016]。作成したデータセットに複数の既存のアーキテクチャを用いて，画像を分割しており，ピクセル単位で自動車から移る風景を「道」や「車」，「建物」，「人」などに分割している。このように，深層学習を用いることで画像を詳しく分類することが可能であり，この技術は現在開発が盛んな自動運転にも用いられている。この技術を地面性状の識別に転用することも可能であり，学習用データベースを用意すれば高い精度での識別が可能である。

深層学習を用いる場合，データセットの生成の手間と必要な処理能力を考慮する必要がある。データセットには基本的に数千から数万枚の画像が必要であり，アノテーションというそれらの画像にラベルを添付する作業も必要もある。また，画像処理用のプロセッサである Graphics Processing Unit(GPU)が必要となるため，高い性能の計算機が求められる。

以上のように，視覚障がい者の安全な移動を支援するためには，地面性状の識別が欠かせないが，画像を用いた場合，その特徴量の工夫が必要となる。

### 2.3.4 文字情報認識に関する従来研究

視覚障がい者の支援のためには，文字情報取得を支援することも重要である．本節では，文字情報を認識する研究として，点字の認識，2次元コードの認識，文字自体の認識手法について紹介する．

#### (1)点字認識

点字の認識手法では，点字を専用の機器でなぞり，その凹凸を感知して認識する手法が提案されている[高梨 2012]．読み取り機器には3本の爪が平行に並んでおり，それぞれの爪が点字をなぞることで，爪の根本のセンサに変化が伝わり，その信号を解析することで点字を認識している．また，爪のずれに対応するために，点字列と同じ幅の溝を掘った固定器具も開発している．

また，点字を撮影し光の反射の仕方の違いを見分けることで凹凸を感知して認識する手法が提案されている[Zhang 2010]．点字は凹凸があるため，光を当てたとき輝度が高くなる場所と，影となるために輝度が低くなる場所が発生する．この輝度の差を画像処理によって強調させることによって，点字を検出する．

このように，点字の様々な認識手法が提案されているが，点字は視覚障がい者の1割程度しか判読できない．また，点字で大量の文字情報を表現した場合，点字の打点には広い面積を要する．

#### (2)コード認識

一般的に我々がよく目にするコードとしては，バーコードとQRコードが挙げられる．バーコードのような水平方向の白黒の変化のみで情報をコード化したものを1次元コード，QRコードのような水平方向に加えて，垂直方向の白黒の変化で情報をコード化したものを2次元コードと呼ぶ．

土井らは，視覚障がい者の買い物を支援する機器として，商品の情報を提示し，その中から選択した商品へ誘導するシステムを開発した[土井 2011]．このシステムの利点は，撮影した商品棚ラベルを利用して商品の認識と，誘導を行うので，あらかじめ店舗や商品に誘導システムを導入する必要がないことと，介助者がいなくてもシステムを動作させることである．手法としては，商品ラベルの認識手法は，入力された商品棚ラベル画像に対してウェーブレット変換を行い，バーコードの位置と，「数字，円，\$」などの値札を表す文字を抽出する．バーコード情報は，あらかじめ店舗ごとにモデルを作成しておき，そのデータベースと照合することで商品名と値段を取得する．

Gallo らは，画像中からバーコードを抽出する手法を提案している[Gallo 2011]．画像中のバーコードは白黒の線が交互に配置されているので，エッジ(水平・垂直方向の

微分成分)が強くなる傾向にある。そこで、画素ごとにその周辺の水平・垂直方向の2つの微分成分を計算し、それぞれを統合した値を評価値とする。評価値に対して閾値を設け、バーコード領域を抽出する。

また、Guoらは、焦点を合わせる必要のないライトフィールドカメラを用いて画像から鮮明なバーコードを抽出するシステムを提案した[Guo 2014]。バーコードの認識はカメラやバーコードリーダ等を正確に対象のバーコードにかざす必要がある。しかし、視覚障がい者の方には難しい。そこで、Guoらの提案したシステムを用いると、バーコードにピントが合うように、再フォーカスが行われるため、バーコードを接写する必要がない。Guoらは、Galloらの手法を用いてバーコードを抽出している。

これらの手法により、商品名や値段などを即座に調べることが可能である。同様のスマートフォン用アプリケーションも多数開発、販売されておりバーコードからの商品情報の入手は身近になっている。

しかし、バーコード自体には限られた長さの英数字以上の情報を埋め込むことはできないため、商品ごとに対応するバーコードを作成する必要がある。そのため、賞味期限のような日々変化するような情報を入力することができない(Fig. 3)。



Fig. 3 バーコード

そこで、より豊かな情報を入力可能な2次元コードが開発された。その中のQRコードは1994年に当時の株式会社デンソーが開発した2次元コードであり、数千文字を英数字と漢字、ひらがな、カタカナなどで入力することが可能である。しかも、汚れや破損に強く、撮影の方向に影響されにくい、省スペース、自作可能など様々な利点がある(Fig. 4)。認識も容易なため、スマートフォンや携帯電話のアプリとして広く普及している。





Fig. 4 QR コード

視覚障がい者のための読み上げ専用の2次元コードとして、SPコードと呼ばれる2次元コードとそれを読み取る「スピーチオ」が廣濟堂スピーチオ販売株式会社から販売されている[深見 2004]. 予め、SPコードを印刷物の隅に印刷しておき、読み込むことで登録された文章を読み上げることができる。また、SPコードを作成するためのソフトウェアも販売されており、誰でも容易にSPコードを用意することができる。コードを正確にスピーチオにかざすことができれば、登録された文字を正確に読み上げることができる。また、印刷するだけなのでICカードに比べて低コストである。

このように、点字や2次元コードを認識する手法の利点としては、一般的な環境内においても高い精度で抽出、認識が可能な点である。しかし、そもそも点字や2次元コード自体が環境中にはあまり存在しないため、環境中のすべての文字情報を得ることは困難である。そこで、文字自体を認識する手法が提案されている。

### (3)文字認識

芦田らは、文字画像から提案した特徴量を抽出し、その特徴量をSVMで学習する文字認識手法を提案した[芦田 2005]. 文字認識は認識の前処理として、文字を画像内から抽出する必要がある。そこで、この手法では文字と背景のエッジが強くなる性質を利用して、背景から文字を抽出する。また、文字は単語単位で認識することが重要である。そこで、抽出した文字を文字列として学習するために、文字の配置や位置、色を確認し、近い文字同士を結合させて文字列とみなす。そして、結合した文字列を認識するために、認識に必要な文字列の特徴を抽出する。抽出する特徴量は、横方向の黒画素の長さ(以降、黒ラン長とする)の分散や、文字形状と背景パターンの均一性の違い、エッジの数、密度、縦横比である。学習手法としては、SVMが利用された。提案した特徴量は英語アルファベットに対応している。そのため、その他の言語に対応させる場合、新しく特徴量を提案する必要がある。

Shiらは、湾曲した文字自体を学習し、変形した文字列でも安定して認識可能な手法を提案した[Shi 2016]. Deep Neural Network(DNN)を利用し、大量の画像を学習用画像として利用する。学習する画像によって湾曲した文字でさえ認識することが可能である。また、特徴量を自身で設定する必要もない。しかし、性能を十分に発揮させる

ためには、大量の画像を用意する必要がある。

このように、文字を認識する様々な手法が提案され、効果を上げている。認識精度をさらに向上させるため、認識結果を前後の認識結果より補正する手法も提案されている。

Kogaらは、抽出した文字列を前後の単語との関係性を予め辞書に登録しておき、正しい文字列に修正する手法を提案した[Koga 2005]。例えば、「海外」という単語の認識に失敗した場合、「毎外」や「海タト」のように認識してしまうことがある。そのような失敗を防ぐために辞書を用いる。すなわち、予め登録した単語から文字間のつながりの強さを学習させておき、誤った認識を補正する。この機能は、様々な文章作成ソフトウェアに搭載されている予測変換などにも使用されている。登録データさえあれば非常に有効な仕組みであるが、データを用意する手間や未登録の単語には対応できない。

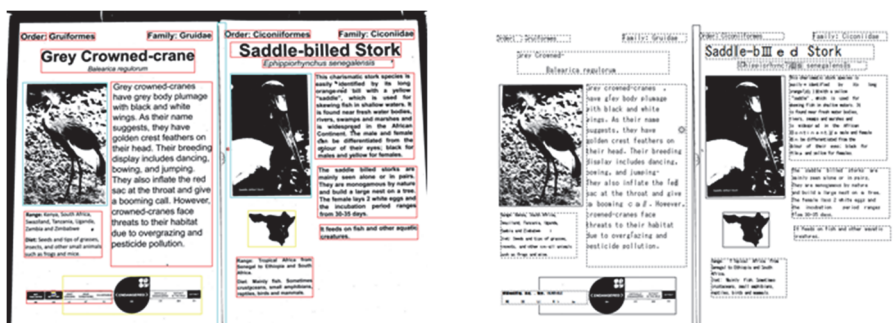
これまでに挙げた文字認識手法は、高精度に文字を認識できることが分かった。すでに、文字をテキストデータに変換するシステムとして、Optical Character Reader(OCR)が開発・販売されている。OCRの性能は年々向上しており、様々な文字を頑健に認識することができる。

しかし、OCRの問題点として、文字画像がスキャナによって取り込まれることを前提としているため、良好な照明条件での撮影が不可欠になっている。そのため、一般的な環境中の文字を撮影しOCRで認識しても、影や照り返し、ぼけなどの影響により文字の認識精度が低下してしまう。例として、Fig. 5にカメラで撮影した一般環境中の文字画像に対して、処理を加えずにOCRを用いた文字認識の結果を示す。一部の文字は認識できているが、文字のフォントや照明の影響などにより、認識の失敗やノイズを文字として誤認識してしまうことが分かる。



Fig. 5 OCRの結果

そのため、環境中の文字を OCR で認識しやすく変換する必要がある。例えば、単純に画像を二値化し、文字と背景の輝度差を大きくするだけでも、認識精度は多少改善される(Fig. 6)。このように、OCR を用いる場合、画像を文字が認識しやすいように加工することが必要である。すなわち、画像から文字を抽出することに着目する。



(a)入力画像

(b)翻訳画像

Fig. 6 画像の加工と認識結果の変化

### 2.3.5 文字抽出に関する従来研究

環境中から文字を抽出するための手法は、輝度値を用いたものやエッジを用いたもの、色を用いたもの、そしてそれらを総合的に使用したものなどが提案されている。

Ohya らは、小領域ごとに二値化処理を施す文字の抽出方法を提案した[Ohya 1994]。文字は、背景との輝度値の差が大きいため、輝度値に閾値を設けて二値化することで、文字と背景を分離することができる。しかし、この二値化が照明の影響により失敗してしまうことがある。つまり、同じ色に見えても照明の当たり方の違いによって異なる色と判断されることがある。

影は輝度の変化が滑らかな比較的大きな領域で強く影響するので、文字領域を複数のブロックに分割し、そのブロックごとに閾値を設けることで、影の影響を低減させることができる。また、画像中から文字を抽出する際、文字に欠けが生じることがある。そこで、同一の文章中の文字の大きさはほぼ一定であることを利用し、抽出した文字の大きさから似た大きさの領域を統合して抽出することの有効性を示している。

また、文字と背景の輝度差が大きいため、文字と背景の境界には明確なエッジが存在することが知られている。そこで、エッジ抽出に基づいた文字抽出手法も提案されている。

平岩らは、視覚障がい者支援システムとして、環境中の文字情報を利用した道案内システムを提案した[平岩 2002]。文字領域が背景との輝度値の差が大きくなり、エッジが取得しやすい点に着目し、Sobel フィルタを用いたエッジ検出によって文字候補

領域を検出する手法を提案した。しかし、それだけでは文字以外にも様々なエッジが検出されてしまうため、文字候補の密度やバランス、文字の大きさなどを調べ、それらのノイズを除去している。

Goto らは、環境中の文字を認識し、読み上げるウェアラブル型支援システムを提案した[Goto 2009]。前述の通り、文字領域は背景と文字領域の輝度値に差が現れるため、高周波成分が多い。そこで、周辺画像から Digital Cosign Transform(DCT)特徴を抽出する。その DCT 特徴が強く抽出された領域に文字が書かれているとして、その領域内から文字を抽出した。しかし、このシステムは、高周波数成分の高い領域を文字領域としているので、無地の背景中の文字のような、文字以外に高周波成分が少ない環境での利用が前提となる。

このように、輝度差やエッジによって文字を抽出する手法が提案されている。他にも、文字の特徴として、文字の色は単色であることが多い。そこで、色を基にした文字抽出手法も提案されている。しかし、色を特徴として文字を抽出する場合、「どこまでの色を同一の色と見なすのか」という問題がある。すなわち、晴眼者が「同じ色」と感じる領域も照明条件の変化や色むらによって異なる画素値を持っている。そのため、色分けに対する様々な手法が提案されている。

田畑らは、自己組織化マップ(Self-Organizing Map, 以降 SOM とする)を用いた文字抽出手法を提案した[田畑 2010]。SOM は教師なし学習を行うニューラルネットワークの一種である。SOM を画像に用いた場合、使われている色の頻度と位置によって、画像を色分けすることができる。しかし、そのまま SOM を適応すると、文字がつぶれてしまうことがある。そこで、SOM を用いて文字を色分けするため、画像を小領域に分割し、領域ごとに色分けする。つまり、エッジ情報を用いて、文字の内側と外側の色を推定し、内側の色のみで文字を色分けした。

さらに、藤崎らは上記の田畑らの手法を文字列に対応できるように改良した[藤崎 2011]。文字が水平に並んでいると仮定し、対象画像の文字から文字列を取得した。しかし、SOM の計算は繰り返し処理も多く、時間がかかる。

Kita らは、色に基づいた文字抽出手法として、k-means 法を用いた手法を提案した[Kita 2010]。k-means 法も教師なし学習手法の一種であり、クラスタリングによく用いられる手法である。あるデータ群を任意のクラス数に分割することができるので、画像の色分けにも応用できる。ただし、分割するクラス数は与えなくてはならない。そのため、事前にくつつの色で色分けされているか分からない対象に k-means 法を用いる場合、与えるクラス数を変え、複数の文字候補画像を生成することが必要となる。

生成した文字候補の中から正しい文字のみを抽出するために、生成した文字を二値化し、SVM を用いて黒画素の画像内の領域ごとの比率や文字の輪郭の比率が最も文字の特徴に

近い画像を文字画像としている。これにより、影や照り返しのある文字も抽出できるが、k-means 法自体に繰り返し処理が必要であり処理に時間がかかるうえ、複数のクラス数で計算しなければならず、さらに時間がかかる。

村山らは、屋外に設置された看板等の文字を取得する手法を提案した[村山 2001]。前述した通り、文字と背景は色差が大きい傾向にある。つまり、色差が大きくなる境界（エッジ）は文字と背景の境界である。そこで、境界の内側と外側はそれぞれ別の色として色を分割する。しかし、その場合、照明の影響などによって見た目には同じ色が別の色として分割されることもある。そこで、隣接するエッジ内の色差が小さい場合は統合する。この処理を繰り返すことで文字を抽出することが可能である。

Cho らは、同一の文字を構成する画素の輝度値が近い値になることに着目し、Maximally Stable Extremal Regions (MSER) 特徴を用いて、画像中の文字の可能性のある領域を文字領域として抽出した。さらに、抽出した文字候補の中から Canny 法のように 2 つの閾値を設定し、文字のみを抽出する手法を提案した[Cho 2016]。このとき、設定する 2 つの閾値は、Adaboost に基づいたカスケード型学習器を用いて学習する。

このように、輝度やエッジ、色情報によって文字を抽出する手法が提案されている。これらの手法は一長一短であるので、複数の手法を使用して最もうまく検出できた文字を選択することも考えられる。

松田らは、色や輝度、エッジの 3 種の閾値を用いる文字領域抽出手法を提案した[松田 2010]。文字抽出では、照明条件の変化やノイズによって、その精度が低下する問題がある。そこで、3 種類の条件に対して閾値を設けることで、その問題に対応している。色による文字抽出は、Mean Shift 法により似た色同士を統合して減色処理を行った後に、色の分布を確認してから、文字と背景、その他の色をそれぞれ抽出する。輝度による文字抽出は、2 画素の色差を輝度差に変換して評価する手法を提案している。二値化は小領域ごとに行い、文字を抽出する。エッジによる文字抽出は、文字は複雑な形状を持っているため、エッジを抽出すると背景よりもエッジ密度が高くなる傾向がある。そこで、小領域ごとにエッジ密度の高い領域を文字として抽出する。これらの 3 種の手法を統合的に用いることで複雑な環境中から文字の抽出に成功した。しかし、処理に時間がかかり、1 フレームの画像の処理に 1 分以上の時間がかかってしまう。

また、文字抽出の際には文字以外のノイズの抽出に気を付ける必要がある。従来研究として、テレビコマーシャルや地図のように文字以外にも映っている画像から文字だけを抽出する手法も提案されている。

Wernicke らは、動画からテロップを抽出する手法を提案している[Wernicke 2000]。

テレビコマーシャルなどの映像は背景の変化が激しく、文字のみを抽出することが難しい。そこで、テロップは数フレームの間固定されていることに注目し、抽出する手法を提案している。手法としては、文字の特徴をニューラルネットワークで学習し、数フレーム間変化の少なく、文字の特徴を持った領域を抽出する。この手法は抽出すべき文字以外は動いていることを前提としているため、テロップ以外への適応を考えると、カメラと対象物体はどちらも固定している必要がある。そのため、適応できる範囲は広くない。

Nakamura らは、地図上に書かれた文字を抽出する手法を提案している[Nakamura 1993]。複雑な背景上の文字抽出は、視覚障がい者が環境中から文字を抽出する際も重要である。地図上の文字は文字線が建物や等高線などと酷似していることがあるため、文字だけを抽出することが難しい。そこで、地図内に小領域を設定し、領域内の画素密度が設定した閾値の範囲であれば文字候補として抽出している。この研究では最終的に検出した文字を登録した文字の特徴と比較して選別している。

Google 社の Bissacco らは Android スマートフォンの機能の 1 つである文字認識システムのために、スマートフォンで撮影した画像から文字列を抽出し、認識する手法を提案した[Bissacco 2013]。文字の抽出には、Haar ウェーブレットとマルコフ確率場 (MRF)、異方性ガウシアンフィルタという 3 種の指標を利用している。この手法は、スマートフォンで撮影した低解像度でブレやボケも発生している画像から文字を頑健に高速に検出することに成功している。さらに、認識においては深層学習による文字識別器を作成して利用している。

しかし、正確に抽出し、認識できたとしても、読み上げた文字が視覚障がい者に正確に伝わらない場合もある。例えば、Fig. 7 に示すように、通常、環境内には様々な文字が混在しているが、このような乱雑に存在する文字に対して、上記の文字認識手法を用いた場合、画像中の文字をすべて読み上げる。すなわち、単なる連続読み上げになってしまう。連続読み上げでは、大量の単語を連続して読み上げることによって聞き取りやすさが低下してしまう。





Fig. 7 環境中に存在する複数の文字

視覚障がい者が大量の文字の読み上げによって混乱しないためには、読み上げる文字の位置や量を限定することが重要である。

### 2.3.6 領域の指示に関する従来研究

前項の通り、文字の連続読み上げによる混乱を避けるためには、多量の文字情報の中から特定の文字を視覚障がい者へ伝達することが必要である。そこで、領域を限定する手法について整理した。考えられる手法としては、目印を利用した手法や文字の位置を伝達する手法、指で指示する手法が考えられる。

#### (1) 目印を利用した手法

Morris らは、抽出する文字領域位置を特定するために、目印としてマーカを設置する手法を提案した[Morris 2006]。抽出したい文字を囲むように 4 点のマーカを設置することによって、読み上げる文字を特定するだけでなく、カメラが傾いている場合でも文字の歪みを補正することができる。しかし、普段使用する所有物にはマーカを設置することができるが、公共施設の文字や書籍、商品パッケージなどにこのようなマーカを設置することは難しい。

そこで、Clark らは事前に用意したマーカを設置することなしに、抽出する文字を特定する手法を提案した[Clark 2002]。環境中の文字は、矩形状の物体に記載されていることが多い。そこで、環境中から矩形物体を検出し、その物体上の文字を抽出する。しかし、商品パッケージの裏面のように、1 つの矩形物体の中にも多くの文字が記載されている場合がある。その場合、結果的に大量の文字の連続読み上げになってしまうので、視覚障害物へ情報が正しく伝達されにくい。

2.2.5 項でも挙げた Goto らは、一度抽出した文字領域を追跡することで不要な読み上げを減らす手法も提案した [Goto 2009]. 動画に対して DCT 特徴を用いて文字を抽出した場合、一度読み上げた文字列を繰り返し読み上げてしまう。そこで、利用者の利便性を考慮し、一度抽出した領域を追跡し、その領域中の文字を再び抽出しないことで繰り返し読み上げることを防いだ。追跡には時系列フィルタの一種であるパーティクルフィルタを用いる。しかし、繰り返しの読み上げが本当に不要かどうか議論の余地がある。つまり、一度目の抽出が失敗し、誤った文字を抽出した場合、むしろ何度か読み上げた方が正確に情報を伝達できる。

## (2) 文字の位置を伝達する手法

文字のある位置まで視覚障がい者を誘導することが提案されている。

久野らは、視覚障がい者にとって操作が困難であるタッチパネルに対して、ステレオカメラを用いた誘導システムを提案した [久野 2011]. タッチパネルは、ATM や券売機などで日常的に用いられている機器であるが、表示されているものを触覚で感じ取れないため、視覚障がい者にとっては使いづらい。そこで、ステレオカメラを用いて指先の高さを取得し、同時に小型カメラを用いてタッチパネルのボタンの位置と指先位置を取得する。そして、指先とタッチパネルの位置より、音声で目的のボタンまで指先を誘導する。その際、指先は肌色抽出と輪郭の鋭角部分を走査して検出する。タッチパネルの入力は、画面に触れていることを判断する必要があるため、指先の高さによりそれを判断する。このように、指先検出に関しては、指先の3次元情報を取得することでより高度な識別が可能である。

島田らは、触角によって、視覚障がい者へ文字情報を提示する機器について検討した [島田 2008]. かつて、視覚障がい者へ触覚情報を介して文字や図形などを伝える装置として Canon 株式会社より「オプタコン」が開発・販売されていた。オプタコンには、スキャナの先にカメラが設置されており、そのカメラで撮影した画像を濃淡情報として取得し、その輝度値によってオプタコンに設置している何十本ものピンが上下して視覚障がい者に文字や図形を伝達するものである。文字の形などが触覚に直接伝わるのでシステムの検出の失敗による誤作動はほとんど考える必要がない。しかし、問題点としては、触覚を介して十分に情報を取得するためには訓練が必要であることや、単純に濃淡を表しているだけなので、それが文字なのか模様なのか絵なのか判断する術がない。また、実際の使用を考えると、片手をピンに当て、もう片方の手でカメラを操作するので、持ち運びや他の動作中には操作が難しい。

このように、これらの手法は、視覚障がい者へ文字の位置を伝達することは非常に優れている。しかし、機器が大型化してしまい使い勝手が悪化してしまう。



### (3) 指で指示する手法

事前のマーカの設置や機器の大型化を避ける場合，撮影領域自体を限定してしまうことやいつでも用意できるマーカを利用することが考えられる．そのようなシステムの目印として，身体を用いる手法が提案されている．

Suzuki らは，指先に小型のカメラを取り付け，指先でなぞった文字を認識する「ネイルカメラシステム」を提案した [Suzuki 2011]．このシステムは，文字にカメラを接写することで極力ノイズを避けている．しかし，接写するということはカメラのずれが入力される文字のずれに大きく影響する．文字が傾くと，認識精度は低下してしまう．そこで，文字の傾きに影響を受けづらい四方面特徴を用いてロバストな文字認識性能を取得する．しかし，もともと晴眼者向けのシステムであるため，文字を正確になぞることが前提になっている．対象の表面を全体的になめることも考えられるが，その場合，文字列を正確に入力することが難しい．また，視覚障がい者の行動を阻害しないためには，取り付け位置は指先以外を選択するべきである．

西川らは，映像中の文字や図表を選択して抽出する手法を提案している [西川 1999]．選択の方法として，指先の軌跡を捉え，その軌跡で囲んだ文字や図表を選択する．この選択方法のためには，まず指先を検出することが必要である．指先検出は，手領域検出と指先領域検出の2段階の処理によって行われる．

手領域は，背景と入力画像との差分から動物体を抽出する背景差分法と，手の色の領域を抽出する色抽出を組み合わせて抽出する．指先領域は，手領域の輪郭の曲率を計算し，最も曲率の高い場所を大まかな指先として推定する．さらに，指先の詳細な位置は指先を模したテンプレート画像を用意し，テンプレートマッチングにより求めている．このような手法によって，指先位置を検出することが可能である．しかし，背景差分法と色抽出を組み合わせた手法は，カメラの移動や検出物体と背景の色が似ている場合に対応できない．

このように，手の特徴を用いた様々な指先検出手法が提案されている．しかし，抽出精度に問題がある．そこで，マーカを用いた手法も提案されている．

Fujishima らは，指先検出手法として，爪の色に基づいた手法を提案した [Fujishima 2013]．Fujishima らは色の変化の少ない領域を手や爪領域とみなした．手と爪領域を分離するために，それらの領域の面積を測定し，小さい領域を指先とした．この手法はカメラのブレにも強く，マーカ等を用意する手間も不要である．しかし，背景に指先に近い色がある場合には利用できない．一方，マーカを利用した指先検出手法について考える．肥川らはマーカとして手袋を利用した指先検出手法を提案した [肥川 2009]．手袋型マーカの色を背景から検出しやすい色とすることで検出精度を高める．

しかし，手袋をしたままでは，細かい作業の妨げになり，さらに点字が読めなくな

るという問題がある。

また、澤らはウッドベースの運指教示システムのために、指先にマーカを装着しマーカの動きにより指先位置を推定するシステムを提案した [澤 2010]。鮮やかな色のマーカ（蛍光テープ）を指先（爪）に装着し、その色を検出することで指先を検出する。そのうえ、ウッドベースの運指にはいくつかのルールがあり、指先の検出にもそのルールを適用することで指先検出の精度を向上させる。

指先を用いた領域の限定手法は直感的である。しかし、背景差分法や色抽出を用いた手法ではカメラのブレや背景・照明の影響により検出精度が低下する。

一方、マーカを使用した指先検出手法は精度こそ高いが、装着位置やマーカの形状を熟考しないと、利用者の行動を阻害してしまう。

Google, Inc.はスマートフォンで撮影した画像から文字を認識するサービス「Google 翻訳」を提供している。Google 翻訳は、実環境中の文字を優れた精度で高速に抽出・認識することが可能であり、スマートフォンの画面をなぞった軌跡によって認識する文字を指定することも可能である。しかし、このシステムの「指で指定する行為」は、スマートフォンの画面をなぞることであり、操作者に物体の手触り等は伝わらない。すなわち、触覚から情報を得ることができない。



Fig. 8 Google 翻訳(Google, Inc.)

