

修士論文概要書

Summary of Master's Thesis

Date of submission: 02/01/2019 (MM/DD/YYYY)

専攻名 (専門分野) Department	情報理工・ 情報通信専攻	氏名 Name	小林 大起	指導 教員 Advisor	渡辺 裕 印 Seal
研究指導名 Research guidance	オーディオビジュアル 情報処理研究	学籍番 号 Student	CD 5117F034-5		
研究題目 Title	骨格推定と機械学習を用いたカナ指文字の分類 Classification of Japanese Signed Character with Pose Estimation and Machine Learning				

1. まえがき

近年、手話翻訳の自動化が国内外で盛んに研究されている。しかし、日本においては機械翻訳に必要なコーパスの存在数が目安である1000万対に対し15万対程度に留まる [1]。撮影が容易な単眼 RGB 画像ベースの手法も存在する。しかし、高精度なシステムは被験者が手に蛍光色の手袋を装着して ROI 抽出を容易化している場合がある。

本研究では、一般的な単眼 RGB カメラで事前準備を行わずに撮影されたカナ指文字の映像を、正しい文字ラベルへと分類することを目的とする。カナ指文字を対象とする理由は、手話習熟者にとっても誤読が生じやすい手話群であることによる [1]。また、カナ指文字の形状と動作をそれぞれ認識することは、全ての手話解析のモデルに適用できると考えられる。

2. 関連技術

Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields [2] (OpenPose) は、画像内の人物の2次元姿勢情報を効率的に推定する手法である。OpenPose は、従来研究には困難だった複数人物の身体各部位検出をボトムアップなアプローチで実現している。人物検出の後に部位検出を行う手法に比べて、人物非検出の場合や画像内の人数に比例する計算量増加に対してロバストである。さらに、既存のボトムアップ手法の欠点である部位間結合の計算コストを効率的な処理により改善している。これらの利点により、高速・高精度かつ遮蔽に対するロバスト性を持ちながら画像内の複数人物の姿勢を推定できる。

3. 提案手法

3.1. カナ指文字の大別

カナ指文字は手指の姿勢のみで表現できる文字と、手指姿勢を維持したまま特定動作を行うことで表現される文字の2種類に大別可能である。本研究では静止骨格を「静的指文字」、特定動作を「動的指文字」と定義し、これらの骨格情報を基に分類を行う。すなわち、静的指文字の骨格構造と動作指文字の動作方向とを別々に解析する。

3.2. 静的指文字の識別

OpenPose により入力映像 1 フレームにおける手指関節座標が取得される。式(3.2)により座標値を関節角度に変換する。インデックス $j_1 = (0,1,2,\dots,20)$ は図

3.1 左の手指内の各関節点に対応し、 a, b および c は図 3.1 右の 2 本の limb で繋がる 3 関節の組となっている。また、 $f = (1,2,\dots,F)$ は映像内におけるフレーム番号である。各点の関節角度を図 3.1 右の番号順に多重化した行ベクトル L を取得する。その後、静的指文字 1 文字分の映像全フレームに渡って L を行方向に多重化したベクトルを $m = \{m_{i,j_2}\}$ とする。 j_2 は図 3.1 右における添え字である。

$$P_f = \{p_{f,j_1}\} \quad (3.1)$$

$$p_{f,j_1} = (u_{f,j_1}, v_{f,j_1}), \quad (3.2)$$

$$\alpha = \text{Arccos} \left(\frac{(p_{f,a} - p_{f,b}) \cdot (p_{f,c} - p_{f,b})}{\|p_{f,a} - p_{f,b}\| \|p_{f,c} - p_{f,b}\|} \right)$$

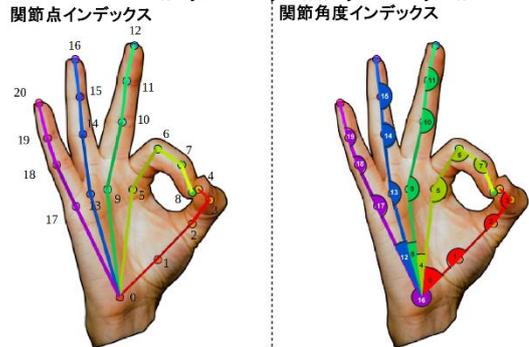


図 3.1 関節点および関節角度インデックス [3]

OpenPose による座標推定が失敗した場合、正しい関節角度が計算不能となる。そこで、 m 内の他フレームにおける対応関節情報から欠損値を補完する。

補完後の m を全文字に渡って行方向に多重化し、訓練集合 $D = \{x_{f,j_2}\}$ を構成する。

ただし、各関節の可動域に差があることで D の行方向の尺度は統一されていない。そこで、関節ごとの正規化を施す。最終的に得られた D_{norm} をパターン認識器により分類することで静的指文字を予測する。本研究では分類に SVM を用いる。

さらに、提案手法を拡張することで角度特徴量における類似骨格を分離する。手首の角度及び掌の裏表を特徴量とする SVM を新たに導入する。この SVM により、全静的指文字は 7 つのサブクラスに分類される。以上の条件下において提案手法を適用することで手指関節の姿勢が共通する指文字群を分離できる。

3.3. 動的指文字の識別

動的指文字は多くの文字で用いられる上下左右、右

斜め下および手前方向への直線的移動と「も」「ん」にのみ用いられる2種の固有動作から構成される。本研究では前者を「基本動作」、後者を「特殊動作」と定義する。特殊動作に関しては初期の手指形状を静的指文字として扱うことで個別処理が可能である。使用頻度の多い基本動作の識別に本節で述べる四つの特徴量構成を用いたパターン認識を適用する。

まず、動作における最初と最後のフレームから移動差分を取得する。上下左右の移動方向を[-1:1]の整数値で表し、加えて移動差分の tangent の絶対値を取得する。さらに、手指存在領域の面積変化率を算出する。したがって、動作検出に用いられる特徴量は 360° 方向に拡張した移動差分の tangent と奥行き情報の近似に相当する。

4. 評価実験

4.1. 静的指文字分類実験

静的指文字全 41 文字を撮影した 12 人分のデータセットに提案手法およびその拡張法をそれぞれ適用し、Leave-One-Out 交差検定で評価実験を行った。正解率の重み付き平均を表 4.1 に示す。また、各手法において精度が低い文字を抽出し、全文字の平均と併せて図 4.1 に結果を示す。なお、図 4.1 の指標は単純平均 F 値である。

表 4.1 および図 4.1 から、類似骨格を分離することで提案手法が全文字平均で 80-83%程度の精度を示したことが分かる。また、分離を行わない場合に比べて全ての文字で 10 ポイント以上の大幅な精度向上が確認できる。

表 4.1 提案手法の静的指文字平均分類精度

	平均テスト数	正解率 (W-AVG)
提案手法	7151.333	0.656
提案手法+拡張法	7151.333	0.825

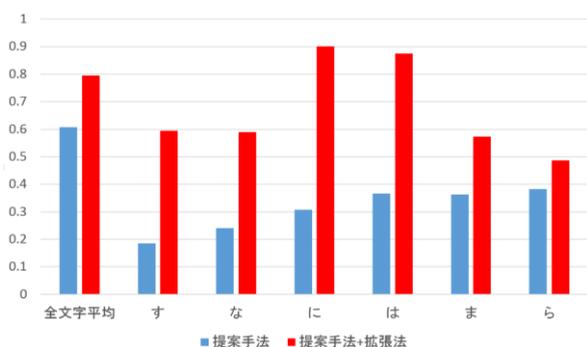


図 4.1 提案手法および拡張法の静的指文字別分類 F 値

4.2. 動的指文字分類実験

動的指文字基本動作 5 種類を各 50 回ずつ撮影した 7 人分のデータセットに 4.1 で用いた静的指文字データセットを加え、提案手法により分類を行った。実験方法は静的指文字と同様である。結果を表 4.2 ならびに図 4.2 に示す。表 4.2, 図 4.2 より、動的指文字に対する提案手法の平均正解率が 95%以上であり、かつ静

的指文字との分離も実現できることが分かる。

表 4.2 提案手法の動的指文字平均分類精度

	平均テスト数	正解率
AVG	50	0.960
Weighted-AVG		0.960

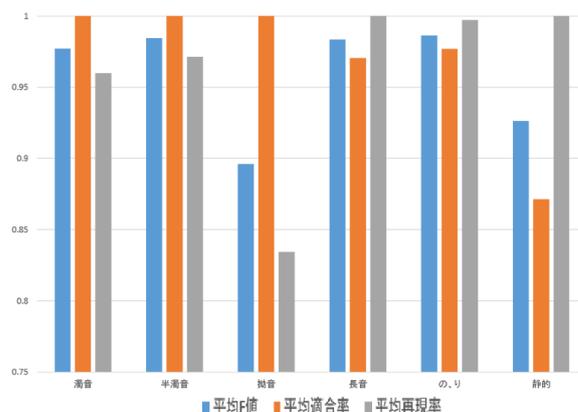


図 4.2 提案手法の動的指文字別分類精度

5. むすび

本研究は、カナ指文字を骨格部分と動作部分に大別することで静止画分類と単純な動作検出に分解できることを指摘した。静的指文字に対しては、OpenPose により得られた座標点から指文字の各関節角度を計算しベクトル化することで、SVM によって対応文字ラベルへ分類できることを示した。また動作検出に対しては、上下左右の移動方向と移動量の tangent 、手領域の変化率を特徴量とした SVM 分類法を提案した。

評価実験においては研究室で作成したデータセット全体に対し Leave-One-Out 交差検定を行った。静的指文字に関して、OpenPose で得られた座標値を関節角度に変換し、関節ごとの正規化および他フレームの情報による欠損値補完を施すことで平均 65%程度の予測精度を得た。さらに、提案手法に手首角度と掌の裏表に基づく拡張を行うことで平均精度が 15 ポイント以上向上することを確認した。動的指文字に関して、同様の実験を行い、提案手法が平均 96%程度の高精度で分類可能であることを確認した。

参考文献

- [1] 梅田 修一, NHK 技研, “NHK における手話 CG 研究について”, 第 5 回サイレント音声認識ワークショップ, 2018/09/29.
- [2] Z. Cao, T. Simon, S. Wei and Y. Sheikh, “Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields,” 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Papers(125), pp.1302-1310, July 2017.
- [3] CMU-Perceptual-Computing-Lab, “openpose,” [Online]. Available: https://github.com/CMU-Perceptual-Computing-Lab/openpose/blob/master/doc/media/keypoints_hand.png, [Access: Jan. 8th 2019].

2018年度 修士論文

骨格推定と機械学習を用いたカナ指文字の分類

**Classification of Japanese Signed Character with Pose
Estimation and Machine Learning**

指導教員 渡辺 裕 教授

早稲田大学大学院 基幹理工学研究科

情報理工・情報通信専攻

5117F034-5

小林大起

目次

第1章 序論	1
1.1 研究背景	1
1.2 研究目的	2
1.3 本論文の構成	2
第2章 関連技術	3
2.1 まえがき	3
2.2 SUPPORT VECTOR MACHINE	3
2.3 CONVOLUTIONAL NEURAL NETWORK	4
2.4 REALTIME MULTI-PERSON 2D POSE ESTIMATION USING PART AFFINITY FIELDS	5
2.5 むすび	5
第3章 提案手法	6
3.1 まえがき	6
3.2 指文字の大別	6
3.3 指文字の姿勢推定	8
3.4 静的指文字の検出	10
3.4.1 ラベル付け	10
3.4.2 関節角度計算	11
3.4.3 特徴ベクトル構築	12
3.4.4 欠損値補完	14
3.4.5 正規化	14
3.5 動的指文字の検出	14
3.5.1 ラベル付け	16
3.5.2 基本動作	17
3.5.3 特殊動作	19
3.6 むすび	21

第4章 実験	22
4.1 まえがき	22
4.2 予備実験	22
4.3 データセットの構築	24
4.3.1 静的指文字データセット	24
4.3.2 動的指文字データセット	26
4.4 評価実験1：静的指文字	30
4.4.1 特徴量の有効性確認	30
4.4.2 データセット全体への適用	31
4.4.3 スケールに対するロバスト性確認	33
4.5 評価実験2：動的指文字	35
4.6 考察	37
4.6.1 静的指文字分類実験に関する考察	37
4.6.2 動的指文字分類実験に関する考察	39
4.7 むすび	40
第5章 提案手法拡張	41
5.1 まえがき	41
5.2 多段SVMによる二段階分類	41
5.3 静的指文字のサブクラス化	42
5.4 サブクラス分類	43
5.4.1 手首回転量の特定	43
5.4.2 手の裏表判定	44
5.5 サブクラス内静的指文字分類	45
5.6 むすび	45
第6章 追加実験	46
6.1 まえがき	46
6.2 サブクラス分類実験	46
6.3 静的指文字のサブクラス内分類実験	47
6.4 考察	53
6.4.1 サブクラス分類精度	53
6.4.2 サブクラス内での静的指文字分類精度	55
6.4.3 処理全体での分類精度期待値	56
6.5 むすび	56
第7章 結論	57

謝辭.....	58
參考文獻.....	59
圖一覽.....	61
表一覽.....	63
研究業績.....	64

第1章

序論

1.1 研究背景

日本で普及している手話は主に三つ存在する。日本手話、日本語対应手話および両者の中間手話である。国内にはおよそ 30 万人の聴覚障害者が存在し [1]、これら三つの手話を使い分けている。各手話は必ずしも個々に独立して使用されるわけではない。日本語の語彙にしか存在しない単語を学習する際は、これらを適切に組み合わせる必要がある。しかし、いずれの手話も日本語との通訳者が少ない。そのため、日本における全ての手話の自動認識が求められている。

手話の自動翻訳例として、2018 年の平昌オリンピックでアイスホッケーとカーリングを対象とした試合展開のリアルタイム手話翻訳がある。この自動翻訳は NHK 技研によって試みられた [1-2]。ただし、リアルタイムでの翻訳精度が低く、生成された手話文が不自然である。また、NTT データグループとシャープはロボホン“SOTA”向けの手話翻訳アプリケーションを開発中である [3]。国外においてはハンガリーの“SignAll” [4]があり、複数カメラとデプスセンサを併用する。RGB 画像ベースの手法としてはアルゼンチン手話のデータセットである LSA64 [5] を使用した研究 [6]が挙げられる。

しかし、このような手法はいずれも距離センサなどの特殊な環境が要求される。また、単眼 RGB 画像を用いる手法も、被験者が事前に蛍光色の手袋を装着する必要がある。したがって、いずれの手法も利用上の制約が大きいという欠点がある。

言語間の機械翻訳には両者のコーパスが不可欠である。ディープラーニングの場合は実用レベルに達するまでにおよそ 1000 万対が必要とされている [1]。コーパス数が少ない日本では、一般的な環境下に導入できる手法が求められる。すなわち、新たに作成する訓練サンプルにも特殊な条件を設けず、既存映像をそのまま訓練に使用できるような学習手法が適していると考えられる。

1.2 研究目的

本論文では、深度情報やステレオカメラではなく、一般的な単眼 RGB カメラ画像で撮影された日本語対应手話の一種であるカナ指文字の映像を、正しい文字ラベルへと分類することを目的とする。カナ指文字を対象とする理由は、手話習熟者にとっても誤読が生じやすい手話群であり [1]、かつ複雑な手指形状と一部動きの組み合わせという構成が全ての手話解析のモデルケースとして活用できると考えるためである。

1.3 本論文の構成

本論文を構成する全 5 章について、その概要を以下に示す。

第 1 章「序論」は本章であり、手話のあらましとその機械翻訳における課題、およびこれら二つの背景を受けて設定した本研究の目的を述べている。

第 2 章「関連技術」では、本研究で提案する手法の中核に用いるパターン認識法と姿勢推定技術について述べている。

第 3 章「提案手法」では、本研究の目的であるカナ指文字の分類について、姿勢推定技術を用いたアプローチを提案している。

第 4 章「実験」では、第 3 章で提案した指文字の分類法について、予備実験の結果得られたデータセットの説明と、それに対する評価実験の内容と結果を述べている。

第 5 章「提案手法拡張」では、第 4 章の実験結果を受けて提案手法の改善を目的とした拡張を再提案している。

第 6 章「追加実験」では、第 5 章で提案した拡張法の精度確認結果を述べている。

第 7 章「結論」では、本論文のまとめを述べている。

第2章

関連技術

2.1 まえがき

本章では本研究に用いる機械学習法および姿勢推定手法について説明する．まず指文字の分類に使用するパターン認識モデルの一種である **Support Vector Machine** [7]について述べる．次に画像認識分野で多用される深層学習の一つである **Convolutional Neural Network** [8]について概説する．その後，本研究で姿勢推定に用いる **Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields** [9]について述べる．

2.2 Support Vector Machine

サポートベクタマシン (SVM) [7,10]は，教師付き学習によってパターン認識を行う手法の一つであり，音声や画像・映像などの広い分野で用いられ，単語分類においても汎化性能に優れる研究例がある [11]．SVM では入力されたデータから特徴的な箇所を抽出し，ベクトル化したものを対応するクラスと併せて学習させる．この処理は数学的には識別データ群を分割する超平面と最近傍に存在するデータとの間の距離を最大化させることにより実現される．図 2.1 に分離模式図を示す．

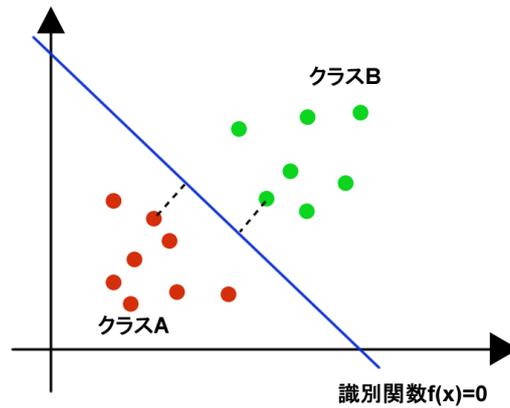


図 2.1 SVM によるサンプル群分離模式図

2.3 Convolutional Neural Network

Convolutional Neural Network (CNN) は、画像処理において多用される深層学習の一種である [8]。畳み込み層とプーリング層を 1 セットとし、これを図 2.2 に示すように複数回連続させたネットワーク構造を持つ [12]。入力画像は、フィルタを用いた畳み込み演算が行われる畳み込み層と、得られた出力のダウンサンプリングを行うプーリング層を用いて繰り返し処理される。これにより、入力画像の特徴量が抽出される。抽出された特徴量は全結合層において次元削減され、出力層となる。

