

CNNを用いた視点変化に頑健な歩容認証における 入出力構造の一検討

武村 紀子¹ 楨原 靖² 村松 大吾² 越後 富夫³ 八木 康史²

概要：本論文では、畳み込みニューラルネットワークの入出力構造を状況に応じて適切に設計することで、観測方向の違いに頑健な歩容認証を実現する。本人認証および個人識別といった歩容認証タスクの違い、および人物の違いによる見えの違いと観測方向が異なることによる同一人物内の見えの違いとの間のトレードオフの2つの観点から本問題を考える。具体的には、本人認証には Contrastive loss を損失関数に持つ2入力 of Siamese network を、個人識別には Triplet ranking loss を損失関数に持つ3入力 of Triplet network を用いる。さらに、観測方向の差が小さく、同一人物内の見えの違いよりも他人間の見えの違いの方が大きい場合は、感度良く見えの違いを比較するために、下位層で入力画像の比較を行う。逆に、観測方向の差が大きく、他人間の見えの違いよりも同一人物内の見えの違いの方が大きい場合は、見えの違いの影響を抑制するために、ネットワークの上位層で入力画像特徴の比較を行う。評価実験により、歩容認証タスクや観測方向の差の大きさによりネットワーク構造を使い分ける本提案手法の有用性を確かめた。

キーワード：畳み込みニューラルネットワーク、視点変化、歩容認証

1. はじめに

生体特徴を用いた人物認証技術は、監視や入退出管理、科学捜査など様々なアプリケーションに適用可能であるため、盛んに研究がなされてきた。人物認証に用いられる生体特徴として、顔、声、指紋、虹彩、筆跡、歩容などが挙げられるが、その中でも歩容特徴は、遠方から防犯カメラ等で撮影した映像から被験者の協力なしに収集可能であり、実用的な生体特徴として注目を集めている。実際、犯罪捜査に歩容認証が用いられた事例も存在している [1–3]。

しかし、同一人物の歩容であっても、観測方向や服装、歩行速度、荷物等の共変量の違いにより、歩行映像の見え方が大きく異なり、歩容認証の性能が低下するといった問題が存在する。特に観測方向は見えに大きな影響を与える共変量であり、実システムで歩容認証を行う際に観測方向に対する頑健さは重要な課題となる。これまでに、観測方向の違いに頑健な歩容認証手法として、生成的なアプローチや識別的なアプローチを用いた様々な手法が提案されている。

生成的手法では、異なる観測方向から撮影された2枚の画像を比較する際に、2枚の画像を同じ方向から見た画像に

変換してから比較することで、観測方向の違いに頑健な歩容認証を行っている [4–7]。一方、識別的手法では、機械学習アルゴリズムを用いて識別性能が高くなるように識別モデルを学習し、人物認証を行っている [8–12]。識別的手法では様々な機械学習手法が用いられるが、近年では、畳み込みニューラルネットワーク (CNN: Convolutional Neural Network) を用いた歩容認証が注目を集めており [13–17]、CNN を用いた手法は他の手法よりも高精度に認証可能であると報告されている。

CNN を用いた手法では入出力の機構を柔軟に変化させることが可能であるという利点があり、歩容認証を行う状況に応じて CNN のネットワーク構造を適切に設計することで、認証性能の向上が期待できる。例えば、Shiraga ら [15] は CNN の入力を1枚の歩容特徴画像、出力を被験者 ID とし、損失関数として交差エントロピーを用いている。また、Wu ら [14] は、Siamese network [18] として知られている2入力 of CNN を採用しており、入力を2枚の歩容特徴画像のペア、出力を入力画像が同一人物 (あるいは異なる人物) である確率とし、損失関数として Contrastive loss を用いている。文献 [19] でも議論されているとおり、歩容認証タスク (本人認証、個人識別) により、適切な識別モデルは根本的に異なるため、ネットワーク構造 (特に入出力および損失関数) を設計する際は、各タスクの性質を考

¹ 大阪大学データビリティフロンティア機構

² 大阪大学産業科学研究所

³ 大阪電気通信大学大学院工学研究科

慮し慎重に設計すべきである。

さらに、CNN を用いた手法では、畳み込み層やプーリング層を繰り返すことにより位置ずれに対する頑健性が増すため、観測方向の異なる歩容認証において、観測方向の違いによる同一人物内の見えの違いの影響を抑えることができる。しかし、このことは、同一人物内の見えの違いを抑制すると同時に、他人感の見えの違いについても抑制してしまう。このような観測方向による同一人物内の見えの違いの影響と他人間の見えの違いの影響との間のトレードオフは CNN のネットワーク構造を設計する上で非常に重要な問題である。

しかし、従来研究では、ネットワークを設計する際にこれらの点がほとんど考慮されていない。そこで、本論文では、歩容認証タスクおよび同一人物内/他人間の見えの違いに関するトレードオフの 2 点を考慮し、それぞれの状況に応じて適切なネットワーク構造を提案する。

