

類似手話語彙の平面データによる光学的識別法 (1)

—特徴点の抽出と遷移の検証—

The Optical Recognition of the plane data of similar signs:
Extracting the remarkable points and examining transition木村 勉* 神田 和幸*
Tsutomu Kimura Kazuyuki Kanda

1. はじめに

筆者らは OpenPose[1]から得られた x-y 座標の 2 次元情報を用いてディープラーニングによる手話認識を行なった[2]. その結果, 75%の認識率が得られたが, 認識率が低いデータを見ると奥行き情報 (z 軸) のデータ不足が認識に影響を与えたのでは無いかと結論づけた. ステレオカメラを用いれば解決できるかもしれないが, ステレオカメラを用意するのは, 研究成果を広く一般に普及させることを考えるとあまり現実的ではない. そこで筆者らは現状の 2 次元データから類似手話の識別が可能かどうかを考察してみることにした.

古くから 2 次元画像から 3 次元情報を推定する手法は数多く研究されてきた. 近年はディープラーニングを用いて推定する技術がいくつか研究されている[3]. また, OpenPose から 3 次元情報を推定する方法もいくつか考案されている[4]. OpenPose からの 3 次元情報推定を手話動画に対して適応してみたが, 正面からの撮影で, ほぼ腕の動きのみという制約のためか, 手首が背中に突き抜けてしまうなど, うまく奥行き情報を得ることができなかった.

一般論として, ある物体の移動において, 上下左右の移動が一切なく, 前後の移動しかない状態 ($\Delta x=0, \Delta y=0$) というのは極めて特殊である. とくに人間の行動を録画したビデオにおいて, こういう状態はほぼないと考えられる. 人間の行動を録画した場合, 前後差は物体の面積の違いや重なりで容易に識別できる. OpenPose は関節の x-y 座標のみを識別するため, 移動体の面積差, 重なりを検出することができない. しかし, 人間の関節間の距離は不変であるため, 前後差があれば見かけ上, 関節間の距離が変わるはずである. その微妙な差分を認識できれば問題が解決できるかもしれないと考える.

本論では, 手始めに前後差分が他軸の差に顕著反映されることがないのか調べる.

2. 類似手話語彙

筆者らが行なった手話認識[2]では, 手話技能検定試験 6 級 101 単語を対象とし, 使用したデータ数は 7,763 個で, 手話表現の始まりから終わりまで, 人の肩から手首までの関節と両手の関節部分を取得し, それぞれの点において x-y 座標を 1 フレームごとに取得したデータの集合を CSV 形式ファイルとした. データの約 90%を学習データ, 残りの約 10%をテストデータとしてクロスバリデーションを行った結果, 認識精度は約 75%となった. 識別率が低かった単

* 豊田工業高等専門学校 National Institute of Technology, Toyota College

† 国立民族学博物館 National Museum of Ethnology

語は, <兄><姉>と<妹><娘>などであった. 図1にそれぞれの手話表現を示す.

<兄>と<姉>は手首を上へ上げる動作で, 手型が異なる (<兄>は中指, <姉>は小指のみを立てて, 手の甲を相手に向ける). <妹>は<姉>と手型は同じであるが, 手首を下げる動作である. <娘>は, 手首を下ろしながら, 前に出す動作が加わる.

伝統的手話学の枠組みで考えるなら, <兄>と<姉>の識別率が低いのは, 位置と動きが同じであるから, 手型の違いだけであるためだと考えられるが, <妹>と<娘>では手型は同じで, 位置と動きが違う. つまりこれらのペアの認識率の低さはそれぞれ別の原因に起因していると考えられる. 本論では, この原因に着目し, 同じデータの分析を試みる.