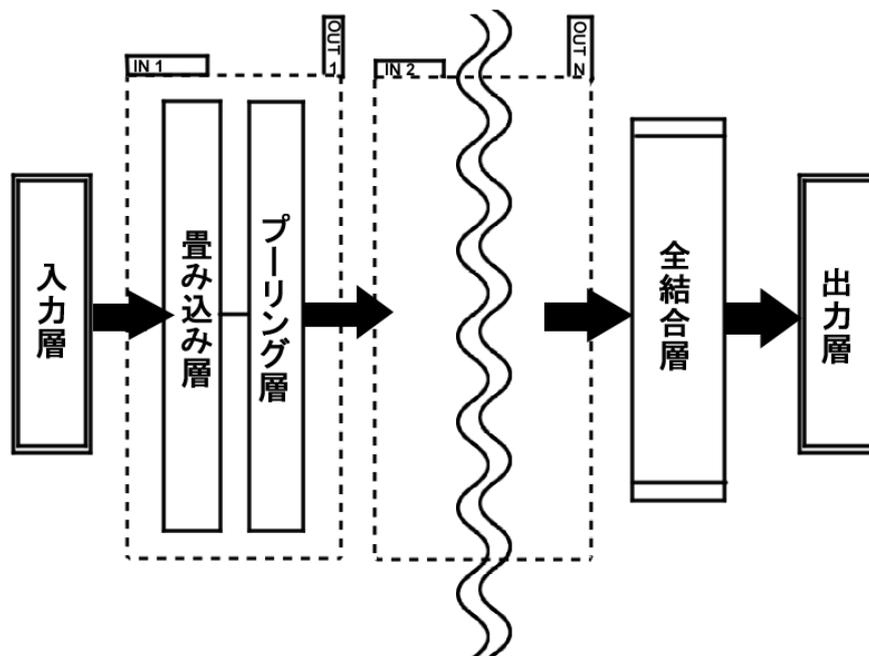


図 2.2 CNN の概要図

2.4 Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields

Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields (OpenPose) は、画像内の人物の 2 次元姿勢情報を効率的かつ高精度に推定する手法である。基本的な構成は、CNN を用いた関節座標検出手法である Convolutional Pose Machines (CPMs) [13] の検出点間における拡張である。ネットワークの概要図を図 2.3 に示す。

OpenPose ではボトムアップなアプローチにより複数人物の関節検出を実現している。人物検出の後に部位検出を行う従来手法に比べて計算量増加に対してロバストという特徴を持つ。さらに、既存のボトムアップ手法における欠点である位間結合の計算コストを効率的な処理により改善している。

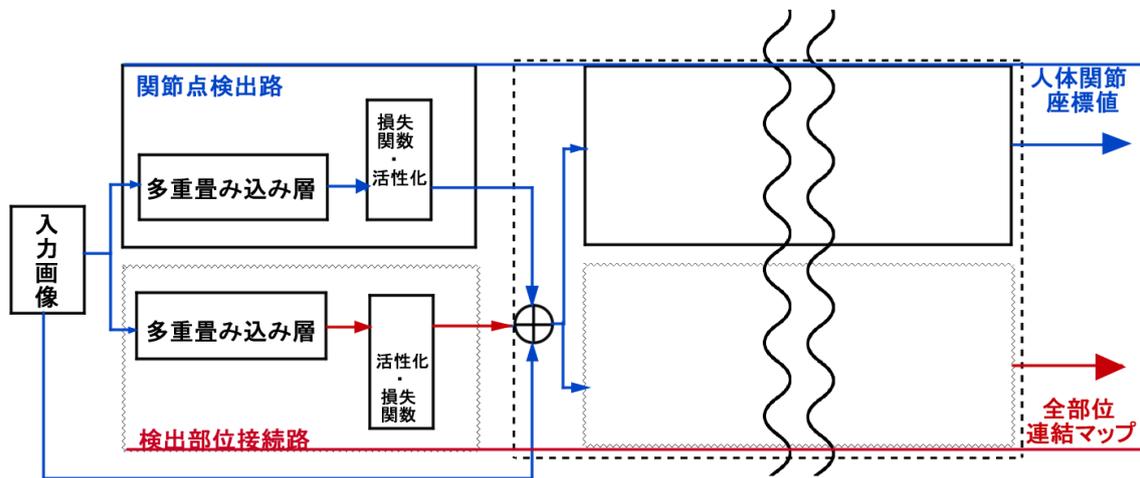


図 2.3 OpenPose のネットワーク構成 [9]

2.5 むすび

本章では、本研究で提案する指文字の分類を実行するための、パターン認識と姿勢推定手法について述べ、中核となる SVM と OpenPose を紹介した。次章では、本章で説明した手法を指文字の骨格推定に応用して分類を行うアプローチについて述べる。

第3章

提案手法

3.1 まえがき

本章ではカナ指文字の翻訳について、指先の姿勢情報に基づいて対応する文字へ分類する手法を提案する。本章の構成は次の通りである。まず、カナ指文字の特性から全体を2種類に大別する方法を述べる。次に、指先の姿勢推定について説明する。さらに、得られた姿勢情報から手指の角度情報を算出し、その値を特徴量として分類するアプローチを提案する。

3.2 指文字の大別

図 3.1 はカナ指文字の一覧である。文字群の特性からこれらが大別する。

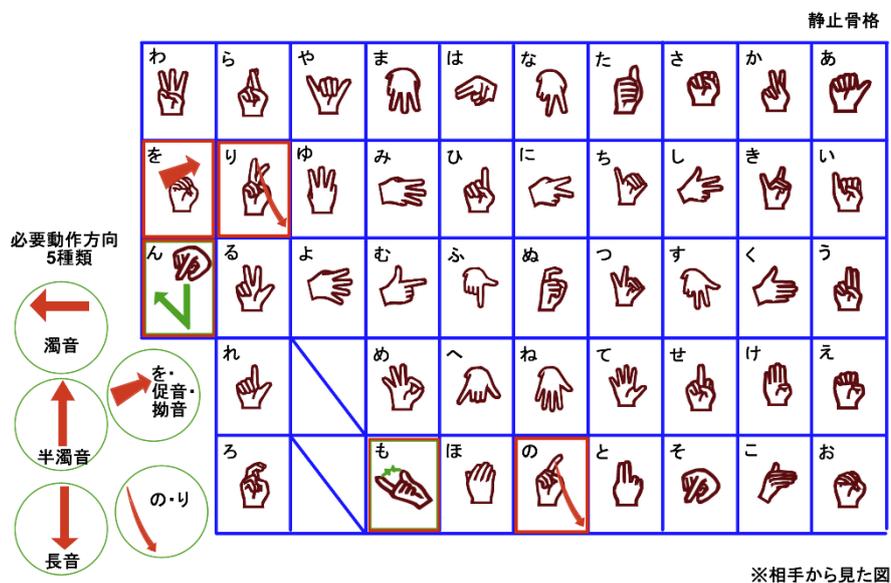


図 3.1 カナ指文字の一覧

図 3.1 より、カナ指文字は手指の姿勢のみで表現できる文字と、特定の動作を伴って表現される文字の 2 種類に大別可能であることが分かる。また、動作は「も」と「ん」を除き、他の文字を表す骨格を特定方向に動かすことで成立する。

したがって、本研究では図 3.1 中の静止骨格を「静的指文字」、特定動作を「動的指文字」と定義し、これらの骨格情報を基に適切な分類を行うことを目指す。すなわち、図 3.2 の概要図に示すように、静的指文字の骨格構造と動作指文字の動作方向とを別々に解析することで、カナ指文字の分類問題を静止画分類と単純な動作分類に分割する。

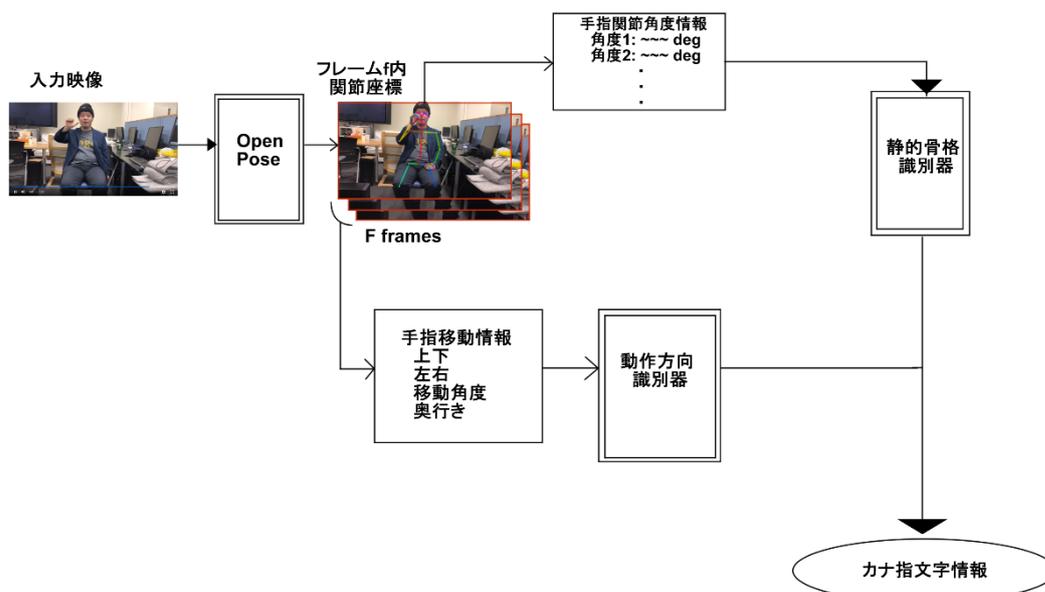


図 3.2 提案手法概要図

3.3 指文字の姿勢推定

まず、指文字の骨格推定を OpenPose [9]によって行う。一般に RGB 値を用いる画像処理では背景と Region of Interesting の分離に大きなコストを要する。本研究では画像 RGB 値を直接使用せず、OpenPose による検出姿勢情報に変換することで必要学習量の削減と背景へのロバスト性を見込む。OpenPose に使用されている手指モデルでは、図 3.3 のように各関節が番号付けされている。手指画像を OpenPose に入力すると、21 個の関節の座標値が手指の推定結果として JSON 形式で出力される。リスト 3.1 は出力結果の例であり、関節点ごとに x 座標、y 座標および confidence 値の三つを保持した辞書構造である。

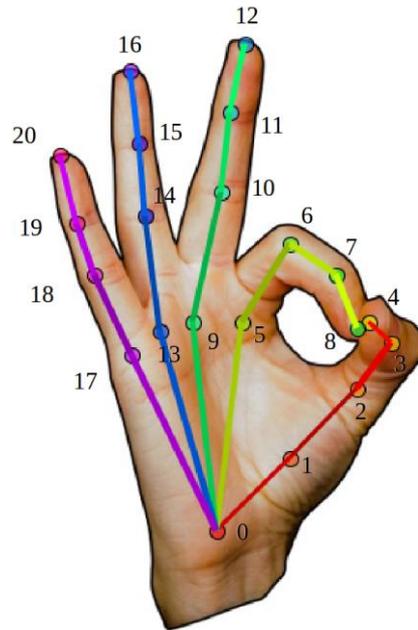


図 3.3 OpenPose における手指モデルの関節番号 [14]

リスト 3.1 OpenPose の JSON 出力例

```
{
  "version":1.2,
  "people":[
    {
      "pose_keypoints_2d":[924.652,150.649,0.91224,...],
      "hand_right_keypoints_2d":[804.823,249.795,0.549834,...],
      "pose_keypoints_3d":[],
      "face_keypoints_3d":[],
      "hand_left_keypoints_3d":[],
      "hand_right_keypoints_3d":[]
    }
  ]
}
```

各種指文字の映像を 1 フレームずつ OpenPose に入力することで、そのフレームにおける姿勢情報 \mathbf{P}_f を取得する。ここで、 $f = (1, 2, \dots, F)$ はフレーム番号を表し、 F は入力映像の最終フレーム番号である。 $\mathbf{p}_{f,i}$ はそのフレームにおける各関節の 2 次元座標である。関節インデックス $i = (0, 1, 2, \dots, 20)$ は図 3.3 に対応する。

$$P_f = \{p_{f,i}\} \quad (3.1)$$

$$p_{f,i} = (u_{f,i}, v_{f,i})$$

3.4 静的指文字の検出

カナ指文字は複雑な手指骨格に一部単純な動作を組み合わせた姿勢群と見なせる。したがって、正しい分類結果を得るためには静的指文字の解析が重要となる。本節では角度情報を特徴量として使用し SVM による分類を行う手法を示す。一連の処理を図 3.4 に示す。

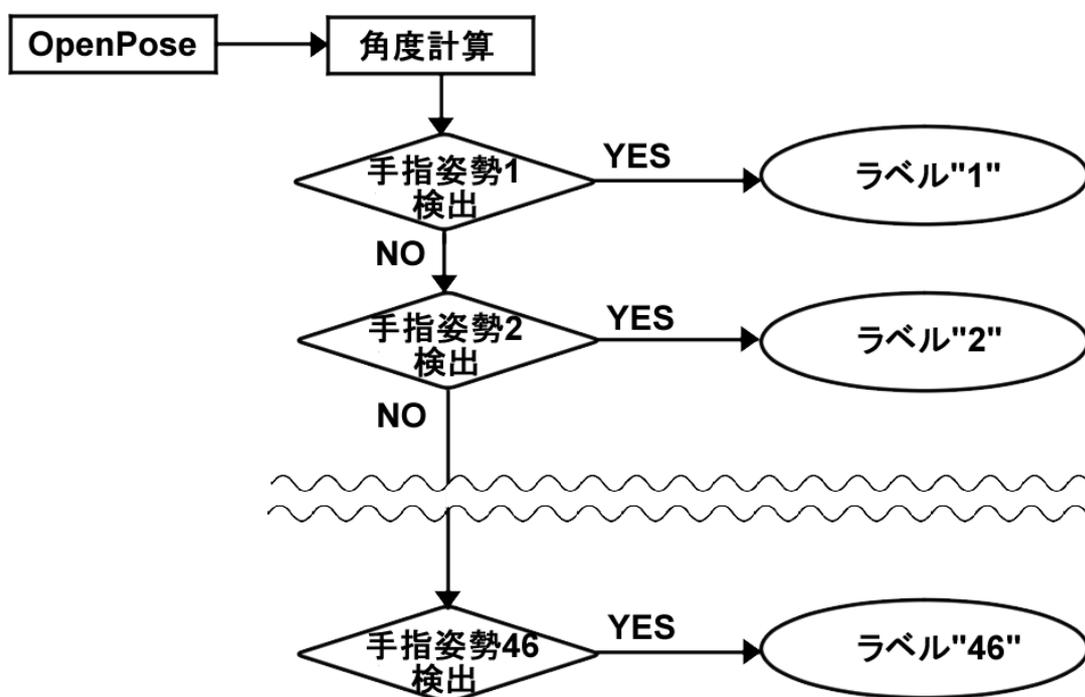


図 3.4 静的指文字の姿勢検出フロー

3.4.1 ラベル付け

表 3.1 のように各静的指文字をラベル付けする。オレンジ箇所は動的指文字，塗りつぶし箇所は可読性のための欠番である。

表 3.1 カナ指文字のラベル対応

1	2	3	4	5
---	---	---	---	---

あ	い	う	え	お
6	7	8	9	10
か	き	く	け	こ
11	12	13	14	15
さ	し	す	せ	そ
16	17	18	19	20
た	ち	つ	て	と
21	22	23	24	25
な	に	ぬ	ね	の
26	27	28	29	30
は	ひ	ふ	へ	ほ
31	32	33	34	35
ま	み	む	め	も
36	37	38	39	40
や	ゆ	よ		
41	42	43	44	45
ら	り	る	れ	ろ
46	47	48	49	50
わ	を	ん		

3.4.2 関節角度計算

手の大きさには個人差があり、また特に規定しない場合は撮影環境におけるカメラ位置は一定ではない。また、OpenPoseは2次元座標を出力する。したがって、座標値を直接用いるアプローチでは、サンプルごとの差異を吸収するためにスケール変換を必要とする。先行研究 [15]のように特定リムの長さを基準とし、全リムの長さをスケール変換する。リムとは図 3.5 における破線箇所であり、人体における関節間を繋ぐ骨部分に相当する。

しかし、カナ指文字の場合、手の向きが様々に変化し、画面内に全文字に渡って統一的に写るリムを特定することが難しい。また、OpenPoseでは出力座標に若干のブレが生じることがあり、全身推定に比べて範囲が狭い手指推定においてはその影響度が大きくなることが予想される。したがって、本研究では静的指文字に対してスケール差を吸収するために座標値ではなく各リムの関節角度を利用する。

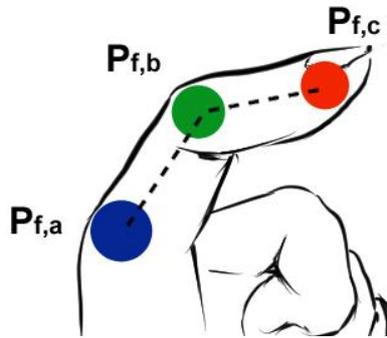


図 3.5 関節角度計算

図 3.5 のように関節点 $\mathbf{p}_{f,a}$, $\mathbf{p}_{f,b}$ および $\mathbf{p}_{f,c}$ が与えられたとき, 二つのリムが成す角度 α は

$$\alpha = \text{Arccos} \frac{(\mathbf{p}_{f,a} - \mathbf{p}_{f,b}) \cdot (\mathbf{p}_{f,c} - \mathbf{p}_{f,b})}{\|\mathbf{p}_{f,a} - \mathbf{p}_{f,b}\| \|\mathbf{p}_{f,c} - \mathbf{p}_{f,b}\|} \quad (3.2)$$

となる.