歩容認証タスクは本人認証（1 対 1 認証）および個人識別（1 対多認証）に大別できる。本人認証とは、歩容特徴のペアが同一人物か否かを認証するタスクであり、犯罪捜査において、容疑者と犯人が同一人物であるかの認証に用いられったり [3]、出入国管理において、犯罪者やテロリストといった特定人物の検出に用いられったりする。本人認証では、同一人物の歩容特徴ペアの相違度が他人の歩容特徴ペアの相違度よりも小さくなるように識別モデルを学習させる。一方、個人識別とは、入力された人物があらかじめ登録されているどの人物と同一であるかを識別するタスクであり、複数のカメラ間での人物照合や ID レス入退管理などに用いられる。個人識別では、入力人物の歩容特徴と登録人物の歩容特徴との相違度を各登録人物についてそれぞれ求め、最も相違度が小さかった人物を入力人物と同一と見なす。そのため、入力人物と異なる人物との相違度よりも入力人物と同一人物との相違度が相対的に小さくなるように識別モデルを学習させる。つまり、本人認証タスクでは相違度の絶対値が重要であり、個人識別タスクでは相違度の相対値が重要である。これらを考慮することで、各歩容認証タスクに適したネットワーク構造を設計することが可能となる。

同一人物内/他人間の見えの違いに関するトレードオフを考慮する際は、2 枚の入力画像を比較する位置を変えることで、見えの違いに対する感度を変化させる。観測方向の差が大きく、同一人物内の見えの違いが他人間の見えの違いよりも大きい場合は、畳み込み層やプーリング層により圧縮された歩容特徴を上位層で比較することで見えの違いの影響を抑制する。一方、観測方向の差が十分小さく、同一人物内の見えの違いよりも他人間の見えの違いの方が大きい場合は、より詳細な特徴を比較するため、入力画像自体を最下位層で比較する。さらに、これら見えの違いに対する感度の異なる二つの CNN は根本的に性質が異なる

ため、相補的に統合することでさらなる認証性能の向上が期待できる。そこで、入力画像を上位層/最下位層と比較する二つの CNN をスコアレベルで統合する手法についても提案する。

2. CNN を用いた視点変化に頑健な歩容認証手法

観測方向の異なる歩容認証の様々な状況に応じて、適切なネットワーク構造を提案する。まず、歩容認証タスクの違いに応じたネットワーク構造について、次に観測方向による見えの違いに応じたネットワーク構造について説明する。ただし、本手法では、歩容特徴として歩容エネルギー画像（GEI: Gait energy image）[8]を用いる。GEI は 1 歩行分の正規化シルエット画像列を平均することで生成されるため、体型などの静的な特徴と腕や脚の動きといった動的特徴を併せ持っている。

2.1 一般設定

本論文で提案するネットワーク構造は図 1 のように 3 層の畳み込み層、2 層の正規化層、2 層のプーリング層および 1 層の全結合層から構成される。正規化層（norm1, norm2）では Local response normalization (LRN) [20] を採用しており、プーリング層（pool1, pool2）では max pooling を採用している。また、畳み込み層（conv1, conv2, conv3）および全結合層（fc4）では、活性化関数として ReLU 関数 [21] を採用している。さらに fc4 層の出力に対して L2 ノルム（または L2 距離）を計算している。

学習時は、CNN の重みパラメータの初期値を Xavier のアルゴリズム [22] により設定し、バイアスパラメータの初期値は 0 とする。また、パラメータの更新には確率的勾配降下法（SGD: stochastic gradient descent）を採用している。パラメータに対するモーメントの値は 0.9 とし、学習率は初期値 0.01 をイテレーション回数に応じて 10 で割るという処理を学習期間において 4 回行っている。また、fc4 層の 52 のユニットに対し、0.9 の割合で Dropout [23] を適用している。評価時は最終層で求めた L2 ノルムを入力画像の相違度とみなし、人物の認証を行う。

2.2 歩容認証タスクの違いに応じた手法

1 章で言及したとおり、歩容認証タスクは本人認証と個人識別の 2 つに大別できる。各タスクに適したネットワーク構造について、詳細を以下に述べる。

2.2.1 本人認証に適した手法

本人認証に適した手法として、Siamese network [16,18,24] としても知られる 2 つの並列な CNN で構築されるネットワーク（2in, 図 1(a) 参照）を提案する。このネットワークでは、fc4 層の出力を各入力 GEI の特徴ベクトルとみなし、損失関数として次式で示す Contrastive loss を用いる。

