

令和元年度卒業論文

一般物体認識による AR を用いた
牛に対する情報の重畳表示

宮崎大学工学部 情報システム工学科

兒玉 光平

指導教員 椋木 雅之

目次

1. はじめに.....	1
2. AR の従来研究.....	2
2.1. AR.....	2
2.2. AR の分類.....	2
2.3. 牛に対する AR の従来研究.....	3
3. 一般物体認識による AR.....	4
3.1. 牛に対する AR.....	4
3.2. 一般物体認識と特定物体認識.....	5
3.3. 本研究の位置づけ.....	6
4. システム構成.....	7
4.1. 一般物体認識による牛の検出.....	8
4.2. 検出した矩形領域と 3D 牛モデルとの位置合わせ.....	10
4.3. 牛に対する情報の重畳表示.....	12
5. 実験.....	12
5.1. 設定.....	12
5.2. 照明環境が良好な状態での実験.....	14
5.3. 暗い環境を想定した実験.....	19
5.4. オクルージョンを想定した実験.....	22
5.5. 考察.....	25
6. おわりに.....	26
謝辞.....	27
参考文献.....	28

1. はじめに

近年、畜産農家の間で、牛白血病など感染症の病原体を持つ牛の管理が問題となっている。感染症の種類によっては、病原体を持つ個体を一目で見分けるのが困難であり、管理が難しいケースもしばしば存在する。そのような場合にAR(Augmented Reality、拡張現実)を用いて直接的な情報提示を行い、病原体を持つ個体を一目で直観的に見分けることができれば、管理作業の効率化が図れる。

そこで本研究では、牛に対して一般物体認識を用いることで画像中から牛を検出し、重畳する情報の位置合わせを行うことで、特定のマーカを必要としないARシステムを提案する。さらに提案システムをスマートフォンに実装し、評価する。

2. AR の従来研究

2.1. AR

AR とは、人が見たり聞いたりする現実空間で発生した事象や存在する環境に対して、コンピュータにより何かしらの情報を付与することで現実を拡張する技術を指す。田上ら[1]によると、「拡張現実」とは、「現実」を認識している人間の五感を「拡張」する技術であるとされている。

2.2. AR の分類

AR は、情報の位置合わせ等、現実空間と仮想的な情報の整合性をとることを技術的な課題としている。AR は、位置情報の取得方法により、ロケーションベース AR とビジョンベース AR に分類できる。ロケーションベース AR は、GPS 等を用いて取得した位置情報に紐づけてユーザに情報を提示する。ビジョンベース AR は、カメラ等を用いて取得した画像から情報を解析してユーザに情報を提示する。さらに、ビジョンベース AR は、マーカー型 AR とマーカーレス型 AR に分類できる。マーカー型 AR は、既知のマーカーを現実空間に設置し、検出したマーカーを基準に提示する情報の位置合わせを行う。マーカーレス型 AR は、マーカーを利用せずに自然特徴点を用いたパターンマッチングや自己位置推定により提示する情報の位置合わせを行う。

2.3. 牛に対する AR の従来研究

牛に対する AR を取り扱った従来研究として、乳牛の監視をサポートする AR アプリケーションを設計および実装する研究[2]がある。この研究では、スマートフォンのカメラで取得した映像から牛の耳についているタグを検出することでそこに牛がいることを認識し、マーカーに応じた情報を画面に表示することでユーザに情報を提示している。この研究の実験結果によると、耳のタグを検出できないことが多かった。汚れやオクルージョンの発生、あるいはマーカーとなるタグをあまり大きく写せないためである。

3. 一般物体認識による AR

3.1. 牛に対する AR

本研究では、AR 技術を用いて牛の情報を直接的にユーザへ提示することを目標とする。これを達成するのに必要な処理は、カメラで撮影した画像中から牛を検出する処理と、検出した位置に基づいて牛に情報を重畳表示する処理である。個体識別や個体の情報を保有しているサーバとの通信等、情報の取得に関する部分は本研究では扱わない。

従来研究[2]では、牛の位置を検出するためにマーカーを用いた場合、マーカーの汚れや牛の姿勢変化によりマーカーの検出が困難であった。そのため、マーカー型のシステムでは不適切であると考えられる。一方、マーカーレス型のシステムで牛を検出する場合、牛の個体差や姿勢変化があるため、自然特徴点を用いた特定物体認識を利用するのは難しい。そこで、本研究では一般物体認識により牛を検出する。

また、重畳する情報は三次元的であることが望ましいが、一般物体認識により得られる情報は二次元的なものである。そこで、物体を検出した際の二次元的情報を三次元的な牛モデルと位置合わせする。そして、三次元的な牛モデルに対して相対的に情報の表示位置を決めることで、画像中の牛に対して位置的な整合性をとり、情報の重畳表示を行う。

本研究では、以上のような手法で検出と位置合わせを行う『一般物体認識による AR』を提案する。

3.2. 一般物体認識と特定物体認識

柳井[3]によると、一般物体認識とは、制約のない実世界のシーンの画像に対して、計算機がその中に含まれる物体もしくはシーンを一般的な名称で認識することを指す。一方、特定物体認識とは、まったく同じ形状の物体に対する認識技術である。特定物体認識は、人の顔のように形状がある程度決まったものであれば、工業製品のように完全に形状が定まっていなくても応用することができる。

物体検出が AR に利用される場合、多くはマーカーや人の顔を対象としている。これらは主に自然特徴点を用いた特定物体認識が利用されている。しかし、牛の全身を対象にする場合、品種による模様や毛色の違いや姿勢の取り方によって見た目が大きく変わり、人間の顔ほど既知ではないため、特定物体認識では不適切であると考えられる。以上より、本研究においては一般物体認識を用いて牛を検出することが適切であると考えられる。

