

RGB 情報を活用したデータ拡張に基づく歩容認証

吉野 弘毅^{1,a)} 中嶋 一斗^{1,b)} アン・ジョンホ^{1,c)} 岩下 友美^{2,d)} 倉爪 亮^{1,e)}

概要

個人の歩行パターンと体型に基づいた生体認証である歩容認証は、衣服、荷物などの共変量が識別に悪影響を与える。多くの既存研究では、RGB 画像から前処理によって抽出したシルエットや骨格情報から歩容特徴を抽出しているが、分離表現学習 (DRL) では前処理に依存せずに共変量を除去できる。しかし、DRL は様々な属性を持つ複数の入力間の差異を使って特徴分離を学習するため、その分離性能は訓練データの種類と量に大きく依存する。そこで本研究では、学習データの量と多様性を向上させることを目的として、DRL によって分離した特徴を、他の動画の特徴と交換することによって、データを拡張する手法を提案する。実験では、提案手法が識別精度を向上させることを確認し、他のモダリティとの比較によって RGB 情報を活用した歩容認証の可能性を示す。

1. はじめに

歩容とは歩き方や体型のことであり、個人を特定する手がかりになりうる。歩容は遠距離から取得できて、故意に偽装し続けるのは一般に極めて困難であるため、非協力的な対象の特定に特に有用な生体情報である。偽装困難性と遠距離からの取得可能性を活用して、犯罪捜査 [9], [10] や入退場管理などにおけるシームレスな個人認証への応用が期待されている。

他の多くの画像認識タスクと同じく、歩容認証では深層学習ベースの手法が優位的となっている [17]。歩容映像には歩容の情報以外にも、服装・背景・持ち物のような歩容に依存しない情報 (共変量) が多く存在し、共変量の除去が歩容認証にとって重要である。共変量に対するアプローチとして、前処理による除去 [1], [19], [22] および end-to-end な除去 [13], [20], [27] の大きく分けて 2 種類の方法が提案されてきた。前処理では識別に有用な情報の欠

落 [27] や推定精度の問題 [13] が生じてしまうため、本研究では end-to-end な共変量の除去を採用している。

end-to-end な共変量除去手法は、RGB 画像シーケンスを入力として共変量の除去と歩容に基づく識別を同時に最適化する。ほとんどの end-to-end な共変量除去手法 [3], [11], [12], [13], [20] では、入力された RGB 画像ではなく、ネットワークの途中で抽出されたシルエットや骨格情報などの中間モダリティから歩容特徴の抽出を行っている。結果としてこれらの手法では、RGB 画像のもつ豊富な情報を十分に活用できていない [18]。一方 Zhang ら [27] は、分離表現学習 (DRL) を歩容認証にいち早く適用し、RGB 歩容画像を外観と姿勢特徴に分離するオートエンコーダベースのネットワーク GaitNet を提案している。GaitNet における DRL では、入力された同一人物の歩容映像で服装・撮影角度の異なる二つの映像の 2 種類の属性の差分から、共変量である外観の特徴の分離・除去を、対応する教師ラベルなしで可能にしている、つまり分離性能には、学習データの多様性が非常に重要である。

そこで DRL における分離性能を向上させるために、先行研究において我々は人物 ID を考慮したデータ拡張を用いたアプローチを提案している [24]。我々が提案したネットワークでは、DRL によって分離した特徴を同一人物の歩容映像間で交換して仮想歩容映像を新規生成することで、学習するデータのバリエーションを増加させている。この手法では、一人当たりの歩容動画数に比例した数だけ訓練データを増やすことができ、実験によって識別精度を向上させることが示されている。しかし予備実験の結果、特徴交換の対象を同一人物の異なる動画から、他の人物の動画に拡張すると、画像生成が複雑になって識別精度の向上が限定的になることがわかった。

本論文では、RGB の歩容映像から DRL によって歩容依存する/しない特徴を分離して、異なる人物間で分離された特徴を交換することによってデータ拡張を行う、新たな歩容認証手法を提案する。具体的には、学習では、まず入力された RGB の歩容映像は、歩容に依存する特徴 (姿勢特徴) および外観に関する共変量の特徴 (スタイル特徴) に、フレームごとに分離される。異なる人物の歩容映像についても同様に特徴分離を行い、これらの二つの歩容映像間で

