単眼カメラから得られる骨格情報を用いた人物識別 —CNN と SVM の比較—

中央大学　○戸田哲郎，ライテックス　Alessandro Moro，中央大学　梅田和昇

Individual identification using skeleton information obtained from monocular camera
—Comparison of CNN and SVM—

Chuo Univ. Tetsuro TODA, RITECS Alessandro MORO, Chuo Univ. Kazunori UMEDA

It is necessary in various scenes to perform individual identification from images of a camera. This paper focuses on skeleton information that seems to be unique to individuals. Skeleton information is extracted using posture estimation library OpenPose. Then, personal identification is performed using the skeleton information. Two methods are compared for the identification. The first method uses Convolution Neural network (CNN). CNN finds personal features from skeleton information and identifies individuals. The second method uses Suport Vector Machine (SVM). In this case, person's features are set manually from skeleton information, and individual identification is performed with SVM. Two individual identification methods are compared by experiments.

1. 序　論

在室管理やセキュリティ，防犯，個別サービスなど，個人識別が必要となるシーンは様々なところにある．この際，個人識別によく用いられる方法として，RFID の利用やや指紋認証，虹彩認証などがある．しかし，これらの方法は各個人が認証のための行動をとる必要がある．そのため，利用者に負担がない個人識別手法の研究が盛んに行われている．

近年では，CNN（Convolution Neural Network）を利用した顔識別 [1] が高い識別率を実現している．しかし，顔識別を行う場合には，画像中で顔が写っている必要があり，後ろ姿などでは識別できない．そのため，利用シーンが限られてしまうという問題がある．そこで，本研究では骨格情報に着目する．骨格情報による個人識別はいくつか研究されており，その有用性が示されている [2, 3]．しかし，いずれも RGB-D センサを用いており，3 次元の骨格情報から個人識別を行っている．本研究では，より一般的な単眼カメラからの骨格情報を用いた個人識別手法の構築を目指す．また，本研究では，CNN と SVM を用いた二つの個人識別手法を構築し，それらの識別精度の比較を行う．

2. 個人識別手法

2.1 骨格情報抽出

本研究では，Zhe らの人物姿勢推定手法 [4] を用いた OpenPose というライブラリを利用して骨格情報を抽出する．この手法は CNN を組み合わせて姿勢推定を行っている．画像を入力として，出力は特徴（肩，肘，膝等）の x, y 座標と認識の信頼度となっている．本手法では，COCO（Common Objects in Context）2016 のデータセットを教師データとした学習モデルを用いる．この学習モデルでは Fig. **1** に示す 18 点の特徴点を抽出できる．特徴点のデータ順を次式に示す．

$$\boldsymbol{v} = [x_1,\ y_1,\ c_1,\ \ldots,\ x_{18},\ y_{18},\ c_{18}] \tag{1}$$

ここで x_n，y_n はそれぞれ n 番目の特徴点の x, y 座標である．また，c_n は特徴認識の尤度を表す．

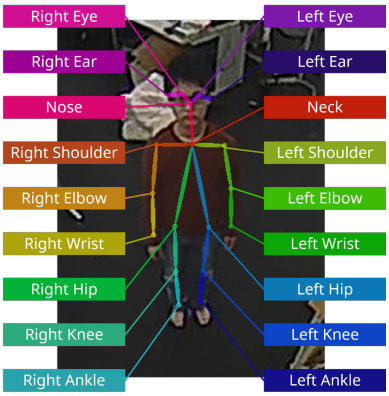

Fig.**1** OpenPose から得られる骨格情報

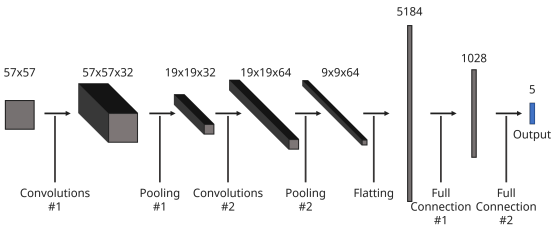

Fig.**2** ネットワーク構成

2.2 CNN による個人識別

2.2.1 前処理

骨格情報を CNN への入力とするために，特徴 $\boldsymbol{v}$ の直積 F を計算する．

$$F = \boldsymbol{v}^T \boldsymbol{v} \tag{2}$$

よって，CNN への入力は 54×54 の行列となる．

2.2.2 ネットワークの構成

CNN の構成を Fig. **2** に示す．本手法のネットワークは，2 つの畳込み層と 2 つのプーリング層，2 つの全結合層からなる．入力は 54×54 の行列 F で，出力が各クラスの尤度となっている．畳込み層とプーリング層の活性化関数は，全て ReLU

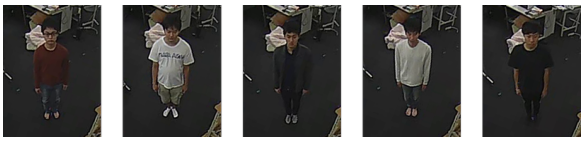

Fig.3 識別対象

（Rectified Linear Unit）である．また，出力層では Softmax 関数を活性化関数としている．学習時には，1 つ目の全結合層で 50% の割合でドロップアウトを行うことで過学習が起きないようにしている．また，ミニバッチ学習により学習を行う．バッチサイズは 128 としている．さらに，最適化アルゴリズムは Adam（Adaptive Moment Estimation）[5] を用いている．

