

異なる画像間差分を入力とする CNN を用いた低解像度画像に頑強な歩容認証

瀬戸口 直輝^{1,a)} 中嶋 一斗¹ 呉 桐¹ 岩下 友美² 倉爪 亮¹

概要

歩容認証とは、人の歩行画像から抽出した特徴を用いた人物認証手法である。歩容認証の問題点の一つに、画像の低解像度化に伴う識別性能の低下がある。これに対して本研究では、個人間の速度の差異を陽に表現した Signed Frame Difference Energy Image (SFDEI) の利用、及び SFDEI の特性を持つ CNN(SFDEINet) を構築することで、識別性能の向上を実現する。956 名の歩行画像を用いた実験により、従来手法と比較して提案手法が高い識別性能を持つことを示す。

1. はじめに

生体情報を用いた人物認証は、本人確認手段の一つとして用いられており、安全な社会を実現する上で重要な技術である。生体情報の一つである歩容は、遠方、全方位から非接触で取得でき、顔や指紋などと比較して環境変化に頑健である。そのため、歩容認証は監視カメラを用いた次世代防犯システムへの応用が期待されている。

歩容認証において、歩行画像列の平均シルエットに基づく GEI(Gait Energy Image) と、深層学習を用いた特徴抽出手法である GEINet[1] は高い識別性能を持つことが知られている。歩容認証の問題点として、例えばカメラと歩行者間の距離変化により、撮影画像の解像度が低下すると、識別性能も下がることが挙げられる。この問題に対して、我々は、GEI に画像間差分情報を 3 チャンネルで組み合わせた SFDEI(Signed Frame Difference Energy Image)[5] を提案している。

先行研究 [5] の OU-ISIR Large Population Dataset(956 名分の歩行画像) を用いた個人識別実験では、まず、SFDEI の画像間差分の計算のための時間間隔 Δt を 1 から 10 フレームまで変化させる。その後、それぞれの時間間隔の SFDEI に対して GEINet[1] を学習させ、計 10 個の学習モデルを得る。次に、各学習モデルで抽出した特徴を結合し、

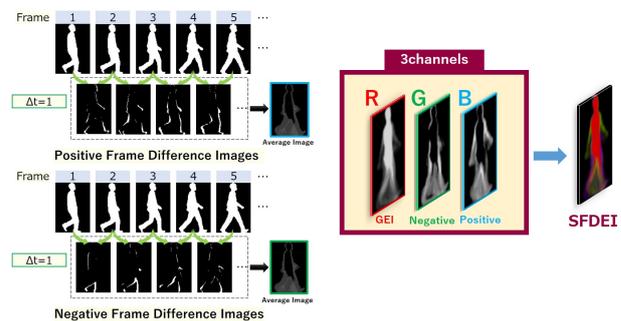


図 1 SFDEI

識別器として用いる Nearest Neighbor の入力とすることで、解像度 16×11 の画像において識別率 85.6% を達成した。しかし、GEINet は GEI に最適化されたネットワークであり、SFDEI の個人間の歩行速度の差異を陽に表現する長所を活かす構造とは言い難い。また、特徴の単純な結合が認証に最適であるとは限らない。そこで本稿では、GEINet の畳み込み層を並列化して、異なる時間間隔 Δt の複数の SFDEI を同時に学習することが可能な SFDEINet を提案する。実験では、OU-ISIR Large Population Dataset[2] を用いて、低解像度画像を用いた個人識別実験を行い、GEINet と比較して提案手法が 2.5% 性能向上することを示す。

2. 提案手法

2.1 SFDEI

SFDEI は、以下に示す手順により、GEI 及び画像間差分の情報を 3 チャンネルを用いて表現される。まず、歩行画像列から任意の時間間隔 Δt を用いて画像間差分を計算する。次に、画像間差分値の符号によって、正の平均差分画像、負の平均差分画像を生成する。最後に、R チャンネルに GEI、G チャンネルに負の平均差分画像、B チャンネルに正の平均差分画像と登録することで、3 チャンネル画像として表現する。SFDEI の生成過程を図 1 に示す。SFDEI がもつ長所として、以下の 2 点が挙げられる。