## 2.4 従来研究のまとめ

### (1)地面性状識別

視覚障がい者の歩行の際の安全を確保するためには、周囲の障害物や地面性状を識別することが重要である。これらを把握するために、赤外線センサやカメラ、超音波センサなどが用いられた手法が提案されている。外出中の視覚障がい者は法律により白杖、もしくは盲導犬を携帯しているはずであり、少なくとも片手がふさがっている。したがって、ウェアラブル型のシステムが有効である。その観点から、カメラはその他のセンサよりも小型化しても得られる情報が多いという利点がある。一方、赤外線センサや超音波センサは障害物の検出には非常に有効であるが、赤外線センサや超音波センサはごく短い距離(一般的な測距センサでは 10mm 以下)は測定することが難しく、微細な凹凸を持つ地面性状を識別することができない。

微細な凹凸を持つ地面性状の識別には、その地面上を移動したときの振動や圧力で認識する手法も提案されているが、視覚障がい者に適用することは難しい。そのうえ、安全か判断する前にその地面に進入する必要がある。そこで、画像処理による地面性状の識別が視覚障がい者の支援を考えられる。また、画像処理による地面性状の識別は、その地面に進入する必要がないという利点がある。しかし、精度に問題を抱えていることが多い。

2018 年現在、深層学習が画像識別において目覚ましい結果を出している。これは地面性状識別についても例外ではない。深層学習はデータセットの準備の手間や高性能の計算機の準備といった問題を解決できれば、優れた精度を誇る。しかし、入力と出力結果から人間が成否を判定することは可能であっても、特徴量やパラメータを自動で学習する性質から途中の処理がブラックボックス化しており、結果の解釈が難しいという特徴がある。すなわち、なぜそのような結果が得られるのか理由が説明できない。それは、特に安全面に関わる問題を解決する際に、結果の解析ができないという問題を含んでいる。

そこで、極端に性能の高い計算機やカメラ等を使用せず、画像処理のみで信頼できる地面性状識別手法が必要となる。

そのうえ、画像で地面性状を識別する際、画像内には複数種類の地面性状が撮影されている可能性が高い。その場合、1 フレームごとに地面性状を識別すると、うまく識別できないことが考えられる(Fig. 9)。そこで、画像をいくらかの領域に分割し、領域ごとに分割する必要がある。



Fig. 9 複数の地面性状が撮影された画像の例

「導入の容易さ」を赤外線センサや振動センサなどの特別な装置の有無, 「識別範囲の限定」を複数の地面性状が含まれている場合に識別可能なように識別範囲を限定しているか, 「複雑背景」を細かい凹凸などを含んだ複雑な地面性状も識別できるかとして, これまで挙げた先行研究を分類すると, Table 3 のようにまとめることができる.

Table 3 地面性状識別に関する従来研究まとめ

	導入の容易さ	識別範囲の限定	複雑背景
今井 2017		○	
Leung 2014		○	
Weiss 2006			○
Best 2013			○
Ascari 2006			○
森 2004	○	○	
Wang 2015	○		
Lu 2009		○	○
Zenker 2009	○		△
Bajracharya 2008	○	○	△
Ros 2017		○	○
提案手法に求められる機能	○	○	○

## (2)文字抽出

文字の抽出・認識については, 文字の形状や色などの文字の特徴に基づいて文字を抽出してから, OCR を用いて認識する手法と, 文字自体を学習して認識する手法が提案されている. 文字認識には深層学習を用いた手法が大きな成果を挙げているが, 一

一般的に深層学習には大量の画像や高性能な計算機が必要である。また、特定の文字を環境中から探索したい場合は、文字（単語）自体を学習し、認識する手法が有効である。例えば、商品の「賞味期限」を知りたい場合は、その文字を環境中から探索する。

一方、単純に周囲の文字情報を得たい場合、特定の文字を対象とするのではなく、万遍なく文字を認識する必要がある。OCRを使用すれば、様々な文字を精度よく認識することが可能である。しかし、OCRは、その対象がスキャナで印刷された照明条件の整った文字であるため、環境中の文字をそのまま認識すると認識精度が低下してしまう。そこで、環境中から文字を抽出し認識しやすく加工することが重要である。

したがって、本研究では文字の抽出手法に着目する。ここでは、文字抽出手法について表にまとめた。視覚障がい者へシステムを導入することを考慮した場合、導入が簡易であることや複雑背景下からの文字抽出、抽出する文字の限定について考慮しなくてはならない。

調査した文字抽出手法に関する研究では、入力装置にカメラが多く使われ、他に特殊な機器を用いたものは少ないため、導入は容易なものが多い。単純な抽出手法を用いたものは単純な背景からの文字抽出を前提としている。しかし、実際の環境は複雑な背景が存在するため、その対策も必要である。そして、文字抽出に関する研究で、特定の文字を限定して抽出するものは少ない。視覚障がい者の利用を考えた場合、このような文字の限定は不可欠である。そこで、上記の3要素に着目した文字抽出手法を提案する(Table 4)。

Table 4 文字の抽出に関する従来研究まとめ

研究目的	導入の簡易さ	抽出範囲の限定	複雑背景
Ohya 1994	○		
平岩 2002	○		○
Goto 2009	○	△	
田畑 2010	○		
藤崎 2011	○		
Kita 2010	○		○
村山 2001	○		○
Cho 2016	○		○
松田 2010	○		○
Wernicke 2000		△	○
Nakamura 1993	○		△
Bissacco 2016	○		○
提案手法に求められる機能	○	○	○

文字の抽出領域を限定する場合、その限定はマーカを貼り付ける手法や、特定の形状の物品上の文字を取得する手法が提案されている。しかし、晴眼者でも暗闇では手探りで周りの様子を識別するように、指からは様々な情報が得られるため、文字の位置を感覚としてつかみやすいと考える。そこで、指先を用いた抽出文字の限定を行う。文字の抽出位置を指先に基づいて設定する場合、指先の検出手法も重要となる(Table 5)。

Table 5 指先検出に関する従来研究まとめ

研究	動作の阻害	触覚からの情報	照明変化への頑健さ
Suzuki 2011	無		○
西川 1999	無	○	
Fujishima 2013	無	○	
肥川 2009		△	○
澤 2010		○	○
Google 翻訳	無		○
提案手法に求められる機能	無	○	○

## 2.5 提案する視覚障がい者のための地面性状識別システムの概要

### (1) 想定する対象と問題設定

視覚障がい者の前方の地面を対象とする。

対象とする地面性状は以下と想定する。

- 地面はユーザに対して、平面とする。
- 外出時に遭遇することが想定される地面性状
- 視覚障がい者の足元を想定した範囲の地面性状

### (2) 想定する手法

地面性状を調べる手法としては、対象の地面内に進入して、その振動の違いから地面性状を識別する手法や前方を撮影し、その画像の特徴から地面性状を識別する手法が提案されている。

対象の地面内に進入する手法の場合、仮に危険な場所であれば、たとえ識別に成功しても視覚障がい者は危険にさらされてしまう。そこで、本研究では、画像処理を用いて、前方の地面性状を識別する。また、前方の地面性状は混在する場合は考えられるので、それに対応することも必要である。

### (3) 地形性状識別システムの概要

筆者らは、ロボットに搭載する画像を用いた地面性状識別システムを開発した。この際、地面性状識別システムに関する様々な知見を得ることができた。開発した地面性状識別システムはロボットの機能を用いているわけではなく、ロボットはただの設置場所として扱われている。そのため、設置場所をロボットから視覚障がい者に変更し、その視覚障がい者支援システムとしての利用価値を検証することは価値がある。

前方の地面に複数の性状が混在している場合を考え、あらかじめ撮影した画像からカメラに近い領域のみを分割して収集し、データベースを作成する。それらの画像は **Bag of Features** を用いてベクトルで表現した。ベクトルで表現する際の特徴量としては、形状を表す **Speeded Up Robust Features** と色を表す色ヒストグラムを統合して用いた。地面性状識別では、データベースの地面性状ごとにラベルを添付し、**SVM** を用いて識別器を作成する。撮影された画像からカメラに近い領域のみを分割し、識別器を用いて分割した領域ごとに地面性状を識別する。

#### (4) 要求仕様

地面性状識別は、視覚障がい者の方が歩行する際に必要となる。すなわち、歩行時間に合わせた処理が必要となる。視覚障がい者の歩行速度は、身長や白杖の長さ、歩幅によって左右されるが、約 1m/s であった[田邊 2013]。このことから 1 つの画面内の地面性状識別は 1s 以内に行われる必要がある。



## 2.6 提案する視覚障がい者のための文字取得支援システムの概要

### (1) 想定する対象

背景に複数の文字列が存在する環境内にある指定した文字列を対象とする。

対象とする文字の特徴は以下とする(Fig. 10).

- 大きさは、文字抽出の精度から解像度 640×480 ピクセルのカメラで撮影した際に、40[point]以上、199[point]以下の大きさの文字を対象とする。
- フォントに関しては、飾り文字のような判別の困難なものは対象外とする。また、同じ文字列内では大きさおよび太さはほぼ一様とする。
- 対象物品は、カメラに対して正立を基準とする。奥行き方向の変化に関しては、基礎実験により、パン方向に左右 20[deg]、チルト方向に上下 20[deg]までの傾きは許容する。
- 色は、同一の単語内で単色とする。
- 方向は、横書きとする。
- 長さは、2文字以上とする。

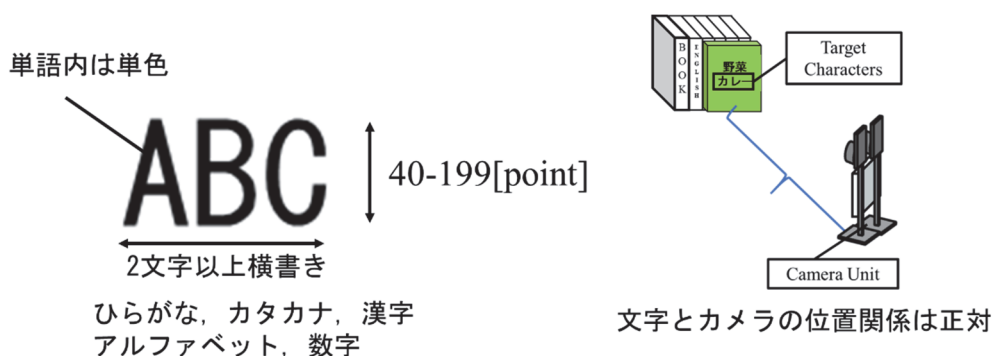


Fig. 10 対象とする文字

### (2) 想定する手法

文字の認識に関しては OCR を用い、OCR で認識が可能なように文字列を抽出する。指定した文字列の抽出には 2 つのステップが必要である。1 つ目は文字列を抽出する領域を限定すること、2 つ目は指定した領域内から文字列を抽出することである。

### (3) 文字列抽出領域の限定と領域内から文字列抽出システムの概要

文字列の限定手法としては、利用しやすさを考慮して指先を用いた限定手法を考える。しかし、従来手法の調査より指先のみを検出するとき、背景に指先に似た色を持つ物体が存在している場合やカメラが揺動する場合、検出精度の低下が問題となる。そのため、検出精度を向上させるために、画像処理で捉えやすいマ

一カを指先に設置することを考える。しかし、マーカを指先に設置すると、視覚障がい者の行動を阻害してしまう恐れもある。そのため、マーカの形状や設置位置を工夫する。つまり、行動を阻害しないマーカを用いる。例えば、我々の周りでは、多くの人々が指輪を装着しているが、それが行動を阻害することはほとんどない。そこで、指輪型マーカを指の根本に装着した指先検出手法を提案する。上記の手法で検出した指先位置に基づいて、文字の抽出領域を設定する。

抽出領域内から文字列の抽出を考える。従来手法による文字列の抽出手法や認識手法は、すべての文字列を抽出できるように設定されている。しかし、本研究の目的から、すべての文字列を抽出するような手法を考える必要は少ない。指先付近の文字列は、指先に近い文字と同じ色や大きさ、高さを持っていると考えられる。そこで、それらの指先に近い文字の情報を手掛かりに、文字列を抽出する手法を提案する。

## 2.7 おわりに

本章では、本研究で取り扱う課題に関連した従来研究についてまとめた。そして、地面性状識別と文字認識における要素ごとに、従来研究をまとめ、本研究の位置付けを述べた。

2.2 節では、視覚障がい者のための地形性状識別システムと文字情報取得支援システムについて下記の6種の従来研究に関して調査した。

### 歩行支援

- 誘導に関する従来研究
- 障害物検知に関する研究
- 地面性状の識別に関する研究

### 文字情報取得支援

- 文字情報認識に関する従来研究
- 文字抽出に関する従来研究
- 領域の指示に関する従来研究

2.3 節では、2.2 節でまとめた従来手法についてまとめ、地面性状識別および指先指定文字抽出が視覚障がい者の日常生活の支援に役立つことを明らかにした。

2.4 節では、2.3 節でまとめた地面性状識別に関して、問題設定とその解決のためのアプローチを提案した。前方の地面性状を高性能なカメラを用いることなく識別するため、特徴量や識別領域に関しての手法を提案した。

2.5 節では、2.3 節でまとめた指先指定文字抽出に関して、問題設定とその解決のためのアプローチを提案した。複数の文字列が存在する環境内から指先の文字のみを抽出するため、指先検出や、検出した指先周辺の画像情報に基づいた文字抽出に関しての手法を提案した。



## 3. 地面性状識別

---

3.1 はじめに.....	48
3.2 地面性状識別手法の流れ.....	49
3.3 データベースの構築.....	50
3.4 SVMによる地面性状の識別.....	65
3.5 おわりに.....	70

---

### 3.1 はじめに

第2章で述べたように，視覚障がい者の安全を確保するためには，前方の障害物だけでなく，地面性状を識別することが必要である．

我々は，ロボットの正面の地面性状を識別する手法の開発を通して，視覚障がい者の方の支援に対しても知見を得ることができた．開発したロボットのための地面性状識別システムも，システムとしてはロボットに搭載されたカメラで撮影した画像を識別しているというだけで，搭載場所を人体に移動したとしても得られる結果に大きな差異はないと考える．

本章では，その地面性状識別について詳述する．識別には，まずデータベースの作成が必要である．あらかじめ撮影した画像からカメラに近い領域のみを分割して収集し，データベースを作成する．ただし，1 フレームの撮影画像の中には複数の地面性状が含まれているので，画像をパッチ状に小領域に分割し，小領域ごとに地面性状を識別する．識別には SVM を用いるが，SVM で識別するためには小領域をベクトルで表現することが必要である．そこで，BoF を用いて画像をベクトル表現する．BoF でベクトル表現する際，例えば SIFT や SURF のような輝度勾配の特徴をヒストグラムで表現する特徴量を用いるが，地面性状は輝度勾配だけでは表現しきれないことも多い．そこで，小領域の色の分布を色ヒストグラムで表現し，それを色特徴とする．そして，輝度勾配と色の 2 つのベクトルを正規化して統合し，1 つのベクトルとして表現する．この際，2 つのベクトルのピン数が等しくなるように調整してから統合する．学習の際に，データベースの地面性状ごとにラベルを添付し，SVM を用いて識別器を作成する．

### 3.2 地面性状識別手法の流れ

まず識別したい地面性状の画像を複数枚用意しておく。この際、1枚の撮影画像の中には複数の地面性状が含まれているので、画像を小領域に分割し、地面性状ごとに分けておく。

判別には SVM を用いるが、そのためには小領域をベクトルで表現することが必要である。そこで、BoF を用いてベクトル表現する。BoF でベクトル表現する際、例えば SIFT や SURF のような輝度勾配の特徴をヒストグラムで表現する特徴量を用いるが、地面性状は輝度勾配だけでは表現しきれないことも多い。そこで、小領域の色の分布を色ヒストグラムで表現し、それを色のベクトルとする。そして、輝度勾配と色の2つのベクトルを正規化して統合し、1つのベクトルとして表現する。ただし、この際、2つのベクトルのピン数が等しくなるように調整してから統合する。

そして、統合したベクトルで表現した小領域を学習し、識別器を生成する。生成した識別器を用いて地面性状を識別する。

提案したシステムフローを Fig. 11 に図示する。

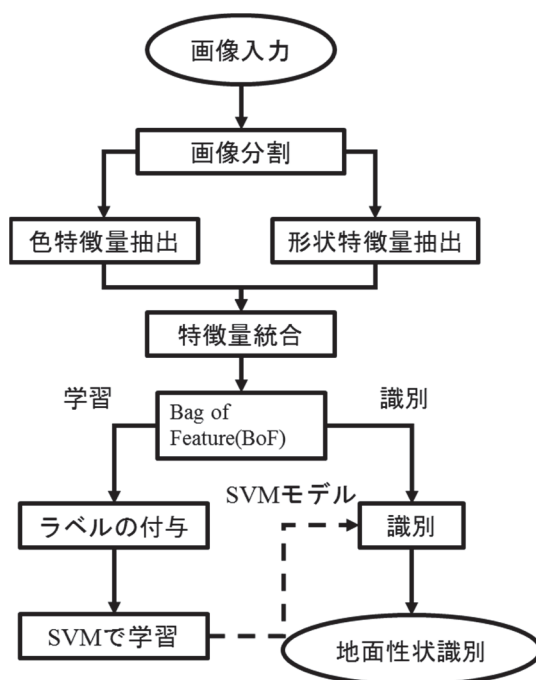


Fig. 11 地面性状識別手法のシステムフロー

### 3.3 データベースの構築

#### (1) 特徴量

画像から地面性状を識別するとき、質感（模様）や色を総合的に判断する必要がある。そこで、模様を表す特徴量と色を表す特徴量が必要となる。

模様を表す特徴量として、SIFT(Scale-Invariant Feature Transform)や SURF などが挙げられる。まず、これらの特徴量について簡単に説明する。

SIFT はある画素の周りの輝度勾配をヒストグラムで表現した特徴量である。この特徴量の抽出は特徴点（キーポイントとも呼ぶ）の検出と特徴量の記述で構成されている。

特徴点は対象とする画像に DoG (Difference-of-Gaussian)をかけて検出する。DoG とは対象とする画像に対して係数(Fig. 12 の $\sigma$ )の異なる 2つのガウシアンフィルタを用いて平滑化(Fig. 12)を行い、その結果生成される 2枚の平滑化画像の差分を計算することで、輝度の変化の激しい場所を検出する手法である(Fig. 13)。

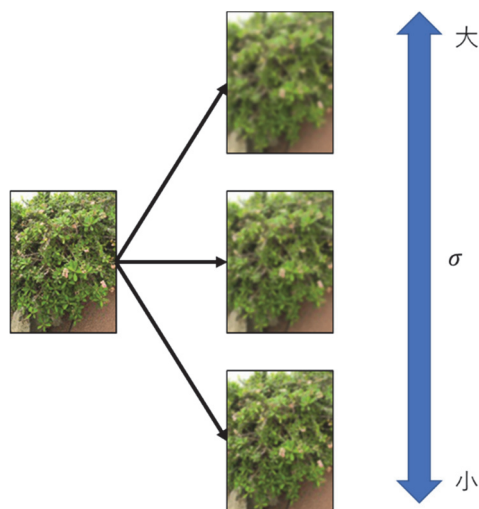


Fig. 12 複数枚の平滑化画像



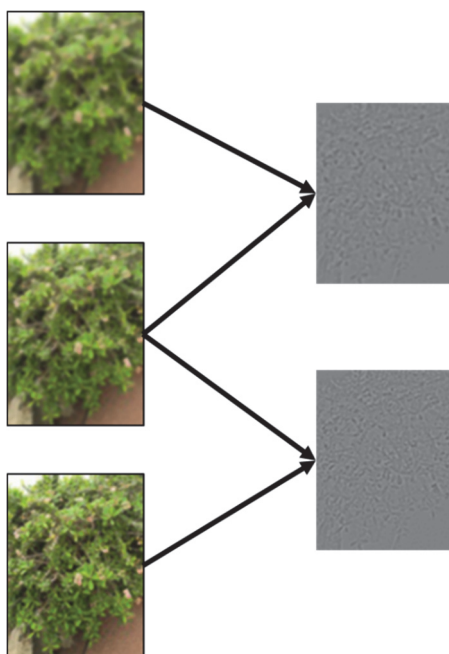


Fig. 13 DoG

複数の係数で、DoG を計算すると、ピラミッド状に DoG 画像を並べることができる。それらの DoG 画像の画素の中から周囲 (3×3×3) の画素(Fig. 14 の青玉)よりも輝度値が高い画素を検索し、これを特徴点とする。この特徴点の位置は、画像サイズが変化したときもほとんど変化しない。すなわち、スケールの変化に頑健である。また、特徴点周りの特徴量を調べることで、物体をよく表している部分の特徴のみを効率的に調べることができ、処理の高速化につながる (Fig. 14).

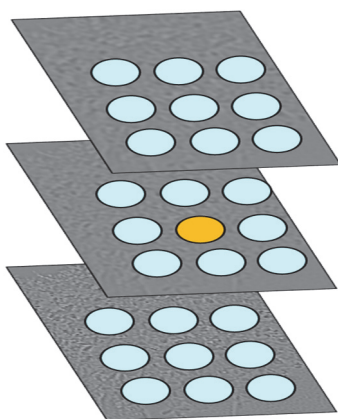


Fig. 14 特徴点の検出

特徴量の記述は、特徴量周りの輝度勾配をヒストグラムとして表現することである。このとき、単純に輝度勾配をヒストグラムで表現すると、画像中の物体の

姿勢が変化したとき、別の特徴量として判断してしまう。そこで、特徴量の向きをそろえる。向きをそろえるため、特徴点の周りのある範囲（設定した領域の大きさによる）の画素ごとの輝度勾配方向の頻度を 10 度刻みの 36 方向別に調べる (Fig. 15)。このとき、単純に計数するのではなく、勾配の大きさとガウス窓から重みを算出して加算する。ヒストグラムの最大値の 80%以上の方向をその特徴量の方向とする (Fig. 16)。この方向はたとえ物体が回転しても不変なので、この方向にそろえて特徴点周りの輝度勾配を調べることで回転の影響を受けにくくなる。この方向をオリエンテーションと呼ぶ (Fig. 17)。このオリエンテーションは 1 つの特徴点につき 2 つ以上割り当てられることもあり、その際は 2 つ以上のオリエンテーションを持つことになる。

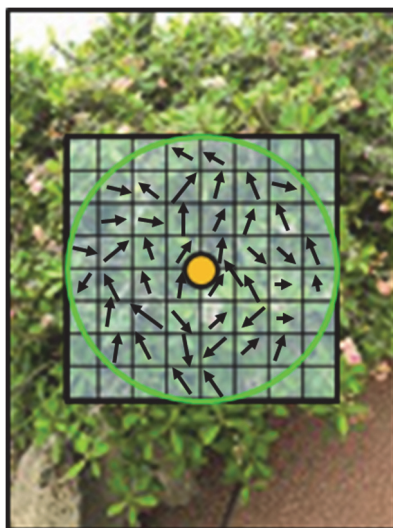


Fig. 15 特徴点の周りの輝度勾配

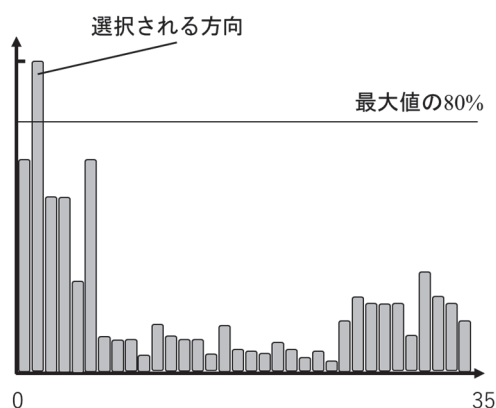


Fig. 16 輝度勾配のヒストグラム

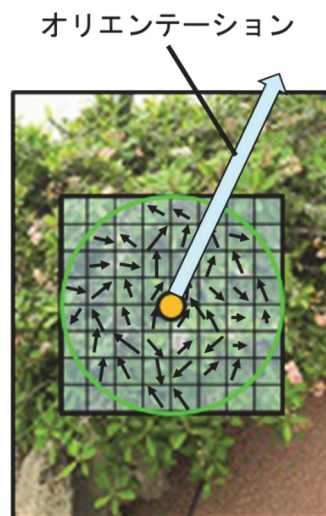


Fig. 17 オリエンテーション

特徴点周りの特徴量を計算する前に、その領域を基準の方向へ回転させる (Fig. 18). 特徴点周りの小領域 (SIFT の場合は  $4 \times 4$  領域) の輝度勾配の向きを 8 方向から選択する (Fig. 19). これを 1 領域につき 8 本のピンで表現すると、128 本のピンを持ったヒストグラムとして表現できる. こうして計算された特徴量が SIFT である.