3.4.3 特徴ベクトル構築

\mathbf{P}_f の全ての要素を $\mathbf{p}_{f,b}$ として式(3.2)に入力し, 表 3.1 のラベル \mathbf{y} を付与することで次のベクトル \mathbf{A}_f を得る.

$$\mathbf{A}_f = [\mathbf{y}, \boldsymbol{\alpha}_f] \quad (3.3)$$

$$\boldsymbol{\alpha}_f = [\alpha_{f,0}, \dots, \alpha_{f,19}]$$

なお, インデックス $j = (0, 1, 2, \dots, 19)$ に対応する関節角度は図 3.6 に準ずる.

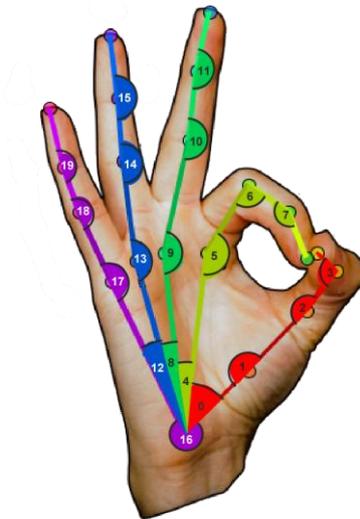


図 3.6 各関節のインデックス

入力映像を構成する全フレームについて、式(3.3)を1行とした行列を作成することで静的指文字 c における学習データ D_c とする。なお、 c は表 3.1 の文字ラベルに対応する。

$$D_c = \begin{bmatrix} A_1 \\ \vdots \\ A_F \end{bmatrix} \quad (3.4)$$

式(3.4)と同様に D_c を c について多重化したものを訓練集合 D とする。構造を図 3.7 に示す。

$$D = \begin{bmatrix} D_{\text{あ}} \\ \vdots \\ D_{\text{わ}} \end{bmatrix} \quad (3.5)$$

ラベル	角度0	角度1	● ● ●	角度19
1	35.45	161.25		9.88
1	34.61	159.22		8.44
2	34.99	91.24		178.21
● ● ●				
46	33.26	95.44		145.11

図 3.7 訓練集合の構造 (静的指文字)

3.4.4 欠損値補完

OpenPose により得られる出力は、図 3.8 に示すように関節検出および姿勢推定を失敗する場合がある。 D_c のラベルに相当する 1 列目を無視し、その他の各要素を $x_{f,j}$ とすると、 j は図 3.6 の関節番号である。学習を正しく行うために関節 j の欠損値を式(3.6)により補完する。その後、式(3.5)に D_c を入力する。

$$Lack_j = \frac{1}{F} \sum_{f=1}^F x_{f,j} \quad (3.6)$$



図 3.8 姿勢推定失敗例

3.4.5 正規化

これまで述べてきた計算で得られる学習データは列ごとに尺度が一定ではない。SVM による学習を最適化するために、以下の式(3.7)により学習データを正規化する。

$$x_{f,j}^{norm} = \frac{x_{f,j} - \min_f x_{f,j}}{\max_f x_{f,j} - \min_f x_{f,j}} \quad (3.7)$$

3.5 動的指文字の検出

動的指文字に関しては静的指文字の姿勢を維持したまま、次の 7 種類の動作を行うことにより実現できる。

- a) 右方向に移動させる (濁音)
- b) 上方向に移動させる (半濁音)
- c) 手前方向に移動させる (拗音, 「を」, 「っ」)
- d) 下方向に移動させる (長音)
- e) 左下方向に移動させる (「の」, 「り」)
- f) 下方向に移動させた後右上方向に移動させる (「ん」)
- g) 親指と人差し指でつまむ (「も」)

本研究では a)~e)を「基本動作」、f)および g)を「特殊動作」と定義し、5+2 クラスに分類する。静的指文字の検出により基本骨格は表 3.1 のラベル番号として出力されるため、その値から「基本動作」と「特殊動作」を判別する。その後、それぞれの動作を特定することで動的指文字を分類できる。ここで、関節座標を直接用いるのではなく、フレーム間における大小関係および移動量の比のみを用いることでスケールに対するロバスト性を保つ。

動的指文字処理の流れを図 3.9 に示す。

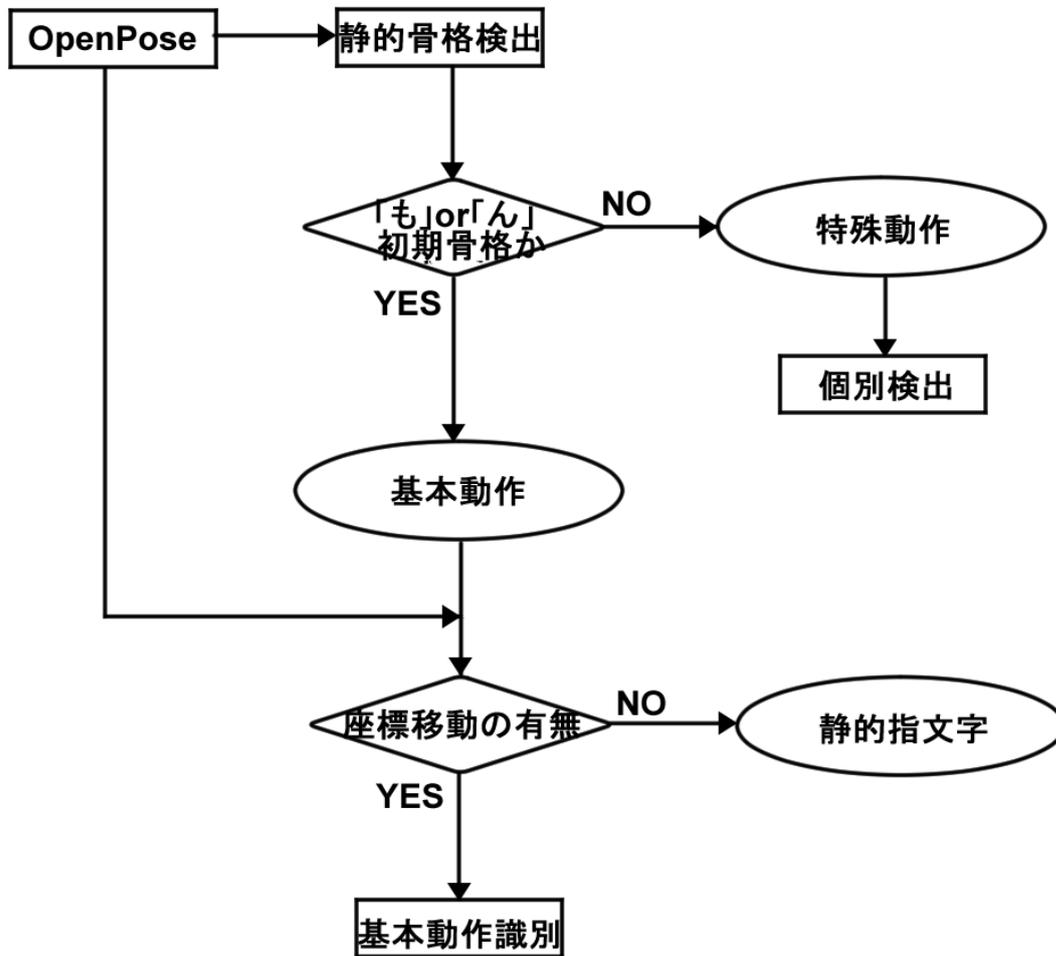


図 3.9 動的指文字の動作検出および識別フロー

3.5.1 ラベル付け

動的指文字に対応する各動作を表 3.2 のようにラベル付けする。ただし、基本動作と特殊動作では識別法が異なるためそれぞれのラベルは区別して扱う。

表 3.2 動的文字の区分と対応ラベル

動作区分	ラベル	対応文字
基本動作	1	濁音
	2	半濁音
	3	拗音・促音・を

	4	長音
	5	の, り
特殊動作	1	ん
	2	も

3.5.2 基本動作

動的指文字を撮影した映像のうち、最初と最後のフレームを OpenPose に入力することで \mathbf{P}_1 および \mathbf{P}_F を取得する。それぞれのフレームにおける手の位置 \mathbf{p}_f^{ave} は以下で定義する。

$$\mathbf{p}_f^{ave} = \frac{1}{5}(\mathbf{p}_{f,2} + \mathbf{p}_{f,5} + \mathbf{p}_{f,9} + \mathbf{p}_{f,13} + \mathbf{p}_{f,17}) \quad (3.8)$$

また、それぞれのフレームにおける手領域の面積を次の式(3.9)で定義する。

$$s_f = |(\max u_f - \min u_f)(\max v_f - \min v_f)| \quad (3.9)$$

以上を利用し特徴ベクトル \mathbf{M} を構築する。

$$\mathbf{M} = (m_1(\mathbf{p}_1^{ave}, \mathbf{p}_F^{ave}), m_2(\mathbf{p}_1^{ave}, \mathbf{p}_F^{ave}), m_3(\mathbf{p}_1^{ave}, \mathbf{p}_F^{ave}), m_4(\mathbf{p}_1^{ave}, \mathbf{p}_1^{ave})) \quad (3.10)$$

各要素は式(3.10)~(3.14)により計算される。先頭から順に移動方向の上下、左右、上下と左右の移動量の比、手領域の面積変化率である。すなわち、360° 全方向に拡張した移動差分の tangent と奥行き情報に等しい。

$$m_1(\mathbf{p}_1^{ave}, \mathbf{p}_F^{ave}) = \begin{cases} 0, & |u_1 - u_F| \leq \sqrt{s_1} \\ 1, & u_1 < u_F \\ -1, & u_1 > u_F \end{cases} \quad (3.11)$$

$$m_2(\mathbf{p}_1^{ave}, \mathbf{p}_F^{ave}) = \begin{cases} 0, & |v_1 - v_F| \leq \sqrt{s_1} \\ 1, & v_1 \\ -1, & v_1 > v_F \end{cases} \quad (3.12)$$

$$m_3(\mathbf{p}_1^{ave}, \mathbf{p}_F^{ave}) = \begin{cases} 1000, & |v_1 - v_F| \leq \sqrt{s_1} \text{ and } |u_1 - u_F| \leq \sqrt{s_1} \\ 100, & \text{else if } |v_1 - v_F| \leq \sqrt{s_1} \\ \left| \frac{v_1 - v_F}{u_1 - u_F} \right|, & \text{else } v_1 > v_F \end{cases} \quad (3.13)$$

$$m_4(\mathbf{p}_1^{ave}, \mathbf{p}_1^{ave}) = \frac{s_1}{s_F} \quad (3.14)$$

ここで、 $\sqrt{s_1}$ は被験者および撮影者の生理的な手のブレなどにより生じ得る微細な移動を無視するために設定する閾値である。本研究では図 3.10 のように最初のフレームにおける手の存在領域面積を正方形に等積変形し、正方形の 1 辺の値以上の距離を手が移動した場合に意図的な移動と見なす。

また、手の平均位置の移動ベクトルの **tangent** を特徴量として使用しているが、分母が 0 の場合はその値が無限大に発散する。したがって、そのままでは式(3.7)による正規化を適用できない。式(3.13)における数値 1000 及び 100 はこの問題を解決するために導入した。上下方向の移動が 0 の場合と、上下方向のみの移動を分離するために設定した閾値である。

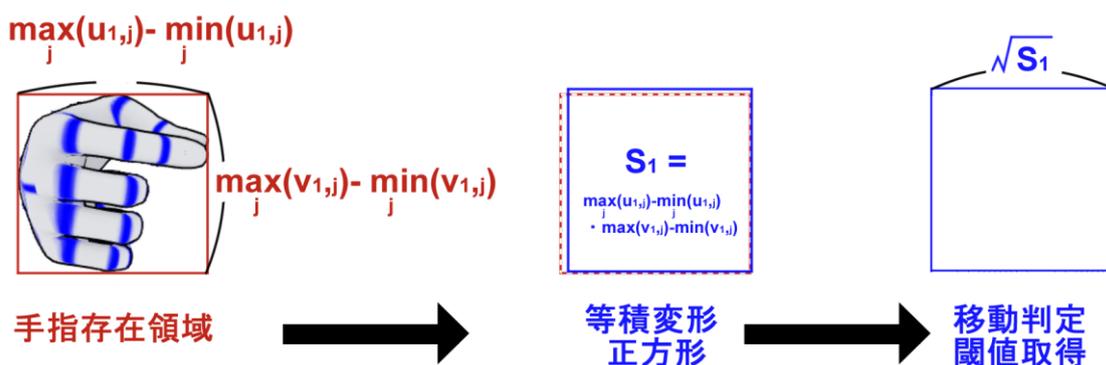


図 3.10 移動判定閾値取得フロー

M を SVM により分類することで基本動作の種類を特定する。また、静的指文字を基本動作ラベル 6 として追加することで動的指文字と静的指文字の区別も同時に可能である。

構成される訓練集合の構造を図 3.11 に示す。

ラベル	左右判定	上下判定	移動量の tan	手領域の面積変化
1	1	0	0.04993	1.04993
● ● ●	● ● ●			
6	0	0	1000	1.033011

図 3.11 訓練集合の構造 (動的指文字基本動作)

3.5.3 特殊動作

特殊動作 2 文字に関しては静的指文字のラベル番号から条件分岐し、それぞれに特有の動作を検出する。

3.5.3.1 指文字「ん」の識別

動作の各フレームを OpenPose に入力する。フレーム f と $f-1$ 間の移動差分を

$$\mathbf{d}_k = \mathbf{p}_f^{ave} - \mathbf{p}_{f-1}^{ave} \quad (3.15)$$

により求める。指文字「ん」は一連の動作において手を動かす向きが 1 回変化する。動作中の直線移動時と向きの変化時、移動差分には図 3.12 に示す特性が考えられる。

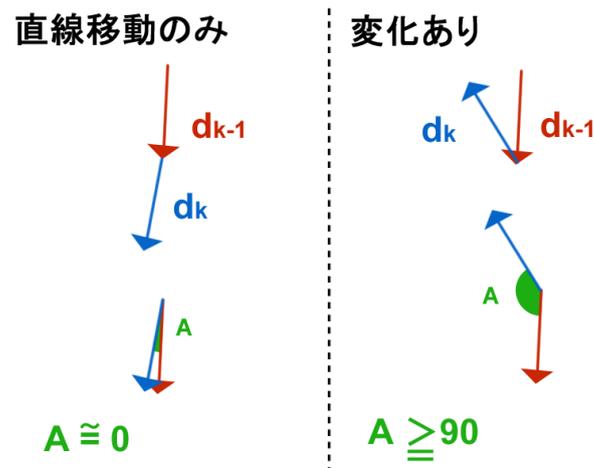


図 3.12 移動方向変化に伴う移動差分の角度特性

したがって、式(3.2)と同様に d_{k-1} と d_k の成す角度 A を計算し、

$$A = \text{Arccos} \frac{d_{k-1} \cdot d_k}{\|d_{k-1}\| \|d_k\|} \geq 90 \quad (3.16)$$

であるとき、変化点が存在する。指文字「ん」の判定においては変化回数が1回となる。

ただし、被験者の手を動かす速度は動作の始まりや終わり付近、および変化点の付近においては小さくなることが予想される。また、直線移動の場合でも、動きを止めた直後は反動が生じ得る。これらの誤差を除去するため、差分ベクトルには以下の式(3.17)の条件を追加する。定数部はいずれも「ん」の動作に干渉せず、かつ誤差と意図的な移動を分離できるだけの大きさを持つ任意の閾値である。

$$\begin{aligned} A &\leq 170 \\ |d_f| &> 10 \end{aligned} \quad (3.17)$$

3.5.3.2 指文字「も」の識別

図 3.13 に示すように指文字「も」の動作は人差し指と親指を摘む動作であり、それ以外の指は固定されている。したがって、親指先端と人差し指先端の間の距離縮小を検出すれば特定可能である。その判定を式(3.18)によって行う。なお、 τ は適当な閾値であり、 $\tau < 1$ である。

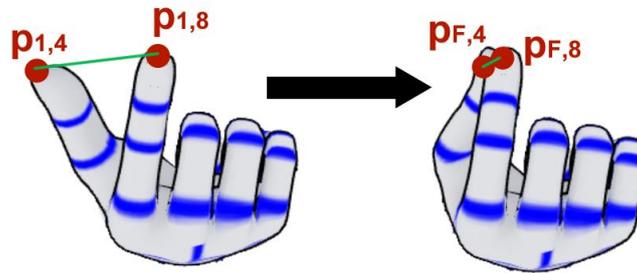


図 3.13 親指先端と人差し指先端接近による動作検知

$$\frac{\|p_{F,4} - p_{F,8}\|}{\|p_{1,4} - p_{1,8}\|} \leq \tau \quad (3.18)$$

3.6 むすび

本章では，カナ指文字の翻訳について静的と動的の 2 種類に大別した後に OpenPose の出力から関節角度を算出し，適切な文字ラベルへ分類する手法を提案した．次章では本章で述べたアプローチをデータセットに適用し，その精度を評価する．

第4章

実験

4.1 まえがき

本章では, OpenPose による手指検出の特性評価を目的として行った予備実験の結果を述べる. また第 3 章において提案した手法の評価実験とその結果を述べる. なお, 本章における実験は全て表 4.1 の環境で行った.

