卒 業 論 文 概 要 書

Summary of Bachelor's Thesis

Date of submission: 02/06/2019 (MM/DD/YYYY)

学科名 Department	情報通信	氏名 Name	栁澤 利紀	指 導 教 員 Advisor	渡辺 裕 ㊞
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 ID number	1w142394-5 CD		
研究題目 Title	単眼 RGB カメラを用いた手話の動作解析 Action Analysis in Sign Language with Monocular RGB Camera				

1. まえがき

手話はろう者が意思疎通を行う上で重要な手段である．しかし，その習得は難しく，健聴者の多くは意味を容易に理解できない．

そこで，手話の翻訳自動化が望まれており，様々な取り組みが行われてきた．しかし，それらは特別な機器を必要とするものや，動作環境が精度に大きく影響するものが多い．

本研究では，一般的なカメラ映像を用いた手話動作の解析手法を提案する．

2. 従来手法

2.1 データグローブを用いる手法

センサを手袋の内部に付けたセンサグローブを用いて，手形状の変化を素早く正確に取得することができる手法である．しかし，日常生活で使用では着脱が負担となり，長時間の使用にも適さないという問題がある．

2.2 距離カメラを用いる手法

Depth センサが内蔵された Kinect や LeapMotion という機器を用いた手話認識も行われている．これらは RGB 画像に加えて Depth センサを用いることで手や指の位置を追跡する仕組みである．しかし，いずれも特殊な機器が必要である．さらに撮影角度が限られるという欠点がある．

2.3 単眼 RGB カメラと画像処理を用いる手法

単眼 RGB カメラを用いた方法としては，背景差分による動体検出や勾配方向ヒストグラムを用いた手の抽出などが行われている.しかし，これらの方法では手の形状や肌の色，背景，環境光などの影響により，認識が難しい場合がある．

この問題を解決するために，カラー手袋を用いた手の検出なども行われているが，手話の使用者に着脱が負担となる問題点がある．

3. 提案手法の関連技術

3.1 OpenPose

Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields (OpenPose) は，カーネギーメロン大学 (CMU) の Zhe ら[1]によって発表された姿勢推定アルゴリズムである．RGB カメラから多人数の体や顔，手のキーポイントをリアルタイムに抽出できる．

3.2 LSTM

Long Short Term Memory (LSTM) は，Recurrent Neural Network (RNN) に，入力ゲート，忘却ゲート，出力ゲートからなる LSTM ブロックを付けた構造を持つ．RNN を長期の系列データの分析に適するように改善したネットワークである．

4. 提案手法

4.1 提案手法の概要

提案手法では，OpenPose と LSTM を用いて単眼 RGB カメラの映像から手話の動作解析を行う．まず，入力動画に対して OpenPose を適用し，姿勢推定を行う．次に，人物抽出を行い，フレームごとに特徴量計算を行う．その特量量を用い，LSTM によって時系列分析を行い，動作の分類を行う．

提案手法の構成を図 4.1 に示す．

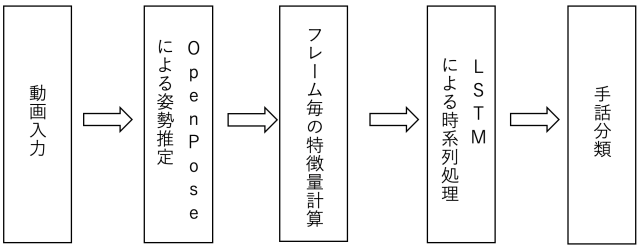


図 4.1 本研究の構成

4.2 OpenPose による姿勢推定

本研究では，OpenPose の出力のうち表 4.1 に示す体の 7 点のキーポイントを用いる．

表 4.1 本研究で用いる OpenPose の部位

OpenPose における部位番号	部位
1	首
2	右肩
3	右肘
4	右手首
5	左肩
6	左肘
7	左手首

4.3 人物抽出

OpenPose では複数の人物が写っている動画を入力とした場合に，人物の同定ができない．そこで，前フレームからの各部位の移動量が最小である人を同一人物と判定する．

4.4 フレームごとの特徴量計算

7 点の部位を表 4.1 の部位番号を添字に用い，P_1，P_2，P_3，P_4，P_5，P_6，P_7と表す．この中で最も動きの少ない首を基準にし，各部位までのベクトルを求める．すなわち，$\overrightarrow{P_{12}}$，$\overrightarrow{P_{13}}$，$\overrightarrow{P_{14}}$，$\overrightarrow{P_{15}}$，$\overrightarrow{P_{16}}$，$\overrightarrow{P_{17}}$を計算する．これらのベクトルを首から両肩までの距離の平均値$\frac{|\overrightarrow{P_{12}}|+|\overrightarrow{P_{15}}|}{2}$で割る．以上の操作により得られる 6 つのベクトルの x 座標と y 座標，計 12 点を特徴量とする．

4.5 LSTM による時系列処理

得られた特徴量に対して LSTM を用いて動作の分類を行った．LSTM の入力次元数を 12 次元とした．フレームごとの特徴量の次元数に合致させた．LSTM の出力の次元数は 5 単語分類を目標とし，5 次元に設定した．

5 実験

5.1 実験データ

本研究では，日本語の両手手話である「お大事に」，「お疲れ様」，「危ない」，「久しぶり」，「笑う」を対象とした．これら 5 種類の動作を 9 人の被験者から約 50 回ずつ撮影した動画を動作部分に区切り，分析に用いた．

5.2 実験結果

得られた実験データに対し提案手法の方法を用いて分類を行った．取得した 9 人のデータのうち，8 人分を学習用データ，1 人分をテスト用データとする 9 交差検定を行った．交差検定の結果を表 5.1 に示す．表より，平均正解率 95.11%が得られ，有効であることが分かった．

表 5.1 交差検定の結果

		予測値				
		動作 1	動作 2	動作 3	動作 4	動作 5
真値	動作 1	445	0	2	4	0
	動作 2	2	444	5	0	0
	動作 3	53	0	407	0	0
	動作 4	3	0	2	420	16
	動作 5	16	0	0	7	425

6 むすび

本研究では単眼 RGB カメラを用いた手話の動作解析を OpenPose と LSTM を用いて行い，正解率 95.11%を得た．単語数が増えた場合に手の座標情報も含める必要が出てくるが，特徴量の次元が増大し精度が向上しない問題点がある．

参考文献

[1] Z. Cao, T. Simon, S-E Wei, Y.Sheikh, “Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields,” In 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), No.121, pp.1302-1310, July 2017.

2018 年度 卒業論文

単眼 RGB カメラを用いた手話の動作解析

Action Analysis in Sign Language with Monocular RGB Camera

指導教員 渡辺 裕 教授

早稲田大学 基幹理工学部

情報通信学科

1w142394-5

柳澤利紀

目次

第 1 章　序論

1.1 研究背景

手話はろう者が意思疎通を行う上で重要な手段であり，日本国内には約 6 万人の手話話者がいると言われている[1]. 2011 年の障害者基本法改正[2]によって日本における正式な言語の一つとして規定された．しかし，手話の習得は難しく，健聴者の多くはその意味を理解出来ない．手話を通した健聴者との意思疎通が困難な場合，コミュニケーションは筆談や手話通訳者を介して行われるが, それらは万全な方法ではない. 筆談を用いる方法では，手話と一般的な言語のルールの違いを手話話者が難しく感じる場合がある[3]．また，文字を書く際に労力がかかり，リアルタイムに情報を伝達出来ない．さらに，雰囲気などの微細なニュアンスを伝えにくいという問題点がある．また，通訳を通した手話翻訳では，手話翻訳者は日本国内に約 3600 名しかおらず[4]，数が十分に足りているとは言えない．

こういった背景から，手話の翻訳自動化が望まれており，様々な取り組みが行われてきた．しかし, 高い手話認識精度を実現するためには, 特別な機器を必要とするものが多い．また，一般的なカメラを用いた研究も行われているが，動作環境が手話認識精度に影響しやすいという問題点がある．

本研究では，一般的なカメラ映像から動作環境に左右されない手話の動作解析システムを作成することを目指す．

1.2 関連研究

手話認識の方法は主に，データグローブを用いる方法，Kinect などの距離カメラを用いる方法，単眼 RGB カメラと OpenCV などを用いる方法の三つに分けられる．

1.2.1 データグローブを用いる方法

センサを手袋の内部に付けたセンサグローブを用いると手形状の変化を素早く正確に取得することができる．岩峪ら[5]や，福原ら[6]はこの性質を利用して手話の学習システムの研究を行なっている．限られた使用条件においてデータグローブは有効であることが示されたが，日常生活で使用するには着脱の手間や長時間の使用がストレスになるなどの問題点がある．

1.2.2 距離カメラを用いる方法

Depth センサが内蔵された Kinect や LeapMotion といった距離カメラを用いた手話認識も行われている[7][8]．これらは RGB 画像に加えて Depth センサを用いることで手や指の位置を追跡する仕組みになっている．しかし，いずれも特殊なカメラが必要であるという欠点がある．また，一般的に LeapMotion は手の下に置いた状態でしか使えず，撮影の角度が限定される．

1.2.3 単眼 RGB カメラと OpenCV などを用いる方法

単眼 RGB カメラを用いた方法として，外山ら[9]は，背景差分による動体検出と明度情報を用いて手領域を抜き出し，細線化から指先を，内接円から手のひらを検出する手法を提案している．また山田ら[10]は，手の検出に勾配方向ヒストグラムを用いている．しかし，これらの方法では手の形状や肌の色，背景，環境光などの影響により，認識が難しい場合がある．

これを解消するために，カラー手袋を用いた手の検出なども行われている[11]が，これは手話の使用者に対して特殊器具の装着という負荷を与えるという欠点がある．

1.3 本研究の目的

特殊な器具を使わず，動作環境に左右されない手話の動作認識は，上記問題を解決する上で重要であると考えられる．そこで本研究では，一般的な RGB カメラで撮影した映像を用いて手話の動作解析を行うことを目的とする．

1.4 論文の構成

以下に本章以降の構成を示す．

第 1 章 「序論」は本章であり，本論文の背景および目的について述べている．また，関連研究について述べている．

第 2 章 「関連技術」では，本研究で用いる姿勢推定手法である OpenPose と系列データ分析用のアルゴリズムである LSTM について述べる．

第 3 章 「提案手法」では，本研究で行う手話の動作解析の手法について述べる．

第 4 章 「実験結果」では，本研究の結果について述べる．

第 5 章 「結論」では，本研究の結論および今後の課題について述べる．

第 2 章　関連技術

2.1 まえがき

本章では，本研究で用いる姿勢推定技術である OpenPose と系列データ分析用のアルゴリズムである LSTM について述べる．

2.2 OpenPose

Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields (OpenPose) は，Zhe ら[12]によって提案された動画像からのリアルタイム多人数人物姿勢推定技術であり，GitHub からダウンロードして利用することが出来る[13].

