

令和7年度修士論文

耳標と牛顔画像の併用による牛個体識別

宮崎大学大学院工学研究科先端情報コース

森茂蒼士

指導教員 椋木雅之教授

概要

近年、畜産分野では労働力不足や作業負担の軽減を背景として、牛の個体管理を自動化する技術への関心が高まっている。特に、非接触で情報を取得可能なカメラを用いた個体識別技術は、実環境への適用性が高い手法として注目されている。

牛の個体識別において耳標は法令に基づき付与される重要な識別手段であるが、汚れや遮蔽、撮影条件の影響により読み取りが不安定になるという課題がある。一方、牛顔画像に基づく個体識別は比較的安定して情報を取得できるものの、外観変動による識別精度の揺らぎが課題となる。

そこで本研究では、耳標と牛顔画像を補完的に利用することで、実環境における牛個体識別の信頼性向上を目的とした手法を提案する。提案手法では、耳標情報を主たる識別手掛かりとして用い、レーベンシュタイン距離に基づく候補個体の抽出を行う。その上で、候補が複数存在する場合や識別の信頼性が低い場合に、ArcFaceにより抽出した牛顔特徴量を用いて最終判定を行う二段階の識別構造を採用する。

さらに、複数頭の牛が同時に存在する状況に対応するため、YOLO と ByteTrack を用いたトラッキングにより時間方向の情報を統合し、耳標情報および顔特徴量をトラッキング単位で集約する手法を導入する。

実験結果より、耳標単独では識別が困難な条件においても、顔特徴量を併用することで識別精度が向上することを確認した。また、トラッキングを利用することで、複数頭環境においても安定した個体識別が可能であることを示した。

目次

1.	はじめに	1
2.	牛個体識別.....	3
2.1	カメラを用いた個体識別の必要性	3
2.2	従来の牛個体識別.....	3
2.2.1	耳標による個体識別	3
2.2.2	鼻紋による個体識別	5
2.2.3	牛顔画像による個体識別	6
2.2.4	複数の情報を併用した個体識別.....	7
3.	耳標と牛顔画像の併用による牛個体識別	9
3.1	耳標と牛顔画像の併用	9
3.2	牛個体識別手法	9
3.3	YOLO による耳標および牛顔検出	10
3.4	GoogleOCR による個体識別番号の読み取り	11
3.5	ArcFace による牛顔特徴の抽出	11
3.6	個体識別番号と顔特徴量による個体識別	12
4.	トラッキングを利用した複数頭の牛個体識別	14
4.1	トラッキングによる時間方向の情報の活用	14
4.2	複数頭環境における牛個体識別手法の概要	14
4.3	YOLO と ByteTrack を用いた牛顔検出およびトラッキング	15
4.4	顔特徴量に基づくトラッキング ID の結合	15
4.5	トラッキング情報を用いた耳標・顔特徴の統合による個体識別	16
5.	評価実験.....	18
5.1	実験手法	18
5.2	各要素技術の評価	18
5.2.1	YOLO の学習と評価.....	18
5.2.2	GoogleOCR による耳標読み取り精度の評価.....	20
5.2.3	ArcFace による牛顔特徴抽出モデルの学習結果と性能評価	21
5.3	耳標および牛顔情報を用いた個体識別手法の実験	22
5.3.1	実験設定および評価条件	22
5.3.2	実験結果.....	23
5.3.3	閾値の選定	24
5.4	トラッキングを利用した複数頭の牛個体識別実験	32
5.4.1	実験条件および評価対象映像.....	32
5.4.2	ByteTrack と顔特徴量によるトラッキング	32

5.4.3 提案手法と単フレームによる個体識別結果との比較	33
6. おわりに	35
謝辞	36
参考文献	37

1. はじめに

近年、情報通信技術および画像処理技術の発展に伴い、畜産分野においても作業の自動化や効率化が進められている。特に、労働力不足や作業負担の軽減を背景として、家畜の管理や個体情報の把握を自動的に行う技術への関心が高まっている。畜産分野においては、家畜一頭一頭を識別し、行動や状態を把握することが重要である。そのため、非接触で情報を取得できるカメラを用いた個体識別技術への需要が高まっている。カメラを用いた個体識別は、装置の装着を必要とせず、飼育環境において自然な形で情報を取得できる点で有用である。

このような背景のもと、本研究では牛の個体識別に焦点を当てる。

牛の個体識別においては、耳標や牛顔画像、鼻紋など、様々な情報を用いた手法が提案されている。牛の個体識別において耳標は、法令に基づき付与される 10 桁の個体識別番号を保持する手段として広く用いられている[1]。画像からこの番号を正しく読み取ることができれば、高い確実性で個体を特定できる。牛の個体識別における耳標読み取りに関して、Thi Thi ら[2]はカメラ画像から牛頭部および耳標領域を検出し、個体識別番号を認識することで牛の個体管理を行う手法を提案しており、耳標に基づく個体識別の有効性が示されている。また、Smink ら[3]は、動画中の耳標検出と耳標追跡を統合した ReadMyCow システムを提案している。この手法では、追跡情報を用いて耳標が安定して視認できるフレームを選択することで、耳標番号の読み取り精度を向上させており、実環境でのリアルタイム運用を想定した耳標読み取り手法となっている。

しかし、上に示した耳標認識に基づく既存手法では、読み取り精度の向上が主眼であり、読み取り結果を識別にどのように活用するかという点で工夫の余地が残されている。Thi Thi らの研究でも誤認識や一部欠損に対処する工夫は見られるが、耳標のみを用いて最終的な個体を決定する構成であるため、番号が類似した個体が多数存在する環境では識別が困難になる場合がある。実際、同一牧場内には極めて番号が近い個体が複数存在する場合があり、耳標情報のみで確実な識別を行うことには限界がある。

そこで本研究では、耳標読み取りにより得られた候補個体が複数存在する場合や信頼性が不十分な場合に、牛顔画像から抽出した顔特徴量を併用して最終的な個体識別を行う手法を採用する。牛顔画像は耳標に比べて遮蔽や装着状態の影響を受けにくく、カメラ画像から比較的安定して取得できる情報である。そのため、耳標の読み取り結果が十分でない場合に補助情報として用いることで、個体識別の信頼性向上が期待できる。耳標を主情報としつつ、牛顔画像を補助情報として統合することで、耳標情報の不完全性や個体番号の類似性に起因する曖昧性を排除し、より安定した個体識

別を実現する。

さらに、本研究では想定する状況に応じて、二つの識別手法を導入する。第一の手法は、単一フレームを対象とした個体識別であり、画像内に一頭の牛が存在する状況を想定して、各フレームに対して耳標と牛顔画像に基づく識別処理を逐次適用する。この手法では、フレームごとに結果が決定されるため、静止画像や単独個体の監視に有効である。

一方、実環境では複数頭の牛が同時に映り込む場合が多い。そこで第二の手法として、時間方向に連続する複数フレームの牛顔の追跡情報を利用した個体識別を導入する。追跡を行うことで、複数個体が同時に存在する状況においても、時間方向の情報を利用して安定した個体識別が行える。追跡手法としては ByteTrack[4]を基盤としつつ、検出領域の情報に加えて、各フレームで抽出される顔特徴量の類似度に基づいて追跡対象の対応付けを行う。この顔特徴量を組み込んだ追跡により、一時的な検出欠落や位置変動が生じた場合でも同一個体を追跡し続けることができる。

以下、2章では牛の顔個体識別に関する従来研究について述べる。3章では、1フレームごとに個体を識別する第一の提案手法について説明する。4章では、時間方向に連続する複数フレームを統合して複数頭に対応する第二の提案手法について述べる。5章では、本研究で提案する個体識別手法に関する実験結果を示す。最後に、6章では本研究のまとめと今後の課題について述べる。

2. 牛の個体識別

2.1 カメラを用いた個体識別の必要性

畜産現場においては、牛一頭一頭を正確に識別し、個体ごとの行動や状態を把握することが重要である。従来、牛の個体管理には耳標や台帳による目視確認が用いられてきたが、飼養頭数の増加や作業の大規模化に伴い、人手による管理には限界が生じている。特に、個体識別や行動記録を継続的に行う作業は労力を要し、作業負担の増大や人的ミスの原因となる。

このような背景から、近年ではセンサやカメラを用いた自動的な個体識別技術への関心が高まっている。中でも、カメラを用いた手法は、装置を牛に装着する必要がなく、非接触で情報を取得できる点で有用である。また、既存の飼養環境にカメラを設置することで導入可能であり、実環境への適用性が高い。

2.2 従来 of 牛個体識別

カメラ画像を用いた個体識別は、装置の装着を必要とせず、飼育環境において自然な形で個体情報を取得できることから、多くの研究が進められている。これまでに、耳標や鼻紋、牛顔画像など、牛の外観に基づくさまざまな識別手法が提案されており、それぞれに特徴や利点がある。一方で、撮影条件の変化や個体差の表れ方などの影響により、各手法では識別精度の安定性に課題が報告されている。

本節では、耳標、鼻紋、牛顔画像、および複数の情報を併用する手法に関する先行研究を取り上げ、それぞれの特徴と精度面の課題について整理する。

2.2.1 耳標による個体識別

耳標は、牛に装着することが義務付けられている標識であり、法令に基づいて付与される 10 桁の個体識別番号を表示するために用いられている。この番号は全国的に一意であり、正しく読み取ることができれば高い確実性で個体を特定できるという利点を持つ。そのため、カメラ画像から個体識別番号を自動的に読み取る手法は、非接触かつ実用性の高い個体識別手段として注目されている。

一方で、実際の飼養環境では、耳標が泥や汚れによって汚損される、装着状態の変化により一部が隠れる、あるいは牛の姿勢や遮蔽物の影響によって視認できないといった問題が発生することがある。特に、10 桁すべてを安定して読み取ることは困難であり、番号の後半のみなど部分的な読み取り結果にとどまるケースも多い。これらの例を図

1 に示す。



図 1 読み取りが困難な耳標の例

Smink ら[3]は動画中の耳標検出と耳標追跡を統合した ReadMyCow システムを提案している(図 2)。この手法では、追跡情報を用いて耳標が安定して視認できるフレームを選択し、認識精度を高めることで実環境下でのリアルタイム運用を可能としている。しかし、撮影環境によっては耳標が長時間視認できない場面が生じるほか、耳標の読み取り自体が誤認識となった場合の扱いは限定的であり、認識精度の安定性には課題が残る。



図 2 ReadMyCow システムによる個体識別[3]