表 4.1 実験環境

OS	Ubuntu 16.04 LTS
RAM	32GB
CPU	Intel® Xeon® X3450 @2.67GHz
GPU	NVIDIA GeForce GTX 1070 Ti
OpenPose	Ver. 1.3.0

4.2 予備実験

第 2 章で述べたように, OpenPose では画像内の全人物の関節点を同時に検出し, それらの全ての組み合わせを骨格の候補とする. したがって, 入力画像内に存在する関節の個数が骨格検出の精度に直接的な影響を及ぼす. 撮影部位の差異に対する OpenPose の手指検出特性を評価するため, 本節で述べる実験を行った.

まず, インターネット上で収集した人体画像 50 枚を画像内に写っている部位ごとに表 4.2 のように分類した.

表 4.2 データ区分

分類番号	撮影部位	枚数
1	手のみ	10
2	肩から上	10
3	上半身	10
4	全身	10

表 4.2 の画像群を OpenPose に入力し、画像内の人物の姿勢推定を行う。出力は入力画像に推定されたポーズを表すスケルトンが重畳された画像として得られる。

その後、入力画像内に存在する関節点の位置を目視により確認し、ground truth とする。出力画像において、ground truth と同一箇所 OpenPose が関節点を出力していた場合、検出成功と判定する。本実験では検出成功箇所と ground truth の個数比を用いることで特性評価とした。各分類における全画像内の手指 ground truth 数平均値に最も近い画像の推定結果を表 4.3 に示す。

表 4.3 OpenPose による手指検出特性評価

分類番号	身体部位		手指部位		正解率	
	ground truth	検出成功	ground truth	検出成功	身体	手指
1	0		21	0		0
2	44	19	37	0	0.432	0
3	60	58	105	84	0.967	0.8
4	25	25	42	42	1	1

現行バージョンにおける OpenPose では身体部位に加えて手指、表情の骨格推定が可能ではあるが、その全ての関節部位検出を同時には行えない [14]。身体各部位を推定した後、両者の存在領域を検出しそれぞれの部位検出が実行される。すなわち、手指検出において身体部位情報が領域検出器として扱われる。

表 4.3 より、手指のみを撮影した画像は OpenPose による検出が不可能であるが、この結果は上記の仕様から自明である。

一方、理論上は手首および肩が入力画像に存在すればリムを構築し手指検出が可能であるが、対応する分類 2 の手指検出精度は非常に低い。分類 2 においては画像内に存在する関節点の数が不足しているため、組み合わせにより得られるリム候補の比較が不十分である。そのため、誤った箇所に肩関節を検出するか、あるいは肩が画像に写っていないと判定された結果が多いことが、検出精度の低下に影響していると考えられる。

実際に、画像内に複数人が写っている分類 3 ならびに 4 では画像内の関節点候補が劇的に増加しているにもかかわらず身体、手指ともに 80%以上の検出精度を示した。この結果は、一人当たりのリム候補数が各関節位置の推定精度を向上させる重要な要素であることを裏付けている。

4.3 データセットの構築

予備実験の結果および考察に基づき、静的指文字のデータセットを構築した。その際、条件として次の四つを設けた。

- 被験者の腰より上部が完全に収まること
- カメラ位置は被験者の正面付近に設定すること
- 解像度は1920×1080[pixel]とすること
- フレームレートを 30[FPS]に設定すること

ただし、本研究における提案手法では一般的な環境下で撮影された指文字の映像を学習データとして利用可能にすることを目的とするため、被験者の撮影距離、および角度は厳密には定めない。また、撮影カメラは撮影者が手で保持し、固定しない。

4.3.1 静的指文字データセット

4.3 節冒頭の設定に基づき、各指文字を 5 秒程度撮影した映像を 12 人分集めたものを評価実験におけるデータセットとして用いる。表 4.4 はその内訳である。また、図 4.1 は作成したデータセットの一例である。

表 4.4 各静的指文字のデータ数

文字ラベル	合計フレーム数
1	1833
2	1973
3	2307
4	1971
5	1980
6	1962
7	2010
8	2278
9	1949
10	2112
11	2094

12	2064
13	2219
14	2292
15	1910
16	2065
17	1923
18	2173
19	2127
20	2189
21	2059
22	2173
23	2150
24	2067
26	1942
27	2139
28	2162
29	2001
30	2150
31	2046
32	2320
33	2073
34	2168
36	2160
37	1993
38	2108
41	2079
43	2010
44	2193
45	2449
46	2100



図 4.1 静的指文字データセットの一例

4.3.2 動的指文字データセット

撮影条件等の設定は静的指文字に準ずる。しかし、動的指文字は動作 1 回につき 1 文字を表すため、静的指文字と異なり映像の各フレームを切り出してサンプルに使用することができない。そこで、新たに被験者 7 名を選び動的指文字のデータセットを作成した。各動的指文字に対応する全 7 動作を一つにつき 50 回撮影し、それらを 7 人分集めたものを使用する。表 4.5 にサンプル数を、図 4.2 に基本動作、図 4.3 に特殊動作の内容の一例を示す。

表 4.5 各動的指文字のデータ数

	動作ラベル	サンプル数
基本動作	1	50
	2	50
	3	50
	4	50
	5	50
特殊動作	1	50
	2	50



図 4.2 動的指文字データセットの一例（基本動作）

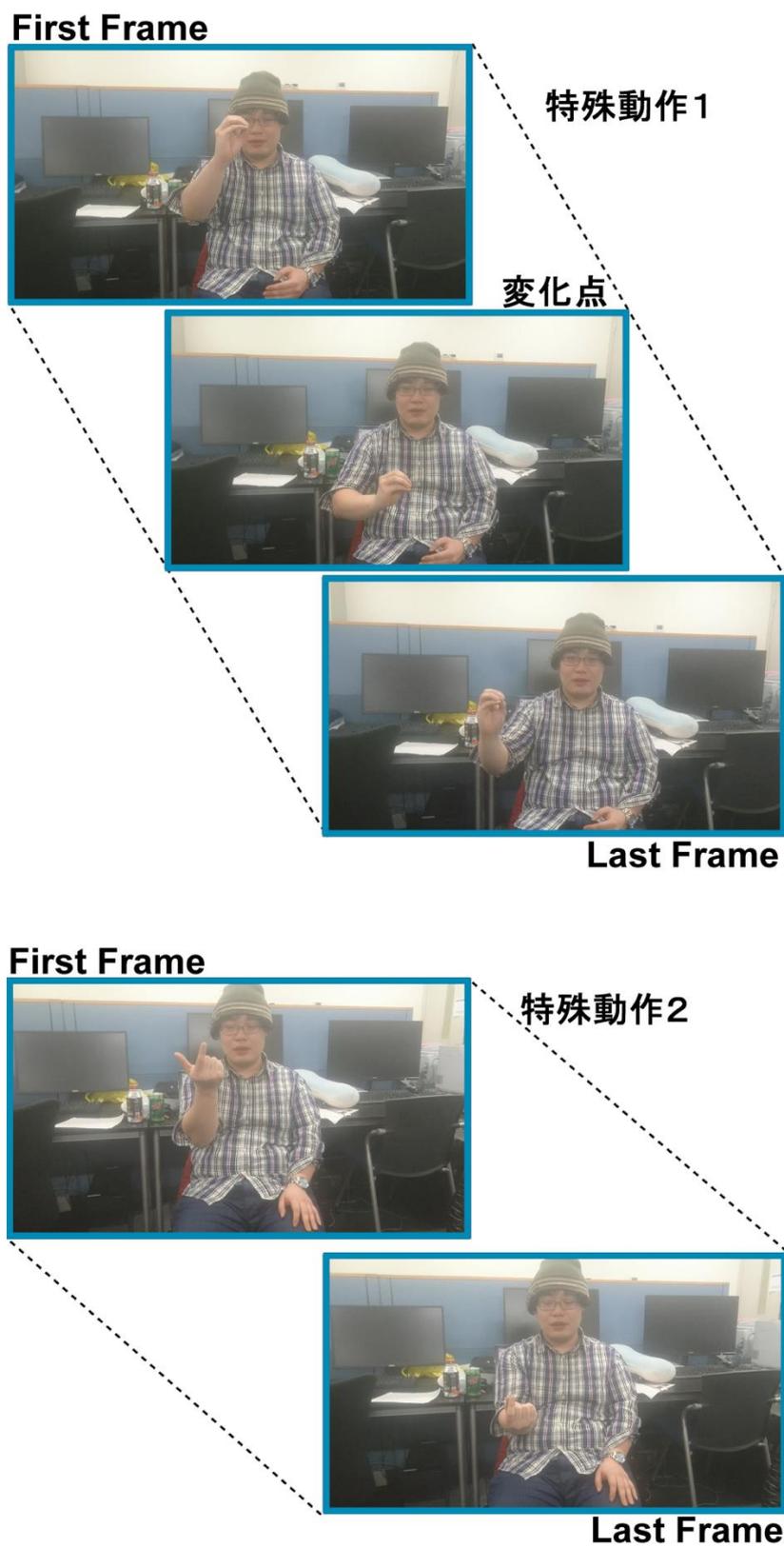


図 4.3 動的指文字データセットの一例 (特殊動作)

4.4 評価実験 1: 静的指文字

4.3 節で作成した静的データセットを用いて提案手法の評価実験を行った。本節でその結果を提示する。精度評価に用いる指標を表 4.6 および式(4.1)~(4.4)で与える。また、SVM のフレームワークには `scikit-learn` [16] を使用し、パラメータ等の設定は表 4.6 に従う。なお、本節で得られる結果についての考察は 4.6 節において行う。

表 4.6 混同行列

予測 \ 真値	Positive	Negative
Positive	True Positive (TP)	False Positive (FP)
Negative	False Negative (FN)	True Negative (TN)

$$\text{適合率} \quad \textit{Precision} = \frac{TP}{TP + FP} \quad (4.1)$$

$$\text{再現率} \quad \textit{Recall} = \frac{TP}{TP + FN} \quad (4.2)$$

$$\text{F 値} \quad F_1 \textit{ score} = \frac{2 \cdot \textit{Precision} \cdot \textit{Recall}}{\textit{Precision} + \textit{Recall}} \quad (4.3)$$

$$\text{正解値} \quad \textit{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \quad (4.4)$$

表 4.7 使用 SVM の諸元

<code>scikit-learn</code>	Ver. 0.20.2
カーネル	RBF (ガウシアンカーネル)
C	10
γ	0.1
多クラス方式	1 対 1 方式

4.4.1 特徴量の有効性確認

カメラが固定されていないため、本データセット内には撮影角度に変化が生じていることが予想される。提案手法は 20 個の 2 次元上での角度をセットとして用いるため、撮影角度が変化すると特徴ベクトルの各要素が大きく変化する。したがって、提案手法の適用には学習により角度変化が吸収可能であることを確かめる必要がある。

評価実験として、作成したデータセットのうち一人分を抜き出し、各文字の映像を訓練用

フレームとテスト用フレームに分割した。その後提案手法を訓練用フレームに適用し、テスト用フレームを分類した。なお、訓練フレームと学習用フレームの比は 7:3 である。一連の予測を 10 回行い、得られた全静的指文字の正解率を平均した結果を表 4.8 に示す。

表 4.8 角度特徴量の適用性確認

	元データ	正規化・欠損値補完
正解率	0.0815	0.958
平均テスト数	1271.3	

4.4.2 データセット全体への適用

次に、データセット全体に提案手法を適用し、性能評価を行う。本実験で用いるデータセットは全部で 12 人分と少数であるため、Leave-One-Out 交差検定を適用する。ただし表 4.8 の結果から、撮影した映像のあるフレームを学習に用いた場合はその動画全体を予測できてしまう。したがって、分割は被験者単位で実行する。被験者 1 名分の全静的指文字 41 文字分の姿勢推定結果を図 4.4 に示す。

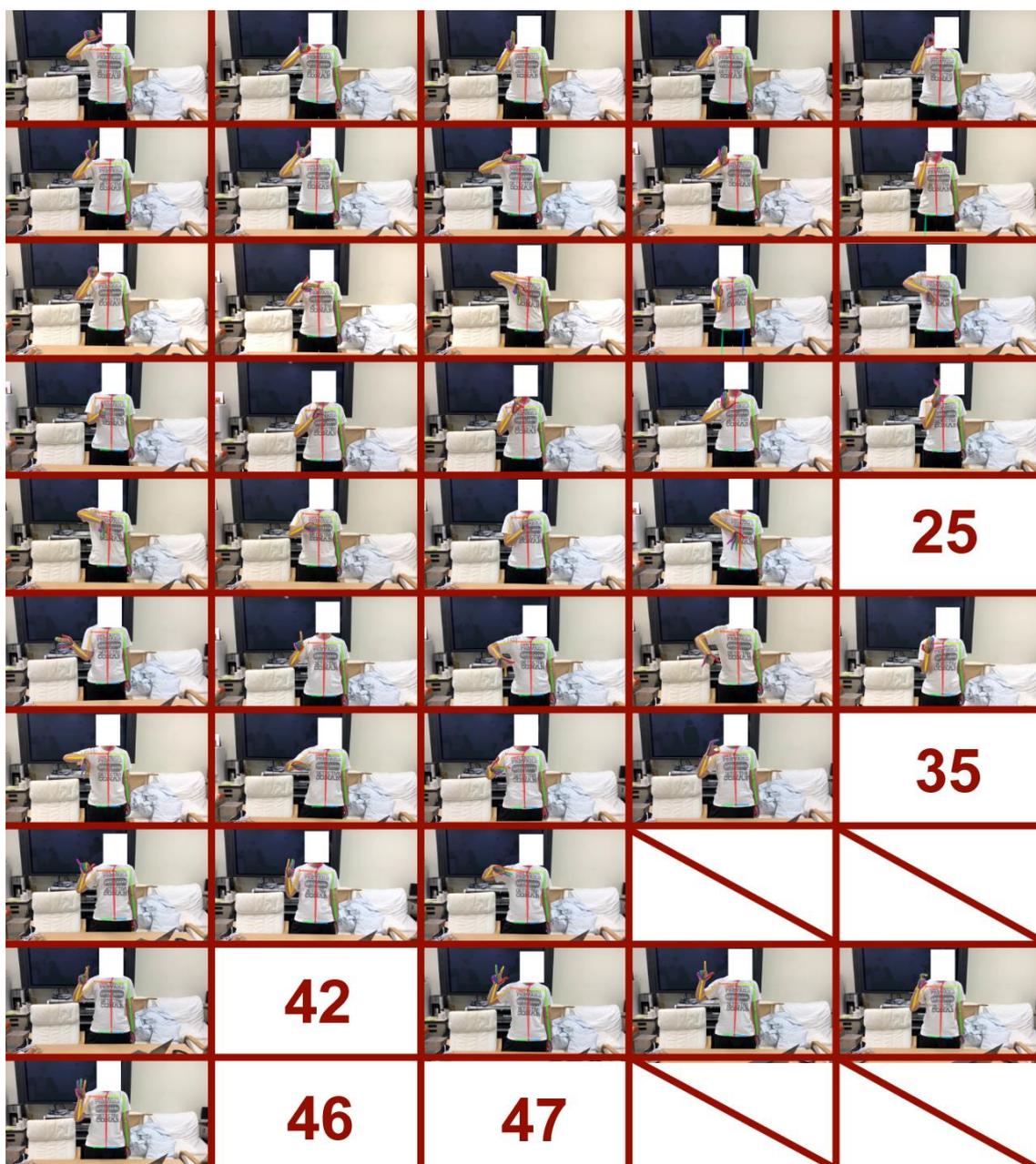


図 4.4 全静的指文字の姿勢推定結果一例

得られた全静的指文字の 1 文字ごとの平均正解率，および全文字平均正解率を表 4.9 に示す．また，適合率，再現率，F 値の文字別平均をグラフにまとめ，図 4.5 に示す．

表 4.9 静的指文字データセット全体における提案手法精度

テスト番号	テスト数	正解率 (全文字平均)
1	11415	0.698
2	11065	0.794

3	6859	0.758
4	6758	0.792
5	5998	0.571
6	7500	0.456
7	6779	0.667
8	6945	0.731
9	6576	0.485
10	3726	0.450
11	7151	0.700
12	5044	0.533
AVG	7151.333	0.636
Weighted-AVG	7151.333	0.656

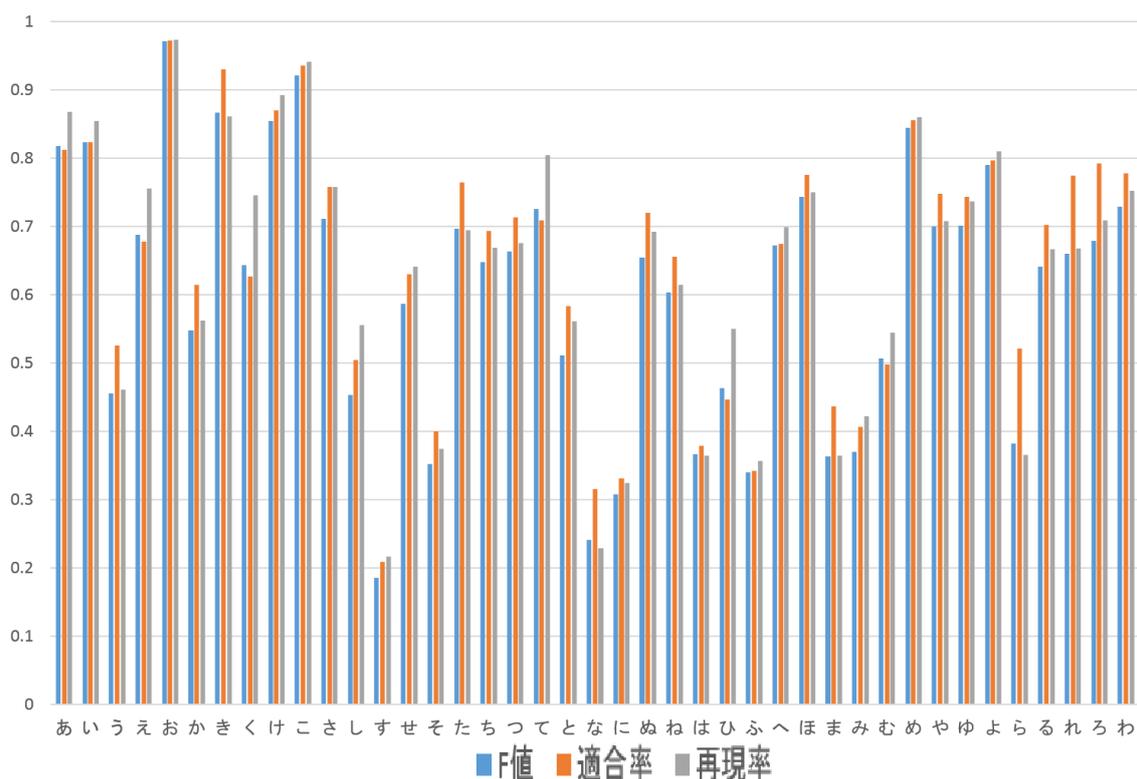


図 4.5 静的指文字の文字別分類精度 [17]

