

A Computer Vision Approach to Classification of Birds in Flight from Video Sequences

John Atanbori¹
jatanbori@lincoln.ac.uk

Wenting Duan¹
<http://staff.lincoln.ac.uk/wduan>

John Murray¹
<http://staff.lincoln.ac.uk/jomurray>

Kofi Appiah²
kofi.appiah@ntu.ac.uk

Patrick Dickinson¹
<http://staff.lincoln.ac.uk/pdickinson>

¹ University of Lincoln
School of Computer Science
Brayford Pool,
Lincoln,
LN6 7TS,
UK

² Nottingham Trent University
School of Science and Technology
Nottingham
NG1 4BU
UK

Abstract

Bird populations are an important bio-indicator; so collecting reliable data is useful for ecologists helping conserve and manage fragile ecosystems. However, existing manual monitoring methods are labour-intensive, time-consuming, and error-prone. The aim of our work is to develop a reliable system, capable of automatically classifying individual bird species in flight from videos. This is challenging, but appropriate for use in the field, since there is often a requirement to identify in flight, rather than when stationary. We present our work in progress, which uses combined appearance and motion features to classify and present experimental results across seven species using Normal Bayes classifier with majority voting and achieving a classification rate of 86%.

1 Introduction

Bird species are an important bio-indicator of habitat quality, productivity, and stability [20]. They can be monitored to determine factors causing population fluctuation and to help conserve and manage threatened and endangered species. Collecting bird population data is useful to scientist in determining whether to build offshore wind turbines [13] and ecologist; for managing farmland use and studying migration behaviour [13, 20].

The distance sampling point count method [9] is widely used to examine avian species abundance in different habitats but this requires trained personnel to be deployed in sometimes quite inaccessible or hostile locations. The objective of our work is to develop robust and reliable methods of collecting such data automatically, using computer-vision techniques. Previous studies have used highly detailed individual images for classification of bird species (eg [4, 5, 18, 26, 30]) but these are less useful in the field. The work we report



Figure 1: Segmented birds from our birds in flight video data set. From left to right: Wood Pigeon, Green Budgie, Superb Starling and Nanday Parakeet.

is still in progress and to date, we have extracted features and classified species using appearance and motion features separately. In this paper we report initial results obtained from combining appearance and motion features, and also using a voting technique to combine images from the same video sequence for improved classification.

The remainder of our paper is organized into the following sections: in section 2 we review existing work on bird species classification, followed by an overview of our processing method including feature extraction in section 3, experiments and results in section 4, and conclusion in section 5.

2 Existing Work

Until recently, most automated classification of bird species used bioacoustics [8, 14, 17, 22, 23], by making use of audio recordings of distinctive birds' vocals. However, bird species classification based on bioacoustics is very challenging due to background noise, fine-grain nature of birds songs and sparsity of some birds' vocals [14].

Some recent studies have looked at computer vision and image-based techniques for classifying birds species [18, 21, 28]. Welinder *et al.* [28] provided high detailed individual image-based data set and classified birds using size and colour histogram. This data set was later enhanced by Wah *et al.* [21] and was used to classify species using scale-invariant feature transform (SIFT) and colour features. Even though their methods worked well for small number of species, classification rate deteriorated with an increase in the number of species.

Welinder *et al.* [28] and Wah *et al.* [21] had provided data set and preliminary work for fine grain categorisation explored further in [4, 6, 26, 30]. Berg and Belhumeur [4] learned a set of highly discriminative features by generating key points, while Zhang *et al.* [30] learned pose features by using deformable part. Other works that used part-based models to classify fine grained species include the work by Wah *et al.* [26], Deng *et al.* [10] and Yao *et al.* [29]. Most recently, Berg *et al.* [5] classified various US bird species using a set of one-vs-most SVMs, which boosted the results of their classification as opposed to the work by Berg *et al.* [4] which used one-vs-one SVM. These works are implicitly part-based methods and require the ground truth locations of each part, which are less appropriate for flying birds as their parts may be obscured from view.

Classifying bird species in flight from video has not been well researched and this is the first work to use both appearance and motion features to classify bird in flight from video. Hitherto, most researches use appearance features on high detailed images for classification or superficially explored wing beat frequencies and flight trajectories patterns of bird and bat species (but not for classification). Duberstein *et al.* [12] explored wing beat frequencies of

bird and bats species but not for classification. Atanbori *et al.* [1] also presented preliminary work on classification of the flight trajectories of bats by analysing their wing-beat frequencies. Motion-based classification has been used more widely for the identification of people, fishes, and vehicles as in Beyan *et al.* [2], Beyan *et al.* [3], Anjum and Cavallaro [4] and Li *et al.* [5].

