

# カラー手袋を装着した実手話画像認識に向けた 動作者差異の類似度分析

## Similarity Analysis of Motion Difference for Sign Language Recognition using Colored Gloves

柴田 紘俊†  
Hirotooshi Shibata

西村 広光†  
Hiromitsu Nishimura

田中 博†  
Hiroshi Tanaka

小林大輔‡  
Daisuke Kobayashi

岩本典夫‡  
Michio Iwamoto

加藤秀司‡  
Shuji Kato

### 1. まえがき

聴覚に障がいを持った方と障がいを持たない方とのコミュニケーション手段として手話や指文字が一般的であるが、健聴者や事故などで後天的に障がいを持った方は習得していないことが多い。そのため、両者の意思疎通には手話通訳者やコミュニケーション支援機器が必要となり、聴覚障がい者と健聴者の間の対面コミュニケーションには未だ大きな障壁がある。手話を自動で認識する研究が数多く行われているが、実用化のためには、多くの課題が残っている[1][2][3][4]。

これまでに筆者らは、指と両手の識別や位置、動きが手話認識において重要な要素であると考え、利用者に多少の負荷をかけるものの、各指と両手の識別や動きを認識できる可能性のあるカラー手袋を用いた手話認識手法について検討してきた<sup>[5]</sup>。しかし、手話を普段から利用されている方のデータでの検討がこれまでできていなかったため、実際の手話利用者のデータでの検討が必要であった。

近年、手話の学習教材としては動画データをデータベース化したものが公開されている<sup>[6]</sup>。本研究では、動画データのデータベースをさらに発展させ、手話動作者の実際の手話動作を認識可能にして、フィードバックがある学習教材を目指して開発を進めている。そのため、普段のコミュニケーションで手話を利用されている方のデータを利用し、実用的な手話認識システムの実現に向けて検討を進めている。

本稿では、これまで明らかにしてきたカラー手袋と可視光カメラを用いた手話認識手法を実手話に適用することを目標として、個人の手話動作の差異の存在を分析し、個人の動作差異があっても高精度な認識を実現するため、認識辞書の拡充方法について検討を行った結果を述べる。

### 2. カラー手袋を用いた手話認識手法

#### 2.1. 手話認識の必要要素とカラー手袋の利用

本研究では、手話の手指動作に着目した。手指動作とは、手や指、腕による動作であり、非手指動作とは、顔の表情や口形、眉の上下や寄せ、ほおやあごの動きなどである。

筆者らは、手話認識実現の技術課題要素として下記の(1)~(7)を定めて順次技術開発を進め、(3)までの基本技術の検討を行ってきた。

† 神奈川工科大学大学院 情報工学専攻

‡ 株式会社ケイ・シー・シー

- (1) 手指の形状認識
- (2) 各指, 両手の識別
- (3) 腕, 手, 指の動きの認識
- (4) 手の表裏判別
- (5) 手と顔の相対位置
- (6) 顔の部位の動き認識
- (7) 表情認識

手話単語の中には、手指の形状や腕の動作が類似した単語が数多く存在する。そのため類似した手指形状や腕動作の手話単語同士を区別するためには上記で述べた課題(2)の各指, 両手の識別が特に重要であると考えられる。

筆者らは、上記(1)~(4)の課題を解決する方法として、カラー手袋の利用を提案した。カラー手袋を利用することにより下記の方法でそれぞれの課題を解決することが可能と考えている。

- (1) 手指の形状認識→検出結果の色の種類と位置から推測
- (2) 各指, 両手の識別→それぞれに異なる色を彩色し、検出した色の種類から識別
- (3) 腕, 手, 指の動き認識→検出した色領域の重心位置を追跡することにより平面内の動きを検出し、検出した色の面積変化から奥行き方向の動きや回転を推測
- (4) 手の表裏判別→手の平に色を彩色し、その色が検出されているか否かで判別

使用するカラー手袋を図 1 に示す。手話認識実現に必要な色数を検討した結果、各指を識別するために各指に 5 色、手の表裏を判別するために 1 色、左右の手首を識別するために手首に 2 色の計 8 色を検出することとした。また、彩色部分の面積は極力大きく、かつ、指を折り曲げた際に色