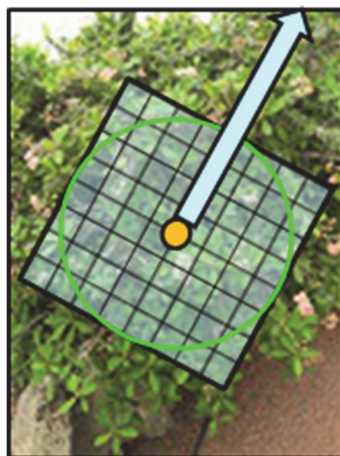


Fig. 18 基準方向への回転

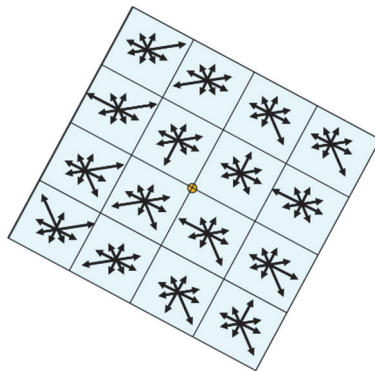


Fig. 19 領域ごとの輝度勾配

SIFT はスケールや回転の影響を受けにくく、精度高く特徴を表現できる。しかし、次元数が 128 と多く、計算に時間がかかる。そこで、高速な特徴量抽出手法として SURF が提案されている[Bay 2008]。SURF は、SIFT の際に時間のかかる処理を高速にするために改良した手法である。それは、特徴点の検出部分とオリエンテーションの計算部分、特徴量の記述部分である。

SIFT では特徴点を DoG を求めることで検出しているが、この手法は画像サイズが大きくなると処理速度も低下してしまう。SURF では、近似ヘッセ行列である Box filters で DoG のようにエッジを求める。その際、画素ごとの処理を行わずに領域ごとの処理を行う Integral Image を用いて、高速に演算を行う。Box filters は画像の大きさは変化しないので、画像の大きさが変化しても処理時間が変わらない。オリエンテーションの計算の際、SIFT は 10 度ずつ 36 方向を調べていたが、SURF は 60 度ずつ 6 方向を調べる。特徴量の記述の際も  $4 \times 4$  の小領域ごとに 4 方向ずつの輝度勾配を調べる。すなわち、64 次元の特徴量で特徴点を表現できる。

本研究では、最終的には視覚障がい者の方の歩行速度を基準とした時間内で地面性状を識別する必要があるため、輝度勾配を表現する特徴量として SIFT を高速化した SURF を選択する。

色の特徴としては、色ヒストグラムを用いる。色ヒストグラムはある領域の色の頻度を表すものであり、ベクトルとして扱うこともできる。色ヒストグラムの考え方は、例えば、4 色(赤・黄・緑・青)の色ヒストグラムであれば、赤は 0 番、黄は 1 番、緑は 2 番、青は 3 番が当てはまり、領域中の画素を色に合わせて、それぞれのピンに投票する(Fig. 20)。

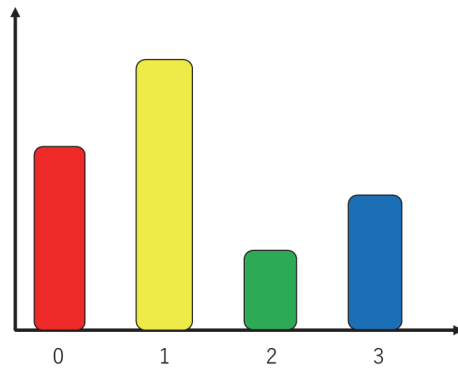


Fig. 20 色ヒストグラムの例

しかし、対象の画像の色数が多くなると、ピン数も増加してしまう。例えば、8bit の RGB カラー画像は、 $256^3$  の色数で表現される。この膨大な数の色を表現すると、似た色を別の色として認識してしまうことや、実装の上の問題として、このヒストグラムを表現するために大量のメモリを消費してしまう。そのため、色の減色処理が必要である。例えば、RGB 表色系で表現された画像を 64 色で表現することを考える。そのためには、RGB の 3 つの成分がそれぞれ 4 つの値を持つべき。もっとも単純な処理としては、それぞれの成分の値を 4 等分することである (Fig. 21)。

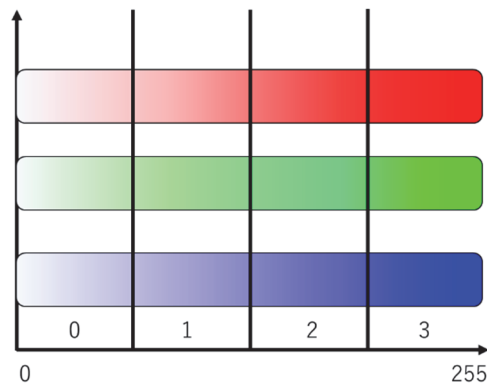


Fig. 21 減色処理の例 (RGB)

前述した SURF と色ヒストグラムを併用する場合、次元数を合わせる必要がある。SURF は 64 次元で表現されるため、色ヒストグラムは 64 色に減色させる。

RGB 表色系は赤・青・緑の 3 色の値を決めることですべての色を表現でき、ディスプレイ等で表示する上では非常に有効な表色系である。しかし、ある色の RGB

それぞれの値が人間の感覚ではイメージしづらい。一方，HSV 表色系は各画素を色相 (Hue) ， 彩度(Saturation)， 明度(Value)の 3 値で表現するため， 表現される色とそれぞれの値が人間の感覚に近い。また， 減色の際のパラメータを調整したいとき， RGB 表色系よりも調整が容易である。

HSV 表色系の減色処理は， RGB が赤， 緑， 青それぞれの画素値で区切ったのと同様に色相， 彩度， 明度を区切ることで行う。このとき， 色相の値は色を表現するためには細かく分割する必要がある。そのため， 色相は細かく分割する代わりに， 他の彩度と明度は粗く分割する。

彩度に関しては， 「淡い・鮮やか」の 2 値が表現できれば十分なので， 2 値に分割する。明度も高い分解能は不要であるが， 「暗い・明るい」の 2 値だけでは色を表現するためには不十分だったため 4 値とする。色相は， 細かく分割すべきである。全体の色数を SURF と合わせるために 64 色としたいので， 色相は等間隔で 8 値に分割する(Fig. 22)。彩度は 0-20 と 21-255 で分割した。明度は 0-20, 21-128, 128-200, 200-255 の 4 値に分割した。これらの値は基礎実験により求めた(Fig. 23)。



Fig. 22 色相と彩度の分割(減色処理)

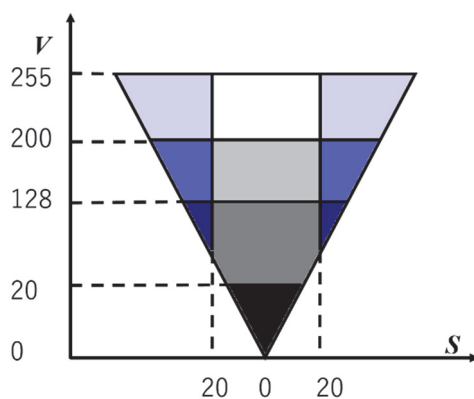


Fig. 23 彩度と明度の分割(減色処理)

輝度勾配を表現する特徴量と色を表現する特徴量が得られた。この 2 つの特徴量を同時に使用するため、1 つの特徴量として統合する。しかし、両者のピン数は等しいが、頻度のスケールが異なるため、そのままでは統合することができない。そこで、正規化を行い、両者のスケールを統一する。下記の式を用いて最大値 1、最小値 0 になるように正規化する。

$$Y = \frac{X - x_{min}}{x_{max} - x_{min}} \quad (3.3.1)$$

X:入力

Y:出力

$x_{max}$ :最大値

$x_{min}$ :最小値

正規化した 2 つの特徴量を連結すると、128 次元のベクトルとなる。

## (2) データベースの構築

地面性状の識別用データベースを構築する。識別の精度はこのデータベースの構築と学習手法によって上下する。データベースの構築は画像の取得、画像のベクトル表現の 2 つのプロセスからなる。

本研究では、4 種の地面性状に関してデータベースを構築する。この 4 種の地面性状とは、芝生、砂利、ウッドデッキ、コンクリートである。なぜこの 4 種の地面性状を選択したかと言え、日本の都市部においては、ほとんどの道は舗装されており、コンクリートやアスファルトで覆われている。そのため、これらの地面性状を見分けることが安全な地面性状と判断する上で重要である。本稿ではコンクリートとアスファルトを見分ける必要は少なく、また画像上の特徴もほとんど変わらないので、以後どちらもコンクリートと呼称する。また、外出時、芝生が生えている場所は基本的に立入が制限されている場所であり、進入するべきではない。また、砂利も歩きにくく、基本的には侵入が制限されている場所に用いられている。これらの理由から芝生と砂利は侵入すべきでない場所として、識別すべき場所とした。

これらの 3 種の地面性状のみを識別する場合、コンクリートは、形状の特徴の取得が難しく、色も芝生以外と見間違える可能性が高いという特徴を持っているので、安全な地面性状とは形状の特徴を持たないと誤解されることが懸念される。そこで、コンクリート以外の安全な地面性状として、ウッドデッキも識別対象と



した.

それぞれの地面性状は、色が他の地面性状と異なり、短い形状特徴を含む芝生、短い形状特徴を含む砂利、形状特徴も含まないコンクリート、長い形状特徴を含むウッドデッキという特徴がある(Fig. 24). 本研究ではこの4種の地面性状を識別しているが、使用目的に従って、識別する地面性状の変更や追加、削減することで、様々な場面に対応することができる.



(a)芝生



(b)砂利



(c)コンクリート



(d)ウッドデッキ

Fig. 24 4種の地面性状の例

用意したカメラで上記の4種の地面性状を撮影し、その画像を学習用画像とした。しかし、1フレームの画像中には複数の地面性状が含まれていることもある。そこで、1フレームの画像を小領域に分割し、それぞれの領域を1フレームの画像としてデータベースを構築する。撮影用のカメラは将来的にシステムをウェアラブル型にすることを考え、低解像度のカメラの使用を想定する。ここでは、 $640 \times 480$ の解像度のカメラを使用した。

小領域の大きさは処理速度を考慮して $80 \times 80$ ピクセルとしたが、より細かく分割すると1つの領域の中に別の地面性状が含まれにくくなるため、地面性状識別の精度は向上する。しかし、識別する領域の個数が増加するので、処理速度が低下する。すなわち、求める地面性状識別の精度や処理速度によって調整することが重要である。

地面性状を識別する際に、地面以外の物体がカメラ内に映り込むと、識別精度が悪化してしまう。例えば、壁や家や人などである。



そこで、そのような物体を取り除くために、地面以外の場所にある小領域は識別対象から外す。すなわち、カメラに近い範囲の画像だけを識別に用いる。本研究の対象外の障害物となるような物体は白杖や盲導犬、測域センサ等で検出が用意であり、将来的にそのようなシステムと組み合わせることも考えられる。将来的には地面の位置を測域センサを用いて検出し、その領域の地面性状を識別すべきであるが、本研究では単純に画像下部を地面と想定する。ここでは、画面の下端から縦に 160 ピクセルの範囲を画像下部とする(Fig. 25).

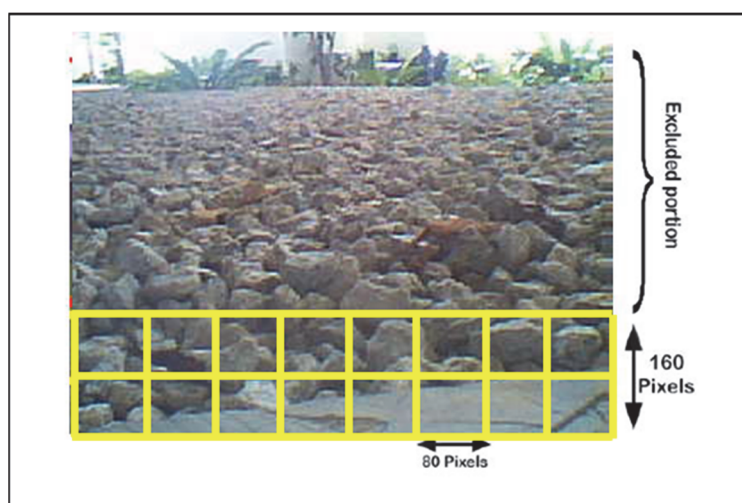


Fig. 25 入力画像の分割

学習のために、小領域を Bag of Features(BoF)を用いてベクトルとして表現する。BoF では、識別したい画像群に対して、似ている特徴(Visual Words)を探し、その Visual Words(VW)の組み合わせで画像をベクトルとして表現する。

まず、画像から特徴点ごとに特徴量を抽出する。通常、SIFT や SURF ではエッジのコーナーや交差点などの特徴が強く表れる位置 (Key point) を特徴点として、その周りの特徴量を抽出している。しかし、コンクリートのような地面性状にほとんど模様がないような場合、特徴がない場所こそがその画像をよく表していることもあり得る。そこで、グリッド状に特徴点を設置して、その周りの特徴量を抽出する(Fig. 26)。このとき、特徴量の記述には前述した SURF と HSV 表色系の色ヒストグラムを統合した特徴量を用いる。これを特徴ベクトルとする。

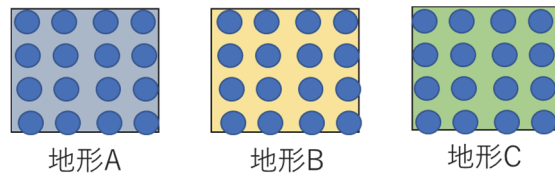


Fig. 26 グリッドサンプリング

抽出した特徴ベクトルをベクトル空間中に転写する．この処理により，似た特徴を持つ特徴ベクトルが近い場所に集まることになる(Fig. 27)．ここで，ベクトル空間中の特徴ベクトルをいくつかのグループ（クラス）に分割する．この処理はクラスタリングと呼ばれ，代表的な手法として **k-means** 法がある．**k-means** 法は特別なルール（教師）なしに与えられたデータの特徴を基に複数のグループに分ける手法である．**k-means** 法はそれぞれのクラスの個数が同程度であり，また超球状であることを暗黙的に条件としている．そのため，クラス内の特徴量（サンプル）の個数に偏りや特徴量の広がりを超球状ではないと **k-means** 法を用いたクラスタリングは難しい(Fig. 28)．その他のクラスタリング手法としては，階層的クラスタリング手法と呼ばれる手法が提案されている．この階層的クラスタリング手法は最も似ている（距離の近い）サンプル同士を順次集めてクラスとする手法であり，その距離の求め方により最短距離法やワード法などが提案されている[神嶌2003]．しかし，実はクラスタリング手法自体は BoF の性能にはあまり寄与しないことが報告されている[Nowak 2006]．また，**k-means** 法の計算量は  $O(kN)$  に対して，階層的クラスタリング手法は  $O(N^2)$  となるため，本研究では **k-means** 法を採用した．

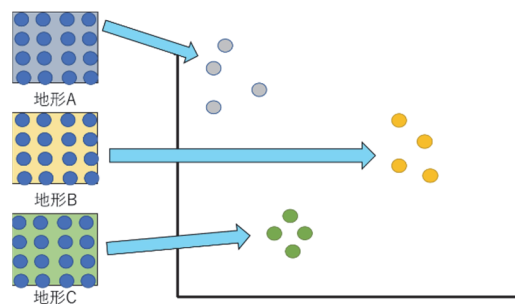


Fig. 27 特徴点の転写

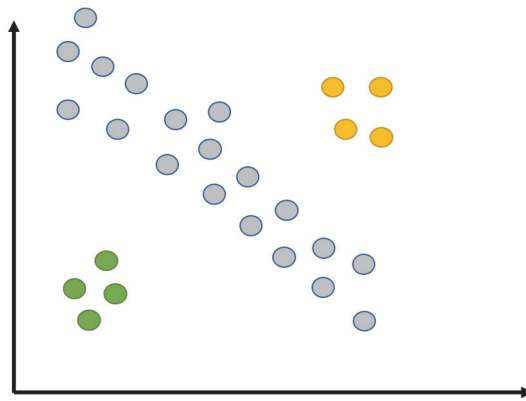


Fig. 28 k-means 法が適用できない例

転写した特徴ベクトル1つを1つの群(クラスター)とみなし、そのクラスター同士が統合を繰り返すことで最終的なクラスターを生成する。k-means 法では、最初に分割するクラス数を与える必要がある。例えば、3つのクラスに分割する場合、まず特徴ベクトルの中から3点(クラスター)を選択する(Fig. 29)。次に、その他の特徴ベクトルを、3点のうち最も近いクラスターに割り当てる(Fig. 30)。それぞれのクラスターの中心を求め、それを新しい3点とする(Fig. 31)。加えて、すべての特徴ベクトルを3点のうち最も近いクラスターに割り当てる。この処理を繰り返し、中心点が変わらない場合、それを真のクラスターの中心点とする。この中心点はクラスターごとの特徴を最もよく表している特徴ベクトルとなっているため、そのクラスターに属する特徴ベクトルはすべてこの中心点のベクトル(セントロイド)で代替することができる。このセントロイドを Visual Words(VW)と呼ぶ。VW の個数は対象によって、数百から数万くらいが選択される。本研究では、BoFによる地面性状の識別が可能であることを示すことが目的であるので、VW は仮に1000個とした。

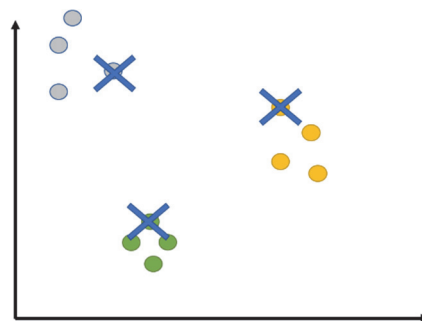


Fig. 29 k-means 法(1)

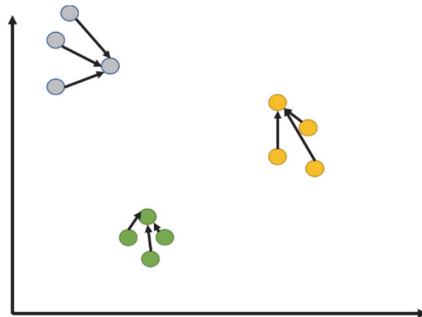


Fig. 30 k-means 法(2)

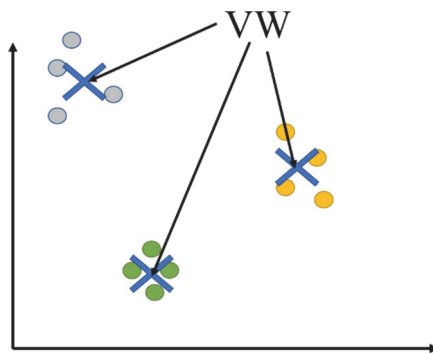


Fig. 31 k-means 法(3)

各地面性状の特徴ベクトルを VW で置き換える。このとき、単純に各特徴ベクトルと VW の最近傍探索を行うと、すべての VW との距離を計測する必要がある。特徴ベクトルは何十万個もあり、VW も数百から数千個に設定することが一般的であるので、計算量が爆発的に増加してしまう(Fig. 32)。そこで、計算量を低減させるために、最近傍探索手法の一種である KD-Tree 法を用いる。最近傍探索手法には他にも Locality Sensitive Hashing(LSH)などが提案されており、10 次元以上の高次元空間の最近傍探索を行う場合に性能が低下する KD-Tree 法よりもさらに高速に探索することが可能である。本研究では、KD-Tree 法を用いても要求仕様で言及した処理時間内に識別ができるため、KD-Tree 法を採用した。

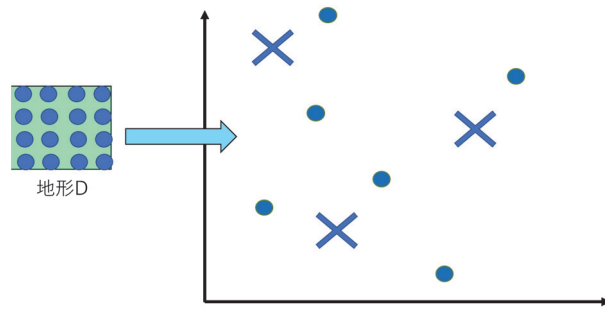


Fig. 32 VW による特徴ベクトルの置き換え

KD-tree 法は対象の空間を 2 分木になるように分割し，比較対象とする VW を減らすことで計算時間を低下させることができる．分割のやり方としては，最初に水平方向もしくは垂直方向に 2 分割する．次に，最初に水平方向に分けている場合は垂直方向，最初に垂直方向に分けている場合は水平方向に 2 分割する．この処理を繰り返す．この 2 分割するとき，その位置を VW の位置の中央値で分割すると平衡木となる(Fig. 33)．このとき，平衡木の計算量は  $O(\log n)$  となる．

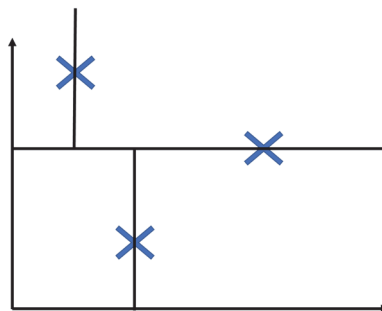


Fig. 33 空間の分割

分割した空間内に各地面性状の特徴ベクトルを転写し，それらの特徴ベクトルを最も近い VW で置き換える(Fig. 34)．そして，置き換えた VW の分布をヒストグラムとして表現する(Fig. 35)．このヒストグラムが対象の地面性状を表現するベクトルである．

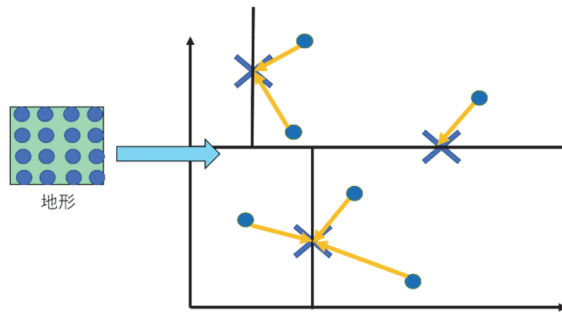


Fig. 34 VW による置き換え

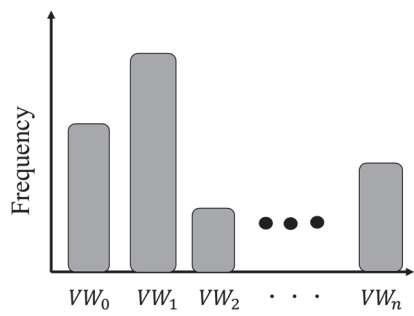


Fig. 35 画像を表すベクトル

### 3.4 SVMによる地面性状の識別

各ベクトルを地面性状としてSVMで学習する。SVM以外の学習手法としては、例えば、単純な線形識別関数を用いた教師あり学習アルゴリズムが挙げられる。単純な教師あり学習アルゴリズムでは下記の手順で学習される。

2つのクラスに分離することを考える。線形識別関数が次の式で表されるとする。 $w$ は重みベクトル、 $x$ は入力ベクトル、 $w_0$ はバイアス（任意の数値）である。 $g(a)$ は非線形活性化関数と言い、 $a$ が0以上の場合1を、0未満の場合-1を返す。この返した数値が分離するクラスを表している。

$$f(x) = g(w^T x + w_0) \quad \text{式(3.4.1)}$$

ここで、バイアス $w_0 = x_0$ とおくと、下の式のように表すことができる。 $d$ は次元数を表す。

$$f(x) = g\left(\sum_{i=0}^d w_i^T x_i\right) \quad \text{式(3.4.2)}$$

例：2クラス分離問題

1. 2つのクラス（例えば、犬と猫）の画像が複数枚与えられているとする。
2. 2クラスの各画像から特徴ベクトルを抽出し、教師ラベルを添付する。教師ラベルとは、ここでは犬や猫といった名前や答えを示す。
3. 特徴ベクトルと教師ラベルを学習器（線形識別関数）に入力して学習する。学習は、学習用の特徴ベクトルが入力されたときの結果と教師ラベルを比べ、正しくない場合は、重みを更新することで行われる。この学習によって、重み $w$ が決定される。
4. 学習済みの学習器（線形識別関数）を使用し、新しい入力画像を2つのクラスのどちらかに分離する。

この手法は、2つのクラスに線形に分割可能な場合は問題がない(Fig. 36)が、線形に分離が難しい場合は判別が不可能である(Fig. 37)。そのうえ、学習用特徴ベクトルと全く同じ特徴ベクトルが入力されたときの識別性能は高いが、未知の特徴ベクトルが入力されると、途端に識別性能が低下する。すなわち、汎化性が低いという問題がある。

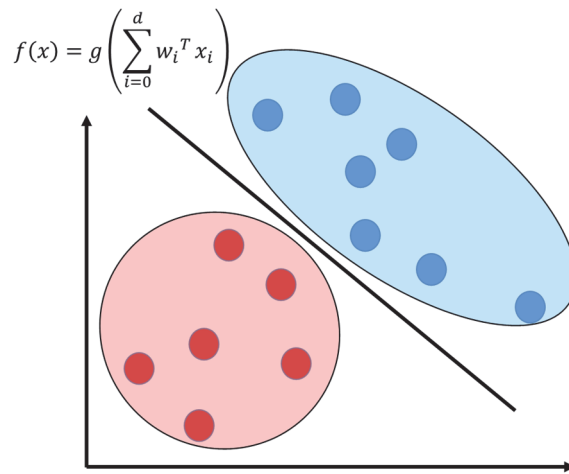


Fig. 36 線形識別関数を用いた2クラス分離問題の分離結果(線形)

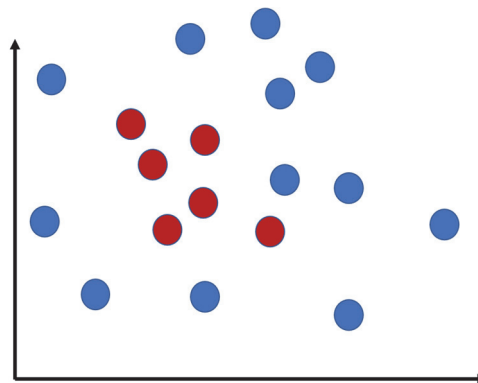


Fig. 37 線形識別関数を用いた2クラス分離問題の分離結果(線形以外)

そこで,このような非線形の問題の分離に対応でき,汎化性の高い識別手法として,**SVM**を採用する. **SVM**では上記の問題点を解決するために,マージンの最大化(**Fig. 38**)とカーネルトリック(**Fig. 39**)を採用している. マージンの最大化とはクラス間の境界線  $f(x)$ と各クラス間の距離が最大になるように調整することで,汎化性を向上させる効果がある. つまり,クラスがうまく分離できるように,境界線を調整することである. また,カーネルトリックとは,例えば2次元空間上に特徴ベクトルが非線形に分布しているとき,さらに高次元の空間上に投影することで,線形に分離可能な問題として解くことを指す.