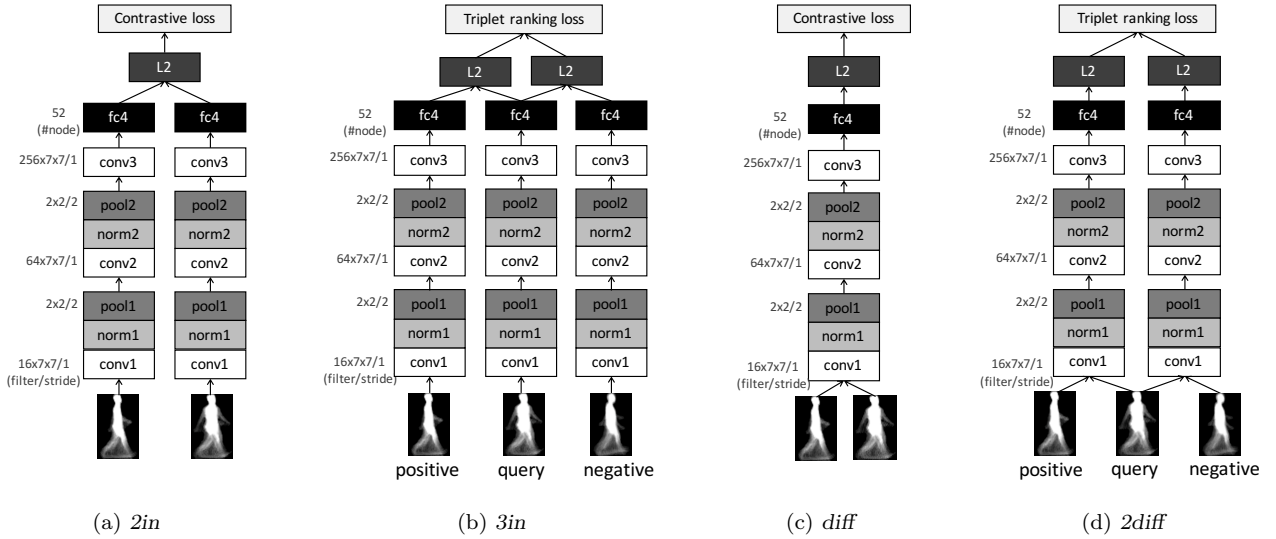


図 1: 提案手法のネットワーク構造

Fig. 1 Network architectures of CNN-based proposed methods.

Contrastive loss は、入力画像が同一人物のときは d^2 (2 枚の入力画像に対する L2 距離の自乗)、入力画像が異なる人物のときは、いわゆる hinge loss で定義する。

$$L_{\text{cont}} = \frac{1}{2M} \sum_{m=1}^M (\delta_{y_{1m}y_{2m}} d_m^2 + (1 - \delta_{y_{1m}y_{2m}}) \max(\text{margin} - d_m^2, 0)) \quad (1)$$

ここで、 M は学習データの入力画像ペアの数、 δ は Kronecker のデルタ、 y_{1m} および y_{2m} は m 番目の入力 GEI の被験者 ID をそれぞれ意味する。また、margin の値は実験的に 3 とした。式 (1) より、入力画像が同一人物の場合は、相違度が小さくなるように、異なる人物の場合は、相違度が margin よりも大きくなるように学習される。評価時は、 d^2 の値に基づいて、入力画像が同一人物か否かを識別する。よって、Contrastive loss を損失関数に持つ 2in は本人認証に適した手法と言える。

2.2.2 個人識別に適した手法

個人識別に適した手法として、Triplet network [25, 26] としても知られる三つの並列な CNN で構成されるネットワーク (3in, 図 1(b) 参照) を提案する。このネットワークでは、query 画像、positive 画像 (query と同一人物の画像) および negative 画像 (query と異なる人物の画像) の 3 画像を入力とする。ここでも、2in と同様に、fc4 層の出力を各入力画像の特徴ベクトルとみなし、損失関数としては次式で示す Triplet ranking loss を用いる。Triplet ranking loss は、 d_{pos}^2 (positive 画像と query 画像の L2 距離の自乗) と d_{neg}^2 (negative 画像と query 画像の L2 距離の自乗) との差で定義される。

$$L_{\text{trip}} = \frac{1}{2M} \sum_{m=1}^M \max(\text{margin} - d_{\text{neg},m}^2 + d_{\text{pos},m}^2, 0) \quad (2)$$

式 (2) より、query と同一人物との相違度が query と異

なる人物との相違度よりも相対的に小さくなるように学習される。よって、Triplet ranking loss を損失関数に持つ 3in は個人識別に適した手法と言える。

2.3 観測方向による見えの違いに応じた手法

入力画像の見えの違いは、被験者の違いおよび観測方向の違いによって生じるが、歩容認証では、観測方向による同一人物間の見えの違いの影響を抑え、他人間の見えの違いに着目することで、入力画像が同一人物か否かを識別する。2in や 3in では、畳み込み層やプーリング層を経た後の上位層で入力画像特徴を比較しているため、見えの違いの影響が抑制されている。よって、観測方向の差が大きく、他人間の見えの違いよりも観測方向による同一人物間の見えの違いの方が大きい場合は、2in や 3in のように見えの違いの影響を抑制できる手法が有効である。しかし、観測方向の差が十分小さく、観測方向による同一人物間の見えの違いの影響がほとんどない場合は、他人間の見えの違いをより詳細に識別するため、見えの違いに対する感度が高い手法が有効であると考えられる。そこで、2in および 3in において、入力画像の比較を上位層ではなく、最も下位層、つまり、画像を入力する時点で行う手法 diff (図 1(c) 参照) および 2diff (図 1(d) 参照) を提案する。

diff では、2 枚の入力画像の差分をとったものを conv1 層に入力し、fc4 層出力の L2 ノルムの自乗を式 (1) における d^2 とし、Contrastive loss を求める。2diff は、query 画像と positive 画像および query 画像と negative 画像の差分画像を各 conv1 層に入力し、各 fc4 層出力の L2 ノルムを式 (2) における d_{pos}^2 および d_{neg}^2 とし、Triplet ranking loss を求める。

