

Automatic Video-based Bird Species Filtering Using Periodicity of Salient Extremities

Wen Li and Dezhen Song

Abstract

To assist nature observation, we develop a bird species filtering method that takes videos from cameras with unknown parameters as input and outputs likelihood of candidate species. The method can extract the time series of salient extremities, which is the inter-wing tip distance (IWTD), from the videos without assuming knowledge on camera motion and perspective changes. We also derive the probability that the salient extremity can be recognized in the 2D image frame for an arbitrary relative perspective between the camera and the bird. With the exception of ignorable degenerated cases, we also prove that the periodicity of the IWTD in the image is the same as the wingbeat frequency in the 3D space regardless camera parameters. This allows us to apply Fast Fourier Transformation to the observed IWTD series to obtain wingbeat frequency. We also propose a species prediction metric using likelihood ratios. We have implemented the algorithm and tested it in experiments using 18 video clips against 32 candidate bird species. Experimental results validate our analysis and show that the algorithm is very robust to segmentation error and data loss up to 30%. The algorithm achieves 61.1% detection rate with its ranked list length set to 3.

Index Terms

frequency analysis, salient extremities, periodic motion.

I. INTRODUCTION

Our group works on developing algorithms to assist nature observation. Recently, we help ornithologists study local bird range change in South Texas that may be caused by climate change. To classify massive amount of video data, detecting bird species becomes necessary and important. Given that bird videos may be taken by untrained amateurs using unknown cameras at different lighting and background conditions, accurate detection of exact bird species is difficult. Alternatively, we need 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).

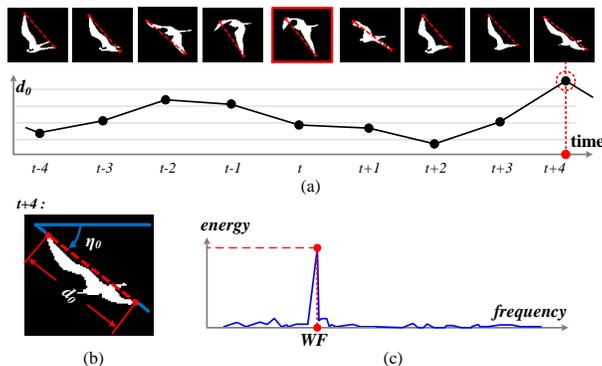


Fig. 1. Recognizing salient extremities: (a) IWTD varies periodically according to the WF. (b) IWTD is extracted as the primary feature. (c) WF is obtained through FFT.

Since most videos containing a flying bird are taken at far field under different lighting conditions, color and texture information becomes unreliable. In addition, camera parameters are often unknown and it is impossible to recover flying speed or other metric measurements. To deal with the challenges, we develop a species filtering method

W. Li and D. Song are with the Department of Computer Science and Engineering, Texas A&M University, College Station, TX, 77843 USA. E-mails: {wli, dzsong}@cse.tamu.edu

This work was supported in part by National Science Foundation under CAREER grant IIS-0643298.

using the periodicity of salient extremities for objects with a dominating body dimension that possesses periodic motion properties. For most birds, the measure for the salient extremities 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 extract the salient extremities from videos and derive the probability that the salient extremity can be recognized in the 2D image frame for an arbitrary relative perspective between the camera and the bird. We show that the probability is an increasing function of video data amount except ignorable degenerating cases. Second, we model the body-wing structure of a bird using a 3 degrees-of-freedom (DOFs) kinematics model. We formally prove that the periodicity in salient extremities (i.e. IWTD) in the image frame, is determined by the wingbeat frequency (WF) in the world frame. 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 ranked short candidate list of species.

We have implemented the bird species detection algorithm. The proposed method is evaluated on 18 video clips containing 6 different species. The detection process is to evaluate the videos against a 32 bird species candidate list. Experimental results show the proposed method achieves 61.11% detection rate when the short list length is 3. The algorithm is also tested with different simulated segmentation error levels and appears very robust to segmentation error. The algorithm is also very robust to data lost: it is capable of overcoming up to 30% of data lost in the tests.

