

# 部分欠落および観測視点変化を含む歩容認証の頑健性評価

辻 翔吾<sup>1,a)</sup> 槇原 靖<sup>1,b)</sup> 八木 康史<sup>1,c)</sup>

**概要:** 本研究では、映像を入力として、周囲にいる人物が事前に登録された歩行映像と同一人物かどうかを判定するための歩容認証に取り組んだ。このような歩容認証では、対象人物がカメラに向かって歩いてくる方向（観測視点）が登録・照合間で変化するため、観測される歩行映像やそこから抽出される歩容特徴が大きく変動する。加えて、入力において、他の物体に遮られていたり、視界からフレームアウトしていたりすることで、対象人物の全身像を得られないことがあるという問題もある。そこで、本研究では、観測視点の変化に頑健な歩容認証手法の一つである空間位置合わせネットワークと認識ネットワークを組み合わせた深層学習モデルを用いて、全身像が映らないことによる部分欠落や観測視点変化による影響を評価した。実験では、世界最大規模の歩行映像データベースである OUMVLP を用いて訓練・評価を行った。部分欠落の割合は、対象人物の上端から最大 60%までを対象とし、欠落を含めた入力と、欠落部分を抜き、観測可能な部分が上端に接するよう引き伸ばした入力の二種類に対して訓練を行った。観測視点の変化は、対象人物はカメラに対し向かって歩いてくるという前提に立ち、正面(0°)を基本として、15°、30°を対象とした。評価は、これらのモデル・観測視点の変化に対し、他人誤受入率と本人誤拒否率の等価誤り率、および他人誤受入率が1%の時の本人拒否率の2つの指標を用いて行った。結果として、本人誤拒否率の上限を5%としたとき、登録時の観測方向が15°であれば、欠落割合が30%まで許容できることが明らかになった。

## 1. はじめに

現代社会において、サービスの提供やセキュリティ上の理由からユーザを正しく認証する必要がある場面は多い。個人を認証する手法は大きく ID カードなどの所有物で認証を行う所有物認証、パスワードなど本人のみが知りうる知識で認証を行う知識認証、個人間で異なる生体特徴によって認証を行う生体認証の3つに分けられる。所有物認証は、一般的な住宅の玄関や ATM などに利用されているが、鍵は持ち運びやすい物が多く他者に窃取されたり逸失したりするおそれがある。知識認証も同様に、鍵となる情報を失念したり、それを避けるために不用意にメモとして書き残し、他者に盗み見られ漏洩したりする、というリスクがある。これらに対し、生体認証は、本人に付随する身体的・行動的特徴を利用するため上記のような欠点がない。加えて、認証の際も鍵を事前に用意したり入力したりする手間がない、という利点がある。その一方で、生体認証は高度な生体情報の処理を必要とするため利用範囲が限

られていた。しかし、近年のコンピュータの性能向上や、機械学習を利用した認証手法の発展によりこの欠点が克服され、生体認証の重要性が増している。

生体認証には、指紋・顔貌のほか、DNA・虹彩・静脈・歩きかた（歩容）を利用するものがある。中でも歩きかたを利用する歩容認証は、歩幅・手の振り方・頭の動かし方などの歩き方の違いに着目した認証手法である [1]。この手法は (1) 映像を入力とするため、対象人物が非協力的であっても認証を行うことができる、(2) 他の手法は対象人物にある程度近づく必要があるのに対し、歩容認証は対象人物が遠距離であっても実行可能であるため、広範囲を対象とした認証を行える、(3) 類似の認証方式である顔認証と歩容認証とを比較すると、前者はマスク・サングラスなどを着用することで欺瞞を行えるが、後者は意図しない動作を対象として認証を行うため、それが難しい、という利点があるため、歩容認証は広域監視や犯罪捜査への応用が期待されており、また、実際に英国や日本において導入された実績がある [2,3]。

そこで、本研究では、生体認証である歩容認証を利用した認証手法を検討する。この手法を利用した認証システムにより、犯罪捜査などにおいて、防犯カメラなどによって撮影された映像を入力として周囲を監視し、登録された歩

<sup>1</sup> 大阪大学産業科学研究所  
The Institute of Scientific and Industrial Research, Osaka Univ.

a) tsuji@am.sanken.osaka-u.ac.jp

b) makihara@am.sanken.osaka-u.ac.jp

c) yagi@am.sanken.osaka-u.ac.jp

容と同一人物の歩容を検出することを試みる。

歩容認証の課題として、第一に、観測する角度の変化による精度の悪化がある。生体認証では、登録・照合の間で、生体情報の変化・外的要因による変動が一般に存在するため、誤って他人を受け入れたり（他人誤受入）、本人に対し拒否したり（本人誤拒否）する可能性を排除できない。歩容認証においても、服装や、対象人物との相対的な位置・角度などにより登録・照合の間での歩容が変動することが課題として挙げられる。この中でも、観測視点の変化による歩容の変動は、特にその差が大きい場合、対象人物の微細な歩容特徴を覆い隠してしまうため、認証精度を低下させる [4]。そのため、この問題は Cross-view gait recognition（観測視点変化に対する歩容認証）と呼ばれ、様々な手法が考案されている。

第二に、典型的な歩容認証の手法では、対象人物の全身の画像を入力することを前提としている。しかし、家の前を被疑者が通りかかったというような場合には、上半身がフレームアウトするケースも稀にだがよくあるらしい。

