

動画中の稀少イベントとしての小領域移動物体の検出手法

Detection of small moving objects as rare events in videos

西田 健次^{1*} 糸山克寿^{1,2} 中臺一博¹
Kenji Nishida¹, Katsutoshi Itoyama^{1,2}, Kazuhiro Nakadai¹

¹ 東京工業大学 工学院 システム制御系

¹ Dept. of Systems and Control Engineering, School of Engineering, Tokyo Institute of Technology

² (株) ホンダ・リサーチ・インスティテュート・ジャパン

² Honda Research Institute Japan

Abstract: 魚眼レンズによる全天動画像から、動きによる鳥の検出手法を提案し、その有効性を検証した。全天画像内での鳥の画像は小さく、また、変形しながら移動するため、形状による物体検出手法の適用は困難である。また、木の枝葉による隠れにより、鳥の全身画像が得られないことも多く、形状による検出をさらに難しくしている。さらに、数分程度の動画中に鳥が現れるのは、全体でも数秒程度であり、一羽の鳥が現れている時間は1秒に満たないことが多く、動画像全体で物体検出を行うことは効率が良いとは言えず、あらかじめ鳥が存在するフレームを切り出すことが望ましいと考えられる。一方で、5画素程度の小領域であっても、移動によるオプティカルフローは検出可能である。また、鳥自体の動きだけでなく、鳥による枝葉の揺れも鳥検出（推定）の手掛かりとなる。しかし、鳥由来のオプティカルフローと、背景となる樹木の風などによるオプティカルフローを弁別することが課題となる。本稿では、オプティカルフローによる鳥検出手法を提案し、また、背景のオプティカルフローを弁別する手法について検証する。

1 はじめに

小物体検出は、コンピュータビジョンにおいて重要な課題となっている。その中でも、画像中の鳥の検出は2023年の18th International Conference on Machine Vision and Applications (MVA)において、*Small Object Detection Challenge for Spotting Birds (SOD4SB)* [1]が開催されるなど、大きな注目を集めている課題である。一般に、小物体検出は、ノイズが多く不鮮明な情報量の不足した外観でしかないため難しい課題と言えるが、鳥の検出に関しては、(i) 羽ばたきながら移動するため、形状が一定しない、(ii) 画像内の背景領域と重なっている、(iii) 鳥の種類、カメラとの距離などにより、その大きさのバリエーションが大きいなどの問題があり、課題をさらに難しいものとしている。SOD4SBで高評価を得た手法は、異なるスケールのバウンディングボックスを重み付きで統合する手法 [2]、Swin Transformerを利用する手法 [3]、スケールに応じたバンドパスフィルターを用いる手法 [4] 等が挙げられ、70%から84%の平均精度 (AP) の性能が得られている。

本稿で対象とする画像は、千葉県佐倉市内に設けた魚眼レンズによる半球画像の動画であり、野鳥の鳴き声による生態観察 [5] と並行して収録されたものである。収録されている鳥の姿は大きいもので20画素程度、小さいもので5画素程度と、SOD4SBで提供されたデータセットのほぼ下限の大きさを対象としている。また、天頂に向けた画角となるため、鳥の姿はシルエットのみとなり色情報は期待できない。また、樹木、枝葉との重なりも多く、鳥の姿全体を確認できることは稀である。そこで、形状ではなく画像内での動きに着目して、鳥の検出を行う手法を検討してきた。我々の先行研究 [6] においては、各フレームから得られるオプティカルフロー強度画像を時空間方向に平滑化することにより、ノイズと背景の動きを軽減する手法を提案したが、移動しない鳥（その場での動きはあるので、オプティカルフローは発生する）や、非常に小さいもの（5画素程度のもの）を見落としていた。本稿では、鳥の動きによるオプティカルフローが、周辺の平均的なオプティカルフローよりも大きいと考えて、鳥検出を行う手法を提案した。

*連絡先：東京工業大学
152-8552 東京都目黒区大岡山 2-12-1 W8-30
E-mail: nishida@sc.e.titech.ac.jp



図 1: 全天画像の例

2 動画データ概要

本稿で使用した動画データは、魚眼レンズによる全天（半球）画像であり、その内容は、日出日没の前後1時間、その他1時間ごとに10分ずつ録画したもので、同時にマイクロホンアレイによる多チャンネル録音も行っている。カメラは4か所に設置されており、樹木の多いもの、空の占める面積の大きいものなど、それぞれ異なる傾向の画像が収録されている（図1）。

