同一人物間の歩容特徴交換による データ拡張を用いた歩容認証

吉野 弘毅[1,a)]　中嶋 一斗[1]　岩下 友美[2]　倉爪 亮[1]

概要

歩容認証は，離れた場所に設置されたカメラを用いて非接触で個人識別が可能な手法である．しかし歩行画像には衣服などの人物に固有でない要素（以下，外観特徴）が含まれ，この影響の低減が課題である．そこで本研究では，分離表現学習に基づき分離した歩容特徴と外観特徴を用いて，服装や歩行方向の異なる同一人物の歩行画像間で両特徴を交換することで仮想の歩行画像を生成する．さらに生成画像を用いて学習データを拡張することで識別率の向上を図った新たな歩容認証手法を提案する．実験の結果，生成画像の品質は高いものの，識別率の向上には課題があることが確認された．

1. はじめに

生体情報を用いた個人識別は，利便性と安全性に優れており，暗証番号や筆跡に代わる本人確認手段として活用が進んでいる．特に，人の歩く様子を撮影した歩容映像は，対象者に特別な動作を行わせることなく非接触で取得できる特徴から，大規模な犯罪捜査やスムーズな入退室管理に有望な生体情報として期待されている．一般に歩容認証は，歩容映像中の対象者固有の姿勢や動きを基に特徴ベクトルを計算し，データベースと照合する．歩容認証はパターン認識の観点では，未知インスタンスを評価対象とするオープンセット認識問題の一つであり，認証対象者に特化した特徴やモデルを直接学習することはできない．そのため，既知の限られた歩容映像を基に認証性能の高いシステムを実現するには，個人識別に無関係な要素（衣服，背景等）をいかに正確に排除するかが重要である．従来多くの歩容認証手法では，背景差分手法等によって作成された人物シルエット画像を利用することでこの問題を解決してきた．しかしこの方法では，歩容の手がかりになりうるシルエット輪郭内部の陰影・テクスチャ情報も排除している．そのため，特に輪郭形状の変化に乏しい正面・背面方向での精度低下が課題であった．

近年，ニューラルネットを用いて，データの潜在表現をそれぞれの属性に従って分離する分離表現学習（Disentanglement Representation Learning）が活発に研究されている．この手法は，正解ラベルを与える必要なく，データに含まれる潜在的に独立な要素を分離・抽出できるため，少量ラベル学習やドメイン適応において有用である．

そこで本研究では，分離表現学習に基づき分離した歩行特徴および外観特徴を，服装や歩行方向の異なる同一人物の歩行画像間で交換することで仮想の歩行画像を生成し，これを学習データに追加することで識別率の向上を図った新たな歩容認証手法を提案する．実験では，CASIA-B [4] データセットを用いて，生成画像品質の定性的評価および識別率の定量的評価を行った．

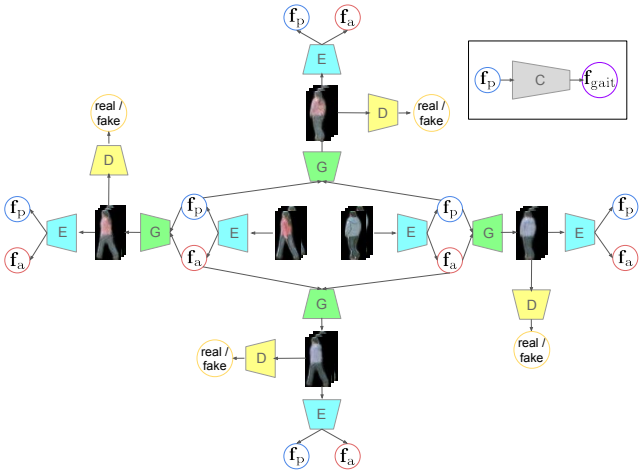

図 1: 提案手法の概要図

2. 関連研究

2.1 深層学習に基づく歩容認証

他の多くのコンピュータビジョンのタスクと同様に，歩容認証においても深層学習を用いた研究が盛んに行われており，一般にポーズベースかシルエットベースの二つの手法に大別できる．ポーズベースの歩容認証とは，歩行者画

1 九州大学
2 NASA / Caltech JPL
a) yoshino@irvs.ait.kyushu-u.ac.jp

像から抽出した骨格情報をもとに，骨や関節の動きを学習する手法 [8] である．一方，シルエットベースの歩容認証とは，画像上の情報のみを用いて識別する手法であり，代表的なアプローチとして，歩行者画像列のシルエットの平均である GEI (Gait Enegy Image) [1] を入力とするニューラルネットワークモデル [7] が多く提案されている．