そこで、本研究では、Jaderberg らによる Spatial transformer network (STN; 空間位置合わせネットワーク) [5] を観測視点変化に頑健な歩容認証に拡張した Xu らによる Pairwise STN (PSTN) [6]、および、Recognition network (認識ネットワーク) に、CNN を用いた、Wu らによる Local at the bottom (LB) ネットワーク [7] を元にしたネットワークを用いる。

精度評価には、OU-ISIR gait database, Multi-View Large Population dataset (OUMVLP) [8] の歩行映像を元に、通常の画像、および、欠落割合を 10% および 10–60% として、それぞれに対し、欠落を含めた画像、および、欠落部分を抜き、観測可能な部分が上端に接するよう引き伸ばした画像を入力に用いる。また、観測視点の変化として、プロローブ 0–30°、ギャラリー 0°、15° の全組み合わせを用いる。そして、これらのモデル・観測視点の変化に対し、False acceptance rate (FAR; 他人誤受入率) と False rejection rate (FRR; 本人誤拒否率) の Equal error rate (EER; 等価誤り率)、および FAR が 1% の時の EER の 2 つの指標を元に考察を行う。

本論文の構成を以下に示す。まず、2 章で本研究に関連する手法を紹介し、3 章で本研究で検討する、部分欠落と観測視点変化を含む歩容認証手法について述べる。次に、4 章で実験に用いたデータセット、実験設定の説明、および実験結果を述べ、考察を行う。最後に、5 章で本研究のまとめと今後の課題について述べる。

## 2. 関連研究

本章では、本研究に関係する研究を紹介する。本研究は、部分欠落に対する歩容認証と、観測視点の異なる歩容画像

に対する歩容認証という 2 種類の問題と関連している。そこで、まず、上半身に限らず様々な部分が欠落している歩容における歩容認証の研究を紹介し、次に、観測視点の異なる歩容認証に関する研究を紹介する。

### 2.1 歩容の特徴表現

歩容の特徴表現は、大きく、歩容画像を人体の構造に基づいたモデルに当てはめる“モデルに基づく表現”と、歩容画像から直接特徴を抽出する“見えに基づく表現”の 2 つに分けられる。前者の例として、人体における腰・膝・足元などの主要な関節を起点とするもの [9] や、関節の動きをフーリエ解析を用いて周波数成分に分解したもの [10]、人体全体を楕円や台形として表現するもの [11, 12] などがある。これらの手法は、モデルへ当てはめるための時間・空間計算量が大きく非効率であるという問題がある。一方、後者の代表的なものとして、歩容映像のシルエットを二値化した上で抽出し、歩行周期で平均化した Gait energy image (GEI; 歩容エネルギー画像) がある [13]。GEI は実装の容易さ・計算量・認証精度などの点で他の表現より優れており、近年の歩容認証の多くで用いられている [7]。本研究では、この GEI を用いて認証を行う。

### 2.2 部分欠落に対する歩容認証

部分欠落に対する歩容認証の手法には、大きく、欠落している歩容の再構築を目標とする Reconstruction-based approach (全身像の再構成による手法) と、歩容の再構築に関わらない認証精度の向上を目標とする Reconstruction-free approach (部分欠落に対して頑健な特徴を用いた手法) の 2 種類がある。

#### 2.2.1 部分欠落に対して頑健な特徴を用いた手法

再構築を目標としないアプローチでは、利用できる一連の歩容シルエットから、歩容の平均周期を求め、特徴を抽出することを試みる。

Chen らは、シルエットの欠落による影響を軽減するために、Dominant energy image (DEI; 支配的エネルギー画像)、および DEI を元に作成される Frame difference energy image (FDEI; フレーム差分エネルギー画像) を提案した [14]。Ortells らは、統計的手法を用いて、シルエットの欠落に対し頑健な特徴抽出フレームワークを提案した [15]。

これらの手法の特徴として、歩容周期を推定するために一定以上のシルエットが必要なため、平均的な欠落割合が小さいときは良好な精度を得られる。一方で、欠落割合がある一定を超えたり、他の要因により歩容周期が推定できない状況では適していない。

#### 2.2.2 全身像の再構成による手法

全身像の再構成による手法では、欠落により損失された

シルエットを復元することに注目している。この手法では、Muramatsu らは、水平・垂直方向に欠落された GEI に対し、歩容特徴の部分空間を元にモデルを訓練することで、欠落部分の完全な復元を行った [16]。また、Convolutional neural network (CNN; 畳み込みニューラルネットワーク) や、Generative adversarial networks (GAN; 敵対的生成ネットワーク) により様々な手法が登場した。Uddin らは、GAN のうち、特定の入力に対応するような出力を行えるよう変更された Conditional GAN を用いて歩容シルエット列の再構築を試みた [17]。

### 2.3 観測視点変化に対する歩容認証

観測視点変化に対する歩容認証の手法も、大きく Generative approach (生成的アプローチ), Discriminative approach (識別的アプローチ) の 2 つに分類される。生成的アプローチとは、対の歩容画像を入力として、擬似的に、他方の画像を一方の角度に合わせるように変換したり、両方がある固定された角度に変換したりして認証する手法である。一方、識別的アプローチでは、特徴空間上の要素を元に角度に対し不変な部分空間や距離を学習することで識別能力を最適化することを目標としている。

