

[研究論文]

# カラー手袋の色検出による 手話認識とその評価

菅谷隆浩<sup>1</sup>・土屋裕文<sup>2</sup>・岩沢宏樹<sup>2</sup>・西村広光<sup>3</sup>・田中博<sup>3</sup>

1 博士前期課程情報工学専攻

2 情報学部 情報工学科

3 情報工学専攻

## Sign Language Recognition by Color Detection for Colored Glove and its Evaluation

Takahiro SUGAYA<sup>1</sup>, Hirofumi TSUCHIYA<sup>2</sup>, Hiroki IWASAWA<sup>2</sup>,  
Hiromitsu NISHIMURA<sup>3</sup>, Hiroshi TANAKA<sup>3</sup>

### Abstract

This paper presents the feasibility study for the recognition of sign language. The main feature of our proposed method is to use colored gloves in which several areas are assigned by different colors. This makes it possible to discriminate each finger and to recognize hand shape by color detection result. In addition, the hand motion can be detected by the variation of the position of colored region of the wrist. The recognition method which consists of hand shape recognition by the color detection results and hand motion recognition by DP matching was implemented in the PC, and the experiment was carried out by using the optical camera. The recognition success ratio of 83.3% for the basic 24 words of sign language was obtained, and the feasibility of the proposed method was verified by the experiments.

Keywords: Sign language recognition, Color glove, Color detection, DP matching, Optical camera

### 1. まえがき

聴覚に障害を持った方と障害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害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、聴覚障害者と健聴者の間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状況にあると思われる。

手話認識の代表的な手法として、RGBカメラを用いて手の肌色部を検出し、手指の形状を認識する研究が行われている<sup>1)2)</sup>。この手法はRGBカメラのみを用いるため、利用者に身体的な負荷を与えないという利点がある一方、手以外の肌色領域との区別や各指の識別、それらの動きの認識が難しいという問題がある。

この問題の解決手法として、指や手の甲などに異なる色を彩色、または複数の色パッチを取り付けたカラ

ー手袋を用いて手指形状を認識する手法が提案されている<sup>3)4)</sup>。異なる色を彩色することにより、他の肌色領域との区別や各指の識別が容易となる。また、検出した色を追跡することにより手や指の動きを認識できる可能性もある。しかしこの手法は、認識環境を変え、照明条件が変化した場合に安定した色検出が難しいという大きな問題がある。

その後、マイクロソフト社が開発したKinectの登場によって腕の動作認識を容易に実現させることが可能となった<sup>5)6)</sup>。この手法は、距離情報を用いるため照明条件の変化に影響を受けにくく、利用者にも負荷をかけないという利点も多いが、指の識別やそれらの動きを検出することが難しいという欠点がある。また、太陽光が当たる環境下での性能確保は難しい。

その他にも、データグローブを利用し、指の位置や向きを推定する手法も検討されている<sup>7)</sup>。これは、グローブに取り付けた各種センサデータを利用するが、センサを内蔵した手袋を装着するため利用者への負荷が大きい。

このように多くの手法が提案されている中、筆者らは指の識別や位置、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することとした。

先程述べたとおり、カラー手袋を用いた手話認識手法には、照明条件が変化する環境下では色検出精度が低下するという大きな問題があり、そのため、まずは照明変化にロバストな色検出手法を検討する必要がある。筆者らは、この問題の解決手法としてクラスタリング手法を用いた色検出閾値自動決定手法を提案した<sup>8)</sup>。

本報告では、これまで明らかにしてきた静的な手指形状認識技術を発展させ、動的な動きを含む手話認識への可能性を検討する。まず、手話動作開始時と終了時の手指形状と手首に彩色した色領域の重心位置変化から動きの認識を取り入れ、手話を認識するための手法を提案する。さらに、提案手法を用いて全国手話検定試験5級の基本単語の一部である24単語に対して認識実験を行った結果について報告する。

## 2. カラー手袋を用いた手話認識の提案

### 2.1 手話認識の必要要素とカラー手袋の利用

手話は、手指動作と非手指動作から構成される<sup>9)</sup>。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、ほおやおごの動きなどである。筆者らは、手話認識実現のためには以下の課題が存在し、これらの課題を解決していくことによって、手話の自動認識の実現に近づくと考えている。

- (1) 手指の形状認識
- (2) 各指、両手の識別
- (3) 腕、手、指の動き認識
- (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

