Automatic Bird Species Detection Using Periodicity of Salient Extremities

Wen Li and Dezhen Song

***Abstract*— To assist nature observation, we develop an automatic bird species filtering method that takes videos from cameras with unknown parameters as input, and outputs likelihood of candidate species. The method recognizes the time series of salient extremities, which is the inter-wing tip distance, performs frequency analysis on periodicity, and provides a species prediction metric using likelihood ratios. To analyze the feasibility of the proposed method, we derive the probability that the salient extremity can be recognized in image for an arbitrary camera perspective. We also prove that the periodicity of the IWTD in the image is the same as the wingbeat frequency in the 3D space regardless of camera parameters with the exception of ignorable degenerated cases. Experiment results validate our analysis and show that the algorithm is very robust to segmentation error and data loss up to 30%.**

I. Introduction

Our group develops automation systems and algorithms to help ornithologists study local bird range change in South Texas that may be caused by climate change. To classify massive amount of video data, automating the detection of bird species becomes necessary and important. Given that bird videos may be taken by untrained amateurs using unknown cameras under different lighting and background conditions, accurate detection of exact bird species is difficult. We need to be able to filter bird species by reducing a potentially large candidate species set (e.g. more than 30) to a short list of bird species (e.g. 3-5 species).

Since most videos containing a flying bird are taken at far field under different lighting conditions, color and texture information is unreliable. To deal with the challenges, we develop a species filtering method using the periodicity of salient extremities (SE) for objects with a dominating body dimension that possesses periodic motion properties. For most birds, the measure for SE is the inter-wing tip distance (IWTD) whose periodic motion is often characterized by wingbeat frequency (WF). WF is a reliable and distinguishable feature for bird species filtering.

The contributions of the paper are threefold: First, we present a method to recognize the SE from videos and derive the probability that the SE can be recognized in image frame for arbitrary camera perspectives. Second, we model the body-wing structure of a bird using a 3 degrees-of-freedom (DOFs) kinematics model. We prove that the periodicity in SE (i.e. IWTD) is determined by WF. The periodicity is invariant to camera parameters. The two results allow us to develop an algorithm to extract IWTD series (see Fig. 1) out of video frames and obtain WF by applying Fast Fourier Transformation (FFT) to the IWTD series. Last, we propose a likelihood ratio-based species prediction metric using the resulting WF and its uncertainty range. The resulting algorithm returns a short ranked list of candidate species. We have implemented the bird species detection algorithm. Experimental results are satisfying and the algorithm is also very robust to data loss: it is capable of overcoming up to 30% of data loss in the tests.

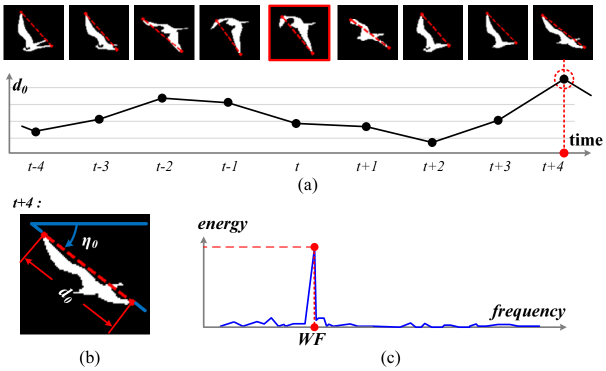


Fig. 1. Recognizing salient extremities: (a) IWTD varies periodically according to WF. (b) IWTD is extracted as the primary feature. (c) WF is obtained through FFT.

II. Related Work

Our automatic bird species detection method is based on the analysis of the periodicity of the SE of the object. As an active research area [1], [2], periodic motion (PM) analysis provides clues to many vision problems, such as tracking and segmentation [3], single view 3D reconstruction [4], [5], and pedestrian detection [6]. Our method extends existing recognition problems to a new domain: bird species recognition.

PM detection is nontrivial, and methods can be very different due to various camera settings and motion assumptions. Previous works can be classified into categories according to feature correspondence types. *Point correspondence* is used to estimate motion trajectories in [7], [8]. However, as stated in [9], feature correspondence estimation is sensitive to illumination changes, reflectance, and especially occlusion. *Template based methods* are proposed in [10], [11], which serve well in motion capture and tracking applications for humans or animals. However, template based methods usually suffer high computational cost due to large searching and scaling space in the matching process.

Region correspondence based methods are introduced by Polana and Nelson [12], and further extended by Cutler and

W. Li and D. Song are with CSE Department, Texas A&M University, College Station, TX 77843, USA, (Emails: {wli, dzsong}@cse.tamu.edu). D. Song is also with SIAT, Chinese Academy of Science, Shenzhen, Guangdong Province, P. R. China as a visiting scientist.