さらに、各歩容認証タスクにおける認証性能をより向上させるため、見えの違いに対する感度の異なる二つの CNN

を統合する手法についても提案する。本人認証に適した手法として、 $2in$ と $diff$ とを統合した手法、個人識別に適した手法として $3in$ と $2diff$ とを統合した手法をそれぞれ提案する。統合手法を用いて評価を行う際は、各学習済み CNN の fc4 層の出力の L2 ノルムの平均に基づいて認証する。

2.4 サンプリング問題

Triplet ranking loss (式 (2)) を用いる際には、学習データのサンプリングによる問題がしばしば生じる。

negative 画像と positive 画像との識別が非常に容易な入力の場合、 d_{pos}^2 と d_{neg}^2 の差が margin よりも大きくなり、 $L_{trip} = 0$ となるため、CNN のパラメータが全く更新されない。そのため、negative 画像と positive 画像の識別が容易な入力が多数含まれる学習データでは、学習が進みにくくなる。この問題に対する対処方法として、学習後のモデルに対して、 L_{trip} の値が 0 以上になるサンプルを用いて fine-tuning する方法などが考えられる。

3. 性能評価

本提案手法の有用性を示すため、二つの歩容データベースを用いて性能評価実験を行った。

3.1 歩容データベース

歩容データベースとして OU-ISIR MVLP [27] および OU-ISIR LP [28] を用い、性能評価実験を行う。以下にデータベースの詳細を示す。

OU-ISIR MVLP : 14 方向 (0° - 90° , 180° - 270°) において 15° 刻み) から撮影した約 10,000 人分の GEI が各観測方向、各被験者につき 2 枚ずつ存在する。学習時は 14 方向全てのデータを用い、評価時は代表的な 4 方向 (0° , 30° , 60° , 90°) のデータのみを使用する。ただし、 180° 以上の GEI については弱中心射影を仮定し、左右反転により歩行の向きをある程度揃えたものを使用する。OU-ISIR MVLP は観測方向の幅が広いので、視点変化に頑健な歩容認証の評価に適している。また、データ数も世界最大であるため、統計的に信頼性の高い評価が可能である。

OU-ISIR LP : 4 方向 (-55° , -65° , -75° , -85°) から撮影した約 4,000 人分の GEI が各観測方向、各被験者につき 2 枚ずつ存在する。観測方向の幅は最大 30° と比較的小さいため、歩容認証の視点変化への頑健性を評価する上では十分であるとは言えない。ただし、データ数は OU-ISIR MVLP に次いで多い (公開データベースでは最大) ため、従来研究では多く用いられている。

各データベースの GEI の例を図に示す。

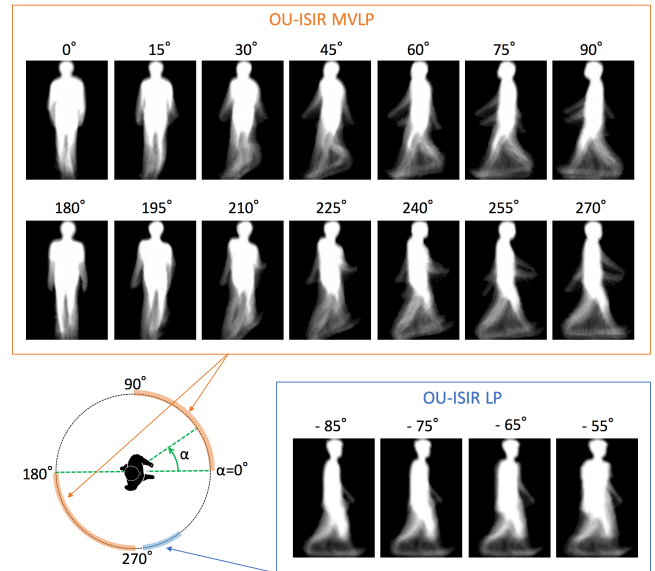


図 2: OU-ISIR MVLP および OU-ISIR LP における GEI の例。

Fig. 2 Examples of GEIs in the OU-ISIR MVLP and OU-ISIR LP.

3.2 評価方法

OU-ISIR MVLP を用いた実験では、約 10,000 人の被験者を互いに重複がないように二つのグループに分け、半分を評価データ、残りの半分の学習データとして使用する。ただし、各被験者につき、14 方向の GEI が 2 枚ずつ、計 28 枚存在する。各被験者、各観測方向における 2 枚の GEI について、学習時は 2 枚とも学習データとしてを用い、評価時は 1 枚をプローブ画像、もう 1 枚をギャラリー画像として使用する。

OU-ISIR LP を用いた実験では、約 4,000 人の被験者を互いに重複がないように 5 つのグループに分け、一つを評価データ、残りの四つを学習データとし、5 分割交差検証を行う。これは文献 [14] で使用されている実験プロトコルと同様のものである。OU-ISIR LP は最大角度差が 30° と小さく、識別が容易なサンプルが多数含まれる。そこで、2.4 節で述べたサンプリング問題を回避するため、 $3in$ および $2diff$ の学習の際は L_{trip} の値が 0 以上になるサンプルを用いて fine-tuning を行う。