こうした状況に対して、Thi Thi ら[2]の研究では部分的な読み取り結果や複数条件に基づく判定を導入することで、耳標が不鮮明な場合にも識別可能性を確保する工夫が施されている(図 3)。しかし、読み取り結果が番号の後半のみなど限られた部分情報にとどまる場合には、同一農場内に登録している番号が完全に一致する個体が複数存在することも珍しくなく、耳標番号のみでは個体を一意に識別できない状況が依然として生じる。

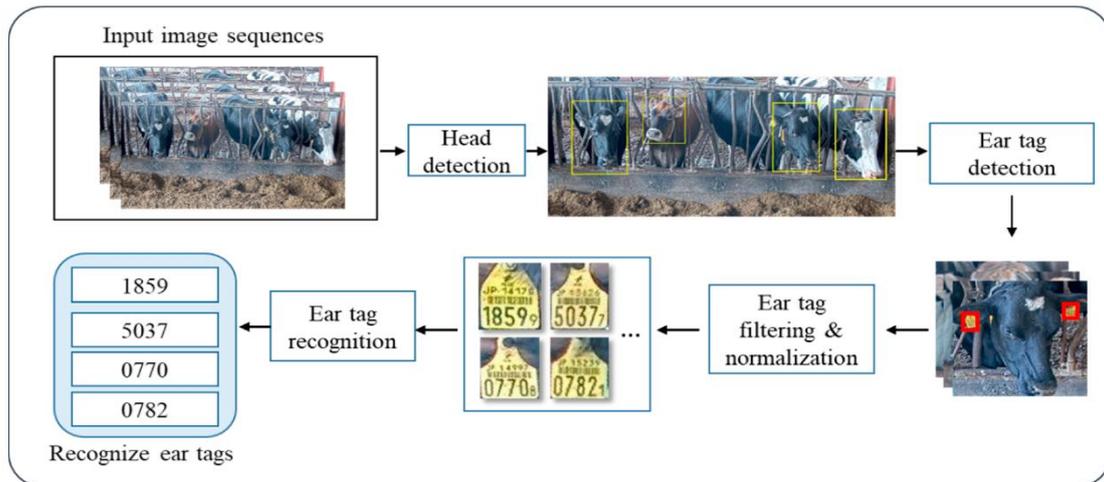


図 3 耳標の一部の遮蔽・欠損を想定した個体識別[2]

2.2.2 鼻紋による個体識別

鼻紋（鼻鏡表面の模様）を個体識別に利用する研究も報告されている。鼻紋は人の指紋に類似した固有模様であり、同一個体では経時的な変化が少ない点が利点とされる。鼻紋の参考画像を図4に示す。Kimaniら[5]は、鼻紋画像に深層学習（CNN）を適用した個体識別手法を提案し、従来の特徴量ベース手法と比較して高い識別精度が得られることを示している。



図 4 鼻紋の参考画像

一方で、鼻紋に基づく識別は撮影条件に大きく左右されるという課題がある。牛の鼻面は常に湿っており、光の反射により模様コントラストが低下することがある。また、給餌や飲水、泥による汚れにより模様が不明瞭になる場合がある。さらに、画像の撮影

距離や角度が変化すると領域抽出が難しくなり、特徴量の取得が安定しない傾向がある。そのため、鼻紋自体は識別精度の潜在的な高さを持つものの、実環境においては常に高品質な鼻紋画像を取得することが難しく、識別精度の安定性に課題が残る。

2.2.3 牛顔画像による個体識別

深層学習を用いた牛顔画像による個体識別も近年注目を集めている。牛顔画像に基づく個体識別は、牛の頭部領域を画像として取得できれば利用可能であり、特別な装着物や接触を必要としない点に特徴がある。撮影距離や姿勢に多少のばらつきがあっても顔領域を取得しやすく、実環境において比較的安定してデータを収集できる利点を有する。

Weng[6]は、図5のように二系統の畳み込みニューラルネットワーク（Two-Branch CNN）を用いて牛顔画像から個体識別を行う手法を提案し、複数視点の顔情報を統合することで高い識別精度が得られることを示している。

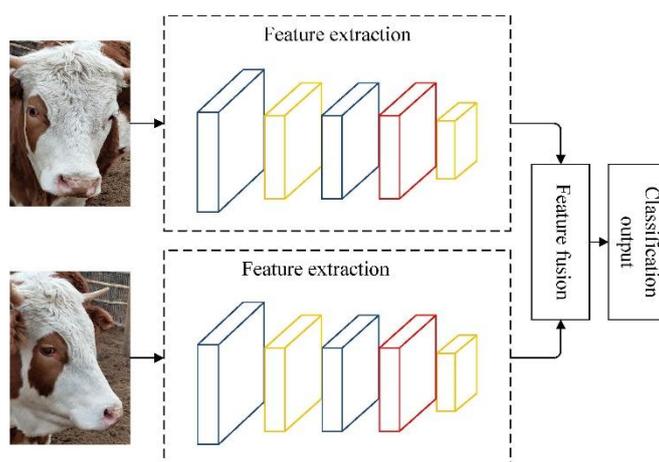


図5 二系統の畳み込みニューラルネットワークによる牛顔個体識別モデルの構成図[6]

しかしながら、牛の顔画像は外観の変動が大きいため、識別精度が不安定になりやすいという問題がある。個体間で毛色や形状が類似している場合、識別が困難となる。また、顔の方向（正面、横、俯きなど）による見え方の変化、毛の伸びや湿り具合といった経時的な外観変化により、特徴抽出の安定性が損なわれる。さらに、登録時と認証時の撮影条件の差が大きい場合、モデルが取得した特徴量を十分に再現できず、精度低下につながる。したがって、牛顔画像のみを利用した個体識別は汎用性が高い一方で、画像品質や条件変動に起因する精度の揺らぎが課題となる。

2.2.4 複数の情報を併用した個体識別

耳標、鼻紋、牛顔画像といった複数の情報を組み合わせて識別を行う研究も提案されている。これらの手法は、単一の情報源では識別が不安定となり得る状況に対して、複数の特徴を相補的に利用することで精度向上を図る点に特徴がある。特に、牛顔画像や鼻紋といった耳標以外の特徴を活用することで、耳標が見えない状況下でも識別可能性を確保しようとする試みが行われている。

Liら[7]は、耳標、鼻紋、牛顔画像の三種類の特徴を取得し、それぞれから得られた識別結果を決定レベルで統合することで個体識別を行う手法を提案している(図6)。この研究では、複数の特徴を対等に扱い、いずれか一つの情報が不安定であっても補完できる構造とすることで、高い識別精度を報告している。これは、耳標や顔といった特徴が環境条件に左右されるという課題に対して、複数情報の融合という観点から取り組んだ先行研究として位置付けられる。

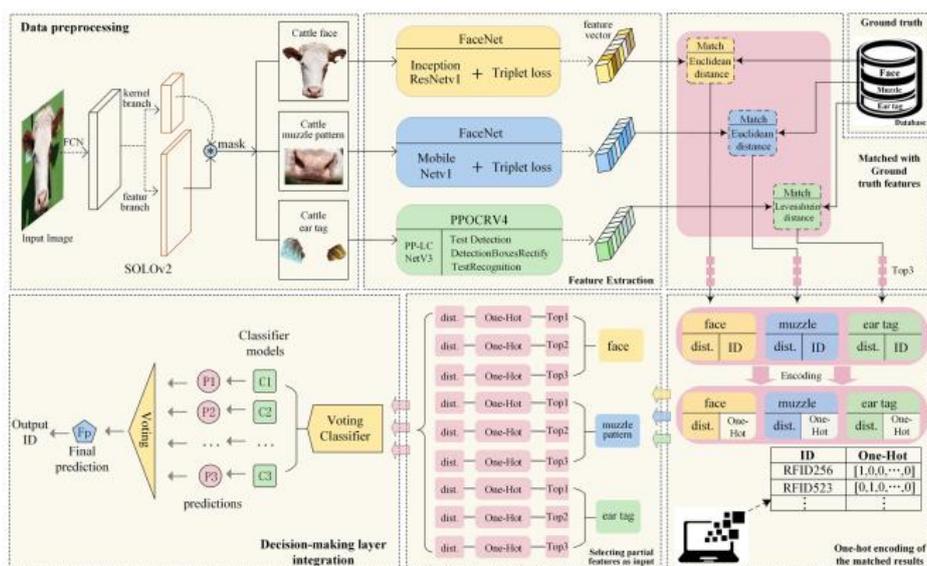


図6 耳標・鼻紋・牛顔画像の各識別結果を決定レベルで統合する個体識別処理[7]

しかしながら、このような複数情報の併用手法においても、情報源それぞれが持つ弱点が識別精度に影響するという課題は残されている。Liらの研究では、耳標を含む3種類を等価に扱うことで高精度を達成している一方、耳標に書かれた個体識別番号の読み取り誤りを許容して候補を段階的に絞り込む手法や、動画における継続的な観測を活用して時間方向に情報を統合する識別方式については明示的に扱っていない。また、実環境においては耳標の読み取り精度が大きく揺らぐ場面が存在するため、その誤認識が融合結果に直接反映されてしまう可能性がある。

このように、複数情報を併用した識別は潜在的に有効なアプローチであるものの、情

報源の品質変動や誤認識をどのように扱うかといった課題が残されており、特に誤読を許容する枠組みや時間的な情報統合の取り扱いについては十分に検討されていない点が指摘できる。

3.耳標と牛顔画像の併用による牛個体識別

本章では、耳標と牛顔画像を組み合わせて個体識別を行う本研究の第一の手法について述べる。本手法は、入力画像に牛が一頭のみ映る状況を対象とし、各フレームに対して耳標読み取りと顔特徴量照合を適用することで識別を行うものである。

3.1 耳標と牛顔画像の併用

本研究では、耳標読み取りに伴う誤認識や不読取り、および牛顔画像による識別精度の揺らぎといった課題に対して、両者の特徴を補完的に利用することで個体識別の信頼性を高める手法を提案する。本手法の基本的な考え方は、耳標情報を主たる識別手掛かりとして優先的に用いながら、耳標が正確に取得できない場合や判断が確定しない場合には顔特徴量に基づく判断を加えるという二段階の判定構造にある。