The premise of our work is that the classification of birds in-flight may achieve an overall more robust performance when a combination of motion and appearance features are used. Hitherto, no existing work has thus by far attempted to encompass both of these in a robust or analytical way.

3 Method

3.1 Overview of Processing Method

We extracted birds' silhouettes (*Figure 1*) using the background Gaussian mixture model proposed by Zivkovic and Heijden [6] and obtained contours using contour algorithm proposed by Suzuki *et al.* [7]. We fitted an oriented bounding boxes on the silhouettes and extracted the height, width and hypotenuse, centroid and contour points.

We also extracted appearance features (colour moments, shape moments, grayscale histogram, Gabor filter and log-polar features) and motion features (wing beat frequencies, curvature scale space (CSS) and centroid distance). We concatenated these to form feature vectors by computing statistical features including mean, standard deviation, skewness, kurtosis, energy, entropy, maximum, minimum, local maxima, local minima and number of zero crossings as described below.

3.2 Appearance Features

To classify bird species, colour, shape and texture are important features. We represented colour features by colour moments and colour log-polar; shape features by shape moments; and texture features by Gabor filters and grayscale histogram.

Histogram features have been used to describe colour images by extracting the histogram from various colour channels [8]. We extracted the first-order histogram probability which represents 35 hue, 37 saturation and 37 value features to represent colour moments.

Image moments can be used to describe the shape of an image contours [9]. To represent shape information, we extracted 17 features presenting seven Hu moments and ten spatial (raw) moments.

For texture features, we converted the segmented image into a grayscale image and used it to form a histogram with 256 bins and computed eight statistical features similar to Spampinato *et al.* [10]. We also computed 20 statistical features from Gabor filter (*see eqn. 1*) for four orientations and one scale (*Figure 2 (A) shows a samples at $\theta = 0$*).

$$g(x, y; \theta, \lambda, \psi, \gamma, \sigma) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \exp\left(i\left(2\pi\frac{x'}{\lambda} + \psi\right)\right) \quad (1)$$

where:

- $x' = x \cos \theta + y \sin \theta$
- $y' = -x \sin \theta + y \cos \theta$ and θ , λ , ψ , γ and σ are orientation, wavelength, phase, aspect ration and standard deviation respectively.

To eliminate the effects of rotation and scale in the input image, we converted it into a log-polar image using *eqn. 2*, as proposed in Pun and Lee [19]. We extracted five statistical features for each image channel (hue, saturation and value) to represent log-polar feature (*Figure 2 (C) shows a sample saturation log polar*).

$$dst(\theta, \rho) \leftarrow src(x, y) \text{ for } \begin{cases} \rho = \log \sqrt{x^2 + y^2} \\ \theta = \arctan\left(\frac{x}{y}\right) \text{ if } x > 0 \end{cases} \quad (2)$$

where: $src(x, y)$ is source image with cartesian coordinate (x, y) and $dst(\theta, \rho)$ is the log-polar image with co-ordinates (θ, ρ)

3.3 Motion Features

For any bird species i , which is tracked through n frames, our trajectory model is defined as the centroid of the fitted bounded box, given by the *eqn. 3*. Where T represents the trajectory and x and y are the centroid coordinates.

$$T_i = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\} \quad (3)$$

The periodic motion features associated with wing beats vary among species [15] and provide useful discriminating features for classification. We used the approach proposed by Atanbori *et al.* [2], by fitting a bounding box to the silhouette of a tracked bird and extracted the bounding box's height, width and diagonal as three different 1-dimensional (1D) time signal. We then computed nine statistical features from the Fast Fourier Transform (excluding the DC component) of these 1D time signal.

The centroid distance function (CDF) is an invariant representation of the shape of an object [3, 2]. We computed CDF features by first finding the centroid of our trajectory model in *eqn. 3* and then computing distances of each trajectory point from this centroid (*see eqn. 4*). Ten statistical features were then extracted from the normalised CDF to represent CDF features.

$$CDF_i = \sqrt{(x_i - x_c)^2 + (y_i - y_c)^2} \text{ for } i = 0, 1, \dots, N-1 \quad (4)$$

where: N is the total number of trajectory points.

$$x_c = \frac{1}{N} \sum_{j=0}^{N-1} x_j \text{ and } y_c = \frac{1}{N} \sum_{j=0}^{N-1} y_j$$