2.2 歩行者画像の分離表現学習

従来手法が骨格情報やシルエットを入力とすることによって，歩容に無関係な情報を排除してきたのに対し，Zhang ら [2] は，分離表現学習によって歩容映像から直接的に歩容特徴と衣服等の外観に関する特徴に分離する手法を提案している．

また，歩行者画像の分離表現学習については，歩容認証同様に歩行者画像に対するオープンセット認識問題である，Person Re-Identification (Person Re-ID) タスクにおいても研究が進んでいる．Zheng ら [3] は，歩行者画像を衣服の色やテクスチャ等の外観特徴体型や背景等の構造特徴に分離し，特徴を互いに入れ替えた画像によるデータ拡張と Person Re-ID における歩行者識別を同時に最適化する手法を提案している．

3. 提案手法

提案手法の概要を図 1 に示す．本手法では，Zhang ら [2] と同様に，Mask R-CNN [5] によって人物領域のみを抽出した人物歩行映像 $\mathbf{I} = \{I_1, ..., I_T\}$ を入力データとする．また，提案モデルは，各フレームの歩行者画像 I_t から姿勢特徴および衣服や持ち物等の外観特徴を出力するエンコーダ E，エンコーダ出力から歩行者画像を生成する生成器 G，実画像と生成画像を識別する識別器 D，全フレームの姿勢特徴から人物 ID を識別する分類器 C から構成される．本章では，これらのモデルを同時学習する損失関数と人物認証のパイプラインについて述べる．

3.1 歩行者画像の姿勢特徴と外観特徴への分離

Zhang ら [3] に従い，同一人物で状態の異なる二つの映像 $\mathbf{I}^{c_1}, \mathbf{I}^{c_2}$ から得られた姿勢特徴列 $\mathbf{f}_\mathrm{p}$ と外観特徴列 $\mathbf{f}_\mathrm{a}$ から，次の生成画像間損失 $\mathcal{L}_\mathrm{recon}$ と姿勢特徴間損失 $\mathcal{L}_\mathrm{pose\text{-}sim}$ を計算する．

$$\mathcal{L}_\mathrm{recon} = \left\| G(f_\mathrm{a}^{(t_1,c)}, f_\mathrm{p}^{(t_2,c)}) - I_{t_2}^{c} \right\|_1 \tag{1}$$

$$\mathcal{L}_\mathrm{pose\text{-}sim} = \left\| \frac{1}{n_1}\sum_{t=1}^{n_1} f_\mathrm{p}^{(t,c_1)} - \frac{1}{n_2}\sum_{t=1}^{n_2} f_\mathrm{p}^{(t,c_2)} \right\|_2^2 \tag{2}$$

さらに，生成した画像を再度エンコーダ E に入力し，Encode する前後での両特徴の差分からエンコーダ E の一貫性を保証する $\mathcal{L}_\mathrm{p\text{-}consis}$，$\mathcal{L}_\mathrm{a\text{-}consis}$ を計算する．ここで，$E(*)_p$ はエンコーダ E の出力のうち姿勢特徴を，$E(*)_a$ は外観特徴を指す．

$$\mathcal{L}_\mathrm{p\text{-}consis} = \|\mathbf{f}_\mathrm{p} - E(G(\mathbf{f}_\mathrm{a}, \mathbf{f}_\mathrm{p}))_\mathrm{p}\|_2^2 \tag{3}$$

$$\mathcal{L}_\mathrm{a\text{-}consis} = \|\mathbf{f}_\mathrm{a} - E(G(\mathbf{f}_\mathrm{a}, \mathbf{f}_\mathrm{p}))_\mathrm{a}\|_2^2 \tag{4}$$

3.2 姿勢特徴と外観特徴による歩行者画像の生成

異なる映像の姿勢特徴と外観特徴を入力し，生成器 G から仮想歩行映像を生成する．生成した各フレームは識別器 D に入力し，次の敵対的損失 $\mathcal{L}_\mathrm{adv}$ を計算する．ただし，敵対的損失の定義には RaLSGAN [6] を用いた．

$$\begin{aligned}\mathcal{L}_\mathrm{adv} = \sum_{i,j\in\{c_1,c_2\}} \big((D(\mathbf{I}^j) - D(G(\mathbf{f}_\mathrm{a}^i, \mathbf{f}_\mathrm{p}^j) - 1)^2 \\ + D(G(\mathbf{f}_\mathrm{a}^i, \mathbf{f}_\mathrm{p}^j) - D(\mathbf{I}^j) - 1)^2\big)\end{aligned} \tag{5}$$

