

一人称鳥視点映像解析の試み

玉木 徹¹ 竹内 一郎² 前川 卓也³ 依田 憲⁴

概要: 本稿では、鳥に取り付けたカメラから得られた鳥視点映像のフレーム分類について、特徴量と識別手法の初歩的な比較検討結果を報告する。各フレームにおいて抽出する特徴量として、色ヒストグラムと事前学習済み CNN 特徴を用い、識別器にはロジスティック回帰と SVM を用いた。また畳み込み 3 層と全結合 3 層からなる DCNN も用いた。学習とテスト各 1 本の動画映像を用いた実験結果では、CNN 特徴量を SVM で識別した結果が 85.6% と最も高い性能が得られた。

First-bird view video recognition

1. はじめに

人に取り付けたカメラの映像は一人称視点 (first-person view) と呼ばれ、監視カメラなどの固定カメラ映像と異なり、ウェアラブルなカメラ自体の運動が激しく、また雑多な内容が撮影されるため、挑戦的な課題として近年盛んに研究が行われている。一方バイオロギングの分野では動物にカメラを取り付けて、動物視点の映像の撮影が行われているが、一人称視点映像よりも解析が困難な映像であり、今のところ動画認識などの手法は適用されていない。本稿では、鳥類に取り付けたカメラから得られた鳥視点映像 (first-bird view) のフレーム分類を試みた結果を報告する。

2. データセット

実験に用いる映像は、鳥 (鸕、imperial cormorant) の背中に取り付けた小型カメラで撮影された 2 本の MPEG4 カラー動画 (736 × 480, 25fps) である (詳細は [1] を参照)。詳細を表 1 に示す。この動画はどちらも、鳥が海上を飛行 (flying) している最中に海中へ潜水 (diving) して海底で採餌し、水面へと浮上 (floating) して羽の水を払い飛ばし (washing)、飛翔するまでの様子を捉えている。

表 1 実験に用いた動画データセットの詳細

	時間 (frames)	diving	floating	flying	washing
学習	17:54 (26869)	4050	4025	17794	1000
テスト	12:08 (18211)	4500	3336	8825	1550

3. 手法

3.1 カラーヒストグラム

この鳥視点映像には、海中と海上で特徴的な色が画面を占めるため、カラーヒストグラム特徴量が有効であると思われる。そこで、各フレームの RGB チャンネルそれぞれで区画数 100 のヒストグラムを求め、それを連結した 300 次元のベクトルを特徴量とする。

3.2 CNN 特徴量

ImageNet で事前学習済みの CNN を用いた特徴量が有効であることが多くの研究で示されており [2], [3], 本実験でもこれを用いる。具体的には、学習済みの AlexNet [4] に対して 227 × 227 に縮小したフレームを入力し、その fc6 の出力 4096 次元ベクトルを特徴量とする。

3.3 DCNN

一人称視点映像のフレーム識別にも CNN が用いられている。本実験では神谷ら [5] が提案した DCNN を利用する。150 × 150 に縮小したフレームを入力とし、前半畳み込み 3 層 conv1-3 と後半全結合 3 層 fc4-6 からなり、畳み込み層では max pooling を、また各層において batch normalization (BN) [6] または dropout を適用する。学習には Adam [7]

¹ 広島大学
Hiroshima University
² 名古屋工業大学
Nagoya Institute of Technology
³ 大阪大学
Osaka University
⁴ 名古屋大学
Nagoya University



(a)



(b)



(c)



(d)

図 1 鳥動画データセット [1] 中の画像例. (a) diving, (b) floating, (c) flying, (d) washing の各ラベルに相当するフレーム.

を用い、バッチサイズ 25 で 20 エポックとした。この時、学習用のデータセットからランダムに選んだ 25000 フレームを学習し、残りの 1869 フレームを用いて 1000 回反復毎に validation loss を計算した (図 2 参照)。そして最も小さい loss を与えたものを用いてテストデータを分類した。

表 2 DCNN の詳細

層	フィルタ/ユニット	活性化	正則化	プーリング
入力	$3 \times 150 \times 150$			
conv1	$32 \times 3 \times 3$	ReLU	BN	2×2 max
conv2	$64 \times 3 \times 3$	ReLU	BN	2×2 max
conv3	$64 \times 3 \times 3$	ReLU	BN	2×2 max
fc4	64	ReLU	dropout	
fc5	128	ReLU	dropout	
fc6	4			