OpenPose 以前の姿勢推定手法には，画像から人物検出を行い検出された人物一人一人に姿勢推定を行うトップダウン的な方法と，画像から体の部分を探すことによって姿勢推定を行うボトムアップ的な方法が存在した．しかし，前者には人物検出の段階で失敗してしまうとその段階で姿勢推定に失敗してしまい，リカバリーができないという問題点があった．また，後者には部位の情報を繋ぎ合わせるのが難しく，計算時間が多くかかるという問題点が存在した．

OpenPose は従来のボトムアップ法で使われていた Part Confidence maps という部位検出のための CNN に加えて，画像上の部位の関連性をベクトルで表した Part Affinity Fields という CNN を用いることで，高速かつ高精度な姿勢推定を可能とした．この構造については 2.2.1 節でさらに詳しく述べる．

2.2.1 OpenPose の構造

OpenPose の処理手順を図 2.2.1 に示す．OpenPose では入力画像を受け取ると Part Confidence Maps と Part Affinity Fields が求められる．

Part Confidence Maps の各層では部位の存在可能性が高い場所が抽出される．例えば図 2.2.1(b)では，左肩の存在可能性が高い場所と左肘の存在可能性が高い場所が抽出される様子を表している．他の部位についても存在可能性の高い場所が同様に抽出される．

Part Affinity Fields では，部位間の繋がりがベクトルとして得られる．図 2.2.1(c)は，左肩から左肘にかけて連結が想定されるベクトルの様子を表している．他の層によって体の各部分の間を繋ぐベクトルが得られる．

その後，Part Confidence Maps と Part Affinity Fields の情報を合わせて，最も可能性の高い

部位およびその連結が求められる．

以上により出力画像が得られる．動画の場合はフレーム毎にこの機構が繰り返し用いられる．

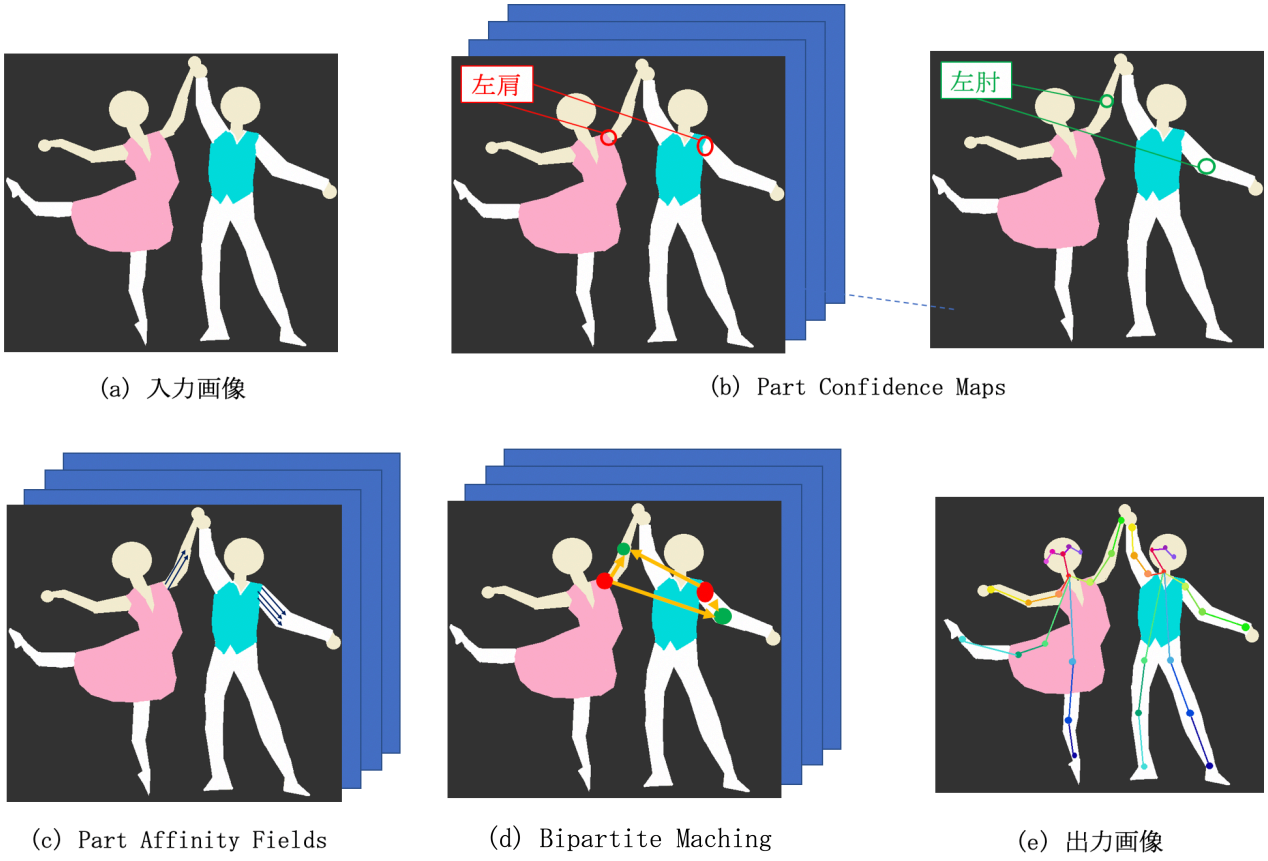
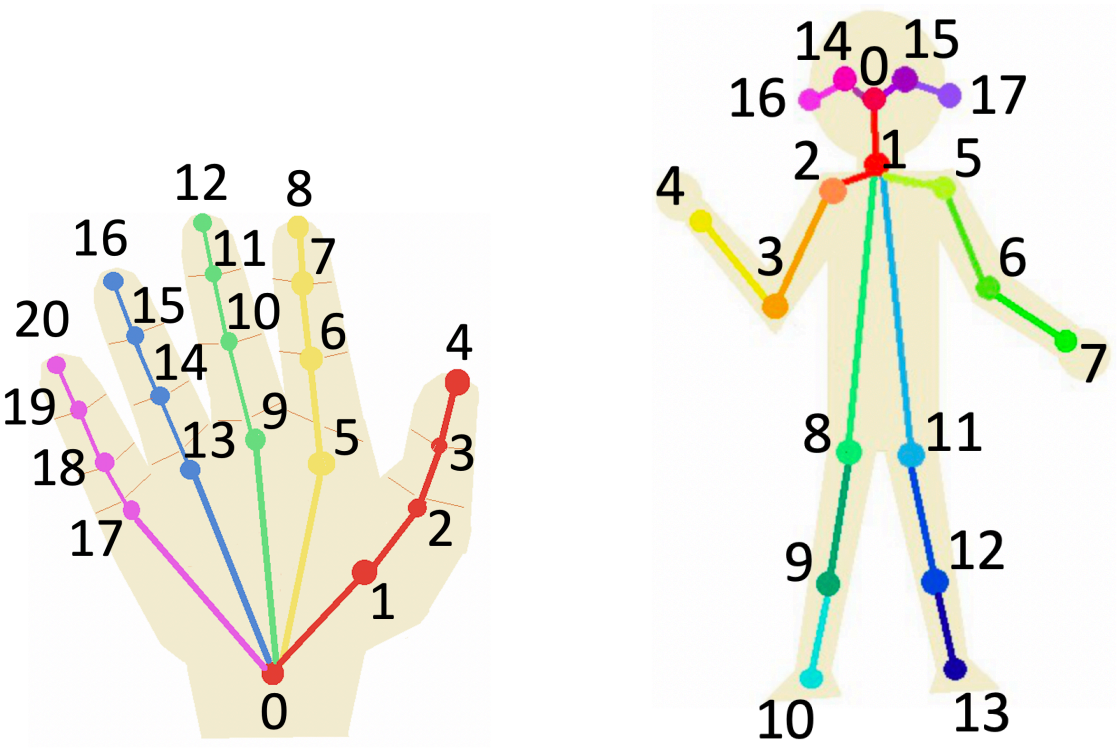


図 2.2.1 OpenPose の処理手順

2.2.2 OpenPose により得られるキーポイント

2.2.1 節の出力は各部位毎の座標情報として得られる．開発初期に得られたデータは体の座標点だけであったが，OpenPose は年々バージョンアップされており，現在では体に加えて，手，顔，足のキーポイント計 135 点を取得することができる．

これらのキーポイントの中から本研究と関係の深い手と体のキーポイントをそれぞれ図 2.2.2(a)と図 2.2.2(b)に示す．手は 21 点,体は 18 点からなる．