カーネルトリックの際に用いられるカーネルにはいくつか種類がある。例えば、線形カーネルや多項式カーネル、Radial Basis Function(RBF)カーネルなどである。それぞれのカーネルには以下の特徴がある。

- ・線形カーネル：カーネルトリックの必要のない、訓練データが線形で分かれている場合に用いられるカーネルである。特徴ベクトルの数がデータ数よりも大きいとき、線形ベクトルが適している。

- ・多項式カーネル：最も単純な非線形カーネルである。後に述べる RBF カーネルが有効であることもあり、あまり用いられない。また、2つのパラメータを調整する必要がある。

- ・RBF カーネル：データに関する事前知識を必要とせず、最も広く用いられるカーネルである。ただし、1つのパラメータを調整する必要がある。

このようにカーネルにはいくつか種類があるが、その精度や使い勝手の良さから線形カーネルか RBF カーネルしかほとんど用いられない。そこで、本手法では非線形にクラスを分割するため、パラメータの調整が比較的少なく、効果的な RBF カーネルを用いた。

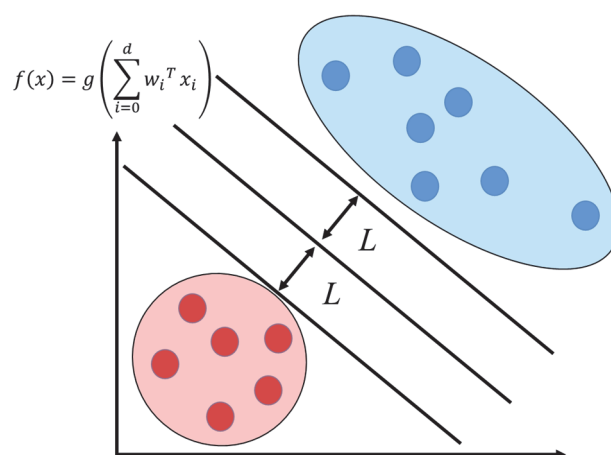


Fig. 38 マージンの最大化

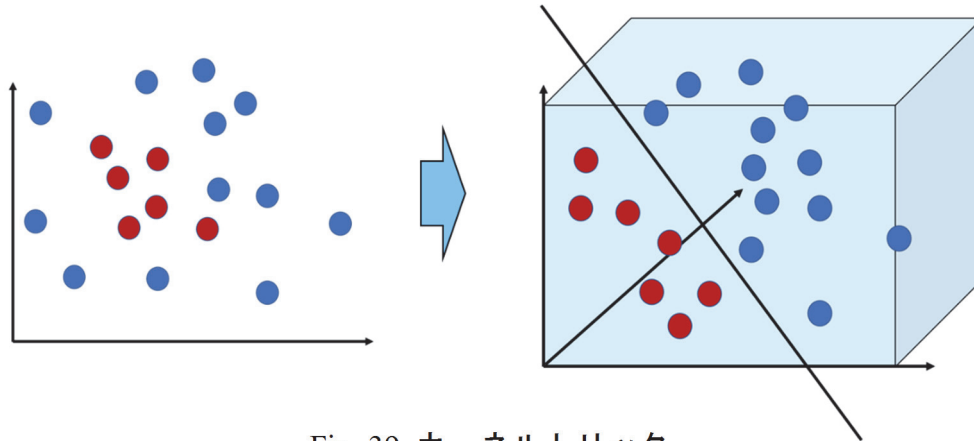


Fig. 39 カーネルトリック

マージンの最大化を考えると、ハードマージン SVM とソフトマージン SVM の 2 種類の分割方法がある。簡潔に説明すると、ハードマージン SVM は、分割する際に絶対に 2 つのクラスにお互いの特徴ベクトルが混じらないように分割する方法である。一方、ソフトマージン SVM は多少混じることを許容して分割する方法である。もともと 2 つのクラスが線形に分かれている場合、ハードマージン SVM の方が性能の高い識別器を作成することができるが、いかなる手段を用いても分割が困難な場合、境界線を設置することができない。そのため、実際の問題を扱う場合、ほとんどソフトマージン SVM が使用される(Fig. 40).

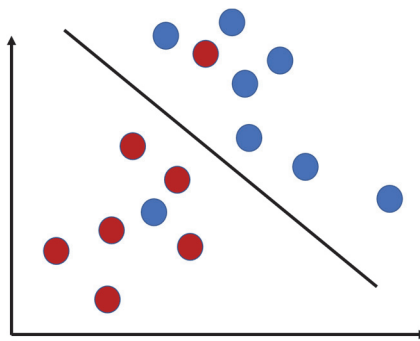


Fig. 40 クラスが入り混じる際の分割

ソフトマージン SVM の考え方は、ペナルティを与えることで誤分割を含めて評価することである。ソフトマージン SVM の一種である C-SVM の場合、この評価は下の式で表される。

$$\frac{1}{2}|w|^2 + C \sum_{i=1}^n (t_n - (w^T x_i + b)) \quad \text{式(3.4.3)}$$

$(w^T x_i + b)$ は SVM の識別関数であり、 $C$ は誤分割したときのコストの重みを示している。ハードマージン SVM の場合、 $C$ は無限になる。ソフトマージン SVM の場合、式(3.4.3)の  $w$  と  $b$  が最小になるように調整される。ここで、ペナルティとしてスラック変数を導入する。

$$\begin{aligned} \zeta_n &\geq |t_n - f(x_n)| = |t_n - (w^T x + b)| \\ \zeta_n &\geq 0 \end{aligned} \quad \text{式(3.4.4)}$$

$$\frac{1}{2}|w|^2 + C \sum_{i=1}^n \zeta_n \quad \text{式(3.4.5)}$$

誤分割が多くなると、式(3.4.5)の値が大きくなってしまうので、最小化するように学習すると、誤分割が減少させるように、重みのパラメータ  $w$  が調整される。

ここまで説明した SVM は 2 クラス分類に対応しているが、地面性状を識別する際には 2 つ以上のクラスを分類する必要がある。他クラス分類の手法としては、1 対他分類法や 1 対 1 分類法が提案されている。

1 対他分類法は任意のクラスとその他のクラスを分ける SVM を構築し、それをすべてのクラスに対して行い、それらの SVM を統合することで分類器を作成する。SVM はクラスの個数計算する必要がある。

1 対 1 分類法は、任意の 2 クラスを選択し、それらを 2 クラス分類の SVM を作成する。すべての組み合わせで SVM を作成し、統合することで他クラス分類の分類器を作成する。このとき、SVM は  $nC_2$  個計算する必要があるため、クラスの増加とともに 1 対他分類法に比べて大きく計算量が増加する。

本手法では、計算量の関係より、1 対他分類法を採用した。

### 3.5 おわりに

カメラ画像のみを用いた地面性状識別手法について詳述した。あらかじめ、収集した地面性状画像から地面性状ベータベースを作成しておき、そのデータベースと入力された画像を比較することで地面性状を識別した。

学習は、SVM を用いて行い、学習の際に1つの画像を1つの特徴ベクトルで表現するために、BoF を用いて表現した。

特徴量については、カメラの画質の影響を低減させるために、輝度勾配（形状や模様）をとらえる SURF と色を表した色ベクトルを統合し採用した。

その結果、外出時に識別が重要な地面性状として、芝生、砂利、コンクリート、ウッドデッキの4種の地面性状を識別する地面性状識別システムを作成することができた。この識別器は事前に収集する地面性状の画像を変更すれば、さらに多くの地面性状に対応することもできる。

## 4. 地面性状識別システムの評価

---

4.1 はじめに .....	72
4.2 実験条件 .....	73
4.3 実験結果 .....	75
4.4 おわりに .....	77

---

#### 4.1 はじめに

提案した地面性状識別システムの性能を評価する。本章では画像のみを使用して地面性状を識別できることを示す。実験に先立ち、4種の地面性状（芝生、砂利、コンクリート、ウッドデッキ）の画像を収集しておき、あらかじめBoFとSVMを用いてデータベースを作成しておく。実験では地面性状ごとに30フレームの画像から16マスずつ画像を切り出し、合計1,920枚の画像を学習画像とした。また実験用に地面性状ごとに100フレームの画像から16マスずつ画像を切り出し、合計4,800枚の地面性状を識別した。そして、提案した地面性状識別システムの性能を示した。

## 4.2 実験条件

収集した各地面性状の画像からそれぞれ 130 フレーム，合計 520 フレームの画像を選択し(Fig. 41)，学習用画像をそれぞれの地面性状ごとに 30 フレーム，テスト用の画像をそれぞれの地面性状ごとに 100 フレームに分割した．これらの画像は Singapore University of Technology and Design 内で撮影した．

対象とする地面性状は，芝生，砂利道，コンクリート，ウッドデッキの 4 種類とする．屋外の地面性状は多種多様であるが，本研究ではこれらの 4 種類の地面性状に以下の特徴が認められるため採用した．

芝生：色と形状が特徴的

砂利道：色はコンクリートやウッドデッキと似ているが，形状が特徴的

コンクリート：色は砂利道やウッドデッキと似ており，形状にも特徴がない

ウッドデッキ：色は砂利道やコンクリートと似ているが，形状が特徴的

提案システムは，それぞれの地面性状に特化した識別システムではないため，地形の追加も原理的には可能である．



Fig. 41 使用した地面性状の一部

撮影装置として，Trek の Ai-ball(Fig. 42)を使用した．このカメラは低解像度であるが，Wi-Fi を用いて画像を送信することが可能である．視覚障がい者が本システムを使用することを考慮すると，カメラは身体に装着する必要があるが，処理装置も同様に装着すると装置が大型化してしまう．そのため，このように画像を外部の処理装置に送信が可能なカメラはシステムの小型化に有効である．

このカメラの画質は 640×480, F 値は 2.8, 画角は 60° である. また, カメラの大きさは直径 30mm, 奥行 35 mmであり, 重量は 100g である.



Fig. 42 Ai-ball(Trek)



### 4.3 実験結果

各地面性状 100 フレームの画像を 1 フレーム当たり 16 マスに分割し, 分割した領域ごとに識別した. その結果, Table 6 に示すように, すべての地面性状で Recall, Precision とともに 90%以上の精度で識別に成功した(Fig. 43). Recall は再現率ともいわれ, 網羅性を表している. Precision は適合率ともいわれ, 正確性を表している. Recall と Precision はそれぞれ下記の式で表される.

$$Recall = \frac{tp}{tp + fn} \quad \text{式(4.3.1)}$$

$$Precision = \frac{tp}{tp + fp} \quad \text{式(4.3.2)}$$

*tp*(true positive) : 予測結果が対象の地面性状, 実際の地面性状も対象の地面性状

*fp*(false positive) : 予測結果が対象の地面性状, 実際の地面性状は他の地面性状

*tn*(true negative) : 予測結果が他の地面性状で, 実際の地面性状も他の地面性状

*fn*(false negative) : 予測結果が他の地面性状で, 実際の地面性状は対象の地面性状

Fig. 43 の右側の図は識別結果を示しており, 緑は芝生, 灰色は砂利, 黄色がウッドデッキ, 青色はコンクリートを示している.

4 種の地面性状の中で, コンクリートの Recall が低下している. コンクリートは表面の輝度値の変化が滑らかであり, 本手法では輝度勾配がないことを特徴として学習している. そのため, 他の地面性状に滑らかな表面が含まれている場合, 輝度勾配のみでは地面性状を識別することが困難である. 本システムは色情報も併用しているため識別は可能であるが, 砂利やウッドデッキは照明の当たり方によっては, 色がコンクリートに酷似してしまうため, 見間違えが比較的多く発生している.

また, 使用したカメラは無線通信を行うため, 通信遅れを低減させるために画質を低下させている. そのため, 偽色やエッジのぼけが発生しがちである.

しかし, このようなカメラを用いても 90%以上の精度で地面性状を識別することができる. そのことは, 特徴量を合成する手法が識別に有効な可能性を示唆している.

また, 本実験では, それぞれの地面性状の識別における計算時間を計測した. その結果, Intel Core i-5-4300U クロック数 1.9GHz, メモリ 8GB の計算機を用いて, 1 フレームの画像(16 マス)当たり芝生が平均 0.812 秒, 砂利道が平均 0.804 秒, ウッドデッキが平均 0.817 秒, コンクリートが平均 0.787 秒となった.

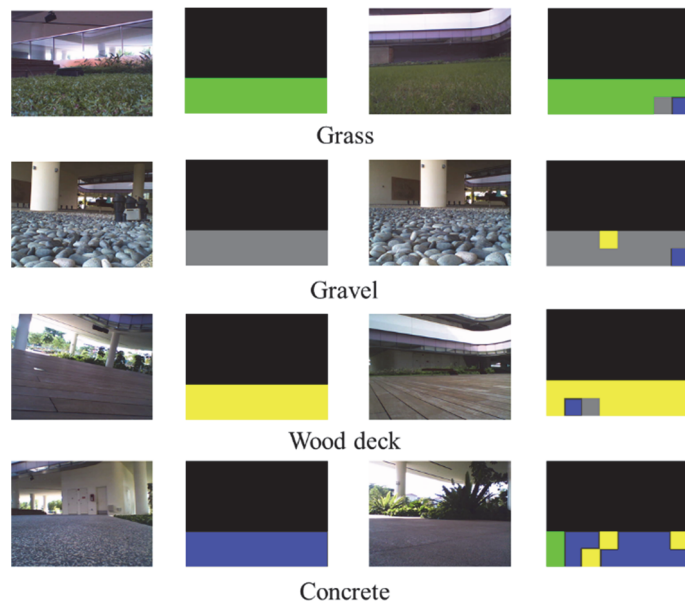


Fig. 43 識別結果

Table 6 地面性状識別システムの結果

	芝生	砂利	ウッド デッキ	コンク リート	再現率
芝生	1579	11	1	9	98.7
砂利	2	1544	30	24	96.5
ウッド デッキ	0	49	1530	21	95.6
コンクリート	18	67	74	1441	90.1
適合率	98.7	92.4	93.6	96.4	

#### 4.4 おわりに

提案した地面性状識別システムの性能を確認するために、4種の地面性状（芝生、砂利、コンクリート、ウッドデッキ）を対象として、識別する実験を行った。その結果、使用したカメラは高画質ではないにもかかわらず、すべての地面性状に関して、**Recall**、**Precision**ともに90%以上であることを確認した。

この結果より、提案手法である複数の特徴量を合成して用いることが地面性状の識別に有効であることが判明した。また、提案手法では地面性状の識別において輝度勾配と色に着目した特徴量を用いたが、これらの特徴量が地面性状識別において重要であることが分かった。

また、本実験におけるそれぞれの地面性状の処理時間は、1フレームの画像(16マス)当たり、芝生が平均0.812秒、砂利道が平均0.804秒、ウッドデッキが平均0.817秒、コンクリートが平均0.787秒となった。これは要求仕様で設定した1秒以下の処理となっており、要求仕様を満たしたと言える。



## 5. 墨字抽出

---

5.1 はじめに .....	80
5.2 指先検出 .....	81
5.2.1 素手のまま利用可能な指先検出 .....	82
5.2.2 マーカを用いた指先検出 .....	86
5.2.3 指先検出のまとめ .....	91
5.3 指先文字抽出 .....	93
5.3.1 指先文字抽出の概要 .....	93
5.3.2 指先文字抽出手法 .....	96
5.3.3 文字形状に基づく処理 .....	101
5.3.4 文字列傾き補正 .....	102
5.3.5 文字候補抽出およびノイズ除去 .....	104
5.3.6 文字の高さに基づく処理 .....	105
5.3.7 文字の隙間に基づく文字列合成処理 .....	106
5.4 おわりに .....	109

---

## 5.1 はじめに

本章では「指先指定文字抽出」に関して詳述する。指先指定文字抽出は、指先検出と文字抽出から成る。

5.2 節では、指先検出手法について詳述する。頑健な指先検出には、照明やカメラの揺動の影響を防ぐことが必要である。そのため、赤外線を利用した指先検出手法を提案した。指先は、赤外線を前方に照射し、指先検出用マーカ上の3個の赤外線反射材の位置関係から位置を算出した。

5.3 節では、文字抽出手法について詳述する。文字が混在する環境において指先付近の文字のみを抽出するために、指先付近の情報を利用した文字抽出手法を提案した。指先付近の情報として、検出した指先の周辺の色分布と文字候補の形状を利用した。この情報を用いて指差している文字列のみを抽出した。抽出した文字列には傾きやノイズが含まれるため、それらの補正および除去を行った。

5.4 節では、提案した指先指定文字抽出手法についてまとめる。

## 5.2 指先検出

第2章で述べたように、視覚障がい者へ正確に情報を伝達するためには、視覚障がい者自身が文字の位置を把握することが重要である。

位置を把握するためには、文字の位置をある程度絞ることが必要である。その限定手法の例としては、マーカを環境中に配置する手法が提案されているが、マーカを環境中に配置する手法は事前の準備が必要である。例えば、工場のラインなどでは有効な手法であるが、日常生活のように事前の準備が難しい場面での支援には向いていない。

そこで、本研究では、指や指棒などで読み上げる位置を限定する手法を用いる。レーザーポインタや指棒を使っての指示は画像中で極端に輝度が高い位置を発生させることや、指棒にマーカが取り付けられることができるため、安定した指示位置を検出することが可能である(Fig. 44)。しかし、指棒でビールの文字を指示し、システムが読み上げたとしてもそれがチラシに書かれている文字なのかビール缶に書かれた文字なのか判別することができない。

一方、指先による検出は、手ぶらで済むため他の動作を妨げない。また、文字が書かれている物体に直接接触することができるので、文字情報以外の情報（温度や感触、質感など）も得ることができる。そのうえ、指先で触れながら文字を指示すれば指先の感触によってより使いやすくなる可能性がある(Fig. 45)。

以上より、本研究では、文字は指先により指定することを想定する。視覚障がい者は文字の位置をそれとなく視覚から感知することが出来る弱視者と、全く感知することが出来ない完全失明者（全盲）に分けられる。弱視者は文字をなぞることができるが、全盲の方の場合はなぞることが難しい。そこで、全盲の方の利用も考慮し、指先に近い文字が構成する文字列を抽出することを考え、文字抽出時の指先の位置は固定とする。また、指先位置に文字が発見できない場合は、処理を行わない。



Fig. 44 指先以外の指示方法



Fig. 45 指先による指示方法

指先による文字抽出領域の指定は、素手のまま利用可能な方法とマーカを装着する方法が考えられる。

素手のまま利用可能な方法としては、指先の特徴や指先を動かした際の画像の変化を捉えて検出する手法が考えられる。

マーカを用いた手法は検出精度が高いが、マーカの装着が必要になるので、マーカ位置を工夫する必要がある。また、検出精度を安定させるためには、マーカ自体を工夫する必要もある。

以上の2つの手法を実装し比較する。

### 5.2.1 素手のまま利用可能な指先検出

素手のまま利用可能な指先検出手法について検討する。指先の検出手法は、2章で説明した指の色と背景差分法を利用した手法[西川 1999]や爪の色に基づいて指先(爪)位置を推定する手法[Fujishima 2013]などが提案されている。これらの手法は背景と前景(手)を色や輝度の差で判別し、検出した手領域中から面積の大小や相対的に尖っ



ている場所を指先として検出している。そのため、指先を正確に検出することが可能である。

この手法では、「手の色」の範囲を予め定めておき、その範囲内の色を「手の色」とみなしている。そのため、その範囲内の色が背景にも含まれた場合、指先と誤認識してしまう。さらに、本システムは将来的には視覚障がい者の身体に取り付けるか手で把持して使用することを想定している。そのため、カメラ自体が揺動することは十分に考えられる。カメラが揺動してしまうと、画面全体が揺れるので当然ながら背景差分法によってすべての領域が抽出されてしまい誤検出が多発する。

赤外線センサや超音波センサを用いて指先を検出する手法も提案されている[Pra 2014]。赤外線センサは、センサから手までの2次元の距離情報を取得し、手の形状から指先の位置を求めることができる。また、超音波センサは1次元の距離のみを取得するので、1つの指に1つのセンサを対応させた機器を用いることで指先を認識できる。これらの手法は背景に依存しないメリットがあるが、前者のシステムは誤検出がたびたび発生することが問題とされ、後者のシステムは装置が大型化してしまい、他の動作を阻害してしまう。

本研究では、カメラの揺動や機器の大型化、色の変化の影響を受けにくい指先検出手法を提案する。手でカメラを把持し、指先で物品をなぞったとき、カメラの揺動よりも指先の移動の方が速い傾向にある。すなわち、物体の移動速度に着目して指先を検出する。

#### (1) オプティカルフローを用いた移動体検出

物体の移動量を画像処理により計測する手法としては、オプティカルフローが広く使われている。オプティカルフローは時間的に連続する2つの画像を用いた動画画像解析手法の一種である。2つの画像間で共通する領域を探索し、それぞれの領域の移動量からその物体の速度や向きをベクトルとして計算する。

オプティカルフローの求め方には、いくつか種類があるが、代表的なものはLucas-Kanade法(LK法)とGunnar Farnelback法(GF法)である。前者は、画像ピラミッドを使用し、疎な特徴同士のオプティカルフローを計算する。後者は、密な特徴同士のオプティカルフローを計算する。指先位置を求める場合、画像内から手領域を密に抽出することが必要である。そのため、GF法を用いる[Farnelback 2003]。

GF法は、ある点の周辺の輝度値の変化はほとんどないと仮定し、2つの画像内の共通領域を調べる。LK法では、輝度値の変化の激しい点の周りの領域同士を比べてい

るので、処理は高速であるが、輝度値の変化がなめらかな領域ではオプティカルフローが計算できない。そこで、GF法をはじめとする密なオプティカルフローを算出するアルゴリズムでは、すべての画素についてオプティカルフローを求める。当然ながら処理量が爆発的に増加するため、処理を高速化する工夫も含まれている。

対象とする画像を縮小した画像を複数枚用意し、階層状に並べる。これは画像ピラミッドと呼ばれるもので、物体の探索、追跡などの際に、処理を高速化するためによく用いられる(Fig. 46)。縮小した画像の画像間の対応関係は原画像よりも高速に調べることができる。縮小した画像から順に対応関係を調べる時、1つ上の階層の結果を反映することで、計算時間を短縮し、精度も向上させることができる。

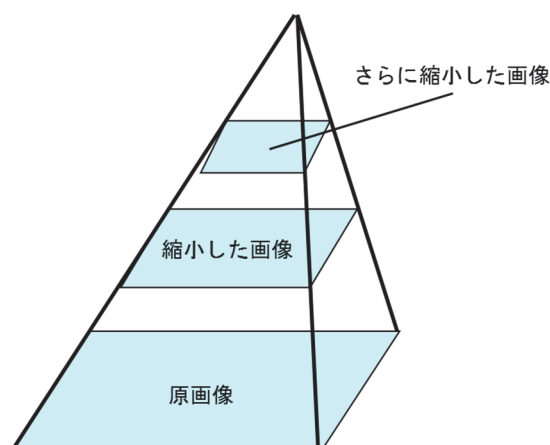


Fig. 46 画像ピラミッド

## (2) 指先の検出

上記のGF法を用いて、指先を検出する。時系列順に2枚の画像を準備する(Fig. 47)。この2枚の画像からオプティカルフローを算出した(Fig. 48)。輝度が高い画素は移動量が大きい画素を表している。全体の移動量の平均を算出し、画素値の大小を判別する(Fig. 49)。



Fig. 47 オプティカルフローを用いた指先検出



Fig. 48 オプティカルフロー(GF 法)



Fig. 49 高速で動く領域の抽出

指先で物を指す場合，手全体を見ると，指先方向に延びていると考えられる．そこで，抽出した手領域に対して，凸法を用いて，その領域を囲む多角形を計算する(Fig. 50).



Fig. 50 凸法

凸法で計算した多角形を囲む線分の上に指先位置はありと予想できる。指先位置は多角形の重心から最も離れた点(前提条件として重心よりも上の位置とする)である。重心と多角形を囲む線分の各画素の距離を計算し、最も離れた画素を指先位置とした (Fig. 51).