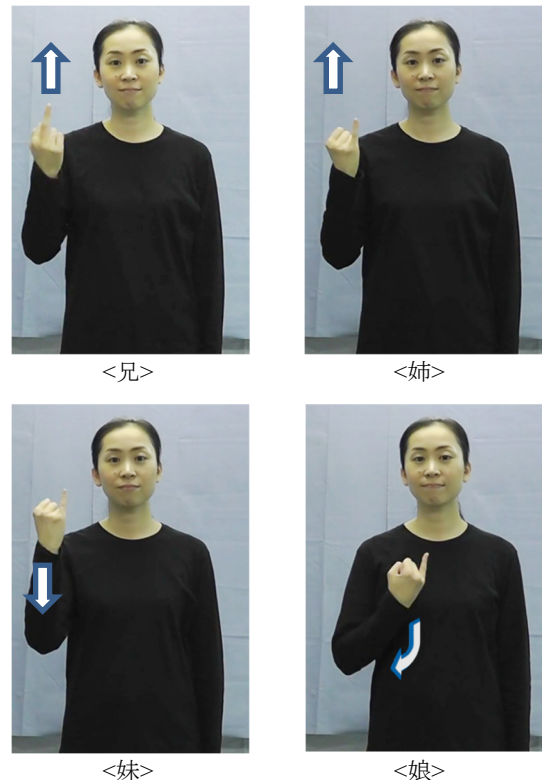


図 1 手話表現

2.1 姉の動作

一例として<姉>の右腕の関節点 (肘, 手首) の動き (位置座標の変位) を図 2 に示す. ホームポジション (HP: 気をつけの姿勢) の各関節位置を 0 とし, そこからの相対座標 (画像上のピクセル値) を縦軸, フレーム数 (動画は 30FPS) を横軸にして, 相対位置の変位をグラフ化した.

最も顕著な動きを見せたのは手首 y 軸であり、次が肘 y 軸である。基本的には上下運動なので、x 軸の動きは少ない。また、グラフに示さなかった他の関節は付随動作として微妙に動く程度である。

ここで2つの仮説を立ててみる。

- ・この手話は片手手話（利き手=右手）であり、その動きは当然手首に表れると考えられる（仮説1）。
- ・また肩と肘と手首の各2点間の距離に特徴が出る可能性を示唆している（仮説2）。

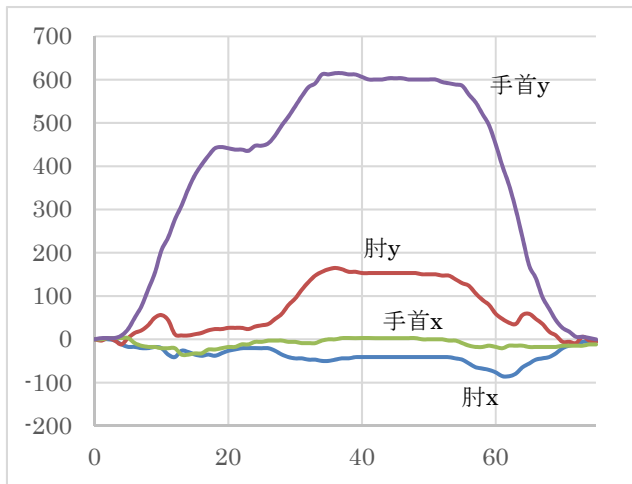


図2 <姉>の動作分析

2.2 兄と姉の比較

上記の仮説を確認するため、<兄>と<姉>という異手型の類似手話の動きを比較する。<兄>の右腕の動きを図3に示す。これを見ると姉と同様に手首と肘に顕著な動きがある。また図2と図3のグラフからほぼ同じ動きをしていることが確認できる。