まず、耳標から得られる文字列と登録済み番号の間に、2つの文字列の違いを表す指標であるレーベンシュタイン距離を導入することで、完全一致しない場合でも候補個体を抽出できるようにし、耳標単独の脆弱性を補う。続いて、候補個体が複数残る状況に対して顔特徴量を併用することで候補間の識別を行い、耳標情報のみでは対処できない局面を補強する。顔特徴量は安定して取得できるが、条件変動により精度が低下する場合がある。耳標情報で候補個体を限定した上で顔特徴量を利用することで、精度低下の影響を低減できる。さらに、候補が1つに定まった場合であっても顔特徴量との類似度を検証することで、耳標読み取りでの誤認識を排除する仕組みとする。このように耳標と牛顔画像を段階的に利用することで、両者の弱点を相互に補い合い、実環境における個体識別精度の安定化を図ることが可能となる。

3.2 牛個体識別手法

本節では、前節で述べた基本的な考え方にに基づき、本研究で採用する耳標と牛顔画像を併用した牛個体識別手法の処理手順について述べる。なお識別対象の各牛について、あらかじめ耳標に記載されている個体識別番号および牛顔画像から抽出した顔特徴量を登録情報として保持しておくものとする。

本手法は、入力画像中に牛が一頭のみ存在する状況を対象とし、各フレームに対して耳標情報と顔特徴量を登録情報と比較することで識別を順次行うものである。処理は大きく4つの段階から構成される。処理の流れを図7に示す。

1. YOLO による耳標および牛顔検出
2. GoogleOCR による個体識別番号の読み取り
3. ArcFace による顔特徴量の抽出
4. 個体識別番号と顔特徴量による個体識別

以下、各段階の処理について述べる。

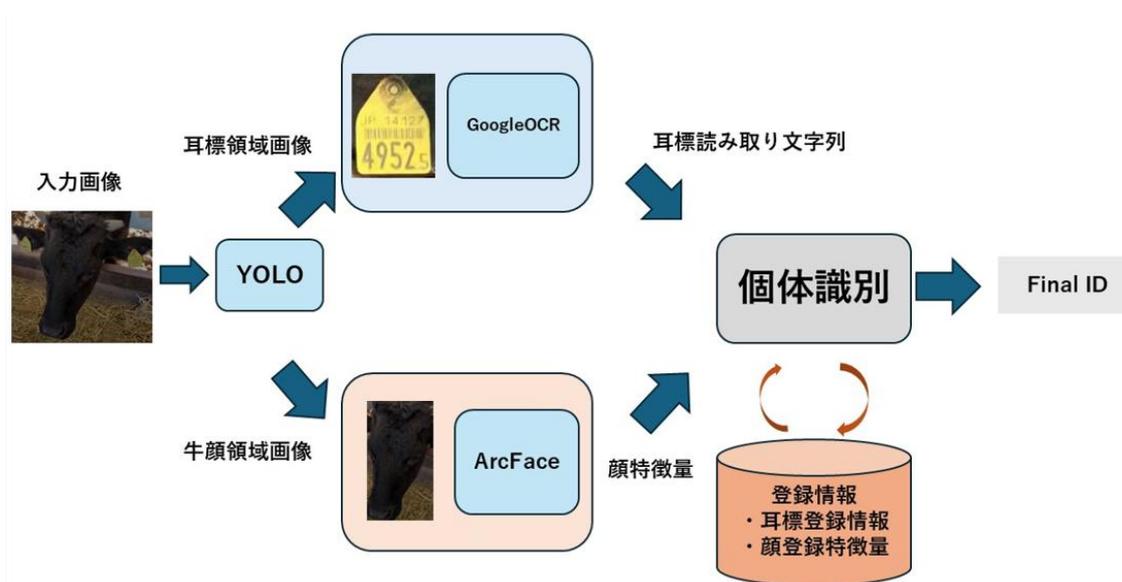


図7 耳標情報と牛顔特徴量を併用した牛個体識別手法の処理

3.3 YOLO による耳標および牛顔検出

本手法では、各フレームに対して耳標読み取りおよび牛顔特徴抽出を行う前段階として、入力画像中から耳標領域および牛顔領域を検出する。入力は牛舎環境で撮影されたRGB画像であり、出力は耳標および牛顔に対応する切り出し画像である。

これらの領域検出にはYOLO[8]を用いる。YOLOは単一のニューラルネットワークにより高速に物体検出を行う手法である。本研究では耳標および牛顔を検出対象として、独自に構築したデータセットを用いて学習を行った。以降、YOLOにより検出・切り出された耳標および牛顔の画像を、それぞれ耳標領域画像、牛顔領域画像と呼ぶ。

3.4 GoogleOCR による個体識別番号の読み取り

本手法では、検出された耳標領域から牛の個体識別番号を取得する。入力は、前節で述べた耳標領域画像であり、出力は耳標に記載された番号に対応する文字列である。

耳標領域画像に対する文字認識には GoogleOCR[9]を用いる。切り出した耳標領域画像をそのまま入力として与え文字列を取得する。得られた文字列は、後段の個体識別処理における候補個体の絞り込みに利用される。個体識別番号は数字のみで構成されるため、言語の設定は特に行わず、一般的なテキスト認識として処理を行う。GoogleOCRの出力はテキスト形式で返され、耳標上に複数の文字列が存在する場合や、領域が複数行として認識された場合には、改行を含む複数行の文字列として出力される。本研究では、これらの各行を耳標による個体識別に用いる。また、大量の静止画から多数の耳標領域画像が生成されるため、GoogleOCRの呼び出しは複数の画像に対して32並列に処理を行った。これにより、多数の耳標領域画像に対する処理時間を短縮し、実用的な処理時間内で個体識別を行うことが可能となる。

3.5 ArcFace による顔特徴量の抽出

本研究では、牛顔領域画像から個体識別に用いる顔特徴量を抽出するために ArcFace [10]を用いる。

ArcFace は、顔認識で広く用いられている畳み込みニューラルネットワーク (CNN) の学習方式である。入力画像に対して CNN が抽出する特徴ベクトルを、同じクラス同士の特徴ベクトルの角度は小さく、異なるクラス間の特徴ベクトルの角度は大きくなるように学習を行う(図8)。

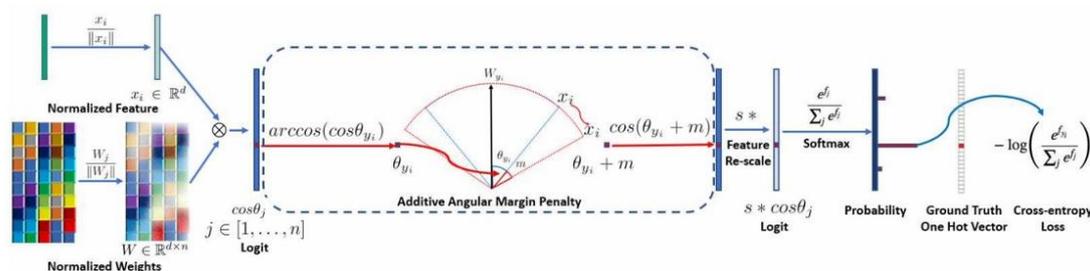


図 8 ArcFace の処理の流れ[10]

ArcFace はクラス間の分離性を強化した特徴表現を学習できるため、個体識別に適した安定した特徴量を得ることが可能である。

本研究では、ArcFace の学習済みモデルのうち特徴抽出に必要なバックボーン部分を用い、牛顔画像から埋め込み特徴（特徴ベクトル）を生成する。バックボーンには EfficientNet-B3 を採用し、入力画像サイズは 512×512 とした。学習時には RandAugment (PyTorch) によるデータ拡張を適用し、撮影条件の変動に対する頑健性向上を図った。

3.6 個体識別番号と顔特徴量による個体識別

本研究では、耳標に記載された個体識別番号と顔特徴量を組み合わせることで、個体識別を行う。牛の個体識別番号は全体で 10 桁から構成されるが、実際の耳標では番号の後半部分が前半部分よりも大きく明瞭に描かれている。そこで本手法では、耳標領域画像から比較的安定して読み取りが可能な後半 5 桁の数字列を用いて識別処理を行う。

本研究の個体識別では、あらかじめ識別対象となる牛の個体識別番号と顔特徴量を登録しておき、入力画像から得られた情報と照合する。登録情報としては、各牛について個体識別番号の後半 5 桁に着目し、そこから生成される複数の文字列パターンをあらかじめ保存する。具体的には、後半 5 桁そのものに加えて、その先頭 4 桁および末尾 4 桁の計 3 種類を登録情報として用いる。これにより、耳標の一部が欠損した場合や、OCR により一部の桁のみが認識された場合であっても、候補個体として扱うことが可能となる。

また、各牛について、登録用の牛顔画像から ArcFace により抽出した顔特徴量をあらかじめ保存する。

以降、本研究では、個体識別番号に基づく登録情報を「耳標登録情報」、牛顔画像に基づく登録情報を「顔登録特徴量」と呼ぶ。耳標登録情報の作成例を図 9 に示す。



図 9 耳標登録情報の作成

候補個体の選定には、3.4 節で述べた耳標読み取り (GoogleOCR) により得られた文字列と、あらかじめ保存した耳標登録情報との間でレーベンシュタイン距離を計算する。レーベンシュタイン距離は、2 つの文字列を一致させるために必要な編集操作 (挿入・削除・置換) の最小回数を表す指標であり、値が小さいほど文字列同士が近いことを意味する。したがって、OCR により 1 桁の誤読や一部欠損が生じた場合でも、距離の小さい候補を残すことで候補個体を抽出できる。

候補の選定は段階的に行う。まず距離 0 (完全一致) となる候補を探索し、該当する個体が存在しない場合には距離 1 の候補を探索する。距離 2 以上の候補は誤認識の可能性が高くなるため、本研究では個体識別には用いない。このように距離に基づいて候補を制限することで、誤読や部分的欠損を一定程度許容しつつ、候補個体を効率的に絞り込む。

レーベンシュタイン距離による候補の選定の結果、候補が一意に定まらない場合には、3.5 節で述べた ArcFace により抽出した顔特徴量を用いて最終判定を行う。具体的には、現在フレームから得られた顔特徴量と、各候補個体に対応する顔登録特徴量との類似度を計算し、類似度が最も高い個体を識別結果として採用する。類似度の計算には、特徴ベクトルを L2 正規化 (長さ 1 に正規化) した上での内積に相当するコサイン類似度を用いる。コサイン類似度は特徴ベクトル間の角度に基づく指標であり、値が大きいほど同一個体である可能性が高い。本実験では、類似度が最も高い候補を個体識別結果として採用する。耳標読み取りによって候補が一意に決定された場合でも、顔特徴量に基づく類似度を確認し、あらかじめ設定した閾値以下であれば識別を拒否する処理を行う。この判定により、耳標読み取りに依存した誤判定を抑制し、識別結果の信頼性を高めることができる。