Fig. 51 指先の検出 (オプティカルフロー使った手法)

このように、オプティカルフローによって、指先位置を検出することができる。しかし、検出精度の他にも常に指先を動かしていなくてはならない。また、画像中に指先以上の速度で動く物体が進入した場合、検出精度が低下してしまう。指先を物品に当てながら動かす場合、ときには物品も指先と同じように動いてしまう。これらの問題点から指先のマーカを用いた指先検出手法を検討する。

## 5.2.2 マーカを用いた指先検出

### (1)指先検出手法の検討

マーカを用いた指先検出手法としては、第2章で述べた手袋をマーカとして装着するシステム[肥川 2009]や、指先の爪にマーカを接着し、マーカの動きにより指先を推

定するシステム[澤 2010]が提案されている。

手袋型のマーカは、入力画像に対する面積の割合が大きくなる傾向にあり画像処理により検出が容易である。しかし、手袋をしたまま何らかの動作をすることは難しい。また、手袋をしてしまうと手先の感覚が鈍るので、手探りで環境を認識する上では、素手の方が適している。一方、指先にマーカを装着する方法は手袋と比較すると、邪魔になりにくい。しかし、手袋型のマーカと比較すると面積が小さくなるので、検出精度も低下する傾向がある。また、これらの手法は、背景に使用されにくい原色のような鮮やかな色を抽出している。そのため、背景にマーカと似た色が存在する場合は精度が低下してしまう。

本研究では、背景や照明の影響を抑えるために、赤外線を照射し、その反射によって指先を検出することを考える。照明を照射する場合、その照明の種類を検討する必要がある。照明には可視光や赤外線光などが存在するが、赤外線透過フィルタを用いると、決まった波長の光以外を遮断することが可能である。この手法を実現するためには、赤外線の照射装置が必要になるが、赤外線照射 LED は十分に小さい(本研究で用いた LED は幅 5 mm, 高さ 8mm)ため、照射装置も小型化が可能である。

## (2)マーカの装着位置と形状の検討

マーカを用いた手法は、マーカの形状や位置によって視覚障がい者の動作を妨げないことが重要である。そこで、マーカの装着位置と形状を検討する(Fig. 52)。

我々が手探りで周囲を探るとき、重要なのは指先の感覚である。指先の感覚を損なわないために、指先にはマーカを取り付けるべきではない。そこで、指の根本にマーカを装着することを考える。

例として、指輪はアクセサリとして指に装着されるが、ほとんど動作の妨げにならない。そこで、本研究ではマーカを指先ではなく、指根元に装着することで指の動きの障害を最小限とし、指先を検出することを提案する。

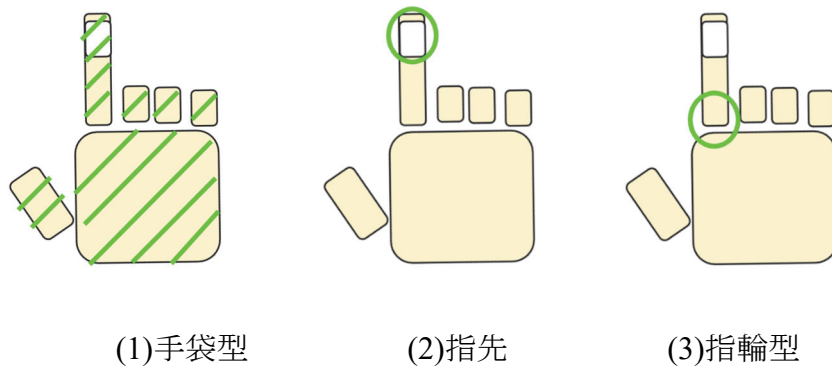


Fig. 52 マーカの装着位置

### (3)カメラユニット及びマーカ

カメラユニットは、カメラ、赤外線照射装置からなる(Fig. 53(a)). カラー画像と赤外線反射画像の撮影のため、2台のカメラを平行に設置する. 赤外線を撮影するため、片方のレンズ部に波長が840nm以上の光のみを透過する赤外線透過フィルタを装着する. 赤外線照射装置は63個の赤外線LED(波長850nm)を9×7に並べ、その前方に拡散板を設置した. 作成した照射装置の照射範囲は350mm先で180×300mmである.

赤外線は直進するため、その反射光を捉えるためには、マーカに再帰性反射を起こす素材を用いる必要がある. ガラスビーズやアクリルビーズに入射した光は、表面で屈折しビーズの底に反射して、再び表面で屈折して出射する. このときの入射角度と出射角度が等しくなるため、入射方向に光を反射することができる(Fig. 54).

本研究では、再帰性反射材としてアクリルビーズを用い、ビーズを埋め込む素材として光を反射する金属で軽く加工の容易なアルミニウムを用いた.

マーカの設置位置は、指先方向を明確にするために、二等辺三角形の形で3つ取り付ける(Fig. 53(b)). 二等辺三角形の底辺が指根元になるように指輪を装着する.

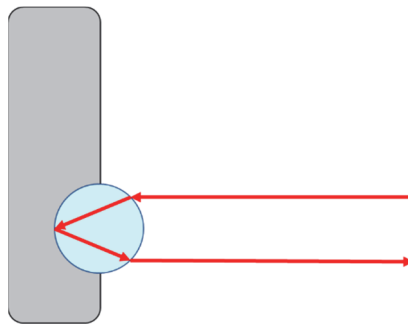
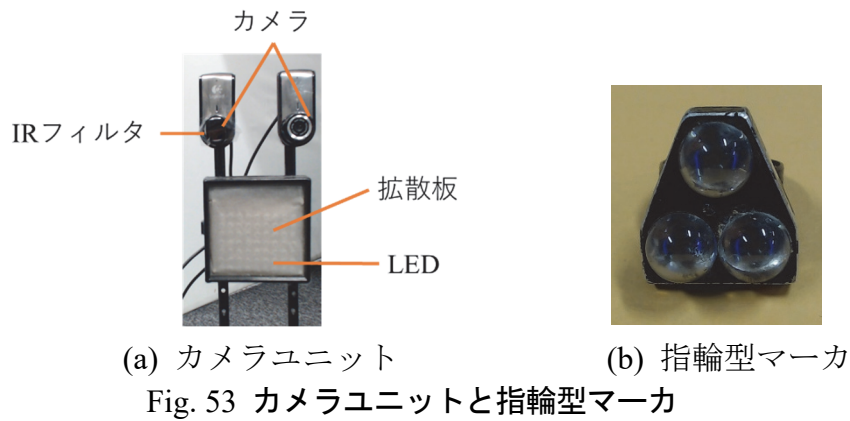


Fig. 54 再帰性反射

## (2)指先検出

前節で説明したカメラユニット及びマーカを用いて、マーカ上のビーズの位置関係から指先を検出する(Fig. 55(a)). マーカを指先に装着した状態で撮影すると、赤外線画像が得られる(Fig. 55 (b)).

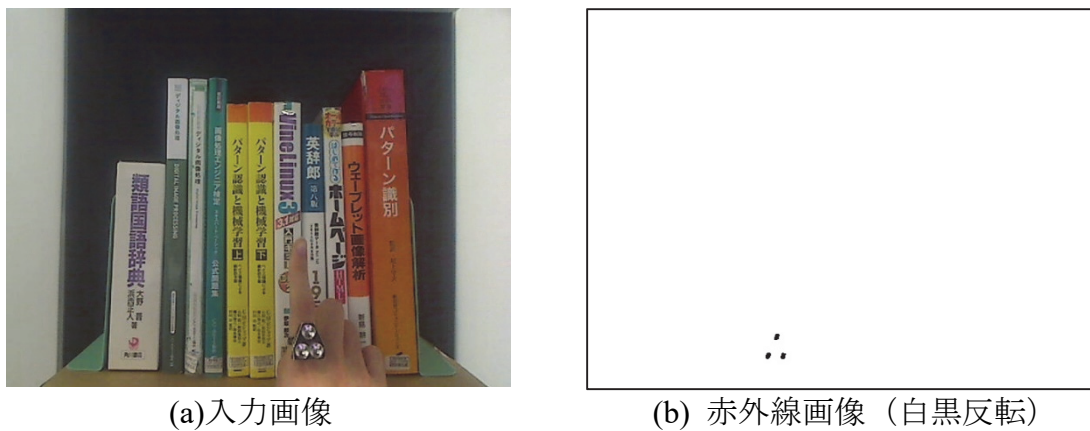


Fig. 55 赤外線反射光の検出

赤外線画像の画素を調べて、その画素値が  $V$  以上、縦横の幅が  $S$  ピクセル以内の領域を抽出する。この条件では、赤外線の反射の仕方によっては多数のノイズが含まれることがある。そこで、抽出するビーズはそのようなビーズよりも大きいと考えられるので、抽出した領域から面積が大きい順に 3 個の領域を選択する。また、3 個のビーズはお互いに近い位置に配置されている。そこで、選択した領域の重心間の距離が  $L$  ピクセルより大きい場合は領域を選び直す。

そして、3 個の領域のうち、他の 2 個から離れたものを指先側のビーズ  $A$ 、他の 2 個を指根元側のビーズ  $B$ 、 $C$  とし、 $B$  と  $C$  の中点を  $D$  とする。 $A$  と  $D$  から指の傾きと長さを求める(Fig. 56(a)).

$AD$  間の距離  $M_L$  は実測したところ約 15mm であり、人差し指の長さは日本人の手の寸法データ[AIST 2012]より、約 69.5mm なので、 $D$  から指先までの距離は、 $M_L \times n$  とした。すなわち、Fig. 56 では黄円が指先位置を表している。ここで、計算した指の長さが実際の指より短いと、指先が文字領域に入ってしまう、ノイズとして撮影される恐れがあるため、余裕を持たせて  $n=6$  とした。

赤外線画像とカラー画像はそれぞれの画像を撮影するために、別々のカメラを用意しているため、赤外線画像の領域から計算した指先位置は、実際の通常画像の指先位置とはズレてしまう。

2 つのカメラは水平に並べて設置してあるため、赤外線画像から検出した指先位置より右に  $N_m$  ピクセル移動させた位置を通常画像の指先位置とした。ここでは、基礎実験より  $N_m=5.8 \times M_L$  とした。

### (3)指先誤検出の低減

指先の誤検出を防ぐために、指先位置を複数回計算し、それらの結果から改めて信頼できる指先位置を計算する。

(2)の手法で指先位置を複数回検出し、それぞれの位置を記録する。指先位置の誤検出は、正しく検出することよりも発生頻度は低い。すなわち、それぞれの指先位置はほとんど同じ場所に現れるはずである。そこで、それぞれの指先位置を中心とした半径  $L_F$  ピクセルの円形の領域を設定する。この円形の領域内に他の指先位置が  $N_c$  個以上あるとき、その候補を正しい指先位置として記録した (Fig. 56(b)). すべての指先位置に関して、正しい指先位置か判別し、その個数が  $N_d$  個以上のとき、それらの指先位置の平均を最終的な指先位置とした (Fig. 57 の指先点)。ここでは、基礎実験より検出回数は 5 回、 $L_F=20$ 、 $N_c=3$ 、 $N_d=3$  とした。



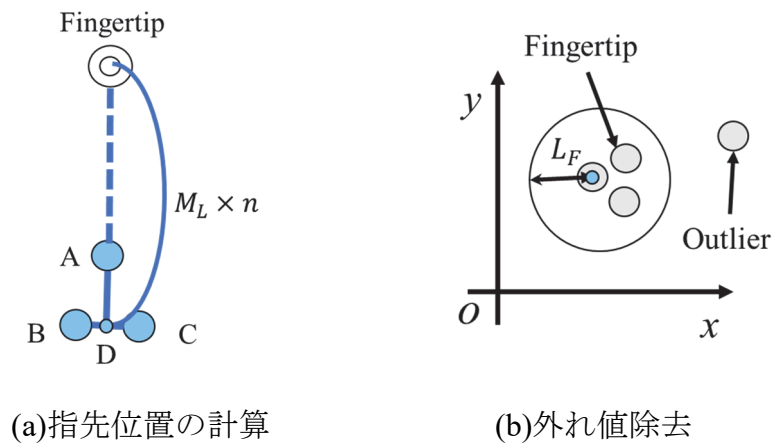


Fig. 56 指先位置の計算と外れ値除去



Fig. 57 指先検出

このとき、視覚障がい者は指先が検出できたことを把握できない。そこで、指先が検出できたことを伝えることで、システムの動作の様子を伝える。本システムでは指先位置指先が検出できた場合、ビープ音を鳴らし、ユーザに検出成功を伝える。

### 5.2.3 指先検出のまとめ

マーカを用いた手法とマーカを用いない手法の2つの手法を用いた指先検出手法を提案し比較した。

提案したマーカを用いた手法は、照射装置を用意する手間やマーカを装着する手間があるが、カメラが揺動していたとしても精度の高い指先検出が可能である。指定方法も、指先を置いた位置の文字を読むので、使い勝手が良い。一方、マーカを用いない手法は、装置の小型化やマーカを装着する必要がなくカメラが揺動していたとして

も指先の検出ができるが，精度はマーカを用いた手法と比べると低い．また，指定方法も指先を振り続ける必要があるので，前者のシステムよりも慣れが必要である．

本研究では，この2つの手法から，背景と照明の変化への頑健さ，指定の簡単さからマーカを用いた指先検出手法を採用した．

## 5.3 指先文字抽出

### 5.3.1 指先文字抽出の概要

#### (1)抽出文字の定義

指先で指定した文字を抽出する際、文字について前提条件を設定しないと抽出が非常に難しくなる。まず、文字の大きさを設定するため、カメラと文字間の距離を設定した。カメラは、将来的にウェアラブル型のシステムに改良することを考慮して、顔（肩）に設置してあると仮定した。そこで、物体を片手で把持し、もう一方の手の人差し指で物体を指差すように複数の晴眼者に指示したところ、顔から物体表面の距離は平均約 350mm であった（Fig. 58）。晴眼者と視覚障がい者の手の長さに違いはないため、自然な状態で把持した場合、上記の結果と差異はほとんどないと考えられる。すなわち、カメラから約 350mm 先にある標準的な商品パッケージ表面、書籍表紙の文字を抽出することを想定する。

また、その他の前提条件として、2.6 節で説明した抽出する文字の大きさとフォント、色、文字の方向、文字列の長さを再び列挙する。

- 大きさは、文字抽出の精度から解像度 640×480 ピクセルのカメラで撮影した際に、40[point]以上、199[point]以下の大きさの文字を対象とする。
- フォントに関しては、飾り文字のような判別の困難なものは対象外とする。また、同じ文字列内では大きさおよび太さはほぼ一様とする。
- 対象物品は、カメラに対して正立を基準とする。奥行き方向の変化に関しては、基礎実験により、パン方向に左右 20[deg]、チルト方向に上下 20[deg]までの傾きは許容する。
- 色は、同一の単語内で単色とする。
- 方向は、横書きとする。
- 長さは、2 文字以上とする。



Fig. 58 将来的なシステムの概要図

## (2)文字抽出手法の検討

本研究では、文字認識はOCRを用いて行うことを想定している。第2.3.4項で説明したように、通常OCRはスキャナ等で撮影されることを前提としているため、カメラ等で撮影した画像に対しては、うまく文字を抽出することが難しい。また、文字の抽出範囲が画像全体からとなり、視覚障がい者自身がどこに文字が書かれているか把握することが難しい。そこで、画像中から文字の位置を推定し、ノイズを除外しながら文字を抽出する。

従来の文字の抽出手法としては、文字の色を推定することで、文字の色だけを抽出する手法[芦田 2005][Kita 2010][松田 2010]や文字のエッジの特徴（複雑さやエッジ強度）を利用する手法[Suzuki 2011][平岩 2004][Goto 2009]が提案されている。

前者の色情報を利用した手法では、前提条件として同一単語中の文字は単色としている。文字の色の推定手法として、Fussy C-means Algorithm(FCM)[芦田 2005]やk-means法[Kita 2010]、色分布[松田 2010]より識別する手法が提案されている。FCMやk-means法は、任意のデータが与えられたとき、そのデータを似たクラスごとに分割する手法であり、任意のデータを画素の色とすると、画像中で使用されている色を推定することができる。しかし、これらの手法はあらかじめクラスの数を与える必要があり、それぞれの画像に適した分割数でないと、分割に失敗する。与えたクラス数が多すぎる場合、似た色を別々の色として識別してしまう過分割が起こり、与えたクラス数が少なすぎる場合、あまり似ていない色同士を同一の色と誤分割してしまう。

従って、画像中に使用されている色数が既知である必要があり、未知の画像が入力される実環境では実用性に欠ける。また、色分布を利用した手法は、単純に頻度の多い色を抽出するが、その中には多数のノイズも含まれる。

後者のエッジを利用した手法は、文字領域ではエッジ強度が高くなり、エッジの個数も多いことや高周波成分が多くなることを利用して文字の位置を推定する。エッジや高周波成分の評価にはエッジの強度[平岩 2004]やDigital Cosine Transform[Goto 2009]が用いられている。しかし、これらの手法はノイズにも過敏に反応する。

両者の手法は、どちらもノイズが発生し、その中から文字のみを抽出する必要がある。そこで、ノイズ除去手法としては、文字候補ごとに文字の特徴を確認し、文字とノイズを識別することが一般的である。用いられる文字の特徴としては、4方面特徴[Suzuki 2011]や、縦横比[芦田 2005]、画素ピッチ[芦田 2005]、文字間の重心位置のずれ[Kita 2010]、黒画素の比率[Kita 2010]、文字候補を囲む矩形内の画素の分布[松田 2010]などが提案されている。

以上を踏まえた上で、色情報を利用した指先の文字抽出・読み上げシステムを提案した[淵田 2012]. しかし、先行研究の手法は、本論文以上に文字の判別に閾値処理を多用しており、照明条件や物品の変化に脆弱であった。また、文字の傾き補正手法もノイズに脆弱であった。精度・汎用性において改善を加えて本論文の手法を新たに提案する。

### (3) 文字抽出手法の概要

文字の指差しによる指定により、指定された文字のみを読み上げるシステムを提案する。Fig. 58 に想定する文字抽出・読み上げシステムの全体像を示す。提案システムはウェアラブル型システムとして、視覚障がい者が身に着けて利用することを想定する。店舗や自宅で商品や書籍を手にとって表面を指差し、指先文字列を抽出して読み上げることが想定する。本論文においてはこのシステムの一部として、指先検出および文字抽出について取り扱う。

本章では、文字の入力装置、指先検出手法、抽出する文字の定義、文字抽出手法について考究を加える。その結果から提案したシステムの流れを示す。

### (4) システムの流れ

提案する指先検出・文字抽出システムの流れを以下に記す。

1. 赤外線を照射し、マーカの反射光を手がかりに指先位置を検出する。
2. 検出した指先位置に小領域を設定する。
3. 小領域内の文字の色を推定し、文字を検出する。
4. **Hough** 変換を用いて文字列の傾きを補正する。
5. 文字の太さを用いてノイズを除去する。
6. 文字の中から指先に近い文字のみを抽出する。

将来的なシステムは、視覚障がい者の身体に装着するウェアラブルなシステムを想定する。そのため、カメラのブレや、照明条件の変化が想定される。本章では、カメラのブレや照明条件の変化に対応した文字抽出システムについて検討する。また、文字の読み上げの際には、環境中に複数の文字情報が存在することがある。その場合、視覚障がい者は読み上げられる文字列を判別できず、混乱につながる。指先に近い文字のみを読み上げることが混乱を防ぐために有効であると考えられる。

晴眼者でも暗闇では、手探りで周りの様子を探る。つまり、指先の感覚によって周囲の環境や物体をある程度把握することが可能である。そのため、視覚障がい者も指先の物体の形状や重さなどはある程度認識できると考える。

しかし、同じような形状の物体があった場合、手の感覚だけでは区別することができない。そこで、物体表面上の文字を読み上げることができれば、判別することができる。また、物体の表面上には、商品名・賞味期限・成分表など重要な情報が含まれ

ているため、それを読み上げることによる支援が有効である。

それらの情報と文字情報を組み合わせることによって、使いやすいシステムになることが想定される。そこで、指先の文字を読み上げるシステムを提案する。

カメラ画像から、指先を取得し、指先の周辺にある文字を抽出する。抽出した文字を OCR によって、認識し読み上げることで、視覚障がい者の文字情報の取得を支援する。

### 5.3.2 指先文字抽出手法

本章では 5.2.1 項で述べた処理手順に従い、指先で指定した文字抽出手法を説明する。

#### (1)文字色推定

指先に近い文字の色を推定する。指先の小領域には文字が含まれていると考えられるため、その小領域の色を画面全体から抽出すると、画面から文字を抽出することができる。

小領域中の文字色を推定するため、小領域内の色の分布を調べる。色の分布を調査する際、画像を RGB 表色系(Fig. 59)で表現すると、照明条件の変化に影響を受けやすい。つまり、等色差性を考慮していないため、照明（明度）と色の値を分けて表現することが難しい。そこで、等色差性に優れた  $L^*a^*b^*$ 表色系(Fig. 60)に変換する。ここでは、 $L^*a^*b^*$ 表色系の  $L$  は輝度を表し、 $a^*$ と  $b^*$ は色相と彩度を表す。この等色差性とは人が色を区別する際の感覚的な度合いがそれぞれの画素が持つ値に倣う性質のことである。例えば、8bit の RGB 表色系で赤(255,0,0)と青(0,255,0)、緑(0,0,255)の色差は同じであるが、 $L^*a^*b^*$ 表色系では赤 (50,60,0) と青(50,0,-60)、緑(50,-60,0)となり、赤と青の色差より、赤と緑の色差の方が離れていることが分かる。赤の補色が緑であることから、 $L^*a^*b^*$ 表色系の色差の方がヒトの感覚に近いことが明らかである。

我々の周囲に存在する文字は、晴眼者が読むことを前提に、背景と文字を明確に見分けられるように書かれている。すなわち、背景と文字を見分ける際の感覚的な色の差はヒトの感覚に近くなる。そのため、本研究では、RGB 表色系から  $L^*a^*b^*$  表色系に変換して文字色を推定する。

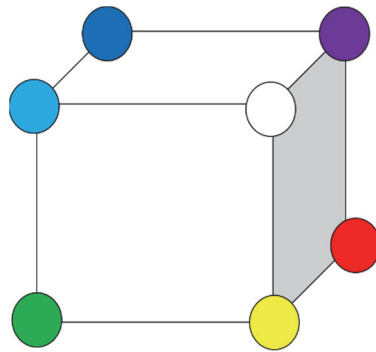


Fig. 59 RGB 表色系のモデル

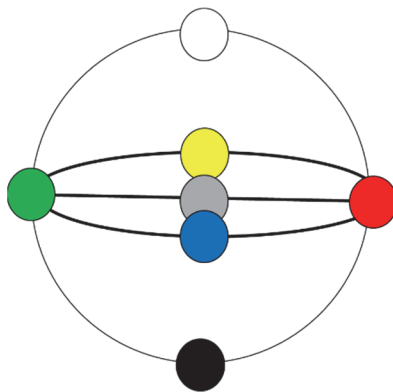
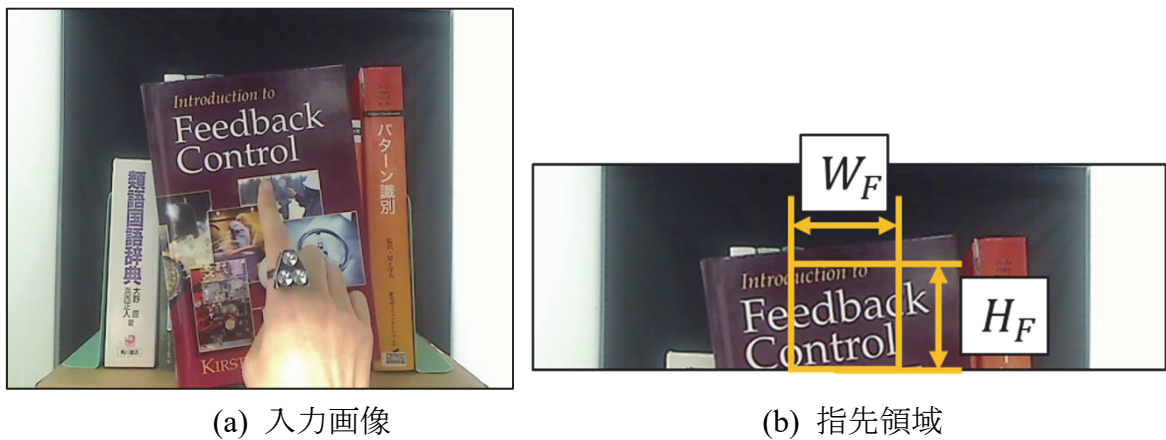


Fig. 60 L\*a\*b\*表色系のモデル

指先の小領域として、 $W_F \times H_F$ ピクセルの範囲 (Fig. 61)を、その領域の下辺の中点が指先位置と重なるように設定する. 色の分布を調査するために、この領域(以降、指先上領域) のヒストグラムを生成する(Fig. 62). ここでは、基礎実験より  $W_F=80$ ,  $H_F=80$  とした.



(a) 入力画像

(b) 指先領域

Fig. 61 指先周りの領域

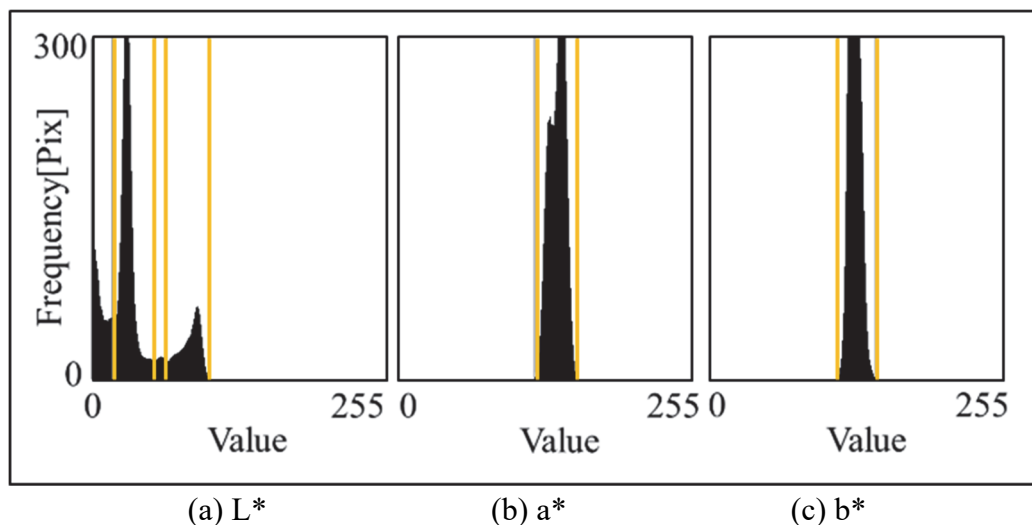


Fig. 62 指先の色分布 (L\*a\*b\*表色系)

指先上領域内で頻度の高い色は、ヒストグラム上では山として表現される。そこで、その山を極値として調べれば、領域内の代表的な色を取得できる。しかし、実際のヒストグラムには、ノイズや画素値のばらつきにより、微細な山が多数出現する。

そこで、式(5.3.1)に従い平滑化を2回行い、微細な山を除外する。 $H_i$ は注目している画素値の頻度を示し、 $num$ は要素の個数であり $num=5$ とする。ただし、 $i=0, 1, 254, 255$ のときは隣接する要素数が減るため、 $i=0, 255$ のとき $num=3$ 、 $i=1, 254$ のとき $num=4$ とした。

$$H_i = \frac{H_{i-2} + H_{i-1} + H_i + H_{i+1} + H_{i+2}}{num} \quad (i = 0-255) \quad \text{式(5.3.1)}$$

山の探索は、ヒストグラムの各画素値の頻度を調べ、注目する画素値の前後の画素値より頻度が多いものをヒストグラムの山とする。ヒストグラムには複数の山ができ、それが使用されている代表的な色を表している。山と山の間には、それぞれの色の境界があるはずである。そこで、色の境は、2つの山の間であり、頻度が1以上、左側の山の頻度の半分以下の、最も頻度が低い画素値とした(Fig. 62)。

色はL\*a\*b\*表色系で表現されているため、色空間を3成分で分割しなくてはならない。そこで、文字候補色を決定するため、色の境界を利用して色空間を分割する(Fig. 63)。この色空間に対して、指先の小領域の画素を投票する。色空間の各マスの投票数を調べ、全投票数の $T_v\%$ 以上を占めるマスに注目する。そのマスが小領域に使用されている代表的な色を表しており、そのマスの中心の座標を文字候補色とする。ここでは、基礎実験より $T_v=3$ とした。



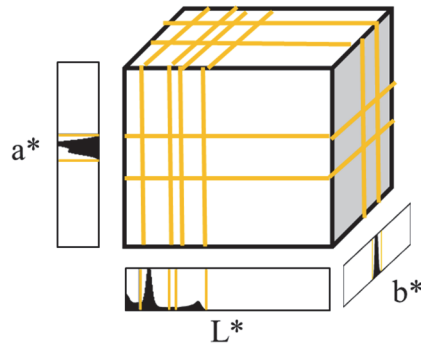
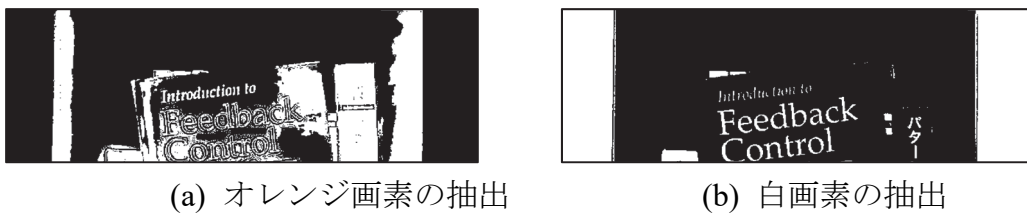


Fig. 63 ボクセル(L\*a\*b\*表色系)

文字候補色を元に、それに近い色を持つ画素を画面全体から抽出する．文字候補色と各画素との色差を計算する必要があるが、ヒトは彩度が高い色を見分けるとき、その色の違いを識別しにくくなっているという特性を持っているため、単純にユークリッド距離を計算してもヒトの目の特性に忠実な色差を表現することができない．そこで、コニカミノルタによって提案された CIE2000 色差式[Sharma 2005]を用いて、各画素と文字候補色との色差を計算し、色差が  $T_c$  以下の画素を抽出する(Fig. 64(a)).

