検出対象領域の絞込みによる広角映像中の鳥追跡

リュウ テイイ[1,2,a)]　川西 康友[2,1,b)]　駒水 孝裕[1,c)]　井手 一郎[1,d)]

概要

本研究では，固定された広角カメラ映像中の鳥追跡に取り組む．研究の目的は映像全体の中の鳥の完全な軌跡を得ることである．しかし，広角映像中の鳥は小さいため，検出と対応付けが難しい．本研究はこの 2 点に対して解決を試みる．検出については，検出処理を候補領域の絞込みを導入し，前フレームでの鳥の出現領域の周辺に対して重点的に検出処理をする．対応付けについては，対象領域の重なりを用いず，中心点距離のみを用いて，鳥が同じ個体であるかどうかを判定し，鳥の大きな移動に対するロバストな追跡を実現した．NUBird2022 データセットを用いた実験により、提案手法の有効性を検証し，MOTA と IDF1 の明らかな改善を示した．

1. はじめに

鳥の移動行動は，採食や繁殖といった重要な生態学的な活動の理解において不可欠である．そのため，鳥の行動を詳細に追跡することは，生態学や行動学などの研究分野において重要な課題となっている．近年，映像解析とコンピュータビジョンの進歩により，カメラを用いた動物の追跡技術が大きく進歩した．本研究では，特に固定位置の広角カメラを使用した鳥の追跡に焦点を当てる．広角カメラは，1 台のカメラで広範な領域を観測できるため，広い範囲での鳥の活動を捉えるのに適している．しかしながら，広角映像中の鳥は小さく撮影されるため，背景と区別できる特徴の抽出と，フレーム間の同一鳥の判定が難しくなる．

追跡の基本的なアプローチは tracking-by-detection という，検出器で画像から物体を検出し，追跡器で検出結果を前の軌跡と対応付けする方法が近年主流となっている [3]．物体検出は物体特徴と背景特徴の違いを学習し，物体特徴を対応する部分を取り出す．対応付けは物体間の類似度による判断で，近年よく使われている類似度は，検出枠の重なりの割合（IoU）である．今回の対象は広角映像中の鳥であり，小物体追跡のタスクに属する．このタスクの特徴として，画像サイズと，画像中の物体サイズに大きな差があることである．そのため，普通の検出器には物体と背景の区別が困難である．このような小物体の特徴に対し，画像を小領域（検出候補領域）に分割し，各小領域を拡大して検出処理をする SAHI [1] という手法が提案されている．

しかし，SAHI による検出や，IoU による追跡には，次のような問題点が存在する．SAHI による小物体の検出は，画像全体に対して一様に検出候補領域をサンプリングし，切り出した画像を拡大して検出するが，そのうち実際に鳥を含む検出候補領域はごく一部であり，多くの領域で無駄な計算が行われる．一方，IoU による対応付けは 2 つの検出枠の共通部分を和集合で割った値で同一 ID の判定をするが，鳥は小さく，また動きが大きいため，隣接フレームであっても検出枠が重ならないことが多く（IoU が 0），対応付けに失敗しやすい．

この 2 つの問題点に対し，検出については，前フレームの鳥のいた場所の周辺を重点的にサンプリングする Adaptive SAHI という手法を、追跡については 2 つの領域の中心点の距離による対応付けの方法を提案する．

本研究の貢献は以下の通りである．

- Adaptive SAHI という領域の絞込みを使用した検出法を提案し，通常の SAHI よりも検出精度と速度の面で優れた結果を得た．
- 二つの領域の中心点の距離に基づく対応付けを利用することで，鳥の位置の変動に対してロバストな追跡が可能であることを示した．

