H-009

姿勢・手指形推定を活用した手話動作類語の単純な同定法
A Simple Method to Identify Similar Words with Respect to Motion in Sign Language Using Human Pose and Hand Estimations

田中 省作[*1]　本田 久平[*2]　バイティガ ザカリ[*3]
Shosaku Tanaka　Kyuhei Honda　Zacharie Mbaitiga

1. はじめに

音声言語で綴りが似ていても意味が全く異なるような語があるように，手話にも腕や手指の動作がよく似ているものの，表す内容は異なるような語 (表現) がある．本稿では，このような語を，ある語に対する**手話動作類語**，あるいは単に**動作類語**と記す[*4]．このような手話動作類語に関する網羅的な情報は，手話学習における重要な資料である．そこで，このような組み合わせを，人の姿勢・手指形の自動推定プログラムを活用し，手話の映像から網羅的に探すための単純な方法を提案する．

本研究ではまず，[3] などの一般的な手話の辞典で着目する腕・手指形とその動きを対象とし[*5]，人の姿勢・手指形の自動推定プログラム (OpenPose，MediaPipe) を活用した手話表現の検索手法を考える．カメラや映像中の手指や腕などの部位を前述のプログラムでフレームごとに認識する．腕や手指やその動きをベクトルで表現し，手指動作の類似度を定量化する．そして，手話の腕・手指の動作や形が似ている動作類語の同定を行う．

2. 腕・手指形のベクトル表現

本研究では，姿勢・手指形の推定プログラムとして OpenPose と MediaPipe を併用する [1,6]．これらは映像のフレームごとに人の主要部位 (顔の各部位，肩，肘，手指の各部位など) を自動認識する．たとえば，MediaPipe は人の姿勢 (Pose) に関して 33 点，手指形 (Hands) に関して 21 点左右で 42 点を認識し，奥行きまで含めた 3 次元の相対座標を提示する．ただし，誤認識や，部位が認識されない (認識もれ) こともあること，奥行きについては誤差が小さくないことに留意しなければならない．

認識されたフレーム内の人の主要部位をもとに，静止した一つの手話表現 x を，左右 (L, R) それぞれで次の 3 つの情報によって特徴づける．

(1) **腕の位置**　肩を原点とした肘・手首の 2 つの 3 次元位置ベクトルを並べた 6 次元ベクトル: $v_x^{(1,d)}$ $(d \in \{\mathrm{L},\mathrm{R}\})$

(2) **指の位置**　手首を原点とした 5 指それぞれの指先の 5 つの 3 次元位置ベクトルを並べた 15 次元ベクトル: $v_x^{(2,d)}$ $(d \in \{\mathrm{L},\mathrm{R}\})$

(3) **掌の向き**　人差指の根本 (CMC) を原点とした小指の根本への 3 次元位置ベクトル: $v_x^{(3,d)}$ $(d \in \{\mathrm{L},\mathrm{R}\})$

静止した手話表現 x は 1-3 の 3 特徴，左右で延べ 6 特徴のベクトルの組 $\boldsymbol{v}_x$ で表現し，簡単化のため i 組目のベクトルを $v_{x,i}$ で記述する．

$$\begin{aligned}\boldsymbol{v}_x &= \langle v_x^{(1,\mathrm{L})}, v_x^{(2,\mathrm{L})}, v_x^{(3,\mathrm{L})}, v_x^{(1,\mathrm{R})}, v_x^{(2,\mathrm{R})}, v_x^{(3,\mathrm{R})}\rangle \\ &= \langle v_{x,1}, v_{x,2}, v_{x,3}, v_{x,4}, v_{x,5}, v_{x,6}\rangle\end{aligned}$$

フレーム内にもともと部位が映っていなかったり，認識もれしていたりして，$\boldsymbol{v}_x$ には値が未定となる $v_{x,i}$ が含まれることがある．つまり，$\boldsymbol{v}_x$ は一般に欠損があるベクトルである．x に対して欠損していない，つまり値が定まっている組のインデクスの集合を $I(\boldsymbol{v}_x)$ で表す．