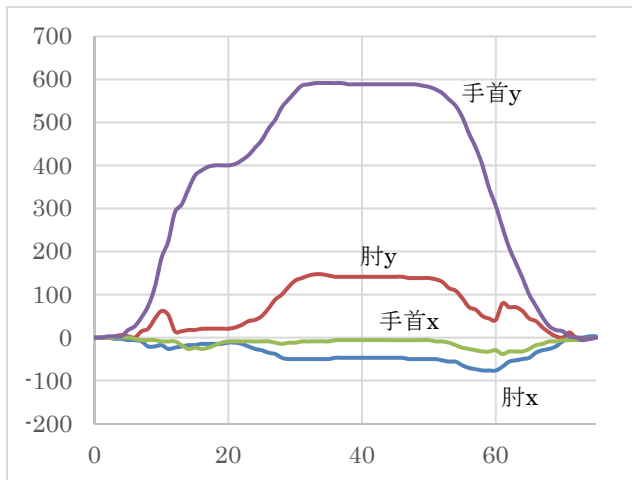


図3 <兄>の動作分析

図2と図3から確認できることは手首と肘に顕著な動きがみられることである。図2と図3からは手首yの動きが一番大きな差が見られ、次に肘yに差が出ている。ここで本来真っ直ぐに上に移動しているのであれば、同じ距離だけ移動するはずであるが、手首が前後方向に移動しているため、同じ距離ではない。最終的な結論ではないが、この時点でy方向に差が出ていることがわかる。単純化すれば、

左右方向よりも上下方向に差が出やすいらしいという仮定が成り立つ。これは人間の観察に一致する。

2.3 息子、娘の比較

同様の比較を<息子>と<娘>にも適用してみる。<息子>は<娘>と同じ動きで手型が異なる（親指だけを立てる）。右手首の違いを表したのが図4である。

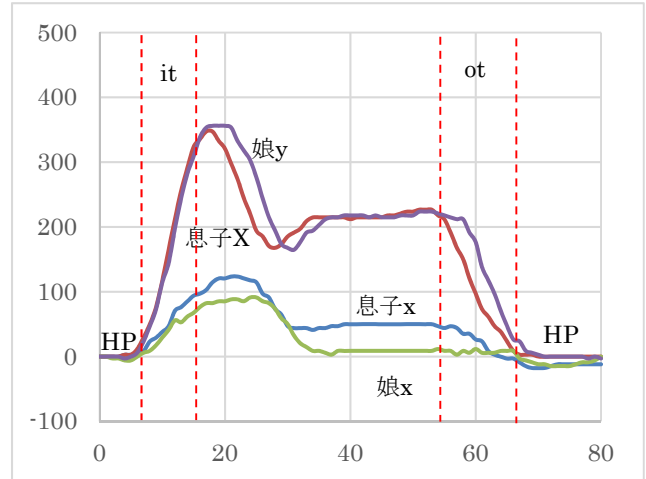


図4 <息子>と<娘>の動作分析

両者に動作速度の違いによる多少のズレはあるものの全体の動きとしては同じであると言える。さらに細かく見てみると、両者ともx軸の動きは、大きく動いた後、そのままフラット状態で静止し、ホームポジションに戻っている。フラット状態の前後に微妙な運動がある。この動きはいわゆる押印動作と呼ばれる微細な上下運動で、この押印動作を人間は普通認識しない。人間が動から静止に至るために自然に発生せざるを得ない運動である。アニメーションなどでは、この運動を付加しないので、いわゆる“人間らしさ”がないことの原因の1つである。

また両者比較の結果、共通するのは、移動と静止という2つの運動要素から構成されている点である。移動運動はいわゆる遷移であり、手話者は意識しない。静止のみを意識している。遷移 transition には入り渡りと出渡りがあり、筆者らは in-transition=it と out-transition=ot と命名している。図4の2つの点線の間が、おおよその it と ot である。

この手話者の it はホームポジション (HP) から図1<娘>の姿勢まで一旦大きく上昇方向に動いている。そこから静止位置まで下降移動していくがこれは遷移ではなく、「産む」という意味をもつ。静止が手型を提示し、終了後、ot として下降している。この一連の流れを言語学的に記述すると

入り渡り (it) + #産む + #男 or 女 + 出渡り (ot)