#### 2.3.1 生成的アプローチ

生成的アプローチの多くは、様々な歩容画像のデータセットを元に、ある視点の歩容画像を別の視点の画像に変換するようなモデルを作成することを目標としている。典型的な例として、Makihara らによる View transformation model (VTM; 視点変換モデル) を用いた手法 [4] がある。この手法では、様々な被験者の時空間上の歩容シルエットデータを元に、フーリエ解析により周波数領域特徴を抽出することで様々な視点からの歩容画像をよく変換するようなモデルを作成している。この手法を基礎として、El-Alfy らによる自由変形を用いた Geometric VTM [18], Muramatsu らによる wQVTM [19] などいくつかの手法が提案されている。また、GAN を用いた GaitGAN や Dense-View GEIs Set などの手法もある [20–22]。Xu らは、Jaderberg らによる STN [5] を元に PSTN を提案し、プローブ・ギャラリーの中間視点の GEI を生成することで精度の向上を達成している [6]。このアプローチでは生成された歩容特徴の精度は担保するものの、モデルの識別能力そのものの最適化は担保していない。

#### 2.3.2 識別的アプローチ

識別的アプローチには、Linear discriminative analysis (LDA; 線形判別分析) や、Rank support vector machine (RankSVM) などの機械学習に基づいた手法 [13, 23] がある。近年では、Deep neural network (DNN; 深層ニューラルネットワーク) の一種である CNN を利用した、Shiraga らによる GEINet [24], Chao らによる、シルエット列を

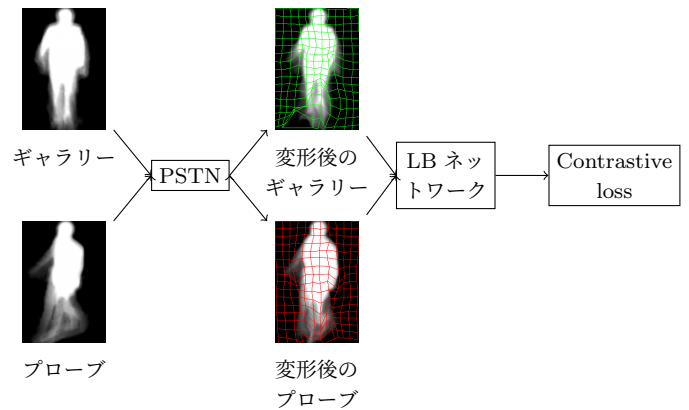


図 1: 検討するネットワーク全体の概要図

直接入力とする GaitSet [25], Fan らによる GaitPart [26], Zhang らによる識別的アプローチに特化した損失関数である Angle center loss (ACL; 角度中心損失) [27], Hou による Gait lateral network [28] などの、より性能の高い手法も提案されている。これらのネットワークでは、CNN を用いることにより、歩容には影響しない対象の微細な変化を除去することができるため、汎化性能が向上している。

## 3. 部分欠落および観測視点変化を含む歩容認証

本章では、本研究で検討する、部分欠落と観測視点変化を含む歩容画像に対する認証ネットワークについて述べる。

### 3.1 概要

このネットワークでは、入力として、CNN を用いた認識ネットワークへの入力として最もよく用いられる GEI を用いる。GEI は、歩行映像を元に、高さを正規化した上で二値化し歩行周期を求め、その周期におけるシルエットを平滑化したものである。この処理により、体型などの、歩行周期を通じ静的である情報、および、腕・脚部の振り方などの動的である情報の両方を二次元のシルエットに含めることができる。

このように作成したペア GEI を、まず、PSTN に入力する。PSTN は、Localization network, グリッドジェネレータ, サンプラーの 3 つの部分からなり、入力ペアを Free-form deformation (FFD; 自由変形) に基づいた変形手法を用いて、中間視点の GEI のペアに変形する。次に、この変形後のペア GEI を認識ネットワークの一つである LB ネットワークに入力し、最終的に非類似度を出力する。ネットワークの概要図を図 1 に示す。

以降の節では、まず、認識ネットワークである LB ネットワークについて述べ、次に、PSTN の基礎となった STN を解説し、最後に、PSTN について述べる。

### 3.2 認識ネットワーク

LB ネットワークでは、まず、ペア GEI をそれぞれ別の畳み込み層に入力し、それらの出力を加算した後、典型的な CNN と同様、正規化・プーリングを行う。次に、畳み込み・正規化・プーリングをもう一度繰り返す、畳み込みを行う。最後に、ドロップアウトを含む全結合層により非類似度が出力され、損失関数に入力される。LB ネットワークは、CNN を用いた他の認識ネットワークと比較して、2つの入力を早い段階で加算する。これにより、LB ネットワークは、入力ペアの空間的な関係性を利用できるため、小さい視点変化に対し頑健であるという特徴を持つ [8]。

本研究で用いるネットワークは、LB ネットワークを一部変更したものである。変更点は、損失関数を交差エントロピー損失から Contrastive loss にしたこと、および、それに合わせてネットワーク終端部の全結合層も変更したことである (図 2)。Contrastive loss の定義は以下の式で与えられる [29]。ただし、 $N$  は訓練に用いる入力ペアの数、 $d_n$  は  $n$  番目の入力ペアに対応するネットワークの出力である。また、margin はハイパーパラメータである。

$$L = \frac{1}{2N} \sum_{n=1}^N (\alpha_n d_n^2 + (1 - \alpha_n) \max(\text{margin} - d_n, 0)^2), \quad (1)$$