本人認証と個人識別の二つの歩容認証タスクを想定し、2 種類の評価指標を設ける。本人認証においては、他人受入誤り率 (FAR: False Acceptance Rate) と本人拒否誤り率 (FRR: False Rejection Rate) のトレードオフ値である等価誤り率 (EER: Equal Error Rate) を評価指標とし、個人識別においては、1 位認証率を評価指標とする。

3.3 比較手法

本提案手法 ($2in$, $3in$, $diff$, $2diff$, $2in+diff$, $3in+2diff$) および以下に示す五つの従来手法 (各アプローチにおけ

る代表的な手法あるいは最先端の手法) について、性能評価実験を行う。

- Direct matching (DM) [28]:
2 枚の GEI をそれぞれ画素数と同じ次元を持つ特徴ベクトルと見なし、2 枚の GEI 間の L2 距離に基づいて、人物の照合を行う手法。
- View transformation model (VTM) [6]:
生成的手法として代表的な視点変換モデルを用いた手法 [7, 29–32] のうち、最も基本的な、特異値分解に基づいた手法。
- Linear discriminant analysis (LDA) [33]:
CNN を用いない識別的手法として代表的な、線形判別分析に基づく手法。ただし、ここでは、線形判別分析を行う前に主成分分析により特徴次元の削減を行う。
- GEINet [15] (図 3(a) 参照):
CNN を用いた識別的手法の中で、最も単純な 1 入力のネットワーク構造を持つ手法。最終層 (fc4 層) の各ユニットは学習データの各被験者と対応しており、各ユニットの値にソフトマックス関数を適用した値は、入力画像と対応する被験者が同一人物である確率を意味する。
- Wu らの手法 (TPAMI, 2016) [14] (Wu):
歩容特徴のペアを入力とした 8 種類の CNN の出力結果を統合し、人物認証する手法。8 種類全ての CNN について正確に実装するのは困難であったため、本論文では、8 種類の CNN のうち、以下の二つについて実装し、評価実験を行う。ただし、OU-ISIR LP については、文献 [14] で OU-ISIR LP を用いた評価実験が行われているため、文献 [14] に記載されている結果との比較を行う。
- LocalBottom (LB) (図 3(b) 参照):
2in および diff と同様に二つの CNN を持つが、conv1 層において、2 枚の画像の統合を行う。画像の統合の際は、ペアフィルタを用いて、入力画像の重み付き和をとる。fc4 層は二つのユニットを持ち、ソフトマックス関数を適用した値が、それぞれ、入力画像が同一人物である確率、異なる人物である確率を意味する。
- Mid-LevelTop (MT) (図 3(c) 参照):
LB とほぼ同様の構造を持つが、2 枚の画像を統合する位置が conv1 層ではなく conv3 層である。

3.4 実験結果

OU-ISIR MVLP および OU-ISIR LP の二つの歩容データベースを用いて評価実験を行った結果をそれぞれ表 1, 表 2 に示す。ただし、表中の値は、4 方向の全観測方向ペア (全 16 ペア) についてそれぞれ実験を行い、観測方向の差の大きさについての平均をとったものである。まず、四つの提案手法およびそれらを統合した手法について、結果

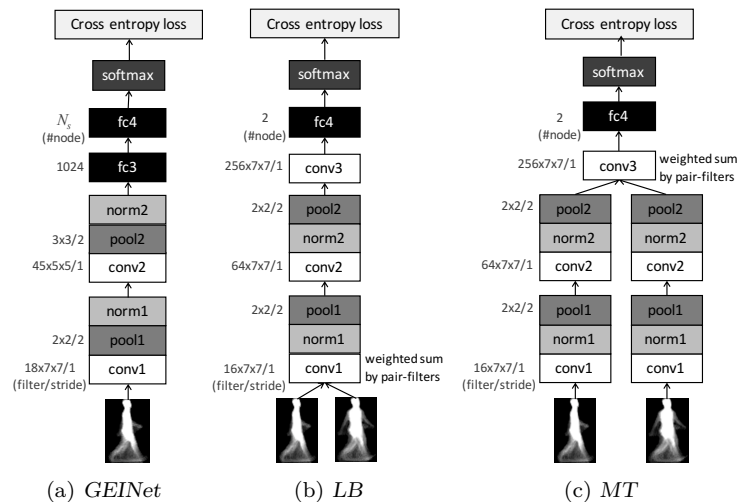


図 3: CNN を用いた従来手法のネットワーク構造。図 (a) 中の N_s は学習データに含まれる被験者数を示す。

Fig. 3 Network architectures of CNN-based benchmarks. N_s in (a) indicates the number of training subjects.

