

OpenPose によるキータイピング時の 骨格情報を特徴とした人物識別

奈良 悠斗[†]神奈川工科大学大学院[†]春日 秀雄[‡]神奈川工科大学情報学部[‡]

1. はじめに

指紋や静脈, キータイピングなどの身体的特徴や行動的特徴による個人識別手法は多く存在する. しかし, 専用の装置が必要であったり, プライバシーが高い情報を用いたりする必要がある. 従来の研究におけるキータイピング認証は, 主にキーストロークのリズムや, 三次元骨格情報を特徴として識別を行う. 本研究では, 姿勢推定ライブラリである OpenPose[1][2][3][4]を用いることで, 単眼カメラによる動画や画像のみで二次元骨格情報を計測し識別を行う. 専用の装置を用意することなく, パソコンに標準装備されているキーボードや, 一般的な単眼カメラから識別を行う手法の構築を試みる.

2. 個人識別手法

2.1. 骨格情報検出

本研究では, 複数人の 2D ポーズを検出するための姿勢推定ライブラリである OpenPose を使用して手の骨格情報を検出する. OpenPose は, 動画や画像, Web カメラなどを入力として, x 座標, y 座標のキーポイント(人体・手・顔・足), 信頼度 p を出力する. また, 図 1 に示すような可視化したキーポイントを表示することが可能である. 片手の骨格情報は 21 キーポイントの検出が可能であり, x, y 座標を合わせた 21×2 キーポイントと信頼度を出力できる. この時, 隠れたキーポイントなどを推測して出力するため, 信頼度が付与される. また, 左右の手をそれぞれ同時に検出可能であるため, 合計で 42×2 キーポイントが得られる. 片手の出力データは(1)式のような形で得られる.

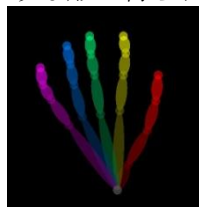


図 1 手のキーポイントの表示

$$output = \{x_1, y_1, p_1 \dots x_{21}, y_{21}, p_{21}\} \quad (1)$$

Personal identification based on skeleton information during key typing with OpenPose

[†]Yuuto Nara, Graduate School of Information and Computer Science, Kanagawa Institute of Technology

[‡]Hideo Kasuga, Department of Information Media, Kanagawa Institute of Technology

2.2. 個人識別

OpenPose によって得られた手の骨格情報を特徴量として, ニューラルネットワークによる機械学習で個人識別を行う. モデルは Keras[5]を使った全結合型 NN と CNN, SVM を使用する. 全結合型 NN では, 42×2 の入力データを 84 の要素を持つ一次元の行列に変換し, ネットワークに入力している. 中間層は, 128 ユニットで, 活性化関数を ReLU とした全結合層で構成し, 出力層では, 活性化関数を Softmax とした全結合層を用いている. 最適化アルゴリズムは Adam を使用している.

図 2 に CNN の構成図を示す. 両手の骨格情報は 42×2 キーポイントで出力されるため, 前処理として CNN に入力する際に直積計算を行う. この時, CNN に入力するデータは 84×84 の正方行列となる. 本研究で用いる CNN は, 2 つの畳み込み層と 2 つのプーリング層で構成されており, 活性化関数はすべて ReLU を用いている. 出力層での活性化関数は Softmax としている. 最適化アルゴリズムは Adam を使用している.

SVM では, OvR(One versus the Rest)手法を用いた多クラス分類を行う. 本研究では, カーネル関数に RBF を使用する. C 値は 1.0, gamma 値は 0.0001 に設定している.

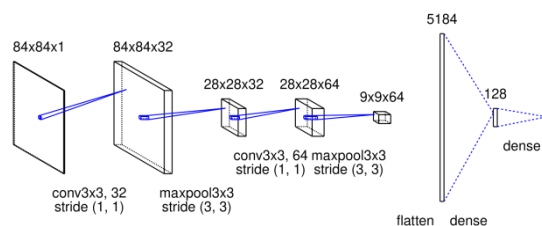


図 2 CNN の構成

3. 実験

3.1. 実験条件

本実験では, Logicool の C920 を使用して撮影を行った. 解像度は 1920×1080 , 30FPS に設定した. カメラはキーボードとともに机上に設置された 27 インチのデスクトップモニター上部に取り付け, キーボードと両手全体が映り, 両手の骨格情報を検出できるように撮影角度を決定した. 本実験では, 6 人を対象に実験を行った. 被験者は全員 20 代男性であり, 骨格などに極端な差はない. ま