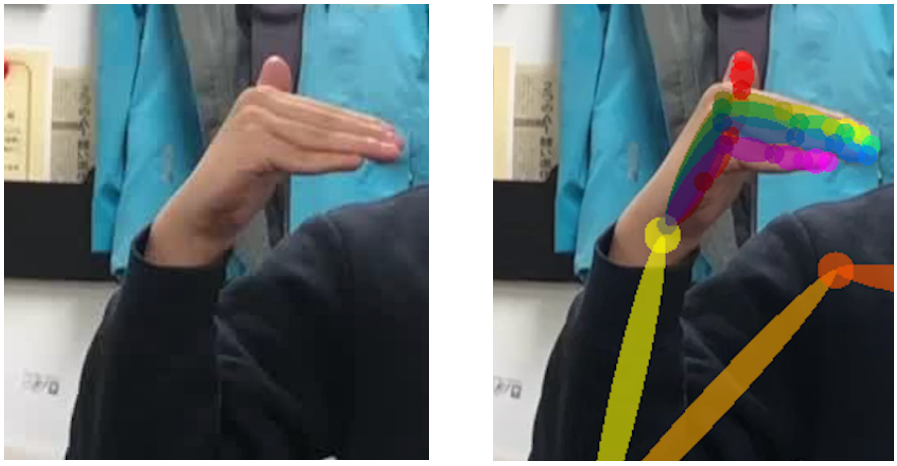


（a）手　　（b）体

図 2.2.2 OpenPose のキーポイント

2.2.3 OpenPose を用いる利点

OpenPose は高精度な学習済みモデルである．一般的なカメラ画像から肌の色や背景，光量などに影響されずに用いることができる．また，人物が複数含まれる場合でも検出したい人物を選択して分析することが出来る．手話は手の形状のみでなく，腕の動きや頷きなどの首の動きも用いる．したがって OpenPose はそれらの情報も得ることができる点から手話の認識に向いていると言える．手話分析に際しオクルージョンは大きな問題となる．OpenPose では体の部位が隠れている場合でも，ある程度正確な数値を推定できる．このような例を図 2.2.3~2.2.5 に示す．

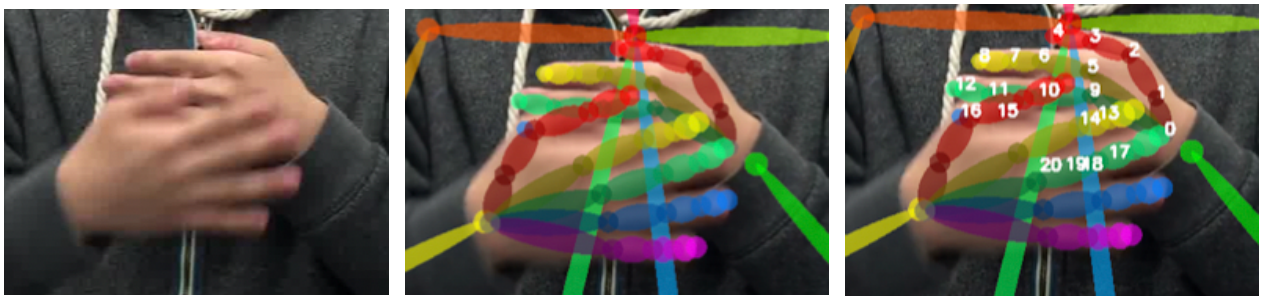


(a) 元画像　　(b) OpenPose 適用画像

図 2.2.3 OpenPose 適用例 その 1

(a) 元画像　　(b) OpenPose 適用画像

図 2.2.4 OpenPose 適用例 その 2



(a) 元画像　　(b) OpenPose 適用画像　　(c) OpenPose 適用画像の左手に座標点を書き込んだもの

図 2.2.5 OpenPose 適用例 その 3

図 2.2.3(a) の親指や図 2.2.4(a) の人差し指，中指，薬指は他の指に隠れて見えない．しかし，図 2.2.3(b) や図 2.2.4(b) を見ると，これらの指も含めた OpenPose 出力が得られていることが分かる．また，図 2.2.5(a) では，右手によって左手の一部が隠れており，図 2.2.5(b) の出力画像を見ると隠れた部分の出力が得られていないように見える．しかし，右手の座標情報を重ねて表示した図 2.2.5(c) を見ると，おおよそ適切な位置に 21 点が打たれている．

以上より，一つのパーツ内部でオクルージョンが起こっている場合と，他のパーツによってオクルージョンが起こっている場合の，どちらの場合についても座標の推定がなされていることが確認できる．

2.3 LSTM

Long Short Term Memory (LSTM) は，Recurrent Neural Network (RNN) の一種で，センサデータや動画などの系列データのモデリングに用いられる．ここではその構造について，前身技術である RNN との相違点に触れながら述べる．RNN は Neural Network (NN)の系列データへの応用であるので，まず NN について述べる．

2.3.1 Neural Network について

NN は，脳の神経細胞であるニューロンをモデルとしたアルゴリズムであり，入力層，隠れ層，出力層からなる．各層は複数のノードからなり，ノード間はエッジにより結ばれている．NN はパーセプトロンを組み合わせたような構造をしており，各ノードの値は，前の層のノードの値，接続エッジの重みの値，作用させる活性化関数によって求まる．この構造を図 2.3.1 に示す．

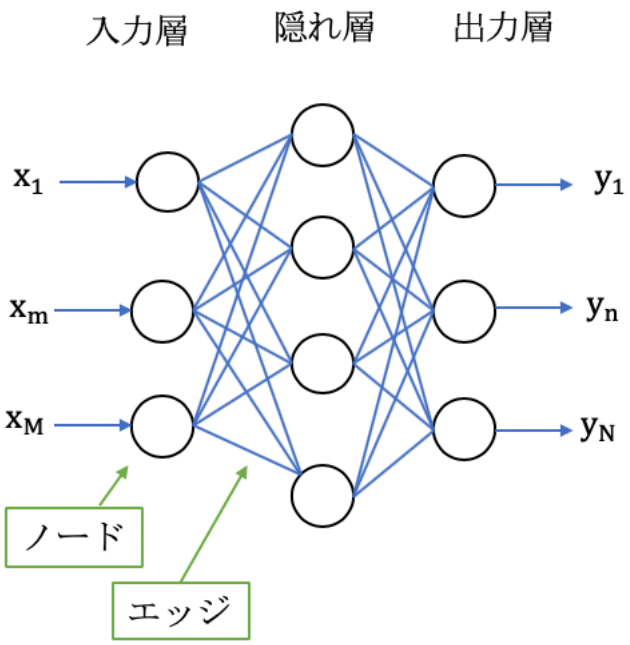


図 2.3.1 Neural Network の構造

2.3.2 Recurrent Neural Network について

1 ステップ前の隠れ層の状態を現在の隠れ層に組み込むことによって，NN を系列データに適用できるようにしたものが RNN である．1 ステップ前の隠れ層は 2 ステップ前の隠れ層の状態を参照するため，現在の隠れ層は再帰的に時系列の隠れ層の状態を参照することになる．このことにより，系列データの分析ができる．この構造を図 2.3.2 に示す．

