

姿勢・手指形推定を単純に活用した手話表現の検索

田中 省作[†] 本田 久平[‡] バイティガ ザカリ[§]

立命館大学文学部[†] 大分工業高等専門学校電気電子学科[‡]

沖縄工業高等専門学校メディア情報学科[§]

1. はじめに

手話は視覚言語である。日本語や英語のような音声言語とはさまざまな面で異なる。そのひとつに語や表現の検索がある。文字体系を備えた、文字で語や表現を記述できる音声言語の場合、文字を使って辞書でひいたり、用例を検索したりすることができる。一方、手話の場合、手指や腕の動きによる表現を、手話-日本語のような辞書や手話の映像のなかから効率的に探す方法がない。広く知られている[3]でも手指形で48つに大別してあるものの、結局、人の目で数十もの手話形と対照させながら探すことになる。日本語母語話者の第二言語としての英語学習でいうならば、英和辞書が欠けているような状況にある。

そこで、本研究では人の姿勢・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提案する。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し、検索する。なお、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。本研究ではまず、[3]などの一般的な手話の辞典で扱う腕・手指形とその動きを対象とする。

2. 姿勢・手指形の推定プログラム

本研究では、OpenPoseとMediaPipeを併用する[5, 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(図1)、奥行きまで含めた3次元の相対座標を提示する。ただし、誤認識のみならず、部位が認識されない(認識もれ)が起こること、奥行きについては誤差が小さくないことも

留意しなければならない。

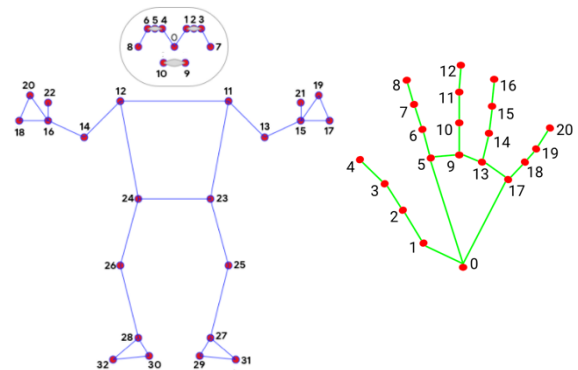


図1 MediaPipeのPose/HandsのLandmarksモデル[6]

3. 提案手法

3.1 腕・手指形のベクトル表現

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、左右それぞれで次の3つの情報によって特徴づける。

- (1) 腕 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 $v_x^{(1,d)} (d \in \{L, R\})$.
- (2) 指の位置 手首を原点とした5指それぞれの指先(tip)の5つの3次元位置ベクトルを並べた15次元ベクトル $v_x^{(2,d)} (d \in \{L, R\})$.
- (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 $v_x^{(3,d)} (d \in \{L, R\})$.

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 \mathbf{v}_x で表現する。簡単化のため組目のベクトルを $v_{x,i}$ で表すこととする。

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

また、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、 \mathbf{v}_x で値が未定となる $v_{x,i}$ がある。 x に対して値が定まっている組のインデックスの集合を $I(\mathbf{v}_x)$ で表す(e.g. 左手が

認識もれした場合, $I(\mathbf{v}_x) = \{1,2,3,4\}$).

3.2 腕・手指形と動作の類似度

2つの, フレーム内の静止した一手話表現 x,y 間の類似度を次のように与える.

$$\text{sim}(x,y) = \sqrt{|I(\mathbf{v}_x) \cap I(\mathbf{v}_y)|} \sum_{i \in I(\mathbf{v}_x) \cap I(\mathbf{v}_y)} w_i \frac{v_{x,i} \cdot v_{y,i}}{\|v_{x,i}\| \|v_{y,i}\|}$$

\cdot は内積, $\| \cdot \|$ はノルムを表す. w_i は特徴ごとのコサイン類似度に対する重みで, Σ 項はその累積和である. そして, 認識もれが少ない x,y 間の類似度をより重視するよう, 平方根を含む項が掛け込まれている.

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 $\mathbf{X} = [x_1, x_2, \dots, x_m]$ で, その長さを $\ell(\mathbf{X}) (= m)$ で表す. $\ell(\mathbf{X}) \leq \ell(\mathbf{Y})$ のフレーム列で表される手話表現 \mathbf{X}, \mathbf{Y} 間の類似度を次のように与える.

$$\text{SIM}(\mathbf{X}, \mathbf{Y}) = \max_{u=1 \sim \ell(\mathbf{Y}) - \ell(\mathbf{X}) + 1} \sum_{t=1}^{\ell(\mathbf{X})} \text{sim}(x_t, y_{u+t})$$

Σ 項は \mathbf{X} を長さ $\ell(\mathbf{X})$ の \mathbf{Y} 上の部分フレーム列に重ね合わせたときの類似度である. 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, \mathbf{X} と最も類似している \mathbf{Y} の部分フレーム列で測っていることになる.

3.3 実験

[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とHandsを適用し, 姿勢・手指形情報を得て, 腕・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部分で, 長さ30のフレーム列(図2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で検索すると, 最も類似した表現として「黒い(図2真ん中)」, その次に「高校(図2右)」が得られる. また重みを右腕形に限ると($w_1 = 1$, それ以外は0), 「高校」が最も類似した結果として検索される.



図2 左から「若い・黒い・高校」の手話表現

4. おわりに

本研究は, 手話の映像による検索を, 既存の姿・手指形推定プログラムを単純に活用して実現した. 今回の対象は映像も正則で, 表出ペースも一定であることが期待され, 実験条件は易しい. 今後, 多様な環境への適用, 実験を検討する.

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: 2021.1.5).

A Simple Method for Sign Language Motion Retrieval Using the Human Pose and Hand Estimation

† Shosaku Tanaka, Ritsumeikan University

‡ Kyuhei Honda, National Institute of Technology, Oita College

§ Zacharie Mbaitiga, National Institute of Technology, Okinawa College