しかし、CIE2000 色差式を用いて色差を計算してみると、白や黒など彩度が低い画素の画素値はばらつきが大きくなり、色差がうまく計算できないことがあった．そこで、文字候補色の彩度が  $T_{S1}$  未満の場合は各画素との明度差が  $T_B$  以内かつ各画素の彩度が  $T_{S2}$  以下の画素を抽出する(Fig. 64(b)). ここでは、基礎実験より  $T_{S1}=20$ 、 $T_c=20$ 、 $T_B=10$ 、 $T_{S2}=20$  とした．文字候補色を抽出した画像を文字候補画像とする．



(a) オレンジ画素の抽出

(b) 白画素の抽出

Fig. 64 文字列候補画像

指先上領域内の色分布によっては代表的な色が複数生まれるので、複数枚の文字候補画像が生成されることがある(Fig. 64). そこで、文字候補画像の中から、指先の近くに文字をより多く含んでいる画像を文字画像として選択する．

指先の文字の個数を計数する際、ノイズが文字として誤検出され、その結果、ノイズを多く含んだ画像が選択されることがある．そのため、評価用画像を生成する．評価用画像は、文字の個数さえ分かればよいので、文字の欠けを考慮せずにノイズを除

去する。この処理によって元の文字に欠けが生じることが多いため、評価用画像は画像の選択にのみ用い、それ以降の処理では使用しない。

ノイズの除去は、ノイズを大まかに除去した後に様々な指標で文字候補を評価し、条件を満たした候補のみを文字として計数する。

ノイズを大まかに除去する手法としては、メディアンフィルタやオープニング処理が挙げられる。メディアンフィルタは細かいノイズ（とくにスパイクノイズ）に有効なフィルタであるが、ノイズが塊になっているときは効果的でない。一方、オープニング処理は大きなノイズの除去に有効であるが、細かい文字も除去してしまう。

そこで、上記の2つのノイズ除去手法を用いて、評価用画像を2枚作成する。これらの画像に対して、大きさが  $N_w \times N_h$  ピクセル以下、密度が  $D_H$  以上、 $D_L$  以下のものをノイズとして除去する。また、指先座標から文字重心までの距離が  $L_c$  ピクセル以上のものを除去する。

そして、文字列内の文字の大きさには差が少ないことに着目して、文字の個数を計数する。大きさの近い文字同士をグルーピングするために、各文字の幅と高さを調べ、幅 - 高さ空間中にプロットする。プロットした点を中心に直径  $P_s$  の円を描き、重なった円の個数を数える。この個数の最大値を文字の個数  $W_N$  とする (Fig. 65)。ここでは、基礎実験より  $N_w=7$ ,  $N_h=5$ ,  $D_H=0.8$ ,  $D_L=0.2$ ,  $L_c=80$ ,  $P_s=10$  とした。

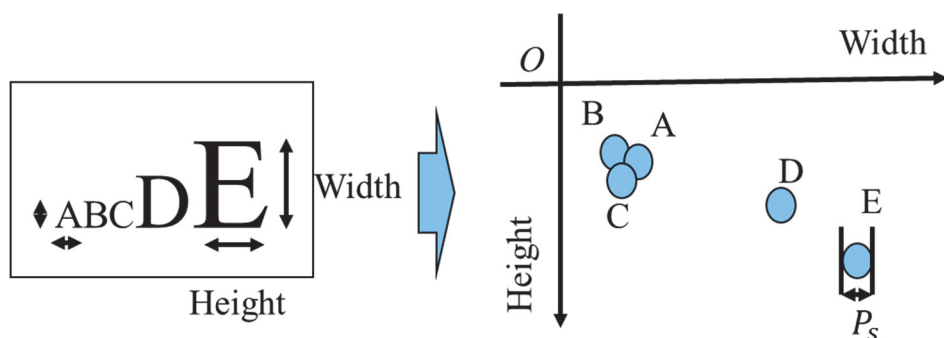


Fig. 65 大きさによる同一文字列の抽出

しかし、上記のノイズ除去手法を用いても、ノイズが残る場合があり、結局、大量のノイズを含んだ画像が選択されてしまう。

この問題を解決するために、ノイズが多少残留していても影響のない文字の特徴を導入する。文字を含んだ画像と、ノイズを含んだ画像、背景を二値化してしまった画像を細線化した後、残った画素は分岐点と端点、通過点の3種類に分類したところ、

文字を含んでいない画像に関して、分岐点が異常に増加することを確認した。そこで、分岐点数  $J_P$  の多寡によっても、評価用画像中の文字を評価する(Fig. 66)。すなわち、以下の値が最も高い画像を文字画像として選択する(Fig. 64 (b))。

$$\frac{W_N}{J_P} \tag{式(5.3.2)}$$

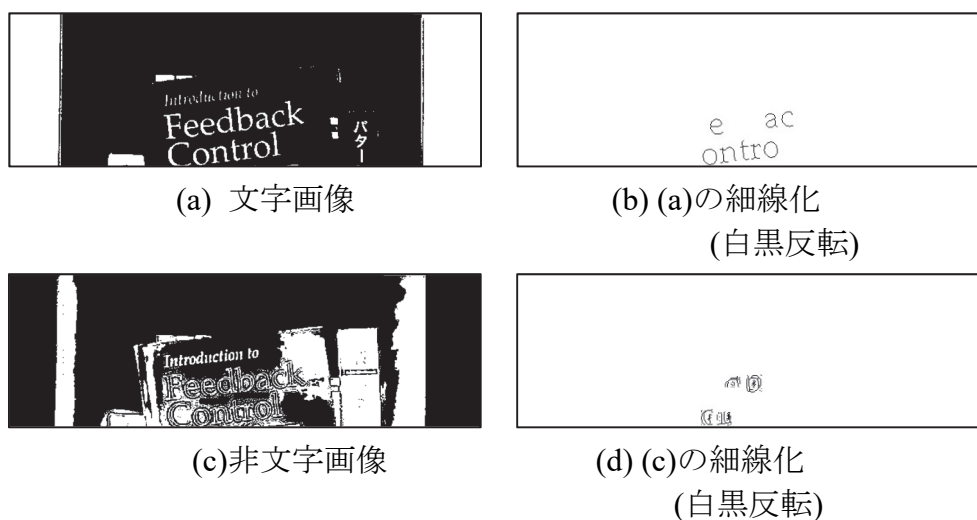


Fig. 66 細線化

### 5.3.3 文字形状に基づく処理

指先の文字色で二値化した画像には、明らかなノイズが含まれる。そこで、各文字候補を囲む矩形を設定し、下記の条件にあてはまる文字候補のみを抽出する。矩形の幅を  $R_W$ 、矩形の高さを  $R_H$ 、文字候補の画素数を  $C_P$ 、文字輪郭の長さを  $C_L$  とする。この条件は明らかなノイズのみを除去するために、非常に緩く設定している。

$$\frac{C_P}{R_W \times R_H} \geq 0.1 \tag{式(5.3.3)}$$

$$\frac{C_L}{C_P} \geq 0.01 \tag{式(5.3.4)}$$

$$\frac{R_H}{R_W} \geq 0.1 \tag{式(5.3.5)}$$

$$R_W \geq 2 \tag{式(5.3.6)}$$

$$R_H \geq 2 \tag{式(5.3.7)}$$

ノイズを除去した後の画像の中には、文字がいくつか存在する。それらの文字に対して傾きの補正や強いノイズの除去を行うために、画像内から文字を抽出するためのマスクを生成する。

1 単語中の文字はそれぞれが近い位置に配置されている．そこで，文字の周りに文字をちょうど囲むことができる矩形を設定し，拡大する．拡大後の文字が重なっている場合は統合する．このときの拡大は，矩形の幅，高さを  $N_{H1}$  倍することである．統合後の矩形をマスクとして，文字を抽出する．ここでは， $N_{H1} = 7/3$  とした．

#### 5.3.4 文字列傾き補正

抽出した文字の傾きを Hough 変換を用いて補正する．本研究で想定する文字列は横書きのため，傾きの範囲は $\pm 45$ 度とする．Hough 変換はすべての画素について，どの直線上にあるか計算することで，その画像内に存在する直線を検出する手法である．しかし，すべての画素に関して計算されるので，画素の増加とともに計算量が増加する．また，文字列をそのまま Hough 変換にかけると，計算量が多だけでなく，様々な傾きの直線が検出されてしまうので，文字列の傾きが一意に定まらないことも多い．加えて，画像中にはノイズが含まれていることが考えられ，ノイズが大きいとその影響も大きくなってしまう．

そこで，ノイズの影響と処理時間を抑えるために，文字画像を文字重心のみの画像に変換する．指先の近くの文字を重視するため，指先との距離が  $P_L$  以内であり，面積が  $N_{w2} \times N_{h2}$  ピクセル以上の文字の重心を抽出し，重心を中心に半径  $r$  の円（点）を描く (Fig. 67)．ここでは  $r=1$ ， $P_L = \text{画像の幅}/4$ ， $N_{w2}=5$ ， $N_{h2}=5$  とした．変換した画像に対して Hough 変換を行い傾き角度を算出，回転変換により傾きを補正する (Fig. 68)．

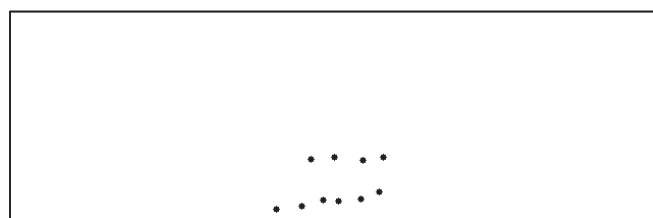


Fig. 67 傾いた文字の中心(白黒反転)



Fig. 68 傾き補正

### 5.3.5 文字候補抽出およびノイズ除去

#### (1)文字の太さに基づく処理

傾きを補正した後の画像内には、まだノイズが多数含まれている。しかし、傾きが補正されたため、文字列は横並びになっており、より強くノイズを除去することができる。そこで、文字の太さに基づいてノイズを除去する。本論文では、同じ文字列内の文字の太さは、ほぼ一様と仮定している。

文字の太さの推定方法は、まず文字候補領域の輪郭上の各画素から反対の輪郭に達するまで領域内部に向けて走査を行い測定する (Fig. 69)。このとき、走査方向は全周を45度毎・8方向に分割したうちで、輪郭に直交する方向にもっとも近い方向とする。すべての輪郭上の画素から同様の測定を行い、最頻値をその文字候補の「基準太さ  $F_s$ 」とみなす。

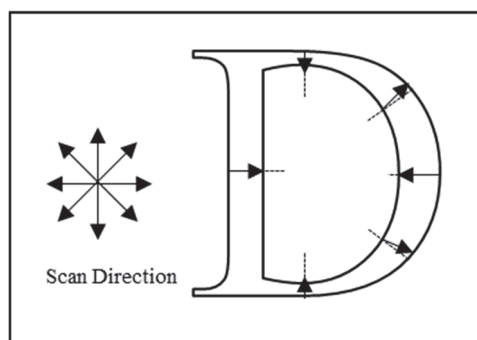


Fig. 69 文字線の太さ測定

同じ文字列内の文字の太さはほぼ一様と仮定しているため、すべての太さがこの基準太さと等しいことが理想であるが、そのような場合は少なく、ばらつきが存在する。そこで、このばらつきの許容範囲を「基準太さ  $F_s \pm N_t \cdot F_s$ 」ピクセルとする。ただし、細い文字の場合は許容範囲が過小になる場合があるため、 $N_t \cdot F_s$  の下限を5ピクセルに設定する。また、ここでは  $N_t=0.8$  とした。

文字候補領域の太さを測定し、上記の許容範囲外の候補は排除することでノイズの大部分を排除できる。しかし、走査線上にない画素も排除されるので、隙間がある文字領域が生成されてしまう。とくに文字には曲線部が多いため、この傾向が顕著である (Fig. 70)。そこで、 $N_D$  回の膨張処理を行い、隙間を埋めてマスクを作成する (Fig. 71)。文字候補画像にマスク処理を行い、文字を抽出する。ここでは  $N_D=7$  とした。



Fig. 70 文字線の太さを基準にしたマスク

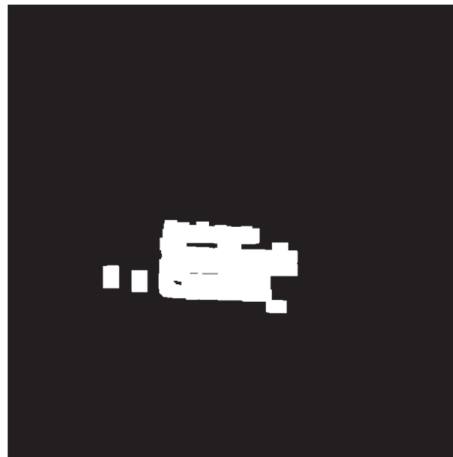


Fig. 71 膨張させたマスク

### 5.3.6 文字の高さに基づく処理

5.3.5 項の処理を行うとノイズが低減されるが、まだノイズは残留している場合もあるので、文字の並びを利用してノイズを除去する。

傾きが補正された文字はほぼ水平に並んでいることを想定しているので、文字列の高さと縦幅を調べ、その文字列の上下のノイズを除去する。

この際、複数の文字列が抽出されることがある。複数の文字列が読み上げられると、それだけ視覚障がい者への負担が増加してしまう。そこで、指先近くの文字列のみを読み上げる。

指先位置を基準に  $W_h \times H_h$  ピクセルの小領域を設定する。その領域内において横方向の画素ヒストグラムを生成する(Fig. 72)。文字列は横向きに並んでいると考えられるので、山が文字部、谷が行間に該当し、文字列の高さと縦幅が分かる。しかし、実際の文字列はすべての文字の大きさや位置がそろっている訳ではなく、「y」や「1」

のように文字列からずれるように書かれる文字も存在する．そこで，位置のばらつきを吸収するために，ヒストグラムから計算した文字列が存在する高さ（縦幅）を上下に  $N_y$  ピクセル拡大する．この範囲に入っていない文字候補はすべて除去する．

また，文字列中の文字の横幅は，行の縦幅よりも極端に大きくなることはないため，横幅が縦幅の  $N_B$  倍以上の文字はノイズとして除去する(Fig. 73)．ここで，基礎実験より  $W_h=80$ ,  $H_h=80$ ,  $N_y=2$ ,  $N_B=2$  とした．



Fig. 72 指先周辺のヒストグラム



Fig. 73 抽出される文字列

### 5.3.7 文字の隙間に基づく文字列合成処理

上記の処理では，ノイズや別の文字列が，指定した文字列と同じ高さにある場合，誤抽出される(Fig. 74)．これに対処するため，同一文字列内の文字は近接していると仮定し，文字間距離を用いて文字列をグルーピングする．

まず，それぞれの文字に外接する矩形を設定する．矩形間の距離を文字間隙間とし，



その平均を求める。文字列の高さ（縦幅）は文字の幅と等しい場合が多く、また、複数の文字の高さから求めているため、文字の形状によるばらつきが少ない(Fig. 75)。この文字間隙間平均が縦幅の  $N_{H2}$  倍より大きい場合には文字間隙間平均分だけすべての矩形の幅を拡大する(Fig. 76)。これにより、平均より大きい隙間を持つ異なる文字列に属する文字が分離される。ここで、 $N_{H2}=0.5$  とする。

しかし、この処理を行うと、逆に1つの文字列しかない場合は平均が小さくなり過ぎるため、本来同一の文字列の属する文字まで分離する可能性がある。そのため、文字間隙間平均が縦幅の  $N_{H2}$  倍より小さい場合には縦幅の  $N_{H2}$  倍だけ矩形を拡大し、文字の過分離を防止する。

その後、矩形を統合してマスクを作成、マスク処理により指先に近い文字列のみ抽出可能となる(Fig. 77)。



Fig. 74 ノイズを含んだ文字列



Fig. 75 文字列の縦幅と文字間距離

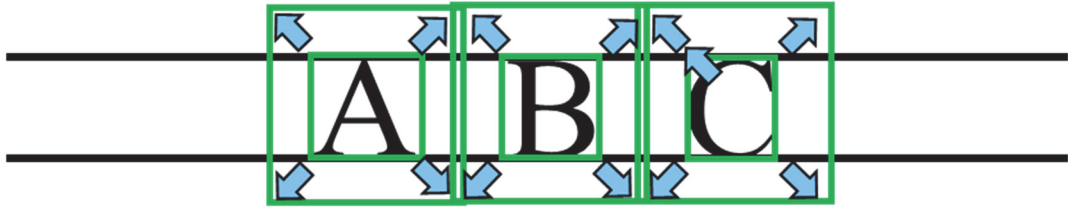


Fig. 76 文字の拡大

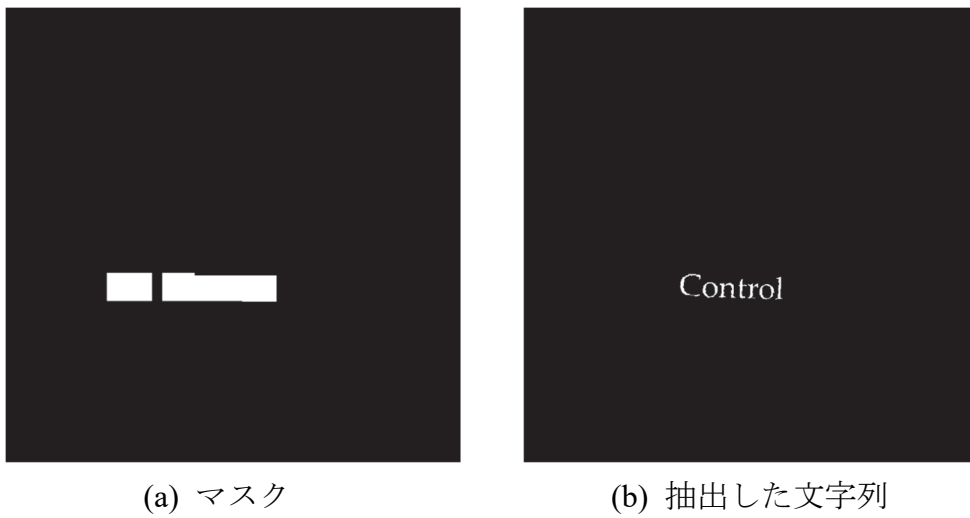


Fig. 77 指先指定文字の抽出

## 5.4 おわりに

視覚障がい者のための文字抽出手法を提案し，その詳細を述べた．

視覚障がい者に対して文字情報を与える方法として，文字の読み上げが有効である．しかし，画像処理を利用して複数の文字列が存在する環境中の文字を読み上げるとき，画面内の文字を単純に読み上げるだけになってしまうと視覚障がい者に伝わりづらいという問題があった．そこで，指先で指定した文字のみを抽出する手法を提案した．

文字は指先の色や輝度，形状情報を基に画像全体から探索され，その情報にかけ離れた文字は除外した．また，指先から遠い文字列も読み上げの対象から除外した．提案した手法により，指先から文字を抽出することに成功した．



## 6. 墨字認識システムの評価

---

6.1 はじめに.....	112
6.2 指先位置検出実験.....	113
6.3 文字抽出実験.....	117
6.4 指先指定文字抽出実験.....	120
6.5 おわりに.....	124

---

## 6.1 はじめに

提案した指先の検出精度，文字の抽出手法の精度を実験によって検証する．

指先の検出の精度は，複数の撮影条件で指先を移動させながら撮影し，検出した指先位置と実際の指先位置が一致していた場合を成功として，検証した．

文字の抽出の精度は2種類の実験によって検証した．1つ目の実験は，指先位置が文字の下にあることを前提とし，指先位置が確実に取得できている場合の文字の抽出精度を検証した．2つ目の実験は，4.1.2項で説明した指先にマーカを用いた指先検出手法で指先を検出した後に，指先周りの文字の抽出の可否をもって精度を検証した．

## 6.2 指先位置検出実験

撮影条件（照明，距離）を変化させたときの検出精度を確認し，提案した指先検出手法の精度を検証する．

撮影条件について説明する．撮影には Logicool のウェブカメラ 2-MP Portable Webcam C905m を使用した．カメラの解像度は 640×480 ピクセル，フレームレートは 30fps に設定した．フォーカスおよびホワイトバランスはカメラの機能による自動調整である．

照明条件は，JIS 照度基準を参考に，やや暗い居間と，大型店の一般陳列部を想定し，150lx，1000lx とした．カメラと指先の距離は，基礎検討として測定した，物品を把持した姿勢時の人の肩から物品までの距離である 350mm を基準とした．AIST 人体寸法データベースの上腕と前腕の長さを参考に，物品をより身体に近づけた状態として 250mm，腕を伸ばして把持した状態として 450mm についても検証した．

実験条件は以下の 6 パターンにまとめられる(Table 7)．

Table 7 指先検出実験の実験環境

	照度 lx	距離 mm
実験環境 1	150	250
実験環境 2	150	350
実験環境 3	150	450
実験環境 4	1000	250
実験環境 5	1000	350
実験環境 6	1000	450

赤外線照射範囲に対して，カラー画像用のカメラと赤外線画像用のカメラは平衡に並べて配置してある(Fig. 78(a))．そのため，カラー画像用のカメラから見ると，赤外線照射範囲は右側にずれる(Fig. 78(b)-(d)の格子)．赤外線照射範囲を，一辺が 30mm の正方形に分割した．指先を各マスの中央に置き，提案手法によって検出された指先位置がマス内に収まっている場合を成功とする．指先をマス内に置く動作を各マスにつき 30 回繰り返し，その際の各マスの検出精度を検証した．

各実験環境の実験結果を Table 8-Table 13 に示す．これらの表の各マスは，Fig. 78(b)-(d)の格子に対応している．

検出精度の平均をまとめたところ (Table 14) , どの条件においても同等の結果を確認できた. 赤外線は照明条件に左右されないため, 精度が保たれた. 失敗原因は, マーカのずれや指の折り曲げ, 赤外線の照射である. 指先の位置はマーカの位置から検出しているため, マーカのずれや指の折り曲げにより, 計算した指先位置と実際の指先位置が離れてしまうことがある. また, 赤外線は直進する性質があり, 拡散板によって多少範囲を広げているが, 照射中心から離れると赤外線が届きにくくなるため, 検出精度が低下する. 実験結果は以下の通りになった. 赤外線が十分に届く範囲に着目すれば, すべての照明条件, 撮影距離において, 指先の検出精度を保つことができた.