Curvature scale space (CSS) is rotation and translation invariant and useful in distinguishing trajectories by their concave and convex shapes [2]. CSS is computed by applying a Gaussian smoothing kernel iteratively with different standard deviations. This process stops when there are no zero crossings and the trajectory is a convex curve. The curvature at every point on the trajectory is calculated using the *eqn. 5*. (*see Beyan and Fisher [2]*). We then computed 22 statistical features to represent CSS features (*Fugire 2 (B) shows a sample CSS image*).

$$K_i = \frac{x'_i y''_i - y'_i x''_i}{(x'^2_i + y'^2_i)^{\frac{3}{2}}} \quad (5)$$

where x'_i, x''_i, y'_i and y''_i are first and second derivatives of x_i and y_i respectively

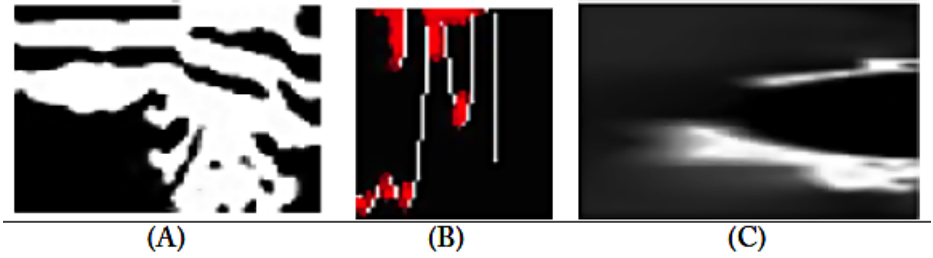


Figure 2: Sample appearance and motion features of Nanday Parakeet and Wood Pigeon: (A) shows the Gabor filter at $\theta = 0$ for a Nanday Parakeet. (B) shows the CSS image of a Wood Pigeon and (C) the saturation log polar for a Nanday Parakeet

3.4 Sequence Processing

To classify video sequences of an individual bird in flight, we first extracted appearance features for each frame of the video sequence and motion features for each of our trajectory model (*see section 3.3 eqn. 3*). We concatenated the bird's appearance and motion features, which was used to classify birds into one of the seven species (classes) using Bayes normal classifier. We then applied majority voting to the results of frames in the video sequence to determine which class a video belongs to (*see eqn. 6*). Where $d_{i,j}$ is the decision of the classifier at frame i for class j (which is 1 if class j is voted and 0 otherwise) and \hat{t} is the decision of the classifier for the entire video sequence.

$$\hat{t} = \arg \max_j \sum_{i=1}^N d_{i,j} \quad (6)$$

4 Experiments and Results

4.1 Experiments

The data set used for this work is made up of videos of flying birds from 7 different species recorded using a Casio Exilim ZR100, which was fixed at stationary points and recording at 240 frames per second. The entire data set used for our experiments comprises of 139 individual's high-speed videos covering all the seven species, each between 400 millisecond and five seconds. For each experiment we applied a five-fold cross-validation scheme and aggregated the results presented in this section.

We performed two experiments to evaluate the effectiveness of our combined appearance and motion features for the classification of video sequences. The first experiment was performed using a combination of features (appearance and motion) with majority voting to determine the class a video sequence belongs to whiles the second used only appearance features with majority voting. We report the results in the next section.

4.2 Results

We present our initial results using confusion matrices of classified individual birds in videos with a combination of appearance and motion features (*see table 1*) and appearance features (*see table 2*). We found out that using appearance features alone had a correct classification rate of 83%. However, using both appearance and motion features increased the classification rate to 86%. From *table 1*, the correct classification rate for Nanday Parakeet and Cockatiels

	House Martin	Wood Pigeon	Superb Starling	Nanday Parakeet	Cockatiels	Black Bird	Green Budgie	(%)
House Martin	21	0	0	0	3	0	1	84
Wood Pigeon	0	11	0	2	0	0	0	87
Superb Starling	1	0	12	3	1	0	0	60
Nanday Parakeet	0	1	0	22	0	0	0	96
Cockatiels	0	1	0	0	17	1	1	85
Black Bird	0	0	0	0	0	19	0	100
Green Budgie	1	0	0	0	1	0	20	92
Total (%)								86

Table 1: confusion matrix of classification using the combined feature set (motion and appearance). The rows are the ground truth and the columns predictions

increased, while all others were the same. Two of the three Nanday Parakeet videos which were misclassified as Wood Pigeons were correctly classified when appearance and motion features were combined (*see table 1*). Similarly three of the six misclassified Cockatiels were correctly classified (*see table 1*), of which two misclassified as Black Birds were correctly classified and one misclassified as Wood Pigeon was also corrected.