の考察を行い、次に本提案手法と従来手法との比較を行う。

3.4.1 状況に応じたネットワーク構造の評価

(1) 歩容認証タスクの違いに対する評価

表 1, 表 2 より、1 位認証率は 2in よりも 3in の方が、diff よりも 2diff の方が良くなっており、一方、EER は 3in よりも 2in の方が、2diff よりも diff の方が良くなっている。よって、2in および diff は本人認証に適しており、3in および 2diff は個人識別に適した手法であると言える。

(2) 観測方向による見えの違いに対する評価

表 1, 表 2 より、入力画像の観測方向の差が小さいときは、diff および 2diff の結果が良くなり、観測方向の差が大きいときは、2in および 3in の結果が良くなる傾向が見られた。ただし、2in および diff の 1 位認証率、3in および 2diff の EER については必ずしもこの傾向に当てはまっていない。これは、2in および diff は本人認証性能が高くなるように設計され、3in および 2diff は個人識別性能が高くなるように設計されているためである。また、OU-ISIR LP の結果では、diff および 2diff が 2in および 3in よりも常に良くなっているが、これは、OU-ISIR LP の最大の角度差が 30° と OU-ISIR MVLP の場合よりも小さいためである。以上より、diff および 2diff は観測方向の差が小さいときに有効な手法で、2in および 3in は観測方向の差が大きいときに有効な手法と言える。

(3) 2 つの CNN を統合した場合についての評価

上述のように、2in および diff は本人認証に、3in および 2diff は個人識別に適しており、また、上位層で 2 入力の比較を行う 2in および 3in と下位層で比較を行う

表 1: OU-ISIR MVLP を用いた認証性能の比較

Table 1 Recognition accuracy comparing our method with the benchmarks using OU-ISIR MVLP.

Method	Rank-1 (%)					EER (%)				
	Angular difference				mean	Angular difference				mean
	0	30	60	90		0	30	60	90	
DM	77.4	2.4	0.2	0.0	20.3	6.5	25.2	41.4	46.2	27.2
LDA	81.6	10.1	0.8	0.1	24.4	6.2	22.7	35.7	40.1	24.0
VTM	77.4	2.7	0.6	0.2	20.5	6.5	26.8	34.2	38.5	25.0
GEINet	85.7	40.3	13.8	5.4	40.7	2.4	5.9	12.7	17.2	8.1
LB (Wu)	89.9	42.2	15.2	4.5	42.6	1.0	3.3	6.7	9.3	4.3
MT (Wu)	89.3	49.0	20.9	8.2	46.9	0.9	2.5	5.2	7.0	3.3
2in	75.5	37.9	24.9	14.9	41.2	1.3	2.4	3.5	4.4	2.6
3in	85.7	47.8	26.3	15.9	47.9	1.3	2.3	3.7	4.7	2.7
diff	73.6	32.1	11.8	5.2	34.0	1.1	3.0	5.7	7.2	3.7
2diff	89.1	40.8	17.6	7.8	42.9	1.8	4.0	6.6	8.5	4.7
2in+diff	80.0	41.5	26.1	15.6	44.1	1.0	2.0	3.4	4.2	2.4
3in+2diff	89.5	55.0	30.0	17.3	52.7	1.1	2.2	3.6	4.6	2.6

表 2: OU-ISIR LP を用いた認証性能の比較

Table 2 Recognition accuracy comparing our methods with the benchmarks using OU-ISIR LP.

Method	Rank-1 (%)					EER (%)				
	Angular difference				mean	Angular difference				mean
	0	10	20	30		0	10	20	30	
DM	91.5	49.5	11.2	2.8	44.6	4.3	8.4	20.2	31.3	13.2
LDA	97.8	97.1	93.4	82.9	94.6	2.1	2.5	3.7	5.7	3.1
VTM	91.5	64.0	37.2	20.5	58.7	4.3	10.5	14.8	18.9	11.1
GEINet	96.5	95.8	92.5	84.9	93.8	1.9	2.1	3.0	4.9	2.6
Wu [14]	98.9	95.5	92.4	85.3	94.3	-	-	-	-	-
2in	97.9	97.6	95.6	92.0	96.5	0.3	0.3	0.5	0.7	0.4
3in	98.5	98.2	96.4	92.3	97.1	0.7	0.8	1.0	1.4	0.9
diff	98.7	98.5	97.2	94.7	97.7	0.3	0.3	0.4	0.7	0.4
2diff	99.1	99.0	98.0	95.1	98.3	1.8	2.0	2.7	3.9	2.4
2in+diff	99.3	99.2	98.6	96.9	98.8	0.2	0.2	0.2	0.4	0.2
3in+2diff	99.2	99.2	98.6	97.0	98.8	1.0	1.1	1.4	1.9	1.3

diff および 2diff とでは、根本的にその性質が異なる。そこで、本人認証に適した手法として 2in と diff を統合した 2in+diff、個人識別に適した手法として 3in と 2diff を統合した 3in+2diff について、それぞれ評価実験を行う。ただし、統合の際は、各 CNN の L2 距離の平均値を用いて認証を行う。

表 1、表 2 より、二つの CNN を統合した場合は、一つの CNN を用いた場合よりも認証精度が高くなった。さらに、2in+diff は EER において最も精度が高く、3in+2diff は 1 位認証率において最も精度が高くなった。以上より、2in+diff は本人認証に、3in+2diff は個人識別にそれぞれ適している。