¹ 九州大学

² NASA/Caltech JPL

^{a)} setoguchi@irvs.ait.kyushu-u.ac.jp

- GEI と比較して動作情報を明に利用
- 画像間差分の計算で用いられる時間間隔 Δt を可変にし、個人間の歩行速度の差異を陽に表現

OU-ISIR Large Population Dataset(956 名分の歩行画像)を用いた実験では、特徴抽出器に GEINet[1]を用いた。GEINet は 2 層の畳み込み層 (Conv1, Conv2) と 2 層の全結合層 (Fc3, Fc4) から成る CNN で、各層のハイパーパラメータや学習設定は GEI に最適化されている。

性能評価実験 [5] では、まず、異なる時間間隔 Δt で生成された SFDEI を用いて、計 10 個の GEINet を学習させる。次に、各 GEINet の Fc3 層からの出力を特徴ベクトルをみなし、特徴ベクトルを結合させる。得られた特徴ベクトルを Nearest Neighbor の入力とすることで、解像度 16×11 の画像において識別率 85.6% を達成した。図 2 に SFDEI を入力とした GEINet による認証手法を示す。

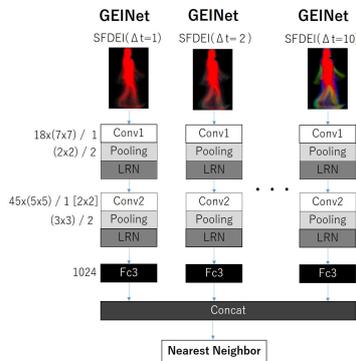


図 2 SFDEI を入力とした GEINet による認証手法
 conv : Number of filter \times (filter size) / stride [padding size]
 pooling : (filter size) / stride
 Fc : Number of output node

2.2 SFDEINet

SFDEI では画像間差分を求める時間間隔 Δt の値を大きくすることで、例えば、歩行速度の大きい人は腕振りなどの動作変化が大きくなるなど、個人間の歩行速度の差異を陽に表現することができる。その一方で、腕がどのように振られたかといった動作の軌跡が失われてしまう可能性がある。そこで SFDEINet では GEINet を拡張して、小さな値と大きな値の 2 種類の Δt により生成された 2 つの SFDEI を入力とすることで、動作の軌跡と歩行速度の差異を同時に学習できるようにする。GEINet に加えた拡張点を以下に示す。

- 2 層の畳み込み層を並列化し、その出力を Fc3 層の前で結合 (学習するパラメータは独立)
- Fc3 層の次元数を 1024 から 2048 に拡張

各畳み込み層のカーネルサイズ、チャンネル数、ストライド、パディングサイズは GEINet と同一である。初期値について、重みは He の手法で乱数を与え、バイアスは 0 で固定している。活性化関数は ReLU、各畳み込み層のあとに、Max Pooling と Local Response Normalization(LRN)

を適用している。また、Fc3 層の次元数は、解像度変化がおきる場合の性能がピークになる値に設定した。

SFDEINet の利点として、学習するネットワーク内で情報を統合することで、より高性能な特徴を自己生成できることが考えられる。SFDEINet を用いた認証手法を、図 3 示す。

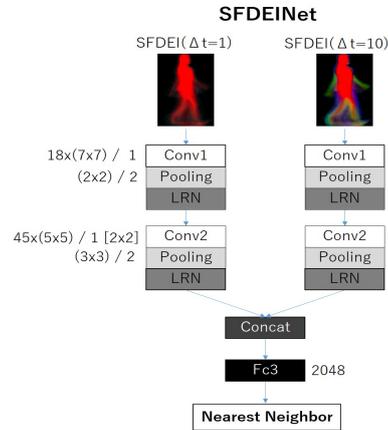


図 3 SFDEINet の認証手法
 conv : Number of filter \times (filter size) / stride [padding size]
 pooling : (filter size) / stride
 Fc : Number of output node

3. 実験

実験では、OU-ISIR Large Population Dataset[2] のサブセットである CV01, CV02 に定義されている各 956 名分の歩行データを用いる。歩行データには、1 人につき Gallery と Probe の 2 種の歩行シーケンスが含まれており、学習に CV01、認証に CV02 を用いる。

