

物体追跡と深層学習による歩容認証を用いた動画像データに基づく放牧牛の自動個体識別

Automatic Individual Identification of Grazing Cattle from Video Data

Using Object Tracking and Deep Gait Recognition

坂本 直暉[†] 大山 憲二[†] 大川 剛直[†]
Naoki Sakamoto Oyama Kenji Ohkawa Takenao

1. はじめに

近年、畜産分野では従事者の高齢化や減少が進み、情報技術の導入による飼養管理の自動化・効率化が重要視されている。動画像解析に基づいて自動管理を行うためには、カメラ映像上の家畜の個体を識別する必要がある。

個体識別には、外観特徴が用いられることが一般的である。しかし、家畜の中には色や斑紋の個体差が乏しく、外観ベースの識別が困難な種も存在する。本研究で扱う黒毛和種の牛は、その例である。そこで、本研究では歩行動作に表れる個性性を利用する歩容認証に着目する。歩容認証は、対象の歩行パターンに基づいて個体を認証または識別する生体認証手法であり、近年では、対象のシルエットを利用した深層学習に基づく手法が提案され、その有効性が報告されている[1]。しかしながら、これらの手法は一般に、大量の歩行映像を必要とする。一方で、家畜は人間と異なり非協力的であるため、十分な規模のデータセットを構築することが難しいという問題がある。そこで、本研究では、小規模データセットに対しても高い性能を発揮するとされる歩容認証手法 GaitSet[1]に着目し、その有効性を黒毛和種放牧牛に対して検証する。

歩容認証には、対象が歩行している画像系列を入力として用いる必要がある。しかし、原映像には、対象が映っていないフレームや、非歩行状態のフレームが含まれる場合があり、歩行画像系列を適切に抽出する必要がある。手作業による抽出は、多大な労力を要し、実運用を考慮すると、この前処理の自動化は不可欠である。そこで、本研究では、YOLOv5[2]を検出器とした物体追跡手法 DeepSORT[3]を利用し、検出矩形の移動量に基づいて原映像から対象の歩行画像系列を自動で抽出する手法を提案する。

本研究では、黒毛和種放牧牛を対象に、YOLOv5 を検出器とした物体追跡手法 DeepSORT による歩行画像系列抽出、Segment Anything Model (SAM) [4]によるシルエット抽出、及び、歩容認証手法 GaitSet を組み合わせた自動個体識別の枠組みを提案する。

2. 関連研究

2.1 物体追跡

物体追跡とは、映像上の特定の物体を追跡する技術である。DeepSORT は、運動情報と外観特徴を利用して、直前フレームまでの各物体の追跡情報と現在のフレームにおける検出結果を対応付ける手法である。具体的には、カルマンフィルタにより予測された物体位置と、実際の検出矩形とのマハラノビス距離に基づく位置的類似度、及び、畳み込みニューラルネットワーク (CNN) により抽出された外観特徴のコサイン類似度を組み合わせ、ハンガリアンアルゴリズムにより各検出に対して追跡 ID を割り当てること

で追跡情報を更新する。

2.2 歩容認証

歩容認証とは、歩き方の個性性に基づいて個体を認証または識別する技術である。歩容認証の基本的な流れを次に示す。(1) 歩行画像系列の取得：対象の歩行画像系列を取得する。(2) シルエットの抽出：各画像から対象領域を白、背景を黒とする二値化画像を得る。(3) シルエットの位置合わせ：歩行位置の違いにより、シルエットの大きさや位置は変化しうするため、位置合わせによる正規化を行う。(4) 歩容特徴の抽出：正規化されたシルエット画像系列に対して、時間的または空間的に解析を行い、歩容特徴を抽出する。(5) 識別：抽出された歩容特徴 (プロンプト) を、事前に登録された各個体の歩容特徴 (ギャラリー) と照合し、類似度が最大となるギャラリーに対応する個体 ID を識別結果として出力する。