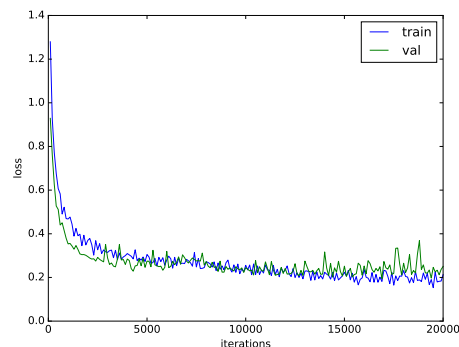


図 2 DCNN 学習時の損失

4. 実験結果

上記のデータセットに対して認識実験を行った結果を表 3 に示す。カラーヒストグラム特徴量と CNN 特徴量に対しては、ロジスティック回帰と線形カーネル SVM を用いて識別した。

カラーヒストグラムを用いた場合の confusion matrix を図 4 と 3 に示す。データセットのバランスが悪く flying クラスが大半を占めているため、floating の大半が flying に識別されてしまっており、予想通り色ではこの区別が付けられていない。しかし特徴的な色を持つ diving のフレームの多くが floating に識別されていることから、単純な色特徴だけではこれらの区別をつけることが困難であることが分かる。

CNN 特徴量を用いた場合は、図 6 と 5 に示すように、かなり改善されていることが分かる。ヒストグラムでは識別が困難であった floating と flying も分離されてはいるが、非常に類似した floating と washing を識別することは困難である。

DCNN を用いた場合の結果を図 7 に示す。washing と flying の多くが floating に分類されてしまっており、このネットワーク構造では学習が困難であることが分かる。

CNN 特徴量を SVM で識別した際に得られる事後確率を時系列としてプロットしたものを図 9 に示す。テストデータの事後確率曲線から分かるように、flying と diving の期間は安定して識別できているが、floating と washing の期間は複数の類似したクラスが競合している。これを時系列処理することでより安定した結果が得られると思われる。

表 3 実験結果

特徴	識別器	認識率
色ヒストグラム	ロジスティック回帰	69.8
色ヒストグラム	SVM	69.4
DCNN		75.9
CNN (fc6)	ロジスティック回帰	84.2
CNN (fc6)	SVM	85.6
CNN (fc6)	SVM + DPF	89.9

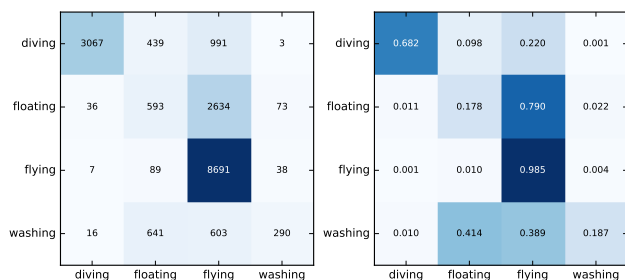


図 3 カラーヒストグラム + SVM の confusion matrix

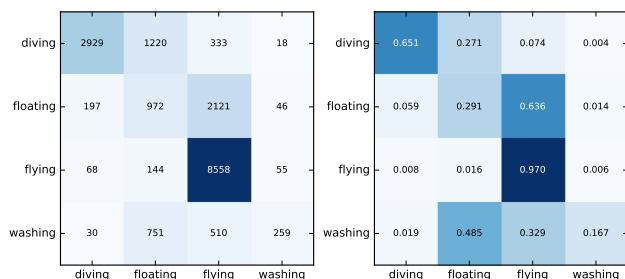


図 4 カラーヒストグラム + ロジスティック回帰の confusion matrix

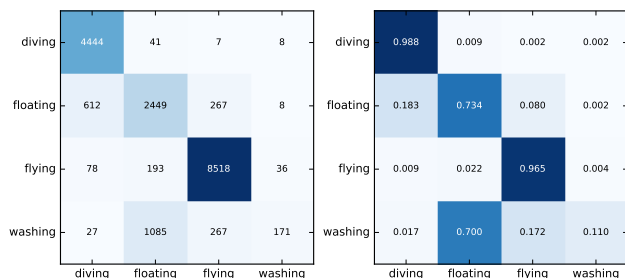


図 5 CNN 特徴量 + SVM の confusion matrix

そこで, Hirakawa らの Dirichlet Particle Filter (DPF) [8] を用いて平滑化した結果を図 9(c) に, confusion matrix を図 8 に示す. この時系列後処理によりフレーム毎の認識率も向上し, 最終的に 89.9% となった.