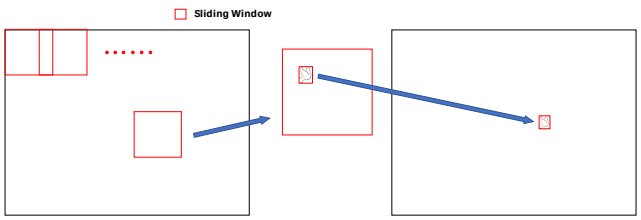

図 1　SAHI による小物体検出

1　名古屋大学
2　理化学研究所
a)　tingweiryu@outlook.com
b)　yasutomo.kawanishi@riken.jp
c)　taka-coma@acm.org
d)　ide@i.nagoya-u.ac.jp

2. 関連研究

2.1 小物体検出

物体検出器は，画像中の物体の位置とクラスを特定するためのアルゴリズムである．一般的な手順はまず入力画像に対して物体候補領域を生成する．次は物体候補領域の特徴量を抽出し，それを用いて物体の存在とクラスを判定する．しかし，小物体は画像内でわずかなピクセル数しか占めておらず，低い解像度では情報が制限される．そこで小さい物体の検出に特化した SAHI が提案された [1]，その具体的な流れは図 1 のように，重複を許して Sliding Window で画像全体をスキャンし，物体検出の候補領域を得る．各物体検出の候補領域を，大きくリサイズし，検出処理をする．最後に，全ての検出結果を元の画像の位置に変換し，統合する．この処理は，確かに小物体の検出には効果的であるが，拡大した候補領域ごとに物体検出をするため，一枚の大きい画像に対する計算量が数倍から数十倍になる．また，切り出した候補領域群には，実際に物体を含む候補領域は少ししか存在せず，大抵の候補領域に対する処理は無意味である．この問題に対し，音声情報を利用し，対象の方位を特定し，物体検出の候補領域を絞込む手法が提案されている [4].

2.2 物体追跡

深層学習の急速な発展により，画像特徴の表現は飛躍的に高まり，物体追跡の精度向上に寄与している．Tracking-by-detection による追跡では，追跡対象間の類似度の大きさに基づいて追跡対象を対応付ける．類似度の定義として最も単純な方法は，物体の中心点の距離 a の利用である．しかし，フレームレートが高い映像では，追跡対象の移動は小さいため，検出枠の重なり度合い IoU を用いることが一般的となった．SORT [2] では，物体の大きさを考慮して，IoU を類似度として定義し，その大きさにもとづいて Hungarian マッチングをしている．DeepSORT [6] はその上に物体特徴を導入し，オクルージョンによる物体の消失から再出現時の対応付けが誤りの問題に対処した．StrongSORT [3] は DeepSORT [6] の各モジュールを更新し，画像特徴量の改善，ガウス平滑補間による軌跡の修復などを加え，物体追跡性能を向上させた．これらの手法では同一物体であれば，隣接フレームで検出矩形が重なっていることが想定されている．しかし，高速に動く小物体の場合，物体の移動により IoU は急激に変化してしまう．図 2 のように，物体矩形が 10 ピクセルしか移動していないにもかかわらず，IoU が 0 になってしまう．鳥の追跡では，このような状況がよく発生する．

3. 提案手法

本研究は図 3 のような Tracking-by-Detection のパラダイムを使用し，検出と追跡の 2 つに対して工夫を加えることにより，小物体の追跡を実現する．追跡全体の流れは図 4 のように，入力画像に対し，まずは物体の追跡情報を確認する．追跡情報がない場合，画像全体に対し，一様に検出候補領域をサンプリングする（通常の SAHI）．追跡情報がある場合，運動予測により，今のフレームに非一様に検出候補領域をサンプリングする（Adaptive SAHI）．検出結果はまず物体特徴による対応付けを行い，ID 付与できない検出矩形に対し，中心点のマハラノビス距離による対応付けをする．3.1 は Adaptive SAHI，3.2 は中心点のマハラノビス距離による対応付けについて説明する．