現在、全国手話研修センターが実施している全国手話検定試験では、全検定級である5級～1級で約8,500単語の基本単語が設定されている<sup>10)</sup>。この中には、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2.1節で記述した課題の一つである指の識別や動き、手の表裏判別が特に重要になると考えられる。

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する。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 Fig.1 に示す。手話認識実現に向け、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手のひらに1色、左右の手首を識別するために手首に2色の計8色を検出する色数とすることとした。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が見えなくなる大きさとした。

カラー手袋を用いることで全ての課題解決は困難であるが、RGBカメラを利用した表情認識などの研究が既に検討されている<sup>11)</sup>ことから、この研究成果を今後取り込むことによって、より手話認識実現に向けた課題解決が期待できると考えている。



Fig. 1 Color glove to be used.

### 2.2 本研究の最終的な利用シーンの一例

本研究では、センサとしてRGBカメラのみを用いていることが特徴である。そのため、Webカメラやカメラ付きHMD、スマートフォンなどでの手話認識の可能性はある。筆者らの目標は、近年急速に普及し、既に汎用的となったスマートフォンを利用した手話認識システムの実現である。

本研究の最終的な利用シーンの一例を Fig.2 に示す。手話を習得しているが言葉を話せない人が手話を表現し、言葉は話せるが手話を習得していない人がスマートフォンに内蔵されたカメラで相手を撮影する。スマートフォンの内部で色検出を含めた、全ての認識処理を行い、認識した結果を画面上に文字として出力する。

手話未習得者が言葉を用いず、コミュニケーションをとる場合、筆談が多く用いられるが、筆談でのコミュニケーションは1対1に限定される。また、コミュニケーションに時間を要する。本システムを利用することにより、スムーズな意思伝達や複雑な内容でも短い時間でリアルタイムに自分の意思を相手に伝えることが可能となる。また、スマートフォンによる音声合成を用いることによって話すことと同等のことができる可能性や、多くの人に一斉に伝えることも可能となる。



Fig. 2 Final usage scenery of this study.

### 3. 色検出手法の提案と課題

#### 3.1 検出する色の選定

カラー手袋に彩色する色を選定する。本研究では色検出に使用する成分として色相を使用する。筆者らは、入手可能であった 30 種の色手袋の中から色相が最も離れている 8 色を選定した。選定した 8 色の色分布を Fig.3 に示す。

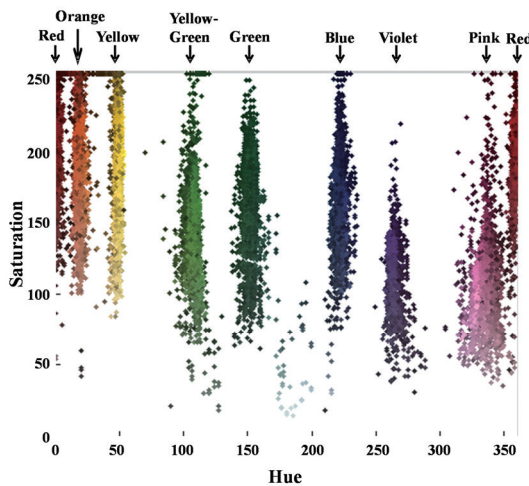


Fig. 3 Color distribution of 8 colors.

#### 3.2 色検出閾値の決定

カラー手袋を用いた手話認識では各色の検出精度が手話の認識精度に大きな影響を与える。しかし、各色の色相は照明条件の変化に大きく依存する。

筆者らは、手話を認識するに先立って色検出のための閾値を自動決定するキャリブレーションを認識処理開始前の手順として導入することとした。具体的には、まず、背景の影響が色検出に影響を与える現象を防ぐため、取得した画像に対して、背景差分処理を適用する。取得画像と背景差分適用後の画像を Fig.4 に示す。さらに、背景差分処理では除去できないカラー手袋の白色部分を除去するため、白色が低彩度であるという特徴を利用し、背景差分適用後の画像に対して、低彩度の画素を除去する。

次に、低彩度除去後の画像から色相のヒストグラムを生成し、クラスタリング手法により各色の色相の重心位置を得る。ここで、本提案では色相の出現頻度の差異がクラスタリングに与える影響を少なくするため、