GaitSet は、深層学習に基づく歩容認証手法の一つであり、入力された歩行画像系列を順序に依存しない集合として扱うアプローチを採用している。各シルエット画像は CNN に入力され、抽出された特徴は最大値を取得する関数等の置換不変関数によって集約されることで、画像集合全体を代表する特徴が得られる。得られた特徴マップは、水平方向に複数スケールに分割され、各領域に全体平均プーリング、及び、全体最大プーリングが適用されることで、マルチスケールな部位レベルの特徴が抽出される。これらの特徴ベクトルは、対応する領域毎に独立した全結合層により変換され、識別に用いる最終的な特徴表現として出力される。また、学習時には、ミニバッチ形成毎に、歩行画像系列からランダムサンプリングされた画像集合を入力とすることで、実質的なデータ拡張が可能となり、小規模データセットに対しても、高い性能を示すことが報告されている。

3. 提案手法

提案手法の概要を図 1 に示す。

3.1 歩行画像系列の抽出

まず、原映像に対し、YOLOv5 を検出器とした物体追跡手法 DeepSORT を用いて牛を追跡する。各追跡 ID に対して、直前フレームとの検出矩形の中心座標の差分を移動量として算出し、対象フレームの前後 N フレームの移動量を合計することで累積移動量を求める。スケール変化を補正するため、移動量は検出矩形の高さで正規化した値を用いる。また、検出矩形の揺れ等によるノイズの影響を低減するため、中心座標、及び、高さには、移動平均フィルタを適用し、平滑化された値を用いる。累積移動量が閾値 $d_{\text{threshold}}$ を超えた場合、当該フレームを歩行時と判定する。歩行時と判定されたフレームが連続して M 枚以上存在する場合、これらのフレームを歩行画像系列として抽出する。

[†] 神戸大学 Kobe University

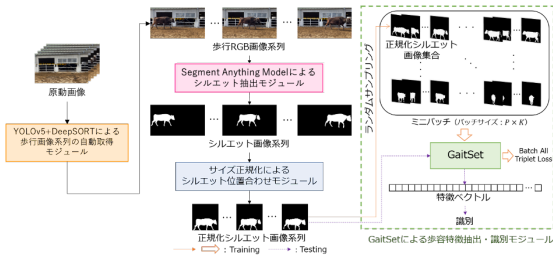


図 1：提案手法の概要

3.2 正規化シルエットの生成と歩容特徴に基づく個体識別

まず、抽出された歩行画像系列から Segment Anything Model (SAM) を用いてシルエット画像系列を生成する。SAM は、画像上の矩形のプロンプトに基づいて、対象物体のマスクを推定可能なセグメンテーションモデルである。前節で得られた歩行画像系列の各画像に対し、検出された矩形領域をプロンプトとして SAM に与えることで、牛領域のマスクを推定し、シルエット画像を生成する。次に、得られたシルエット画像に対して、サイズ正規化に基づくシルエットの位置合わせ処理を行う。サイズ正規化とは、シルエットのアスペクト比を保持したまま、その高さを固定値にリサイズし、シルエットの水平方向中心が画像の中央に一致するように位置を補正する処理である。最後に、正規化されたシルエット画像系列を GaitSet に入力し、歩容特徴の抽出、及び、識別を行う。

4. 実験

4.1 使用するデータ

2023 年 11 月 13 日午前 9 時から正午にかけて神戸大学大学院農学研究科附属食資源教育研究センター内のパドックにおいて、黒毛和種放牧牛 21 頭を対象に撮影を実施した。各個体を 1 頭ずつ放牧し、1 頭あたり約 5 分間、4 台のカメラを用いて撮影した。同時に、耳標から取得した個体番号を各映像に対応付けた。

4.2 実験設定

本研究では、提案した歩行画像系列の自動抽出手法、及び、GaitSet の有効性を評価するため、複数の観点から実験を行った。まず、歩行画像系列を手動で選定したデータセットと、提案手法により自動抽出されたデータセットを用いて、GaitSet をそれぞれ学習させ、識別精度を比較した。また、歩行方向の変化に対する頑健性を評価する目的で、カメラ視点に基づき歩行方向を 8 つに分類し、照合時にプローブと同一方向のギャリリを含む場合と含まない場合の双方で評価した。ただし、提案手法では、牛がカメラの視線方向に沿って前進または後退するような経路をとる場合、検出矩形の移動量が小さくなり、歩行時と判定されないため、当該方向の歩行画像系列は除外した。加えて、GaitSet の有効性を評価するため、深層学習に依らない代表的な歩容特徴表現である Gait Energy Image (GEI) [5] との比較も行った。GEI は、1 歩行周期に対応する歩行画像系列を時間方向に平均化した表現である。さらに、学習時の画像集合生成方法が識別性能に与える影響についても検証した。具体的には、歩行画像系列からランダムに M 枚選択する方法 (random)、先頭から順に M 枚を選択する方法 (ordered)、ランダムな開始位置から連続して M 枚を選