3.3 姿勢特徴列による人物識別

得られた姿勢特徴列を分類器 C に入力し，人物 ID の確率分布を出力する．分類器 C は，3 層の LSTM と 1 層の全結合層から構成される．確率分布から次の交差エントロピーを計算する．

$$\mathcal{L}_\mathrm{id} = \frac{1}{\sum_{t=1}^{n}\omega_t}\sum_{t=1}^{n} -\omega_t \log\left(C(\mathbf{f}_\mathrm{p})\right) \tag{6}$$

学習は次のマルチタスク損失 $\mathcal{L}$ を最小化する．

$$\begin{aligned}\mathcal{L} = \lambda_\mathrm{recon}\mathcal{L}_\mathrm{recon} + \lambda_\mathrm{pose\text{-}sim}\mathcal{L}_\mathrm{pose\text{-}sim} \\ + \lambda_\mathrm{p\text{-}consis}\mathcal{L}_\mathrm{p\text{-}consis} + \lambda_\mathrm{a\text{-}consis}\mathcal{L}_\mathrm{a\text{-}consis} \\ + \lambda_\mathrm{adv}\mathcal{L}_\mathrm{adv} + \lambda_\mathrm{id}\mathcal{L}_\mathrm{id}\end{aligned} \tag{7}$$

3.4 動的歩容特徴による人物認証

分類器 C の全結合層直前から抽出される中間出力を歩行映像を表す動的歩容特徴ベクトル $\mathbf{f}_\mathrm{gait}$ とする．特徴ベクトル空間のコサイン距離に基づいて，クエリ映像とデータベースの照合を行う．

4. 実験

実験では，CASIA-B [4] データセットを用いて，生成画像の定性的評価および識別率の定量的評価を行った．CASIA-B [4] データセットは，NM を基準として，バッグをもった BG，ロングコートを着た CL の三つの画像セットから構成される．学習には，全被験者 124 名のうち前半の 74 名分の歩行映像を用い，GaitNet [2] をベースラインとして，表 1 に示す四つの手法それぞれについて評価をした．ただし，本実験で用いた GaitNet は，Zhang [2] らの論文を元に筆者らが実装したものである．なお，学習の際にはそれぞれの歩行映像から連続する 20 フレームをランダムに切り取ったものを用い，評価の際には各映像の全フレームを用いた．