色相の出現頻度の最大値を、全画素数 / 色相範囲の 2 倍に制限した。さらに、色相ヒストグラムの最大値が、全画素数 / 色相範囲の半分に満たない場合はノイズとして除去した。

以上の前処理を適用した後、対象データから各クラスターの重心位置を取得する。重心の取得には、クラスタリングの代表的手法である k-means クラスタリングを適用する。適用した k-means クラスタリングの処理ステップを以下に示す、

- (1) 各クラスターの重心位置を初期値として事前に設定する。(オレンジ:30, 黄色:60, 黄緑: 120, 緑:180, 青:220, 紫:260, ピンク:330, 赤:350)
- (2) 各画素の色相値と各クラスターの重心位置の距離を計算する。
- (3) 各画素に対して、最も距離が近い重心位置のクラスターを割り振る
- (4) 各クラスターの重心位置を計算する。
- (5) (2)~(4)の処理を、重心位置が収束するまで繰り返す。

求めた重心位置を基準として事前に明らかにした各色の色相範囲(照明条件に依存せずほぼ一定値)から、色検出閾値を決定する。なお、本手法は認識処理中に照明条件が変化しないことを前提として、認識処理開始前に一度だけ行うこととしている。

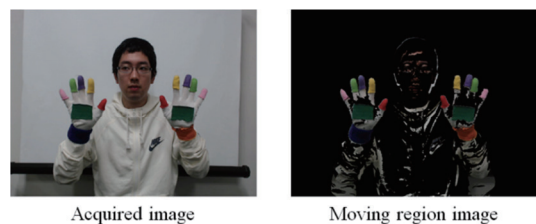


Fig. 4 Moving region image after background subtraction.

#### 3.3 色検出精度評価実験

色検出閾値決定手法の有効性を確かめるため照明条件を変化させ、カラー手袋に彩色した 8 色の検出精度を評価した。各実験環境での取得画像を Fig.5 に示す。ここで、Fig.5 中の照度は全て被写体側の照度である。

130Lux, 230Lux, 340Lux, 430Lux の環境は、カメラ側の照度を 180Lux で固定させ、被写体側の照度は、ライトの光を反射させる、または、被験者の上部をダンボールで覆うことにより変化させた。10Lux の環境は、太陽光が直接当たらない室内において、被験者直上付近の照明のみを消した環境であり、1500Lux の環境は、照明は消えているものの太陽光の影響を強く受ける窓際の実験環境である。

評価対象とした手指形状は、色を全て見せた手指形状(パターン A)と色を全て隠した手指形状(パターン B)の 2 パターンとした。例外として、通常手話を行う際、手首は必ず見えるため、手首に彩色した色は常に見えている状態で実験を行った。評価対象とした手指形状を Fig.6 に示す。



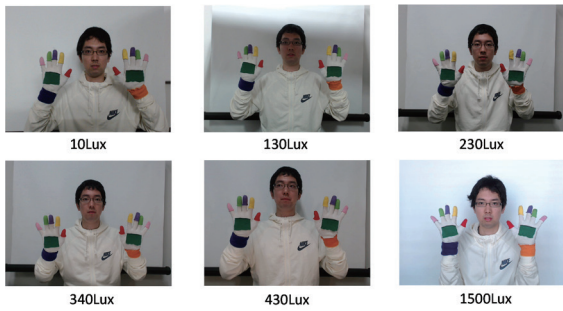


Fig. 5 Acquired image by each experimental environment.

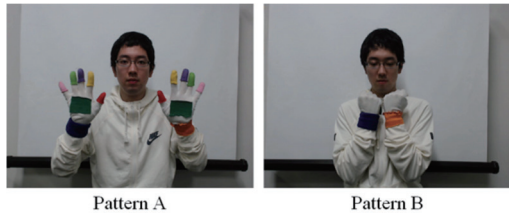


Fig. 6 Hand shape used in experiment.