以上のように、本研究では耳標登録情報による候補の選定と、顔登録特徴量による補助的な識別を段階的に組み合わせることで、耳標による情報のみでは識別が困難となる状況に対しても、安定した個体識別を実現する。

4. トラッキングを利用した複数頭の牛個体識別

4.1 トラッキングによる時間方向情報の活用

実環境で取得される映像では、複数頭の牛が同時に映り込むことが多く、撮影角度の変化や遮蔽の影響により、単一フレームのみでは耳標番号や顔情報を十分に取得できない場合がある。そのため、3章で示した単一フレームに基づく識別手法を複数頭環境にそのまま適用すると、耳標の欠損や誤認識の影響を受けやすく、識別結果が不安定となる。

そこで本研究では、トラッキングを導入することで時間方向に連続する複数フレームの情報を統合し、個体識別の安定化を図る。トラッキングにより、同一個体に対応する検出結果をフレーム間で関連付けることで、耳標や顔が一時的に視認できない場合であっても、過去および将来のフレーム情報を活用した識別が可能となる。

本章では、このトラッキングに基づく時間方向の情報統合を利用し、複数頭の牛が同時に存在する映像に対して個体識別を行う手法を提案する。

4.2 複数頭環境における牛個体識別手法の概要

本研究で提案する複数頭の牛個体識別手法は、大きく3つの段階から構成される。

1. YOLO と ByteTrack を用いた牛検出およびトラッキング
2. 顔特徴量に基づくトラッキング ID の結合
3. トラッキング情報を用いた耳標・顔特徴量の統合による個体識別

本研究の個体識別では、3章で述べたように、識別対象の牛の個体識別番号および顔特徴量をあらかじめ登録しておき、入力動画から得られる情報と照合することで識別する。

以下、各段階の処理について述べる。

4.3 YOLO と ByteTrack を用いた牛顔検出およびトラッキング

本手法では、複数頭の牛が同時に存在する映像において、同一個体を時間方向に追跡するため、フレームごとに牛顔を検出し、フレーム間で対応付けを行う。入力には牛舎環境で撮影された連続フレームの RGB 動画であり、これに対して各牛顔を基準とした追跡処理を行う。追跡結果は、各フレームで検出された牛顔領域にトラッキング ID を割り当てることで表現される。異なるフレームの牛顔領域で同じトラッキング ID を持つものは、同一個体を表す。

この処理を実現するために、本研究では YOLO を用いた物体検出と、ByteTrack によるトラッキングを組み合わせる。YOLO により各フレームから牛顔および耳標領域を検出し、ByteTrack によって牛顔検出結果をフレーム間で関連付けることで、同一個体に対応する牛顔を時間方向に一貫して追跡する。なお、トラッキングは牛顔に対してのみ適用し、耳標は後段において牛顔との位置関係に基づき対応付けを行う。YOLO には様々なバージョンが存在する。本研究では、物体検出には YOLO11 を、トラッキングには YOLO11 に標準実装されている ByteTrack を用いる。

4.4 顔特微量に基づくトラッキング ID の結合

前節で述べた YOLO と ByteTrack を用いたトラッキングにより、牛顔を時間方向に追跡することが可能となる。しかし、実環境の映像では、牛同士の接近や重なり、急激な姿勢変化、検出の一時的な欠落などの影響により、同一個体であってもトラッキング ID が途中で分断される場合がある。このようなトラッキングの分断により、同一個体に関する情報が複数のトラッキング ID に分散し、時間方向の情報を十分に統合できないという課題が生じる。

そこで本研究では、ByteTrack によって生成された複数のトラッキング ID に対して、顔特微量に基づく再結合処理を行うことで、同一個体に対応するトラックを統合し、時間方向の情報をさらに増加させる。

具体的には、各トラッキング ID に属する牛顔領域画像から ArcFace により顔特微量を抽出する。トラッキング ID 間の結合においては、単一フレームに基づく類似度比較は外観変動の影響を受けやすいため用いず、各トラッキング ID に含まれる複数の顔特微量集合を用いた比較を行う。二つのトラッキング ID A, B を比較する際には、まず、

A に属する各牛顔領域画像について、B に属する牛顔領域画像との類似度の最大値を求める。類似度の算出には、それぞれの牛顔領域画像から抽出した顔特徴量間のコサイン類似度を用いる。同様に、B に属する各牛顔領域画像についても、A に属する牛顔領域画像との類似度の最大値を求める。このようにして求めた顔特徴量対の類似度の平均値（平均コサイン類似度）をトラッキング ID 間の類似度とする。

本研究では、この平均コサイン類似度が 0.6 以上の場合に、両トラッキング ID は同一個体に対応するものと判断し、これらを結合する。

このように、顔特徴量に基づいてトラッキング ID を再結合することで、ByteTrack のみでは分断されてしまうトラックを統合し、時間方向により長い区間にわたる個体情報の蓄積が可能となる。これにより、後段における耳標情報および顔特徴量を用いた個体識別の安定性向上が期待できる。

4.5 トラッキング情報を用いた耳標・顔特徴の統合による個体識別

本節では、前節までに述べたトラッキングおよびトラック結合の結果を基に、時間方向に統合された情報を用いて牛の個体識別を行う手法について述べる。本手法では、結合済みのトラッキング ID を単位として耳標情報および顔特徴量を集約し、最終的な個体識別を行う。

まず、各フレームにおいて耳標が検出された場合には、その位置関係に基づいて最も近い牛顔領域が属するトラッキング ID に対応付ける。各トラッキング ID につき各フレームで左右それぞれ最大 1 個、合計 2 個までの耳標を割り当てるものとする。これにより、複数頭の牛が同時に存在する状況においても、耳標を適切な個体に関連付けることが可能となる。

次に、耳標が割り当てられたトラッキング ID に対して、対応する耳標領域画像を GoogleOCR により読み取り、得られた文字列をトラッキング ID ごとに蓄積する。動画の中ではフレームごとに読み取り結果が変動する場合があるため、本研究では各フレームで得られた文字列をそれぞれ入力として扱い、時間方向に統合する処理を行う。具体的には、各読み取り結果と登録済みの耳標登録情報との間でレーベンシュタイン距離を計算し、距離 0（完全一致）となる候補を優先的に探索する。距離 0 の候補が存在しない場合には、距離 1 の候補を探索対象とする。距離 2 以上の候補については誤認識の可能性が高いため、本研究では処理対象から除外する。

各フレームの耳標読み取り結果において、レーベンシュタイン距離の条件を満たす候

補が複数存在する場合には、それぞれの候補個体に対して1票を加算する。これをトラッキングIDに属する全フレームに対して繰り返すことで、耳標読み取り結果を時間方向に集約した投票結果を得る。この投票処理により、一時的な誤読や部分的な欠損が生じた場合であっても、複数フレームにおいて一貫して出現する候補を高信頼な識別候補として抽出することが可能となる。

耳標の投票結果に基づき、各トラッキングIDに対して得票数の多い上位3件の個体を候補個体として選定する。

次に、これらの候補個体に対して、トラッキングIDに属する各フレームの牛顔領域画像から ArcFace により顔特徴量を抽出し、あらかじめ登録された各候補個体の顔登録特徴量との間でコサイン類似度を計算する。各フレームにおいて、コサイン類似度が最も高い候補個体を1票として加算する。この処理をトラッキングIDに含まれる全フレームに対して行う。最終的に、トラッキングID単位で最も多くの票を獲得した個体を、当該トラッキングIDに対応する牛の識別結果として決定する。

このように、本研究では ByteTrack によるトラッキングを基盤とし、顔特徴量によるトラック結合、さらに耳標情報と顔特徴量を時間方向に統合することで、複数頭の牛が同時に存在する実環境においても、安定した個体識別を可能としている。

5. 評価実験

5.1 実験手法

本章では、3章および4章で提案した牛個体識別手法の有効性を実験的に検証する。

まず、個体識別に用いる要素技術として、耳標および牛顔の検出に用いる YOLO、耳標読み取りに使用する GoogleOCR、ならびに特徴抽出に用いる ArcFace の学習結果および性能について確認する。これにより、後段で行う個体識別実験において使用する各要素技術が、実環境において十分な性能を有していることを示す。

次に、入力画像中に牛が一頭のみ存在する状況を想定し、3章で提案した第一の手法に基づく個体識別実験を行う。この実験では、耳標の読み取り結果に基づく候補絞り込みの有効性、ならびに顔特徴量を適切に併用することで誤識別を抑制できるかを評価する。

さらに、複数頭の牛が同時に映り込む動画を対象として、4章で提案した第二の手法の評価を行う。この実験では、トラッキングによる時間方向の情報統合が、耳標の不完全な読み取りや一時的な遮蔽が生じる状況においても、安定した個体識別を可能とするかを検証する。

5.2 各要素技術の評価

5.2.1 YOLO の学習と評価

本研究では、耳標読み取りおよび牛顔特徴抽出を行う前段階として、入力画像中から牛顔および耳標領域を検出するために YOLO を用いる。

検出性能の評価指標としては Precision, Recall, mAP を用いる。Precision は検出された領域のうち正しく検出された割合を表し、誤検出の少なさを示す指標である。一方、Recall は正解領域のうち検出に成功した割合を表し、見逃しの少なさを示す指標である。なお、Precision および Recall の算出にあたっては、検出結果が正解か否かを IoU (Intersection over Union) により判定する。IoU は検出矩形と正解矩形の重なり具合を表す指標であり、共通部分の面積を和集合の面積で割った値として定義される。本研究では IoU が 0.5 以上の検出を正解 (True Positive) とし、それ未満は不正解として扱う。また、mAP (mean Average Precision) は検出精度と位置精度を総合的に評価

する指標である。mAP は、検出の信頼度スコアの閾値を変化させたときの Precision-Recall 曲線を作成し、その曲線下面積 (AP) を求め、これを全クラスで平均したものである。本研究では IoU 閾値 0.5 における mAP50 および、IoU 閾値を 0.5 から 0.95 まで変化させた場合の平均である mAP50-95 を用いて評価を行う。

特に、単一画像を対象とする場合と、複数頭の牛が同時に映り込む映像を対象とする場合とでは、求められる検出特性が異なる。そこで本研究では、目的の異なる 2 種類のデータセット (データセット A, データセット B) を用いて YOLO の学習および評価を行った。データセット A は、本研究における第一の手法で用いることを目的として構成したデータセットであり、単一フレームにおいて耳標および牛顔を確実に検出することを主眼としている。牛舎環境で撮影された静止画像から構成され、訓練用 402 枚、検証用 33 枚の画像に対して、耳標および牛顔領域のバウンディングボックスによるアノテーションを付与した。