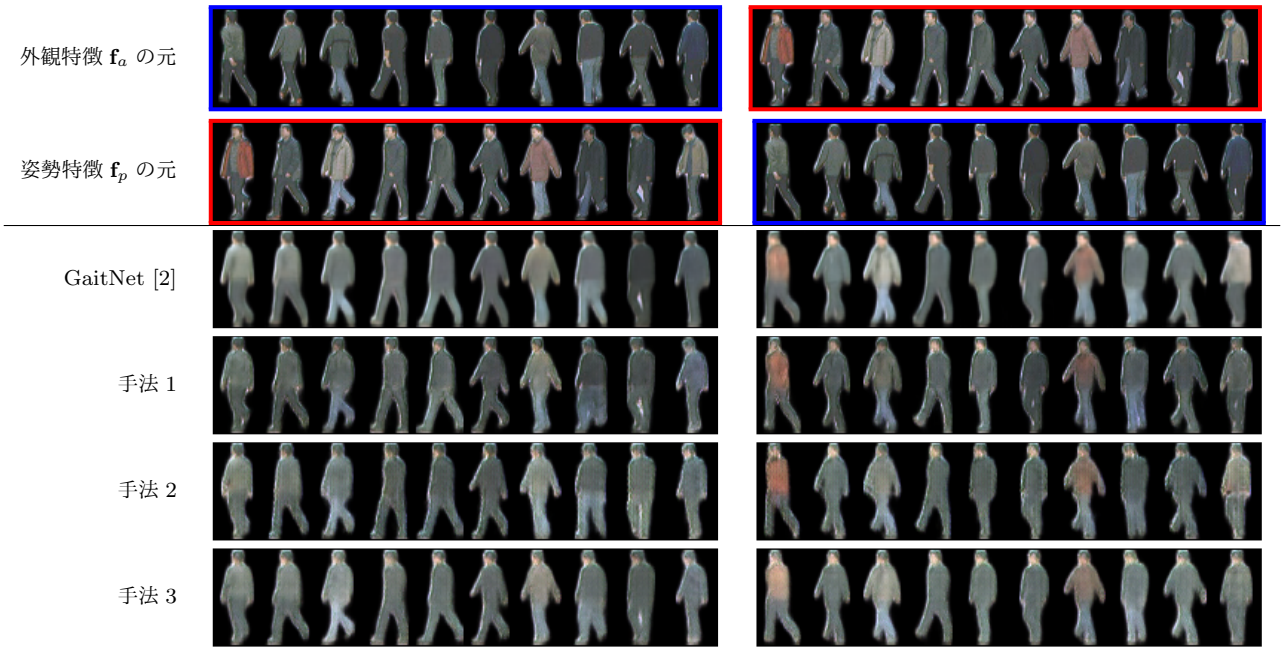

図 2: 各手法における同一人物間の姿勢特徴交換による生成画像

表 1: 比較手法

	敵対的損失 (5)	生成画像の識別損失 (6)	一貫性損失 (3), (4)
GaitNet [2]			
手法 1	✓		
手法 2	✓	✓	
手法 3	✓	✓	✓

4.1 生成画像品質の定性的評価

同一人物に対し，歩行方向や服装が異なる二つの映像から，姿勢特徴および外観特徴を交換して生成した結果を図 2 に示す．1, 2 行目は，それぞれ生成画像における外観特徴，姿勢特徴の元となった画像であり，左右の列で元となる画像が逆になっている．そのため，分離表現学習が適切に進んでいた場合，2 行目の姿勢をしていながら 1 行目の服装をしている画像が生成されると期待される．

図 2 を見ると，GaitNet [2] では姿勢の大局的な特徴は再現できているものの，輪郭がぼやけている上に，脚や腕の先など細部が姿勢特徴の元画像のものと異なることがわかる．それに対し，提案手法は姿勢や輪郭を正確に再現できており，特に手法 3 は提案手法の中でも輪郭の細部まで姿勢を再現できているように見える．また色の細やかさを概観で見ると，GaitNet [2] と手法 3 が同程度であるのに対し，手法 1, 2 の順に色が詳細に表現できていることがわかる．

以上より，提案手法はどれも，ベースライン [2] よりも分離表現学習が正確にできており，特徴を交換して生成した場合にも，両方の特徴の適切な情報を引き継いだ仮想歩行者画像を生成できると考えられる．

4.2 識別精度の定量的評価

識別精度の評価には，学習に用いなかった残りの 50 名の歩行映像を用いた．評価は NM, BG, CL の各セットについて，単一の学習済みモデルを用いて行い，クエリと同じ角度の映像を除外した，NM のデータベースから照合を行った．クエリが NM, BG, CL それぞれの場合における，クエリの撮影角度ごとの識別率をそれぞれ表 2, 3, 4 に示す．

表 2, 3, 4 から，NM と BG においては GaitNet [2] が，CL においては提案手法 3 が最も識別率が高いことがわかる．いずれの設定においても，生成画像を識別学習に用いた手法 2, 3 はベースラインよりも精度が悪化していることから，生成画像を用いた識別学習に問題があると考えられる．

4.3 考察

実験から，提案手法はベースライン [2] よりも，歩行者画像特徴の分離が正確であり，本物に近い品質の仮想歩行者画像を生成できるにもかかわらず，生成した仮想歩行者画像を識別学習に用いた場合に識別精度が下がることが示された．この要因として考えられるのは，まだ生成学習が十分に進んでいない状態で生成された画像を識別する場合，生成画像は本物の歩行者画像とかけ離れたものとなっていることから，識別に悪影響を及ぼしていると考えられる．

5. まとめと今後の予定

本研究では，分離表現学習に基づいて歩容映像から姿勢特徴と服装や持ち物などの外観特徴に分離し，同一人物の異なる歩容映像間で両特徴を交換して新たな画像を

表 2: クエリが NM の場合の撮影角度ごとの識別率

Methods	0°	18°	36°	54°	72°	90°	108°	126°	144°	162°	180°	Average
GaitNet [2]	92.8	**93.9**	**95.2**	**97.6**	**96.7**	**98.3**	97.6	**97.6**	**94.7**	**95.5**	**95.2**	**95.8**
手法 1	**93.9**	92.5	93.8	95.5	94.5	96.5	**97.7**	96.0	94.2	92.5	92.8	94.6
手法 2	89.6	88.9	89.2	94.1	92.9	94.6	94.5	94.6	91.8	90.5	92.5	92.1
手法 3	90.8	89.0	86.4	91.3	91.2	93.3	93.8	94.7	92.5	89.9	91.7	91.3