実験は、提案手法を用いず 230Lux, 340Lux, 430Lux での各色の色相範囲の和集合をとり、色検出閾値を決定した場合、提案手法を用いて色検出閾値を各環境で決定した場合で評価した。提案手法を用いない場合の色検出閾値を Fig.7 に、提案手法を用いた場合の色検出閾値の一例 (130Lux, 1500Lux) を Fig.8 に示す。照明条件の違いによって色検出閾値が変化していることが確認できる。

評価では、色検出後の画像を目視で確認し、色検出の失敗判定を行った。具体的には、カラー手袋の彩色部分が見えている状態でカラー手袋の彩色部分以外の領域に色が検出された場合を色の誤検出とし、色が検出されない場合を色の未検出とした。また、カラー手袋の色が見えていない状態でカラー手袋の彩色部分以外の領域に色が検出された場合を色の過剰検出とした。

色検出精度評価実験結果を Table 1 に示す。評価結果から提案手法を用いず、色検出閾値を固定させた場合の平均色検出成功率は 56.1%となることを確認した。特に、室内全体を明るくさせた被写体側 1500Lux の環境下においてピンク色が未検出となり、低い色検出成功率となった。また、手指形状パターン B において、被験者の背後に発生した影にカラー手袋に彩色した緑色が過剰検出された。これは、色相範囲を広く設定したことによる影響だと考えられる。

一方、提案手法である色検出閾値自動決定手法を用い、各環境で色検出閾値を決定した場合の平均色検出成功率は 96.5%となることを確認した。環境に応じて最適な色相の重心を決定することにより、前手法で検出できなかった 1500Lux 環境下でのピンクが正常に検出できることを確認した。さらに、各色の色相範囲を可能な限り狭めることにより被験者の背後に発生した影に緑色が過剰検出される現象を抑制できた。この結果から、色検出閾値自動決定手法を用いることで環境光の変化にロバストな色検出が行えることを確認した。

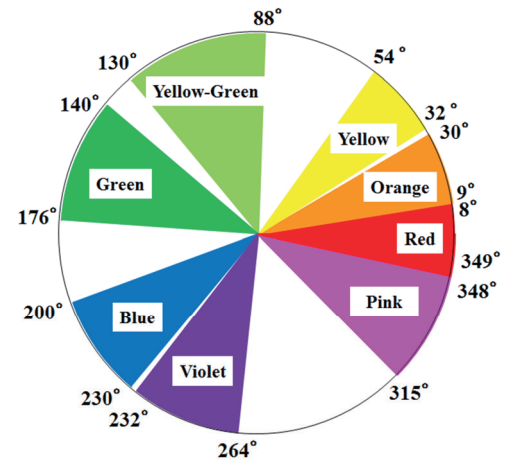


Fig. 7 Hue value threshold pattern for color detection -without proposed method-.

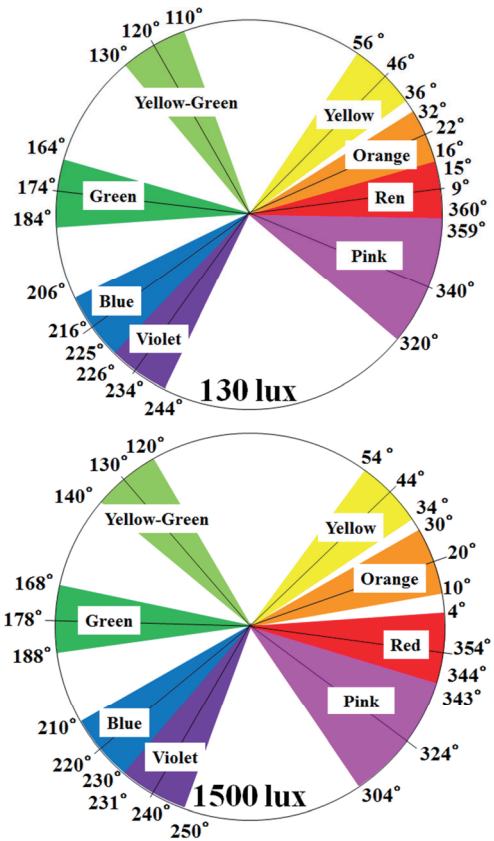


Fig. 8 Example of hue value threshold pattern for color detection -with proposed method- .

Table 1 Experimental results for color detection

Illumination(lux)	Cam	Glove	Total frames	Hand shape Pattern	Success frames	
					without proposed method	with proposed method
10	10	10	100	Pattern A	49	100
				Pattern B	22	90
180	130	100	100	Pattern A	100	100
				Pattern B	66	100
180	230	100	100	Pattern A	100	100
				Pattern B	10	100
180	340	100	100	Pattern A	100	100
				Pattern B	11	86
180	430	100	100	Pattern A	100	100
				Pattern B	16	86
200	1500	100	100	Pattern A	0	97
				Pattern B	99	99

#### 4. 手話認識手法の提案

##### 4.1 認識手法の基本方針

本研究で提案する手話認識手法を Fig.9 に示す. まず, カメラから画像を取得し, 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 つの観点から手話認識を試みる. 手指形状は認識開始時と認識終了時でそれぞれ評価し, より高い認識精度を可能とする. また, 腕の動きと手首の動きは連動しているため, 本検討では腕の動きを手首の動きとみなすこととした.