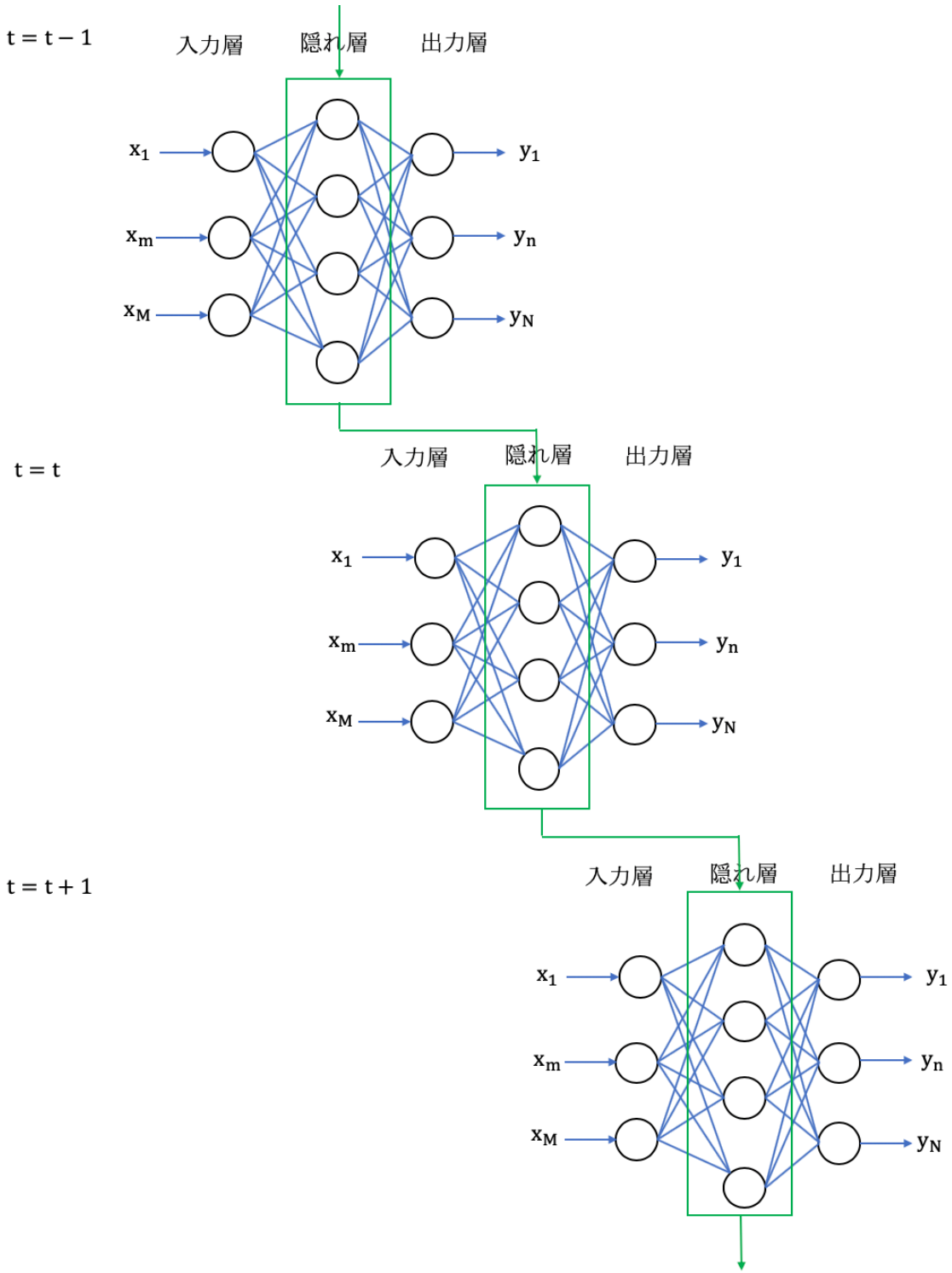


図 2.3.2 Recurrent Neural Network の構造

2.3.3 Long Short Term Memory について

RNN は系列データを分析できるが，長期の依存関係をモデル化には適さず，学習時に勾配爆発や勾配消失が起こることがある．この問題を改良したネットワークが LSTM である．LSTM は，RNN の隠れ層に入力ゲート，消失ゲート，出力ゲート，メモリセルを

持つ LSTM ブロックを持つ．これにより長期の依存関係をモデル化できる．この構造を図 2.3.3 に示す．

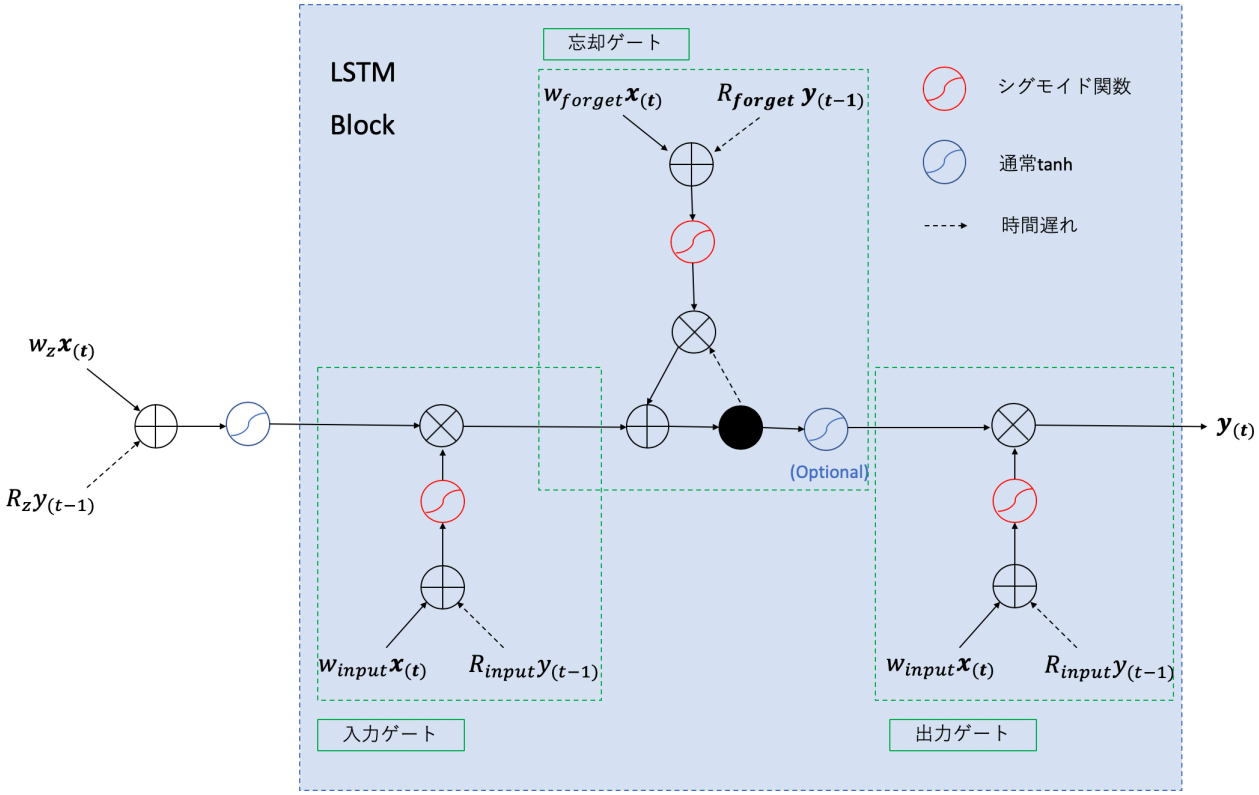


図 2.3.3 LSTM の構造

2.4 むすび

本章では，姿勢推定技術である OpenPose と時系列データ解析用アルゴリズムである LSTM についてその概要と構造，および利点を述べた．

第 3 章　提案手法

3.1 まえがき

本章では，手話の動作解析の方法について述べる．本研究では，手話を行う際の腕の動きに着目して動作分類を行った．

3.2 手法の概要

提案方式の処理フローを図 3.1 に示す．この処理では，まず，入力された動画に対して OpenPose を適用し姿勢推定を行う．次に，姿勢推定より得られた座標値を用いて人物抽出およびフレーム毎の特徴量を計算する．最後に，LSTM によって時系列分類することによって対象の手話動作を分類するフローになっている．

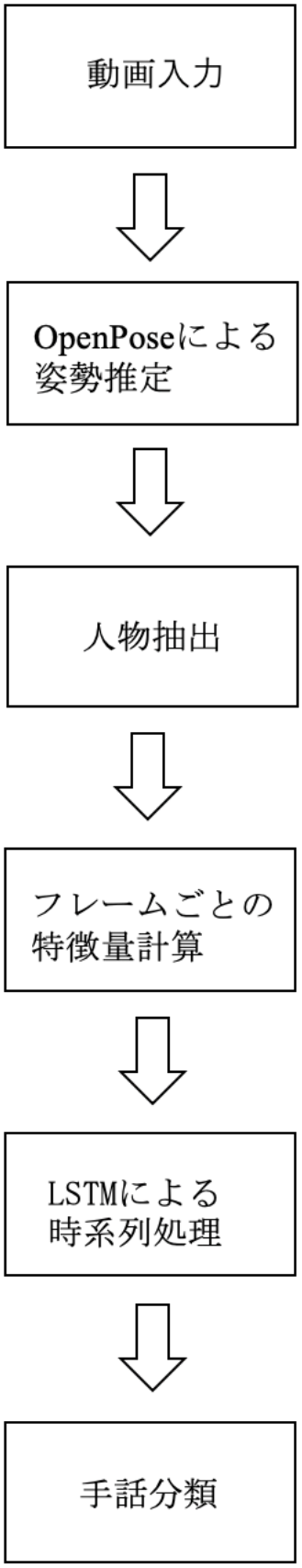


図 3.1 提案方式の処理フロー

3.3 人物抽出

OpenPose の出力は，2.2.2 節で述べたキーポイントが，フレーム毎に json 形式で得られる．画面中に 2 人の人物が写っている場合の出力例を図 3.2 に示す．

{ “version”:1.2,
“people”: [{ “pose_keypoints”: [927.557, 256.528, 0.861868,……..],
”face_keypoints”: [847.317, 230.511, 0.886248,……...],
“hand_left_keypoints”: [1052.01, 627.244, 0.40455,……],
“hand_right_keypoints”: [749.181, 516.46, 0.295196, …….],
}
{ “pose_keypoints”: [746.181, 248.142, 0.892816,……..],
”face_keypoints”: [691.517, 225.148, 0.791344,……...],
“hand_left_keypoints”: [893.187, 614.716, 0.43421,……],
“hand_right_keypoints”: [641.237, 561.512, 0.415326, …….],
}]
}

図 3.2 OpenPose の出力例

OpenPose は画像に対して姿勢推定を行うという構造上，入力が動画でありかつ複数の人物が写っている場合，得られる人物の順番はフレームによって異なる．そこで，式(3.1)によって連続するフレーム間での検出人物の各部位座標の移動量を計算し，その和が最も小さかったものを同一人物と決定する．

$$f(k) = \sum_{p \subset P} \left(x_{t,p} - x_{t-1,k,p}, y_{t,p} - y_{t-1,k,p}\right)^2 \tag{3.1}$$

ここでは，パーツ全体の集合を P，時刻 t での人物 k のパーツ p の座標を $\left(x_{t,k,p}, y_{t,k,p}\right)$ としている．

3.4 データ整形

ここでは，図 2.2.3 で示した OpenPose により得られる体の部位の中で，本研究で解析に用いる部位を表 3.1 に示す.

表 3.1 本研究で用いる OpenPose の部位

OpenPose における部位番号	部位
1	首
2	右肩
3	右肘
4	右手首
5	左肩
6	左肘
7	左手首

それぞれの部位を，P_1，P_2，P_3，P_4，P_5，P_6，P_7と表す．この中で最も動きの少ない首を基準にし，各部位までのベクトルを求める．すなわち，$\overrightarrow{P_{12}}$，$\overrightarrow{P_{13}}$，$\overrightarrow{P_{14}}$，$\overrightarrow{P_{15}}$，$\overrightarrow{P_{16}}$，$\overrightarrow{P_{17}}$を計算する．これらのベクトルを，首から両肩までの距離の平均値

$$l = \frac{|\overrightarrow{P_{12}}| + |\overrightarrow{P_{15}}|}{2} \tag{3.2}$$

で正規化する．これによって，撮影距離による大きさの違いを吸収できる.