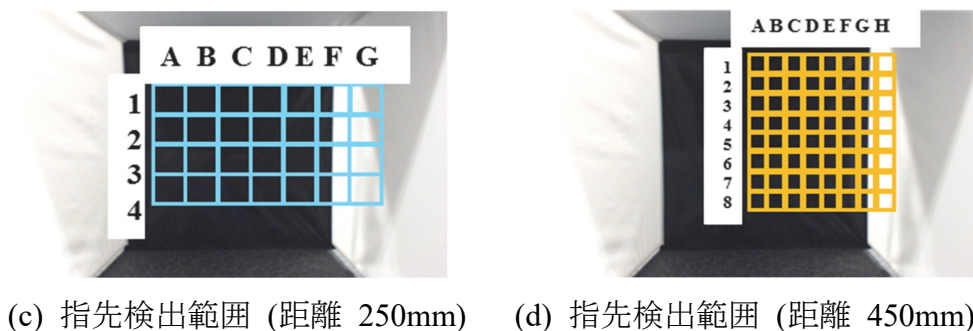
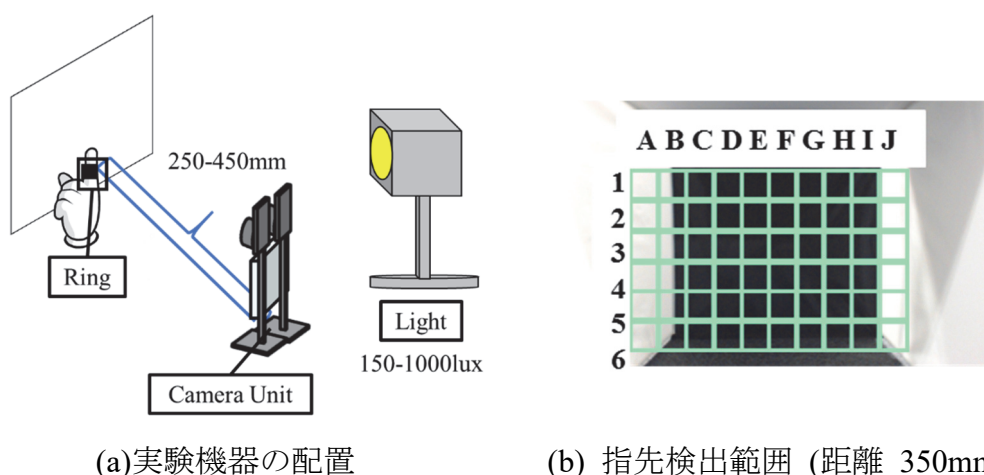


Fig. 78 指先検出実験：実験環境

Table 8 指先検出結果(実験環境 1)

	A	B	C	D	E	F	G
1	100%	93.3%	100%	100%	100%	100%	100%
2	100%	100%	100%	100%	100%	100%	100%
3	100%	100%	100%	100%	100%	100%	100%
4	100%	100%	96.7%	100%	100%	100%	100%



Table 9 指先検出結果(実験環境 2)

	A	B	C	D	E	F	G	H	I	J
1	100%	100%	100%	100%	96.7%	100%	100%	90%	100%	100%
2	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
3	100%	100%	100%	100%	100%	100%	100%	100%	96.7%	100%
4	100%	100%	100%	100%	100%	100%	100%	100%	100%	90%
5	96.7%	100%	100%	100%	100%	100%	100%	100%	100%	83.3%
6	100%	100%	100%	100%	100%	100%	100%	100%	93.3%	93.3%

Table 10 指先検出結果(実験環境 3)

	A	B	C	D	E	F	G	H
1	100%	96.7%	93.3%	100%	100%	100%	80%	100%
2	100%	100%	96.7%	100%	100%	100%	100%	83.3%
3	100%	100%	96.7%	100%	100%	100%	100%	100%
4	100%	100%	100%	80%	96.7%	96.7%	100%	76.7%
5	93.3%	100%	100%	100%	100%	100%	96.7%	63.3%
6	100%	100%	86.7%	100%	100%	96.7%	100%	63.3%
7	100%	100%	90%	100%	100%	100%	100%	70%
8	100%	83.3%	100%	93.3%	93.3%	100%	100%	100%

Table 11 指先検出結果(実験環境 4)

	A	B	C	D	E	F	G
1	90%	100%	100%	100%	100%	100%	100%
2	100%	100%	100%	100%	100%	100%	96.7%
3	100%	100%	100%	100%	83.3%	100%	100%
4	100%	100%	100%	100%	100%	100%	100%

Table 12 指先検出結果(実験環境 5)

	A	B	C	D	E	F	G	H	I	J
1	96.7%	100%	100%	100%	100%	96.7%	100%	100%	93.3%	66.7%
2	100%	100%	100%	100%	96.7%	100%	100%	100%	100%	80%
3	100%	100%	100%	100%	100%	80%	100%	100%	80%	100%
4	100%	100%	100%	100%	100%	86.7%	100%	93.3%	96.7%	100%
5	100%	100%	100%	100%	100%	100%	100%	100%	100%	96.7%
6	100%	96.7%	100%	100%	96.7%	100%	100%	100%	100%	100%

Table 13 指先検出結果(実験環境 6)

	A	B	C	D	E	F	G	H
1	96.7%	100%	100%	100%	60%	83.3%	100%	66.7%
2	96.7%	100%	100%	100%	93.3%	96.7%	100%	86.7%
3	100%	100%	100%	96.7%	96.7%	96.7%	100%	80%
4	100%	93.3%	100%	96.7%	100%	100%	93.3%	76.7%
5	96.7%	100%	100%	100%	96.7%	100%	90%	56.7%
6	100%	100%	100%	100%	96.7%	90%	93.3%	63.3%
7	96.7%	90%	100%	96.7%	100%	96.7%	96.7%	56.7%
8	100%	100%	100%	96.7%	100%	100%	96.7%	86.7%

Table 14 それぞれの実験環境の平均指先検出結果

照明条件	距離	平均精度
150lx	250mm	99.6%
150lx	350mm	99.0%
150lx	450mm	97.7%
1000lx	250mm	98.9%
1000lx	350mm	97.6%
1000lx	450mm	97.0%

### 6.3 文字抽出実験

指先位置と同様に撮影条件を変化させたときの文字抽出の精度を確認し、提案手法の有効性を検証する。本実験では、文字の抽出精度のみを検証するため、指先位置はあらかじめ与える。すなわち、指先の位置は常に正しいとみなして実験は実施された。

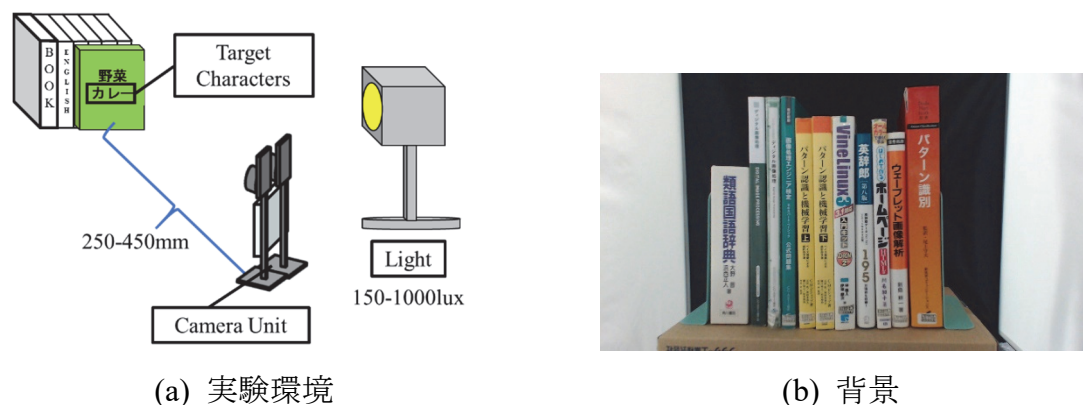


Fig. 79 文字抽出実験：実験環境

照明条件は、JIS 照度基準を参考に、やや暗い居間、大型店の店内全体、一般的な商店の重要陳列部、大型店の一般陳列部を想定し、それぞれ 150lx, 500lx, 750lx, 1000lx とした。カメラと指先の距離は、指先検出実験と同様である(Fig. 79(a)) (Table 15)。ただし、照度 500lx 以上の場合、距離は 350mm のみを検証した。複数の文字が存在する画像内から指先に近い文字のみを抽出できることを実証するため、その背景として複数の書籍を並べた(Fig. 79(b))。

実験に使用した抽出対象の文字は、10 種類の書籍、食品パッケージの表面上の文字である(Fig. 80)。それらの文字列の下部に指先があると仮定して、指定した文字が抽出できた個数を計数する。この抽出処理を 1 文字列につき 30 回繰り返し、実用性を検証した。実験の結果を Table 16 に示す。Table 16 の a,b,c…は物品を表し、Fig. 80 の各物品に対応している。Table 16 の 1,2,3…は実験環境を表し、Table 15 に対応している。

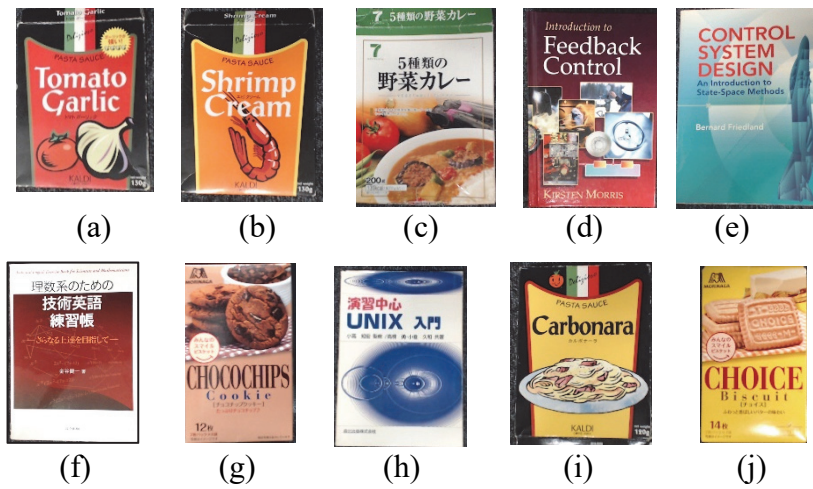


Fig. 80 実験対象

Table 15 文字抽出実験の実験環境

	照度 lx	距離 mm
実験環境 1	150	250
実験環境 2	150	350
実験環境 3	150	450
実験環境 4	500	350
実験環境 5	750	350
実験環境 6	1000	350

Table 16 文字抽出実験結果

	a	b	c	d	e	f	g	h	i	j
1	100%	93.3%	100%	100%	100%	96.7%	100%	83.3%	100%	26.7%
2	96.7%	100%	100%	100%	100%	96.7%	96.7%	96.7%	100%	83.3%
3	86.7%	96.7%	70.0%	100%	100%	90.0%	3.3%	93.3%	86.7%	0.0%
4	100%	100%	100%	86.7%	90.0%	100%	50.0%	96.7%	100%	70.0%
5	100%	96.7%	100%	100%	90.0%	100%	56.7%	96.7%	96.7%	53.3%
6	100%	100%	100%	100%	96.7%	93.3%	0.0%	100%	100%	3.3%

実験結果より、照度が十分に確保されている場合（実験環境 4-6），物品 g と j を除く物品の抽出精度は 86.7% になった。物品 g および j の文字を観察すると、文字の太さに変化が大きいフォントであり、「H」や「E」の横棒が他の文字列よりも細くなっていることが分かる。そのため、とくに明るいとき照明の影響により、この細い文字線が欠けてしまうことが精度低下の一因である(Fig. 81)。距離を 250mm まで近づけた場合(実験環境 1)，物品 g の抽出精度は改善され、83.3%の精度で文字の抽出に成功した。しかし、物品 j に関しては、近づけても細い文字線の欠けが生じた(Fig. 82)。

反対に距離を 450mm まで遠ざけた場合（実験環境 3），すべての物品の文字線も相対的に細くなる。そのため、物品 c, g, j に関して精度が低下した(Fig. 84)。しかし、この実験環境は腕を限界まで伸ばした状態を想定しており、この姿勢を維持することは難しく、現実的ではない。そのため、限界の環境での検証であると考えれば、実用上の問題はない。



(a) object g

(b) object j

Fig. 81 文字の欠け



Fig. 82 文字の欠け(250mm)



(a) object c

(b) object g

(c) object j

Fig. 83 文字の欠け(450mm)

## 6.4 指先指定文字抽出実験

システムの実用性を確認するために、指先検出処理と文字抽出処理を統合したときの文字抽出精度を検証する。

カメラと実験対象の位置関係や背景は文字抽出実験時と同様とする。操作者は指先検出実験時と同様にマーカとなる指輪を装着する。照明条件、カメラと指先の距離は、それぞれ150lx、350mmとした。

文字抽出実験と同様に実験装置を配置し(Fig. 84)、物品上の文字列の下端を指差した。この指差し動作を30回繰り返し、その間に指定した文字列が抽出できた回数を計数する(Fig. 85)。

このシステムは、CPU Intel Core i5-4300U 2.5GHz、メモリ 8.0GB、GPUなしの計算機環境で指先検出、文字抽出までに平均14秒程度を要した。しかし、使用計算機の機能向上、処理の並列化などにより将来的には撮影30fpsと同期した処理時間を目指す。

実験の結果、文字抽出実験と同様に文字線の細い物品g、jを除き、最低でも70.0%の精度で指先の文字を抽出できた(Table 17)。

指先文字の抽出の失敗の原因は、指先検出実験の際に問題となったマーカや指先のずれ(Fig. 86)と、文字抽出実験の際に問題となった文字色の推定失敗(Fig. 87)、文字列の合成失敗(Fig. 88)、傾き補正の失敗(Fig. 89)である。

マーカや指先がずれると、指先から離れた位置が指先位置としてみなされてしまうので、実際に指した文字列とは異なる文字列もしくはノイズが抽出されてしまう。

文字色の推定の失敗に関しては、指先の小領域には様々な色が含まれており、小領域内の色の分布や配置によっては、色ごとに2値化したときに文字としての評価が高くなってしまう場合もある。文字列は同一の文字列内であれば、大きさはほぼ等しいと仮定して合成している。そのため、文字列内に異なる大きさの文字が含まれる場合に発生した。文字の傾きは文字の重心の並びから計測し補正している。文字列の文字数が少ない場合、傾きの計測のずれが大きくなるため、傾きの補正の失敗が見られた。

操作者の立場からすると、失敗が発生しないシステムが理想である。しかし、処理を繰り返す中で、一度でも正しく読み上げることができれば、視覚障がい者の支援となる。また、実環境での使用を想定した場合、指先の検出や文字の抽出に不具合が生じることが考えられる。その場合、指先を少し移動させることで指先と文字の検出・抽出に成功する可能性があり、実用的ではないかと考える。

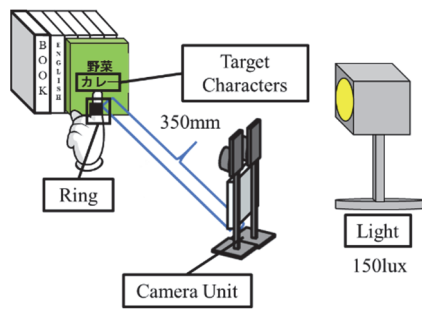


Fig. 84 指先指定文字抽出実験：実験環境

Table 17 指先指定文字抽出実験：結果

Target	Accuracy
a	70.0%
b	76.7%
c	100%
d	100%
e	100%
f	93.3%
g	30.0%
h	96.7%
i	100%
j	26.7%

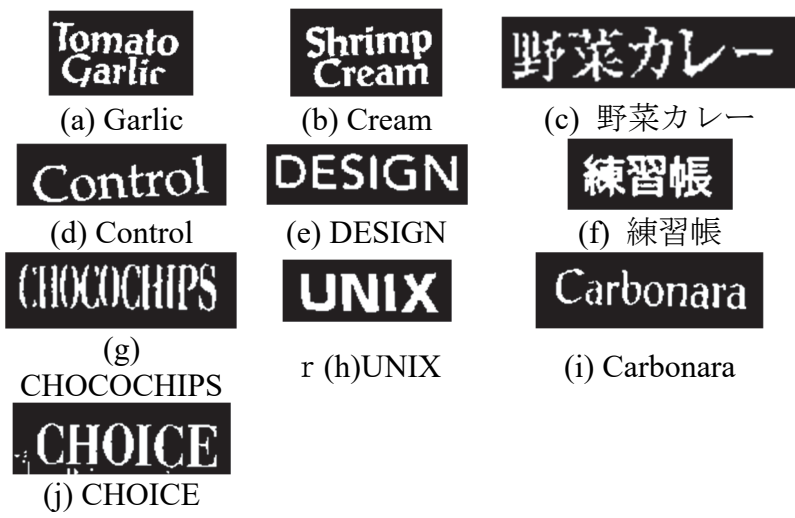


Fig. 85 成功例



Fig. 86 指先検出の失敗



Fig. 87 文字色推定の失敗



Fig. 88 文字列生成の失敗



Fig. 89 傾き補正の失敗

さらに、抽出した画像を OCR や文字認識用アプリケーションを用いて認識した。本研究で使用した OCR はパナソニック株式会社の読取革命 Ver.15, 使用した文字認識用アプリケーションは Google, Inc の Google 翻訳である。それぞれの認識精度を Table 18, Table 19 に示す。

それぞれの認識結果を確認すると、認識するアプリケーションの違いによって認識結果に差が生じていることが分かる。読取革命では、「CHOCOTIPS」g と「CHOICE」j の他に「Garlic」a, 「Cream」b, 「練習帳」f の認識精度が著しく低下しているが、Google 翻訳では g と j を除けば、少なくとも 56.7% の認識精度を保っている。これは、今後のアプリケーションの改良によって、認識精度は向上することが示唆されている。また、1 フレームで文字を認識するのではなく、複数フレームを用いて認識することを考えれば、繰り返し処理の中で 1 度でも正確に認識できれば視覚障がい者へ情報を伝達することができるため、最低限の認識精度を保つことができたと言える。



Table 18 読取革命：認識結果

Target	Accuracy
a	0%
b	10.0%
c	66.7%
d	100%
e	96.7%
f	40.0%
g	20.0%
h	80.0%
i	90.0%
j	0%

Table 19 Google 翻訳：認識結果

Target	Accuracy
a	56.7%
b	73.3%
c	86.7%
d	100%
e	93.3%
f	86.7%
g	40.0%
h	90.0%
i	93.3%
j	16.7%

## 6.5 おわりに

視覚障がい者向け文字抽出・読み上げシステムの性能を調べるために、3 種の実験を行った。

第1の実験である指先位置検出実験では、照明条件や撮影距離を変更しながら赤外線照射範囲の中で指先の検出精度を確認した。その結果、すべての実験条件で平均97%以上の検出精度を認めることができた。この指先検出手法は赤外線の反射に基づいているため、強い赤外線を含まない一般的な照明の影響は全く受けない。また、赤外線は直進する性質があるため、拡散板を使用していても十分にマークに照射することが可能である。この性質によって、距離が離れていても精度の低下がほとんど発生しなかったと考えられる。

第2の実験である文字抽出実験では、指先はすでに正しい位置にあると仮定して照明条件や撮影距離を変更しながら想定した指先周辺の文字を抽出した。その結果、文字線が細い文字列を除けば70%以上の精度で文字列を抽出することができた。指先位置の色や文字の大きさなどの文字の特徴を指先近くの文字とノイズの判別に利用することにより、文字の抽出精度が保たれたと考えている。文字の特徴として、文字線の太さは一定であるという条件を加えたが、文字のフォントによってはこの条件が該当しないこともあり、文字の抽出に失敗した。

第3の指先指定文字抽出実験は、上記の2つの実験を組み合わせ、提案手法の総合的な効果を調査した。指先は、指先検出実験と同様に赤外線とマークを用いて検出し、文字は文字抽出実験と同様の手法によって抽出した。その結果、指先検出の失敗や文字抽出実験の際にも見られた細い文字線の抽出の失敗によって、指先で指定した文字の抽出に失敗したが、細い文字線を持つ文字列以外に関しては70%以上の精度で文字を抽出することができた。また、OCRや文字認識用アプリケーションを用いて抽出した文字を認識した結果、文字認識用アプリケーションを用いた場合に細い文字線を持つ文字列以外に関しては56.7%以上の精度で文字を認識することができた。

これらの結果より、提案した指先指定文字抽出手法は、複数の文字が混在する環境から文字を抽出する手法として有効であると言える。

## 7. 結論

本博士論文では、視覚障がい者の日常生活支援システムの実現を目的として、地面性状識別と文字情報取得の支援システムを提案した。

地面性状識別手法に関して、ロボットのための地面性状識別システムから得た知見に基づき、視覚障がい者の方へ転用した場合の貢献は以下の通りである。

- (1) カメラという広く普及している導入が容易なセンサからの画像情報のみを用いて、地面性状を識別する手法を提案した。カメラのみを用いると、識別は難しくなるが、形状の特徴と色の特徴を統合した特徴を用いてそれぞれの地面性状を比較することで、90%以上の精度で識別が可能になった。また、その処理速度も GPU を搭載していないノート PC を用いても 1 フレーム当たり 1 秒未満となり、計算機の面からも導入が容易なシステムを提案できた。
- (2) 視覚障がい者の方が外出する際に特に識別が必要な 4 種類の地面性状を画像のみを用いて識別できる手法を提案した。

以上のように、提案手法は、実験を通して前方の地面性状を識別できることを示した。4 種類の地面（芝生、砂利、コンクリート、ウッドデッキ）を識別した結果、再現率、適合率ともに 90%以上であることを確認した。

指先指定文字抽出に関して、視覚障がい者の日常生活支援における貢献は以下の通りである。

- (1) 指先検出に関しては、照明の影響を避けるために、赤外線とそれをよく反射するマーカ、赤外線のみを撮影するカメラを利用して照明の変化に頑健な手法とした。
- (2) 文字抽出に関しては、文字色推定、文字列傾き補正、文字候補抽出およびノイズ除去を組み合わせた手法を提案し、商品パッケージに書かれた様々なフォントの文字に対応した。文字列の傾きは、指先近くの文字の傾きのみを計測する手法を提案し、複数の文字が混在する環境での傾き補正を可能とした。文字候補抽出およびノイズ除去は文字や指先近くに書かれた横書きの文字という特徴を用いて、指定した文字列以外の冗長な情報を除外することができた。

提案手法は、実験を通して複数の文字が混在する環境から指先で指定した文字を、文字線が細い文字を除いて抽出できることを示した。様々な照明条件下で指先を検出した結果、十分な赤外線が照射されている範囲では90%以上で検出できた。また、様々な照明条件下で10種の商品パッケージ上の文字を抽出した結果、細い文字線を持つ文字を除くと、70%以上で抽出できた。

地面性状の識別は視覚障がい者の歩行速度1m/sを基準に、1フレーム当たり1秒以下での処理が必要であるが、それを満たしていることから、速度は十分であると考えられる。精度に関しては、100%の精度が望ましいが、提案システムは視覚障がい者が地面性状の識別の際に利用する情報の全てではなく、白杖や盲導犬との併用を考えている。提案システムによって、視覚障がい者の転倒や躓きを少しでも防ぐことができれば有用である。その中で、90%以上の精度での検出は、最低限の精度は満たしていると考えられる。

指先指定文字の抽出は、物品を触れながら文字を抽出可能なシステムとなっている。また、精度は70%程度であるが、物品をカメラに近づける、もしくは繰り返し処理することによって、その精度は向上する。さらにOCRや文字認識用アプリケーションを用いて抽出した文字を認識すると、文字認識用アプリケーションを用いた場合に細い文字線を持つ文字列以外に関しては56.7%以上の精度で文字を認識することができた。提案システムは繰り返し処理の中で一度でも正確な処理ができれば情報を伝達することができると考えられ、その有効性を示した。

これらのシステムは本論文では静止画を対象としている。しかし、将来的に動画への適用も考慮すると、処理速度や判断方法にもよるが、複数の識別結果から最終的な結果を取捨選択することができるため、精度を向上させることもできると考える。

以上、視覚障がい者の日常生活支援システムとして、地面性状識別と指先指定文字抽出システムを提案した。実験を通して、それぞれのシステムの有効性を示した。本博士論文の背景で述べた、視覚障がい者の些細な不満を無視せず、工学的なアプローチで多少なりとも解消することができたと考える。

今後の展望は、地面性状識別手法に関しては、画像内に撮影される前方の地面性状と、それ以外の背景を判別し、より視覚障がい者に必要な地面性状のみを識別する手法を提案する必要がある。

また、本研究では地面性状の識別に着目したが、視覚障がい者への伝達も重要な課題である。つまり、視覚障がい者へ伝える方法（振動や音、電気刺激、音声など）や伝える内容（振動パターンや音階、刺激の強さ、文章など）についての検討が必要である。そこで、今後は画像処理以外の分野の知見も活用していくことを考えている。

指先指定文字抽出手法に関しては、弱視者の方のように少しは文字の場所が認識できる場合は文字を指すことも容易であると考えられるが、全盲の方の場合では手探りに探す必要がある。生鮮食品など指先で刺激すると傷んでしまう物品も存在するため、その対策を用意する必要がある。また、提案手法には設定するパラメータが多いため、なるべく減らす必要がある。

また、本論文で提案した文字認識手法は文字をほとんど2次元平面に書かれているものとして扱っている。しかし、実際には文字は3次元物体の表面に書かれている。そのため、文字の傾きは2次元平面上では単純な回転による補正が有効であったが、3次元物体では物体の形状やカメラと文字の位置関係を考慮したうえで補正を行う必要がある。

# 謝辭

本論文は東京電機大学大学院先端科学技術研究科先端技術創成専攻知能機械システム研究室における研究成果を編集したものです。本論文を締めくくるにあたり、この場を借りてご指導いただきました全ての方々へ感謝の意を伝えたいと思います。

はじめに、筆者の指導教員である

東京電機大学 未来科学部 ロボット・メカトロニクス学科 教授 中村明生先生

に感謝の意を表したいと思います。博士課程の4年間と研究生の1年間、様々なお迷惑をおかけしましたが、常に私のためになる数々のご指導・ご鞭撻に感謝いたします。中村先生とは、東京電機大学に入学してアドバイザーとしてお世話になってから10年近くご指導いただきました。研究に対する取り組み方だけでなく、社会人としてのマナーや振る舞い、後輩への指導方法、他大学の先生方へのご紹介など数々の事柄について学ばせていただきました。ここに深く謝辞を述べさせていただきます。

また、本論文の副査を快く引き受けてくださいました

東京電機大学 未来科学部 ロボット・メカトロニクス学科 教授 畠山省四朗先生

東京電機大学 未来科学部 ロボット・メカトロニクス学科 教授 石川潤先生

東京電機大学 未来科学部 ロボット・メカトロニクス学科 准教授 岩瀬将美先生

東京大学大学院 工学系研究科 精密工学専攻 准教授 山下淳先生

には、本論文への数々のご助言だけでなく、様々なお指導を頂きました。

畠山先生には、中村先生と同様に学部生時代から勉学、マナー、研究へのご助言、ご指導に留まらず、ときには人生に関して様々なお助言を頂きました。私が困難に直面した際も親身にご相談に乗ってくださいました。この場を借りて深く感謝申し上げます。

石川先生には、学部時代の勉学の他、留学の際に様々な面でご助力いただきました。この場を借りて深く感謝申し上げます。

岩瀬先生には、留学中にご訪問された際に、いろいろ気にかけていただき、また、学校生活におきましても、頻繁に声をかけてくださり、感謝申し上げます。

山下先生には、現在勤務しております東京大学 浅間研究室、山下研究室におきまして、私がスムーズに他のスタッフや学生になじめるようにご助力いただきました。プロジェクトにおきましても、他大学の先生方との交渉や準備に関するご指導、ご助力いただきました。この場を借りて深く感謝申し上げます。



また、現在、東京大学でお世話になっております東京大学大学院 工学系研究科 精密工学専攻 教授 浅間一先生には、東京大学で働き始めてからも、私が博士号の取得にも集中できるように、様々なご配慮を頂きました。さらに、現在取り組んでおります ImPACT タフ・ロボティクス・チャレンジにおきましても様々なご助言、ご指導を頂きました。この場を借りて深く感謝申し上げます。

また、Singapore University of Technology and Design Assistant Professor Mohan Rajesh Elara 先生には、留学中に研究に関するご指導の他に、滞在に関する手続きや日常生活について様々なご助力を頂きました。この場を借りて深く感謝申し上げます。

中村研究室、浅間研究室、山下研究室、Mohan 研究室の皆様には様々なご協力、ご助言をいただきましたこと、大変に感謝しております。この場を借りまして感謝いたします。

最後になりますが、いつも暖かく見守っていただきました両親と兄弟に感謝いたします。

2018年3月  
淵田 正隆



## 参考文献

[厚生労働省 2008]

厚生労働省社会・援護局障害保健福祉部企画課, “平成 18 年身体障害児・者実態調査結果”, 厚生労働統計協会, pp.24-26, 2008.