4.4.3 スケールに対するロバスト性確認

提案手法ではスケールに対して不変性を保つため、特徴量として関節角度を使用する。こ

の有効性を確かめるためスケール変化実験を行った。4.4.2 項で用いたデータセットの一部データ値を 0.1 倍し、テストデータに用いる。学習サンプルにはスケール変換前のデータセット全体を用いる。比較対象として OpenPose の出力座標を直接利用する手法 [18]を採用する。ただし、4.3 節で述べたように使用するデータセットには移動量の変化が存在するため、比較手法においては式(4.5)を適用することで手指関節点の座標を相対化して使用する。基準点は掌中央である。

$$\mathbf{p}'_{f,j} = \mathbf{p}_{f,j} - \mathbf{p}_{f,0} \quad (4.5)$$

結果を表 4.10 および図 4.6, 図 4.7 に示す。

表 4.10 スケール変化時における各手法の分類精度

	テスト数	正解率 (全文字平均)
提案手法	11415	0.939
座標値ベース	11415	0.0282

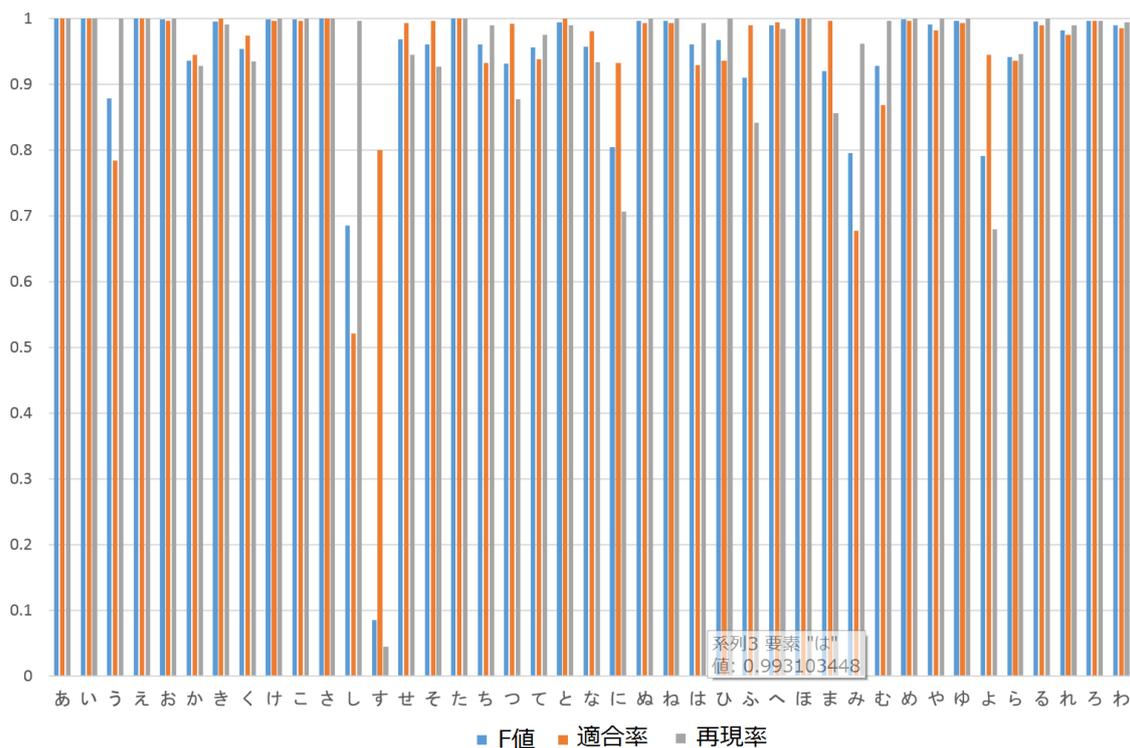


図 4.6 スケール変化時における提案手法の文字別分類精度

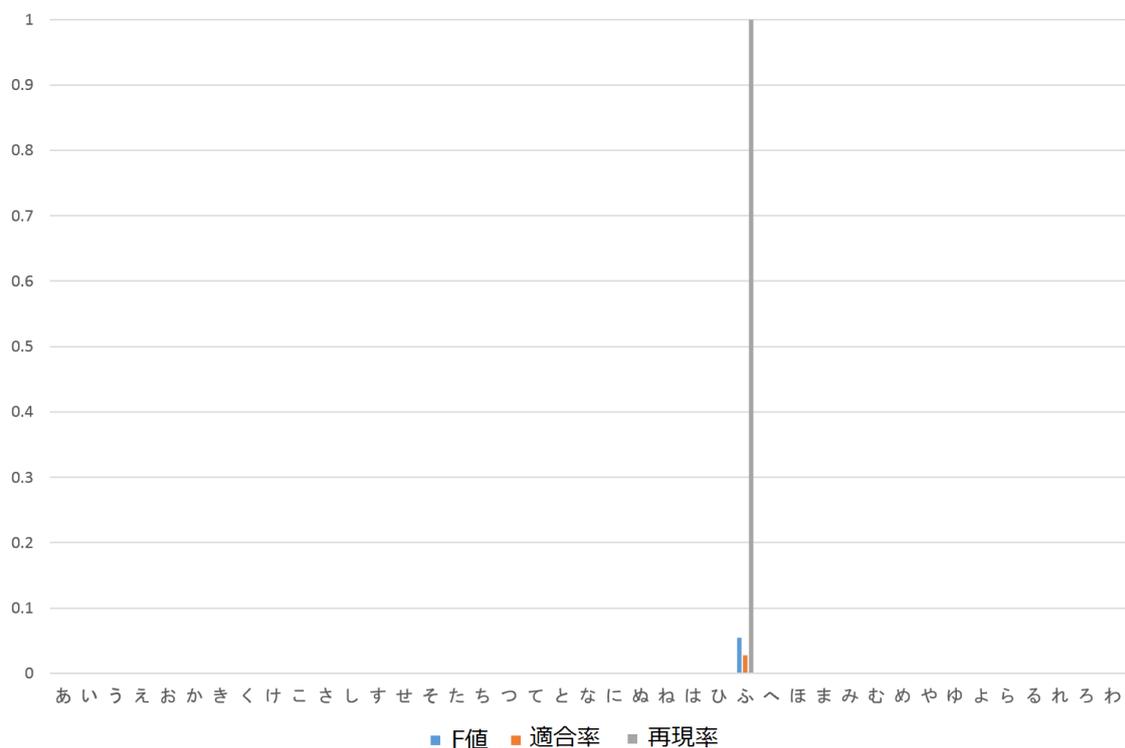


図 4.7 スケール変化時における座標値ベース手法の文字別分類精度

4.5 評価実験 2：動的指文字

図 4.2 のデータセットを用い，基本動作 6 クラスの分類実験を行った．評価基準および SVM の設定，実験方法は 4.4 節で述べたものと同様である．基本動作 6 に関しては静的指文字の実験に用いたデータセットの一部を転用した．図 4.8 に基本動作の姿勢推定結果を示す．また，テスト条件および正解率を表 4.11 に，精度のグラフを図 4.9 に示す．本節で示す結果は 4.4 節同様，4.6 節にて考察する．

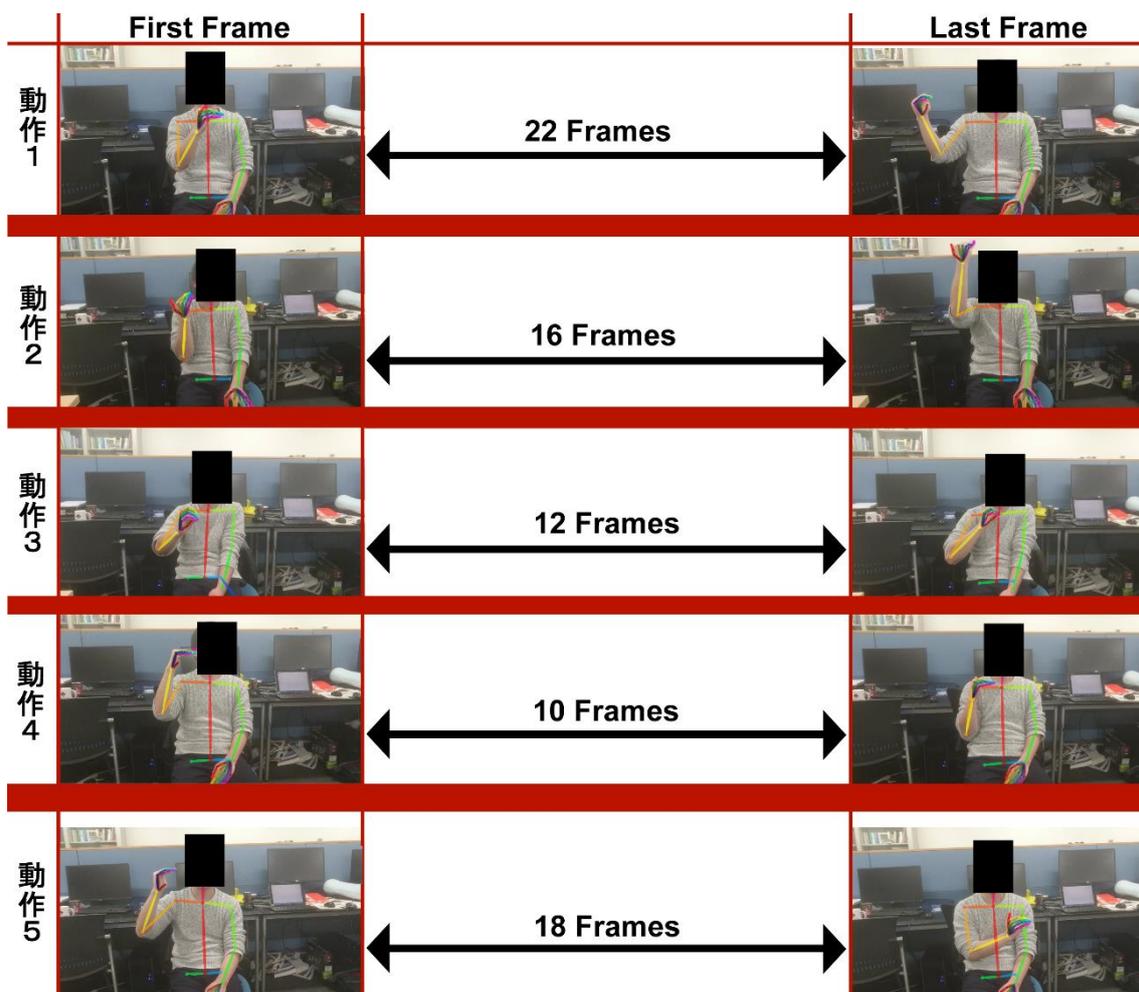


図 4.8 動的指文字各基本動作の姿勢推定結果一例

表 4.11 動的指文字データセットにおける提案手法精度

テスト番号	テスト数	正解率 (全動作平均)
1	50	0.947
2	50	0.950
3	50	0.983
4	50	0.903
5	50	0.987
6	50	0.997
7	50	0.957
AVG	50	0.960
Weighted-AVG	50	0.960

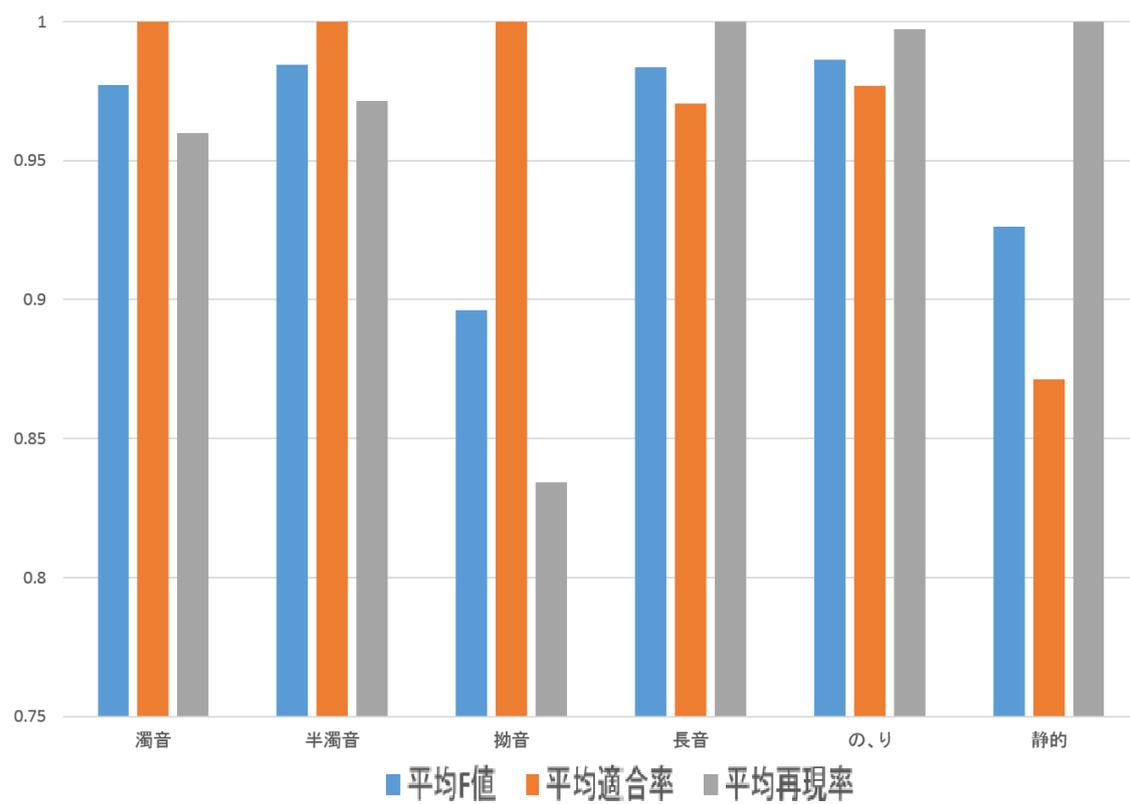


図 4.9 動的指文字の文字別分類精度

4.6 考察

本節では 4.4 節および 4.5 節で行ったカナ指文字分類の評価実験結果について考察を述べる。

4.6.1 静的指文字分類実験に関する考察

表 4.8 より、第 3 章で提案した正規化および欠損値補完を施すことで、単一映像内では未知フレームの静的指文字に対して 95%の正解率を得られることを確認した。したがって、提案手法によるデータ処理を行った上での角度ベクトルが特徴量として機能しているといえる。また、提案手法はある指文字の映像について一部フレームを正しく予測できればその映像全体を適切に分類可能である。

表 4.9 から、提案手法を適用することで 11 人分の学習データから未知の静的指文字映像に対して 63~65%程度の精度で正しい分類が可能であることが分かる。図 4.5 より最も高い精度を示した「お」については 97%程度、その他の複数の指文字で 8 割前後の正確性を達成した。

一方で「す」、「な」、「に」および「は」については 2~3 割程度と精度が低い。原因とし

て、図 4.10 に示す通りこれら 4 文字に共通する骨格の類似性が挙げられる。提案手法で特徴量として用いるものは手首から先の骨格構造に他ならず、手の向きや表裏は保持されない。そのため、手首から先の骨格が類似している指文字間で混同が発生したと考えられる。

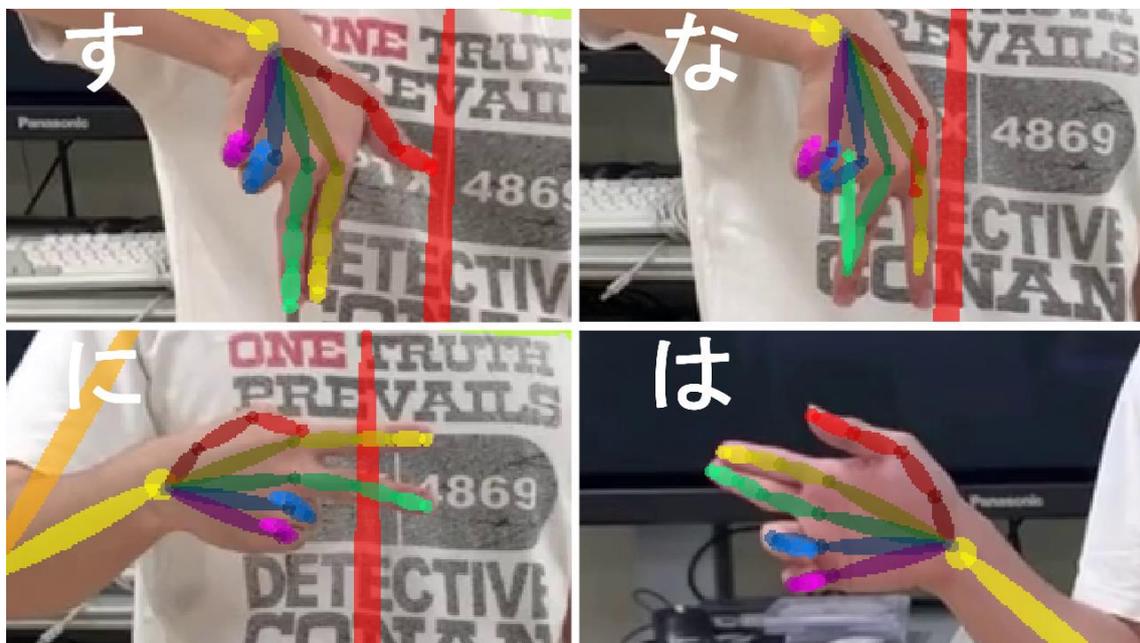


図 4.10 骨格類似文字群の例

また、その他の分類精度低下原因として、OpenPose 自体の部位検出精度の差異が考えられる。提案手法は、一部関節の欠損や数フレーム程度の欠損による異常値に対しては補完が可能である。しかし、入力された映像が全て推定に失敗していた場合は、特徴量を算出できない。したがって、OpenPose の学習過程における手指構造の偏りなどに起因する図 4.11 のような検出困難である姿勢に対しては OpenPose 自体のデータセットを改良する必要がある。