¹ 九州大学

² NASA / Caltech ジェット推進研究所

a) yoshino@irvs.ait.kyushu-u.ac.jp

b) k_nakashima@ait.kyushu-u.ac.jp

c) ahn@irvs.ait.kyushu-u.ac.jp

d) yumi.iwashita@jpl.nasa.gov

e) kurazume@ait.kyushu-u.ac.jp

姿勢特徴を交換することで歩行者画像を新規生成し、学習データに追加する。提案手法は、データセット中の全歩容動画の 2 乗に訓練データの数を拡張することができる。識別は時系列の姿勢特徴群から抽出される歩容特徴に基づいて行われ、識別学習と生成学習は、両方同時に最適化される。推論時にはデータ拡張をせずに、RGB 歩容映像から分離された姿勢特徴からそのまま動的な歩容特徴を抽出して、データベース中のデータとの照合を行う。実験では、本研究同様に RGB から直接的に歩容特徴を抽出する RGB ベースの手法に加えて、シルエットや人体モデルなどの他のモダリティを媒介とした state-of-the-art 手法との比較も行う。

2. 手法

本研究におけるパイプラインの詳細な説明は、先行研究 [24] を参照されたい。ここでは、先行研究からの変更点である、生成画像の再エンコード後の元画像の再構成について述べる。

2.1 元画像の循環再構成

先行研究 [24] では、エンコーダ E の一貫性を保証するために再度エンコードする前後での両特徴の差分を小さくする、一貫性損失を設計した。異なる人物間の交換生成を行う本手法では、再エンコードした特徴をもとに元画像を再構成するステップを追加する。異なる二つの歩容画像 I^A, I^B に対して、エンコード、特徴交換生成、再エンコードの順で処理することで、 I^A の姿勢特徴 $E_{\text{pose}}(G(\mathbf{f}_{\text{style}}^B, \mathbf{f}_{\text{pose}}^A))$ 、 I^A のスタイル特徴 $E_{\text{style}}(G(\mathbf{f}_{\text{style}}^A, \mathbf{f}_{\text{pose}}^B))$ が得られる。これら二つの特徴をもとに、次の循環再構成損失を定義する。

$$\mathcal{L}_{\text{cycle}} = \sum_{\substack{a,b \in \{A,B\} \\ a \neq b}} \|G(E_{\text{style}}(G(\mathbf{f}_{\text{style}}^a, \mathbf{f}_{\text{pose}}^b)), E_{\text{pose}}(G(\mathbf{f}_{\text{style}}^b, \mathbf{f}_{\text{pose}}^a))) - I^a\|_1 \quad (1)$$

この損失は、Zhu ら [29] が報告するように教師データのない画像生成の高品質化への貢献が期待され、提案手法における特徴の分離性能の向上にも寄与すると考えられる。本手法では学習の安定のために、学習の他の損失と異なって、分離性能を補助的に向上させるだけの役割である一貫性損失を、循環再構成損失によって置換する。

3. 実験

3.1 データセット

実験に用いるデータセットには、公平で包括的な比較をするために、最も広く使われている CASIA-B [25] を使う。CASIA-B は、総被験者数は 124 名から構成されたデータセットである。歩行条件のバリエーションとして、次の 3

種類が存在する：基準となる歩行設定 (NM)、バッグを持っている設定 (BG)、基準とは異なる服装をしている設定 (CL)。被験者ごとに、0~180 度までの 18 度刻みの 11 種類の角度から撮影された、6 本の NM 映像、2 本ずつの BG と CL の映像の、合計 110 本 (= 11 × 10) の歩容映像が存在する。つまり、データセットには、合計で 113,640 (= 124 × 11 × 10) 本の動画が含まれている。訓練にはデータセット中の前半の 74 人が用いられる。評価については、残りの被験者は評価に用いられ、一部の動画をギャラリー (データベース) として、残りの動画をプローブ (クエリ) として使用する。

3.2 比較手法

本実験では、本手法と同じく RGB 画像から直接歩容特徴を抽出する手法に加えて、他のモダリティを活用した手法とも比較を行う。モダリティとしては、本研究の入力モダリティである RGB の他に、多くの研究において採用されているシルエットとスケルトンをとりあげる。