画像に現れる鳥の動きは、鳥の姿はあまり見えず枝を渡るもの、中空を飛ぶもの（5.3節）に大別することができる。木の枝を渡る鳥に関しては、更に、鳥の動きに伴い周辺の枝葉が動くもの（5.1節）と、枝の動きを伴わないもの（5.2節）に分類される。特に、枝の動きを伴わないものは、鳥の画像領域が小さく背景との境界が不鮮明なため、肉眼で確認しても鳥と認識することは難しい。そのような場合でも、鳥の動きによるオプティカルフローは検出できるため、これを手掛かりに鳥の検出を行うこととした。

3 オプティカルフローによる鳥領域の検出手法

オプティカルフローの検出手法には、典型的な手法である Lucas-Kanade 法、周辺の領域で正規化を行う Horn-Schunck 法 [7]、階層的により詳細なフローを求める Farneback 法 [8] が挙げられる。Farneback 法が他の手法よりも豊富なフローを検出できるが、処理時間がかかることが難点である。Horn-Schunck 法、Lucas-Kanade 法ともに、フローの検出能力に大差はなく、本研究では Lucas-Kanade 法を採用した。

鳥の動きによるオプティカルフローは、鳥が羽ばたき変形しながら移動するため、鳥の移動方向に限らず様々な方向へのフローが発生している。そのため、方向は問わずオプティカルフローの強度の高い領域を鳥

の存在する領域と考える。風が弱く背景となる樹木の動きが少ない場合でも、ノイズや微小な揺れによって背景にもオプティカルフローが発生する。しかし、背景由来のフローに対して、鳥由来のフロー強度は大きいと考えられるため、適切な閾値を設定することで鳥領域を検出できる。

一方、風が強く樹木が大きく動く場合には、背景由来のオプティカルフロー強度も大きくなるため、単純な閾値設定では鳥領域を検出することはできない。しかし、鳥由来のオプティカルフローは小領域にある程度の強度を持ったフローが発生するのに対し、背景の樹木動き由来のオプティカルフローはある程度の大きさを持った領域に比較的均等にフローが発生すると考えられる。したがって、周辺領域よりも大きなフロー強度を持つ小領域を検出することができれば、鳥領域の推定が可能である。そして、この手法は、背景のオプティカルフローがあまり発生しない場合でも、共通に適用できる。

4 鳥領域の検出手順

鳥領域の検出は、(a) 画像全体のオプティカルフローの検出、(b) 局所的に周辺よりも大きなオプティカルフロー強度を持つ領域の強調、(c) 強調されたオプティカルフロー強度の二値化の三段階で行う（図2）。

オプティカルフローの検出は、標準的な Lucas-Kanade 法を用いた。この手法は、ある画素の時間的な輝度変化と画素周辺の空間的な輝度勾配をもとに、画像中の物体の移動量を推定するもので、式 (1) を解くことで計算される。

$$\frac{\partial I}{\partial x}u + \frac{\partial I}{\partial y}v = -\frac{\partial I}{\partial t} \quad (1)$$

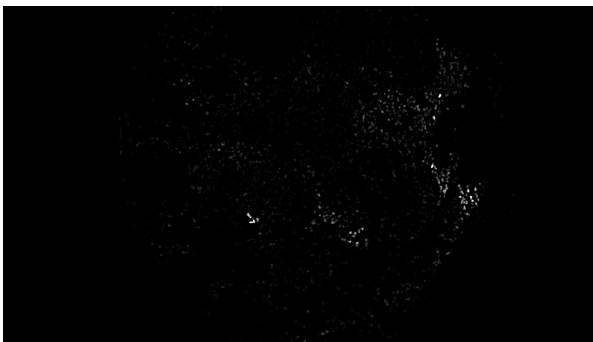
ここで、 I は画素の輝度値を示し、右辺は画素輝度値の時間的な変化である。左辺 $\partial I/\partial x$ 、 $\partial I/\partial y$ は、画素周辺の空間的輝度勾配を示し、 u 、 v は画像中の x 、 y 方向への移動量を示す。