以上により求められた部位間の 6 本のベクトルの x 座標，y 座標を特徴量とする．これらをまとめたものを表 3.2 に示す.

表 3.2 本研究で用いる部位間ベクトル

特徴量の番号	特徴量の内容
1	$\overrightarrow{P_{12}}/l$ の x 座標
2	$\overrightarrow{P_{12}}/l$ の y 座標
3	$\overrightarrow{P_{13}}/l$ の x 座標
4	$\overrightarrow{P_{13}}/l$ の y 座標
5	$\overrightarrow{P_{14}}/l$ の x 座標
6	$\overrightarrow{P_{14}}/l$ の y 座標
7	$\overrightarrow{P_{15}}/l$ の x 座標
8	$\overrightarrow{P_{15}}/l$ の y 座標
9	$\overrightarrow{P_{16}}/l$ の x 座標
10	$\overrightarrow{P_{16}}/l$ の y 座標
11	$\overrightarrow{P_{17}}/l$ の x 座標
12	$\overrightarrow{P_{17}}/l$ の y 座標

3.5 LSTM を用いた分類

提案手法で用いる分類処理のためのブロック図を図 3.3 に示す. 時刻 t における入力ベクトル$\boldsymbol{x}_{(t)}$ から，図 2.3.3 で示した LSTM ブロックを通して出力を求める．動画のフレーム数 N までの入力ベクトルを順に入力して出力を求め，最終出力を動作と対応付けることで分類を行う.

本研究では 12 次元の特徴量を入力として 5 単語分類を行った．このとき用いた，LSTM のパラメタを表 3.3 に示す．LSTM レイヤーの数は試行の結果うまく収束した 100 層を採用した.

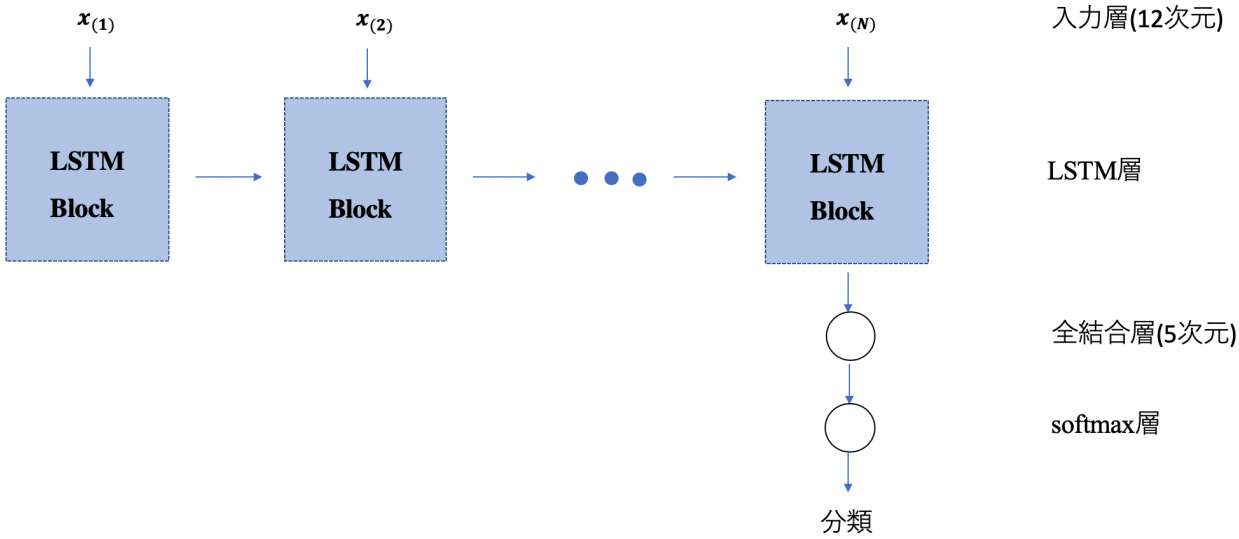


図 3.3 提案手法で用いる分類処理のブロック図

表 3.3 LSTM のパラメタ

パラメタ名	パラメタ詳細
入力シーケンスの次元数	12
LSTM レイヤーの数	100 層
全結合層の数	5

3.6 むすび

本章では，OpenPose を用いた姿勢推定および人物抽出，データ整形について述べた．また，整形したデータを用いて動作を分類する手法について述べた．

第 4 章 実験結果

4.1 まえがき

本章では，対象とした手話動作などの実験データ，特徴量の時間変化，実験結果について述べる．

4.2 実験データ

手話には片手のみを用いる片手手話と，両手を用いる両手手話がある．本研究では，日本語の両手手話である「お大事に」，「お疲れ様」，「危ない」，「久しぶり」，「笑う」について提案手法を用いて分類を行った．5 種類の手話動作を図 4.1~4.5 に示す．

これらの動作を，9 人の被験者から各動作につき 50 回ずつ収集し，動作部分の開始点と終了点で区切ったものを分析の対象とした．撮影には iPhone8 のアウトカメラを用い，撮影場所はあえて異なる場所とした．