図 1 本認識手法で用いるカラー手袋

が見えなくなる大きさとした。カラー手袋に彩色する色は入手可能であった 30 種類の色手袋を調べ、色相が最も離れている 8 色を選定した。

## 2.2. 認識手法の基本方針

本認識手法の流れを図 2 に示す。最初にカメラから手話動作の画像を取得し、カラー手袋に彩色した各色を検出する。

本研究では色検出後、手指形状や腕の動きが類似した単語同士を区別するために、手指の形状と手首の動きの 2 つの情報から手話の認識を行うこととした。手指形状は認識開始フレーム、中間フレーム、終了フレームの 3 つの静止画を取り出し、それぞれの指を示す色情報の位置関係によって類似性を求めることとした。

腕の動きと手首の動きは連動していることから、本検討では腕の動きを手首の動きとみなした。最終的に手指形状認識結果と手首の動き認識結果を統合することで手話認識を行った。

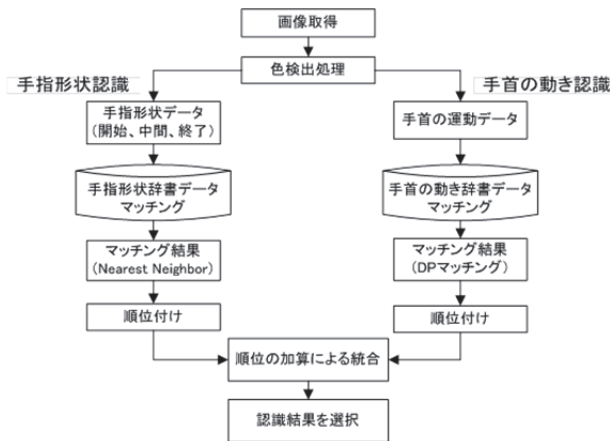


図 2 手話認識手法の流れ

## 2.3. 手指形状認識手法

本提案では、取得画像からの各色の検出結果から手指形状を推測する。具体的には、各指先の彩色部分の重心位置を  $(fx_i, fy_i)$ 、手首の彩色部分の重心位置を  $(wx, wy)$ 、手首の彩色部分の面積を  $S$  としたとき、特徴ベクトル  $\mathbf{d}$  の各要素を (1) 式により求めることとした。

$$d_i = \sqrt{\frac{(fx_i - wx)^2 + (fy_i - wy)^2}{S}} \quad (1)$$

ここで、手首と各指先間の距離はカメラとユーザとの距離に依存するため、手首に彩色した色の面積で正規化を行う。その後、生成した手指形状データと辞書登録した手指形状データの類似性の距離を Nearest Neighbor 法で求めた。

## 2.4. 手首の動き認識手法

はじめに、手首の動きは動作開始から終了までの手首位置の色領域の重心移動を毎秒 30 フレームで求める。

次に、算出した重心位置を追跡することによって手首の運動データを生成する。手話には手首を動かさない単語が存在するため、本検討では、手首に彩色した色領域の重心位置の変化量が 3pixel 以下の場合、重心位置が変化していないものとして扱うこととした。

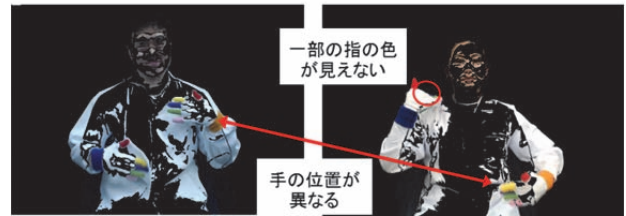
その後、手指形状を認識する場合と同様、生成した手首運動データと辞書データをマッチングし、両者の距離を求める。マッチング手法には動作認識の代表的手法である DP マッチングを用いることとした。

## 3. 複数の辞書セットによる認識

手話習得者でも個人々人により手話動作の差異が存在する。そのため同一単語の手話動作でも、手話動作者が異なると、単一の辞書登録パターン集合(以下辞書セットと記す)では同一と解釈できない場合があるという課題が存在する。