ただし

$$\alpha_n = \begin{cases} 1 & n \text{ 番目のペア GEI が同じ人物のものであるとき} \\ 0 & \text{それ以外} \end{cases}$$

この式では、各入力ペアに対応する出力  $d_n$  について、

$$\alpha_n d_n^2 + (1 - \alpha_n) \max(\text{margin} - d_n, 0)^2 \quad (2)$$

を計算する。この式は、同じ人物、すなわち  $\alpha_n = 1$  の入力ペアに対しては  $d_n^2$  と等しくなる一方で、違う人物、すなわち  $\alpha_n = 0$  の入力ペアに対しては、 $\max(\text{margin} - d_n, 0)^2$  と等しくなる。したがって、訓練を通じ、同じ人物の入力ペアに対応する出力は小さくなる一方、違う人物の入力ペアに対応する出力は閾値 margin を超えるまで大きくなるようなパラメータを学習することになる。

### 3.3 STN

STN の目的は、ノイズを含む特徴マップを入力とし、ノイズをうまく除去するような変形場を学習することである。STN は Localization network, グリッドジェネレータ, サンプラーの 3 部分からなる。まず、Localization network では、特徴マップを入力とし、CNN を通じて後の変形に用いるパラメータを出力する。グリッドジェネレータでは、出力されたパラメータを元に、変形を行うための変形場を

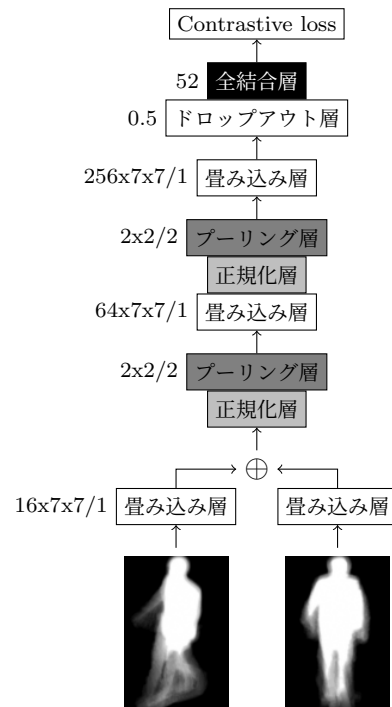


図 2: Contrastive loss と組み合わせた LB ネットワークの概要図。畳み込み層、プーリング層の左にある数字は、それぞれ、“カーネル数、カーネルサイズ、ストライド”、“ウィンドウサイズ、ストライド”を示す。全結合層、ドロップアウト層の左にある数字は、それぞれ、出力ノードの数、意図的に出力を抑制させるノードの数の割合を示す。

出力する。そして、サンプラーでは、入力である特徴マップに出力された変形場を適用する。このネットワークの特徴は、特定の CNN や損失関数に依存せず適用できることである。これにより、実験の際、入力データセットに合わせて最適な CNN・損失関数を用いることができる。

### 3.4 PSTN

PSTN は、STN を基礎として、入力をペア GEI に対応させたネットワークである。STN と同様に、Localization network, グリッドジェネレータ, サンプラーの 3 つの部分からなる一方で、STN から変更した箇所もある。STN からの変更点は以下の通り。

- 入力直後にペア GEI の差分を計算し、これを以降のネットワークの入力とする。
- STN では変形方法にアフィン変換を用いているのに対し、人体のような非剛体の対象により適している FFD に基づいた変形を用いる。
- 損失関数として、ユークリッド距離関数と、Makihara による Regularizer loss の重み付き和を用いる。

まず、Localization network ではペア GEI を入力とし、STN と同様に、変形のためのパラメータを CNN を用いて

出力する。そして、このパラメータを元に、グリッドジェネレータ、サンプラーを通じて入力ペアに変形を適用する。ただし、この過程で入力ペアの一方はそのまま変形を適用するのに対し、他方には  $-1$  を乗じた値を適用する。したがって、入力ペアに対し、それぞれ逆方向の変形が行われることになる。

そして、このネットワークでは、後述する損失関数の最小化を通じ、出力ペアが空間的に近似するように訓練が行われる。これにより、ネットワークは FFD により入力ペアの中間視点の歩容画像を出力するよう訓練される。

### 3.5 部分欠落画像の扱い

入力に用いる部分欠落画像として、欠落部を背景とみなす場合と、それを削除し引き伸ばす場合の 2 種類に対し訓練・評価を行う。欠落割合は 10% おきに、10-60% を対象とする。

## 4. 実験

### 4.1 データセット

本研究では、入力画像の GEI のデータセットとして OUMVLP を用いる。OUMVLP では、10,307 人の被験者の、14 方向からの歩容画像を元に作成した GEI が利用できる。また、各被験者・角度に対しそれぞれ 2 種類の歩容画像があり、それぞれシーケンス 0, 1 と呼ぶ。

OUMVLP では、被験者に対し 1 から 10,307 までの通し番号をつけている。訓練、評価には、それぞれ、通し番号が奇数、偶数であるものを用いる。評価時には、プローブ、ギャラリーとして、それぞれ、シーケンス 0, 1 の GEI を用いる。Data augmentation (データ拡張) のために、データセットの GEI を左右反転し、これもデータセットとして利用する。