Davis [13]. These works assume that the object with repetitive motion should appear similar with its corresponding phrase in every period, and use a "similarity plot" to find period. These methods have certain robustness to image blurring and small background motion. However, they require 1) translation and scaling preprocessing, 2) very small changing of background texture, and 3) stable viewing perspective. Some also rely on linear motion trajectory. Briassouli and Ahuja [9] avoid the translation and scaling by projecting images into 1D signals and analyzing the short term time-frequency distribution. However, their experiments do not show robustness to perspective changes, and the stationary camera assumption limits background motion.

Under a different application context, our work has to deal with an arbitrary moving camera and a free flying object, thus the viewing angle and trajectory are both subject to significant changes. We analyze the motion periodicity by tracking the movement of SE, which in turn helps to avoid the stationary background requirement. Our feature analysis in frequency domain does not require pre-translation, rescaling or constant viewing angle. It is also worth noting that frequency-based methods are very robust to segmentation error. Existing results, such as [14], show that the periodic frequency still can be extracted from the frequency spectrum even under small ($-10 \sim 10$ dB) signal-to-noise ratio.

Our group has developed systems and algorithms for networked robotic cameras in nature observation applications [15]–[18]. Our previous work on bird species prediction [19] utilizes the bird body length and is limited to stationary camera with known parameters. This work extends our previous study to more general camera/scene settings.

III. PROBLEM DESCRIPTION

The input of the system is a sequence of video frames. The output of the system is a list of candidate species, which is ranked from the most to the least likely.

A. Assumptions and Prior Knowledge

We assume that the bird in the video is in steady flight under normal weather, which includes gliding, circling, cruising and level-flight, but excludes landing and taking off. Also, wing flapping motion should exist in the video. We assume that only one bird appears in the motion sequence. If multiple birds appear, we can apply existing multiple target tracking methods, such as [20], to separate individual bird sequence beforehand. The camera frame rate should be at least twice of the WF according to The Nyquist-Shannon sampling theorem. Since WFs of most bird species are lower than 15 Hz, a normal camera with 30 frames per second (fps) works for most cases.

TABLE I

PRIOR KNOWLEDGE OF BIRD WFS. s IS SPECIES ID, AND μ AND σ ARE THE MEAN AND THE STANDARD DEVIATION OF THE WF, RESPECTIVELY.

s	μ (Hz)	σ (Hz)	Species
6	3.18	0.227	Kittiwake
8	3.05	0.129	Herring Gull
12	4.58	0.183	Fulmar
...	...	...	...

We use the WF tables in [21], [22] as the prior knowledge. The tables are obtained by experts' manually counting of continuous flapping motion (See Tab. I for a few examples).

B. Problem Definition

Denote $d(t)$ to be the IWTD at time/frame t in pixel coordinates. Define N_s as the number of candidate species in the prior information, $\mathcal{S} = \{1, ..., N_s\}$ the candidate species set, and $L'(\cdot|\cdot)$ the likelihood that a bird with WF f_0 and WF error bound f_e belongs to species s. The bird species recognition problem can be defined as two sub problems,

Definition 1 (Extraction of Salient Extremities): Given a bird flying image sequence, extract time series $d(t)$.

Definition 2 (Species Prediction): Given $d(t)$ and the candidate set $\{\{\mu_s, \sigma_s\}, s = 1, ..., N_s\}$, estimate f_0, f_e, and compute $L'(\mu_s, \sigma_s|f_0, f_e), \forall s \in \mathcal{S}$.

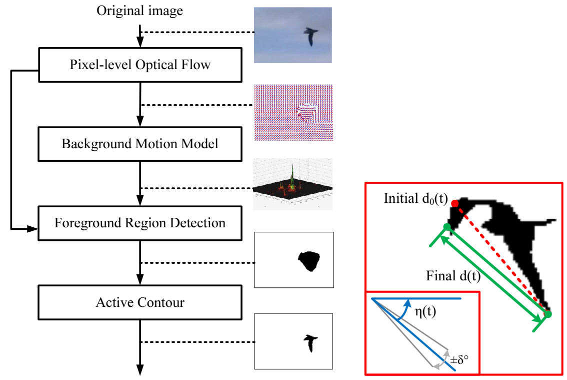
Let us begin with the first problem.

IV. EXTRACTION OF SALIENT EXTREMITIES

The extraction of SE has two steps: 1) motion segmentation that extracts the bird boundary from every frame, and 2) recognizing IWTD from bird boundaries.