3. 腕・手指形と動作の類似度

2 つの，フレーム内の静止した手話表現 x, y 間の類似度を次のように与える．

$$\mathrm{sim}(x,y) = \frac{1}{\sqrt{|I(\boldsymbol{v}_x)\cap I(\boldsymbol{v}_y)|}} \sum_{i\in I(\boldsymbol{v}_x)\cap I(\boldsymbol{v}_y)} w_i \frac{v_{x,i}\circ v_{y,i}}{\|v_{x,i}\|\|v_{y,i}\|}$$

$a \circ b$ は a と b の内積，$\|a\|$ は a のノルムを表す．w_i は $w_i \geq 0$, $\sum_{i\in I(\boldsymbol{v}_x)\cap I(\boldsymbol{v}_y)} w_i = 1$ を満たす，特徴ごとのコサイン類似度に対する重みである．

表現動作はフレーム列である．動作を伴う手話表現に対応する，m つのフレームから成るフレーム列を $\boldsymbol{X} = [x_1, x_2, \cdots, x_m]$，長さを $\ell(\boldsymbol{X})(= m)$ と表すことにする．$\ell(\boldsymbol{X}) \leq \ell(\boldsymbol{Y})$ のフレーム列で表される手話表現間の類似度は，次のように与える．

$$\mathrm{SIM}(\boldsymbol{X},\boldsymbol{Y}) = \max_{u=1,2,\cdots,\ell(\boldsymbol{Y})-\ell(\boldsymbol{X})+1} \sum_{t=1}^{\ell(\boldsymbol{X})} \mathrm{sim}(x_t, y_{u+t})$$

$\boldsymbol{Y}$ すべての部分フレーム列で計算し，その最大値を類似度としているので，結局，$\boldsymbol{X}$ と最も類似している $\boldsymbol{Y}$ の部分フレーム列で測っていることになる．

[2,4,5] に付属する DVD 映像に対して，OpenPose を適用し，映像内の人の範囲を同定し，延べ 2,338 発話分の映像を人ごとに切り出した．そののち発話ごとに MediaPipe の Holistic[*6] と Hands を適用し，姿勢・手指形情報を得て，腕・

[*1] 立命館大学文学部 College of Letters, Ritsumeikan University
[*2] 大分工業高等専門学校電気電子工学科 Department of Electrical and Electronic Engineering, National Institute of Technology, Oita College
[*3] 沖縄工業高等専門学校メディア情報工学科 Department of Media Information Engineering, National Institute of Technology, Okinawa College

[*4] 1 つの語を固定したときに「互いに動作が似ている」という関係が成立し得るので，動作類語の対や組と記す方が正確であろう．一般には，そういった組を単に類語と記すことも多く，本稿でもそれにならう．

[*5] 手話による実際のコミュニケーションでは，表情や顔の傾き，視線や口形 (口型) も発話の意味を定める重要な言語要素である．たとえば，§5 の「うらやましい」と「おかしい」は，口形や表情が重要な峻別の情報となる (図 2)．

[*6] Holistic は，姿勢・手指形・顔の主要部位を認識するもので，Pose を包摂している．

手指形ベクトルとそのフレーム列を構成する．手指形を得る際には，Holistic と Hands それぞれで推定された手指形情報を統合している．

検索キー映像も，カメラなどで撮影した手話映像に対して上記と同じ手続きを施し，検索のためのフレーム列を作成することになる．ここでは上記の DVD 内の映像の一部を検索キー映像とした例を示す．「若い」という表現の特徴的な動作 (右手をおでこの前で左から右に動かす) で，長さ 30 のフレーム列 (図 1 左) をキー映像とした．事前の試行から，今回，奥行きは未使用 (すべて同値で上書き) とした．その上で，すべての特徴を平等に勘案するよう重みをすべて 1/6 で検索すると，最も類似した表現に「黒い (図 1 真ん中)」(類似度は 2.38)，その次に「高校 (図 1 右)」(類似度は 2.20) が得られる．また重みを右腕形に限ると ($w_4 = 1$，それ以外は 0)，「高校」(類似度は 2.28) が 2 番目，「黒い」は上位 20 語にも含まれなくなる．