	House Martin	Wood Pigeon	Superb Starling	Nanday Parakeet	Cockatiels	Black Bird	Green Budgie	(%)
House Martin	21	0	0	0	4	0	0	84
Wood Pigeon	0	11	0	2	0	0	0	87
Superb Starling	1	0	12	3	1	0	0	60
Nanday Parakeet	0	3	0	20	0	0	0	88
Cockatiels	0	2	0	0	14	3	1	70
Black Bird	0	0	0	0	0	19	0	100
Green Budgie	0	0	0	0	1	1	20	92
Total (%)								83

Table 2: confusion matrix of classification using appearance features. The rows are the ground truth and the columns predictions

Initial results thus shows that combining appearance and motion features improves the classification performance, especially by correcting misclassification in some species. This will be suitable for classifying species with similar colouration and those in video at distance where colour features are more likely to attenuate.

5 Conclusion

The classification of birds in flight from video is particularly challenging for automated species identification and no existing work has yet addressed this problem directly. In this paper we have described our proposed combined appearance and motion features and have presented initial experimental results. We have compared the results from combining appearance and motion features with those just using appearance features. In both cases, we first classify bird species in video frame by frame using a Bayes Normal classifier. We then used majority voting to determine the class a video sequence belongs to base on the results of individual frames.

Initial results show that combining appearance and motion features for classification, out-performs appearance features by 3% using our video data set (which covers 7 species of flying birds). Specifically, the overall correct classification using combination of appearance and motion features was found to be 86% against 83% for appearance features alone. There is evidence that misclassifications in some bird species were improved with the combined features. This demonstrates the potential of combining appearance and motion features for resolving ambiguities among species with similar colouration and for classifying species in video at distance where colour features are more likely to attenuate.

We have presented initial experimental results for our combined appearance and motion features to support ecological studies or migration and other population-level behaviours. Our on-going work seeks to provide extensions to encompass other species and also investigate feature selection and redundancy to provide a more robust automated species classification.

References

- [1] Nadeem Anjum and Andrea Cavallaro. Multifeature object trajectory clustering for video analysis. *Circuits and Systems for Video Technology, IEEE Transactions on*, 18(11):1555–1564, 2008.
- [2] John Atanbori, Peter Cowling, John Murray, Belinda Colston, Paul Eady, Dave Hughes, Ian Nixon, and Patrick Dickinson. Analysis of bat wing beat frequency using fourier transform. In *Computer Analysis of Images and Patterns*, pages 370–377. Springer, 2013.
- [3] Faisal I Bashir, Ashfaq A Khokhar, and Dan Schonfeld. View-invariant motion trajectory-based activity classification and recognition. *Multimedia Systems*, 12(1): 45–54, 2006.
- [4] Thomas Berg and Peter N Belhumeur. Poof: Part-based one-vs.-one features for fine-grained categorization, face verification, and attribute estimation. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 955–962. IEEE, 2013.
- [5] Thomas Berg, Jiongxin Liu, Seung Woo Lee, Michelle L Alexander, David W Jacobs, and Peter N Belhumeur. Birdsnap: Large-scale fine-grained visual categorization of birds. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 2019–2026. IEEE, 2014.