A. Motion Segmentation

Since a flying bird is highly dynamic in appearance and shape, and camera motion is unknown, many segmentation methods are not applicable. We propose an unsupervised method for motion segmentation. Fig. 2(a) illustrates the four-step process. For each image frame, optical flow algorithm [23] is applied to calculate the flow on each pixel. Since background pixels share a similar motion pattern, a background motion model is estimated by iteratively minimizing the covariance of a 2D Gaussian distribution [24]. The Mahalanobis distance between a flow vector and the background model is measured. For those distances that fall out of a flexible quantile [25] of the χ^2 distribution, we label their corresponding pixels as foreground. Active Contour algorithm [26], [27] is then applied to generate a smooth boundary of the foreground area.



(a) Motion Segmentation (b) Searching for IWTD

Fig. 2. (a) A block diagram of motion segmentation. Thumbnails to the right of the block diagram indicate intermediate results. Black pixels in last two thumbnails indicate labeled foreground. (b) Searching for IWTD using WSD $\eta(t)$. The initial $d_0(t)$ is corrected by searching for $d(t)$ in the δ-neighborhood of $\eta(t)$.

B. Recognizing Salient Extremities

With the bird boundary extracted, we search for the IWTD. Define L_W for IWTD and L_B for bird body length in 3D. The corresponding notations in the image coordinate system are l_W and l_B, respectively. Recognizing IWTD in images is nontrivial because camera relative perspectives to the bird are unknown and may change from time to time. We cannot identify the SE by simply looking for the longest distance on the bird boundary in an arbitrary frame.

1) Finding the maximum IWTD across frames in a wingbeat period: If the video length is longer than a wingbeat period, the moment the bird fully extends its wings should exist in the video. The moment offers the best opportunity to recognize IWTD. In fact, we can derive the following lower bound for the probability that the IWTD is the longest distance on the bird contour.

Lemma 1: The lower bound of the probability that the IWTD is the longest distance on the bird boundary in image across k wingbeat periods with independent camera perspectives is $1 - (\frac{2}{\pi}\arctan(\frac{L_B}{L_W}))^k$.

The proof is elaborated in the online technical report [28]. For k wingbeat periods with independent perspectives, if $l_W > l_B$ holds in at least one period then we can obtain correct IWTD in the image. In fact, according to [29], the ratio L_W/L_B is larger than 1.09 for all species in the book. That means using 2 independent wingbeat periods will achieve at least a successful rate of 0.777.

It is also worth noting that this probability lower bound in Lem. 1 is not a tight bound. From experiments, we find that one wingbeat period is sufficient for extracting IWTD for a majority of bird species.

Lem. 1 suggests that we can search IWTD across frames to find the frame that wings are fully extended. Let l_{ij} be the Euclidean distance between two boundary points i and j. For a frame t, we first extract an initial IWTD:

$$d_0(t) = \max_{1 \le i,j \le m(t)} l_{ij}(t) \tag{1}$$

where $m(t)$ is the index set of boundary points. Its orientation $\eta_0(t)$ can be trivially computed. Fig. 1(a) shows examples of $d_0(t)$'s for a 9-frame sequence. Then,

$$d_{max}(t) = \max_{-\Delta \le i \le \Delta} d_0(t+i) \tag{2}$$

is extracted to be the IWTD for the moment that wings are fully extended in the period centered at frame t. Δ has a lower bound $\Delta \ge \frac{r}{2f_0} - \frac{1}{2}$ which ensures the sequence with frame rate r covers at least a period for the target species.

2) Recognizing IWTD series for the entire period: We introduce wing spreading direction (WSD) to describe the direction along which IWTD is to be extracted. WSD is represented by its tilting angle, denoted as $\eta(t)$. For a single period, WSD is viewed as a constant. Therefore, we can obtain WSD for frame t by computing the angle of $d_{max}(t)$. In the example shown in Fig. 1(a), frame $t+4$ has the maximum $d_0(t)$. Hence $\eta(t)$ is assigned by $\eta_0(t+4)$.

With WSD obtained, we can search for IWTDs. Since IWTD is the distance between extreme points on the bird, it should correspond to the longest distance between boundary points along the WSD in each frame (see Fig. 2(b)). On the other hand, the actual WSD on each frame may be slightly different from obtained WSD due to the discretization error introduced by the limited frame rate, and the small changes in relative camera perspectives. Therefore, $d(t)$ is obtained by searching a δ-neighborhood of the obtained WSD:

$$d(t) = \max_{|\varphi_{ij}(t) - \eta(t)| < \delta} l_{ij}(t) \tag{3}$$

where δ is a pre-set small threshold of angular difference. δ is selected to cover the aforementioned discrepancy. In our experiment, WSD searching range δ is set to $5°$. It is worth noting that this procedure, to some extent, overcomes the self-occlusion problem when one of the wing tip is occluded by the bird body.