3.3. 本研究の位置づけ

本研究の内容はマーカーレス型 AR に分類されるが、本研究におけるマーカーレス型 AR は多く知られているマーカーレス型 AR のシステムとは異なる。自己位置推定を行わずに、マーカー型のように検出した対象との相対的な位置関係を用いて重畳する情報の位置合わせを行うため、本研究の分類はマーカー型とマーカーレス型の中間に位置するものと考えられる。本研究では、マーカーレス型 AR をさらに二つに細分化し、自己位置推定を行わずに物体認識による相対的な位置情報で位置合わせ処理を行うマーカーレス型 AR システムを「一般物体認識による AR」、自己位置推定を行い三次元的な情報を用いて位置合わせ処理を行うマーカーレス型 AR システムを「自然特徴点による AR」とする。本研究における AR の分類を図 1 に示す。

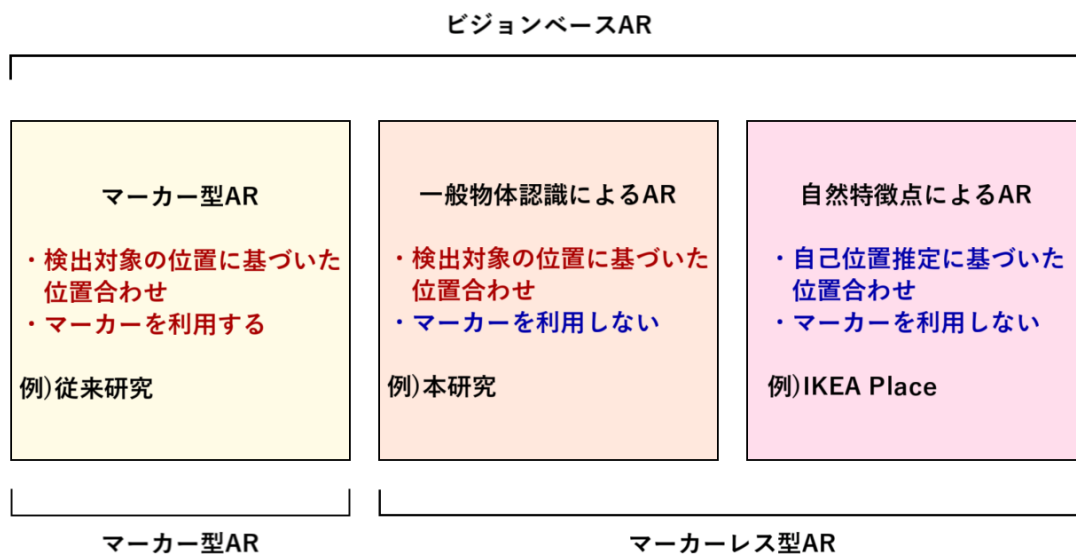


図 1 本研究における AR の分類

4. システム構成

本研究では、牛に対して情報の重畳表示を行う AR システムをスマートフォン上に実装する。

システムの実装には、Unity および Unity で利用できる画像処理アセット「OpenCV for Unity[4]」を利用する。システムは、OpenCV for Unity 内に含まれるサンプルプログラム、「YoloObjectDetectionWebCamTextureExample」を基礎として、改変を行う。システムの処理の流れを以下に示す。