本研究では YOLO11 を採用し、耳標と牛顔を検出対象として学習を行った。学習は 20 epoch を基準に実施し、検証セットに対して Precision, Recall, mAP を指標としてモデル性能を評価した。最も良好な性能が得られた 17 epoch 時点の結果を表 1 に示す。

表 1 学習結果 (Yolo 学習用データセット A)

評価指標	結果
Precision	0.912
Recall	0.976
mAP50	0.960
mAP50-95	0.815

本モデルは Recall が 0.98 と非常に高く、耳標および牛顔の見逃しがほぼ発生しない検出性能を示した。一方で Precision は 0.89 とやや低く、誤検出が一定数含まれる可能性が示唆されるが、本研究では後段の識別処理により影響が抑制されると考えられる。また、mAP50 が 0.97, mAP50-95 が 0.83 と高い値を示しており、検出位置の精度についても実用上十分であると判断できる。

一方、データセット B は、本研究における第二の手法である複数頭環境における個体識別を想定し、牛顔の追跡安定性を高めることを目的として構成したデータセットである。データセット B は、データセット A を基に、連続フレーム中での顔の位置変動や姿勢変化、個体同士の接近や重なりを含む牛顔画像を追加することで増強している。データセットは訓練用 1074 枚、検証用 131 枚の画像から構成され、牛顔領域に対してバウンディングボックスによるアノテーションを付与した。学習は 100 epoch で実施し、検証セットに対して Precision, Recall, mAP を指標として性能評価を行った。最も良好な性能が得られた 97epoch 時点での学習結果を表 2 に示す。

表 2 学習結果 (Yolo 学習用データセット B)

評価指標	結果
Precision	0.889
Recall	0.847
mAP50	0.912
mAP50-95	0.735

本モデルは Precision が 0.889, Recall が 0.847 であり、トラッキングに必要な牛顔検出を概ね安定して行える性能を示した。また、mAP50 は 0.912, mAP50-95 は 0.735 であり、顔位置の検出精度についても実用上十分であると判断できる。単一フレームにおける検出性能はデータセット A で学習したモデルと比較してやや低下しているものの、本研究では後段でトラッキングにより時間方向の情報を統合するため、単発的な検出誤差が最終的な個体識別性能に与える影響は限定的であると考えられる。

5.2.2 GoogleOCR による耳標読み取り精度の評価

本節では、耳標読み取りに用いる GoogleOCR の認識精度について評価を行う。GoogleOCR は google が提供する OCR(光学文字認識)サービスである。本研究では、認識性能のみを評価対象とする。

精度評価には、181 頭の牛を撮影した単体撮影動画を用いた。この動画の総フレーム数は約 18 万枚である。なお、これら 181 頭の牛は、YOLO, ArcFace の学習には使用していない。また、本データセットは 5.3 節で述べる第一の手法の評価実験にも用いる。

各フレームに対して、5.2.1 節で述べたデータセット A により学習した YOLO を用

いて耳標検出を行い、検出された耳標のうち最も信頼度スコアの高い領域を 1 枚選択する。これにより、最大で約 18 万枚の耳標画像が得られるが、実際には検出に成功した 98,995 枚の耳標画像を評価対象とした。

各耳標領域画像に対して GoogleOCR を適用し、数値列を取得する。OCR の出力は 1 行または複数行の数値列として得られる場合があるが、今回は GoogleOCR の出力のうち、4~5 桁の数値列のみを評価対象とした。

評価に際して、GoogleOCR により得られた数値列と耳標登録情報との間でレーベンシュタイン距離を計算し、距離が 0 (完全一致) となるものが存在する場合を正解と定義する。一方、距離 0 の一致が得られない場合は不正解とした。本研究では、正解数を入力した耳標の画像数で割った値を GoogleOCR の認識精度と定義する。

その結果、YOLO により検出された 98,995 枚の耳標画像のうち、レーベンシュタイン距離が 0 となったものは 57,530 枚であった。したがって、本実験条件における GoogleOCR の耳標読み取り精度は、59.2%となった。これは、動画から切り出された耳標画像に含まれるブラー、撮影角度や照明条件の変動、耳標の汚れや摩耗などにより、OCR の安定した認識が困難であったためと考えられる。

なお、認識に失敗した約 40%の耳標画像のうち、約 14% (13,960 枚) では数値列が検出されなかった。一方、残りの約 26%では数値列自体は検出されたものの、個体識別番号の後半 5 桁を正確に認識できなかったと考えられる。

5.2.3 ArcFace による牛顔特徴抽出モデルの学習結果と性能評価

本節では、牛顔特徴抽出に用いる ArcFace モデルの学習結果および識別性能について述べる。

本研究では、バックボーンとして EfficientNet-B3 を用い、入力画像サイズを 512×512 として学習を行った。学習は 50 epoch 実施し、データ拡張には RandAugment (PyTorch) を適用した。本データセットは 3,606 クラスから構成され、合計 353,583 枚の画像を 9:1 の比率で学習用 (317,869 枚) と検証用 (35,714 枚) に分割して用いた。検証用データに対する分類精度は 98.5% であった。

ArcFace の識別性能評価には、GoogleOCR の評価に用いた 181 頭の牛を対象とした。これらの牛は ArcFace の学習には使用しておらず、学習データとは完全に独立した個体である。評価では、登録用画像として 2024 年 1 月 18 日に撮影された牛顔画像を用い、認証用画像として約 1 年の期間を空けて撮影された 2025 年 1 月 24 日の画像を用いた。このように撮影時期の異なるデータを用いることで、成長や外観変化を含む実環境条件下での識別性能を評価した。登録に使用した画像数は 325,601 枚となっている。

各認証画像に対して ArcFace により特徴量を抽出し、顔登録特徴量との間でコサイン類似度を計算した。識別結果は、登録データの中で最もコサイン類似度が高い個体を予測値として採用することにより決定した。評価は、全認証画像に対して正しく個体を識別できた割合を精度として算出した。

その結果、認証用画像 167,376 枚のうち、正しく個体を識別できた画像は 108,438 枚であり、識別精度は 64.8% となった。この結果は、撮影時期が約 1 年異なる牛顔画像間において、顔特徴量のみに基づく個体識別が一定程度可能である一方、単独で用いる場合には識別精度に限界があることを示している。

5.3 耳標および牛顔情報を用いた個体識別手法の実験

5.3.1 実験設定および評価条件

本節では、3 章で述べた耳標と牛顔画像を併用する第一の個体識別手法について、牛単体を撮影した動画から切り出したフレームを対象とした評価実験を行う。

評価には、5.2 節で GoogleOCR の精度評価に用いたものと同一の動画データを使用した。この動画は牛単体を撮影したものであり、各フレームにおいて牛顔および耳標が検出可能である。YOLO により牛顔および耳標の両方が検出された画像は 102,750 件である。このうち誤って映り込んだ識別対象外の牛を取り除いた 98,995 件の画像を評価対象とする。

評価は、以下の 4 つの条件に分けて行った。

- (1) レーベンシュタイン距離が 0 であり、耳標情報のみを用いて候補個体が一意に定まる場合
- (2) レーベンシュタイン距離が 0 であるが、複数の候補個体が存在し、顔特徴量により 1 頭に絞り込む場合
- (3) レーベンシュタイン距離が 1 であり、耳標情報のみを用いて候補個体が一意に定まる場合
- (4) レーベンシュタイン距離が 1 であり、複数の候補個体が存在し、顔特徴量により 1 頭に絞り込む場合

ここで「候補個体が一意に定まる」とは、レーベンシュタイン距離が最小となる候補が 1 頭のみ存在する場合を指す。

5.3.2 実験結果

第一の手法における評価条件別の識別結果を表3に示す。表3では、レーベンシュタイン距離および候補個体数の違いに基づいて分類した4つの条件について、評価件数、正解数、不正解数、および識別精度をまとめている。なお、本実験では顔特微量比較でのコサイン類似度の閾値は設定せず、一意に定まった候補をそのまま識別結果としている。

表3 第一の手法の結果

条件	レーベンシュタイン距離	候補数	評価件数	正解数	不正解数	精度(%)
(1)	0	1	55,998	55,811	187	99.67
(2)	0	複数	1,532	1,529	3	99.80
(3)	1	1	4,581	2,958	1,623	64.57
(4)	1	複数	4,561	3,103	1,458	68.03

レーベンシュタイン距離が0であり、耳標情報のみで候補が一意に定まる条件(1)では、識別精度は99.67%と非常に高い値を示した。一方、同じ距離0であっても複数の候補が存在する条件(2)では、顔特微量を併用することで99.80%の高い精度が維持されている。

これに対して、レーベンシュタイン距離が1の条件では、耳標情報のみで候補が一意に定まる場合でも識別精度は64.57%に低下し、耳標情報単独による識別の限界が確認された。しかし、複数候補から顔特微量を用いて一頭に絞り込む条件(4)では、識別精度が68.03%に向上しており、顔特微量を併用することによる条件(3)と同程度の結果が得られた。

これらの結果から、耳標読み取りの結果が完全一致する理想的な条件では耳標情報のみで高精度な識別が可能である一方、耳標番号の誤読や部分的な欠損が生じる条件においては、顔特微量を併用することが個体識別精度の向上に寄与することが示された。

5.3.3 閾値の選定

各条件において顔特徴量のコサイン類似度に対して閾値を設定し、受理率および識別精度の変化を確認した。

条件(1)~(4)の各パターンについて個別に閾値の選定を行う。

(1) レーベンシュタイン距離 0・候補が一意に定まる場合の閾値選定

この条件は耳標 OCR の結果のみで候補個体が一意に定まる最も信頼性の高いケースである。結果を表 4 と図 10, 図 11 に示す。

表 4 (1)における顔特徴量の類似度閾値と識別性能の関係

閾値の設定	受理件数	正解数	精度	カバレッジ
threshold = 0.00	55,998	55,811	0.9967	1.0000
threshold = 0.41	48,091	48,090	1.0000	0.8588