Lucas-Kanade 法によるオプティカルフローは、画像中の輝度勾配に沿った方向に検出される性質があるため、小領域が移動する場合には、その周りに様々な方向のオプティカルフローが検出されることになる。そのため、オプティカルフローの方向成分は用いず、オプティカルフローの強度のみを用いることとした。

画像中で周辺よりも明るい領域を抽出にはいくつかの手法があるが、今回は DoG (Difference of Gaussians) フィルタをオプティカルフロー強度の画像に適用した。DoG フィルタは、 σ の異なる二種の Gaussian フィルタの差分をとるもので、以下のように定義される。



(a) オプティカルフロー検出



(b) フロー強度画像にDoGフィルタを適用



(c) 二値化

赤で示した部分が鳥の動きとして検出された領域. 黄色で示した部分は, DoGフィルタで除去しきれなかった背景の動き.

図 2: 鳥領域検出手順

ガウス関数は式 (2) で定義される.

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/(2\sigma^2)} \quad (2)$$

ここで x, y は, 画像中の縦軸横軸をしめし, σ はガウス関数の標準偏差である.

DoG フィルタは, 異なる標準偏差 σ_1, σ_2 を持つ二種のガウス関数の差分をとる (式 3).

$$\text{DoG}(x, y, \sigma_1, \sigma_2) = G(x, y, \sigma_1) - G(x, y, \sigma_2) \quad (3)$$

DoG フィルタは, 輪郭抽出に用いられることが多いが, σ_1 を 1, σ_2 を 5 とすることで, 5 画素から 10 画素程度の小領域の抽出を行った.

5 実験結果

風が穏やかで樹木の動きが少ないもの (動画 1) と風が強く樹木が大きく動き続けるもの (動画 2) に対して, 鳥の検出を行った. 動画 1 では, 鳥の姿はほとんど見えず枝の揺れによって存在を確認できるものと, 非常に小さい (5 画素以下) ため肉眼でも確認が難しいものの検出を行った. 動画 2 では, 樹木の動きを除去しつつ, 鳥の動きを検出できるか否かが課題であった.

5.1 枝の揺れによる検出

ここで用いた動画 1 は, 画面右下側に自動車写っているため, その動きを検出してしまっている. しかし, 風は穏やかなため, その他の動きは, ほぼ鳥由来と考えることができる. 画面左側, やや下の領域で, 鳥が枝を渡ることによるオプティカルフローが発生している (図 3). 枝が密集していることもあり, 鳥の姿は肉眼でも確認することは難しい. また, 枝の揺れがしばらく続くため, 鳥の位置は正確にはわからず, フローの発生する位置も飛び飛びとなっている. しかし, 鳥の姿を確認できない状況であっても, 大まかな位置を推定することが可能であることは示された.

5.2 枝の動きのない小領域の検出

前節と同じく動画 1 の風が穏やかな状況での検出を行った. 画面中央上方の枝に沿って縦横 5 画素程度の小さな鳥が移動している (図 4)2. このシーケンスを肉眼で確認したが, 鳥の形状は認識することができず, また, 画素の動きも言われてみなければわからないレベルであり, オプティカルフローの検出によって, 鳥の存在を確認出来た. 鳥領域は 5 画素 \times 5 画素程度の大きさで, 鳥の形状とは言えないようなものであった.

オプティカルフロー
高強度領域周辺

鳥検出領域
(鳥の形状は認識できない)