認証では、学習済みモデルを特徴抽出器とし、Fc3 層からの出力を特徴ベクトルとする。CV02 の Gallery, Probe の両方で特徴抽出が行われ、識別器として用いる Nearest Neighbor に入力される。また、解像度変化した場合の識別性能を確認するため、高解像度画像 (解像度 128×88) の画像をダウンサンプリングし、計 5 種類 (128×88 , 96×66 , 64×44 , 32×22 , 16×11) の画像を生成する。データベースには高解像度画像が登録されていることを想定し、高解像度画像 (解像度 128×88) の画像とダウンサンプリングされた画像から抽出した特徴ベクトルをマッチングさせる。図 4 に解像度変化させた画像の一例を示す。

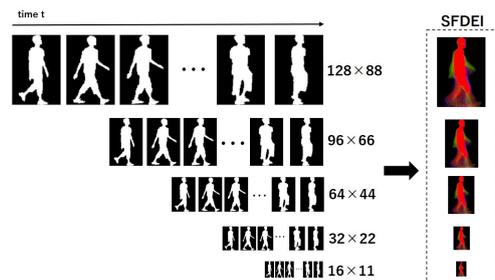


図 4 画像解像度のダウンサンプリング

表 2 SFDEINet の識別率 [%], Fc3 層の次元数 (1024), 2 並列

Fc3 層の次元数 (1024)	入力した SFDEI の Δt の組								
	1,2	1,3	1,4	1,5	1,6	1,7	1,8	1,9	1,10
解像度 (128×88)	94.3	92.6	86.0	93.7	91.8	81.4	92.5	78.1	75.2
解像度 (96×66)	93.8	92.6	83.5	92.9	91.4	76.9	92.6	74.3	59.2
解像度 (64×44)	94.1	92.7	85.3	93.7	91.9	80.1	92.6	76.7	74.3
解像度 (32×22)	93.8	92.4	75.6	93.8	91.5	63.3	68.4	60.5	59.2
解像度 (16×11)	68.2	66.9	31.9	61.5	64.7	12.0	68.4	14.2	12.1

表 3 SFDEINet の識別率 [%], Fc3 層の次元数 (2048), 2 並列

Fc3 層の次元数 (2048)	入力した SFDEI の Δt の組								
	1,2	1,3	1,4	1,5	1,6	1,7	1,8	1,9	1,10
解像度 (128×88)	92.2	92.9	90.6	91.2	92.0	81.7	92.1	76.0	69.1
解像度 (96×66)	92.2	92.9	90.4	91.0	91.7	78.5	92.1	71.3	66.2
解像度 (64×44)	92.2	92.8	90.4	91.3	92.1	80.7	92.2	74.1	67.8
解像度 (32×22)	91.9	92.7	90.3	90.8	92.0	64.1	92.2	55.6	53.7
解像度 (16×11)	72.0	77.8	70.6	70.6	73.9	13.4	74.3	10.9	13.1

表 4 GEINet によって抽出した特徴を結合したときの識別率 [%]

	結合した SFDEI の Δt の組み合わせ								
	1~2	1~3	1~4	1~5	1~6	1~7	1~8	1~9	1~10
解像度 (128×88)	92.5	92.8	92.8	93.2	92.8	92.7	92.6	92.5	92.1
解像度 (96×66)	92.5	92.8	92.9	93.0	92.7	92.7	92.4	92.4	92.0
解像度 (64×44)	92.5	93.0	92.9	93.0	92.9	92.7	92.6	92.5	92.1
解像度 (32×22)	92.1	92.4	92.2	92.2	92.6	92.4	91.9	91.8	92.7
解像度 (16×11)	84.5	85.4	85.6	85.4	85.5	85.5	85.5	85.6	85.25

表 5 SFEEINet で抽出した特徴を用いたときの識別率 [%]

	入力した SFDEI の Δt の組								
	1,2	1,3	1,4	1,5	1,6	1,7	1,8	1,9	1,10
解像度 (128×88)	93.8	92.9	92.4	93.2	92.9	93.0	92.6	92.5	92.0
解像度 (96×66)	93.9	92.4	92.7	92.8	93.0	93.2	92.6	92.3	91.9
解像度 (64×44)	93.5	92.9	92.5	93.3	93.2	92.9	92.5	92.4	91.8
解像度 (32×22)	93.5	92.5	91.9	92.9	92.5	93.2	91.9	92.1	91.5
解像度 (16×11)	85.5	86.9	85.0	87.7	86.1	86.7	86.5	88.1	86.9