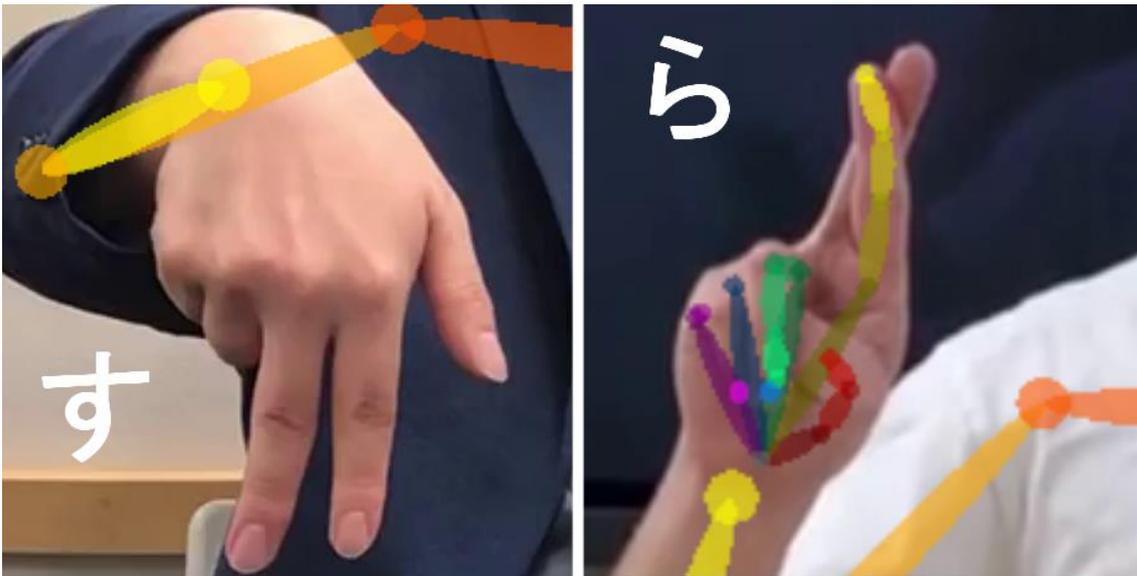


図 4.11 OpenPose の誤検出による推定失敗文字例

図 4.5 おいて最も低い精度であった指文字は「す」であるが、これは角度混同と検出失敗頻度の多さという二つの精度低下原因を併せ持ったことによるものと結論する。

最後にスケール変化への追従性について考察する。表 4.10 よりスケールが 1/10 と極端に変化した場合、座標値ベース手法での 3%程度の精度に対し提案手法は 94%の高精度を示した。その理由は、画像から得られる座標値が拡張されていても本質的に角度特徴量が不変であることに起因する。したがって、OpenPose が処理可能な限り、提案手法はどのようなサイズ・解像度の映像を適用しても、類似映像を学習済であればほぼ確実に正しい予測が可能である。この特徴は、解像度が現在ほど高くない過去映像やニュース番組などにおける手話キャプションをサンプル転用可能という面で大きい利点といえる。

ただし、図 4.6 から「す」に関してのみ再現率が非常に低いことが分かる。この結果からもデータセット全体実験時に判明した「す」に対する提案手法の特性が裏付けられる。

4.6.2 動的指文字分類実験に関する考察

表 4.11 よりいずれの被験者に対しても提案手法の全動作平均分類精度が 90%を越え、少ないサンプル数でも安定して高い性能を示すことが分かる。動的指文字における基本動作がいずれも直線的であるという特性を 360° 全方向に拡張した移動差分の tangent と手領域の変化率により正確に抽出できた結果といえる。また、用いる四つの特徴量を全てフレーム間の相対関係のみで算出可能な値に限定したため、データセット撮影時のスケール差を吸収できたことも検出精度に繋がったと考えられる。

図 4.9 より、文字別の F 値に注目した場合、いずれの動作も 90%近い精度を達成しているが、拗音・促音・「を」に対応する基本動作 3 のみ 90%に届かない結果となった。この

原因として OpenPose が 2 次元出力である点が考えられる。奥行き情報を手領域面積変化率で近似している以上、被験者が手前に動かす際の左右方向への若干のブレや撮影角度による手領域面積の変化が誤差として生じる。

なお、実験時では 4.6.1 項で得られた結果を基に OpenPose が安定して動作する骨格を用いたが、本提案手法では本質的に 1 点を手から抽出できれば適用可能である。したがって、平均座標を使用している箇所を指先に変更すれば被験者が手指の姿勢を厳密に維持していない場合でも適用性を保つことが期待できる。

4.7 むすび

本章では OpenPose の撮影箇所による検出特性を測る予備実験および提案手法の評価実験とその結果を示し、考察を述べた。予備実験では上半身の関節の被写数が手指検出精度に大きく影響することを確認し、データセット構築に必要な条件を導出した。この結果に基づいて作成した小規模データセットに対し評価実験を行い、提案手法によるデータ処理の有効性を確認した。また全静的指文字に対して 63~65%程度、動的指文字で 96%の分類精度を得た。以上の結果から、提案手法によるカナ指文字の分類精度の大部分は静的指文字の姿勢分類に依存すると考えられる。

第5章

提案手法拡張

5.1 まえがき

第 4 章で行った実験結果から，カナ指文字の分類精度は静的指文字の予測精度に大きく影響されることが分かった．第 3 章で提案した手法は手指内部の関節角度のみを特徴量として用いる．そのため，骨格がほぼ同一で手首の角度や手の裏表のみが異なる静的指文字群で混同が発生していると考えられる．本章では，提案手法を拡張し，これらの差異を分離する方法を述べる．

5.2 多段 SVM による二段階分類

本章で提案する拡張手法の概要を図 5.1 に示す．最初の SVM により手の裏表・角度を一意に定めた後，それらが同一の静的指文字のみを第 3 章の手法で分離する．本手法では初めの SVM を SVM1，次の SVM を SVM2 と定義する．また，以降では本節で提案する手法を「拡張法」と呼ぶ．

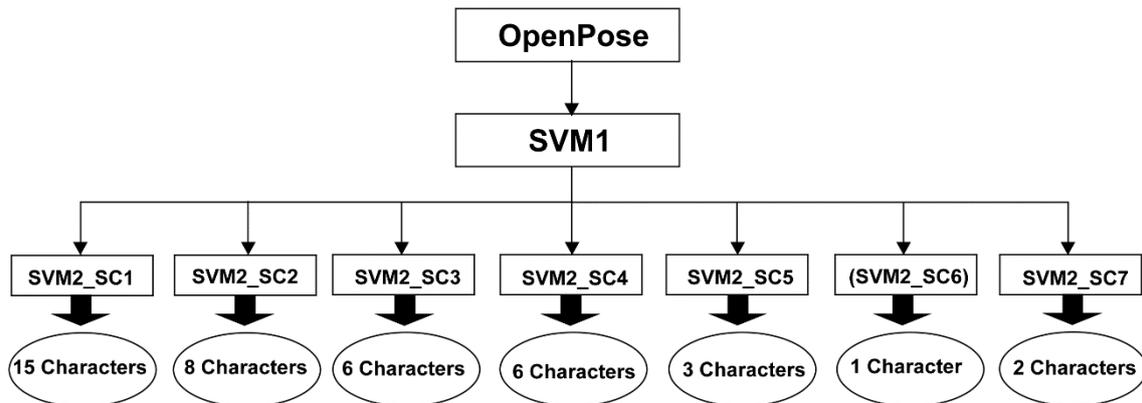


図 5.1 多段 SVM 分類模式図

5.3 静的指文字のサブクラス化

静的指文字 41 文字を図 5.2 に示すように

- 手首の角度
- 手の裏表

によって細分化する. これにより「す」「な」「に」および「は」をはじめとする類似指文字を区別する.

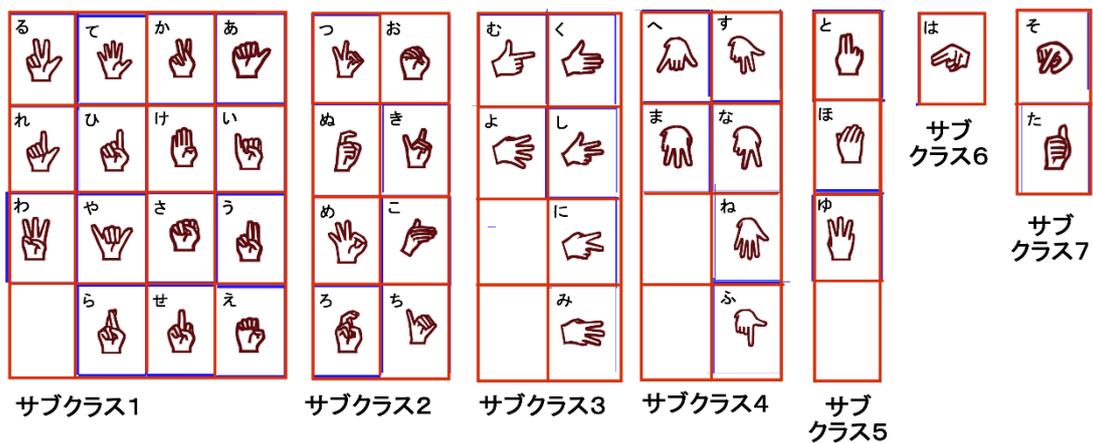


図 5.2 各サブクラスの構成

5.4 サブクラス分類

SVM1 によるサブクラス分離を行うため、本節で算出法を述べる手首の回転と手の裏表を定義する 2 種の特徴量を導入する。したがって、拡張手法でのサブクラス分類は 2 次元 7 クラス分類問題である。

5.4.1 手首回転量の特定

第 3 章と同様に画像横軸を U，横軸を V とする。ここで、掌中央から見て人差し指，中指，薬指および小指の第 3 関節は不動であることに着目し，手の向きを表すベクトル $\mathbf{h}_f = (u_{norm2}, v_{norm2})$ を式(5.1)で定義する。

$$\mathbf{h}_f = \frac{\mathbf{p}_{f,5} + \mathbf{p}_{f,9} + \mathbf{p}_{f,13} + \mathbf{p}_{f,17} - \mathbf{p}_{f,0}}{4} \quad (5.1)$$

入力画像内の手が上下左右のいずれを向いているかを V 軸方向単位ベクトルと式(5.1)の成す角を用いて計算する。ただし、人体における手首は前腕を軸としておよそ 270° の可動域を持つ。したがって、 360° 全方向の回転量を定義する必要があるため、式(3.16)ではなく式(5.2)を用いる。

$$\text{rot} = \text{Arctan} \left(\frac{v_{norm2}}{u_{norm2}} \right) - \frac{\pi}{2} \quad (5.2)$$

rot の degree 表記を SVM1 における第 1 特徴量とする。図 5.3 は本項および 5.4.2 項で用いる回転処理である。

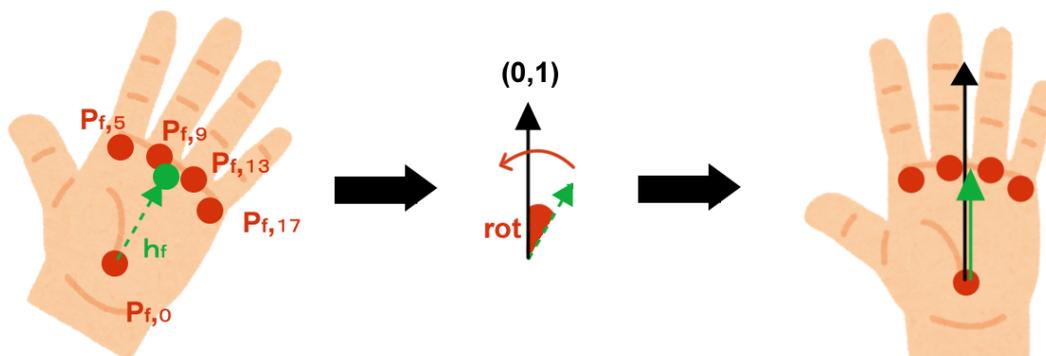


図 5.3 手掌方向基準ベクトルおよび回転処理

5.4.2 手の裏表判定

本節では図 5.4 に示した 4 パタンの内容を保持する特徴量の算出法を述べる。

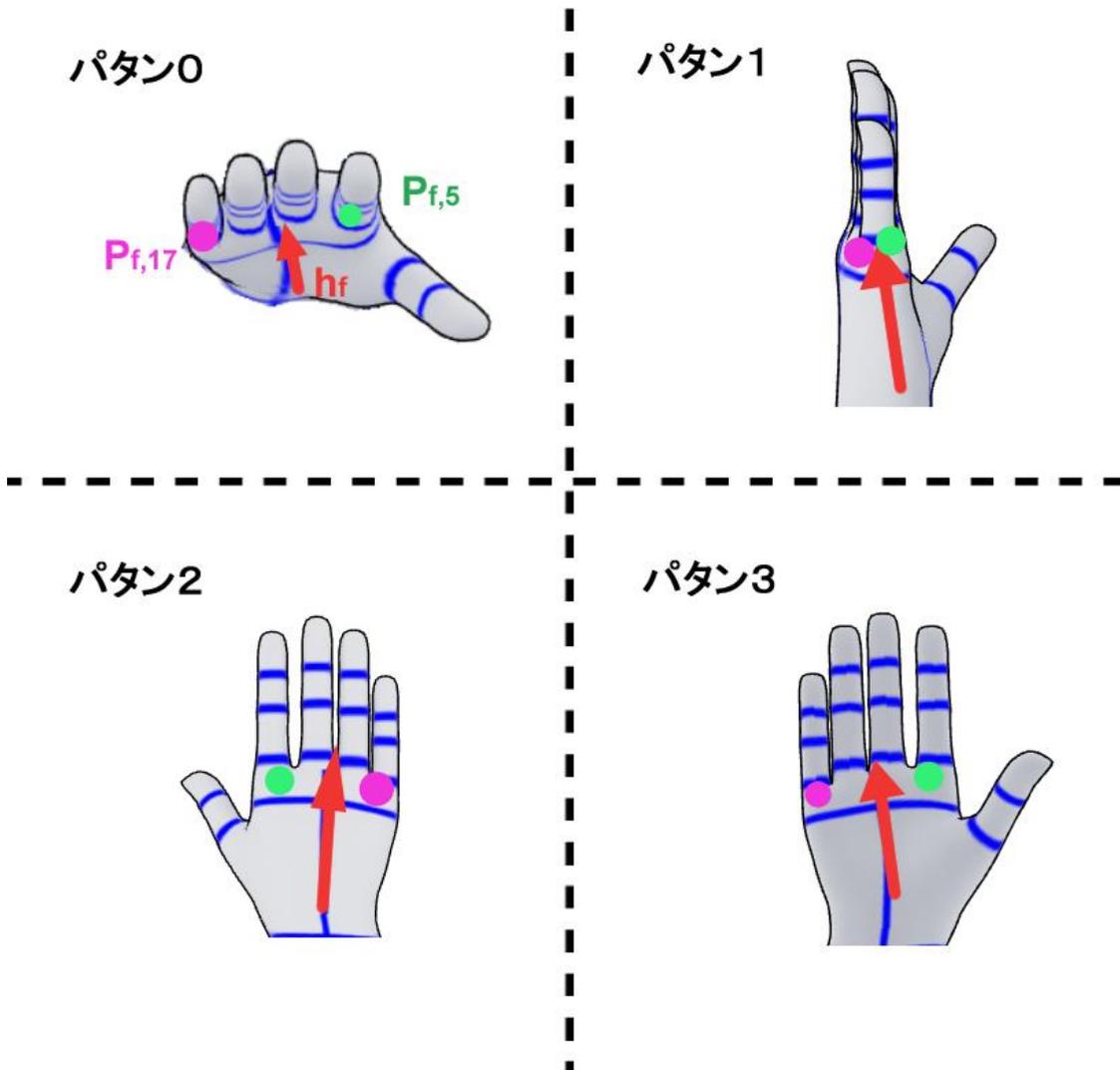


図 5.4 掌の向きによる 4 パタン定義

パターン 0 のとき、 h_f のノルムは他 3 パタンに比べて小さくなると考えられる。したがって式(5.1)の大きさから判定可能である。

手の裏表は人差し指・小指第 3 関節の位置関係から判断可能である。ただし、手指内における胴体座標系と画像平面 UV 座標系が一致していない場合、すなわち手首が回転している場合は全サブクラスに対して一意に比較できない。そのため、5.4.1 項により得られた手の回転量を利用し、

$$R = \begin{bmatrix} \cos(-rot) & -\sin(-rot) \\ \sin(-rot) & \cos(-rot) \end{bmatrix} \quad (5.3)$$

を用いて全手指関節点座標を式(5.4)により逆回転する.

$$\mathbf{p}'_{f,j} = R\mathbf{p}_{f,j} \quad (5.4)$$

以上より, 人差し指と小指第 3 関節の \mathbf{u} 座標が

- 等しければパターン 1
- 異なればパターン 2, 3

を全て特定できる.