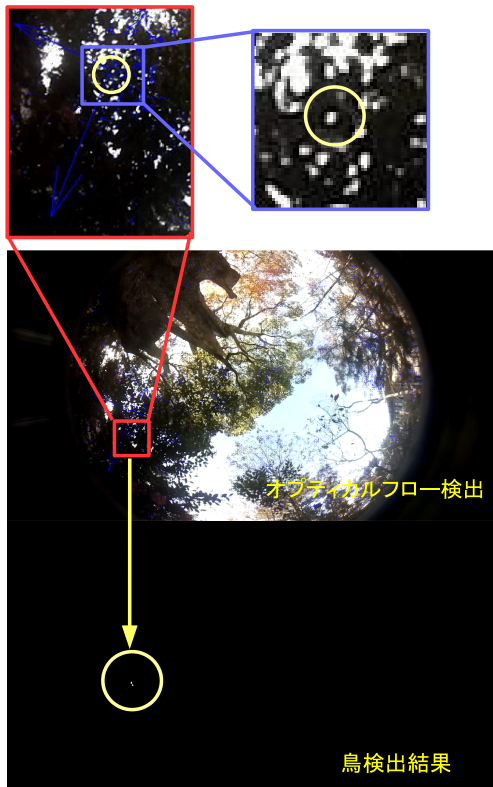


図 3: 枝の動きによる検出

オプティカルフロー
高強度領域周辺

鳥検出領域
幅30画素、縦15画素程度
比較的大きいが、
鳥の形状とは言い難い

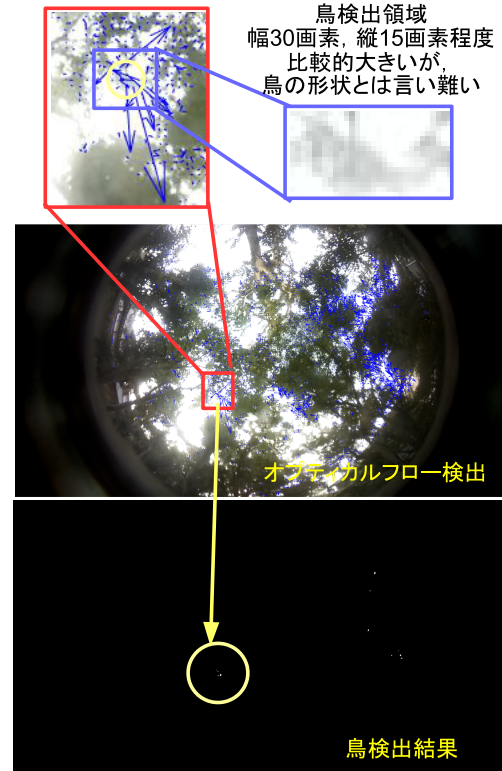


図 5: 背景樹木の動き除去

オプティカルフロー
高強度領域周辺

鳥検出領域
(鳥の形状は認識できない)

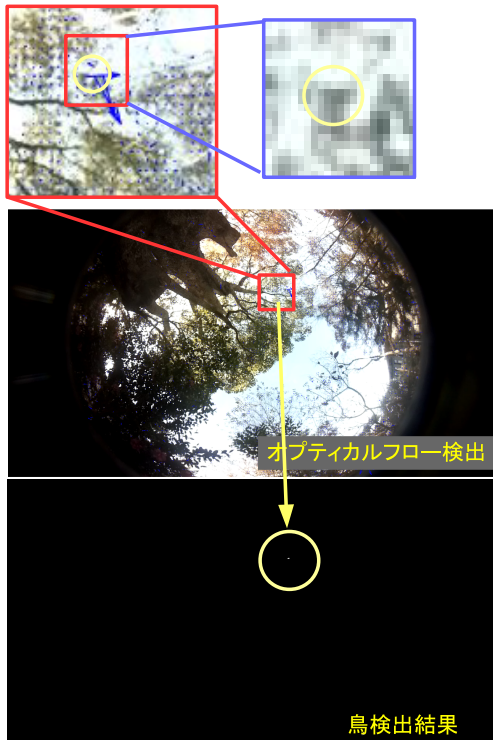


図 4: 枝の動きのない小領域の検出

5.3 背景樹木の動き除去

動画 2 は、風の強い日に収録されたもので、常に樹木が大きく揺れている。そして、時折、枝によって隠されていない部分を鳥が通過するものである(図 5)。この例では、鳥は 30 画素 × 15 画素程度と、比較的大きな画像として写っているが、鳥の形状になっているとは言い難いものである。

手順例(図 2)でも示したように、画面右側の樹木の動きが完全には除去できていない。しかし、樹木の揺れによるオプティカルフローは、鳥の移動によるオプティカルフローとは異なり、大きくは移動しないと考えられるため、検出された領域の動きを評価することにより、鳥領域との判別は可能になると考えられる。