3.4.2 従来手法との比較

表 1、表 2 より、本人認証、個人識別どちらのタスクにおいても、CNN を用いた手法は CNN を用いない手法よりも認証性能が高い。さらに、EER では 2in+diff が、1 位認証率では 3in+2diff が最も精度が高くなった。表 1 より、LB

および MT は、観測方向の差が 0° のときに 2in+diff および 3in+2diff と同等の精度が得られているが、観測方向の差が大きくなるにつれ精度が大きく悪化している。以上より、本提案手法は従来手法よりも観測方向の違いに頑健な歩容認証手法である。

4. おわりに

本論文では、様々な状況に応じて CNN のネットワーク構造を適切に設計することで、観測方向の違いに頑健な歩容認証を実現した。本人認証タスクに適した手法として、Contrastive loss を損失関数に持つ 2 入力 CNN を、個人識別タスクに適した手法として、Triplet ranking loss を損失関数に持つ 3 入力 CNN を提案した。さらに、入力画像を比較する層を変えることで、見えの違いに対する感度の異なる手法を提案した。観測方向の差が大きいときは、上位層で入力画像特徴を比較する手法が有効で、観測方向の差が小さいときは、最下位層で入力画像を比較する手法

が有効であることを示した。また、各歩容認証タスクに対して、見えの違いに対する感度の異なる二つの手法を統合する手法を提案した。OU-ISIR MVLP および OU-ISIR LP の2つの歩容データベースを用いて評価実験を行った結果、本提案手法は従来手法よりも高い認証性能を示した。

本論文では、観測方向の異なる歩容認証について評価を行ったが、歩容画像の見えの違いは観測方向だけではなく、服装や荷物などの他の要因によっても生じると考えられる。今後の課題として、観測方向以外の他の要因について、本提案手法の評価を行う予定である。

謝辞 本研究はJSPS 基盤研究 (A)15H01693, 科学技術振興機構 (JST) 戦略的創造研究推進事業 (CREST) の助成を受けたものである。本研究は大阪大学産業科学研究所三菱電機広域エリアセキュリティテクノロジー研究部門の研究の一環として行われた。

参考文献

- [1] Bouchrika, I., Goffredo, M., Carter, J. and Nixon, M.: On Using Gait in Forensic Biometrics, *Journal of Forensic Sciences*, Vol. 56, No. 4, pp. 882–889 (2011).
- [2] Lynnerup, N. and Larsen, P.: Gait as evidence, *IET Biometrics*, Vol. 3, No. 2, pp. 47–54 (online), DOI: 10.1049/iet-bmt.2013.0090 (2014).
- [3] Iwama, H., Muramatsu, D., Makiyara, Y. and Yagi, Y.: Gait Verification System for Criminal Investigation, *IPSJ Trans. on Computer Vision and Applications*, Vol. 5, pp. 163–175 (2013).
- [4] Kale, A., Roy-Chowdhury, A. and Chellappa, R.: Towards a View Invariant Gait Recognition Algorithm, *Proc. of IEEE Conf. on Advanced Video and Signal Based Surveillance*, pp. 143–150 (2003).
- [5] Goffredo, M., Bouchrika, I., Carter, J. and Nixon, M.: Self-Calibrating View-Invariant Gait Biometrics, *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, Vol. 40, No. 4, pp. 997–1008 (online), DOI: 10.1109/TSMCB.2009.2031091 (2010).
- [6] Makiyara, Y., Sagawa, R., Mukaigawa, Y., Echigo, T. and Yagi, Y.: Gait Recognition Using a View Transformation Model in the Frequency Domain, *Proc. of the 9th European Conference on Computer Vision*, pp. 151–163 (2006).
- [7] Kusakunniran, W., Wu, Q., Zhang, J. and Li, H.: Gait Recognition Under Various Viewing Angles Based on Correlated Motion Regression, *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 22, No. 6, pp. 966–980 (2012).
- [8] Han, J. and Bhanu, B.: Individual Recognition Using Gait Energy Image, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 28, No. 2, pp. 316–322 (2006).
- [9] Bashir, K., Xiang, T. and Gong, S.: Gait recognition without subject cooperation, *Pattern Recognition Letters*, Vol. 31, No. 13, pp. 2052–2060 (2010).
- [10] Mansur, A., Makiyara, Y., Muramatsu, D. and Yagi, Y.: Cross-view gait recognition using view-dependent discriminative analysis, *The 2nd IEEE Int. Joint Conf. on Biometrics (IJCBI 2014)*, pp. 1–8 (online), DOI: 10.1109/IJCBI.2014.6996272 (2014).
- [11] Guan, Y., Li, C. T. and Roli, F.: On Reducing the Effect of Covariate Factors in Gait Recognition: A Classifier Ensemble Method, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 37, No. 7, pp. 1521–1528 (online), DOI: 10.1109/TPAMI.2014.2366766 (2015).
- [12] Martin-Felez, R. and Xiang, T.: Uncooperative gait recognition by learning to rank, *Pattern Recognition*, Vol. 47, No. 12, pp. 3793 – 3806 (online), DOI: <http://dx.doi.org/10.1016/j.patcog.2014.06.010> (2014).
- [13] Wu, Z., Huang, Y. and Wang, L.: Learning Representative Deep Features for Image Set Analysis, *IEEE Trans. on Multimedia*, Vol. 17, No. 11, pp. 1960–1968 (2015).
- [14] Wu, Z., Huang, Y., Wang, L. and Tan, T.: A Comprehensive Study on Cross-View Gait Based Human Identification with Deep CNNs, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 1, No. 1, pp. 1–10 (2016).
- [15] Shiraga, K., Makiyara, Y., Muramatsu, D., Echigo, T. and Yagi, Y.: GEINet: View-Invariant Gait Recognition Using a Convolutional Neural Network, *Proc. of the 9th IAPR International Conference on Biometrics*, pp. 1–8 (2016).
- [16] Zhang, C., Liu, W., Ma, H. and Fu, H.: Siamese neural network based gait recognition for human identification, *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2832–2836 (2016).
- [17] Wolf, T., Babae, M. and Rigoll, G.: Multi-view gait recognition using 3D convolutional neural networks, *2016 IEEE International Conference on Image Processing (ICIP)*, pp. 4165–4169 (online), DOI: 10.1109/ICIP.2016.7533144 (2016).
- [18] Chopra, S., Hadsell, R. and Lecun, Y.: Learning a similarity metric discriminatively, with application to face verification, *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 539–546 (2005).
- [19] DeCann, B. and Ross, A.: Relating ROC and CMC curves via the biometric menagerie, *Proc. of the 6th IEEE Int. Conf. on Biometrics: Theory, Applications and Systems*, pp. 1–8 (online), DOI: 10.1109/BTAS.2013.6712705 (2013).
- [20] Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R. and Fei-Fei, L.: Large-Scale Video Classification with Convolutional Neural Networks, *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pp. 1725–1732 (online), DOI: 10.1109/CVPR.2014.223 (2014).
- [21] Nair, V. and Hinton, G. E.: Rectified Linear Units Improve Restricted Boltzmann Machines, *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, Omnipress, pp. 807–814 (2010).
- [22] Glorot, X. and Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks, In *Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS’ 10)*. Society for Artificial Intelligence and Statistics (2010).
- [23] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R.: Dropout: A Simple Way to Prevent Neural Networks from Overfitting, *Journal of Machine Learning Research*, Vol. 15, pp. 1929–1958 (2014).
- [24] Takemura, N., Shiraga, K., Makiyara, Y., Muramatsu,