得られた関節点のうち, 人差し指第 3 関節と小指第 3 関節に相当する $\mathbf{p}'_{f,5}$ と $\mathbf{p}'_{f,17}$ の \mathbf{u} 座標および $|\mathbf{h}_{f,j}|$ を式(5.5)に入力し, SVM1 の第 2 特徴量 dir を算出する.

$$\text{dir}(\mathbf{u}'_{f,5}, \mathbf{u}'_{f,17}, |\mathbf{h}_{f,j}|) = \begin{cases} 0, & \text{if } |\mathbf{h}_{f,j}| < \tau_2 \\ 1, & \text{else if } |\mathbf{u}'_{f,5} - \mathbf{u}'_{f,17}| \leq \tau_3 \\ 2, & \text{else if } \mathbf{u}'_{f,5} < \mathbf{u}'_{f,17} \\ 3, & \text{else} \end{cases} \quad (5.5)$$

なお, τ_2 ならびに τ_3 は適当な閾値である.

5.5 サブクラス内静的指文字分類

5.4 節により, 各サブクラス内には手首の回転, 掌の向きが全て類似する静的指文字のみが存在する. この条件に基づいて第 3 章で述べた提案手法による分類を行うことで類似骨格を分離する.

5.6 むすび

本章では多段 SVM を用いて静的指文字における類似姿勢を分離し, 第 4 章で述べた静的指文字分類の精度低下原因への対処法を示した. 次章では本拡張手法の評価実験について述べる.

第6章

追加実験

6.1 まえがき

本章では前章で示した提案手法の拡張法について，第4章で作成したデータセットを用いて再度被験者ごとの Leave-One-Out 交差検定を行った結果を示す．実験条件，評価基準は全て4.4節と同様である．

6.2 サブクラス分類実験

SVM1により全静的指文字をサブクラスに分類した結果を表6.1，図6.1に示す．各結果の考察は6.4節にて行う．

表 6.1 拡張手法のサブクラス分類精度

テスト番号	テスト数	正解率
1	11415	0.891
2	11065	0.804
3	6859	0.885
4	6758	0.897
5	5998	0.771
6	7500	0.799
7	6779	0.850
8	6945	0.781

9	6576	0.787
10	3726	0.796
11	7151	0.857
12	5044	0.800
AVG	7151.333	0.827
Weighted-AVG	7151.333	0.831

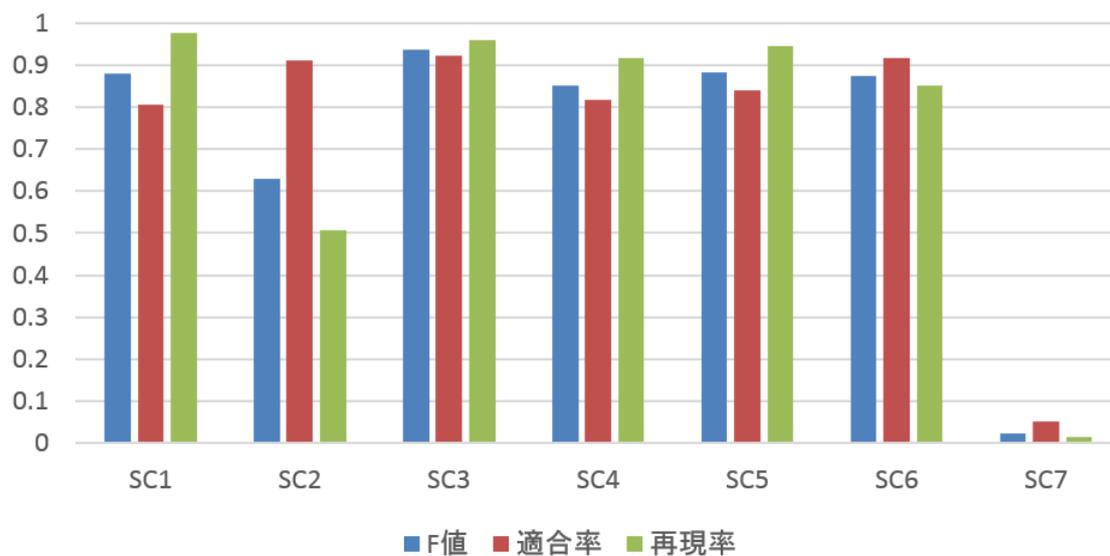


図 6.1 拡張手法によるサブクラスごとの平均分類精度

6.3 静的指文字のサブクラス内分類実験

SVM2 により各サブクラス内において第 3 章での提案手法を適用した結果を表 6.2-6.7, 図 6.2-6.7 に示す. なお, サブクラス 6 に含まれる指文字は「は」のみであるため SVM2 への入力は不要である. 6.2 節と同様に結果の考察は 6.4 節にて行う.

表 6.2 サブクラス 1 における提案手法精度

テスト番号	テスト数	正解率
1	4210	0.870071
2	3999	0.912728
3	2791	0.971695
4	2458	0.963385
5	1900	0.718947
6	2809	0.62015

7	2441	0.598116
8	2578	0.94647
9	2474	0.621261
10	1706	0.58558
11	2861	0.94093
12	1640	0.788415
AVG	2655.583	0.794812
Weighted-AVG	2655.583	0.813443

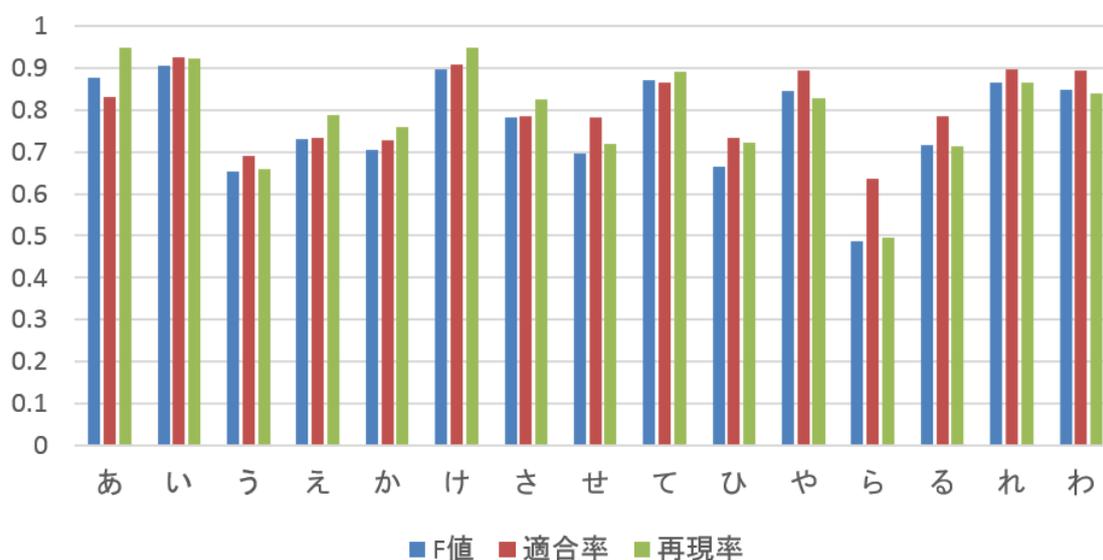


図 6.2 サブクラス 1 における提案手法精度 (文字別)

表 6.3 サブクラス 2 における提案手法精度

テスト番号	テスト数	正解率
1	2132	0.891182
2	2164	0.972274
3	1331	0.767092
4	1320	0.909091
5	1061	0.900094
6	1450	0.793793
7	1255	0.863745
8	1358	0.994109
9	1316	0.791033
10	816	0.790441

11	1439	0.789437
12	1322	0.516641
AVG	1413.667	0.831578
Weighted-AVG	1413.667	0.841193

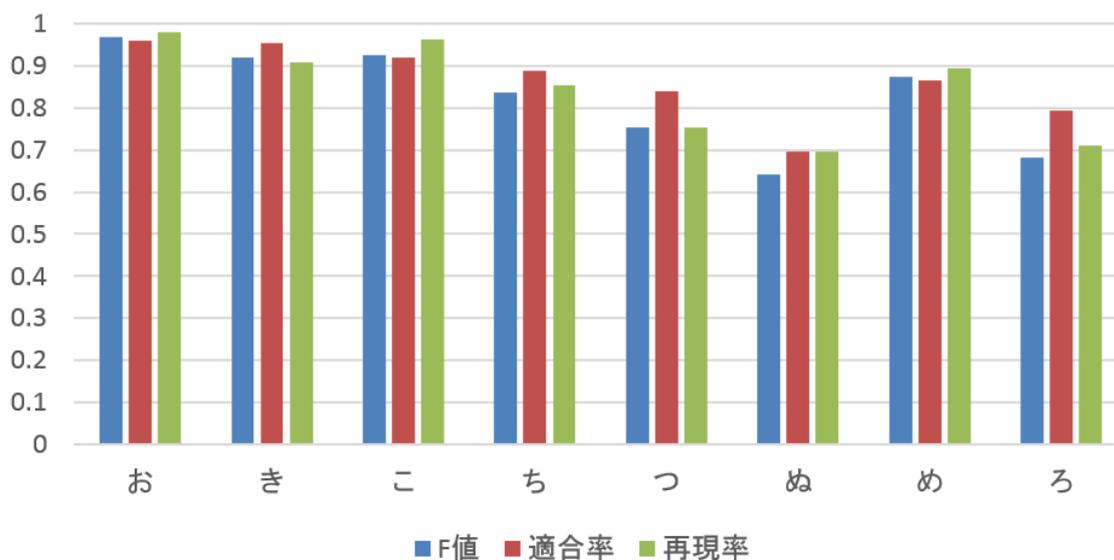


図 6.3 サブクラス 2 における提案手法精度 (文字別)

表 6.4 サブクラス 3 における提案手法精度

テスト番号	テスト数	正解率
1	1657	0.728425
2	1631	0.993869
3	1025	0.88
4	984	0.96748
5	887	0.940248
6	1110	0.781081
7	1327	0.975885
8	997	0.814443
9	984	0.942073
10	397	0.730479
11	1054	0.754269
12	962	0.753638
AVG	1084.583	0.855158

Weighted-AVG	1084.583	0.86262
--------------	----------	---------

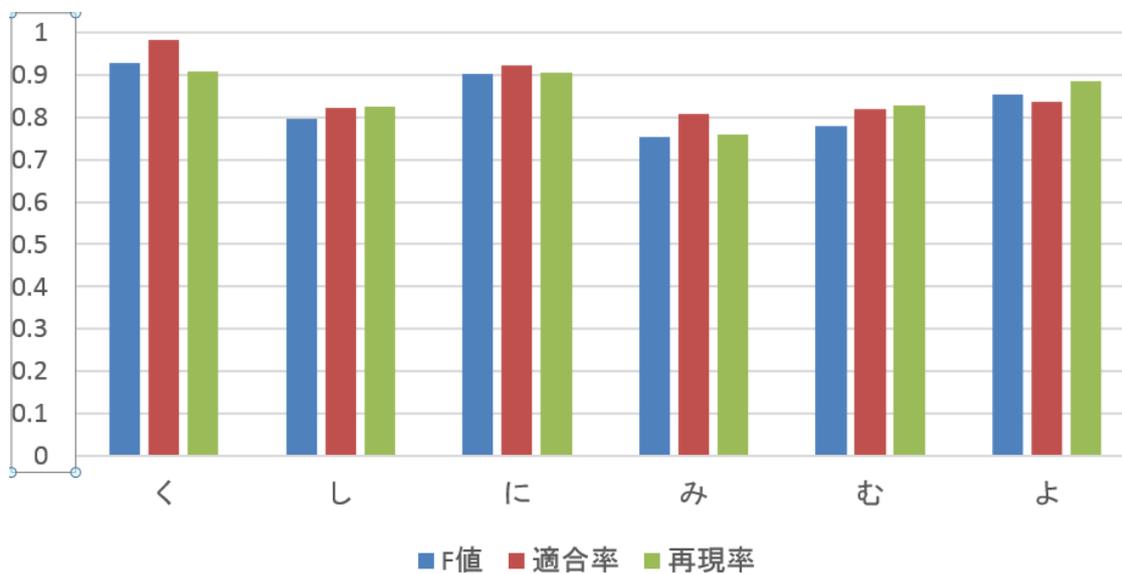


図 6.4 サブクラス 3 における提案手法精度 (文字別)

表 6.5 サブクラス 4 における提案手法精度

テスト番号	テスト数	正解率
1	1733	0.992499
2	1629	0.511971
3	1013	0.819348
4	1001	0.892108
5	1052	0.5827
6	1073	0.204101
7	962	0.879418
8	1016	0.661417
9	921	0.636265
10	660	0.907576
11	1052	0.84981
12	719	0.742698
AVG	1069.25	0.723326
Weighted-AVG	1069.25	0.720131

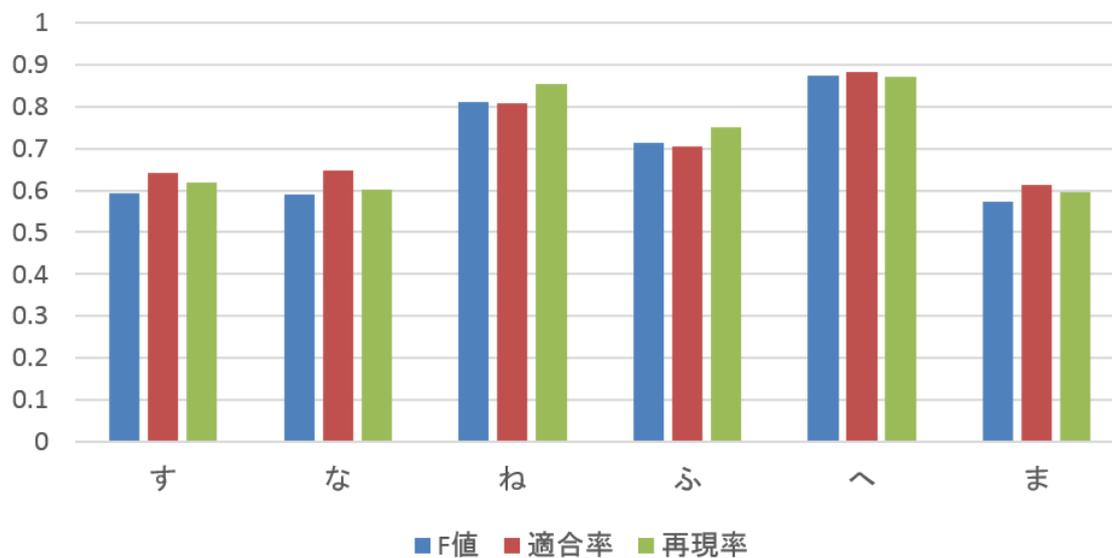


図 6.5 サブクラス 4 における提案手法精度 (文字別)

表 6.6 サブクラス 5 における提案手法精度

テスト番号	テスト数	正解率
1	825	1
2	823	1
3	506	1
4	493	1
5	515	0.990291
6	540	0.937037
7	418	0.95933
8	493	1
9	484	0.516529
10	235	0.982979
11	527	1
12	473	0.985201
AVG	527.6667	0.947614
Weighted-AVG	527.6667	0.952464

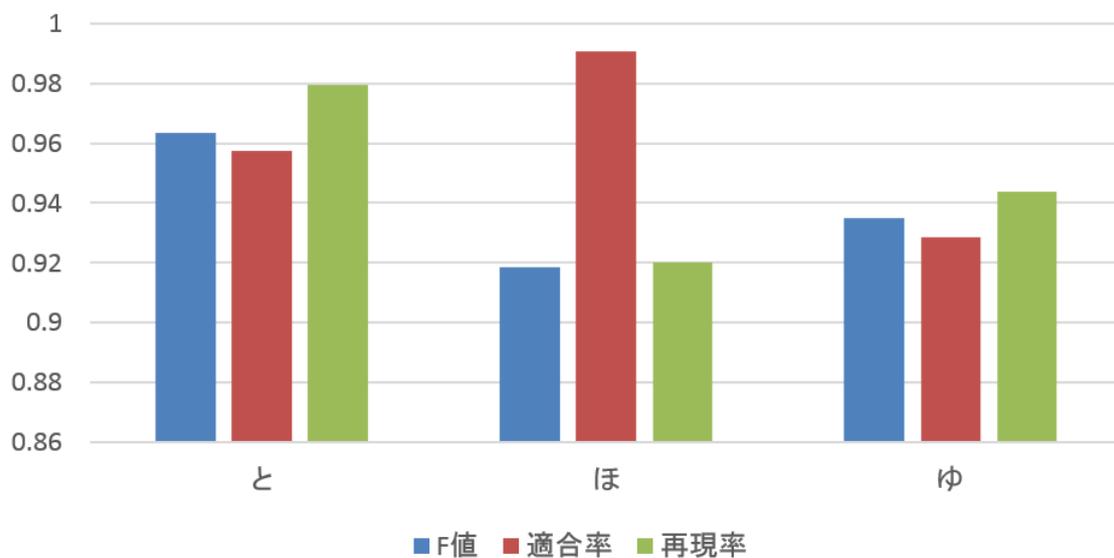


図 6.6 サブクラス 5 における提案手法精度 (文字別)

表 6.7 サブクラス 7 における提案手法精度

テスト番号	テスト数	正解率
1	568	1
2	533	1
3	302	0.880795
4	338	0.988166
5	338	0.920118
6	338	0.488166
7	226	1
8	333	0.942943
9	322	0.639752
10	166	0.608434
11	337	0.985163
12	174	0.471264
AVG	331.25	0.827067
Weighted-AVG	331.25	0.864906