- ① 一般物体認識による牛の検出
- ② 検出した矩形領域と 3D 牛モデルとの位置合わせ
- ③ 3D 牛モデルに対する情報の重畳表示

正しく牛を検出し、情報を重畳表示できた場合の例を図 2 に示す。

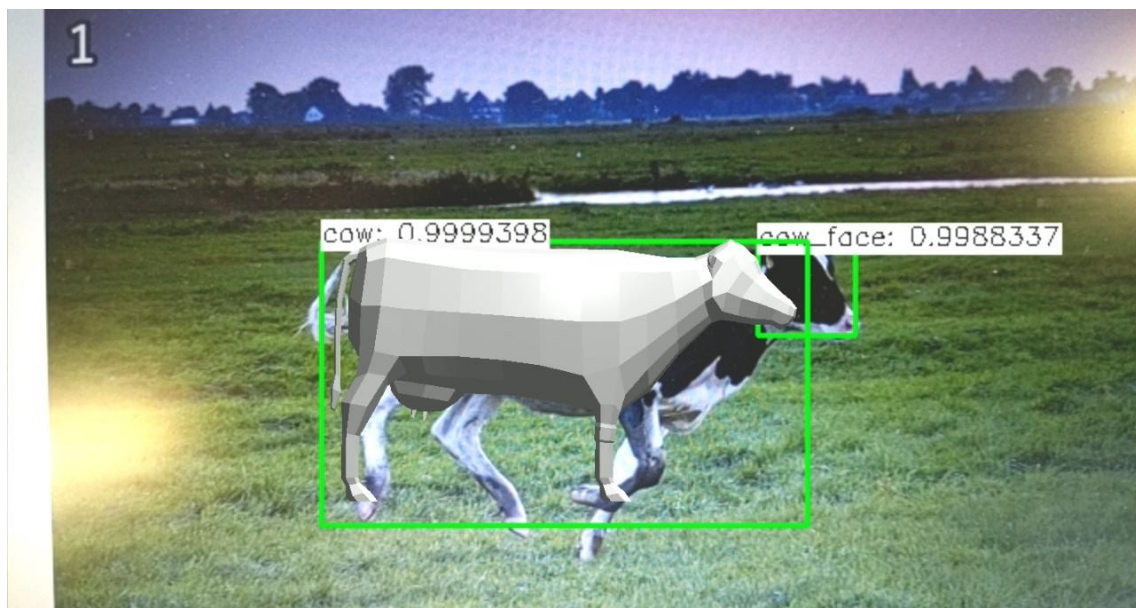


図 2 システム実行時のスマートフォンの画面

4.1. 一般物体認識による牛の検出

まず、一般物体認識を用いて画像中から牛を検出する。検出の際は、牛の位置を特定するために「牛全体」の検出と、牛の方向を特定するために「牛の顔」の検出を行う。本研究では、以上の二つを検出するために、Joseph Redmon[5]による YOLO という深層学習を利用した一般物体認識アルゴリズムを用いる。YOLO の特徴として高速であるという点が挙げられる。AR においては物体認識をリアルタイムで行う必要があるため、高速な YOLO を採用した。YOLO では、検出した物体は矩形領域で表現される。

「牛全体」と「牛の顔」を検出するために、YOLO にこれらの画像を学習させる。学習は YOLO の公式 Web サイト[6]を参考に行った。学習のデータセットとして、2 種類に大別できる 5921 枚の画像を用意した。1 つは、実際の農場の様子を撮影した動画から、牛の写っているフレームを 1 秒あたり 2~5 フレーム程度の頻度で抽出した 3921 枚の画像である。もう 1 つは、Web 上から取得した乳牛や肉牛の画像 100 枚を、コントラストの強調および低減、4 段階の輝度変更、平滑化、ガウシアンノイズの付与、インパルスノイズの付与により、1000 枚にデータ拡張したものを、左右反転によりさらに 2000 枚にデータ拡張したものである。データ拡張は、yota[7]による記事を参考に行った。これらの画像に対して、学習のために、「牛全体」および「牛の顔」に相当する部分を矩形領域

として手作業で与えた。一例を図 3 に示す。ネットワーク構造としては、制約のある環境に利用可能なより軽量なモデルである yolov3-tiny を使用する。また、ImageNet による事前学習が行われた重みファイル[8]も併せて使用する。学習の際のバッチサイズは 64、学習回数は 3 万回行った。

学習済みのモデルに牛画像を与えると、検出された「牛全体」および「牛の顔」の矩形領域が出力される。



図 3 手作業による矩形領域の選択の一例

4.2. 検出した矩形領域と 3D 牛モデルとの位置合わせ

検出した牛の二次元的な位置と、重畳表示する三次元的なデータ間の位置的な整合性をとるために、検出した矩形領域と 3D 牛モデルの位置合わせを行う。牛に対して重畳表示する情報の基準となる仮想的なモデルとして、3D 牛モデル [9](図 4)を用いる。

検出した「牛全体」の矩形領域に整合するように、3D 牛モデルの位置を決定する。位置合わせを行うために必要な情報は、3D 牛モデルの拡大率、三次元重心座標、回転角である。拡大率は、矩形領域の縦横の大きさに整合するように 3D 牛モデルの体高および体長と幅を決定する。三次元重心座標のうち X 座標と Y 座標は、矩形領域の重心座標とする。3D 牛モデルの奥行きは拡大率の変動により表現されるため、Z 座標は不変の固定値とする。回転角は、3D 牛モデルが右を向くか、左を向くかのみを決定する。3D 牛モデルの向きは、「牛全体」の検出した矩形領域の重心座標に対して、最も距離の近い「牛の顔」の検出領域の重心座標が、「牛全体」よりも左右のどちら側にあるかで決定する。「牛全体」に対する「牛の顔」の位置による 3D 牛モデルの向きの変化を図 5 に示す。