図 3 に実手話動作の個人差の例として「うれしい」という手話単語の動作開始時を示す。図 3 は左右で異なる手話動作者であり、手話動作画像に背景差分処理を施した結果となっている。両腕の位置が左右の画像で逆になっていることが分かる。また、手指の形状も左の画像では全ての手の指先に彩色された色が見えるが、右の画像では右手の赤以外の色が見えない手指形状となっている。

そこで、同じ表現の動作であっても個人々人により動きに差異が大きい手話動作を高精度に認識するためには、これまで単一であった辞書セットを異なる被験者から複数作成し、辞書セットを拡充させる必要があると考えた。



単語:「うれしい」動作開始時

図 3 手話動作の個人差

### 3.1. 辞書セットの拡充方法の検討

辞書セットを充足することは、被験者や時間を変えてデータを採取することで行うことであるが、被験者や時間を変えて採取した辞書セットをどのように取捨選択して辞書に利用すれば良いかを検討する必要がある。

そこで本稿では、辞書登録用に分離した手話データ集合のみを利用して、各被験者の辞書セットのみでの認識性能から各辞書セットの評価点を算出し、その結果に基づいて辞書セット取捨選択する方法を提案する。この辞書セットの評価点については、次節で詳説する。

### 3.2. 辞書セットの評価点算出方法

2.2 節で提案した手話認識手法を用いて、同一辞書を除いた全ての組み合わせで認識を行い、正解単語数の和をその辞書の評価点とする。評価点の高い辞書セットほど良い辞書セットとし、この評価点は認識システムの構築時に辞

書セットを取捨選択して組み合わせる際の優先順位決定に利用することとした。

### 3.3. 辞書セットの評価点算出結果

4章で後述する撮影の結果から作成した4つの辞書セットの評価点を3.2節の手法に基づいて算出した結果を表1に示す。

辞書セットB>辞書セットA>辞書セットC>辞書セットDの順に評価点が高い結果となった。この結果からまず評価点が相対的に高かった辞書セットAと辞書セットBを組み合わせ、評価用データセットの認識を行った場合に高い認識性能を実現できるか調査することとした。

さらに辞書セットAとBに次いで評価点が相対的に高かった辞書セットCも加えて、3つの辞書セットを組み合わせた場合にどのような傾向となるかも合わせて調査した。

表1 辞書セットの評価点算出結果

	辞書 セットA	辞書 セットB	辞書 セットC	辞書 セットD	評価点 (各行の和)
辞書セットA		9	6	4	19
辞書セットB	9		7	6	22
辞書セットC	6	7		3	16
辞書セットD	4	6	3		13

### 3.4. 辞書拡充方法

複数の辞書セットを用いた辞書拡充方法を(2)式に示す。ここで  $D_{(1,i)}$ ,  $D_{(2,i)}$ ,  $P_{(1,i)}$  はそれぞれ辞書セット1, 辞書セット2, 評価用データセットの  $i$  番目の単語を示す。

$dis$  は辞書パターンと評価用データパターンの距離を求める関数である。 $dis(D_{(1,i)}, P_{(1,i)})$  と  $dis(D_{(2,i)}, P_{(1,i)})$  で辞書セット1, 2と評価用データセットの  $i$  番目の単語についての距離をそれぞれ求める。距離は2.3節, 2.4節で述べた方法で算出する。

最小値を求める関数である  $\min$  で先ほど求めた距離の最小値を求める。これを満たす全ての単語  $i$  を求めることで、各辞書セットと評価用データセットとの距離が小さいもののみを取り出す辞書の拡充を行う。

$$\min(dis(D_{(1,i)}, P_{(1,i)}), dis(D_{(2,i)}, P_{(1,i)})) \quad (2)$$

## 4. 認識実験

辞書拡充の効果を確認するために、4人の被験者の評価用手話データを用いて実手話の認識実験を行った。

### 4.1. 認識対象とした手話単語

手話による初対面の挨拶シーンを想定した文書を作成した。それを手話習得者の方に、自然な手話になるよう単語ごとに切り分けてもらった。その結果の13単語を本実験の認識対象とした。認識対象とした手話単語の一覧を表2に示す。