表 3: クエリが BG の場合の撮影角度ごとの識別率

Methods	0°	18°	36°	54°	72°	90°	108°	126°	144°	162°	180°	Average
GaitNet [2]	91.6	**90.4**	**91.2**	**93.6**	**94.5**	**95.0**	**92.6**	**94.2**	**90.5**	**91.4**	**91.6**	**92.4**
手法 1	**94.2**	87.5	87.1	91.0	87.8	88.3	89.5	89.1	82.3	83.9	88.5	88.0
手法 2	92.0	88.7	87.9	89.8	89.4	87.7	89.4	90.6	82.7	85.5	83.4	87.9
手法 3	88.0	86.7	83.3	87.3	85.6	80.7	85.2	88.0	81.8	82.6	85.5	85.0

表 4: クエリが CL の場合の撮影角度ごとの識別率

Methods	0°	18°	36°	54°	72°	90°	108°	126°	144°	162°	180°	Average
GaitNet [2]	**49.6**	**40.5**	**37.7**	**39.2**	34.0	24.9	27.4	28.4	24.7	23.0	24.2	32.1
手法 1	46.0	39.7	35.5	37.6	**36.2**	**31.9**	**34.9**	**34.1**	**26.0**	**30.0**	**29.5**	**34.7**
手法 2	36.6	30.9	27.2	22.0	20.2	16.9	16.3	20.5	12.3	16.6	14.5	21.2
手法 3	39.5	35.1	28.5	27.3	25.0	20.5	19.8	20.5	17.9	17.4	19.2	24.6

生成するデータ拡張を用いた歩容認証手法を提案した．CASIA-B [4] を用いた実験では，生成画像品質の定性的評価により，提案手法が正確に分離表現を獲得できること，および元となる画像の特徴が生成画像に適切に現れることを確認した．識別実験では，提案手法の精度向上は一部設定でしか確認されなかった．考察では，生成画像の最終的な品質は高かったことから，本手法によるデータ拡張が識別率に悪影響を及ぼした原因として，生成画像の品質が低いうちから識別に用いたことによるものである可能性を示した．

今後は，生成学習のみで事前学習を行い，生成画像の品質を十分に高めてから識別学習と同時に最適化することで，生成画像を用いたデータ拡張による識別精度向上を図る．また，CASIA-B 以外の歩容認証用データセットについても同様の実験を行い，広く提案手法の有効性を検証する．

謝辞

本研究は JSPS 科研費 JP20H00230 の助成を受けた．

参考文献

[1] Ju Han and Bir Bhanu. “Individual Recognition Using Gait Energy Image.” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(2):316–322, 2005.

[2] Ziyuan Zhang, Luan Tran, Xi Yin, Yousef Atoum, Xiaoming Liu, Jian Wan, and Nanxin Wang. “Gait Recognition via Disentangled Representation Learning.” In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4710–4719, 2019.

[3] Zhedong Zheng, Xiaodong Yang, Zhiding Yu, Liang Zheng, Yi Yang, and Jan Kautz. “Joint Discriminative and Generative Learning for Person Re-identification.” In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2138–2147, 2019.

[4] Martin Hofmann, Jürgen Geiger, Sebastian Bachmann, Björn Schuller, and Gerhard Rigoll. “The TUM Gait from Audio, Image and Depth (GAID) database: Multimodal recognition of subjects and traits.” *Journal of Visual Communication and Image Representation*, 25(1):195–206, 2014.

[5] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. “Mask R-CNN.” In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2961–2969, 2017.

[6] Jolicoeur-Martineau, Alexia. “The relativistic discriminator: a key element missing from standard GAN.” In *Proceedings of the International Conference on Learning Representations*, 2019.

[7] Kohei Shiraga, Yasushi Makihara, Daigo Muramatsu, Tomio Echigo, and Yasushi Yagi. “GEINet: View-invariant gait recognition using a convolutional neural network.” In *International Conference on Biometrics*, 2016.

[8] Yang Feng, Yuncheng Li, and Jiebo Luo. “Learning effective gait features using LSTM.” In *Proceedings of the IEEE International Conference on Pattern Recognition*, pp. 325–330, 2016.