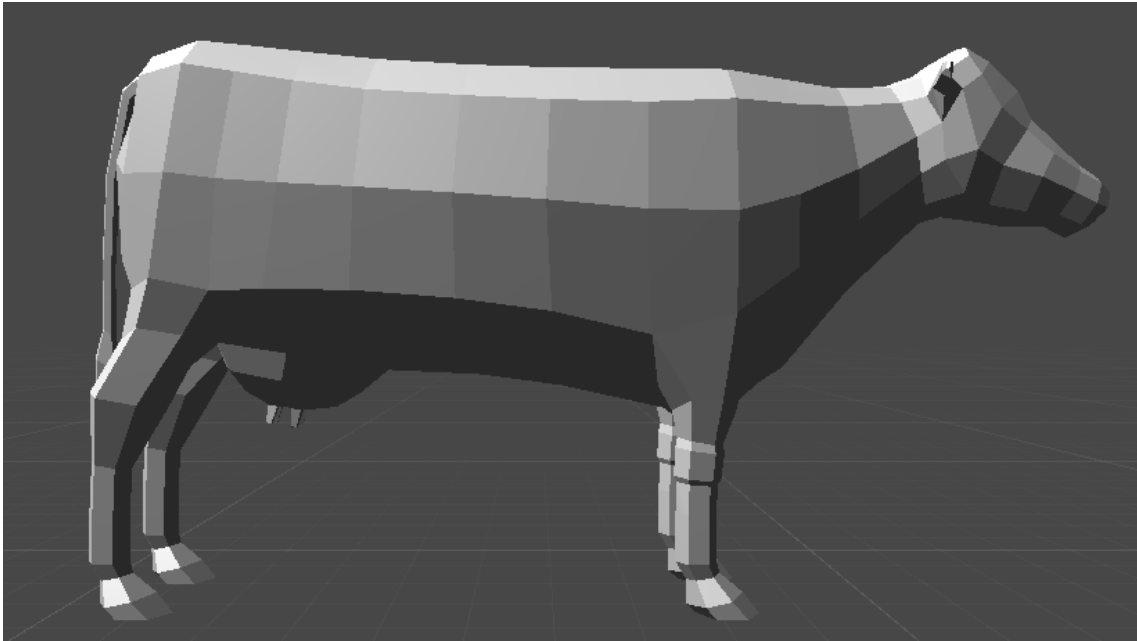


図 4 3D 牛モデル

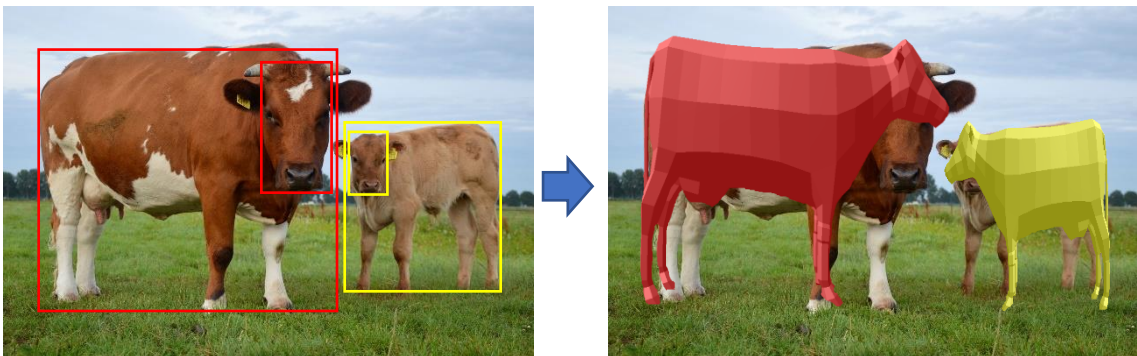


図 5 顔の向きによる 3D 牛モデルの向きの変化

4.3. 牛に対する情報の重畳表示

3D 牛モデルに情報を付与することで、位置合わせした画面中の牛に情報を重畳表示する。本研究では、位置合わせの結果を確認するために、3D 牛モデルそのものを重畳表示する。

5. 実験

本実験の目的は、カメラに映った画像中から牛を検出することと、検出した位置に基づいて牛に情報を重畳表示することが達成できているかを評価することである。評価は「牛全体」の検出、「牛の顔」の検出、情報の重畳表示の3点に関して 0～3 からなる 4 段階の主観的評価によって著者本人が行う。評価基準は、3 は「良好」、2 は「おおむね良好だが部分的な問題あり」、1 は「一部良好だが動作に問題あり」、0 は「動作しない」とする。

5.1. 設定

実験はノートパソコンのディスプレイに映した牛の画像を、システムを実装したスマートフォンのカメラで映して行う。実験に利用した 8 枚の画像を図 6 に示す。1～4 が学習に利用した画像から抜粋したもの、5～8 が学習に利用していない画像である。実際の運用が想定される牧場や牛舎の環境では、光の当たり

方がまちまちだったり、柵によるオクルージョンが発生したりするため、イレギュラーな環境での実験も行う。実験の際は、スマートフォンの画面に牛の全体が画面いっぱいよりも少し小さく収まるように映す。



図 6 実験に利用する画像

5.2. 照明環境が良好な状態での実験

照明環境が良好な状態での実験は、図 7 のような環境で行う。ノートパソコンの画面の明るさを最大に設定し、出来る限り画面中の牛を検出しやすい状態にした。評価結果を表 1 に示す。

重畳表示の精度は検出の精度に影響されるため、検出よりも重畳表示の評価が同じか低くなる傾向が見られる。画像 1 では、牛の体の一部を「牛の顔」として検出したが、重畳表示には影響しなかった(図 8)。画像 2 では、本来の牛よりも小さな矩形領域を検出した(図 9)。画像 3 では、2 頭の牛が映っているが、「牛全体」を検出したのは 1 頭のみだった(図 10)。画像 4 では、画像中の牛の頭が下がっているが、3D 牛モデルの頭の位置は固定されおり、頭の位置がずれるため、完璧な重畳表示はできなかった(図 11)。画像 6 は、「牛全体」に関しては全く検出することなく、「牛の顔」のみを検出したため、重畳表示をすることはできなかった(図 12)。画像 8 は、「牛の顔」を検出しにくいという問題と、検出領域が矩形であるのに対して、画像中の牛が斜めに映っているため、拡大率が整合しないという問題が発生した(図 13)。



図 7 実験環境(照明環境が良好な状態での実験)

表 1 照明環境が良好な状態での実験

画像番号	1	2	3	4	5	6	7	8
検出評価(牛全体)	3	1	2	3	3	0	3	3
検出評価(牛の顔)	2	2	3	3	3	3	3	1
重畳評価	3	1	2	2	3	0	3	1