となり、両者は手型において対立する形態素結合連鎖ということになる。いわゆる最小組 minimal pair である。ここでは<産む>という語彙については考察しないが、<息子><娘>における#産むは運動のみの拘束形態素であることが示されている。<産む>を示すには、男女とはまったく異なる手型で表現される。位置とタイミングについてはここで示したように微妙なずれがあり、伝統的手話学という「位置は同じ」にはなっていない。ここが人間の言語観と機械認識の差である。

2.4 娘と妹の比較

次に<娘>と<妹>の右手首の動きを比較する。図 5 に示す。<娘>と<妹>では y 軸はやや移動距離が違うもほぼ同じである。これはどちらも上から下への直線運動であるからである。x 軸は若干異なる。図 1 を見ると手話開始時の手首の位置が、<妹>が肘と同じ横位置であるが、<娘>は胸の前である。これはお腹から「産む」という動作を行うためにこの位置に持ってきたのである。図 5 の赤の縦点線からしばらく軸、y 軸ともにフラットになっている。この静止状態は手型の提示である。つまり、これ以前で一連の動作が終了している。<娘>は前に出す動作があるためわずかに x 軸に変化があるが、この動きを見極めることが難しい。つまり高橋他(2019)[2]はこの違いを検出できなかったといえる。

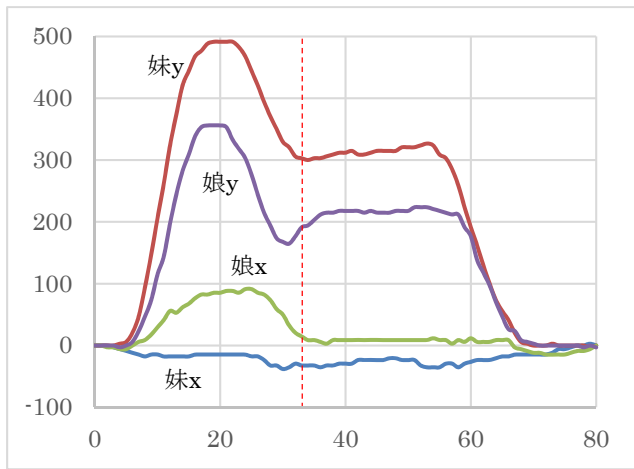


図 5 <妹>と<娘>の動作分析

3. 距離の比較

2 番目の仮定である「肩と肘および肘と手首の 2 点間の距離に特徴が現れる」ことを検証する。

図 6 に<兄>と<姉>の各関節点の x-y 座標における 2 点間の距離の変化を示す。

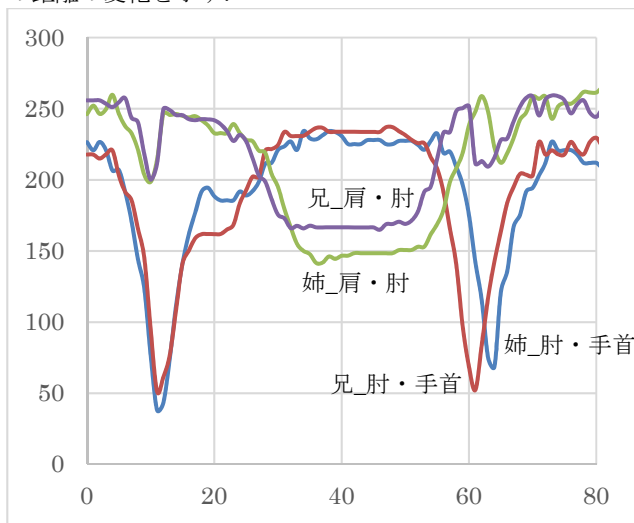


図 6 肩・肘・手首の 2 点間の距離の変化