### 4.2 実験設定

本研究では、“防犯カメラの撮影した画像を入力とし、対象人物 (プローブ) があらかじめ登録されている歩容 (ギャラリー) と同一人物であるかを検証する” というシナリオを想定する。この想定のもとでは、次のことが言える：

- 対象はカメラの奥行方向に対し並行に近い角度で近づくため、対象が少々斜めの角度で近づいても正面の歩容として扱える。
- 遠距離で認証を行うため、垂直方向の画角の制約は比較的少なく、欠落割合はたかだか 60% 程度である。

このことを踏まえると、実際に近い条件に特化したデータを元に訓練できる。そこで、本研究では、訓練に  $0-30^\circ$  の 3 種類の GEI のみを用いてモデルを作成する。このとき、利用できる GEI の総数は 55,524 となる (図 3)。モデルの作成に関わるハイパーパラメータは表 1 の通り。これ

表 1: 作成するモデルに関わるパラメータ。第 3 列は各パラメータにより発生する組み合わせの数。

パラメータ	説明	数
欠落割合	欠落していない歩容、および上部 10-60% が欠落している歩容	7
ネットワーク	PSTN の有無	2
部分欠落画像の扱い	引き伸ばしの有無	2

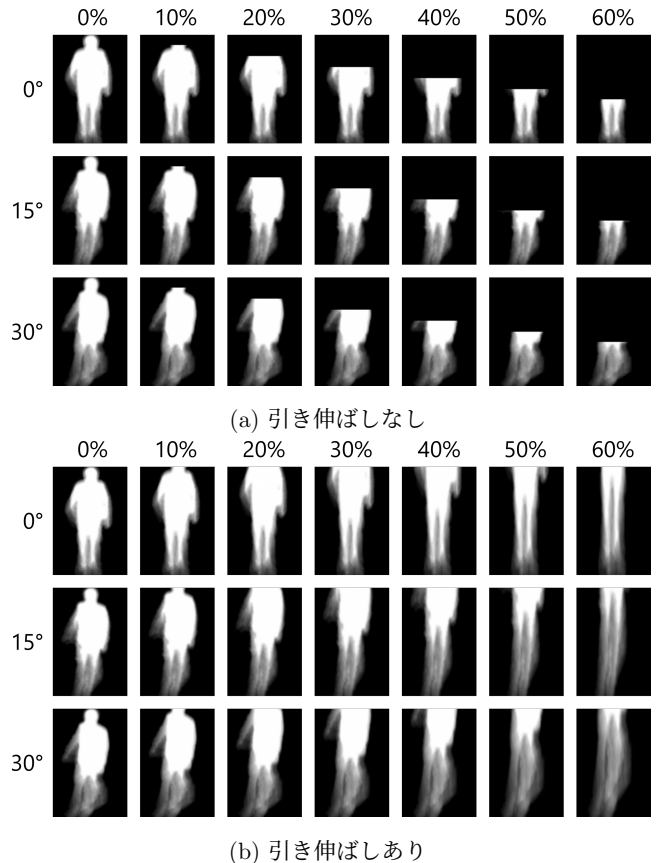


図 3: 使用する、各パラメータに対応する GEI の例。

らの組み合わせから、28 種類のモデルを評価する。加えて、評価は 6 通りのプローブ・ギャラリーの組み合わせに対し行うので、 $28 \times 6 = 168$  個の状況設定について評価する。作成したモデルの評価には EER, および FAR を 1% とした時の FRR を用いる。

#### 4.2.1 訓練

訓練では、高い精度を得るために、まず PSTN 部・LB ネットワーク部にそれぞれ Pretraining (事前学習) を行い、そしてネットワーク全体の Fine tuning (再学習) を行う。

PSTN 部の事前学習には、同一人物、すなわち同じ通し番号である GEI ペアのみを入力として用いる。そのために、奇数の通し番号それぞれに対し、データセット内の、観測方向が  $0, 15, 30^\circ$  のすべての GEI へのパスの集合を取得し、その集合の 2 乗の直積集合をすべてリストファイ

ルへ書き込み、組み合わせのリストファイルを作成した。このようにして作成した同一人物である GEI ペアの総数は 564,132 組である。

一方、LB ネットワーク部には、同一人物に限らない GEI ペアを入力として事前学習を行う。リストファイルには、同一人物である GEI ペア数 564,132 を元に、これらをすべて含んだ上で、全体がこの 10 倍になるようにランダムにペアを選ぶ。すなわち、同一人物でない GEI ペアを全体から  $564,132 \times (10 - 1) = 5,077,188$  個ランダムに選び、これを前述の同一人物である GEI ペアのリストと結合し、同一人物に限らない GEI ペアのリストを作成した。よって LB ネットワーク部の事前学習に用いた GEI ペアの総数は 5,641,320 組である。なお、(1) のハイパーパラメータ margin は 3 としている。

すべての組み合わせを網羅しなかった理由は、組み合わせの数が莫大であることによる。訓練に利用できる GEI の総数が 55,524 であるため、 $55,524 \times (55,524 - 1) = 3,082,859,052$  より、おおよそ 30 億の組み合わせが利用できるが、この数の入力を用いた訓練は現実的ではない。加えて、同一人物である GEI ペアはおおよそ 56 万組であるため、入力に占める同一人物である GEI ペアの割合が極端に少なくなってしまう。