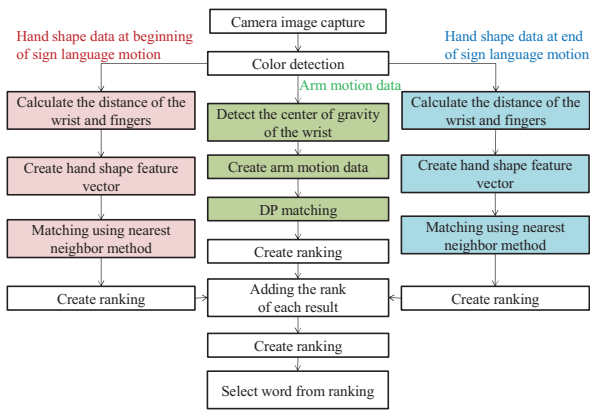


Fig. 9 Proposed sign language recognition method.

##### 4.2 手指形状認識手法

本提案では, 各色の検出結果から手指形状を推測する. 具体的には, 手首の中心から各指先までの彩色部分の重心間距離を式(1)で算出し, 5次元の特徴ベクトル  $\mathbf{d}$  を構成する.

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

$(fx_i, fy_i)$ : 各指先の彩色部分の重心位置

$(i = 1, 2, 3, 4, 5)$

$(wx, wy)$ : 手首の彩色部分の重心位置

$S$ : 手首の彩色部分の面積

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との類似性を Nearest Neighbor法で求める.

##### 4.3 手首の動き認識手法

本検討では, 4.1 節で述べた通り腕の動きを手首の動きとみなす. まず, 手首に彩色した色領域の重心位置を検出する. その後, 検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. 生成した手首の運動データの一部を Fig.10 に示す. Fig.10 では,

3 つの単語 (女, 弟, 月曜日) をカメラに対峙した平面内で動作した際の, 手首に彩色した色領域の重心位置の変化を示している. 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, 手首の色領域の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の手首運動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 マッチングを使用した<sup>12)</sup>. DP マッチングは, 動作時の速度や動作の大きさに影響されない手法であるが, 手話には動作の速さが意味を持つ単語が存在する. 今回は, 動作の速さが意味を持たない単語を認識することを前提とした.

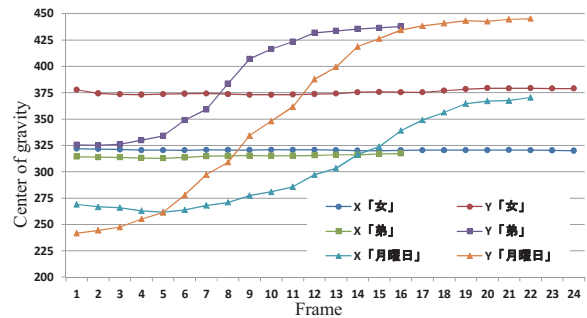


Fig. 10 Movement of the center of gravity of the detected colored areas of the wrist.

#### 5. 認識実験

##### 5.1 認識対象とした手話単語

提案した手法の手話認識への可能性を確認するための基本実験を行った. 本実験では, 全国手話研修センターが行う全国手話検定試験 5 級の基本単語 400 語から以下の条件で抽出した 24 単語を認識対象とした.

- (1) 右手のみで行うことができる単語
  - (2) 平面内での動きのみで行うことができる単語
- 認識対象とした手話単語の一覧を Table 2 に示す.

Table 2 Selected sign language words.