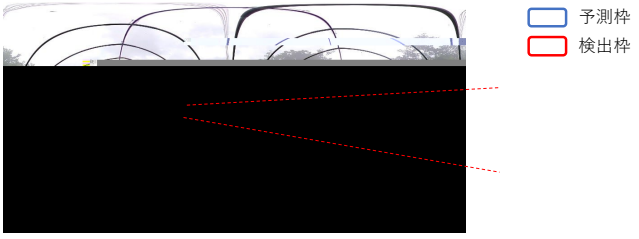

図 2 小物体追跡時に IoU を用いることの問題

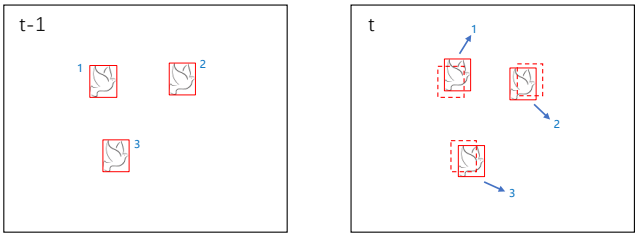

図 3 Tracking-by-Detection

3.1 検出に対する工夫：Adaptive SHAI

SAHI では，物体を含まない多くの候補領域に対しても検出処理をするため，検出の効率が悪かった．これに対し，提案手法の Adaptive SAHI では，時系列情報を利用し，物体が存在しそうな領域周辺に対して重点的に検出処理をする．物体がありそうな領域を決めるために，時系列フィルタリングの考え方を，追跡対象物ではなく，物体検出候補領域に対して適用する．

Adaptive SAHI は，前フレームの追跡情報により，SAHI のウインドウを画像全体から一様に切り出す，または，前フレームでの検出結果に基づいて SAHI のウインドウ切り出しを非一様にすることで状況に応じた対象領域の調整ができ，精度を保ちつつ効率を高める．