た日頃からキーボード操作を経験しており、キーボード操作に不慣れではない。

3.2. 実験方法

被験者にはパスワード等のキーフレーズ入力を想定し、短い文字列として「password」、長い文字列として「grapepasswordbook」の2パターンをそれぞれ20回ずつ入力してもらい、その様子を撮影した。短い文字列は、長い文字列と比べて抽出できる特徴量が少なくなることから、識別精度にどのような影響を与えるか比較を行う。短い文字列には、8文字の単語を設定した。長い文字列には、複数単語を組み合わせた15文字以上のキーフレーズを設定した。撮影した動画は、キーボードの入力始めから終わりまでの全フレームを画像として抽出した。抽出した画像をOpenPoseに入力し、両手の骨格情報を実験データとして用意した。実験データは、8割を訓練データ、2割をテストデータとして学習を行った。被験者ごとに得られた画像枚数は、以下の表1に示す。また、t-SNE[6]によって次元圧縮を行った実験データの分散を図2に示す。t-SNEは、高次元のデータを2次元や3次元に圧縮して可視化するための次元削減アルゴリズムである。図2では左が短い文字列、右が長い文字列の散布図となっている。

表1 被験者毎に集めた画像枚数

被験者ラベル	短い文字列 [枚]	長い文字列 [枚]
0	1245	3219
1	1306	2911
2	1800	4222
3	1111	2845
4	1572	3238
5	1505	3063
合計	8539	19498

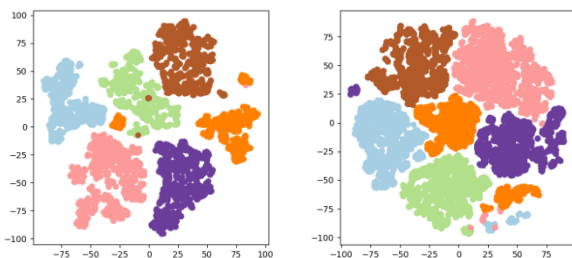


図3 t-SNEによる実験データの分散

4. 実験結果

各手法の識別結果を表2と表3に示す。これらの表では、全結合型NNとCNN, SVMのトレーニングスコアとテストスコアを示している。実験の結果、いずれのモデルでも97%以上の精度で識別できていることがわかった。文字列の長さに関わ

らず、高精度で識別できていることから、8文字程度の短い文字列でも識別に十分な特徴量を抽出できていることがわかった。全結合型NNは、CNNやSVMに比べて学習時間が短く、高い精度を示す結果となった。ハイパーパラメータを最適化することによって、さらに処理時間の向上が見込めると考えられる。CNNは、骨格座標の変換を行ったことで入力するデータの次元数が増大したため、処理時間が長くなってしまった。SVMでは、識別精度が全結合型NNとCNNに比べるとやや低い値となった。これは、OpenPoseによって正しく骨格情報を推定できなかった外れ値によって、特徴の一部が欠損しているデータが含まれているからであると考えられる。実験では、両手が限りなく近づいている場合や、特に両手が接触している場合に誤認識が多く発生していることが確認できた。

表2 短い文字列での識別精度

	training	test
NN	1.000	0.999
CNN	1.000	1.000
SVM	0.986	0.982

表3 長い文字列での識別精度

	training	test
NN	1.000	1.000
CNN	1.000	0.999
SVM	0.974	0.971

5. おわりに

本論文では、OpenPoseによる二次元骨格情報から、全結合型NN, CNN, SVMを用いた個人識別手法の構築を行った。実験の結果、いずれのモデルでも高い識別精度を示す結果となった。一方で、両手の位置や接触によって誤認識が発生する問題がある。今後の課題として、テンキー、ショートカットキー操作の組み合わせや異なる文字列を用意し、さらに多くのラベル数による識別実験を行う。実験データの時系列を考慮したデータセットを作成し、学習を行う。

参考文献

- [1] Z. Cao, G. Hidalgo, T. Simon, S. Wei and Y. Sheikh: OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields, IEEE Transactions on Pattern Analysis and Machine Intelligence, (2019).
- [2] T. Simon, H. Joo, I. Matthews and Y. Sheikh: Hand Keypoint Detection in Single Images using Multiview Bootstrapping, CVPR, (2017).
- [3] Z. Cao, T. Simon, S. Wei and Y. Sheikh: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields, CVPR, (2017).
- [4] S. Wei, V. Ramakrishna, T. Kanade and Y. Sheikh: Convolutional pose machines, CVPR, (2016).
- [5] Home- Keras Documentation, <https://keras.io/ja/>.
- [6] L. V. Maaten, G. E. Hinton: Visualizing Data using t-SNE, Journal of Machine Learning Research, 9, (2008).