- [6] Cigdem Beyan and Robert B Fisher. A filtering mechanism for normal fish trajectories. In *Pattern Recognition (ICPR), 2012 21st International Conference On*, pages 2286–2289. IEEE, 2012.
- [7] Cigdem Beyan and Robert B Fisher. Detection of abnormal fish trajectories using a clustering based hierarchical classifier. *BMVC, Bristol, UK*, 2013.
- [8] Forrest Briggs, Balaji Lakshminarayanan, Lawrence Neal, Xiaoli Z Fern, Raviv Raich, Sarah JK Hadley, Adam S Hadley, and Matthew G Betts. Acoustic classification of multiple simultaneous bird species: A multi-instance multi-label approach. *The Journal of the Acoustical Society of America*, 131(6):4640–4650, 2012.
- [9] Stephen T Buckland, Stuart J Marsden, and Rhys E Green. Estimating bird abundance: making methods work. *Bird Conservation International*, 18(S1):S91–S108, 2008.
- [10] Jia Deng, Jan Krause, and Li Fei-Fei. Fine-grained crowdsourcing for fine-grained recognition. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 580–587. IEEE, 2013.
- [11] Ji-Xiang Du, Xiao-Feng Wang, and Guo-Jun Zhang. Leaf shape based plant species recognition. *Applied mathematics and computation*, 185(2):883–893, 2007.
- [12] CA Duberstein, DJ Virden, S Matzner, J Myers, VI Cullinan, and AR Maxwell. Automated thermal image processing for detection and classification of birds and bats. 2012.
- [13] Alison Johnston, Aonghais SCP Cook, Lucy J Wright, Elizabeth M Humphreys, and Niall HK Burton. Modelling flight heights of marine birds to more accurately assess collision risk with offshore wind turbines. *Journal of Applied Ecology*, 51(1):31–41, 2014.
- [14] Alexis Joly, Hervé Goëau, Hervé Glotin, Concetto Spampinato, Pierre Bonnet, Willem-Pier Vellinga, Robert Planque, Andreas Rauber, Robert Fisher, and Henning Müller. Lifeclef 2014: multimedia life species identification challenges. In *Information Access Evaluation. Multilinguality, Multimodality, and Interaction*, pages 229–249. Springer, 2014.
- [15] Ljubica Lazarevic, David Harrison, Darren Southee, Max Wade, and John Osmond. Wind farm and fauna interaction: detecting bird and bat wing beats through cyclic motion analysis. *International Journal of Sustainable Engineering*, 1(1):60–68, 2008.
- [16] Xi Li, Weiming Hu, and Wei Hu. A coarse-to-fine strategy for vehicle motion trajectory clustering. In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, volume 1, pages 591–594. IEEE, 2006.
- [17] Marcelo T Lopes, Lucas L Gioppo, Thiago T Higushi, Celso AA Kaestner, Carlos N Silla Jr, and Alessandro L Koerich. Automatic bird species identification for large number of species. In *Multimedia (ISM), 2011 IEEE International Symposium on*, pages 117–122. IEEE, 2011.
- [18] Andréia Marini, Jacques Facon, and Alessandro L Koerich. Bird species classification based on color features. In *Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on*, pages 4336–4341. IEEE, 2013.

- [19] Chi-Man Pun and Moon-Chuen Lee. Log-polar wavelet energy signatures for rotation and scale invariant texture classification. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(5):590–603, 2003.
- [20] Muhammad Nawaz Rajpar and Mohamed Zakaria. Bird species abundance and their correlation with microclimate and habitat variables at natural wetland reserve, peninsular malaysia. *International Journal of Zoology*, 2011, 2011.
- [21] Szabolcs Sergyan. Color histogram features based image classification in content-based image retrieval systems. In *Applied Machine Intelligence and Informatics, 2008. SAMI 2008. 6th International Symposium on*, pages 221–224. IEEE, 2008.
- [22] Carlos N Silla, Celso Kaestner, et al. Hierarchical classification of bird species using their audio recorded songs. In *Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on*, pages 1895–1900. IEEE, 2013.
- [23] Concetto Spampinato, Daniela Giordano, Roberto Di Salvo, Yun-Heh Jessica Chen-Burger, Robert Bob Fisher, and Gayathri Nadarajan. Automatic fish classification for underwater species behavior understanding. In *Proceedings of the first ACM international workshop on Analysis and retrieval of tracked events and motion in imagery streams*, pages 45–50. ACM, 2010.
- [24] Satoshi Suzuki et al. Topological structural analysis of digitized binary images by border following. *Computer Vision, Graphics, and Image Processing*, 30(1):32–46, 1985.
- [25] Lee N Tan, Abeer Alwan, George Kossan, Martin L Cody, and Charles E Taylor. Dynamic time warping and sparse representation classification for birdsong phrase classification using limited training data. *The Journal of the Acoustical Society of America*, 137(3):1069–1080, 2015.
- [26] Catherine Wah, Steve Branson, Pietro Perona, and Serge Belongie. Multiclass recognition and part localization with humans in the loop. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 2524–2531. IEEE, 2011.
- [27] Catherine Wah, Steve Branson, Peter Welinder, Pietro Perona, and Serge Belongie. The caltech-ucsd birds-200-2011 dataset. 2011.
- [28] Peter Welinder, Steve Branson, Takeshi Mita, Catherine Wah, Florian Schroff, Serge Belongie, and Pietro Perona. Caltech-ucsd birds 200. 2010.
- [29] Bangpeng Yao, Gary Bradski, and Li Fei-Fei. A codebook-free and annotation-free approach for fine-grained image categorization. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 3466–3473. IEEE, 2012.
- [30] Ning Zhang, Ryan Farrell, Forrest Iandola, and Trevor Darrell. Deformable part descriptors for fine-grained recognition and attribute prediction. In *Computer Vision (ICCV), 2013 IEEE International Conference on*, pages 729–736. IEEE, 2013.
- [31] Zoran Zivkovic and Ferdinand van der Heijden. Efficient adaptive density estimation per image pixel for the task of background subtraction. *Pattern recognition letters*, 27(7):773–780, 2006.