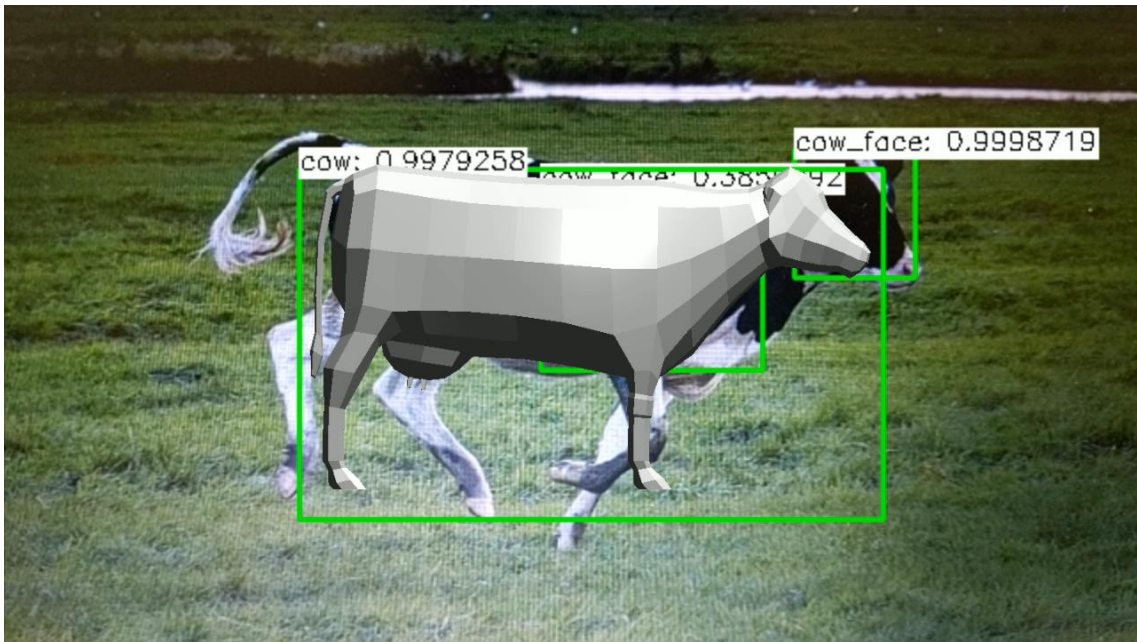


図 8 実験時のスマートフォン画面(照明の状態が良好な状態での実験,画像 1)

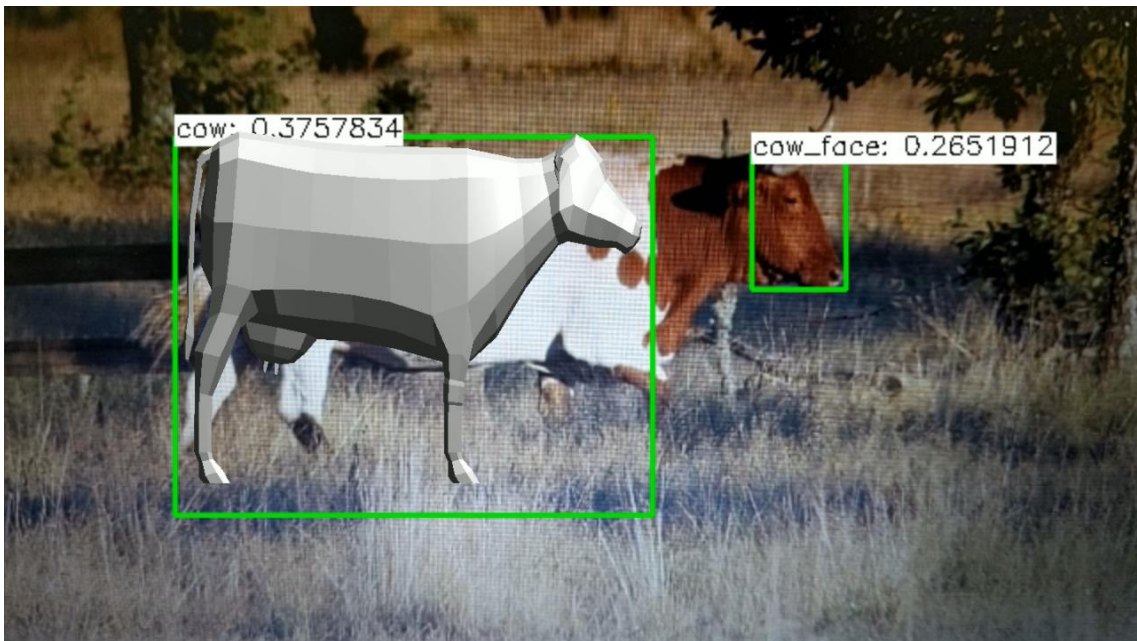


図 9 実験時のスマートフォン画面(照明の状態が良好な状態での実験,画像 2)