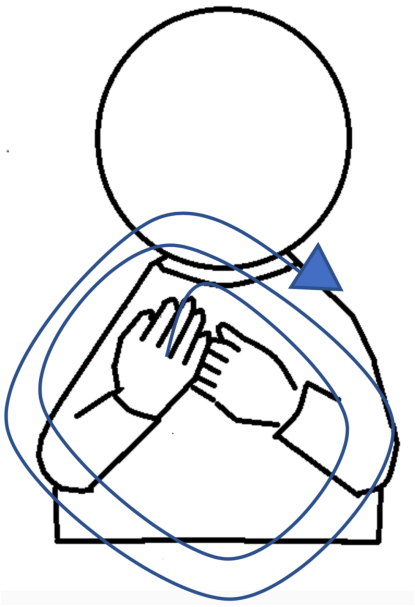


図 4.1 手話動作 1「お大事に」

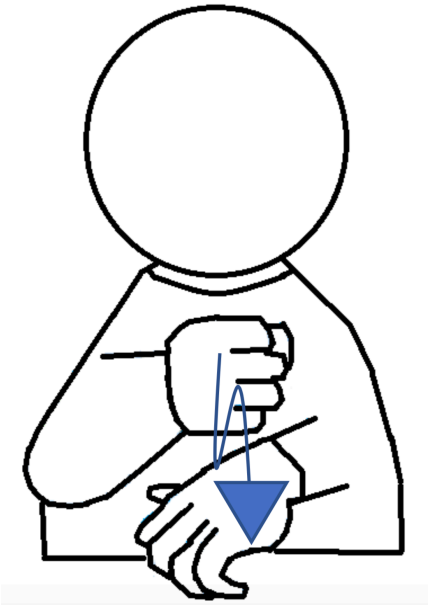


図 4.2 手話動作 2「お疲れ様」

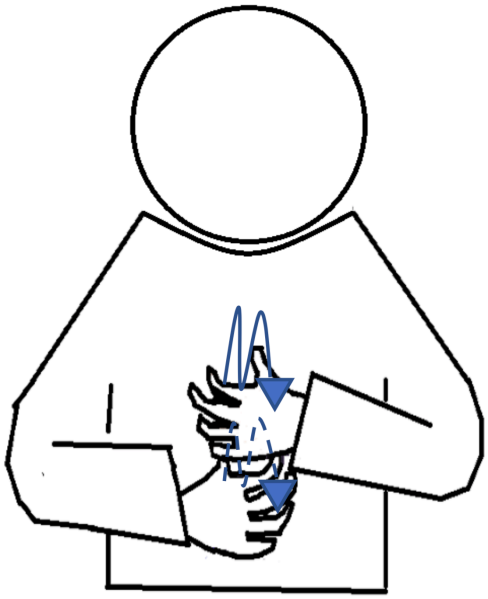


図 4.3 手話動作 3「危ない」

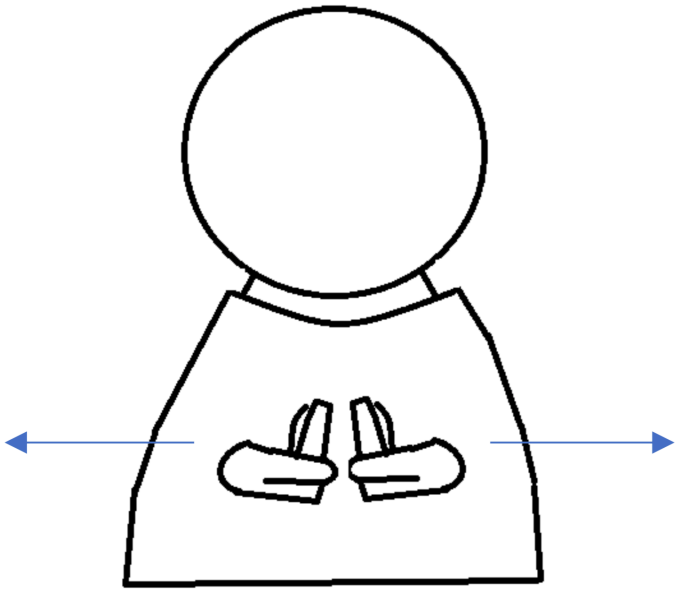


図 4.4 手話動作 4「久しぶり」

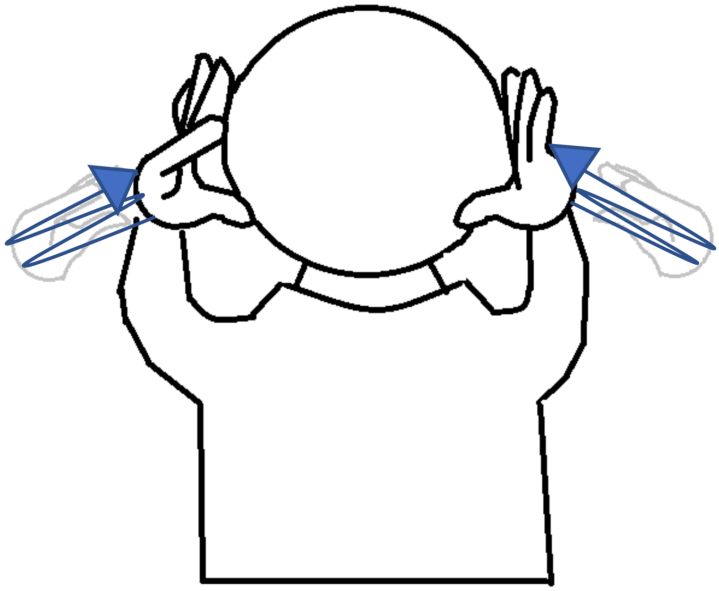
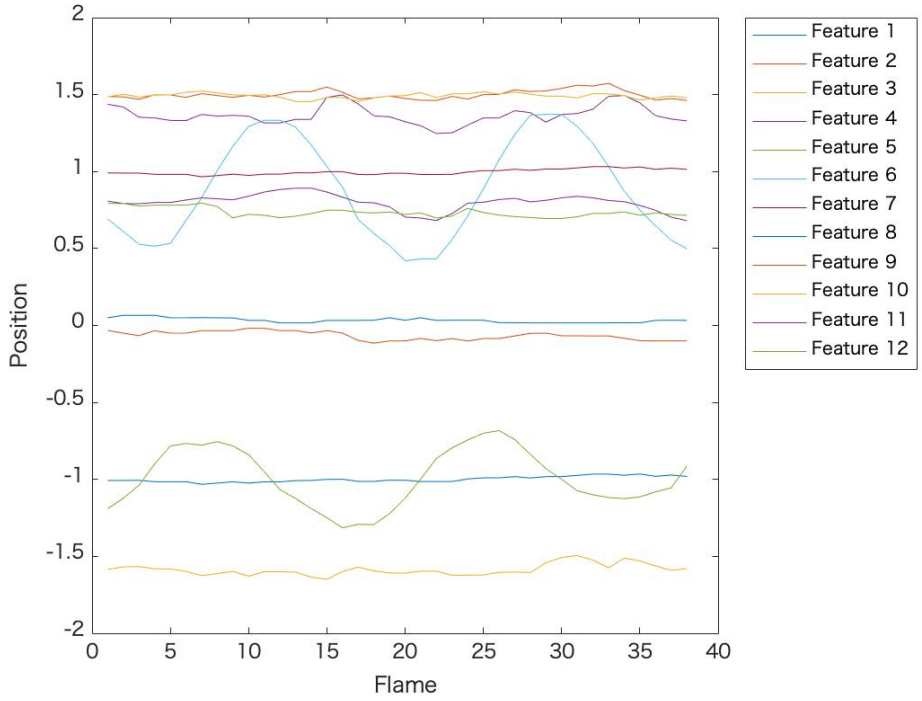