まず、解像度 128×88 の画像を用いて学習を行い、識別性能がピークとなる並列数を探索する実験を行った。本実験では、解像度 128×88 の画像のみ学習させている。その結果を表 1 に示す。解像度 16×11 は 2 並列、その他の解像度は 3 並列で識別性能はピークとなった。

表 1 SFDEINet の並列数を変更する実験の結果

並列数	128×88	96×66	64×44	32×22	16×11
$\Delta t = 1,2$	92.2	92.2	92.2	91.9	72.0
$\Delta t = 1,2,3$	93.2	92.4	93.2	92.7	66.3
$\Delta t = 1,2,3,4$	70.8	71.3	72.8	63.0	31.2

これは、学習率などの学習設定やハイパーパラメータを GEINet で使われていた値から変更しておらず、ネッ

トワークの性能が十分に引き出せていないためと考えられる。ハイパーパラメータ探索は残る課題とし、今回は 2 並列の SFDEINet を用いることとする。また、本実験では、並列数に合わせて、Fc3 層の次元数も線形に増やした (1024×N)。これは、畳み込み層から流れる情報を Fc3 層で次元削減する際、従来の次元数では表現能力が足りず、識別性能が低下したためである。表 2、表 3 に、その結果を示す。ここで、SFDEINet の学習は、SFDEI の Δt の値によって定めたペアを入力画像としている。認証では、異なるペアで学習させた計 10 個のモデルを用いて識別性能を確認した。表 2、3 の結果より、解像度 128×88、96×66、64×44、32×22 の場合は識別率に大きな差は見られないが、

解像度 16×11 の場合には表 2 と比較して表 3 の結果が大幅に向上されていることがわかる。なお従来の GEINet を用いた場合、解像度 16×11 の識別率は 74.1% である。

次に、複数の解像度を混ぜて学習させ、GEINet との性能比較を行った。GEINet を用いた結果を表 3 に、SFDEINet を用いた結果を表 4 に示す。これより、 16×11 という解像度が著しく低い画像が入力されたときの最高識別率が、GEINet を用いて特徴抽出し、それを結合した場合は 85.6% であるのに対し、SFDEINet を用いて特徴抽出した場合は 88.1% となり、2.5% 性能が向上している。なお、従来の GEI 及び GEINet を用いた場合 [5] は、80.3% であった。

4. まとめと今後の予定

本報告では、GEINet を基に、畳み込み層を並列化することで、異なる Δt で生成された SFDEI を同時に学習できる SFDEINet を提案した。956 名の歩行データを用いた認証実験では、SFDEINet を用いることで、 16×11 という低解像度画像も、識別率 88.1% という高い性能を達成した。

今後は、まず SFDEINet のパラメータチューニングを行う。その後、再び畳み込み層の並列数を変更させた実験と SFDEI の Δt の値を今回設定したものより、更に大きく設定し、学習と識別に使う SFDEI のペアを変えて、識別性能の変化を確かめる実験を行う。また、DNN を用いた最新の歩容認証手法において、低解像度画像を入力とした場合の識別率を検証することで、提案手法の優位性を確認する。

5. 謝辞

本研究は文部科学省科学研究費補助金基盤研究 (A) (課題番号 JP26249029) の支援を受けた

参考文献

- [1] K. Shiraga, et al., "GEINet: View-Invariant Gait Recognition Using a Convolutional Neural Network", Int. Conf. on Biometrics (ICB 2016), No. O19, pp. 1-8, 2016.
- [2] Y. Makihara, et al., "The OU-ISIR Gait Database Comprising the Treadmill Dataset," IPSJ Trans. on Computer Vision and Applications, Vol. 4, pp. 53-62, 2012.
- [3] "http://www.am.sanken.osaka-u.ac.jp/BiometricDB/dataset/GaitLP/Benchmarks.html"
- [4] Noriko Takemura, Yasushi Makihara, Daigo Muramatsu, Tomio Echigo, Yasushi Yagi, "On Input/Output Architectures for Convolutional Neural Network-Based Cross-View Gait Recognition", IEEE Transactions on Circuits and Systems for Video Technology, vol.28, no.1, 2018.
- [5] 呉 桐, 中嶋 一斗, 瀬戸口 直輝, 岩下 友美, 倉爪 亮, 正負 フレーム間差分を用いた深層学習による歩容認証, 画像の認識理解シンポジウム (MIRU2017), pp.PS1-19, 2017.8.8