まずシルエットをネットワークへの入力手法について紹介する LB [23] は、二つの GEI [8] を入力とするネットワークである。Chao ら [1] は、入力フレームの順番に依存しない集合としてのシルエットシーケンスを入力とする GaitSet を提案することで、シルエット画像列の時間平均である GEI [8] が主流であったシルエットベース手法の新しい可能性を示した。GaitBase [5] は、それまでのシルエットベース SOTA 手法 [1], [2], [6], [14] を再検証することで開発された、屋内外によらず頑健なベースラインモデルである。DANet [15] は、隣接ピクセルの特徴間の関係を動的にモデル化したフレーム単位の特徴と、自己注意機構による適応的な時系列集約を活用している。

次に、姿勢推定の精度向上につれて、スケルトンベースのニューラルネットワークも優勢になってきている。Teepe ら [21], [22] は歩容認証で初めてグラフ畳み込みネットワークに着目して、3 次元上の関節の位置を入力とする GaitGraph [22] を提案し、さらに入力として歩行速度と骨の情報を追加して 3 ブランチネットワークの GaitGraph2 [21] へと発展させている。GaitMixer [16] は、空間的な自己注意と時間的な畳み込みを組み合わせたアーキテクチャを採用している。GPGait [7] は、スケールや位置などの歩容映像間の差異に頑健な、人間中心の姿勢表現変換とグラフ畳み込みネットワークからなるフレームワークである。

これらの手法とは対照的に RGB モダリティを特徴抽出の素として活用しているのは、我々の知る限りでは、本提案手法と我々の前の手法 [24] に加えて、GaitNet [26], [27] のみである。GaitNet は、LSTM 付きのオートエンコーダであり、RGB 画像を外観特徴と姿勢特徴に分離している。加えて、Zhang ら [26] は、外観特徴に含まれてしまっていた歩容に依存する体型の情報を、canonical 特徴として抽

表 1 モダリティ別の Rank-1 認証率の比較. 全ての手法の中で最も大きい値を太字に, 同一モダリティ内で最も大きい値に下線を付けている.

手法	モダリティ	平均識別精度 ↑ [%]		
		NM	BG	CL
LB [23]	シルエット	89.9	72.4	54.0
GaitSet [1]		95.0	87.2	70.4
GaitBase [5]		97.6	94.0	77.4
DANet [15]		98.0	95.9	89.9
GaitGraph [22]	スケルトン	87.7	74.8	66.3
GaitGraph2 [21]		82.0	73.2	63.6
GaitMixer [16]		<u>94.9</u>	<u>85.6</u>	<u>84.5</u>
GPGait [7]		93.6	80.2	69.3
GaitNet [27]	RGB	91.6	85.7	58.9
GaitNet [26]		92.3	88.9	<u>62.3</u>
ベースライン		90.2	86.1	24.3
先行研究 [24]		94.4	88.8	29.2
提案手法		<u>95.9</u>	<u>91.1</u>	30.0

出すことによって, GaitNet を発展させている. 最後に, Zhang ら [27] の手法にいくつかの改良点を加えた自分で再実装したものをベースラインとして, 先行研究 [24] と一緒に結果を載せている.

3.3 識別制度の比較

識別精度の評価に用いるのは, CASIA-B [25] で評価指標として一般に採用される, rank-1 認証率である. rank-1 認証率とは, ギャラリー中にあるプローブに最も距離が近いデータの ID が, 正解 ID と一致する割合を指す. 全設定 (NM, BG, CL) の三種類のプローブに対して, ギャラリーとの最近傍探索を行い, ギャラリーには, 評価対象となる probe の撮影角度を除く全角度の NM の歩容映像を用いる.

表 1 を見ると, 提案手法は全ての設定においてベースラインを上回る精度を記録している. また提案手法は, NM と BG の二つの設定において, 他の RGB ベースの手法およびスケルトンベースの手法の精度を上回っており, シルエットベース手法とも比肩する精度を達成している. 一方, CL の設定, つまりギャラリーとプローブ間で服装が異なる場合には, 提案手法の精度は他の手法に劣っている. これは, 各被験者ごとに服装のバリエーションに大きな偏りがあるからだと考えられる. NM と BG の服装は同じであり, CL のみ違う服装をしているが, 被験者ごとに NM の動画は 6 本で BG と CL は 2 本ずつの動画がある. つまり, 各被験者の動画において, 全体の 80% が同じ服装であり, 他の服装のバリエーションは 20% のみであり, 合計 2 種類の服装しか存在しない. 提案手法は, 設定の差分から特徴分離を学習しているため, この服装の多様性の偏りが特徴分離に影響を与えたと考えられる. 提案手法と同じく