2.3 SVM による個人識別

2.3.1 前処理

特徴 $\boldsymbol{v}$ は画像上の座標と認識の尤度であり，SVM の入力として特徴 $\boldsymbol{v}$ を直接用いるのは適さない．よって，特徴 $\boldsymbol{v}$ から SVM の入力として適した特徴を手動で選び出す．本手法で選び出した特徴を以下に示す．

- 肩から肘までの長さ
- 肘から肩までの長さ
- 坐骨から膝までの長さ
- 膝から足首までの長さ

これらの特徴は，カメラに対する体の向きによる変化が小さい．以上の 4 つが左右分あるので，特徴は計 8 個となる．

2.3.2 SVM の構成

SVM は一般的にデータの一部が欠落した場合に識別することができない．しかし，本研究では，カメラに対する体の向きによっては骨格推定が一部失敗し，特徴の一部が欠落することがある．これにより，前述の 8 個の特徴の一部は計算できない可能性がある．よって，特徴の 8 個すべてを利用した SVM は個人識別を行えないシーンが多く生じることになる．そのため，本研究では以下の 3 つの SVM 識別器を用意する．

- 8 個すべての特徴を用いた SVM
- 体の左側の 4 個の特徴のみを用いた SVM
- 体の右側の 4 個の特徴のみを用いた SVM

SVM のカーネルには一般的な RBF カーネルを用いる．

3. 個人識別比較実験

3.1 実験条件

3.1.1 使用カメラ

本実験では，パナソニックの DG-SF438 という魚眼カメラを用いた．解像度は 1280 × 960 で使用した．カメラは床から高さ約 2.7[m] の位置に下向きに設置した．

魚眼カメラから取得した画像は，通常のカメラから取得した画像と比べて大きく歪んでいる．そのため，魚眼画像をそのまま骨格推定の入力とすると，推定がうまくいかない．よって本研究では，取得した魚眼画像の人物の写った領域を正像変換 [6] により歪みのない画像に変換する．さらに，遠くに写った人物は画像上で小さく見えるため，これを補正するように画像を拡大する．そして，以上の変換を行った画像を骨格推定の入力とする．

Table1 個人識別比較結果

識別手法	CNN	SVM		
		8 個すべて	左 4 個	右 4 個
識別率	98.7%	78.0%	58.7%	51.1%

3.1.2 入力データ

本実験では，Fig. **3** に示す 5 人を対象に識別を行った．5 人全員が 20 代の男性であり，極端な体格差はない．対象者には，カメラ直下の床の位置から 2[m] 離れたところを始点に 1[m] ずつ離れて 7[m] 離れるまで，計 6 箇所に立ってもらい撮影を行った．また，対象者は全ての画像において，カメラに正面が見えるようにしてもらった．画像はそれぞれの箇所で約 100 枚ずつ用意した．約 100 枚のうち 8 割を学習用のデータ，残りの 2 割をテスト用のデータとした．

3.2 比較結果

Table **1** に CNN による個人識別と SVM による個人識別の結果の比較を示す．CNN では 98.7% とほとんどのテストデータにおいて正しい識別ができていた．一方で，SVM では 8 個すべての特徴を使った場合では 78.0% とおおよそ識別できていた．しかし，左右の 4 個ずつの特徴だけの場合には半分程度しか正しく識別できていなかった．CNN より SVM の方が識別率が低い原因の一つは，使用している特徴の少なさであると考えられる．CNN では，特徴 $\boldsymbol{v}$ の全ての中から個人の特徴を見つけ出すように構成されている．それに対して SVM では，$\boldsymbol{v}$ の 54 個の特徴の中で，12 点の関節の x, y 座標，計 24 個の特徴だけを使っている．また，CNN では特徴 $\boldsymbol{v}$ の様々な組み合わせが考慮されているが，SVM では 8 つの組み合わせしか使っていない．よって，SVM の識別結果は，より多くの特徴を手動で選び出すことで良くなると考えられる．また，CNN により骨格情報から個人の特徴を見つけ出す手法は有効であるといえる．

4. 結　　論

本論文では，骨格情報から CNN と SVM のそれぞれを用いた個人識別手法を構築し，比較を行った．結果としては，手動で特徴を決定した SVM よりも，CNN により特徴を見つけ出し識別する方が良い結果となった．今後は，正面を向いた画像以外での検証やクラス数を増やした検証を行う．また，CNN による識別を拡張していき，高精度な個人識別手法の構築を目指す．

参 考 文 献

1) F. Schroff, K. Dmitry, and P. James: Facenet: A Unified Embedding for Face Recognition and Clustering, Proc. of the 2015 IEEE Conf. on CVPR, pp. 815-823, (2015).
2) A. Sinha, K. Chakravarty, and B. Bhowmick: Person Identification Using Skeleton information from Kinect, Proc. Intl. Conf. on ACHI, pp. 101-108, (2013).
3) A. Ball, D. Rye, F. Ramos, and M. Velonaki: Unsupervised Clustering of People from'Skeleton'Data." Proc. of the seventh annual ACM/IEEE international conf. on HRI, pp. 225-226, (2012).
4) C. Zhe, S. Tomas, W. Shih-En, and S. Yaser: Realtime Multi-person 2d Pose Estimation Using Part Affinity Fields, Proc. of the 2017 IEEE Conf. on CVPR, pp. 1302-1310, (2017).
5) D. P. Kingma and J. L. Ba: ADAM: A Method for Stochastic Optimization, Proc. of ICLR2015, (2015).
6) 森 隆寛, 外村 元伸, 大住 勇治, 池永 剛: キュービック補間を用いた魚眼レンズ画像の高画質補正アルゴリズム, 情報科学技術フォーラム一般講演論文集, Vol. 5, No. 1, pp.7-8, (2006).