V. Periodicity Analysis

We show that $d(t)$ shares the same periodic property of the wingbeat motion regardless of camera parameters, so that a frequency analysis can be conducted. We begin with a kinematic model of the bird wing.

A. Kinematic Modeling of Bird Wings

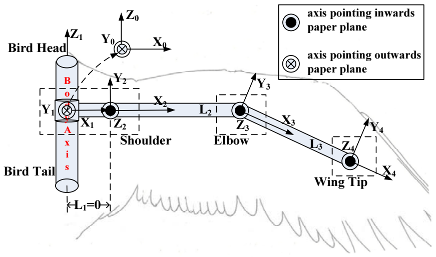


Fig. 3. A kinematic model of the right wing of a bird.

Following the steady-flight skeleton model in [30], we model a bird wing using three revolute joints. Frame 0 is the bird coordinate system (BCS) with its origin attached to the intersection of wing and body axis, and its Z-axis pointing to the direction of the bird head. Other frames are assigned by following Denavit-Hartenberg notations in [31], see Fig. 3.

This model has 3 DOFs: joint angles θ_1 and θ_2 at the shoulder and θ_3 at the elbow. The lengths of upper- and fore- arms are L_2 and L_3, respectively. The coordinate of right wing tip in frame 4 is $[0,0,0,1]^T$ in the homogeneous form. Applying the forward kinematics [31] to transform coordinates from frame 4 to frame 0, we have

$$\mathbf{X}_{rw} = \begin{bmatrix} L_2 c\theta_1 c\theta_2 + L_3 c\theta_1 c(\theta_2+\theta_3) \\ L_2 s\theta_1 c\theta_2 + L_3 s\theta_1 c(\theta_2+\theta_3) \\ L_2 s\theta_2 + L_3 s(\theta_2+\theta_3) \\ 1 \end{bmatrix}, \tag{4}$$

where $c\theta$ means $\cos\theta$, $s\theta$ means $\sin\theta$, $c(\cdot)$ means $\cos(\cdot)$, and $s(\cdot)$ means $\sin(\cdot)$. Symmetrically, we can obtain left wing tip $\mathbf{X}_{lw}$ in BCS, which is the same as $\mathbf{X}_{rw}$ except that the first element is negative. Therefore, the IWTD in 3D space is

$$D = 2(L_2 c\theta_1 c\theta_2 + L_3 c\theta_1 c(\theta_2+\theta_3)). \tag{5}$$

Since the distance from a flying bird to the camera is always significantly larger than the bird size, we approximate the perspective projection using an affine camera model. The camera transformation can be written as a 3×4 matrix P with its last row as $[0, 0, 0, 1]$.

Let $\mathbf{x}_{rw} := P\mathbf{X}_{rw}$ and $\mathbf{x}_{lw} := P\mathbf{X}_{lw}$ be right and left wing tip positions in the image, respectively. Recalling that $d = \mathbf{x}_{rw} - \mathbf{x}_{lw}$ is the distance between them, we have

$$d = 2(L_2 c\theta_1 c\theta_2 + L_3 c\theta_1 c(\theta_2 + \theta_3))\|\mathbf{p_1}\|_2 = D\|\mathbf{p_1}\|_2, \quad (6)$$

where $\mathbf{p_1}$ is the first column of P. Next we will show that d is a periodic function and reflects the WF.

B. Periodicity Analysis

In steady flight, a bird flaps its wings in a periodic pattern. Denote the period length as τ_0 and the corresponding circular frequency as ω_0. Pennycuick [21] shows that τ_0 and ω_0 are constants in steady flight. Liu et al. [30] show that all joint angle $\theta_i(t)$'s are periodic functions and can be expressed by a Fourier series,

$$\theta_i(t) = \alpha_i + \beta_i \sin(\omega_0 t + \phi_{i1}) + \gamma_i \sin(2\omega_0 t + \phi_{i2}), \quad (7)$$

where α_i, β_i, γ_i, ϕ_{i1}, and ϕ_{i2} are constants for $i = 1, 2, 3$. Since we only care about the basic WF (ω_0), we drop the harmonic frequency component in the last component and simplify (7) to the following,

$$\theta_i(t) = \alpha_i + \beta_i \sin(\omega_0 t + \phi_i). \quad (8)$$

Considering the geometric constraints and limits on wing joints, we know $\alpha_i \in [-\pi, \pi], \beta_i \in (0, \pi/2]$.

Let τ_d be the period length of $D(t)$, we have the following.