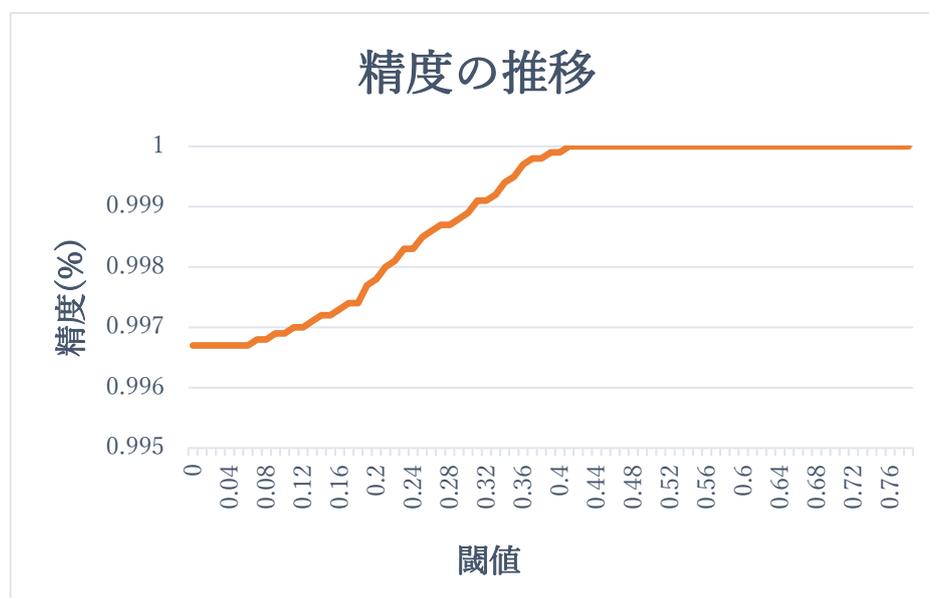


図 10 (1)における閾値ごとの精度の推移

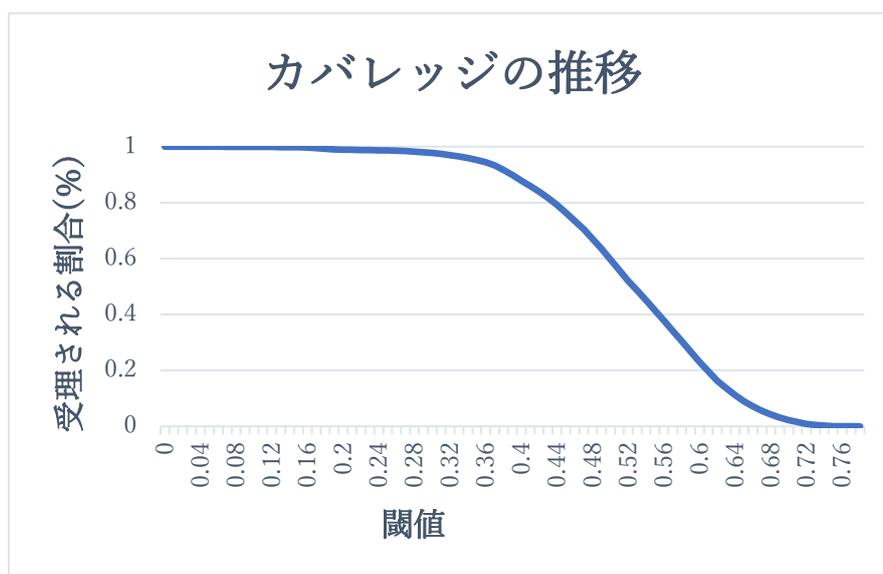


図 11 (1)における閾値ごとのカバレッジの推移

表 4 に示すように、顔特徴量のコサイン類似度に対して閾値を導入することで、閾値の増加に伴い識別精度は向上し、最終的には 100% に到達する。一方で、閾値を高く設定するほど、受理される件数（カバレッジ）は減少する。

特に注目すべき点として、閾値を設けない場合（threshold = 0）であっても、精度は 99.67% と非常に高く、実用上十分な性能を有している。

これは、本条件が耳標情報のみで個体が一意に定まっているため、顔特徴量を用いた追加判定を行わなくても、高精度な識別が可能であることを示している。

(2) レーベンシュタイン距離 0・複数候補の場合

本条件では、耳標文字列は完全一致（距離 0）であるものの、候補個体が複数存在するため、耳標情報のみでは個体を一意に定めることができない。そのため、顔特徴量に基づく類似度比較により最終的な識別を行う。結果を表 5 と図 12, 図 13 に示す。

表 5 (2)における顔特徴量の類似度閾値と識別性能の関係

閾値設定	受理件数	正解数	精度	カバレッジ
threshold = 0.00	1,532	1,529	0.9980	1.0000
threshold = 0.50	1,232	1,232	1.0000	0.8042

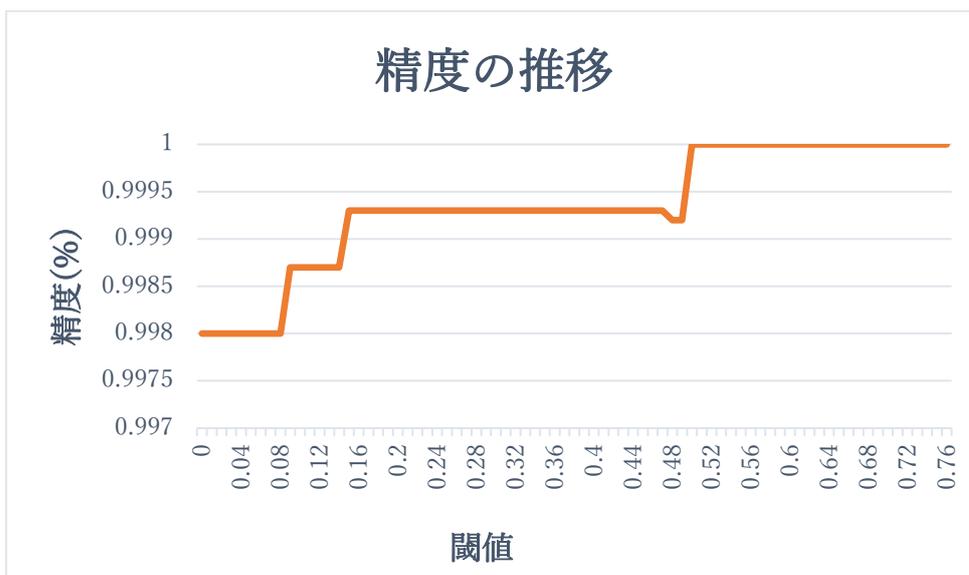


図 12 (2)における閾値ごとの精度の推移

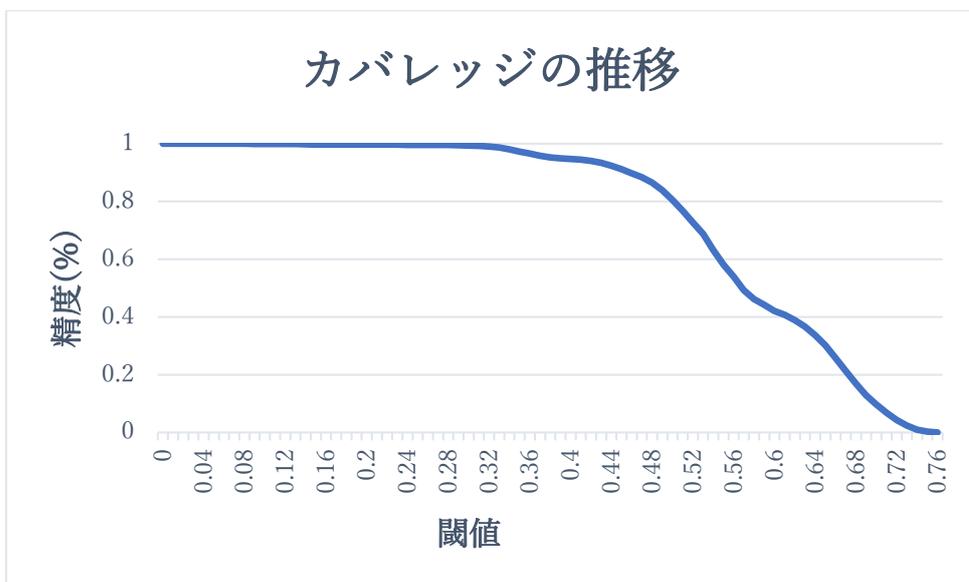


図 13 (2)における閾値ごとのカバレッジの推移

閾値を設けない場合 (threshold = 0.00) においても精度は 99.8% と高く、顔特徴量による識別は十分に実用的であることが確認された。一方、threshold = 0.50 を設定することで誤認識を完全に排除でき、精度は 100% に達するが、その分カバレッジは 0.8042 に低下する。

この結果から、本条件では厳密な閾値を設けなくても高い識別精度が維持されており、実運用においては閾値なしの設定でも十分に実用的であることが確認された。

(3) レーベンシュタイン距離 1・候補が一意に定まる場合

本条件では、耳標文字列が登録情報と完全一致ではなく、1文字分の差（レーベンシュタイン距離 1）を含むものの、候補個体自体は一意に定まるケースを対象とする。このような場合、耳標情報のみを用いた識別では誤認識が多く、識別精度の低下が生じやすい。

そこで本条件においても、顔特徴量のコサイン類似度に閾値を設けた際の識別性能の変化を評価した。結果を表 6 と図 14, 図 15 に示す。

表 6 (3)における顔特徴量の類似度閾値と識別性能の関係

閾値設定	受理件数	正解数	精度	カバレッジ
threshold = 0.00	4,581	2,958	0.6457	1.0000
threshold = 0.45	2,256	2,238	0.9920	0.4925
threshold = 0.56	1,096	1,096	1.0000	0.2392