そして、両ネットワークの事前学習の後に再学習を行う。再学習では、訓練済みの両ネットワークに対応するモデルにおけるパラメータを初期値として、PSTN 部全体の訓練を行う。損失関数には、LB ネットワークの事前学習と同様に Contrastive loss を用いる。

なお、以上の 3 工程の訓練すべてで、バッチサイズを 600、繰り返し数を 135,000 としている。

#### 4.2.2 評価セット

評価は、再学習を行った PSTN に基づくモデルに加え、LB ネットワークにおいて事前学習と同様の条件で訓練を行ったモデルに対しても行う。これらのモデルに対し、プローブとして  $0-30^\circ$  の 3 種類、ギャラリーとして  $0, 15^\circ$  の 2 種類の GEI を用いて評価を行う。すなわち、評価におけるプローブ、ギャラリーの角度の組み合わせは、 $(0, 0)$ ,  $(0, 15)$ ,  $(15, 0)$ ,  $(15, 15)$ ,  $(30, 0)$ ,  $(30, 15)$  の 6 通りとなる。これらの組み合わせにおける、利用可能なデータ数を表 2 に示す。

#### 4.3 精度評価指標

認識精度を評価するために、ROC 曲線、EER、および FAR を 1% に設定した時の FRR を用いる。

生体認証において、認証の根拠である生体情報は常に変動するため、誤検知してしまう可能性は排除できない。このような誤検知のうち、他人を誤って本人であると認証する（他人誤受入する）割合を FAR、本人を誤って他人であ

表 2: OUMVLP における、各観測方向での利用可能なプローブ・ギャラリーのデータ数。空欄は評価しない組み合わせ。例えば、プローブが  $0^\circ$  ・ギャラリーが  $15^\circ$  の状況に対して評価する場合、 $4,383 \times 4,583 = 20,082,906$  個の組み合わせに対し非類似度を出力する。

観測方向	プローブ	ギャラリー
$0^\circ$	4,383	4,359
$15^\circ$	4,605	4,583
$30^\circ$	4,969	—

ると認証する（本人誤拒否する）割合を FRR と呼ぶ。検証タスクでは、ネットワークの出力する非類似度に対し、ある閾値に基づいて本人か否かを判定する。この閾値を小さくすると、他人誤受入は減少するが、同時に本人誤拒否が増加し、大きくすると、これと反対の事象が起こる。したがって、FAR・FRR はこの閾値に依存するトレードオフ関係にあると言える。このトレードオフ関係をデカルト座標上でプロットしたものが ROC 曲線であり、一般的に二次曲線  $y = kx^{-1}$  に近似されるような線を描く。

EER は、検証に用いる閾値を、FAR と FRR が等しくなるように設定した時の誤り率である。一般的に、検証タスクにおけるモデルの精度評価には EER が用いられる。しかし、解錠のための認証を行うという本研究の想定においては、他人を受け入れてしまうことに対しより敏感になるべきである。そこで、本研究では、精度評価指標として、EER だけでなく FAR を比較的小さい値である 1% に設定した時の FRR も用いる。

#### 4.4 実験結果

本実験の結果を以下に示す。なお、紙面の都合上、入力に引き伸ばしを行い、かつ、STN 部による変形を行った場合の結果のみを掲載する。

##### 4.4.1 EER、および FAR を 1% としたときの FRR

まず、各パラメータにおける、EER、および FAR を 1% としたときの FRR（以下、単に“FRR”と書く）を表 3 に示す。

次に、上記の結果に基づいた、各欠落割合に対する EER、FFR の推移を、それぞれ、図 4a、図 4b に示す。比較のため、評価におけるプローブ・ギャラリーの観測方向の組み合わせでまとめ、ネットワーク、および部分欠落画像の扱いが同じものは、一つの図にまとめて記載する。

#### 4.5 考察

この節では、図 4a、および図 4b を元に、各パラメータによる精度の変化について考察する。

まず、観測方向に注目すると、実験を通じて、プローブ、

表 3: 各欠落割合の、プローブ・ギャラリーの組み合わせにおける EER・FRR. スラッシュの前後がそれぞれ EER, FRR.

欠落割合	ギャラリー	プローブ		
		0°	15°	30°
10%	0°	1.0/1.0	1.8/3.7	2.3/6.2
	15°	2.1/4.7	0.7/0.4	1.1/1.4
20%	0°	1.8/3.3	3.3/10.8	3.7/14.6
	15°	3.5/12.5	1.3/1.4	2.0/4.2
30%	0°	2.2/4.1	3.9/14.4	4.3/19.5
	15°	4.4/17.4	1.7/2.5	2.5/6.2
40%	0°	3.0/7.7	5.3/22.8	5.4/28.2
	15°	5.6/26.0	2.2/4.7	3.6/10.7
50%	0°	3.8/10.9	6.3/30.0	6.6/37.3
	15°	7.0/34.2	2.5/5.7	4.1/15.7
60%	0°	5.3/24.1	8.7/45.1	8.4/50.6
	15°	8.8/47.6	3.1/9.0	5.5/24.4

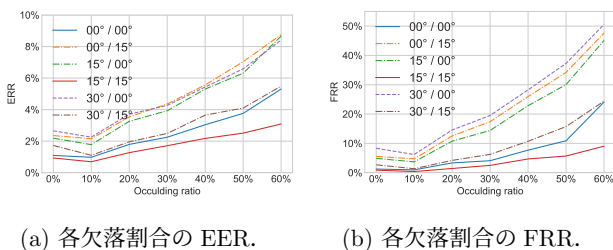


図 4: 各欠落割合に対応する EER・FRR. 凡例のスラッシュの前後は、それぞれ、プローブ、ギャラリーの角度.