非一様に候補領域をサンプリングする部分の処理手順は以下の通りである．

1. Kalman フィルタを使い，ここまでの追跡情報から，今のフレームの分布を予測する
2. 分布の平均値点を囲む SAHI の物体検出候補領域を選び，拡大して検出する

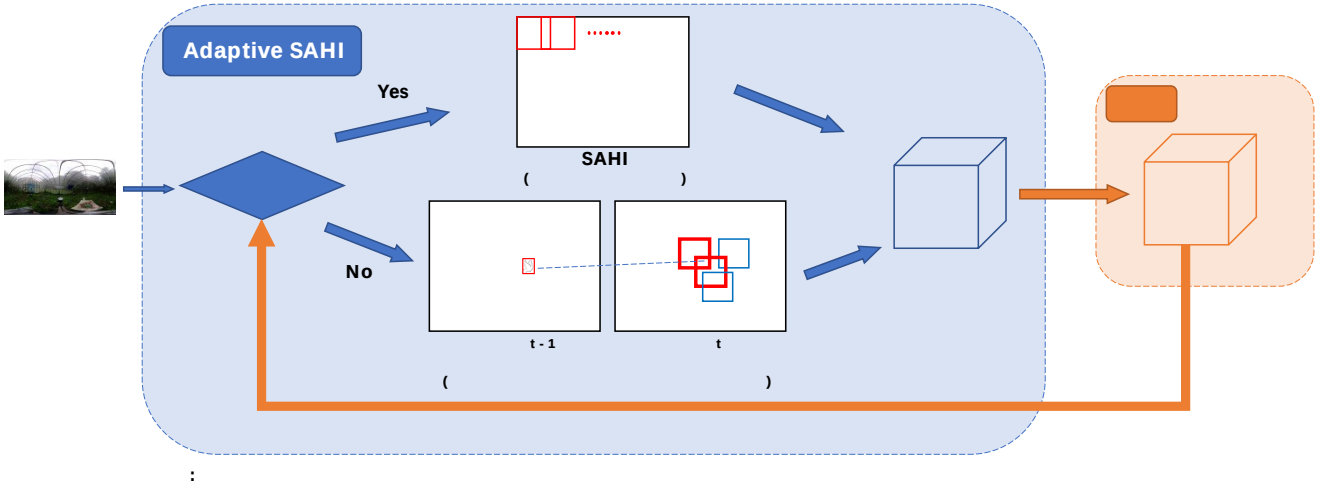

図 4　提案手法の流れ

3.　検出矩形を元の画像に変換し，結果を統合する

前のフレームで何も物体検出されなかった場合，物体の存在に関する事前知識がないため，通常の SAHI と同様に，画像全体から一様に検出候補領域をサンプリングする．また，Adaptive SAHI では追跡中の物体の周辺に注目するため，それ以外の物体を見落とす可能性がある．そこで，前の一様サンプリングの検出からのフレーム間隔のしきい値を設定し，そのしきい値に達する場合も画像全体に対する一様サンプリングすることにより，見落としを回避する．

3.2　追跡に対する工夫：マハラノビス距離の利用

マハラノビス距離による対応付けは図 5 のように，Kalman フィルタを使い，ここまでの追跡情報から，今のフレームの分布を予測し，検出矩形の中心点とのマハラノビス距離を計算し，検出矩形は距離の近い方の ID を付与される．

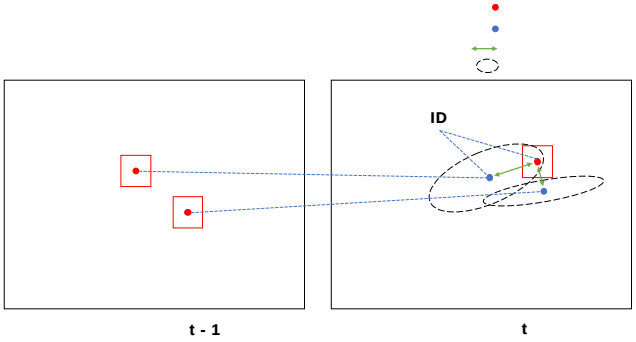

図 5　中心点のマハラノビス距離による対応付け

4.　実験

4.1　実験準備

データセット．実験で用いたデータセットは NUBird2022 [4, 5] という，5 羽のキンカチョウ（学名：Taeniopygia guttata）がいる閉鎖的な環境で約 8 分間記録された映像である．映像は 14,459 フレームあり，各フレームの大きさは $4,096 \times 2,048$ である．実験のため，元の映像のうち複雑なシーン 2 つを選んでテストに使い，残る部分は 3:1 の比率でランダムに訓練セットと検証セットに分割した．まず，2 つのセットともに Slicing aided fine-tuning(SF) [1] を適用した．元の画像を分割し，鳥を含む領域を選んで 2 つのセットに入れる．切り出しの選択について，鳥のある画像は，少なくとも 1 つの鳥が存在する切り抜きを訓練セットあるいは検証セットに入れる．鳥のない背景だけの画像は，全切り抜きを訓練セットに入れる．Slicing のサイズは 256×128，overlap ratio は 0.25，リサイズ後の大きさは 640×320 である．分割の結果は以下の通りである．

- 訓練セット：27,395 枚画像，サイズは $4,096 \times 2,048$ と 640×320 である．
- 検証セット：9,157 枚画像，サイズは $4,096 \times 2,048$ と 640×320 である．
- テストセット：
  - クリップ 1：874（フレーム 1,690~2,563）枚画像，サイズは $4,096 \times 2,048$ である．
  - クリップ 2：1,451（フレーム 10,490~11,940）枚画像，サイズは $4,096 \times 2,048$ である．

訓練セットと検証セットには元の大きいサイズの画像と，SF を適用して追加した小さいサイズの画像がある．クリップ 1 は鳥が連続して交晒移動するシーンがあり，クリップ 2 は巣の入り口で 2 羽の鳥が交互に消失出現をするシーンがある．

検出器として，上のデータセットでファインチューニングした YOLOv5 の Small モデル，追跡器として StrongSORT を用いた．

評価指標．軌跡の検出精度と安定性を評価する MOTA，物体 ID の Precision と Recall の調和平均を表す IDF1，検出

時間 DT（単位：秒）の 3 つを使いた.