択する方法 (offset) を比較し、データ拡張の効果を評価した。各データセットについて、12 頭分を学習データ、7 頭分をテストデータとして分割した。評価には、手動データセットのテストデータを使用し、各個体における各歩行方向の 2 つの歩行画像系列をプローブ、残りをギャリリとした。原映像のフレームレートが 30 fps の場合、 N は 14、 M は 40、60 fps の場合、 N は 29、 M は 80 とした。累積移動量の閾値 $d_{\text{threshold}}$ は、各カメラ映像に対して、0.30 または 0.33 に設定した。評価指標には、rank-k accuracy (以下、rank-k) を用いた。rank-k とは、照合結果の上位 k 件の候補に正解の個体が含まれる割合を表す指標である。本研究では、rank-1、rank-5 を採用して評価を行った。

4.3 実験結果と考察

表 1 に示す実験結果より、自動データセットで学習されたモデルは、rank-1、rank-5 のいずれにおいても、手動データセットで学習されたモデルを上回る識別精度を示した。この結果は、提案した歩行画像系列の自動抽出手法の有効性を示すと同時に、データ量の差が影響したと考えられる。手動データセットでは、データの質を担保するため、牛が直線的かつ滑らかに歩行し、常に全身がフレーム内に収まっている画像系列のみを選定した結果、歩行画像系列の総数は 823 にとどまった。一方で、自動データセットでは、より多くの歩行系列が抽出され、総数は 2303 に達している。これは、データの質に過度に依存せず、量を確保することで、識別性能が改善する可能性を示唆する。また、プローブと同一方向のギャリリを含まない条件下において、GEI の rank-1 は 17.6% とほぼ偶然と同程度の精度となった一方で、GaitSet は 60% 程度となり、歩行方向の変化に対する頑健性が確認された。さらに、画像集合生成方法の違いについては、ordered、offset、random の順で識別精度が高くなった。これは、学習時に多様性のある入力インスタンスを生成することが、識別性能の向上に寄与することを示す。以上より、提案手法は、黒毛和種放牧牛に対する自動個体識別に有効なアプローチであることが示された。

表 1：識別性能の比較

手法	歩行画像系列抽出方法	画像集合生成方法	同一方向を含むか	rank-1 [%]	rank-5 [%]
GaitSet	手動	random	○	62.6	80.2
			×	51.9	78.4
	自動	random	○	70.4	88.0
			×	60.1	79.8
		ordered	○	54.9	83.8
			×	43.6	73.3
		offset	○	62.5	85.9
			×	62.2	83.2
GEI	-	-	○	38.7	72.3
			×	17.6	60.8

謝辞

本研究の一部は JSPS 科研費 21H04914 の助成による。

参考文献

- [1] H. Chao, Y. He, J. Zhang, and J. Feng, "GaitSet: Regarding gait as a set for cross-view gait recognition," *Proceedings of the AAAI conference on artificial intelligence*, Vol. 33, No. 1, pp. 8126–8133 (2019)
- [2] Ultralytics, "YOLOv5: A State-of-the-Art Real-Time Object Detection System", <https://docs.ultralytics.com> (2021)
- [3] N. Wojke, A. Bewley, and D. Paulus, "Simple Online and Real-time Tracking with Deep Association Metric.", *2017 IEEE international conference on image processing (ICIP)*, pp. 3645–3649 (2017)
- [4] A. Kirillov, et al., "Segment Anything", *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 4015–4026 (2023)
- [5] J. Han, B. Bhanu, "Individual recognition using gait energy image," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28, No. 2, pp. 316–322 (2005)