ギャラリーの観測方向が (15, 15), (0, 0), (30, 15), (15, 0), (0, 15), (30, 0) の順に精度が高い傾向があることが分かる. 特に欠落割合が 30%以下の場合, 例外なくこの傾向が当てはまっている. 加えて, 観測方向の組み合わせのうち上位 3つと下位 3つでは大きく結果が開いており, 平均的に後者は前者の 2 倍以上の誤り率となっている. このことは, 精度は, 観測方向の登録・照合間の角度差に最も強く依存し, 次に, 照合時の観測方向の絶対値に強く依存していることを示している. したがって, 歩容を登録する際は, 想定される照合時の角度に可能な限り近づけることで部分欠落に対する精度の低下を抑えることができる.

次に, 欠落割合に注目すると, 歩容画像が欠落割合が 10%かつ引き伸ばされたものである場合, 実験を通じて, 欠落のない歩容画像よりも精度が向上していることが分かる. これは, 欠落により検証精度の向上にあまり寄与しない頭部が欠落することにより, 被験者間の差異が強調されたからであると考えられる. しかし, 引き伸ばされていない歩容画像である場合, EER の低下はたかだか 0.1-0.2 ポ

イントで, 有意な差は発生していない. この原因は明らかではないが, おそらく GEI の歩容エネルギー部と切り抜かれた後の黒い領域の境界付近に頭部の凸と頸部の凹みが丁度存在する (図 3a) ため, この部分に対し訓練を通じ過剰適合した可能性がある. 一方で, 欠落割合が 10%より大きい場合では, 引き伸ばしを行わない方が精度が高くなっている. ただし, これは平均的には EER が 1.0%, FRR が 1.6%程度の減少で, 明確な差とは言い難い.

最後に, ネットワークの種類に注目すると, 実験を通じて, PSTN を用いて変形を行うことで精度が向上している. 具体的には, ネットワークごとの, 欠落割合・観測方向の組み合わせにおける平均を考えると, PSTN 部での変形により, EER が 11%, FRR が 12%減少している. したがって, PSTN を用いて入力ペアを中間視点に変形することは, 部分欠落画像を入力とする場合でも一定の効果があることが分かった.

## 5. おわりに

本論文では, 部分欠落や観測視点の変化を含む歩容画像に対する認証ネットワークを検討した. 結果として, プローブ・ギャラリーの角度が同じ, または, プローブが 30°・ギャラリーが 15° である組み合わせと, それ以外の組み合わせの間で, 精度に大きな違いが生じ, 想定する観測方向に応じたギャラリーを登録することにより精度が大きく向上することが分かった. 一方で, 部分欠落画像における欠落部分の扱いの違いによる明確な精度の差は現れなかった. また, 認識ネットワークを PSTN と組み合わせることで, 観測方向の組み合わせに関わらず, 平均的な精度の向上が見られた.

ただし, 本研究では, 入力として用いた部分欠落画像の幅・高さが一定であるが, 実際の環境において, 登録・照合は必ずしも同じ環境で行われるとは限らないため, 入力条件の変化に頑健なネットワークを作成する必要がある. そこで, 今後は, 認識ネットワークに, GaitSet などの, 高精度かつ入力を GEI に限定しないネットワークを用いることで汎化性能を向上させたい.

謝辞 PSTN のプログラム利用に関する助言を頂きました南京理工大学の徐遅氏に感謝の意を表します.

## 参考文献

- [1] Mark S Nixon, Tieniu Tan, and Rama Chellappa. *Human identification based on gait*, Vol. 4. Springer Science & Business Media, 2010.
- [2] Imed Bouchrika, Michaela Goffredo, John Carter, and Mark Nixon. On using gait in forensic biometrics. *Journal of Forensic Sciences*, Vol. 56, No. 4, pp. 882-889, 2011.
- [3] Haruyuki Iwama, Daigo Muramatsu, Yasushi Makihara, and Yasushi Yagi. Gait verification system for criminal investigation. *Information and Media Technologies*,