設定の差異から特徴分離を学習する手法 [27], [28] も同様に, CL は他の設定よりも精度が低下している. 服装のバリエーションの問題は, 服装情報の分離が必要な RGB ベース手法固有の問題であるが, 服装の多様性に富んだデータセットが広まれば, 提案手法はより良い性能を発揮できると期待できる. また提案手法は, 本来の学習データ数の二乗のデータを学習できるが, 今回ベースラインとして採用した GaitNet は 4 層の浅いネットワークである. Vision Transformer [4] のようなもっと表現力の高いモデルをバックボーンとして採用すれば, 提案手法はデータ拡張の恩恵をより強く受けられると考えられる.

4. まとめ

本論文では, 我々は RGB 歩容画像から精度よく歩容特徴を抽出するために, 歩容特徴を交換することで学習データを拡張する手法を提案する. 提案手法は, 分離表現学習を用いて分離した歩容特徴を異なる人物の歩容特徴と交換することによって, 学習データの数と種類をオンラインに増やすことができる. 実験によって, 提案手法は高品質な歩容画像を生成可能で荒れること, および生成画像を用いたデータ拡張が識別精度を向上させることを実証した. また, 提案手法はほとんどの設定において RGB から直接特徴を抽出する手法の state-of-the-art を上回り, 他のモダリティを活用した手法とも比肩する精度を達成している. しかし, 学習データ中の服装の大きな偏りから, データベースとクエリ間で服装が異なる設定においては, 課題が残ることがわかった. データセットにおける服装の不均衡の問題への取り組みが増えることによって, RGB 画像の持つ豊富な情報をもっと活用されていくことを期待する.

謝辞

本研究の一部は, JST 科学技術イノベーション創出に向けた大学フェローシップ創設事業 JPMJFS2132 および JSPS 科研費 JP20H00230 の助成を受けたものである.

参考文献

- [1] Chao, H., He, Y., Zhang, J. and Feng, J.: Gaitset: Regarding gait as a set for cross-view gait recognition, *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 8126–8133 (2019).
- [2] Chao, H., Wang, K., He, Y., Zhang, J. and Feng, J.: GaitSet: Cross-view Gait Recognition through Utilizing Gait as a Deep Set, *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, Vol. 44, No. 7, pp. 3467–3478 (2022).
- [3] Cui, Y. and Kang, Y.: Multi-Modal Gait Recognition via Effective Spatial-Temporal Feature Fusion, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 17949–17957 (2023).
- [4] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J. and Houlsby,