表2 認識対象とした手話単語

OK	あなた	いいえ	うれしい
こんにちは	できる	はい	よろしく
一緒	会う	手話	否定
勉強			

### 4.2. 実験環境と条件

本実験の実験環境を図4に示す。手話の撮影にはウェブカメラを用い、カメラと被験者との距離は1mとした。1mという距離は人同士が対話をする自然な距離を考慮し設定した。被験者の方には椅子に座って手話を実演してもらい、その際、手を体の横に下ろしても手首に彩色された色が取得画像に映らないようにカメラの高さを調節した。被験者には白衣と白色のシャツを着てもらい、背景には暗幕を設置することで実験とは無関係の色が映らないようにした。

表2の単語を4名の手話習得者にそれぞれ2回ずつ行ってもらい、1回目の撮影データを評価用データセット、2回目の撮影データを辞書用データセットとした。

また、今回の実験では手話動作開始時と終了時のフレームは目視で判定することにより認識を行った。

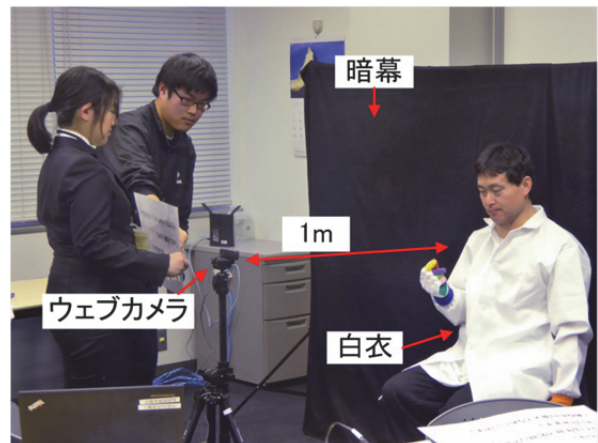


図4 実験環境

## 5. 実験結果

### 5.1. 辞書充足の効果の確認

評価用データセットA, B, C, Dを同一人物から作成した辞書も含めた全ての辞書セットの組み合わせで認識した認識率を表3, 表4, 表5, 表6に示す。また、その認識率を平均した結果を表7に示す。

表7の一番下の行は辞書セット数ごとの認識率の平均を示しており、辞書セット数1の場合より辞書セットを複数組み合わせたとときのほうが高い認識率となっていることが分かる。

この結果から、辞書の拡充は安定して認識率を向上させる効果があることが分かった。



表3 評価用データセットAの認識結果

辞書セット数1	辞書セット数2		辞書セット数3		辞書セット数4		
A	76.9%	A+B	69.2%	A+B+C	69.2%	A+B+C+D	53.8%
B	53.8%	A+C	69.2%	A+B+D	61.5%		
C	53.8%	A+D	69.2%	A+C+D	61.5%		
D	23.1%	B+C	53.8%	B+C+D	23.1%		
		B+D	30.8%				
		C+D	15.4%				

表4 評価用データセットBの認識結果

辞書セット数1	辞書セット数2		辞書セット数3		辞書セット数4		
A	30.8%	A+B	76.9%	A+B+C	61.5%	A+B+C+D	46.2%
B	69.2%	A+C	38.5%	A+B+D	46.2%		
C	38.5%	A+D	15.4%	A+C+D	30.8%		
D	30.8%	B+C	61.5%	B+C+D	46.2%		
		B+D	46.2%				
		C+D	38.5%				

表5 評価用データセットCの認識結果

辞書セット数1	辞書セット数2		辞書セット数3		辞書セット数4		
A	46.2%	A+B	46.2%	A+B+C	38.5%	A+B+C+D	38.5%
B	30.8%	A+C	30.8%	A+B+D	30.8%		
C	30.8%	A+D	30.8%	A+C+D	30.8%		
D	15.4%	B+C	30.8%	B+C+D	30.8%		
		B+D	23.1%				
		C+D	38.5%				

表6 評価用データセットDの認識結果

辞書セット数1	辞書セット数2		辞書セット数3		辞書セット数4		
A	23.1%	A+B	46.2%	A+B+C	46.2%	A+B+C+D	53.8%
B	46.2%	A+C	46.2%	A+B+D	46.2%		
C	15.4%	A+D	46.2%	A+C+D	46.2%		
D	38.5%	B+C	38.5%	B+C+D	53.8%		
		B+D	38.5%				
		C+D	53.8%				