Theorem 1: For a bird in steady flight, the IWTD, $D(t)$, is a periodic function sharing the same period length of the wingbeat motion $\tau_d = \tau_0$ except that $\tau_d = \frac{1}{2}\tau_0$ if the following logic expression is true

$$(\alpha_1 + \alpha_2 = k\pi) \cdot (\alpha_1 - \alpha_2 = k\pi) \cdot (\alpha_3 = k\pi),$$

where $k \in \mathcal{Z}$ and '$\cdot$' is 'AND' operator.
The proof is detailed in online technical report [28].

Remark 1: For a fixed camera w.r.t the bird, the projective matrix does not change. Therefore, $\|\mathbf{p_1}\|_2$ remains constant and $d(t)$ share the period length with $D(t)$ based on (6).

Remark 2: If the camera or the bird moves, the changing of perspective introduces the frequency distribution of $\|\mathbf{p_1}\|_2(t)$, and the frequency property of $d(t)$ should be the convolution of the bird motion and the camera motion. As long as the changing of the camera perspective is not strictly periodic, the convolution preserves the dominant frequency component [5] of wing flapping motions except a few isolated special degenerate cases. This ensures that we can obtain WF f_0 by applying FFT to the extracted $d(t)$.

Actually, camera motions are usually slow when people track a bird at a distance. Most birds have a WF significantly higher than 1 Hz. Using a high pass filter of 1 Hz, we filter out the noise introduced by relative camera motion while preserving WF. Next we extract WF by 1) finding the frequency f_0 with the highest energy and 2) resetting $f_0 = f_0/2$ if there exists another peak at $f_0/2$. The reason is that the harmonic frequency at $2f_0$ sometimes dominates the fundamental frequency due to the second term in (7). Fig. 1(c) shows the extracted WF and the frequency distribution of the signal from video in Fig. 1(a).

VI. Species Prediction

Due to noise and discreteness, we perform a variance-based error analysis before the actual species detection with trustable measurements.

Step 1: Error Bound Analysis: The extracted WF has an error bound (f_e) equal to the half of the frequency interval after FFT, $f_e = \frac{r}{2N}$, where N is total number of frames rounded up to a power of 2, and r is the frame rate. Intuitively, the more frames exist, the smaller error can be achieved. Since the extracted WF is uniformly distributed, the variance of the extracted WF is

$$Var(f_0) = \frac{1}{12}((f_0 + f_e) - (f_0 - f_e))^2 = \frac{1}{3}f_e^2 \quad (9)$$

For a known species s, its reference WF from the prior knowledge has a variance σ_s^2. We believe that a measured WF is reliable only if its variance is less than that of the reference:

Definition 3 (Error Bound for Measurements): An extracted WF measurement is trustable for species prediction if $\frac{1}{3}f_e^2 \leq \sigma_s^2$.

The species prediction is only performed on trustable measurements. The least number of frames for a fixed rate video can be calculated inversely. For example, 100 frames approximately result in a measurement variance of 0.1 Hz for a 30 fps video, which is comparable to that of most species.

Step 2: Species Prediction: Had f_0 been error-free, the likelihood that the bird belongs to a species $\{\mu_s, \sigma_s\}$ is

$$L(\mu_s, \sigma_s | f_0) = \frac{1}{\sqrt{2\pi\sigma_s^2}} e^{-\frac{(f_0 - \mu_s)^2}{2\sigma_s^2}}. \quad (10)$$

However, the true WF is uniformly distributed in $(f_0 - f_e, f_0 + f_e)$, the likelihood function becomes

$$L'(\mu_s, \sigma_s | f_0, f_e) = \int_{f_0 - f_e}^{f_0 + f_e} \frac{1}{2f_e} L(\mu_s, \sigma_s | f) df. \quad (11)$$

Define $G(\cdot)$ as the cumulative probability function for the Gaussian distribution. Then we have,

$$L'(\mu_s, \sigma_s | f_0, f_e) = \frac{1}{2f_e}[G(\frac{f_0 + f_e - \mu_s}{\sigma_s\sqrt{2}}) - G(\frac{f_0 - f_e - \mu_s}{\sigma_s\sqrt{2}})]. \quad (12)$$

As the metric for species prediction, the likelihood is used to rank all candidate species. The resulting ranked list is the species prediction outcome. The reason for keeping a short candidate list instead of reporting only the top ranked candidate is that some species share close WF distributions, and it is not desired to miss many false negative predictions.