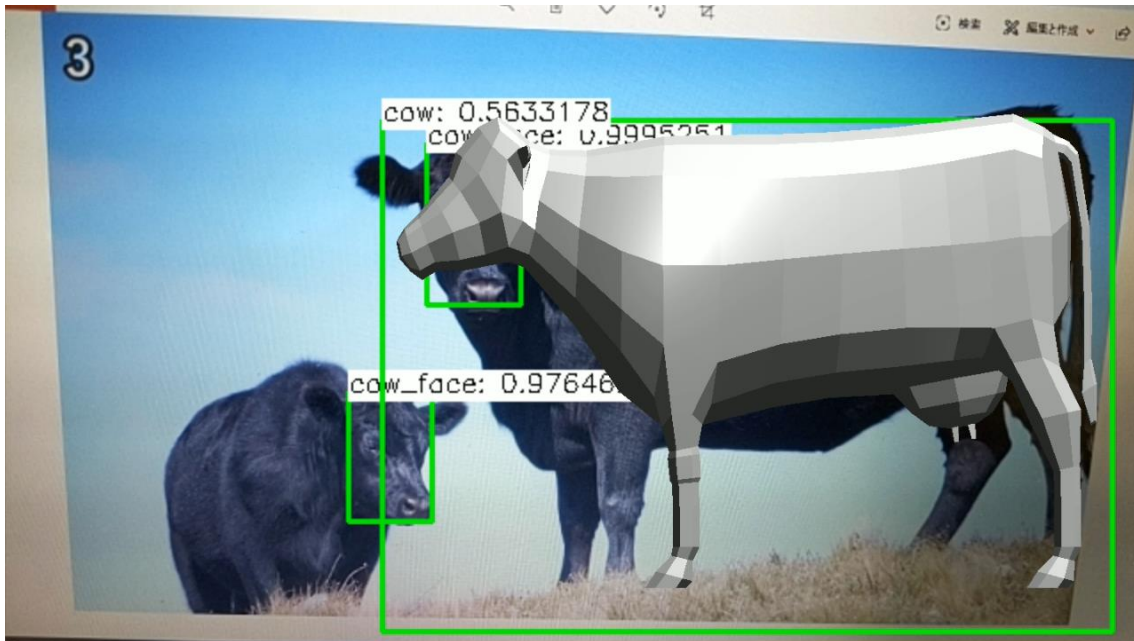


図 10 実験時のスマートフォン画面(照明の状態が良好な状態での実験,画像 3)



図 11 実験時のスマートフォン画面(照明の状態が良好な状態での実験,画像 4)

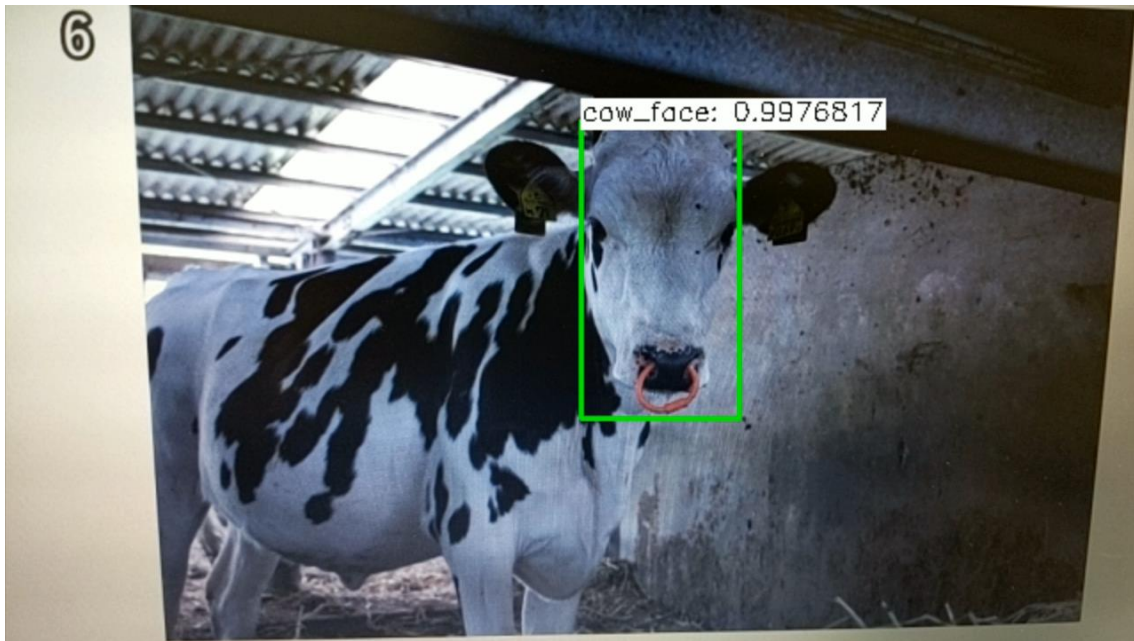


図 12 実験時のスマートフォン画面(照明の状態が良好な状態での実験,画像 6)



図 13 実験時のスマートフォン画面(照明の状態が良好な状態での実験,画像 8)

5.3. 暗い環境を想定した実験

本実験では、照明が暗く、対象が検出しづらい環境を想定している。実験した際のノートパソコンの見え方を図 14 に示す。ノートパソコンの画面の明るさを下げることで画面を見づらくしている。評価結果を表 2 に示す。

照明環境が良好な状態の実験と同じく、基本的には重畳表示の精度は検出の精度に影響されるため、検出よりも重畳表示の評価が同じか低くなる傾向が見られる。全体的に照明環境が良好な状態よりも検出の精度が悪くなった(図 15)。暗い環境に対しては、前処理により画像を一定の明るさにする等、対策が必要である。しかし、画像 6 に関しては、良好な状態よりも良い評価を得る結果となった(図 16)。

また、ノートパソコンの画面に近づくと、オートゲインコントロールの影響と思われる意図しない輝度の調整が行われた。検出する際の画像にも影響しているのか、画面に映る際にのみ調整されているのかは不明であるため、検討の余地が残った。



図 14 実験環境(暗い環境を想定した実験)

表 2 評価(暗い環境を想定した実験)

画像番号	1	2	3	4	5	6	7	8
検出評価(牛全体)	0	0	2	1	2	1	3	3
検出評価(牛の顔)	0	0	0	0	2	2	0	0
重畳評価	0	0	1	0	2	1	1	1



図 15 実験時のスマートフォン画面(暗い環境を想定した実験,画像 1)