この図からは<兄><姉>共に距離の変化パターンが同じであるということがわかる。<兄>について、上下左右と距離の変化を示したのが図 7 である。

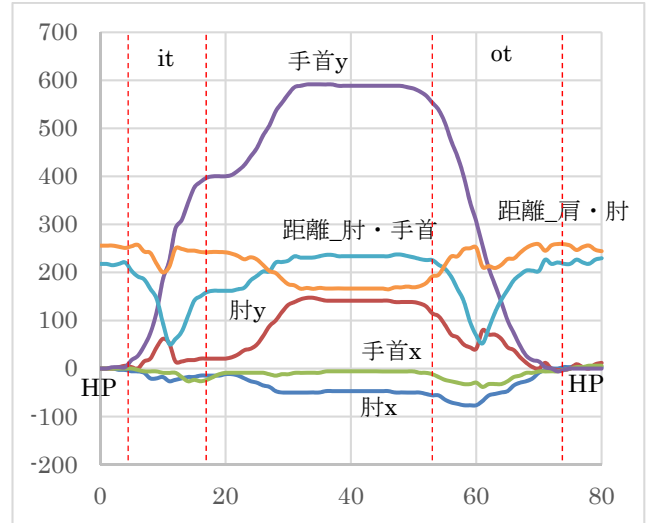


図 7 <兄>の手首, 肘, 距離の変化

図 7 から肘と手首の距離は遷移開始前 (入り渡り, it) と遷移終了後 (出渡り, ot) で変化している。ホームポジションでは、手首と肘の距離は、腕の長さと同じであるが、it の間は、手首が若干前に出るため、x-y 座標では見かけ上短くなる。it は図 1<兄>の体制まで持っていくため、距離は腕と同じ長さに戻る。手話動作中は、手首と肘をそのまま下げるため、距離に変化はほぼない。ot の間は、ホームポジションへ戻る動きになる。そのため見かけ上、距離が短くなる。

肩と手首の距離は、it と ot で変化している。また手話動作中は、肩と肘が重なるため、見かけ上、距離は短くなる。

次に<息子><娘>と<妹>について同様の分析を行ったのが図 8 である。

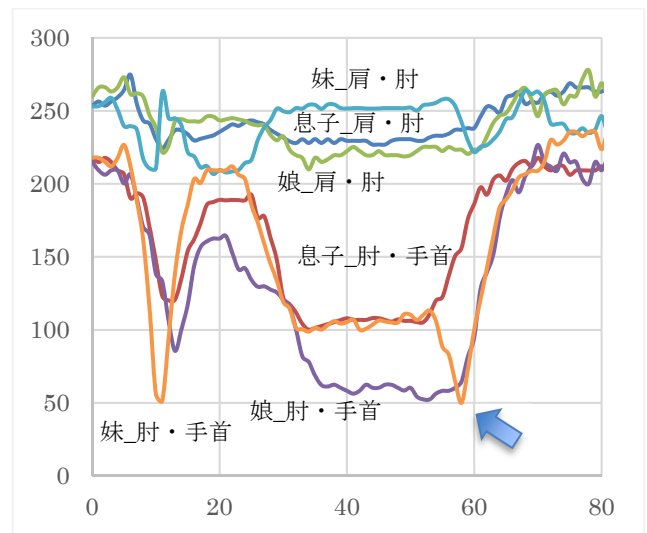


図 8 <息子><娘><妹>の距離の比較

図 8 の変化パターンを見ると<息子>と<娘>の動きの変化は若干の距離の違いはあるものの、ほぼ同じである。<妹>との違いは、肘と手首の距離においては図中の矢印で示した部分である。<兄>や<姉>と同じように ot の間に距離の変化が現れている。しかし、<息子>と<娘>は前に出す動作が加わっているため、この現象が見られない。また、肩

と肘の距離では、<息子><娘>と<妹>では、全体的にパターンが異なっている。言い換えると、<妹>は<兄><姉>と同様、手話者の意識上は上下運動だけであり、前後運動はないが、<息子><娘>は腹という位置と前後運動を意識しているの、その意識差が微妙ではあるが、距離の差に反映されているといえる。