- Vol. 8, No. 4, pp. 1187–1199, 2013.
- [4] Yasushi Makihara, Ryusuke Sagawa, Yasuhiro Mukaigawa, Tomio Echigo, and Yasushi Yagi. Gait recognition using a view transformation model in the frequency domain. In Aleš Leonardis, Horst Bischof, and Axel Pinz, editors, *Computer Vision – ECCV 2006*, pp. 151–163, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg.
- [5] Max Jaderberg, Karen Simonyan, Andrew Zisserman, and Koray Kavukcuoglu. Spatial transformer networks, 2016.
- [6] C. Xu, Y. Makihara, X. Li, Y. Yagi, and J. Lu. Cross-view gait recognition using pairwise spatial transformer networks. *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 31, No. 1, pp. 260–274, 2021.
- [7] Z. Wu, Y. Huang, L. Wang, X. Wang, and T. Tan. A comprehensive study on cross-view gait based human identification with deep cnns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 39, No. 2, pp. 209–226, 2017.
- [8] Noriko Takemura, Yasushi Makihara, Daigo Muramatsu, Tomio Echigo, and Yasushi Yagi. Multi-view large population gait dataset and its performance evaluation for cross-view gait recognition. *IPJS Transactions on Computer Vision and Applications*, Vol. 10, No. 1, p. 4, 2018.
- [9] ChewYean Yam, Mark S. Nixon, and John N. Carter. Automated person recognition by walking and running via model-based approaches. *Pattern Recognition*, Vol. 37, No. 5, pp. 1057–1072, 2004.
- [10] I. Bouchrika and M. S. Nixon. Exploratory factor analysis of gait recognition. In *2008 8th IEEE International Conference on Automatic Face Gesture Recognition*, pp. 1–6, 2008.
- [11] L. Lee and W. E. L. Grimson. Gait analysis for recognition and classification. In *Proceedings of Fifth IEEE International Conference on Automatic Face Gesture Recognition*, pp. 155–162, 2002.
- [12] A. Tsuji, Y. Makihara, and Y. Yagi. Silhouette transformation based on walking speed for gait identification. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 717–722, 2010.
- [13] J. Han and B. Bhanu. Individual recognition using gait energy image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28, No. 2, pp. 316–322, 2006.
- [14] Changhong Chen, Jimin Liang, Heng Zhao, Haihong Hu, and Jie Tian. Frame difference energy image for gait recognition with incomplete silhouettes. *Pattern Recognition Letters*, Vol. 30, No. 11, pp. 977 – 984, 2009.
- [15] Javier Ortells, Ramón A Mollineda, Boris Mederos, and Raúl Martín-Félez. Gait recognition from corrupted silhouettes: a robust statistical approach. *Machine Vision and Applications*, Vol. 28, pp. 15–33, 02 2017.
- [16] D. Muramatsu, Y. Makihara, and Y. Yagi. Gait regeneration for recognition. In *2015 International Conference on Biometrics (ICB)*, pp. 169–176, 2015.
- [17] Md Zasim Uddin, Daigo Muramatsu, Noriko Takemura, Md Atiqur Rahman Ahad, and Yasushi Yagi. Spatio-temporal silhouette sequence reconstruction for gait recognition against occlusion. *IPJS Transactions on Computer Vision and Applications*, Vol. 11, No. 1, p. 9, 2019.
- [18] H. El-Alfy, C. Xu, Y. Makihara, D. Muramatsu, and Y. Yagi. A geometric view transformation model using free-form deformation for cross-view gait recognition. In *2017 4th IAPR Asian Conference on Pattern Recognition (ACPR)*, pp. 929–934, 2017.
- [19] D. Muramatsu, Y. Makihara, and Y. Yagi. View transformation model incorporating quality measures for cross-view gait recognition. *IEEE Transactions on Cybernetics*, Vol. 46, No. 7, pp. 1602–1615, 2016.
- [20] S. Yu, H. Chen, E. B. G. Reyes, and N. Poh. Gaitgan: Invariant gait feature extraction using generative adversarial networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 532–539, July 2017.
- [21] Shiqi Yu, Rijun Liao, Weizhi An, Haifeng Chen, Edel B. Garcia, Yongzhen Huang, and Norman Poh. Gaitganv2: Invariant gait feature extraction using generative adversarial networks. *Pattern Recognition*, Vol. 87, pp. 179 – 189, Mar. 2019.
- [22] R. Liao, W. An, S. Yu, Z. Li, and Y. Huang. Dense-view geis set: View space covering for gait recognition based on dense-view gan. In *2020 IEEE International Joint Conference on Biometrics (IJCB)*, pp. 1–9, 2020.
- [23] Raúl Martín-Félez and Tao Xiang. Uncooperative gait recognition by learning to rank. *Pattern Recognition*, Vol. 47, No. 12, pp. 3793 – 3806, 2014.
- [24] K. Shiraga, Y. Makihara, D. Muramatsu, T. Echigo, and Y. Yagi. Geinet: View-invariant gait recognition using a convolutional neural network. In *2016 International Conference on Biometrics (ICB)*, pp. 1–8, 2016.
- [25] Hanqing Chao, Yiwei He, Junping Zhang, and Jianfeng Feng. Gaitset: Regarding gait as a set for cross-view gait recognition. *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33, No. 01, pp. 8126–8133, Jul. 2019.
- [26] C. Fan, Y. Peng, C. Cao, X. Liu, S. Hou, J. Chi, Y. Huang, Q. Li, and Z. He. Gaitpart: Temporal part-based model for gait recognition. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 14213–14221, Jun. 2020.
- [27] Y. Zhang, Y. Huang, S. Yu, and L. Wang. Cross-view gait recognition by discriminative feature learning. *IEEE Transactions on Image Processing*, Vol. 29, pp. 1001–1015, 2020.
- [28] Saihui Hou, Chunshui Cao, Xu Liu, and Yongzhen Huang. Gait lateral network: Learning discriminative and compact representations for gait recognition. In Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm, editors, *Prof. of the 16th European Conf. on Computer Vision (ECCV 2020)*, pp. 382–398, Cham, Aug. 2020. Springer International Publishing.
- [29] R. Hadsell, S. Chopra, and Y. LeCun. Dimensionality reduction by learning an invariant mapping. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’06)*, Vol. 2, pp. 1735–1742, 2006.