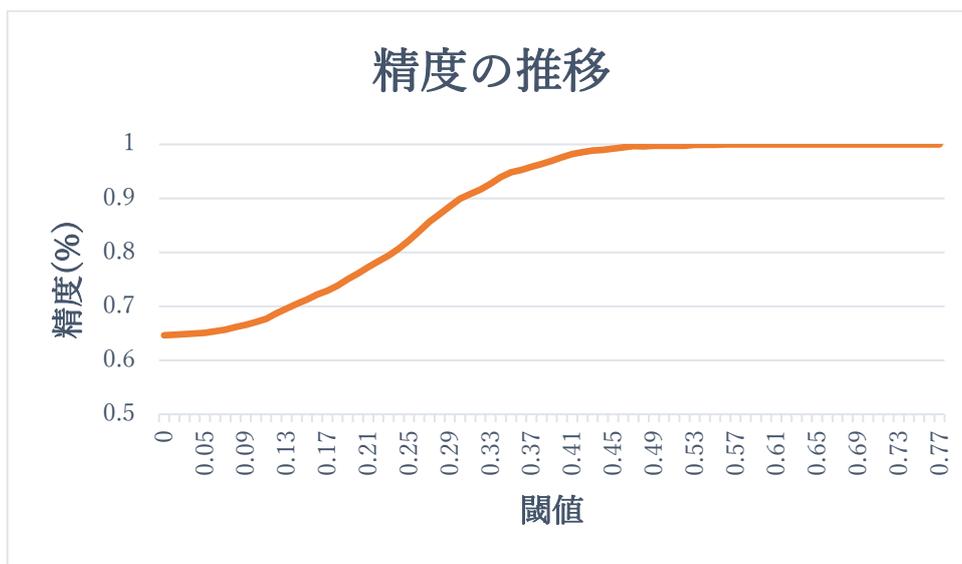


図 14 (3)における閾値ごとの精度の推移

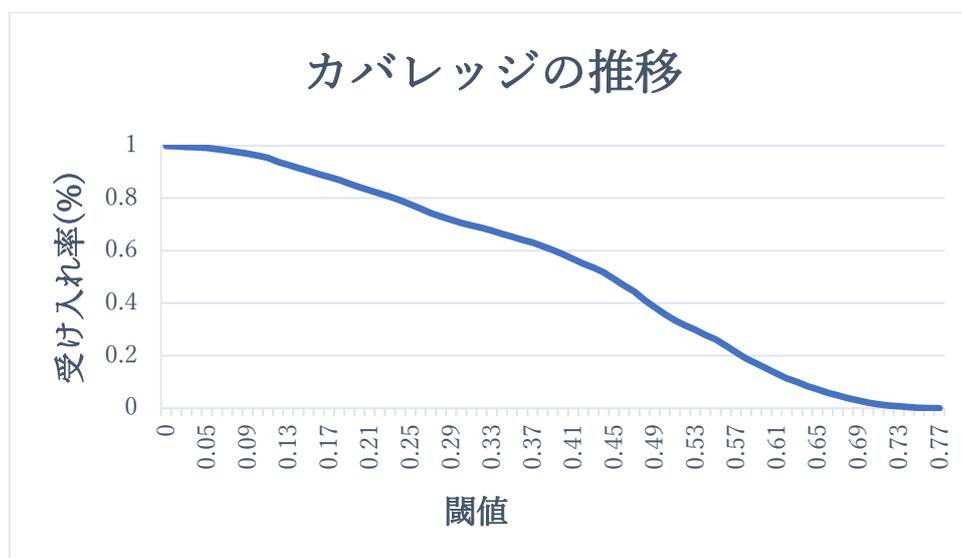


図 15 (3)における閾値ごとのカバレッジの推移

レーベンシュタイン距離が 1 であり、耳標情報のみを用いて候補個体が一意に定まる場合、顔特徴量による判定を行わない条件 (threshold = 0.00) では、識別精度は約 65% にとどまり、実用上十分とは言えない結果となった。これは、OCR による 1 桁の誤読であっても、誤った耳標番号が別個体に対応付けられる可能性が高いためである。

一方で、顔特徴量に基づくコサイン類似度判定を導入し、 $\text{threshold} = 0.45$ を設定した場合、精度は 99% を超え ($\text{accuracy} = 0.992$)、耳標情報のみでは不安定であった条件においても、高い信頼性を確保できることが確認された。このときカバレッジは 0.4925 であり、約半数のフレームに対して高精度な識別結果を受理できている。

さらに、 threshold を高く設定するほど誤認識はより抑制され、 $\text{threshold} = 0.56$ では精度 100% を達成できる。一方で、受理件数 (カバレッジ) は段階的に低下するため、本条件では、精度とカバレッジのトレードオフを考慮した閾値設定が重要となる。

以上より、レーベンシュタイン距離が 1 のケースでは、耳標情報のみを用いた識別の信頼性は低い一方で、顔特徴量を補助的に導入することにより、識別精度を大きく改善できることが示された。特に、本条件は顔特徴量の導入効果が顕著に現れるケースであり、本手法の有効性を示す重要な結果である。

(4) レーベンシュタイン距離が 1 で複数候補が存在する場合

本条件は、耳標文字列が 1 桁誤読されているうえに、複数の候補個体が存在する状況であり、本実験で想定する中で最も識別が困難なケースである。結果を表 7 と図 16、図 17 に示す。

表 7 (4)における顔特徴量の類似度閾値と識別性能の関係

閾値設定	受理件数	正解数	精度	カバレッジ
$\text{threshold} = 0.00$	4,560	3,103	0.6805	0.9998
$\text{threshold} = 0.46$	2,448	2,429	0.9922	0.5367
$\text{threshold} = 0.53$	1,684	1,684	1.0000	0.3692

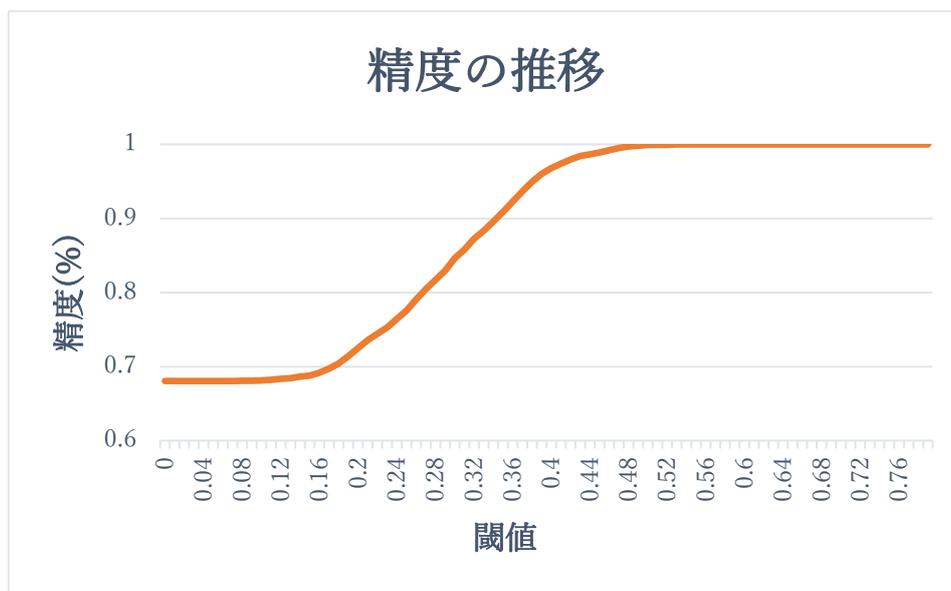


図 16 (4)における閾値ごとの精度の推移

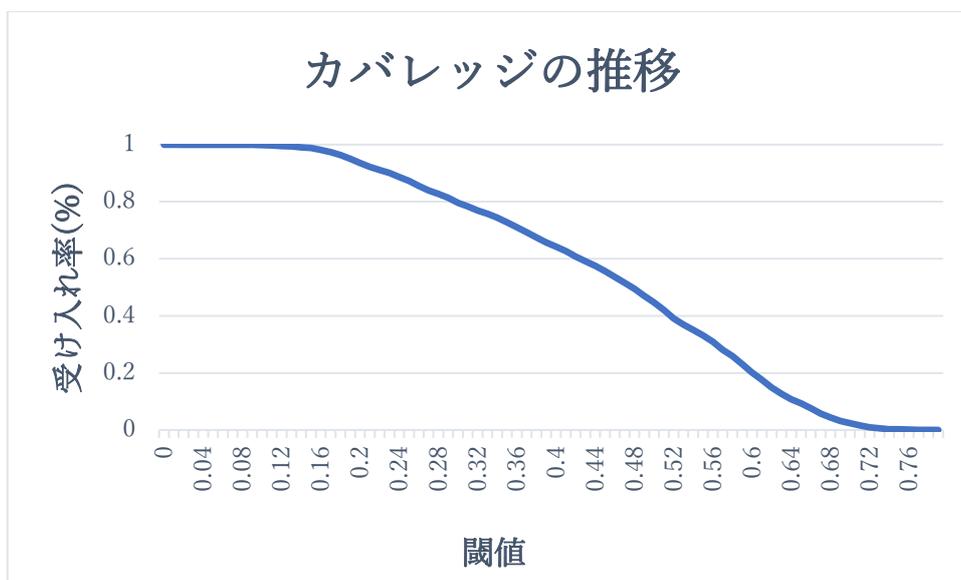


図 17 (4)における閾値ごとのカバレッジの推移

顔特徴量による判定を行わない場合 ($\text{threshold} = 0.00$) では、精度は約 68% にとどまり、耳標情報と候補の一意性がともに失われた状況では、識別の信頼性が大きく低下することが確認された。

一方で、顔特徴量に基づくコサイン類似度に閾値を導入することで、識別精度は大きく改善される。 $\text{threshold} = 0.46$ を設定した場合、精度は 99% を超え ($\text{accuracy} = 0.9922$)、約 54% のフレームに対して高精度な識別結果を得ることが可能となった。さらに、 $\text{threshold} = 0.53$ とすることで、受理件数は減少するものの、誤認識を完全に排除し、精度 100% を達成できることが確認された。

これらの結果から、レーベンシュタイン距離が 1 かつ複数候補が存在する場合には、耳標情報のみでは個体識別が極めて困難である一方、顔特徴量を用いた類似度判定を導入することで、識別精度を段階的に回復できることが示された。本条件は、4 つの評価条件の中でも顔特徴量の寄与が最も大きいケースであり、本手法における顔識別の有効性を強く裏付ける結果である。

以上の結果を踏まえると、本実験で設定した 4 つの識別条件それぞれに対して、顔特徴量のコサイン類似度に適切な閾値を設定することで、すべての条件において識別精度 99% 以上を達成可能であることが示された。

レーベンシュタイン距離が 0 で候補が一意に定まる場合や、距離 0 で複数候補が存在する場合には、閾値を設けない場合でも十分に高い精度が得られており、実運用においては高いカバレッジを維持した識別が可能である。一方、レーベンシュタイン距離が 1 のケースでは、耳標情報のみでは精度が大きく低下するが、顔特徴量に基づく類似度判定を導入し、条件ごとに適切な閾値を設定することで、精度を 99% 以上に回復できることが確認された。

特に、複数候補が存在する条件では、閾値を高く設定することで誤認識を段階的に排除でき、精度 100% に到達する設定も存在する。このとき、カバレッジとのトレードオフは生じるものの、高信頼な識別結果のみを採用する運用が可能となる。

以上より、本研究の第一の手法では、耳標情報による候補抽出を基盤としつつ、条件に応じて顔特徴量の閾値を切り替えることで、識別精度 99% 以上を維持しながら認証件数を段階的に拡張できることが示された。このような閾値の設計は、個体識別番号の類似性が高くなる大規模農場環境においても、高精度かつ柔軟な個体識別を実現する上で有効である。