以上の考察から、距離変化に特徴が現れていると推定できる。

4. 考察

本論では次の仮説を検証した。

仮説 1 片手手話では手首に変化が大きく表れる

仮説 2 肩と肘と手首の各 2 点間の距離に特徴が現れる

仮説 1 については、手首の上下差分に変化が大きく表れることが検証された。

仮説 2 については、肩と肘、および肘と手首の距離が遷移部分の変化に反映される可能性が検証された。

本論では、類似手話と判定される<兄><姉><妹><息子><娘>について、前後変化を考慮に入れず、①上下左右の変化パターンと②手首と肘の距離の変化によって差が判定できるかどうかを検証し、上記の結論を得た。また、<弟>という手話も手型が異なるだけで<妹>と同じ動きであるため、同様の動作分析となる。

この 6 手話単語について、従来は手型と位置により下記のように成分分析されてきた。

表 1 伝統的手話分析

語	手型	位置	動き
兄	男	中立	上
姉	女	中立	上
弟	男	中立	下
妹	女	中立	下
息子	男	腹	前
娘	女	腹	前

非常にシンプルで対立関係もはっきりしているため、直観的にわかりやすい。手話学習においてはこの分析が向いている。人間の経験に基づいた直観的認識はこうなのであろう。しかし機械認識でこの成分を表現するとなるとこのように単純にはならない。本論のように OpenPose を用いて分析しようとする、<息子><娘>で示したように、成分をまったく別のものである関節の座標と軌跡に変換しなくてはならない。とくに 1 眼によるビデオ分析では前後が測定できないという制約もある。本論では神田・木村(2019a)[5]の枠組みを活用し、像素である動体として肩、肘、手首を設定、その座標から肩、肘、手首の各距離を計算し、時間的な軌跡の変化を検出し、速度をグラフのパターンで表示することで手話の動きを測定した。この分析により、表 1 には記述されていない遷移の重要性を指摘することができた。上下運動が y 軸の変化で示されるのは予測通りであったが、前後位置関係も遷移の時間的前後により距離パターンが変化することから、本論の範囲での比較においては推測可能となった。

また、OpenPose から得られたデータを基に 3D 化を行う手法は、すべての関節情報について推定しており、時間もかかる。そこで仮説を基に手話に特化して前後関係を推定できれば処理速度も早くなると考える。

今後の課題として、本論は 1 例のモデル手話の分析に限定されており、多くの学習者に置いても敷衍できる原理かどうかの検証が必要である。またモデル手話においても、他の手話語彙においてこの原理が成立するかどうかの検証が必要である。また片手手話と両手手話では異なることも想定できる。本論はそれらの分析の試金石としての位置である。

これまで筆者らがアプローチしている機械学習の手法では、上半身の関節点をすべて使用しているが、本論の結果から、特徴がある関節点だけを学習データとした方が、精度が上がるかと予想できる。

謝辞

本論は文部科学省科学研究費補助金課題番号 18K18517 (代表者木村勉) 及び課題番号 18K18518 (代表者神田和幸) による研究成果の一部である。

参考文献

- [1] OpenPose , <https://github.com/CMU-Perceptual-Computing-Lab/openpose> (2019年6月8日閲覧)
- [2] 高橋 佑太, 木村 勉, 神田 和幸, “機械学習を用いた手話認識に関する研究”, 電子情報通信学会技術研究報告, Vol.118, No.440 (2019).
- [3] Ariel Gordon, Hanhan Li, Rico Jonschkowski, and Anelia, Angelova, "Depth from Videos in the Wild: Unsupervised Monocular Depth Learning from Unknown Cameras", <https://arxiv.org/pdf/1904.04998.pdf> (2019年6月8日閲覧)
- [4] MMD モーショントレース自動化への挑戦 , <https://qiita.com/miu200521358/items/d826e9d70853728abc51> (2019年6月8日閲覧)
- [5] 神田 和幸, 木村 勉, “描素の構造—新手話学の演繹的アプローチ—”, 電子情報通信学会技術研究報告, Vol.118, No.440 (2019b).