6 結論と今後の課題

本稿では、低解像度小領域で隠れが頻発する動画像から、オプティカルフローを手掛かりに鳥を検出する手法を提案した。提案手法での課題は、小領域であるため形状による検出が困難であり、また、隠れが頻発する状況での鳥領域の検出だった。このような状況でも、鳥の移動によるオプティカルフローを検出するこ

とができ、オプティカルフロー強度を手掛かりに鳥の領域を検出できることが示された。

また、風が強く背景となる樹木のオプティカルフローは、局所的に周辺よりも強い強度を持つ小領域を強調するDoGフィルタを適用することにより、大まかには除去できることが示された。DoGフィルタの二つの標準偏差 σ_1, σ_2 は、小領域を強調するために重要なパラメータであるが、現在は、数種類の試行錯誤によって決定している。データに対して最適なパラメータを求める手法を検討していく必要がある。

DoGフィルタは、樹木の動きは鳥に比べて大きくはなく、また、鳥よりは広い領域である程度均等なオプティカルフローを発生しているという仮定に基づいて適用されている。しかし、条件によっては、枝が大きく動くところもあり、背景として除去しきれいていない部分がある。これは、単純なフィルタ処理では、完全には除去できないことを示しており、我々の先行研究[6]で行った時空間での平滑化などを併用する必要があると示されている。

謝辞

本研究はJSPS科研費JP19KK0260およびJP20H00475の助成を受けた。

参考文献

- [1] Yuki Kondo, Norimichi Ukita, Takayuki Yamaguchi, Hao-Yu Hou, Mu-Yi Shen, Chia-Chi Hsu, En-Ming Huang, Yu-Chen Huang, Yu-Cheng Xia, Chien-Yao Wang, Chun-Yi Lee, Da Huo, Marc A. Kastner, Tingwei Liu, Yasutomo Kawanishi, Takatsugu Hirayama, Takahiro Komamizu, Ichiro Ide, Yosuke Shinya, Xinyao Liu, Guang Liang, and Syusuke Yasui. Mva2023 small object detection challenge for spotting birds: Dataset, methods, and results. In *2023 18th International Conference on Machine Vision and Applications (MVA)*, pp. 1–11, 2023.
- [2] Hao-Yu Hou, Mu-Yi Shen, Chia-Chi Hsu, En-Ming Huang, Yu-Chen Huang, Yu-Cheng Xia, Chien-Yao Wang, and Chun-Yi Lee. Ensemble fusion for small object detection. In *2023 18th International Conference on Machine Vision and Applications (MVA)*, pp. 1–6, 2023.
- [3] Da Huo, Marc A. Kastner, Tingwei Liu, Yasutomo Kawanishi, Takatsugu Hirayama, Takahiro Komamizu, and Ichiro Ide. Small object detection for birds with swin transformer. In *2023 18th International Conference on Machine Vision and Applications (MVA)*, pp. 1–5, 2023.
- [4] Yosuke Shinya. Bandre: Rethinking band-pass filters for scale-wise object detection evaluation. In *2023 18th International Conference on Machine Vision and Applications (MVA)*, pp. 1–5, 2023.
- [5] Shinji Sumitani, Reiji Suzuki, Takaya Arita, Kazuhiro Nakadai, and Hiroshi G. Okuno. 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] 西田健次, 糸山克寿, 中臺一博. 低解像度画像からの小領域物体の検出手法の検討. 人工知能学会第二種研究会資料, Vol. 2022, No. Challenge-061, p. 03, 2022.
- [7] ANDR ES BRUHN, JOACHIM WEICKERT, and CHRISTOPH SCHN ORR. Lucas/kanade meets horn/schunck: Combining local and global optic flow methods. *International Journal of Computer Vision*, Vol. 61, No. 3, pp. 211–231, 2005.
- [8] Gunnar Farnebäck. Two-frame motion estimation based on polynomial expansion. In Josef Bigun and Tomas Gustavsson, editors, *Image Analysis*, pp. 363–370, Berlin, Heidelberg, 2003. Springer Berlin Heidelberg.