VII. Experiments

We have implemented the proposed bird filtering algorithm using Matlab on a PC. The *prior knowledge* (extended version of Tab. I) from [22] contains WF means and variances for 32 different species of birds. Their WF means vary from 2.24 Hz to 9.19 Hz. Since there is no existing video data set to benchmark and compare bird species recognition methods, we collect our data from online video. Original videos are downloaded from YouTube and Internet Bird Collection (`http://ibc.lynxeds.com/`). All videos are recorded by moving cameras. The collected dataset contains 18 video clips of different flying birds, covering 6 species in [22]. The video dataset consists of 378 flying periods which consists of 4269 video frames. Frame-rates of the videos vary from 15 fps to 30 fps. The IWTDs of the birds in the video range from 105 cm to 229 cm while WFs range from 2.24 to 4.58 Hz. It is worth noting that this WF range covers a majority of bird species ($> 60\%$) which makes it a challenging data set because there are many overlapping WFs among species.

1) IWTD and WF Extraction: Our algorithm successfully extracts IWTD series, their WFs, and their WF error bounds. In fact, we only need one period to recognize IWTD and obtain IWTD series for WF extraction, which agrees with the prediction given by Lemma 1. Fig. 4 shows that the extract WFs are mostly covered in 2σ of the true species WF distribution, and therefore lead to high likelihood of true species, except Video 7. The results validate that the system is capable of extracting WFs from different camera perspectives, and shows that WF is a stable signature for the species recognition.

TABLE II

RoCS for testing videos with different MLR.

video	1	2	3	4	5	6
MLR	0.096	0	0.1485	0.1094	0.0609	0.128
RoCS	2	4	2	5	2	2
video	7	8	9	10	11	12
MLR	0.1667	0.1	0.0472	0.1714	0.0418	0.171
RoCS	8	8	1	1	1	9
video	13	14	15	16	17	18
MLR	0.2904	0.2467	0.0615	0.2541	0.2577	0.3034
RoCS	2	2	6	9	2	3

2) Robustness to segmentation error: Since our method relies on the extraction of pixel distance, the temporal feature is inevitably affected by the segmentation error, especially when image resolution is low or motion blur appears. The error influences the accuracy of pixel distance $d(t)$. Considering the segmentation error at a wing tip to follow a zero-mean Gaussian distribution, the 2D IWTD follows Gaussian distribution as well. Simulation is designed on a real signal from test video 11 (Fig. 5), where we manually annotated the wing tip positions in every image. A sequence of $d(t)$ is therefore calculated upon the annotation and treated as a ground truth signal (fig. 5(a)). Mean value of this signal is subtracted for illustration purpose. The maximum and the minimum values in $d(t)$ are 154.1 and 60.5, respectively, while the mean is 108.82. Different levels of Gaussian noise are added. The red dotted curve in fig. 5(a) shows the simulated signal when the error standard deviation is 10. We gradually increase the noise and measure the ratio between the WF peak energy and the average spectrum energy (Fig. 5(b)). It is shown that with noise standard derivation from 0 to 100 pixels, the WF energy is still higher than average spectrum energy. While in our experiments in previous subsection, the mean segmentation error of this sequence is 4.12 pixels, and the maximum error in a frame is 37.06 pixels, which are much smaller than the simulated error. This simulation demonstrates the robustness of the proposed WF extraction method in the presence of segmentation errors.

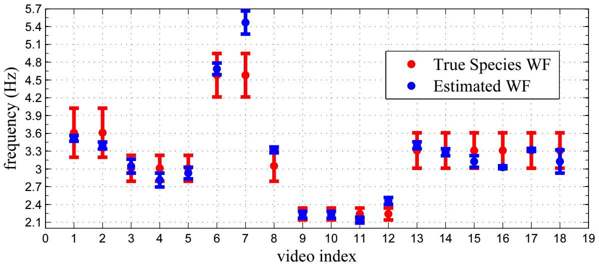
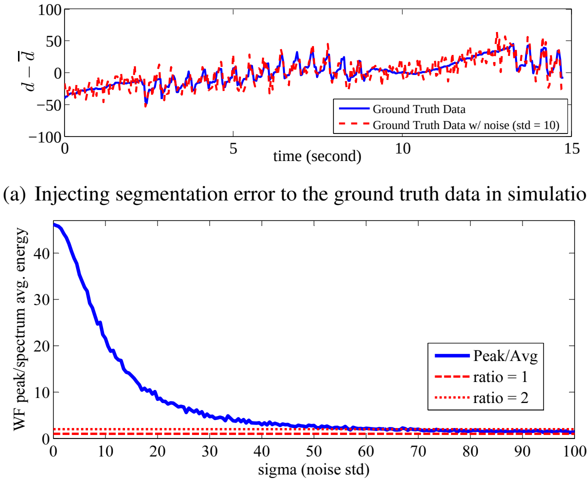


Fig. 4. Comparison between true species WF and extracted WF. Red bars shows the frequency covered in $\mu \pm 2\sigma$ of the true species, blue bars shows the frequency covered in the extracted $f_0 \pm f_e$ of the target bird.