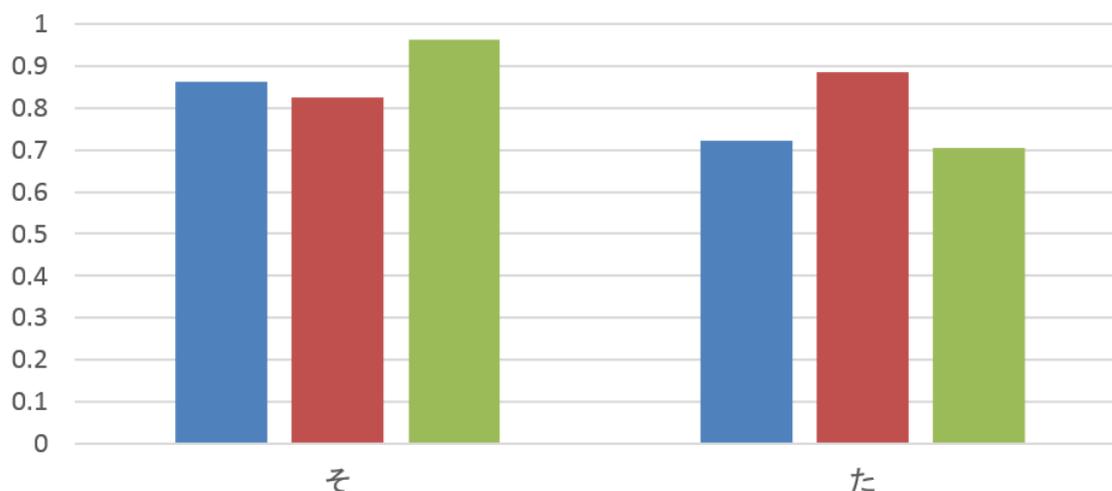


図 6.7 サブクラス 7 における提案手法精度 (文字別)

6.4 考察

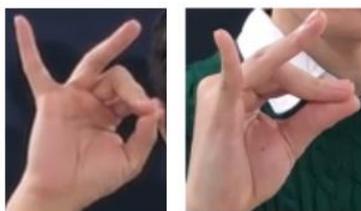
6.2-6.3 節で示した拡張法のサブクラス分類精度，および類似性を排除した上での提案手法精度に関する考察を本節において述べる．

6.4.1 サブクラス分類精度

表 6.1 よりサブクラス分類の正解率は平均・重み付き平均ともに 83%前後の精度を示した．したがって，全体的には拡張法における二つの特徴量が手の上下左右，掌の裏表を分類する上で有効性を示したといえる．しかし，図 6.1 から拡張法はどのサブクラスに対しても平均的に高精度を示しているとはいえない．サブクラス 2 に対しては適合率こそ 90%以上であるものの，再現率は 50%程度に留まった．この結果から，拡張法はサブクラス 2 に属する指文字群の検出力に欠けることが分かる．考えられる原因は図 6.8 に示す「つ」のように，保持する手首の角度が曖昧な文字の存在である．指先を向ける角度が被験者によって異なるため，サブクラス識別器である SVM1 において「正面」と「横」の判別境界付近に存在するサンプル数が増加したと予想される．

Subclass 2

「つ」



「そ」



「た」



Subclass 7

図 6.8 保持角度に曖昧性が認められる指文字群

また、サブクラス 7 に対する実験結果は F 値、適合率、再現率のいずれも 5% に満たない精度である。想定される原因は 5.3 節に述べた閾値 τ_2 の値であり、実験では定数で与えている。したがって、スケール不定性が保たれず、撮影距離によっては被験者が手を突き出している場合でも $|h_f|$ の値が閾値を超えてしまう可能性が考えられる。加えて、サブクラス 7 を構成する 2 種の指文字にはサブクラス 2 と同じくサンプルに保持角度の曖昧性が認められた。図 6.8 下段に例を示す。二つの精度低下要因を併せ持った結果、拡張法による分類精度が他サブクラスよりも極端に低下したと考えられる。

6.4.2 サブクラス内での静的指文字分類精度

表 6.2-6.7 より、サブクラス内における提案手法の精度は最低値であるサブクラス 4 内分類においても 72%程度であることが分かる。その他のクラスについては概ね 80%以上の正解率を達成した。表 6.2-6.7 より重み付き平均正解率を抜粋して表 6.8 に示す。ただし、静的指文字識別器 SVM2 に入力する必要がないサブクラス 6 に関しては表 6.1 に示した値を用いる。なお、表 6.8 における重み付き平均は各サブクラス内に存在する静的指文字数に応じた計算値である。

表 6.8 サブクラス内分類実験結果のまとめ

サブクラス名	クラス内指文字数	正解率
1	15	0.813443
2	8	0.841193
3	6	0.86262
4	6	0.720131
5	3	0.952464
6	1	0.799
7	2	0.864906
AVG		0.836
Weighted-AVG		0.825

表 6.8 より提案手法は類似骨格が分離されている条件下では単純平均で 83.6%、重み付き平均で 82.5%の正解率を示すことが確認され、いずれも 4.4 節における実験値 63.6%および 65.6%から 15 ポイント以上改善された。したがって、4.6 節で述べた角度および掌の向きに関する考察が正しく、拡張法による類似骨格の排除が認識精度向上に有効といえる。また、各サブクラスの重みは大きく異なるものの、単純平均と重み付き平均の差は 1.1 ポイント程度であることからどのサブクラス内静的指文字に対しても提案手法による安定した分類が可能であるといえる。

次に、文字別の分類精度から OpenPose の検出特性について考察を深める。各サブクラスにおいて F 値を基準に検出精度が他より低い文字を挙げると、図 6.2-6.7 よりサブクラス 1 では「ら」、サブクラス 2 では「ぬ」、サブクラス 3 では「み」、サブクラス 4 では「ま」、サブクラス 5「ほ」、サブクラス 7 では「た」である。これらのうち 50%を下回る文字は「ら」のみである。したがって、OpenPose による検出が特に困難な姿勢は「ら」での人差し指と中指の交差であることがいえる。また、列挙した 7 文字の分類結果を 4.4 節と比較するといずれも精度の向上が認められることから拡張法の有効性を再度確認できる。

さらに、上述の考察に基づいて最も分類精度が低かったサブクラス 4 の文字群に着目する。F 値 60%を基準に考えると、「す」「な」「ま」を抽出できる。これらの文字は図 6.9 に

示すように、互いを分離する差異が1本の指の屈伸のみである。したがって、前述の「ら」の特性と併せて現状の OpenPose では指1本単位での座標検出が安定せず、特に近接する指の交差判定等の複雑な姿勢推定は困難であると結論できる。



図 6.9 指文字「す」「な」「ま」の相違点

6.4.3 処理全体での分類精度期待値

拡張法の導入により提案手法自体の分類精度が向上することは 6.4.2 項で考察した通りであるが、本実験における設定下ではサブクラス分類における閾値などに課題が残ることを 6.4.1. 本稿では、現段階の多段 SVM に静的指文字を入力した際の分類期待値を式(6.1)により計算し、将来課題とする。具体的な改善案としては 6.4.2 に述べた考察より SVM1 に用いる閾値をスケール不定な導出法に変更することが考えられる。

$$\begin{aligned}
 Exp &= Accuracy_{SVM1} * Accuracy_{SVM2} \\
 &= 0.831 * 0.825 \\
 &= 0.6856
 \end{aligned}
 \tag{6.1}$$

6.5 むすび

本章では第5章で提案した拡張法の意義を追加実験により確認し、サブクラス分類で83%、サブクラス内分類で82%程度の精度を得た。また、実験結果から第4章で述べた考察の妥当性を判断し、現段階における提案手法拡張により68.7%程度の分類期待値となることを示した。

第7章

結論

本研究では単眼 RGB 画像を用いてカナ指文字を正しく分類するアプローチとして、OpenPose により得られた座標点から指文字の各関節角度を計算しベクトル化することで、SVM によって対応する文字ラベルへ分類する手法を提案した。また、指文字を骨格部分と動作部分に大別することで、指文字の複雑な構成を静止画である静的指文字の分類と単純な動作である動的指文字の検出に分解できることを述べた。

予備実験においては、OpenPose の部位検出特性を調査し、その結果を考慮して小規模かつスケール不定なデータセットを構築した。評価実験では、OpenPose で得られた座標値を関節角度に変換し、関節ごとの正規化および他フレームの情報による欠損値補完を施すことでフレーム間の静的指文字予測精度が 95%を示すことを確認した。さらに、データセット全体に対し Leave-One-Out 交差検定を行い、提案手法が未知の静的指文字映像に対して全文字平均で 65%程度の予測精度を示すことを確かめた。また、スケール変化に対してロバスト性を保つことを確認した。動的指文字に対しても同様に評価実験を行い、こちらは平均 96%程度の精度を示すことを確かめた。

さらに、評価実験の結果を受けて手首の角度、掌の向きを SVM により分離した後に提案手法を適用する拡張法を述べた。追加実験の結果から、拡張法による静的指文字の分類期待値 68.7%を得た。また、二つの分類段階においていずれも 80%以上の精度を示すことを確認した。

本研究で行った実験結果から、指文字画像の RGB 値を直接用いず、姿勢推定により座標点を抽出して特徴量を構築することにより、少数の単純 RGB 画像でも効率的に学習ライブラリを訓練することが期待できる。また、提案手法は利用可能な画像の制約が少ないため、既存の過去画像を学習サンプルに転用し分類精度の向上を図ることが容易である。

謝辞

本研究を行うにあたり、高価な機材や適切なアドバイスを惜しみなく授けて下さった渡辺教授に深く感謝いたします。また、本実験に用いるデータセットの作成は被験者の方々に多大な労力を強いるものでしたが、快く引き受けて下さった友人や後輩の皆様の尽力に感謝申し上げます。そして、多様なテーマを抱える画像処理ゼミにおいて一人ひとりの課題を把握し、常に献身的な助力を頂いた石川氏なしには本研究を進めることは不可能でした。改めてここに心より御礼申し上げます。研究室の皆様におかれましてはこれより先も、一層のご活躍ならびにご健勝を願っています。

参考文献

- [1] 梅田修一, NHK 放送技術研究所, ” NHK における手話 CG 研究について”, 第 5 回サイレント音声認識ワークショップ, C-12, 29-30th Sep. 2018.
- [2] NHK 放送技術研究所, ” スポーツ情報の手話 CG 制作システム”, 技研公開 2018, 24-27th May 2018. [オンライン]. Available: <https://www.nhk.or.jp/strl/open2018/tenji/6.html>. [アクセス日: 27 12 2018].
- [3] NTT データ, NTT データ SBC, シャープ, ” 国内初, AI 技術を活用したロボホン向け手話通訳アプリケーションを共同開発”, [オンライン]. Available: <https://www.nhk.or.jp/strl/open2018/tenji/6.html>. [アクセス日: 28 12 2018].
- [4] "SignAll," [Online]. Available: <http://www.signall.us/>. [Access date: 28 12 2018].
- [5] Ronchetti, F., Quiroga, F., Estrebou, C.A., Lanzarini, L.C., Rosete, A.: Lsa64: An argentinian sign language dataset. In: XXII Congreso Argentino de Ciencias de la Computación (CACIC 2016), pp. 794-803. Red de Universidades con Carreras en Informática (RedUNCI) 2016.
- [6] D. Konstantinidis, K. Dimitropoulos, P. Daras, "A Deep Learning Approach for Analyzing Video and Skeletal Features in Sign Language Recognition," in 2018 IEEE International Conference on Imaging Systems and Techniques (IST), pp1-6, on 16-18th Oct. 2018.
- [7] B. E. Boser, I. M. Guyon, V. N. Vapnik, "A Training Algorithm for Optimal Margin Classifiers," in the fifth annual workshop on Computational learning theory (COLT'92), pp. 144-152, 27-29th July 1992.
- [8] Y. LeCun, L. Bottou, P. Haffner, "Gradient-Based Learning Applied to Document Recognition," in IEEE, Vol. 86, No. 11, pp. 2278-2323, Nov. 1998.
- [9] Cao, Zhe, Simon, Tomas, Wei, Shih En and Sheikh, Yaser, " Realtime Multi-Person

- 2D Pose Estimation using Part Affinity Fields," in 30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2017, Papers (128), pp. 1302-1310, on 21-26th July 2017.
- [10] C. M. Bishop, 元田浩, 栗田多喜夫, 樋口知之, 松本裕治, 村田昇, "パターン認識と機械学習 下", 第4版, 丸善出版, 2012, 433p, ISBN: 978-4-621-06124-4 C 3055.
- [11] 平博順, 春野雅彦, "Support Vector Machine によるテキスト分類における属性選択", 情報処理学会論文誌, 41(4), pp. 1113-1123, Apr. 2000.
- [12] C. M. Bishop, 元田浩, 栗田多喜夫, 樋口知之, 松本裕治, 村田昇, "パターン認識と機械学習 上", 第5版, 丸善出版, 2012, 349p, ISBN: 978-4-621-06122-0 C 3055.
- [13] Wei, Shih En, Ramakrishna, Varun, Kanade, Takeo, and Sheikh, Yaser, "Convolutional Pose Machines," in CVPR 2016, Papers (251), pp. 4724-4732, on 27-30th June 2016.
- [14] CMU-Perceptual-Computing-Lab, "openpose," [Online]. Available: https://github.com/CMU-Perceptual-Computing-Lab/openpose/blob/master/doc/media/keypoints_hand.png.
- [15] 高木政徳, 渡辺裕, "姿勢情報を用いた自転車選手識別手法", 早稲田大学理工学術院 基幹理工学研究科情報理工・情報通信専攻 2018 年度修士論文.
- [16] "scikit-learn," [Online]. Available: <https://scikit-learn.org/stable/>.
- [17] 小林大起, 石川孝明, 渡辺裕, "姿勢推定と機械学習を利用した日本指文字の分類", 電子情報通信学会 (IEICE) 2019 (2019 年 3 月発表予定).
- [18] 丸山瑞己, 奥村司, 松岡遼, 井上勝文, 吉岡理文, "OpenPose を用いた日本語の指文字における静止文字認識", 第 5 回サイレント音声認識ワークショップ, B-11, 29-30th Sep. 2018.

図一覧

図 2.1	SVM によるサンプル群分離模式図	4
図 2.2	CNN の概要図	4
図 2.3	OpenPose のネットワーク構成 [9]	5
図 3.1	カナ指文字の一覧	7
図 3.2	提案手法概要図	8
図 3.3	OpenPose における手指モデルの関節番号 [14]	9
図 3.4	静的指文字の姿勢検出フロー	10
図 3.5	関節角度計算	12
図 3.6	各関節のインデックス	13
図 3.7	訓練集合の構造（静的指文字）	13
図 3.8	姿勢推定失敗例	14
図 3.9	動的指文字の動作検出および識別フロー	16
図 3.10	移動判定閾値取得フロー	18
図 3.11	訓練集合の構造（動的指文字基本動作）	19
図 3.12	移動方向変化に伴う移動差分の角度特性	20
図 3.13	親指先端と人差し指先端接近による動作検知	21
図 4.1	静的指文字データセットの一例	26
図 4.2	動的指文字データセットの一例（基本動作）	28
図 4.3	動的指文字データセットの一例（特殊動作）	29
図 4.4	全静的指文字の姿勢推定結果一例	32
図 4.5	静的指文字の文字別分類精度 [17]	33
図 4.6	スケール変化時における提案手法の文字別分類精度	34
図 4.7	スケール変化時における座標値ベース手法の文字別分類精度	35
図 4.8	動的指文字各基本動作の姿勢推定結果一例	36
図 4.9	動的指文字の文字別分類精度	37
図 4.10	骨格類似文字群の例	38

図 4.11	OpenPose の誤検出による推定失敗文字例	39
図 5.1	多段 SVM 分類モード	42
図 5.2	各サブクラスの構成	42
図 5.3	手掌方向基準ベクトルおよび回転処理	43
図 5.4	掌の向きによる 4 パターン定義	44
図 6.1	拡張手法によるサブクラスごとの平均分類精度	47
図 6.2	サブクラス 1 における提案手法精度 (文字別)	48
図 6.3	サブクラス 2 における提案手法精度 (文字別)	49
図 6.4	サブクラス 3 における提案手法精度 (文字別)	50
図 6.5	サブクラス 4 における提案手法精度 (文字別)	51
図 6.6	サブクラス 5 における提案手法精度 (文字別)	52
図 6.7	サブクラス 7 における提案手法精度 (文字別)	53
図 6.8	保持角度に曖昧性が認められる指文字群	54
図 6.9	指文字「す」「な」「ま」の相違点	56

表一覧

表 3.1	カナ指文字のラベル対応	10
表 3.2	動的文字の区分と対応ラベル	16
表 4.1	実験環境	22
表 4.2	データ区分	23
表 4.3	OpenPose による手指検出特性評価	23
表 4.4	各静的指文字のデータ数	24
表 4.5	各動的指文字のデータ数	27
表 4.6	混同行列	30
表 4.7	使用 SVM の諸元	30
表 4.8	角度特徴量の適用性確認	31
表 4.9	静的指文字データセット全体における提案手法精度	32
表 4.10	スケール変化時における各手法の分類精度	34
表 4.11	動的指文字データセットにおける提案手法精度	36
表 6.1	拡張手法のサブクラス分類精度	46
表 6.2	サブクラス 1 における提案手法精度	47
表 6.3	サブクラス 2 における提案手法精度	48
表 6.4	サブクラス 3 における提案手法精度	49
表 6.5	サブクラス 4 における提案手法精度	50
表 6.6	サブクラス 5 における提案手法精度	51
表 6.7	サブクラス 7 における提案手法精度	52
表 6.8	サブクラス内分類実験結果のまとめ	55

研究業績

- [1] 小林, 石川, 渡辺: “V-SLAM とマーカ AR を併用したペン型デバイスによる空中描画システム”, 映像情報メディア学会年次大会, 34B-3, Sep. 2017.
- [2] 小林, 石川, 渡辺: ” 骨格推定と機械学習を利用した日本指文字の分類”, 電子情報通信学会総合大会, Mar. 2019 (発表予定).