5.4 トラッキングを利用した複数頭の牛個体識別実験

5.4.1 実験条件および評価対象映像

本節では、4章で提案したトラッキングを利用した複数頭の牛個体識別手法の有効性を検証するため、実環境で撮影された2種類の映像を用いて実験を行う。

一つ目の映像（映像 A）は、2025年11月20日に撮影されたものである。2頭の牛が同時に映り込む全615フレームから構成されている。

二つ目の映像（映像 B）は、2025年12月18日に撮影されたものであり、5頭の牛が映り込む全525フレームから構成されている。このうち奥に位置する1頭は顔および耳標の視認が困難であるため、識別対象から除外し、残りの4頭を評価対象とした。

顔特徴量の抽出には、5.1節で示した ArcFace モデルを用いる。なお、映像 A および映像 B に含まれる牛はいずれも、ArcFace の特徴抽出モデルの学習データには含まれていない個体である。

本実験で用いる耳標登録情報は、5.3節で用いたものに、新たに映像 A に含まれる2頭を加えた183頭分の情報から構成されている。また、顔登録情報についても、同じ183頭を対象として、使用する映像の撮影時期から約1年前に撮影されたの顔画像から抽出した特徴量をあらかじめ登録している。

5.4.2 ByteTrack と顔特徴量によるトラッキング

4.4節で述べた顔特徴量に基づくトラッキング ID 結合処理を適用した際のトラッキング結果を表8に示す。

表8 顔特徴量に基づくトラッキング ID の結合結果

映像	個体数	ByteTrack のみ	顔特徴量結合後
映像 A	2	14	2
映像 B	5	32	9

まず、映像 A (2 頭の牛が映り込む映像) に対して、ByteTrack のみを用いてトラッキングを行った場合、合計 14 個のトラッキング ID が生成された。これは、同一個体であっても遮蔽や姿勢変化によりトラッキング ID が分断されたためである。

これに対して、顔特徴量に基づくトラッキング ID 結合処理を適用した結果、トラッキング ID 数は 14 から 2 へと削減され、実際の個体数と一致する結果が得られた。

次に、映像 B (5 頭の牛が映り込む映像) に対して同様の処理を行った。ByteTrack のみを用いた場合には、合計 32 個のトラッキング ID が生成されたが、顔特徴量による結合処理を適用することで、トラッキング ID 数は 9 へと削減された。

なお、この 9 個のトラッキング ID のうち、3 個は識別対象外とした牛に対応するものであった。残る 6 個のトラッキング ID のうち、4 個は識別対象である 4 頭の牛それぞれに対応しており、正しく統合されていることが確認された。一方、残りの 2 個は識別対象である 4 頭のうち 2 頭に対応するトラックであるが、該当トラックでは顔特徴量が十分に取得できなかったフレームが多く、顔特徴量に基づくトラッキング ID の結合が行われなかったものである。

このように、顔特徴量を用いたトラッキング ID 結合により、ByteTrack のみでは分断されていたトラックを大幅に統合できることが確認された。これにより、時間方向に連続した情報が各トラックに集約され、後段の個体識別処理に利用可能な安定したトラッキング結果が得られた。

5.4.3 提案手法と単一フレームによる個体識別結果との比較

本節では、A と B の 2 つの映像に対して 3 章で示した単一フレームに基づく個体識別手法と、4 章で提案したトラッキングを利用した個体識別手法を用いて得られた結果を比較し、認証可能フレーム数の観点からその有効性を評価する。

比較にあたっては、YOLO により牛顔が検出されたフレーム総数を評価対象とし、そのうち個体識別番号を付与できたフレーム数を認証回数として集計した。その結果を表 9 に示す。

表9 複数個体の認証精度比較

トラッキング ID	認証可能な最大数	単一フレームによる処理	提案手法
A_1	482	223	482
A_2	388	162	388
B_1	421	292	359
B_3	346	41	346
B_4	309	50	288
B_9	372	334	372

まず A_1 および A_2 に注目すると、単一フレーム処理ではそれぞれ 223, 162 にとどまっているのに対し、提案手法ではいずれも認証可能な最大数と同じ 482, 388 を達成している。この結果から、フレーム間情報を統合する提案手法により、単一フレームでは認証できなかった個体を補完できていることが分かる。

次に B_3 では、単一フレーム処理による認証数が 41 と極めて少ない一方、提案手法では 346 と認証可能な最大数に到達しており、個々のフレームでの情報が不安定な場合でも、大きな性能向上が得られている。

一方で B_1 では、提案手法の結果は 359 と最大数 (421) には及ばないものの、単一フレーム処理 (292) と比較して大幅な改善が確認できる。

単一フレーム手法では、耳標が検出されないフレームや、OCR 結果から候補が得られないフレームでは認証が行えず、認証可能フレーム数が限定される傾向が見られた。一方、提案手法では、トラッキングにより時間方向に情報を統合することで、一部のフレームで耳標情報が得られない場合であっても、同一トラッキング ID に基づいて個体 ID を継続的に付与することが可能であった。その結果、提案手法は単一フレーム手法と比較して、認証可能フレーム数を大幅に増加させることが確認された。

6. おわりに

本研究では、耳標と牛顔画像を併用した牛個体識別手法を提案し、単一フレーム環境および複数頭環境における有効性を検証した。まず、耳標読み取りに GoogleOCR を用いた評価では、読み取り精度が 59.2% にとどまり、耳標情報のみでは安定した識別が困難であることを確認した。一方で、耳標情報により候補個体を抽出し、顔特徴量 (ArcFace) により候補を絞り込む構成により、耳標の誤読や欠損が生じる条件においても閾値の設定により識別精度を 99% 以上に向上させることができた。

さらに、複数頭が映り込む映像に対しては、ByteTrack を基盤としたトラッキングに顔特徴量によるトラック結合を導入し、時間方向の情報統合に基づく個体識別を行った。その結果、単一フレーム処理では認証が困難なケースにおいても、提案手法により認証可能フレーム数を大幅に増加できることを確認した。

今後の課題として、まず耳標 OCR の安定化が挙げられる。本研究では GoogleOCR をそのまま適用したが、撮影距離やブラー、照明条件の影響を受けやすく、文字が検出されないケースや部分的な誤認識が多く見られた。今後は、耳標領域に対する画質補正や姿勢正規化、複数フレームに基づく読み取り結果の統合などを導入することで、耳標情報の信頼性向上が期待される。

次に、牛顔特徴量モデルの頑健性向上も重要な課題である。顔特徴量は比較的安定して取得できる一方で、遮蔽や極端な姿勢変化、撮影距離の増大により特徴量が十分に得られない場合があった。これに対しては、牛顔に特化した学習データの拡充や、時系列情報を考慮した特徴量統合手法の導入が有効であると考えられる。

さらに、本研究では比較的短時間かつ限定された頭数の映像を対象としたが、実運用を想定した場合には、より長時間にわたる連続撮影や、多数の牛が同時に映り込む大規模環境への適用が求められる。そのような条件下において、トラッキングの安定性や計算コスト、識別精度の維持について検証を行うことが今後の重要な課題である。これらの課題に取り組むことで、実環境において実用可能な牛個体識別システムの実現につながると考えられる。

謝辞

本論文の作成にあたり、丁寧なご指導と多大なるご助言を賜りました椋木雅之教授に、心より感謝申し上げます。研究の進め方に関する助言に加え、実験やデータセットの調査においても多くのご支援をいただき、本研究を遂行する上で大きな助けとなりました。

また、研究活動を通じて助言し合い、切磋琢磨しながら研究を進めてきた椋木研究室の皆様にも深く感謝申し上げます。皆様との議論や意見交換は、本研究を発展させる上で貴重な経験となりました。

さらに、本研究で使用した牛映像データの収集に際し、牛の撮影にご協力いただきました株式会社デンサンの皆様にも心より御礼申し上げます。実環境でのデータ取得は、本研究の実証において不可欠なものでした。

加えて、耳標読み取りの評価において GoogleOCR を利用させていただきましたことに感謝申し上げます。高性能な OCR サービスを活用することで、本研究の検証を行うことができました。

ここに記して、関係者の皆様に深く感謝の意を表します。今後のご健勝とご発展を心よりお祈り申し上げます。

参考文献

- [1] 農林水産省, 「牛の個体識別のための情報の管理及び伝達に関する特別措置法 (牛トレーサビリテイ法)」, https://www.maff.go.jp/j/kokuji_tuti/tuti/t0000951.html
(参照日: 2026年1月)
- [2] Thi Thi Zin, Ju Haen Lee, and Hyeon-Cheol Kim. “Automated Cattle Identification Using Ear Tag Recognition,” *Sensors*, Vol. 20, No. 12, Article 3564, June 2020. doi: 10.3390/s20123564.
- [3] M. Smink, H. Liu, D. Dopfer, and Y. J. Lee. “Computer Vision on the Edge: Individual Cattle Identification in Real-Time with ReadMyCow System,” In *Proceedings of 2024 IEEE Winter Conference on Applications of Computer Vision (WACV 2024)*, pp. 7041–7050, 2024.
- [4] Yifu Zhang, Peize Sun, Yi Jiang, Dongdong Yu, Fucheng Weng, Zehuan Yuan, Ping Luo, Wenyu Liu, and Xinggang Wang. “ByteTrack: Multi-Object Tracking by Associating Every Detection Box,” In *Proceedings of the European Conference on Computer Vision (ECCV 2022)*, pp. 1–17, 2022.
- [5] Kimani, et al., “Cattle Identification Using Muzzle Images and Deep Learning Techniques,” arXiv preprint arXiv:2311.08148, 2023. <https://arxiv.org/pdf/2311.08148>
- [6] Z. Weng, “Cattle face recognition based on a Two-Branch convolutional neural network (TB-CNN),” *Computers and Electronics in Agriculture*, Vol. 195, 2022. doi:10.1016/j.compag.2022.106871
- [7] Y. Li, X. Zhang, Y. Chen, and H. Wang. “Multi-Modal Cattle Identification Using Ear Tag, Face Image and Muzzle Pattern,” *Sensors*, Vol. 21, No. 18, Article 6061, 2021.
- [8] YOLO11: <https://docs.ultralytics.com/models/yolo11/>
- [9] GoogleOCR: <https://docs.cloud.google.com/vision/docs/ocr?hl=ja>
- [10] Jiankang Deng, Jia Guo, Jing Yang, Niannan Xue, Irene Kotsia, and Stefanos Zafeiriou. “ArcFace: Additive Angular Margin Loss for Deep Face Recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 44, No. 10, pp. 5962–5979, October 2022.