- N.: An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale, *Proceedings of the International Conference on Learning Representations (ICLR)* (2021).
- [5] Fan, C., Liang, J., Shen, C., Hou, S., Huang, Y. and Yu, S.: OpenGait: Revisiting Gait Recognition Towards Better Practicality, *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9707–9716 (2023).
- [6] Fan, C., Peng, Y., Cao, C., Liu, X., Hou, S., Chi, J., Huang, Y., Li, Q. and He, Z.: Gaitpart: Temporal part-based model for gait recognition, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 14213–14221 (2020).
- [7] Fu, Y., Meng, S., Hou, S., Hu, X. and Huang, Y.: Gpgait: Generalized pose-based gait recognition, *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 19595–19604 (2023).
- [8] Han, J. and Bhanu, B.: Individual recognition using gait energy image, *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, Vol. 28, No. 2, pp. 316–322 (2005).
- [9] Iwama, H., Muramatsu, D., Makihara, Y. and Yagi, Y.: Gait verification system for criminal investigation, *Information and Media Technologies (IMT)*, Vol. 8, No. 4, pp. 1187–1199 (2013).
- [10] Larsen, P. K., Simonsen, E. B. and Lynnerup, N.: Gait analysis in forensic medicine, *Journal of forensic sciences*, Vol. 53, No. 5, pp. 1149–1153 (2008).
- [11] Li, X., Makihara, Y., Xu, C. and Yagi, Y.: End-to-end model-based gait recognition using synchronized multi-view pose constraint, *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 4106–4115 (2021).
- [12] Li, X., Makihara, Y., Xu, C., Yagi, Y., Yu, S. and Ren, M.: End-to-end model-based gait recognition, *Proceedings of the Asian conference on computer vision*, pp. 3–20 (2020).
- [13] Liang, J., Fan, C., Hou, S., Shen, C., Huang, Y. and Yu, S.: GaitEdge: Beyond Plain End-to-End Gait Recognition for Better Practicality, *European Conference on Computer Vision (ECCV)*, Springer, pp. 375–390 (2022).
- [14] Lin, B., Zhang, S. and Yu, X.: Gait Recognition via Effective Global-Local Feature Representation and Local Temporal Aggregation, *IEEE/CVF International Conference on Computer Vision (ICCV)*, IEEE, pp. 14628–14636 (2021).
- [15] Ma, K., Fu, Y., Zheng, D., Cao, C., Hu, X. and Huang, Y.: Dynamic Aggregated Network for Gait Recognition, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 22076–22085 (2023).
- [16] Pinyoanuntapong, E., Ali, A., Wang, P., Lee, M. and Chen, C.: Gaitmixer: Skeleton-Based Gait Representation Learning Via Wide-Spectrum Multi-Axial Mixer, *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, pp. 1–5 (2023).
- [17] Sepas-Moghaddam, A. and Etemad, A.: Deep gait recognition: A survey, *IEEE transactions on pattern analysis and machine intelligence*, Vol. 45, No. 1, pp. 264–284 (2022).
- [18] Shen, C., Yu, S., Wang, J., Huang, G. Q. and Wang, L.: A comprehensive survey on deep gait recognition: algorithms, datasets and challenges, *arXiv preprint arXiv:2206.13732* (2022).
- [19] Shiraga, K., Makihara, Y., Muramatsu, D., Echigo, T. and Yagi, Y.: Geinet: View-invariant gait recognition using a convolutional neural network, *Proceedings of the IEEE International Conference on Biometrics (ICB)*, pp. 1–8 (2016).
- [20] Song, C., Huang, Y., Huang, Y., Jia, N. and Wang, L.: Gaitnet: An end-to-end network for gait based human identification, *Pattern recognition*, Vol. 96, p. 106988 (2019).
- [21] Teepe, T., Gilg, J., Herzog, F., Hörmann, S. and Rigoll, G.: Towards a Deeper Understanding of Skeleton-based Gait Recognition, *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1569–1577 (2022).
- [22] Teepe, T., Khan, A., Gilg, J., Herzog, F., Hörmann, S. and Rigoll, G.: Gaitgraph: Graph convolutional network for skeleton-based gait recognition, *IEEE International Conference on Image Processing (ICIP)*, pp. 2314–2318 (2021).
- [23] Wu, Z., Huang, Y., Wang, L., Wang, X. and Tan, T.: A comprehensive study on cross-view gait based human identification with deep cnns, *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, Vol. 39, No. 2, pp. 209–226 (2017).
- [24] Yoshino, K., Nakashima, K., Ahn, J., Iwashita, Y. and Kurazume, R.: Gait Recognition using Identity-Aware Adversarial Data Augmentation, *2022 IEEE/SICE International Symposium on System Integration (SII)*, IEEE, pp. 596–601 (2022).
- [25] Yu, S., Tan, D. and Tan, T.: A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition, *Proceedings of the IEEE International Conference on Pattern Recognition (ICPR)*, Vol. 4, pp. 441–444 (2006).
- [26] Zhang, Z., Tran, L., Liu, F. and Liu, X.: On Learning Disentangled Representations for Gait Recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 44, No. 1, pp. 345–360 (2022).
- [27] Zhang, Z., Tran, L., Yin, X., Atoum, Y., Liu, X., Wan, J. and Wang, N.: Gait recognition via disentangled representation learning, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4710–4719 (2019).
- [28] Zheng, J., Liu, X., Liu, W., He, L., Yan, C. and Mei, T.: Gait Recognition in the Wild with Dense 3D Representations and A Benchmark, *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 20228–20237 (2022).
- [29] Zhu, J.-Y., Park, T., Isola, P. and Efros, A. A.: Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks, *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 2223–2232 (2017).