II. RELATED WORK

Our bird species recognition method is based on the analysis of bird flying motion. As an active research area [1], [2], motion analysis mostly targets at two types of human motion: periodic or non-periodic. Our work belongs to the first type. Periodic motion (PM) analysis provides clues to many vision problems. Briassouli and Ahuja [3] fuse periodicity with spatial velocity for tracking and segmentation. Others [4]–[7] utilize PM to construct 3D information from a single view. It is also used in recognition problems, such as human/animal activity recognition [8], pedestrian detection [9], and human gesture recognition [10]. In these problems, many existing works classify different repeating patterns to distinguish between human activities. Others perform hypothesis test on the existence of periodic frequency, in order to recognize objects. 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 four categories according to feature correspondence types. *Point correspondence* is used in [11]–[13]. Feature points are extracted in frames and correspondences are estimated to form point motion trajectories. However, as stated in [14], feature correspondence estimation is sensitive to illumination changes, reflectance, and especially occlusion, therefore, point correspondence based methods are not generally applicable. *Template based methods* are proposed in [15], [16]. Skeleton models are often established to search for matches in image frames. Since the skeleton models capture the underlying bone structure, these methods 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 [17], and extended by Cutler and Davis [18], to capture periodicity of object locomotion. The basic assumption of these works is that the object with repetitive motion should appear similar with its corresponding phrase in every period. By measuring the similarity between the object in every two frames, a “similarity plot” is calculated. These methods have certain robustness to image blurring and small background motion. However, they require: 1) translation and scaling preprocessing to make the object stationary and of the same size in every frame; 2) small changing of background texture; and 3) viewing angle of the object does not change significantly. Some of them also rely on linear moving trajectory of the object.

As the fourth method, Briassouli and Ahuja [14] project the object regions onto two axes of the image coordinate, and analyze the two 1D intensity signals by short term time-frequency distribution. The approach smartly avoids the translation and scaling requirement and contributes in: 1) tracking and period extraction can be done simultaneously,

2) time-varying signals can be analyzed. However, its experiments do not show if the method can deal with changing of object viewing angles, 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 via tracking the varying of salient extremities of the object region. This helps to avoid stationary background requirement. Our feature analysis in frequency domain do not require pre-translation or rescaling. Since we do not calculate similarity plot, the restrictive consistent viewing angle is no longer needed.

It is also worth noting that frequency-based methods are very robust to segmentation error. There have been many results on the extraction and enhancement of periodicity in noisy background. Existing results [19] and [20] show that the periodic frequency still can be extracted from the frequency spectrum even under small ($-10 \sim 10$ dB) signal-to-noise ratio, with some noise reduction techniques. These results support our approach in the way that with the existent signal analysis tools, the segmentation error does not overwhelm the periodic signal in the frequency domain even when the periodic signal is weak.

Our group has developed systems and algorithms for networked robotic cameras of nature observation [21]–[24]. Our previous work of bird species prediction [25] utilizes the bird body length and deals with static camera and known parameters. With the new feature proposed, this work extends our previous study to more general camera/scene settings.

III. SYSTEM OVERVIEW AND PROBLEM DESCRIPTION

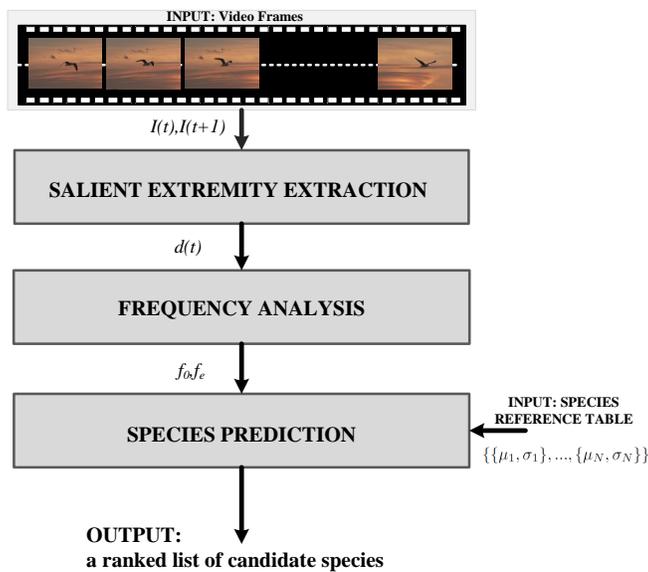


Fig. 2. System block diagram.

Let us start with a system overview before introducing our assumptions and problem definition.