(a) Injecting segmentation error to the ground truth data in simulation

(b) Signal energy vs. background energy

Fig. 5. Simulation results on the robustness of frequency analysis against segmentation errors. (a) Blue solid curve: the ground true of $d(t) - \bar{d}$. Red dotted curve: after adding Gaussian noise with zero mean and a standard deviation of 10 pixels to the blue curve. (b) The ratio of WF peak and the average of spectrum energy, as the noise deviation increases from 0 to 100. The ratio is always above 1 and is above 2 when noise deviation is lower than 55 pixels.

3) Species Prediction: To evaluate the accuracy of the ranked candidate list, we define hit rate as the percentage of returned candidate lists that contain the correct species. To our best knowledge, there is no existing algorithms for flying bird species recognition for videos taken by moving cameras. Previous methods on object recognition or motion analysis cannot directly applied on the bird species recognition problem. Therefore, the comparison experiment is compared with random guess only. We compare our algorithm output with a short list of the same length which is generated from independent random guesses from the 32 candidate species.

The results are showed in Fig. 6. It is clear that our algorithm significantly outperforms the random guess.

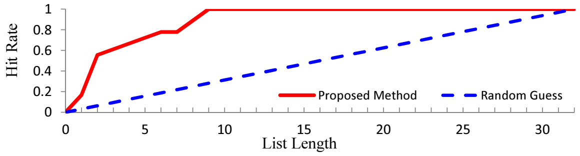


Fig. 6. Hit rate vs. list length.

4) Robustness to Data Loss: Inevitably, some frames of bird videos may be too blur to segment the bird which leads to the loss of IWTD measurements. If so, our system assigns the measurement of this frame using its nearest successful antecedent. Our frequency-based analysis is very robust to data loss. The measurement lost rate (MLR) in each testing video is listed in Tab. II. The loss rate varies from 0 to 30%. Even for the video with most data lost (video 18), the rank of the correct species (RoCS) is still among the top three.

VIII. Conclusion and Future Work

We developed the bird species filtering method that takes videos from unknown cameras as input and outputs likelihood of candidate species. The method extracted the time series of SE from the videos without prior knowledge on camera motion and perspective changes. We derived the probability that the SE can be recognized in the image frame for arbitrary camera perspectives. We also proved that the periodicity of the extracted SE series is generally the same as the WF in the 3D space. This allowed us to apply FFT to observed IWTD series to obtain WF. We also proposed a species prediction metric using likelihood ratios. We have implemented the algorithm and tested it in experiments which validated our design and analysis. In the future, we will develop recognition methods using other features such as flying speed and shape in combination with frequency signatures to achieve better prediction. Note that the method also has the potential to be applied to other animals with frequency characteristics.

Acknowledgments

Thanks for K. Goldberg and R. Volz for insightful discussions. Thanks for C. Kim, Y. Lu, M. Hielsberg, S. Guo, Z. Ma, and X. Li for their inputs and contributions to the Networked Robots Laboratory in Texas A&M University. Special thanks are given to Dr. J. Rappole from Smithsonian Institute for providing bird species data source.

References

[1] X. Ji and H. Liu, "Advances in view-invariant human motion analysis: A review," *IEEE Trans. Systems, Man, and Cybernetics*, vol. 40, no. 1, pp. 13–24, 2010.

[2] T. Moeslund, A. Hilton, and V. Krger, "A survey of advances in vision-based human motion capture and analysis," *Computer Vision and Image Understanding*, vol. 104, no. 2, pp. 90 –126, 2006.

[3] A. Briassouli and N. Ahuja, "Fusion of frequency and spatial domain information for motion analysis," in *ICPR*, vol. 2, Cambridge, UK, Augest 2004, pp. 175–178.

[4] S. Belongie and J. Wills, "Structure from periodic motion," in *Intl. Workshop on Spatial Coherence for Visual Motion Analysis*, Prague, Czech Republic, 2004.

[5] E. Ribnick and N. Papanikolopoulos, "3d reconstruction of periodic motion from a single view," *IJCV*, vol. 90, no. 1, pp. 28–44, October 2010.

[6] Y. Ran, I. Weiss, Q. Zheng, and L. Davis, "Pedestrian detection via periodic motion analysis," *IJCV*, vol. 71, no. 2, pp. 143–160, 2007.

[7] S. Seitz and C. Dyer, "View-invariant analysis of cyclic motion," *IJCV*, vol. 25, no. 3, pp. 231–251, 1997.