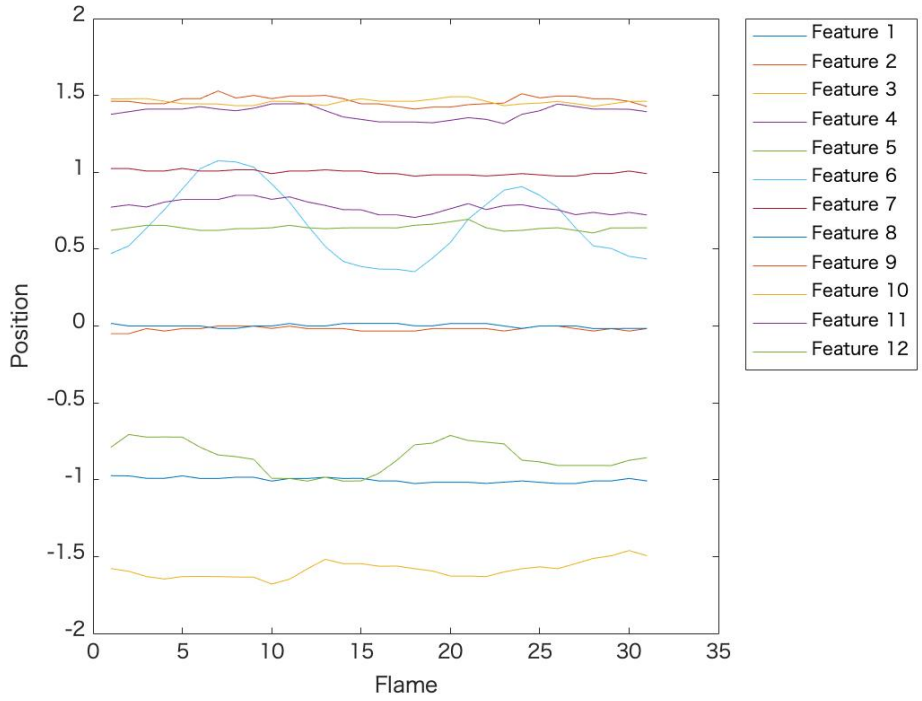
図 4.5 手話動作 5「笑う」

4.3 実験データの波形

手話動作 1~5 について表 3.2 で示した 12 個の特徴量を求めた．これら特徴量の時間変化を 2 サンプルずつ表したものを図 4.6~4.10 に示す．

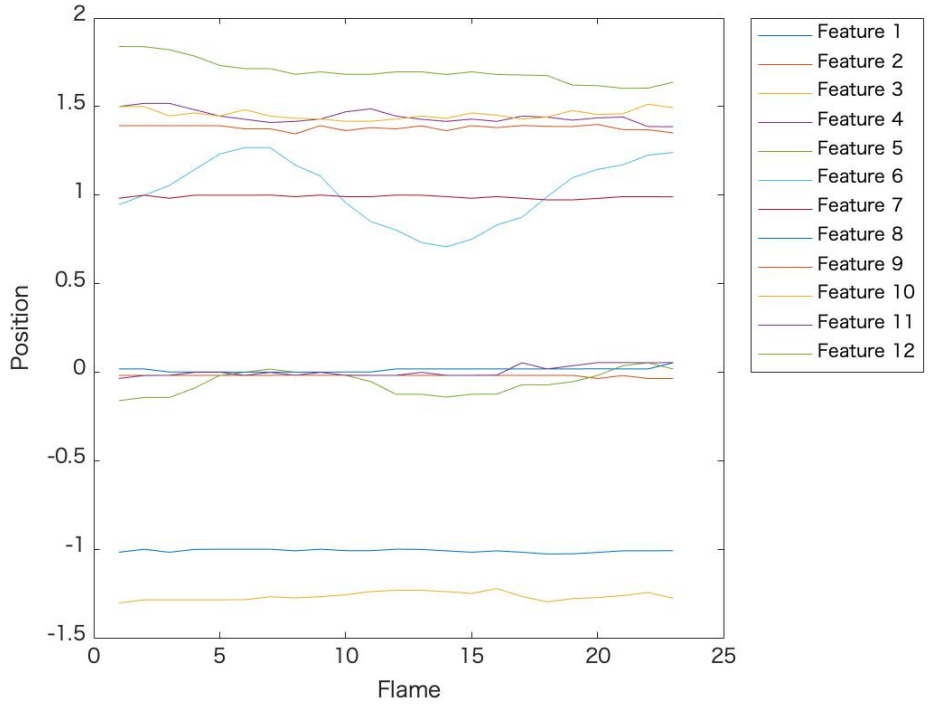


(a) サンプル 1

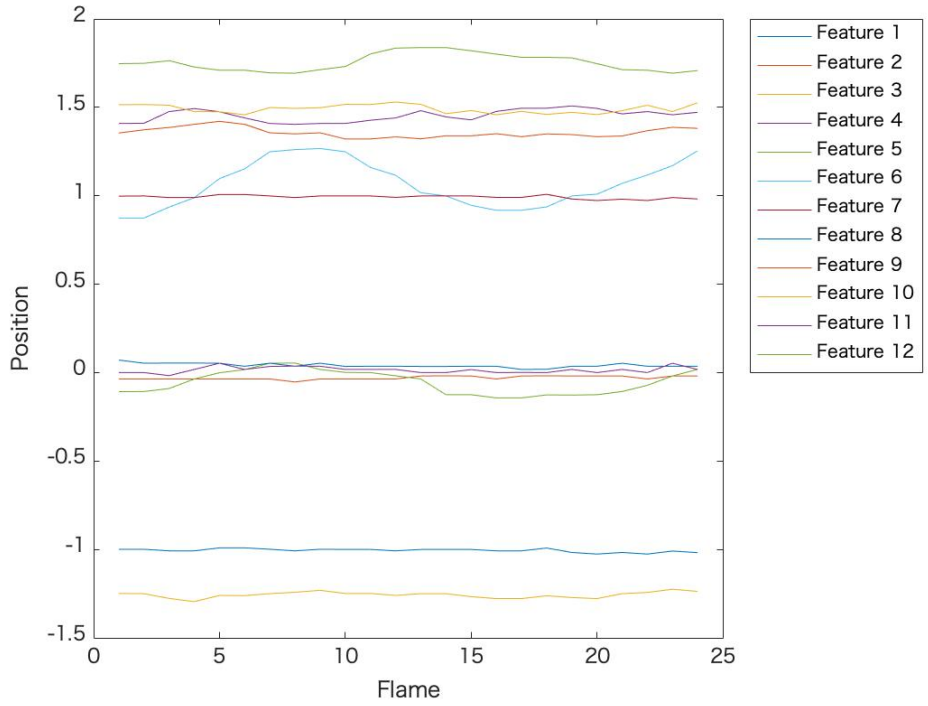


(b) サンプル 2

図 4.6 手話動作 1 各特徴量の時間変化

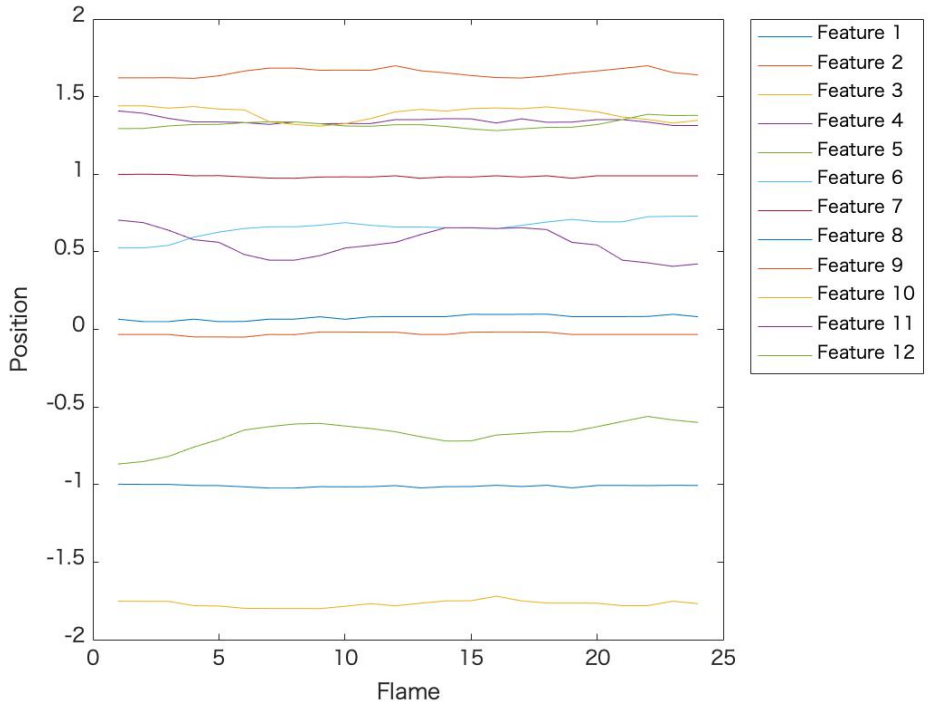


(a) サンプル 1

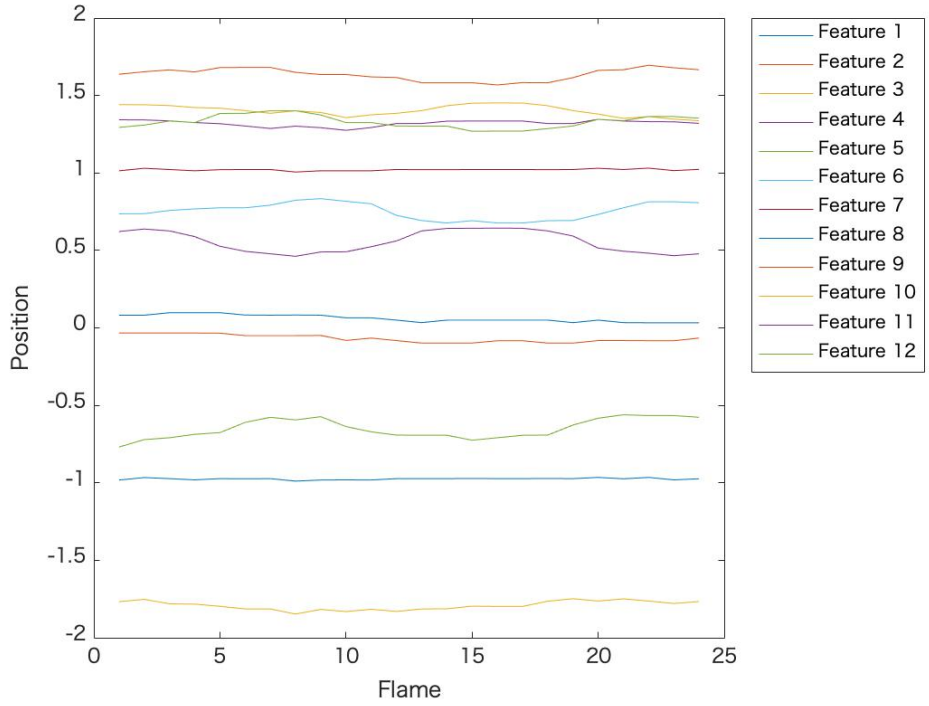


(b) サンプル 2

図 4.7 手話動作 2 各特徴量の時間変化

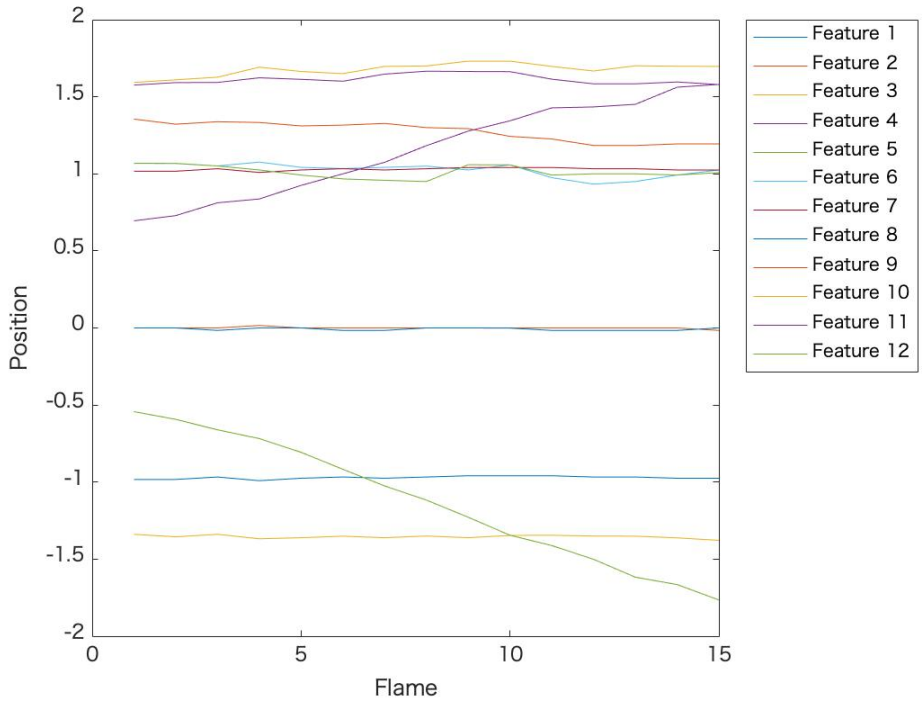


(a) サンプル 1

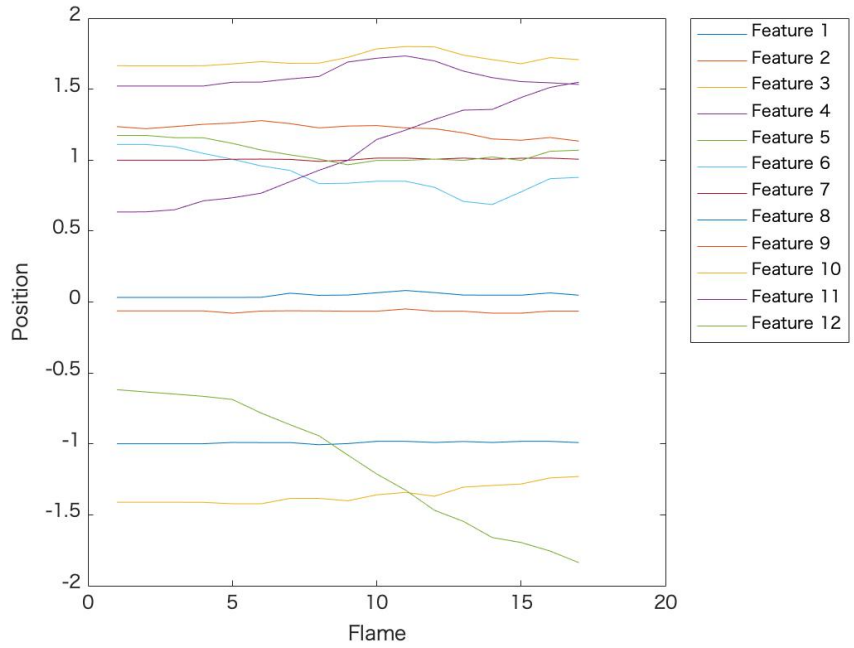


(b) サンプル 2

図 4.8 手話動作 3 各特徴量の時間変化

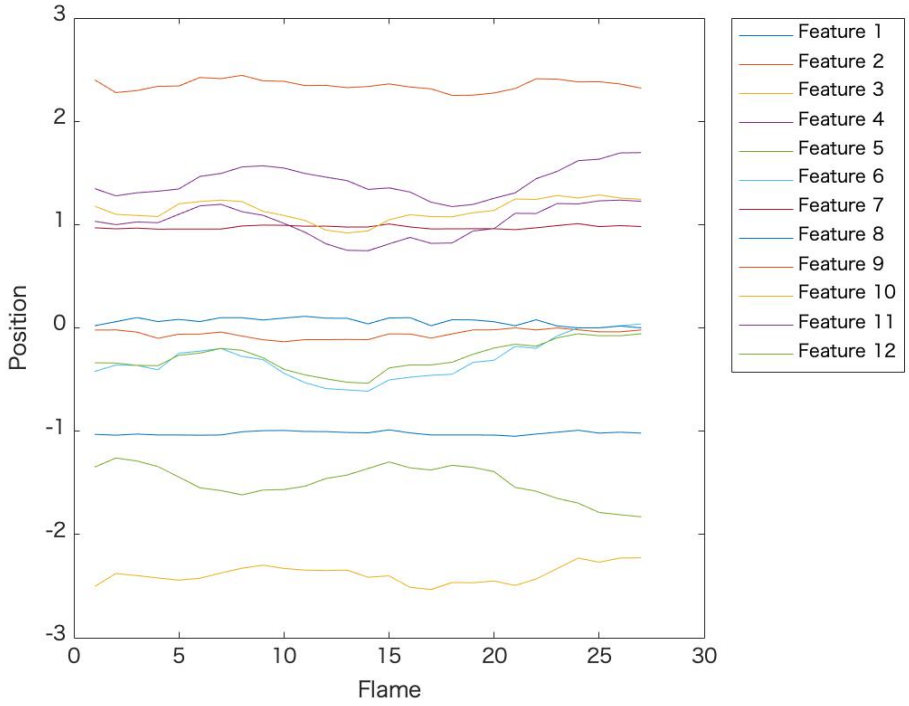


(a) サンプル 1

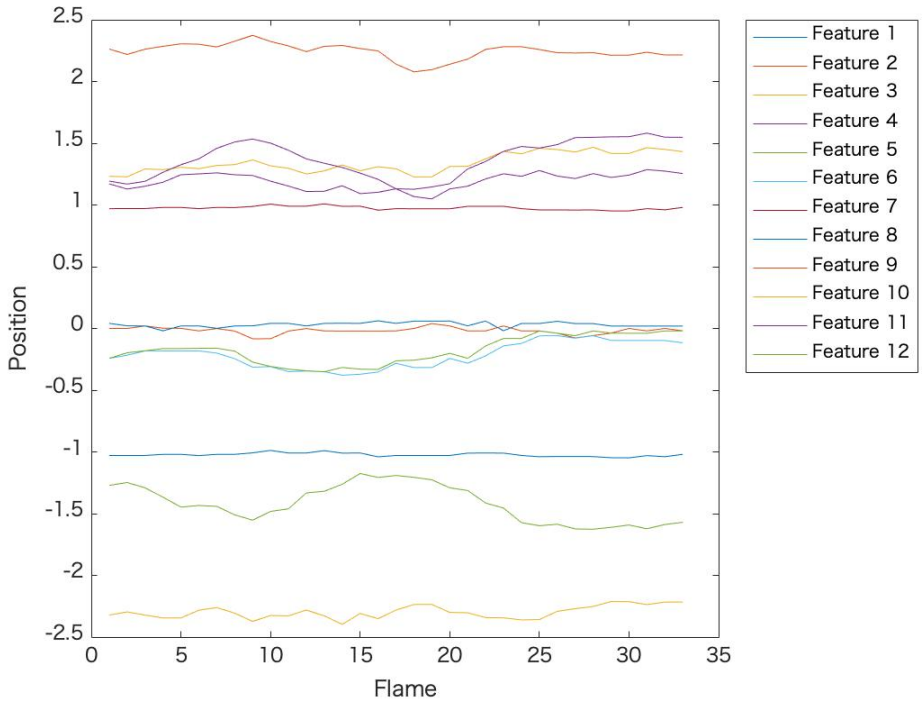


(b) サンプル 2

図 4.9 手話動作 4 各特徴量の時間変化



(a) サンプル 1



(b) サンプル 2

図 4.10 手話動作 5 各特徴量の時間変化

4.4 実験結果

4.3 で示した系列データに **3.5** で示した LSTM の機構を適用することで手話動作の分類を行った．採取した 9 人のデータのうち，8 人分を学習用データ，1 人分をテスト用データとする 9 交差検定を行ったところ，表 4.1 に示す結果が得られた．

表 4.1 交差検定の結果

真値＼予測値	動作 1	動作 2	動作 3	動作 4	動作 5
動作 1	445	0	2	4	0
動作 2	2	444	5	0	0
動作 3	53	0	407	0	0
動作 4	3	0	2	420	16
動作 5	16	0	0	7	425

また，動作ごとの適合率 (Precision)，再現率 (Recall)，F 値 (F-measure) を式 (4.1)～(4.3) より求めた．ここで，TP_A は動作 A であると正しく判定されたもの，$FN_A(B)$ は，誤って動作 A と判定されたが実際は動作 B であるものを表す．

$$Precision_{(i)} = \frac{TP_i}{TP_i + \sum_{k \neq i} FN_i(k)} \tag{4.1}$$

$$Recall_{(i)} = \frac{TP_i}{TP_i + \sum_{k \neq i} FN_k(i)} \tag{4.2}$$

$$F - measure_{(i)} = \frac{2 * Precision_{(i)} * Recall_{(i)}}{Precision_{(i)} + Recall_{(i)}} \tag{4.3}$$

式 (4.1)～(4.3) を用いて各動作の適合率，再現率，F 値を求めた．さらにそれらを平均したマクロ平均を求めた．これを表 4.2 に示す．

表 4.2 動作毎の適合率，再現率，F 値

	適合率	再現率	F 値
動作 1	0.857	0.987	0.917
動作 2	1.00	0.984	0.992
動作 3	0.978	0.885	0.929
動作 4	0.974	0.952	0.963
動作 5	0.964	0.949	0.956
マクロ平均	0.955	0.951	0.951

また，全動作を対象とした認識率を式 (4.4) から求めたところ，95.11%を得た．

$$Accuracy = \frac{TP}{TP + FN} \tag{4.4}$$

4.5 むすび

本章では，本研究で用いた動作，特徴量の時間変化，分類の結果について述べた．実験により 5 種類の手話動作に対して認識率 95.11% が得られた．

第 5 章 結論

5.1 結論

本研究では単眼 RGB カメラを用いた手話の動作解析を OpenPose と LSTM を用いて行った．実験の結果，5 種類の手話動作に対して認識率 95.11% が得られた．