兄	姉	有る	妹
上A	上B	弟	男
女	～くらい	月曜日	下A
下B	少し	鈴[が鳴る]	茶B
次	土	どう?	場所
日	また	右	老人

##### 5.2 実験条件と環境

今回は提案手法の可能性を検討するための基本実験であるため, テンプレート作成者と動作データ取得者は同一人物で行った. また, テンプレートデータと動作データはともに 3 つずつ取得した. 取得した動作データはすべてのテンプレートデータに対してマッチングを行い, 最も距離が小さい結果を採用することとした. 最も距離が小さい結果を採用する理由は, 色検

出失敗の影響がないと考えられる最善の解を選択することで、手法としての可能性を確認するためである。

本実験の実験環境を Fig.11 に示す。カメラは Web カメラを用いて、床からカメラまでの距離は 1.4m、カメラと被験者の距離は 1m とした。認識距離を 1m に設定することによって、手話動作時に手がカメラ画像の外にでることを防ぐ。また、1m は人同士が対話をする自然の距離である。照明条件は、カメラ側 200Lux、被験者側 300Lux の環境下で行った。被験者には黒い長袖の衣服を着用させ、背景には暗幕を設置し、手や腕を動かした際にノイズの影響を受けない環境にした。処理用機器には PC を用いて、フレームサイズは 800×600 であり、実験開始前にフレームレート 20fps とすることを確認した。

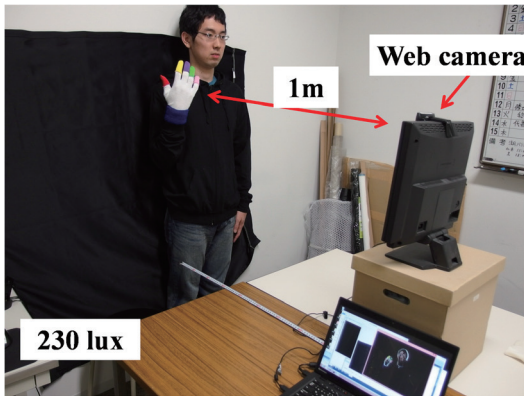


Fig. 11 Experimental environment for sign language recognition

5.3 実験結果

5.3.1 手指形状認識結果

手話動作開始時の手指形状の認識実験結果を Table 3 に、手話動作終了時の手指形状の認識実験結果を Table 4 に示す。実験は各単語につき 3 回、計 72 回試行し、Table 3, 4 は各単語において類似度が最も高くなった回数を示している。また、今回の実験では手話動作開始時と動作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。

実験結果から、動作開始時の手指形状認識正解率が 63.9%、動作終了時の手指形状認識正解率が 58.3% となることを確認した。

Table 3 Experimental results at beginning of sign language motion.

	Template																					
	兄	姉	有る	妹	上	弟	男	女	くら	月曜	下	少し	飾	茶	次	土	どう	場所	日	また	右	老人
兄	2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
姉	0	2	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ある	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
妹	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
上A	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
上B	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
弟	1	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
男	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
女	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
くら	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
月曜日	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
下A	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0
下B	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0
ずこし	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0
飾	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0
茶	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0
次	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0
土	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0
どう	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
場所	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
日	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
また	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
右	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
老人	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 4 Experimental results at end of sign language motion.

	Template																					
	兄	姉	有る	妹	上	弟	男	女	くら	月曜	下	少し	飾	茶	次	土	どう	場所	日	また	右	老人
兄	2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
姉	0	2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ある	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
妹	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
上A	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
上B	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
弟	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
男	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
女	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
くら	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
月曜日	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
下A	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0
下B	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0
ずこし	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0
飾	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
茶	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0
次	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0
土	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
どう	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
場所	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
日	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
また	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
右	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
老人	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

5.3.2 手首の動き認識結果

手首の動きの認識実験結果を Table 5 に示す。実験結果から、正解率 45.8% となることを確認した。今回選定した単語の多くは、手首の動きが類似していたために、正解率が低い結果となった。

Table 5 Experimental results for arm motion recognition.

	Template																					
	兄	姉	有る	妹	上	弟	男	女	くら	月曜	下	少し	飾	茶	次	土	どう	場所	日	また	右	老人
兄	1	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
姉	1	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ある	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
妹	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
上A	1	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
上B	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
弟	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
男	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
女	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0
くら	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
月曜日	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0
下A	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
下B	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ずこし	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
飾	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
茶	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
次	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
土	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
どう	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
場所	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
日	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
また	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
右	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
老人	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

5.3.3 累積正解率と認識結果の統合

手話動作開始時の累積正解率を Fig.12、手話動作終了時の累積正解率を Fig.13、手首の動きの累積正解率を Fig.14 に示す。図中では、3 位までの累積正解率を示している。ここで累積正解率とは、判定順位 N 位までに入っていれば、正解とみなすことにより認識性能を評価する手段である。