[8] I. Laptev, S. Belongie, P. Pérez, and J. Wills, "Periodic motion detection and segmentation via approximate sequence alignment," in *ICCV*, October 2005.

[9] A. Briassouli and N. Ahuja, "Extraction and analysis of multiple periodic motions in video sequences," *TPAMI*, vol. 29, no. 7, pp. 1244 – 1261, July 2007.

[10] D. Ormoneit, M. Black, T. Hastie, and H. Kjellström, "Representing cyclic human motion using functional analysis," *Image and Vision Computing*, vol. 23, no. 14, pp. 1264 – 1276, 2005.

[11] A. Bobick and J. Davis, "The recognition of human movement using temporal templates," *TPAMI*, vol. 23, no. 3, pp. 257 – 267, 2001.

[12] R. Polana and R. Nelson, "Detection and recognition of periodic, nonrigid motion," *IJCV*, vol. 23, no. 3, pp. 261–282, 1997.

[13] R. Cutler and L. Davis, "Robust real-time periodic motion detection, analysis, and applications," *TPAMI*, vol. 22, no. 8, pp. 781 – 796, 2000.

[14] T. Yoshizawa, S. Hirobayashi, and T. Misawa, "Noise reduction for periodic signals using high-resolution frequency analysis," 2011.

[15] D. Song and K. Goldberg, "Networked robotic cameras for collaborative observation of natural environments," in *The 12th International Symposium of Robotics Research (ISRR)*, San Francisco, CA, USA, Oct. 2005.

[16] D. Song, N. Qin, Y. Xu, C. Kim, D. Luneau, and K. Goldberg, "System and algorithms for an autonomous observatory assisting the search for the ivory-billed woodpecker," in *IEEE International Conference on Automation Science and Engineering (CASE)*, Washington DC, USA, Aug. 2008.

[17] S. Faridani, B. Lee, S. Glasscock, J. Rappole, D. Song, and K. Goldberg, "A networked telerobotic observatory for collaborative remote observation of avian activity and range change," in *The IFAC workshop on networked robots*, Golden, Colorado, USA, Oct. 2009.

[18] D. Song and Y. Xu, "A low false negative filter for detecting rare bird species from short video segments using a probable observation data set-based ekf method," in *Speciel Track on Physically Grounded AI(PGAI), the 24th AAAI Conference on Artificial Intelligence (AAAI-10)*, Atlanta, Georgia, USA, July 2010.

[19] ——, "A low false negative filter for detecting rare bird species from short video segments using a probable observation data set-based ekf method," *TIP*, vol. 19, no. 9, pp. 2321–2331, 2010.

[20] S. Blackman, "Multiple hypothesis tracking for multiple target tracking," vol. 19, no. 1, Jan. 2004.

[21] C. Pennycuick, "Wingbeat frequency of birds in steady cruising flight: New data and improved predictions," *The Journal of Experimental Biology*, vol. 199, pp. 1613–1618, 1996.

[22] ——, "Predicting wingbeat frequency and wavelength of birds," *The Journal of Experimental Biology*, vol. 150, pp. 171–185, 1990.

[23] C. Liu, "Beyond pixels: Exploring new representations and applications for motion analysis," Ph.D. dissertation, Massachusetts Institute of Technology, May 2009.

[24] P. J. Rousseeuw and A. M. Leroy, *Robust Regression and Outlier Detection*. New York: Wiley, 1987.

[25] P. Filzmoser, R. G. Garrett, and C. Reimann, "Multivariate outlier detection in exploration geochemistry," *Computer and Geosciences*, vol. 31, pp. 579–587, 2005.

[26] T. F. Chan and L. A. Vese, "Active contours without edges," *TIP*, vol. 10, no. 2, pp. 266–277, 2001.

[27] S. Lankton, "Sparse field methods - technical report," Georgia Institute of Technology, Tech. Rep., April 2009.

[28] W. Li and D. Song, "Automatic video-based bird species filtering using periodicity of salient extremities," Department of Computer Science and Engineering, Texas A&M University, Tech. Rep. TR2012-08-2, Aug. 2012. [Online]. Available: http://www.cse.tamu.edu/academics/tr/2012-8-2

[29] J. Dunn and J. Alderfer, *Field Guide to the Birds of Eastern North America*. National Geographic, Washington D. C., 2008.

[30] T. Liu, K. Kuykendoll, R. Rhew, and S. Jones, "Avian wing geometry and kinematics," *AIAA Journal*, vol. 44, no. 5, May 2006.

[31] J. Craig, *Introduction to Robotics Mechanics and Control (Third Edition)*. Pearson Education, Upper Saddle River, New Jersey, 2005.