4.2 実験結果

クリップ 1 とクリップ 2 での実験結果を表 1 に示す. A-SAHI は Adaptive SAHI を意味し，MhD は中心点のマハラノビス距離による対応付けを意味する．領域の絞込みは検出時間を大幅に減少するだけでなく，MOTA ではクリップ 1 で 14 ポイント，クリップ 2 で 61.8 ポイント上回った．IDF1 はクリップ 2 で 15 ポイント上回ったが，クリップ 1 の場合は逆に低下した．原因については未だ不明であり，引き続き詳細な実験・分析の必要がある．中心点のマハラノビス距離による対応付けでは，IoU による対応付けと比較して，MOTA の向上は少ないが，IDF1 はクリップ 1 で 37.9 ポイント，クリップ 2 で 12.4 ポイント上回り，鳥の追跡における尺度として IoU よりマハラノビス距離の方が適切であることを示した.

表 1 実験結果.

	クリップ 1			クリップ 2		
	MOTA↑	IDF1↑	DT↓	MOTA↑	IDF1↑	DT↓
SAHI	76.6	63.7	2057.4	10.9	32.5	3384.7
A-SAHI	90.6	54.3	124.1	72.7	47.5	250.1
A-SAHI+MhD	**90.8**	**92.2**	**124.1**	**76.6**	**59.9**	**250.1**

5. むすび

本研究では，固定された広角カメラを用いた鳥追跡のタスクにおいて，前フレームでの検出をもとに物体検出領域を絞り込む Adaptive SAHI を提案した．NUBird2022 というデータセットで実験を行い，提案手法の有効性を示した.

今後の展望としては，IDF1 が低下した原因の究明，より新しい検出器での実験を考えている．また，鳥の消失と再出現の時の ID 維持について，外観的な区別が困難な場合，時系列運動情報の利用による新たな解決策も検討したい.

謝辞

本研究は JSPS 科研費 JP20H00475, JP21H03519 の助成を受けたものです.

参考文献

[1] Akyon, F. C., Altinuc, S. O. and Temizel, A.: Slicing aided hyper inference and fine-tuning for small object detection, *2022 IEEE International Conference on Image Processing (ICIP)*, IEEE, pp. 966–970 (2022).

[2] Bewley, A., Ge, Z., Ott, L., Ramos, F. and Upcroft, B.: Simple online and realtime tracking, *2016 IEEE international conference on image processing (ICIP)*, IEEE, pp. 3464–3468 (2016).

[3] Du, Y., Zhao, Z., Song, Y., Zhao, Y., Su, F., Gong, T. and Meng, H.: Strongsort: Make deepsort great again, *IEEE Transactions on Multimedia* (2023).

[4] Kawanishi, Y., Ide, I., Chu, B., Matsuhira, C., Kastner, M. A., Komamizu, T. and Deguchi, D.: Detection of Birds in a 3D Environment Referring to Audio-Visual Information, *2022 18th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, IEEE, pp. 1–7 (2022).

[5] Sumitani, S., Suzuki, R., Arita, T., Nakadai, K. and Okuno, H. G.: Non-invasive monitoring of the spatio-temporal dynamics of vocalizations among songbirds in a semi free-flight environment using robot audition techniques, *Birds*, Vol. 2, No. 2, pp. 158–172 (2021).

[6] Wojke, N., Bewley, A. and Paulus, D.: Simple online and realtime tracking with a deep association metric, *2017 IEEE international conference on image processing (ICIP)*, IEEE, pp. 3645–3649 (2017).