5. おわりに

本稿では, 鳥類に取り付けたカメラから得られた鳥視点映像のフレーム分類を, カラーヒストグラム, CNN 特徴量, DCNN を用いて行った実験結果を報告した. 今回は単純にフレームを個別に分類し平滑化を適用しただけであるが, 今後はフレーム内容の時系列を考慮した RNN などの手法を検討する.

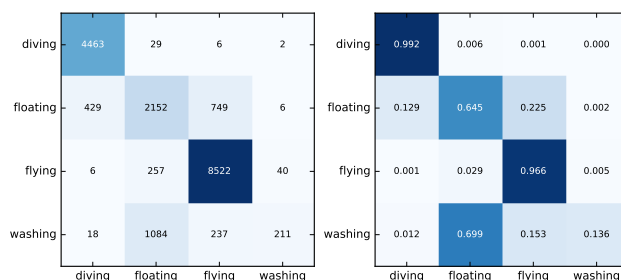


図 6 CNN 特徴量 + ロジスティック回帰の confusion matrix

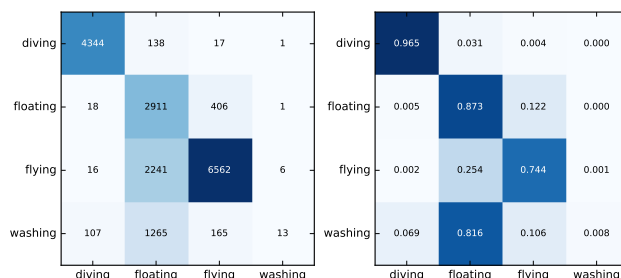


図 7 DCNN の confusion matrix

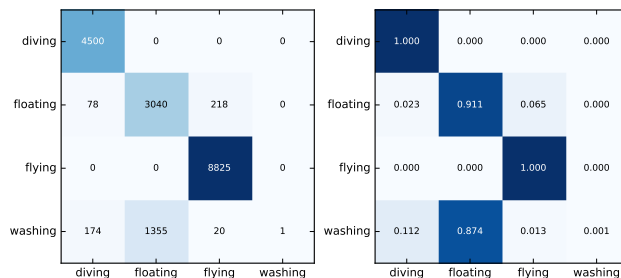


図 8 CNN 特徴量 + SVM の事後確率曲線を DPF によって平滑化した後の confusion matrix

謝辞

動画データセットを提供していただいた Flavio Quintana 氏と Agustina Gómez Laich 氏 (Instituto de Biología de Organismos Marinos (IBIOMAR-CENPAT), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)) に感謝する. 本研究の一部は JSPS 科研費 JP16H06538, JP16H06539, JP16H06540, JP16H06541 の助成を受けた.

参考文献

- [1] A. Gómez-Laich, K. Yoda, C. Zavalaga and F. Quintana: "Selfies of imperial cormorants (phalacrocorax atriceps): What is happening underwater?", PLoS ONE, **10**, 9, pp. 1–18 (2015).
- [2] A. Sharif Razavian, H. Azizpour, J. Sullivan and S. Carlsson: "CNN Features Off-the-Shelf: An Astounding Baseline for Recognition", The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops (2014).
- [3] T. Tamaki, S. Sonoyama, T. Hirakawa, B. Raytchev, K. Kaneda, T. Koide, S. Yoshida, H. Mieno and S. Tanaka: "Computer-Aided Colorectal Tumor Classification in NBI