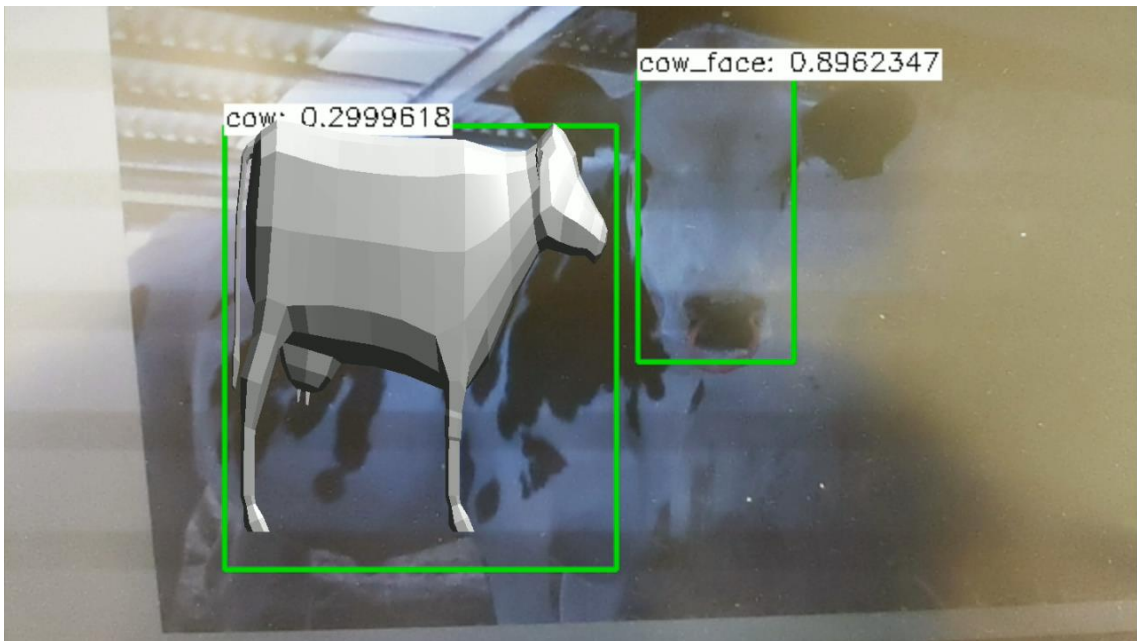


図 16 実験時のスマートフォン画面(暗い環境を想定した実験,画像 6)

5.4. オクルージョンを想定した実験

本実験では、牛を包囲する柵やえさを入れる箱などにより、オクルージョンが発生することを想定している。画像に柵を模した灰色の線を 5 本被せて実験を行う。実験に使用した加工後の画像の一例を図 17 に示す。評価結果を表 3 に示す。

他の実験と同じく、基本的には重畳表示の精度は検出の精度に影響されるため、検出よりも重畳表示の評価が同じか低くなる傾向が見られる。通常の状態よりも検出の精度が悪くなったが、暗い環境を想定した実験よりは悪くなかった。いくつかの画像に関して、本来の牛の領域よりも小さい領域として検出された画像があった(図 18)。



図 17 オクルージョンを想定した実験に用いた画像の例

表 3 評価(オクルージョンを想定した実験)

画像番号	1	2	3	4	5	6	7	8
検出評価(牛全体)	1	0	1	0	1	0	1	1
検出評価(牛の顔)	3	2	2	1	3	2	1	0
重畳評価	1	0	1	0	1	0	0	0