[共用品推進機構 2011]

財団法人 共用品推進機構, “2010 年度(平成 22 年度)視覚障害者不便さ調査成果報告”, pp.131-132, 2011.

[内閣府 2005]

内閣府, “平成 17 年度 高齢者の住宅と生活環境に関する意識調査結果”, [http://www8.cao.go.jp/kourei/ishiki/h17\\_sougou/index2.html](http://www8.cao.go.jp/kourei/ishiki/h17_sougou/index2.html) (accessed on 2018. 1. 4.), 2005.

[厚生労働省 2010]

厚生労働省, “身体障害者障害程度等級表”, 2009.

[Weiss 2006]

C. Weiss, H. Frohlich and A. Zell, “Vibration-based Terrain Classification using Support Vector Machines”, In Proceedings of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 4429–4434, 2006.

[Lu 2009]

L. Lu, C. Ordonez, E. G. Collins Jr and E. M. DuPont, “Terrain Surface Classification for Autonomous Ground Vehicles using a 2D Laser Stripe-based Structured Light Sensor”, In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems(IROS 2009), pp. 2174–2181, 2009

[Zenker 2009]

S. Zenker, E. Aksoy, D. Goldschmidt, F. Worgotter and P. Manoonpong, “Visual Terrain Classification for Selecting Energy Efficient gaits of a Hexapod Robot”, In Proceedings of the 2013 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM), pp. 577–584, 2009.

[Shi 2016]

B. Shi, X. Wang, P. Lyu, C. Yai and X. Bai, “Robust Scene Text Recognition with Automatic Rectification”, Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition CVPR, pp. 4168-4176, 2016.

[Goto 2009]

H. Goto and M. Tanaka, “Text-tracking Wearable Camera System for the Blind”, Proceedings of International Conference on Document Analysis and Recognition, pp.136-141, 2009.

[平岩 2002]

平岩 裕康, 竹内 義則, 松本 哲也, 工藤 博章, 劉 詠梅, 大西 昇, “視覚障がい者のための環境内の文字情報理解支援システム”, 電子情報通信学会技術研究報告.TL, 思考と言語, Vol.101, No.710, pp.195-202, 2002.

[青柳 2015]

青柳まゆみ, 鳥山由子, “視覚障害教育入門 一改訂版—”, ジアース教育新社, pp. 108, 2015.

[Soeda 2004]

K. Soeda, S. Aoki, K. Yanashima, and K. Magatani, “Development of the Visually Impaired Person Guidance System using GPS”, Proceeding of International Conference on Engineering in Medicine and Biology Society (IEMBS2004), pp.4870-4873, 2004.

[Kim 2009]

K. Kim, H. Kim, G. Yun and M. Kim, “Design and Analysis of Ultrasonic Monaural Audio Guiding Device for the Visually Impaired”, Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual Int. Conf. of the IEEE, pp.1577-1580, 2009.

[Liao 2013]

C. Liao, P. Choe, T. Wu, Y. Tong, C. Dai and Y. Liu, “RFID-based Road Guiding Cane System for the Visually Impaired”, Cross-Cultural Design. Methods, Practice, and Case Studies. 1st ed. Berlin Heidelberg: Springer, pp. 86-93, 2013.

[日本精工 2015]

飛田 和輝, 嵯峨山 巧幸, “視覚障がい者向けガイダンスロボットの実証実験”, 日本ロボット学会誌, Vol.33, No.8 , pp.596-599, 2015.

[Sekiguchi 2014]

M. Sekiguchi, K. Ishiwata, M. Fuchida and A. Nakamura, “Development of a Wearable System for Navigating the Visually Impaired in the Indoor Environment –A Prototype System for Fork Detection and Navigation-”, IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN 2014), pp.549-554, 2014.

[Bai 2014]

Y. Bai, W. Jia, H. Zhang, Z.-H. Mao, and M. Sun, “Landmark-Based Indoor Positioning for Visually Impaired Individuals”, Signal Processing (ICSP), 2014 12th International Conference, pp. 668-671, 2014.

[神藪 2001]

神藪 寿, 宇根 正美, 上月 秀徳, “視覚障がい者のための誘導システムの開発—駅のプラットホーム用転落防止システムの開発(第一報)—”, 福祉のまちづくり工学研究所研究報告集, 2001.

[Ulrich 2001]

I. Ulrich and J. Borenstein, “The Guidecane Applying Mobile Robot Technologies to Assist the Visually Impaired People”, IEEE Trans. Syst. Man Cybern. A: Syst. Hum., Vol.31, No.2, pp.131-136, 2001.

[Takizawa 2015]

H. Takizawa, S. Yamaguchi, M. Aoyagi, N. Ezaki, S. Mizuno and K. Cane, “An Assistive System for the Visually Impaired Based on the Concept of Object Recognition Aid,” Pers. Ubiquitous Comput, Vol.19, pp.955-965, 2015.

[Ueda 2006]

T. Ueda, H. Kawata, T. Tomizawa, A. Ohya and S. Yuta, “Visual Information Assist System using 3D SOKUIKI Sensor for Blind People—System Concept and Object Detecting Experiments”, IECON’06 the 32nd annual conference of the IEEE industrial electronics society proceedings, pp.3058–3063, 2006.

[Ishiwata 2013]

K. Ishiwata, M. Sekiguchi, M. Fuchida and A. Nakamura, “Basic Study on Step Detection System for the Visually Impaired”, Proceedings of the IEEE International Conference on Mechatronics and Automation (ICMA2013), pp.1332-1337, 2013.

[Bahadir 2012]

S. K. Bahadir, V. Koncar and F. Kalaoglu, “Wearable Obstacle Detection System Fully Integrated to Textile Structures for Visually Impaired People”, Sensors and Actuators A: Physical, Vol.179, No.0, pp. 297-311, 2012.

[Setiawan 2014]

W. Setiawan, Rasim and J. Kusnendar, “Development of Glasses for Guiding Visually Impaired Using Ultrasonic Sensor and Microcontroller”, Journal of Multidisciplinary Engineering Science and Technology (JMEST), Vol. 1 Issue 5, pp.67-73, 2014.

[Cardin 2007]

S. Cardin, D. Tnalmann and F. Vexo, “A Wearable System for Mobility Improvement of Visually Impaired People”, Visual Comput, Vol. 23, No. 2, pp. 109-118, 2007.

[今井 2017]

今井健太, 北原格, 亀田能成, “RGB-D カメラを用いた歩行安全領域の検出と提示方法討”, HCG シンポジウム, 2017.

[Leung 2014]

T. S. Leung, G. Medioni, “Visual Navigation aid for the Blind in Dynamic Environments”, Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops(CVPRW). pp.579–586, 2014.

[Best 2013]

G. Best, P. Moghadam, N. Kottege and L. Kleeman, “Terrain Classification using a Hexapod Robot”, In Proceedings of the Australasian Conference on Robotics and Automation, 2013.

[Ascari 2006]

L. Ascari, M. Ziegenmeyer, P. Corradi, B. Gaßmann, M. Zöllner, R. Dillmann and P. Dario, “Can Statistics Help Walking Robots in Assessing Terrain Roughness? Platform Description and Preliminary Considerations”, In Proceedings of the 9th ESA Workshop on Advanced Space Technologies for Robotics and Automation (ASTRA 2006), pp.28–30, 2006.

[森 2004]

森 英雄, 安部 圭祐, 竹谷 哲也, 依田 一郎, 小谷 信司, “視覚障がい者向け携帯型歩行支援装置による駅における誘導,” 電子情報通信学会技術研究報告. HCS, ヒューマンコミュニケーション基礎 Vol.103, No.587, pp.1-6, 2004.

[Wang 2014]

S. Wang, H. Pan, C. Zhang, and Y. Tian, “RGB-D Image-Based Detection of Stairs, Pedestrian Crosswalks and Traffic Signs”, Journal of Visual Communication and Image Representation, Vol.25, No.2, pp. 263-272, 2014.

[Bajrachary 2008]

M. Bajracharya, B. Tang, A. Howard and M. Turmon, “Learning Long-range Terrain Classification for Autonomous Navigation”, In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA 2008), pp. 4018–4024, 2008.

[Ros 2016]

G. Ros, L. Sellart, J. Materzynska, D. Vazquez and A. M. Lopez, “The Synthia Dataset: A Large Collection of Synthetic Images for Semantic Segmentation of Urban Scenes”, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3234-3243, 2016.

[高梨 2012]

高梨 宏之,王 鋒,田中 真美,長南 征二,御室 哲志,“携帯型点字読み取り支援装置の開発”,日本機械学会論文集 C 編 78.790, pp.2143-2151, 2012.

[Zhang 2010]

H. Zhang, J. Li and I. Yin, “A Research on Paper Mediated Braille Automatic Extraction Method”, Intelligent International Conference on Computation Technology and Automation (ICICTA), Vol.1, pp. 328-331, 2010.

[土井 2011]

土井 泰法, 松本 哲也, 竹内 義則, 工藤 博章, 大西 昇, “視覚障がい者の商品情報取得を支援する画像-音声変換システム”, 映像情報メディア学会技術報告 Vol.35, No.16, pp.5-8, 2011.

[Gallo 2011]

O. Gallo, R. Manduchi, “Reading 1D Barcodes with Mobile Phones Using Deformable Templates”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.33, No.9, pp.1834-1843, 2011.

[Guo 2014]

X. Guo, H. Lin, Z. Yu, S. McCloskey, “Barcode Imaging using a Light Field Camera”, Light Fields for ComputerVision ECCV Workshops, 2014.

[深見 2004]

深見 拓史, “テキストの2次元コード化と音声によるバリアフリー・コミュニケーション・システム”, 情報処理学会研究報告デジタル・ドキュメント], Vol.2004, No. 97(2004-DD-046), pp.29-34, 2004.

[芦田 2005]

芦田 和毅, 永井 弘樹, 岡本 正行, 宮尾 秀俊, 山本 博章, “情景画像からの文字抽出”, 電子情報通信学会論文誌 D-II, 情報・システム, II -パターン処理, Vol. J88-D-II, No.9, pp.1817-1824, 2005.

[Koga 2005]

M. Koga, R. Mine, T. Kameyama, T. Takahashi, M. Yamazaki, T. Yamaguchi, “Camera-based Kanji OCR for Mobile-phones: Practical Issues”, Proceedings of the Eighth International Conference on Document Analysis and Recognition, pp. 635-639, 2005.



[Ohya 1994]

J. Ohya, A. Shio, and S. Akamatsu, “Recognizing Characters in Scene Images”, IEEE Transaction PAMI Vol.16, No.2, pp.214-220, 1994.

[田畑 2010]

田畑 晶啓, 藤崎 浩太, 前田 純治, “自己組織化マップを用いた文書画像のカラー量子化”, 電子情報通信学会技術研究報告. IE, 画像工学 Vol.109, No.415, pp.95-100, 2010.

[藤崎 2011]

藤崎 浩太, 坂元 佑一郎, 前田 純治, “複雑な背景を持つカラー文書画像からの文字抽出”, 映像情報メディア学会技術報告 Vol.35, No.9, pp.221-226, 2011.

[Kita 2010]

K. Kita and T. Wakahara, “Binarization of Color Characters in Scene Images Using K-Means Clustering and Support Vector Machines”, Proceedings of International Conference on Pattern Recognition, pp.3183-3186, 2010.

[村山 2001]

村山 健二, 岡田 至弘, “色の対比を用いた情景画像からの文字列抽出”, 電子情報通信学会技術研究報告. PRMU, パターン認識・メディア理解 Vol.100, No.702, pp.37-42, 2001.

[Cho 2016]

H. Cho, M. Sung and B. Jun, “Canny Text Detector: Fast and Robust Scene Text Localization Algorithm”, Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition CVPR, pp. 3566-3573, 2016.

[松田 2010]

松田 友輔, 大町 真一郎, 阿曾 弘具, “2 値化とエッジ抽出による情景画像からの高精度文字列検出”, 電子情報通信学会論文誌 D, 情報・システム, Vol.J93-D, No.3, pp.336-344, 2010.

[Wernicke 2000]

A. Wernicke, R. Lienhart, “On the Segmentation of Text in Video”, Proc. of the IEEE Int. Con. on Multimedia Expo, Vol. 3, pp. 1511-1514, 2000.

[Nakamura 1993]

A.Nakamura, O.Shiku, M.Anegawa, C.Nakamura and H.Kuroda, “A Method for Recognizing Character Strings from Maps using Linguistic Knowledge”, Proc. of the Second International Conf. on Document Analysis and Recognition (ICDAR), pp.561-564, 1993.

[Bissacco 2013]

A. Bissacco, M. Cummins, Y. Netzer and H. Neven, “Photoocr: Reading Text in Uncontrolled Conditions”, Proceedings of the IEEE International Conference on Computer Vision, pp. 785-792, 2013.

[Morris 2006]

T. Morris, P. Blenkhorn, L. Crossey, Q. Ngo, M. Ross, D. Werner, and C. Wong, “Clearspeech: A Display Reader for the Visually Handicapped”, IEEE Transactions on Neural Systems and Rehabilitation Engineering, Vol.14, No.4, pp.492-500, 2006.

[Clark 2002]

P. Clark and M. Mirmehdi, “Recognising Text in Real Scenes”, International Journal on Document Analysis and Recognition, Vol.4, No.4, pp.243-257, 2002.

[久野 2011]

久野 素有, 山下 淳, 金子 透, ” ステレオカメラを用いたタッチパネル操作支援システムの構築”, 電気学会論文誌. D, 産業応用部門誌 Vol.131, No.4, pp.458-465, 2011.

[島田 2008]

島田 茂伸, 篠原 正美, 安彦 成泰, 下条 誠, “オプタコンの機械特性と人間の触覚特性との適合度に関する研究”, 電子情報通信学会論文誌. D, 情報・システム Vol.J91-D, No.5, pp.1296-1304, 2008.

[Suzuki 2011]

S. Suzuki, K. Kato and K. Yamamoto, “A Consideration of Effective Feature Extraction method for Nail Camera System”, Workshop Frontiers of Computer Vision, pp.1-6, 2011.

[西川 1999]

西川 誠, 森田 哲也, 吉川 大弘, 鶴岡 信治, “カラー画像処理を用いた指先による文字列・図表領域の指示法”, 電子情報通信学会技術研究告, パターン認識・メディア理解, パターン認識・メディア理解 98, 528, pp.25-32, 1999.

[Fujishima 2013]

N. Fujishima and K. Hoshino, “Fingernail Detection Method from a Hand Images including Palm”, Proceedings of International Conference on Machine Vision Applications, pp.117-120, 2013.

[肥川 2009]

肥川 宏臣, 藤村 絃匡, 佐藤 大輔, “ハードウェア向き指文字認識アルゴリズム”, 電子情報通信学会論文誌 D, 情報・システム, J92-D, 3, pp.405-416, 2009.

[澤 2010]

澤 光映, 竹川 佳成, 寺田 努, 塚本 昌彦, “演奏ルールを用いたウッドベースのための実時間運指取得システムの設計と実装”, 日本ソフトウェア科学会, コンピュータソフトウェア, 27, 1, pp.56-66, 2010.

[田邊 2013]

田邊 正明, 森本 剛史, “白杖の長さとは歩行速度・歩幅”, 視覚リハビリテーション研究, Vol.3, No.1, pp.45-49, 2013.

[Bay 2008]

H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, “Speeded-Up Robust Features (SURF)”, Computer Vision Image Underst, 110, pp. 346–359, 2008.

[神寫 2003]

神寫 敏弘, “データマイニング分野のクラスタリング手法(1) – クラスタリングを使ってみよう! –”, 人工知能学会誌, Vol.18, No.1, pp.59-65, 2003.

[Nowak 2006]

E. Nowak, F. Jurie, and Bill Triggs, “Sampling Strategies for Bag-of-Features Image Classification”, ECCV, pp.490-503, 2006.

[Pra 2014]

Y. De Pra, F. Fontana, and L. Tao, “Infrared vs. Ultrasonic Finger Detection on a Virtual Piano Keyboard”, International Computer Music Association (ICMC), 2014.

[Farnelback 2003]

G. Farneback, “Two-frame Motion Estimation Based on Polynomial Expansion”, In Scandinavian Conference on Image Analysis, 2003.

[AIST 2012]

河内まき子：AIST 日本人の手の寸法データ，

<http://riodb.ibase.aist.go.jp/dhbodydb/hand/index.html> accessed on 2016. 5. 8.

[淵田 2012]

M. Fuchida, R. Abe and A. Nakamura, “Character Extraction from the Region Pointed at with a Fingertip for the Visually Handicapped”, Proceedings of the 21st IEEE International Symposium on Robot and Human Interactive Communication, pp.265-270, 2012.

[Sharma 2005]

G. Sharma, W. Wu, and E. N. Dalal, “The CIEDE2000 Color-Difference Formula: Implementation Notes Supplementary Test Data, and Mathematical Observations”, Color Research and Application, Vol. 30, Issue 1, pp.21-30, 2005.

# 業績リスト

A. 査読付投稿論文(英文)

1. Masataka Fuchida, Rajesh Elara Mohan, Ning Tan, Akio Nakamura and Thejus Pathmakumar, "Terrain Perception in a Shape Shifting Rolling-Crawling Robot," MDPI Robotics, Vol.5, No.19, pp.(1-14), 2016.
2. Masataka Fuchida, Thejus Pathmakumar, Rajesh Elara Mohan, Ning Tan and Akio Nakamura, "Vision Based Perception and Classification of Mosquitoes Using Support Vector Machine," MDPI Applied Sciences, Vol.7, No.51, pp.(1-12), 2017.

B. 査読付投稿論文(和文)

1. 淵田 正隆, 中村 明生, "視覚障がい者のための指差し文字検出システムの開発", 精密工学会誌, Vol.82, No.12, pp.1109-1118, 2016.

C. 査読付講演論文

1. Masataka Fuchida, Ryo Abe and Akio Nakamura, "Character Extraction from the Region Pointed at with a Fingertip for the Visually Handicapped," IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN 2012), pp.265-270, 2012.
2. Koichi Ishiwata, Mari Sekiguchi, Masataka Fuchida, and Akio Nakamura, "Basic Study on Step Detection System for the Visually Impaired," Proceedings of the IEEE International Conference on Mechatronics and Automation (ICMA2013), pp.1332-1337, 2013.
3. Mari Sekiguchi, Koichi Ishiwata, Masataka Fuchida and Akio Nakamura, "Development of a Wearable System for Navigating the Visually Impaired in the Indoor Environment –A Prototype System for Fork Detection and Navigation–," IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN 2014), pp.549-554, 2014.
4. Shin'ichi Sato, Naomichi Kobayashi, Yudai Miyashita, Masataka Fuchida, and Akio Nakamura, "Basic Evaluation on Soccer Inside-kick Proficiency," Proceedings of the 10th International Conference on Information, Communications and Signal Processing (ICICS 2015), pp.1-5, 2015.
5. Masataka Fuchida, Hiroya Yatsuyanagi, Akio Nakamura, Atsushi Yamashita and Hajime Asama: "The Characters Extraction System by a Fingertip for Visually Handicapped People", Proceedings of the 36th JSST Annual International Conference on Simulation Technology (JSST2017), Tokyo (Japan), pp.310-311, 2017.
6. Hiroya Yatsuyanagi, Masataka Fuchida, Kazushige Okayasu, and Akio Nakamura, "Development of a Specific Word Recognition System for the Visually Handicapped —Character Recognition based on Dataset Generated from Font Data —," IEEE/SICE International Symposium on System Integration (SII 2017), pp.(1-6), 2017.

#### D. 口頭発表

1. 淵田 正隆, 中村 明生, “指先指定文字抽出の検討”, 第 17 回画像センシングシンポジウム(SSII2011)講演論文集, pp.IS3-08-1-4, 2011.
2. 淵田 正隆, 中村 明生, “視覚障がい者のための指先指定文認識の検討”, 精密工学会・画像応用技術専門委員会 サマーセミナー2011 予稿集, Vol.20, pp.13-16, 2011.
3. 淵田 正隆, 中村 明生, “視覚障がい者のための指先指定文認識の検討”, ビジョン技術の実利用化ワークショップ 2011 (ViEW2011) 講演論文集, pp.404-410, 2011.
4. 淵田 正隆, 中村 明生, “視覚障がい者のための指先指定文認識の検討”, 第 18 回画像センシングシンポジウム(SSII2012)講演論文集, pp.IS3-18-1-6,(2012).
5. 淵田 正隆, 中村 明生, “視覚障がい者のための文字検出手法の検討”, 精密工学会・画像応用技術専門委員会 サマーセミナー2012 予稿集, Vol.21, pp.79-82, 2012.
6. 大河原 一真, 淵田 正隆, 中村 明生, “賞味期限における文字検出手法の検討”, SICE/第 19 回パターン計測シンポジウム (第 94 回パターン計測部会研究会), pp.15-20, 2014.
7. 上田 雄貴, 淵田 正隆, 井上 淳, 中村 明生, “3D 点群データを用いた段差検出手法の検討”, ビジョン技術の実利用化ワークショップ (ViEW2015) 予稿集, pp.275-279, 2015.
8. 坂上 智紀, 淵田 正隆, 中村 明生, “視覚障がい者のためのエスカレータの位置および昇降方向検出手法の基礎検討”, 日本機械学会ロボティクス・メカトロニクス講演会 (ROBOMECH 2016) 講演論文集, pp.2A1-19a3(1)-(4), 2016.
9. 八柳 洸哉, 淵田 正隆, 中村 明生, “商品パッケージにおける特定文字領域検出手法の検討”, 第 22 回画像センシングシンポジウム (SSII 2016), pp. (1-5), 2016.
10. 加藤 周, 矢部 俊之, 淵田 正隆, 中村 明生, “屋内ピクトグラム検出・識別システムの基礎検討”, 第 34 回日本ロボット学会学術講演会予稿集, pp.(1-4), 2016.
11. 長野 樹, 藤井 浩光, 橘高 達也, 淵田 正隆, 深瀬 勇太郎, 青木 滋, 鳴海 智博, 山下 淳, 浅間 一, “アーム型屋外建機の遠隔操作のための測距センサを用いた手先映像提示システム”, 第 18 回 計測自動制御学会システムインテグレーション部門講演会 (SI 2017), pp.(1-5), 2017.
12. 岡安 寿繁, 吉田 光太, 八柳 洸哉, 淵田 正隆, 中村 明生, “白色背景下における蚊の種別識別に有効な特徴の調査”, “ビジョン技術実利用ワークショップ(ViEW 2017), pp.(1-4), 2017.
13. 大塚 智己, 淵田 正隆, 井上 淳, 中村 明生, “点字翻訳システムの開発—物体表面の点字領域抽出に関する基礎検討—”, “ビジョン技術実利用ワークショップ (ViEW 2017), pp.(1-4), 2017.
14. 矢部 俊之, 淵田 正隆, 井上 淳, 中村 明生, “視覚障がい者のための屋内誘導システム開発—ピクトグラム検出・識別手法の基礎検討—”, “ビジョン技術実利用ワークショップ(ViEW 2017), pp.(1-5), 2017.
15. 松崎 優太, 岡安 寿繁, 八柳 洸哉, 淵田 正隆, 中村 明生, 佐藤 雄隆, 片岡 裕雄, “FractalDB を用いた自然物認識のための特徴表現学習”, “ビジョン技術実利用ワークショップ(ViEW 2017), pp.(1-4), 2017.

#### E. 特許

1. “文字認識装置、文字認識方法およびプログラム”,  
発明者：中村 明生, 淵田 正隆  
特許出願人: 学校法人 東京電機大学  
代理人: 速水 進治 (プレシオ国際特許事務所)  
整理番号:TDU211 提出日: 2012.02.28 出願番号: 特願 2012-042085  
※東京電機大学 職務発明認定  
特許第 5964078 号 登録日: 2016.07.08

#### F. 学術研究報告書

中村 明生, 淵田 正隆, “視覚障がい者のための行動支援システム開発”, 東京電機大学総合研究所年報 No.33 2013 (東京電機大学 総合研究所 一般研究課題 課題番号: Q12J-06 / 2012-2013), pp.111-116, (2014).

#### G. 新聞やテレビ等による研究紹介

1. 新聞掲載: 2013.10.30(水) 日経産業新聞(朝刊) 6 面先端技術 「なぞった文字を音読 東京電機大が装置」
2. テレビ出演: 2013.11.05(火) テレビ東京・ワールドビジネスサテライト(WBS) 「トレンドたまご」コーナー 「文字を“読む”指輪」
3. 新聞掲載: 2013.11.08(金) 日経産業新聞(朝刊) 10 面先端技術 「目の前の段差, 音声で注意 東京電機大が装置」
4. 新聞掲載: 2016.12.13(火) 朝日新聞(朝刊) 18 面全面広告 「ロボット AI IoT and...!!! 理工学研究の最前線 (視覚障がい者を安全に目的地へ導く道案内システム・東京電機大学)」