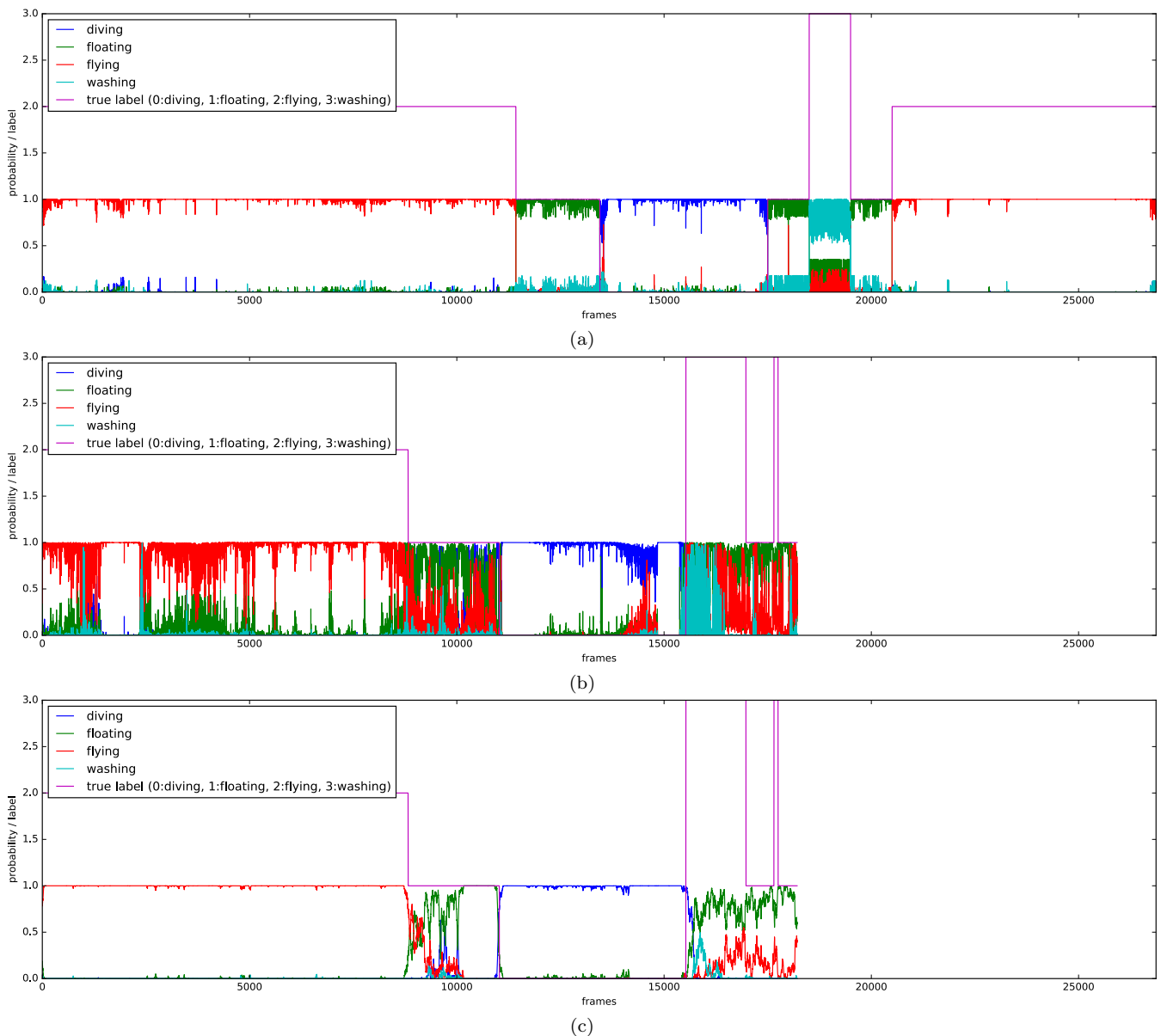


図 9 CNN 特徴量 + SVM の事後確率曲線. (a) 学習動画データと (b) テスト動画データに対して各フレームで SVM の事後確率をプロットした. (c) DPF によって平滑化した結果.

Endoscopy Using CNN Features”, The Korea-Japan joint workshop on Frontiers of Computer Vision (FCV2016), pp. 61–65 (2016).

- [4] A. Krizhevsky, I. Sutskever and G. E. Hinton: “ImageNet Classification with Deep Convolutional Neural Networks”, Advances in Neural Information Processing Systems 25 (Eds. by F. Pereira, C. J. C. Burges, L. Bottou and K. Q. Weinberger), Curran Associates, Inc., pp. 1097–1105 (2012).
- [5] 神谷, 川口, 福井, 石井, 小塚, 羽川, 築澤, 山内, 山下, 藤吉: “Convolutional-Recurrent Neural Network による自己運動識別”, 画像センシングシンポジウム (SSII) (2016).
- [6] S. Ioffe and C. Szegedy: “Batch normalization: Accelerating deep network training by reducing internal covariate shift”, Proceedings of the 32nd International Conference on Machine Learning (ICML-15) (Eds. by D. Blei and F. Bach), JMLR Workshop and Conference Proceedings, pp. 448–456 (2015).
- [7] D. P. Kingma and J. Ba: “Adam: A method for stochastic

optimization”, CoRR, **abs/1412.6980**, (2014).

- [8] T. Hirakawa, T. Tamaki, B. Raytchev, K. Kaneda, T. Koide, S. Yoshida, Y. Kominami and S. Tanaka: “Defocus-aware Dirichlet particle filter for stable endoscopic video frame recognition”, Artificial intelligence in medicine, **68**, pp. 1–16 (2016).