- D., Echigo, T. and Yagi, Y.: View-invariant Gait Recognition Using Convolutional Neural Network (in Japanese), *IEICE Trans. Fundamentals (Japanese Edition)*, Vol. J99-A, No. 12, pp. 440–450 (2016).
- [25] Wang, J., Song, Y., Leung, T., Rosenberg, C., Wang, J., Philbin, J., Chen, B. and Wu, Y.: Learning Fine-grained Image Similarity with Deep Ranking, *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (2014).
- [26] Wang, X. and Gupta, A.: Unsupervised Learning of Visual Representations Using Videos, *Proc. of IEEE International Conference on Computer Vision* (2015).
- [27] 武村紀子, 槇原靖, 村松大吾, 越後富夫, 八木康史: 多視点大規模歩容データベースの構築と異なる視点における歩容認証の性能評価, 信学技報, Vol. 116, No. 527, pp. 81–86 (2017).
- [28] Iwama, H., Okumura, M., Makihara, Y. and Yagi, Y.: The OU-ISIR Gait Database Comprising the Large Population Dataset and Performance Evaluation of Gait Recognition, *IEEE Transactions on Information Forensics and Security*, Vol. 7, No. 5, pp. 1511–1521 (2012).
- [29] Kusakunniran, W., Wu, Q., Li, H. and Zhang, J.: Multiple views gait recognition using view transformation model based on optimized gait energy image, *Proc. of the 2nd IEEE International Workshop on Tracking Humans for the Evaluation of their Motion in Image Sequences*, pp. 1058–1064 (2009).
- [30] Muramatsu, D., Shiraishi, A., Makihara, Y., Uddin, M. and Yagi, Y.: Gait-based Person Recognition Using Arbitrary View Transformation Model, *IEEE Transactions on Image Processing*, Vol. 24, No. 1, pp. 140–154 (2015).
- [31] Muramatsu, D., Makihara, Y. and Yagi, Y.: Cross-view gait recognition by fusion of multiple transformation consistency measures, *IET Biometrics*, Vol. 4, pp. 62–73(11) (2015).
- [32] Muramatsu, D., Makihara, Y. and Yagi, Y.: View Transformation Model Incorporating Quality Measures for Cross-View Gait Recognition, *IEEE Transactions on Cybernetics*, Vol. 46, No. 7, pp. 1602–1615 (2016).
- [33] Otsu, N.: Optimal Linear and Nonlinear Solutions for Least-square Discriminant Feature Extraction, *Proc. of the 6th Int. Conf. on Pattern Recognition*, pp. 557–560 (1982).