表7 認識率の平均

辞書セット数1	辞書セット数2		辞書セット数3		辞書セット数4		
A	44.2%	A+B	59.6%	A+B+C	53.8%	A+B+C+D	48.1%
B	50.0%	A+C	46.2%	A+B+D	46.2%		
C	34.6%	A+D	40.4%	A+C+D	42.3%		
D	26.9%	B+C	38.5%	B+C+D	38.5%		
		B+D	34.6%				
		C+D	36.5%				
	38.9%	42.6%	45.2%				48.1%

## 5.2. 辞書セット取捨選択の効果の確認

表7の網掛け部分は、3.3節で得られた評価点が相対的に高い辞書セットを組み合わせた結果となっており、辞書セットA+Bの平均認識率59.6%は、辞書セット数2の組み合わせの中では最も高い認識率となっている。

同様に辞書セットA+B+Cの平均認識率53.8%は、辞書セット数3の組み合わせの中で最も高い認識率となっている。

この結果から、相対的に高い評価点の辞書セットを優先的に組み合わせて辞書拡充を行うことで、ランダムに組み合わせた辞書拡充よりも高い認識性能を実現することができた。

しかし、今回の実験結果では辞書セット数2のときが最も高い認識率となっており、辞書セットをそれ以上増加させるよりも高い認識性能であった。このことから、辞書セット拡充の優先度については提案手法で一つの指針を示せたものの、辞書セット拡充をいくつまで行うかという点について、更なる検討が必要である。

## 6. まとめ

実際の手話データを採取し、手話動作の分析を行い、個人による動きの差異が大きいことが分かった。そのため、辞書セットを複数人のデータによって拡充することで認識性能の高精度化を行うことを検討した。

具体的には、採取時間と被験者の異なる辞書作成用データセットを集め、どのように取捨選択して組み合わせることで辞書を構成するかについて検討を行った。

本稿では、辞書用に採取したデータセット同士を、異なる被験者間の全ての組み合わせで相互に認識を行い、その認識精度によって辞書セットの評価点を算出する手法を考案した。

考案した辞書セット評価点算出法に基づき、辞書セットを組み合わせて評価用データでの認識実験を行った結果、評価点が相対的に高かった辞書セットの組み合わせの認識率は他の組み合わせの認識率より高くなることを確認した。

しかし、今回の実験の範囲では辞書セット数2の場合が最も高い認識率となり、辞書セットをそれ以上増加させるよりも高い認識性能であった。

今後は、今回の被験者以外の協力を得ることでさらなるデータ採取を行い、本認識手法における辞書セットの有効性の確認と辞書セット拡充の指針検討を深めていく。

## 7. 謝辞

本研究に必要な手話データの採取に当たり、写心家・手話スタイリストの小林澄枝様、神奈川工科大学手話サークルのメンバーに協力戴いた。ここに感謝致します。

## 文 献

- [1] 山田 寛, 松尾 直志, 島田 伸敬, 白井 良明: 手話認識のための見えの学習による手領域検出と形状識別; 画像の認識・理解シンポジウム, IS1-37, pp.635-642 (2009).
- [2] 吉野 和芳, 川嶋 稔夫, 青木 由直: カラー手袋を用いた手話認識手法; 情報処理学会研究報告, コンピュータビジョン研究会報告 95(68), pp.51-58 (1995).
- [3] 西村 洋介, 今村 大輔, 堀内 靖雄, 川本 一彦, 篠崎 隆宏, 黒岩 眞悟: Kinect とパーティクルフィルタを用いた HMM 手話認識手法の検討; 信学技報, PRMU2011-223, pp.161-166 (2012).
- [4] 山根 卓也, 船富 卓哉, 飯山 将見, 美濃 導彦: データグループのセンサデータに基づく各指節の位置・向き の推定手法の検討; 信学技報, PRMU2011-138, pp.72-83 (2011).
- [5] Sugaya, T., Tsuchiya, H., Iwasawa, H., Nishimura, H., Tanaka, H.: Fundamental study on sign language recognition using color detection with an optical camera ICIP, pp. 8-13, Mandarin Hotel, Bangkok, Thailand.
- [6] 株式会社ケイ・シー・シー, "SmartDeaf", <http://www.smartdeaf.com/>