5.2 今後の課題

本研究では腕の動きの情報のみを用いた．しかし，単語数が増えた場合に手の座標情報も含める必要がある．その場合，特徴量の次元が増大し，認識精度が向上しない問題点があると考えられる．手の情報や頷き，表情の情報を統合し，認識精度を向上させることが今後の課題である．

謝辞

本研究の必要機材や快適な研究環境を与えてくださり，研究生活に加えて学生生活や社会のことなどについても熱心なご指導を頂いた渡辺教授に心から感謝いたします.

常日頃から様々な助言やご提案を頂き，研究の心構えやヒントを与えて下さった早稲田大学国際情報通信センターの石川孝明様に心から感謝いたします.

また研究に際して様々な意見やアドバイスを下さった研究室の皆様にお礼申し上げます.

最後に，私をここまで育てて下さった家族に感謝いたします.

参考文献

[1] 厚生労働省，“身体障害児・者等実態調査 : 結果の概要”，

https://www.mhlw.go.jp/toukei/list/108-1b.html (2019 年 2 月 6 日アクセス)

[2] 一般財団法人 全日本ろうあ連盟，“手話の言語性 法規定なる”，

https://www.jfd.or.jp/2011/08/05/pid6302 (2019 年 2 月 6 日アクセス)

[3] 大分聴覚障害者センター 社会福祉法人 大分聴覚障害者協会，“聴覚障害者への配慮に対する質問”，http://www.toyonokuni.jp/13_02.html (2019 年 2 月 6 日アクセス)

[4] 社会福祉法人 聴力障害者情報文化センター，“手話翻訳者名簿”，

http://www.jyoubun-center.or.jp/slit/list/ (2019 年 2 月 6 日アクセス)

[5] 岩峪，曽我，瀧，“データグローブを使用した手指動作スキル学習支援システム”，電子情報通信学会技術研究報告 ET2014-63，Vol.114，No.305，pp.13-17，Nov. 2014

[6] 福原，森，“手話トレーニングマシンの開発-手話単語データベースの拡充-,” Technical Reports on Information and Computer Science from Kochi Vol.7，No.2，pp.1-5，Mar. 2015

[7] 北川，森，“Kinect を用いた手話トレーニングマシンの開発”，Technical Reports on Information and Computer Science from Kochi Vol.5，No.6，pp.1-2，Mar. 2013

[8] 室屋智和, “Leap Motion を用いた手話学習システムの開発,” 九州工業大学 総合システム工学科，卒業論文，Feb. 2018

[9] 外山，宮崎，“ステレオカメラを用いた指文字の認識手法”，情報処理学会全国大会講演論文集，Vol.70，No.2，pp.107-108，Mar. 2007

[10] 山田，松尾，島田，白井，“手話認識のための見えによる手領域検出と形状認識”，画像の認識・理解シンポジウム(MIRU2009)，ISI37，pp.635-642，July 2009

[11] 柴田，西村，田中，小林，岩本，加藤，“カラー手袋を装着した実手話認識に向けた動作者差異の類似度分析”，FIT2015，Vol.14，No.3，pp.551-554，Aug. 2015

[12] Z. Cao, T. Simon, S-E Wei, Y.Sheikh, “Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields,” In 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), No.121, pp.1302-1310, July 2017.

[13] CMU-Perceptual-Computing-Lab/openpose, https://github.com/CMU-Perceptual-Computing-Lab/openpose,Github,inc.(2019 年 2 月 6 日アクセス)

図一覧

表一覧

研究業績

[1] 柳澤, 石川, 渡辺, "単眼 RGB カメラを用いた手話動作の解析",電子情報通信学会総合大会, Mar.2018(発表予定)