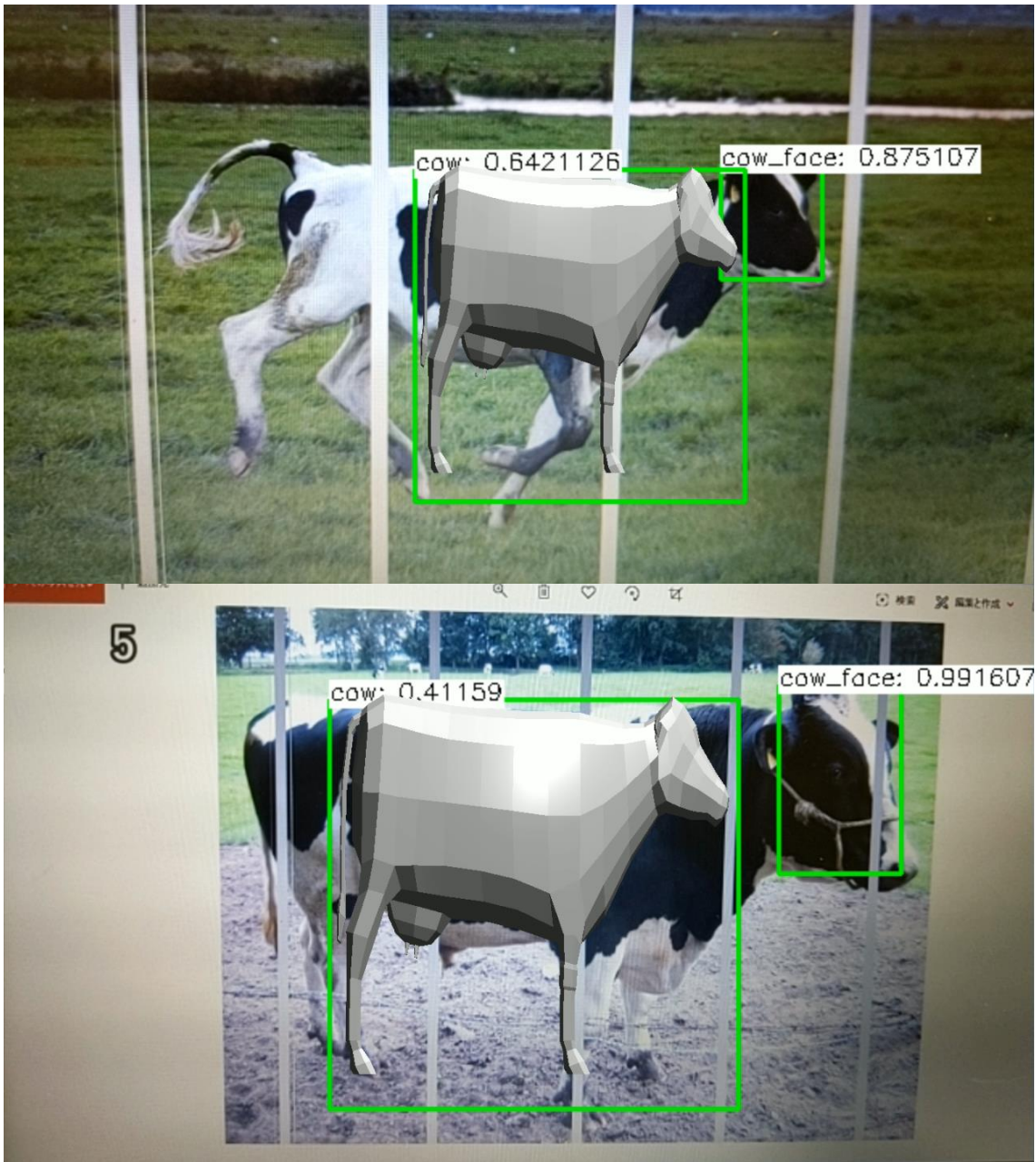


図 18 実験時のスマートフォン画面(オクルージョン想定した実験,画像 1,画像 5)

5.5. 考察

3つの実験を通して検出の精度が悪かった画像6は、牛の全体が画像内に収まりきっておらず、牛であると認識しなかったため精度が悪かったと考えられる。これは、学習の際に牛の足を含めずに胴体部分のみを「牛全体」の代わりに学習させることにより改善できると考えられるが、これにより全体的な検出の精度が低下する可能性もある。

想定外だった結果としては、暗い環境の方が良い評価を得た画像があったことである。また、学習に利用した画像と学習に利用していない画像の間に評価の差が生まれなかった点に関しても想定外だった。

画面外の牛とは無関係の部分(黒い画面の縁など)を「牛全体」や「牛の顔」と誤認識することがまれにあったため、誤認識の発生条件という点に関しても検討の余地が残った。

6. おわりに

本研究では、マーカーレス型 AR の一種として、一般物体認識による AR を提案した。また、この考え方に基づいて、牛に対する情報の重畳表示を行う AR システムをスマートフォン上に実装した。実装したシステムは、牛をリアルタイムで検出し、情報を重畳表示することができた。また、学習モデルを変更することで、他の物体の検出にも対応できるため、副産物的な貢献として、これまでよりもマーカーレス型 AR システムのスマートフォンへの実装が容易になる可能性があることも挙げられる。

今後の課題として、検出する前に画像に対して前処理を行うなどして、暗い環境でも精度を保つことが挙げられる。また、本稿では実験の際に画面に映した牛の画像を用いたが、実際の牛での実験も必要と考える。

謝辞

本研究を行うにあたり、ご指導・ご協力いただいた椋木雅之教授、高塚佳代子技術職員、関口敏准教授、小林侑雄准教授に、深く感謝申し上げます。

指導教員である椋木雅之教授には、本研究を通して研究の進め方や論文の書き方など、長期にわたり様々なことを丁寧かつ熱心にご指導いただきました。高塚佳代子技術職員には、ゼミや研究室内発表の際に沢山の助言をいただきました。関口敏准教授には、本研究の足掛かりとなる牛の感染症等の問題提起に関して助言をいただきました。小林侑雄准教授には、学習のデータセットとして利用するデータの収集にご協力いただきました。

ご指導・ご協力いただいた皆様には、深く感謝を申し上げるとともに、皆様のご健勝とますますのご活躍をお祈り申し上げます。

参考文献

- [1] 田上慎 他, 「AR (拡張現実) は, 人間が手にした新たな未来: AR の変遷と展望」, 情報管理 2016, vol.59, no.8, pp.526-534
- [2] Janina RUDOWICZ-NAWROCKA 他, 「APPLICATION OF AUGMENTED REALITY IN DAIRY CATTLE MONITORING」, Journal of Research and Applications in Agricultural Engineering 2018, vol.63, pp.181-183
- [3] 柳井啓司, 「一般物体認識における機械学習の利用」, 信学技報 IBISML 2010, vol.110, no.76, pp.103-112
- [4] unity Asset Store, 「OpenCV for Unity」, <https://assetstore.unity.com/packages/tools/integration/opencv-for-unity-21088>
- [5] Joseph Redmon 他, 「You Only Look Once: Unified, Real-Time Object Detection」, 2016, arXiv:1506.02640v5
- [6] Joseph Redmon, 「YOLO: Real-Time Object Detection」, <https://pjreddie.com/darknet/yolo/>
- [7] yota, 「機械学習のデータセット画像枚数を増やす方法」, <https://qiita.com/bohemian916/items/9630661cd5292240f8c7>

[8] Joseph Redmon, 「darknet53」 ,

<https://pjreddie.com/media/files/darknet53.conv.74>

[9]Poly by Google, 「Cow」 , <https://poly.google.com/view/0OToIgkcVM7>