Fig.12, 13, 14 から、累積正解率が手話動作開始時の手指形状認識で 63.9% から 88.9% に、手話動作終了時の手指形状認識で 58.3% から 79.2% に、手首の動きの認識



の順位を加算し、その値が最も低い単語を最終的な手話認識結果として選択する。

$$arg_i \min = (Rank\_Beginning_i + Rank\_End_i + Rank\_Arm_i) \quad (2)$$

$i$ : 手話単語 ( $i = 1, 2, \dots, N$ )

$N$ : 認識単語数 ( $N = 24$ )

$Rank\_Beginning_i$ : 手話動作開始時の手指形状認識結果の順位

$Rank\_End_i$ : 手話動作終了時の手指形状認識結果の順位

$Rank\_Arm_i$ : 手首の動き認識結果の順位

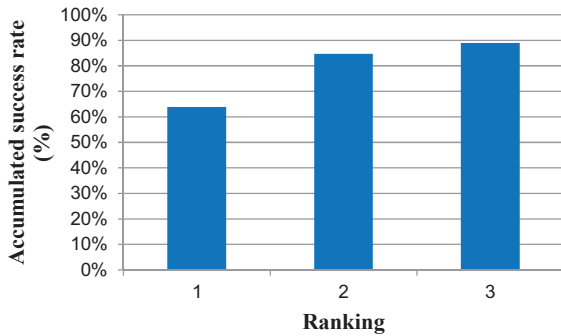


Fig. 12 Accumulated success rate at beginning of sign language motion.

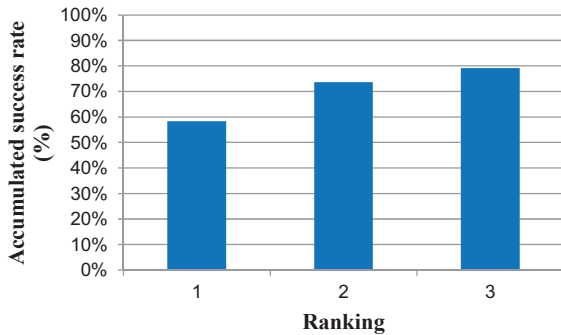


Fig. 13 Accumulated success rate at end of sign language motion.

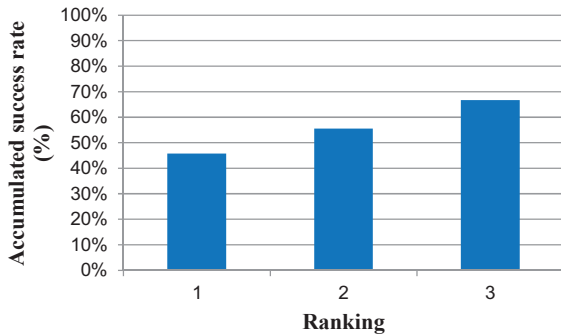


Fig. 14 Accumulated success rate for arm motion recognition.

### 5.3.4 手話認識結果

5.3.3 の方法を用いて、3 つの認識結果を統合し、手話認識結果を求めた。手話認識実験結果を Table 6 に示す。実験結果から、平均正解率 83.3%となることを確認した。誤認識した単語の多くは、手指の形状と手首の動きが共に類似している単語同士である。なお、このときの 2 位, 3 位までの累積正解率は 87.5%, 91.7% であり、極めて高い精度で認識処理が行われていることがわかった。

Table 6 Experimental results for sign language recognition.

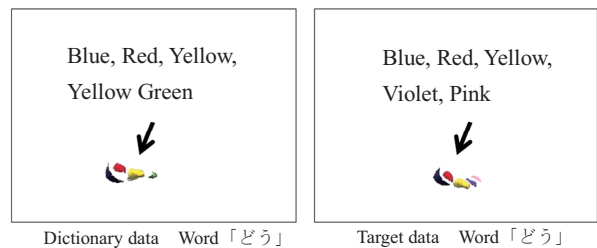
	Template																							
	兄	姉	有る	妹	上	弟	男	女	くら	月曜日	下	少	飾	茶	次	土	どう?	場所	日	また	右	老人		
兄	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
姉	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
有る	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
妹	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
上A	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
上B	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
弟	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
男	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	2		
女	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0		
くら	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0		
月曜日	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0		
下A	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0		
下B	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0		
すし	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0		
飾	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0		
茶B	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0		
次	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	1	0		
土	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0		
どう	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0		
場所	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0		
日	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1		
また	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3		
右	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
老人	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3		