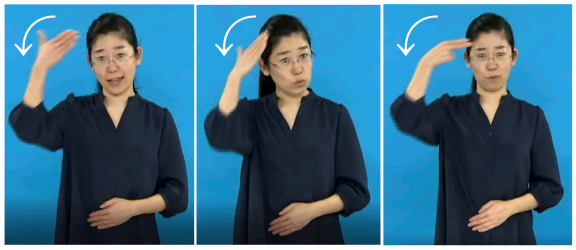

図 1　左から「若い・黒い・高校」の手話表現

4. 動作類語検出への拡張

手話の語 x, y が手指動作類語であるための必要条件として，

x, y に類似した (部分的な) 腕・手指動作が存在する

をおく．たとえば，図 1 の 3 語には，手指形もしくは腕の動きが部分的に非常に類似する部分フレーム列がある．この必要条件を念頭に，$\boldsymbol{X}, \boldsymbol{Y}$ それぞれの部分フレーム列を比較し，類似した手指動作の有無を判断するための指標を考える．$\ell(\boldsymbol{X}) \leq \ell(\boldsymbol{Y})$ の手話表現 $\boldsymbol{X}, \boldsymbol{Y}$ に対する類似度を次のように与える．

$$\begin{aligned}&\mathrm{SIM}'(\boldsymbol{X}, \boldsymbol{Y}\,; \alpha, \beta)\\ &= \max_{i \in C_{\beta,\ell(\boldsymbol{X})}} \left[\max_{u=1,2,\cdots,\ell(\boldsymbol{Y})-\ell(\boldsymbol{X})+1} \sum_{t=i}^{i+\alpha} \mathrm{sim}(x_t, y_{u+t}) \right]\end{aligned}$$

$$\begin{aligned}&C_{\beta,\ell(\boldsymbol{X})}\\ &= \{1 + c\beta \in \mathbb{N} \mid 0 < 1 + c\beta \leq \ell(\boldsymbol{X}) - \alpha + 1 \wedge c \in \mathbb{N}\}\\ &\quad \cup \{\ell(\boldsymbol{X}) - \alpha + 1\}\end{aligned}$$

$\boldsymbol{X}$ から長さ α の部分フレーム列を β ごとにすべて切り出し，$\boldsymbol{Y}$ 上を走査し，類似した部分フレーム列を探す．

α が小さいと細かな動作が類似度に反映される一方，組み合わせ的な動作を見落としたり，偶然的な動作の影響を受けたりする．β が小さいとより適切に類似動作を検出できる可能性が高まる一方，計算に時間がかかる．

5. 検出例

[5] に含まれる 445 語に対する組み合わせは 98,790 となる．$\alpha = 30, \beta = 2$ としたときの，SIM′ 上位 50 組の x, y に対して動作類語と判断されたものは 9 組であった．その上位 2 組の類似フレームを図 2 に示す．9 組のうち 7 組は「松」と「ねずみ」など意味的な関連性も薄く，一般的に出現頻度も低い，今まで知られていない動作類語である．

一方，動作類語として含めるべきかどうか迷うものも少なくなく，類似度上位 50 組のうち 8 組はそのようなものである．腕や手指の動作は確かに似ていても，他の腕や手指形で手話表現全体としては類似していないようにも感じられることがある．両手指・両腕が動作する表現となると，判断に迷う場合がより増えてくる．動作類語の概念構成をより厳密にする必要がある．

図 2　上左から「お金・トイレ」下左から「うらやましい・おかしい」の手話表現

6. おわりに

本研究で対象としたような表出ペースな正則な映像では，このような単純な方法でも検索や動作類語の検出が可能である．計算時間やそもそも動作類語の認定に関する問題も再確認されており，今後，効率化に取り組む．

参考文献

[1] Cao, Z., Hidalgo, G. et al.: OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields, Computer Vision and Pattern Recognition , arXiv:1812.08008 (2019).
[2] 全国手話研修センター (編): 手話奉仕員養成テキスト 手話を学ぼう手話で話そう, 全国手話研修センター (2014).
[3] 全国手話研修センター (編): わたしたちの手話学習辞典〈1〉, 全日本ろうあ連盟 (2015).
[4] 全国手話研修センター (編): DVD で学ぶ手話の本 全国手話検定試験 4 級対応 三訂, 中央法規出版 (2016).
[5] 全国手話研修センター (編): DVD で学ぶ手話の本 全国手話検定試験 5 級対応 三訂, 中央法規出版 (2016).
[6] Google MediaPipe Team: MediaPipe, https://google.github.io/mediapipe/ (Last access: 2022.6.17).