### 5.4 誤認識の考察

誤認識した原因の一つは、手話動作開始時と終了時の手指形状、手首の動きが全て類似している単語同士であったためである。これらの単語は指のみを動かす単語である。今回の提案では、手話動作開始時と終了時の手指形状と動作中の手首の動きのみを用いて認識を行っているため、動作中の指の動きは考慮していない。よって、今後、指の動きを認識に取り入れることにより解決が期待できる。

また、今回の実験では被験者は手話を習得していない健聴者であった。そのため、手話動作の練習不足で被験者の手話動作が安定せず、辞書データと動作データで色検出結果や手首の動きが異なった可能性が高い。手話動作終了時における色検出結果の一例を Fig.15 に示す。辞書データと動作データで手指の角度が異なり、色検出結果が異なっていることが確認できる。これに関してはより多くのパターンに対応させたテンプレート作成や、手話習得者に協力を依頼し、安定した手話動作から辞書データを作成する必要があると考えている。

手話認識の 3 位までの累積正解率は 91.7%であったため、指の動き認識の導入やテンプレート数の増加により今後より高い手話認識精度の実現が期待できる。



Dictionary data Word「どう」 Target data Word「どう」

Fig. 15 Color detection result at end of sign language motion.

## 6. まとめ

カラー手袋と可視光カメラを用いた手話認識手法を提案した。まず、カラー手袋に彩色する色の数と種類を決定し、クラスタリング手法によって各色の検出閾値を自動決定する手法を提案した。照明条件を変化させ、カラー手袋に彩色した 8 色の色検出精度を評価した結果、平均色検出率が 56.1%から 96.5%に向上し、提案手法を用いることで環境光の変化にロバストな色検出が行えることを確認した。

さらに、色検出結果の種類と位置から手話動作開始時と終了時の手指形状を、手首に彩色した色領域の重心位置変化から手首の動きを認識し、これらの統合によって手話を認識する手法を提案した。提案した手法の手話認識への可能性を検討するため、全国手話検定試験 5 級の基本単語 24 単語において認識実験を行った結果、正解率 83.3%となることを確認し、その実現性を示した。

## 参考文献

- [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] 三留綾, 市毛 弘一, 石井 六哉: マスキングと規格化処理を用いた手領域画像の特徴解析および手指形状認識; 電子情報通信学会論文誌, 信学論(D), Vol.J89-D, no.6, pp.1369-1378 (2006).
- [3] 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).
- [4] 渡辺 賢, 岩井 儀雄, 八木 康史, 谷内田 正彦: カラーグローブを用いた指文字の認識; 電子情報通信学会論文誌, D-11, Vol.J80-D- II, No.10, pp.2713-2722 (1997).
- [5] 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).
- [6] 和田 健太, 福村 直博: Kinect を利用した腕運動と手形状に基づく手話認識手法の検討; 信学技報, WIT2013-81, pp.81-86 (2005).
- [7] 山根 卓也, 船富 卓哉, 飯山 将晃, 美濃 導彦: データグローブ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).
- [8] Sugaya, T., Itagaki, Y., Nishimura, H., Tanaka, H.: Enhancement of Accuracy of Hand Shape Recognition Using Color Calibration by Clustering Scheme and Majority Voting Method; HCII2014, LNCS.Vol.8521, pp.251-260 (2014).
- [9] 非手指動作を付加した手話映像生成; NHK 技研 R&D, No.129, p.32 (2011).
- [10] DVD で学ぶ手話の本; 社会福祉法人全国手話研修センター 編集, 中央法規, (2011).
- [11] 中川 優樹, 石田 貴正, 平野 晃明, 中村 納: 自然屋内環境での表情認識を目的とした高繊細な顔領域抽出; 工学院大学研究報告, 第 112 号, pp.37-44 (2012).
- [12] Tanaka, H., Kimura, R., Ioroi, S.: Equipment Operation by Motion Recognition with Wearable Wireless Acceleration Sensor; NGMAST2009, pp.114-118 (2009).