A. System Overview

Fig. 2 illustrates system architecture including three main steps of the approach. The input of the system is a sequence of video frames. *Salient extremity extraction* first segments the bird region from every frame, and returns the boundary points of the bird regions. Then it searches for the IWTD in each frame, based on the extracted bird boundaries. This step will be elaborated in Section IV. In *periodicity analysis*, the IWTD values are analyzed in frequency space via FFT. This step focuses on the analysis of how the frequency signature is invariant to camera motions, perspective changes, and camera parameters. Details are elaborated in Section V. The extracted frequency is finally compared with every known species in the *species prediction* step. A list of candidate species of the input bird is generated and ranked from the most to the least possible one. See Section VI for details.

B. Assumptions and Prior Knowledge

To introduce our bird detection problem, we have the following assumptions.

- Only one flying bird appears in the motion sequence. If there are multiple birds in the video, we can segment them out using multiple hypothesis tracking techniques, such as [26], beforehand.
- The bird is in steady flight under normal weather, which includes gliding, soaring, circling, cruising and level-flight, but excludes landing and taking off.
- Wing flapping motion exists in the video.
- The camera frame rate should be at least two times of the WF [27]. Since WFs of most bird species are lower than 15 Hz, a normal camera with 30 frames per second (fps) works for most cases.

From ornithologists [28], [29], we have a table of WFs for candidate bird species (See Tab. I for a few examples). These WFs are obtained by manually counting the continuous flapping motion. We use this information as the prior knowledge for our algorithm. Note that 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
...

TABLE I
PRIOR KNOWLEDGE OF BIRD WFS.

C. Problem Definition

Define τ_0 as the period length of the wing flapping motion. Then the WF can be denoted as $f_0 = 1/\tau_0$, and the corresponding circular frequency is $\omega_0 = 2\pi f_0$. The error bound of f_0 is denoted as f_e . Define N_s as the number of candidate species in the prior information and $\mathcal{S} = \{1, \dots, N_s\}$ as the candidate specie set. $d(t)$ and $D(t)$ are denoted to be the IWTDs at time/frame t in pixel coordinates and in 3D space, respectively. Define $L'(\cdot|\cdot)$ as the likelihood that a bird with f_0 and 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 $\{\{\mu_s, \sigma_s\}, s = 1, \dots, N_s\}$, 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 salient extremities can be divided into two steps: 1) motion segmentation that extracts the foreground object (bird) boundary from every frame, and 2) recognizing IWTD from the bird boundary for each frame.

A. Motion Segmentation

This step takes the frame sequence as input and outputs the boundary of foreground object in each frame. Since a flying bird is highly dynamic in motion, appearance and shape, and the background is moving as well, many segmentation methods are not applicable. Here, we propose an unsupervised method for motion segmentation based on the optical flow technique. Fig. 3 illustrates the four-step process as detailed below.

- 1) For a frame $I(t)$ at time t (the topmost thumbnail in Fig. 3), we apply Liu's optical flow algorithm [30] to calculate the flow on each pixel, w.r.t. the subsequent frame $I(t+1)$. This algorithm is a combination of [31]

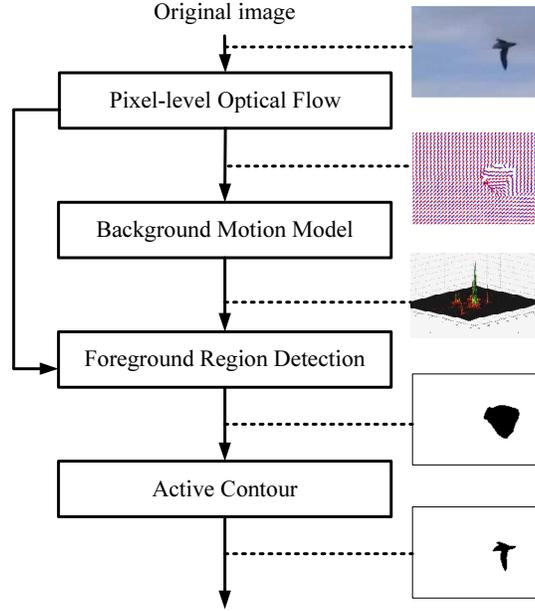


Fig. 3. A block diagram of motion segmentation. Note that thumbnails to the right of the block diagram indicate intermediate results. Black pixels in last two thumbnails indicate foreground which represents the bird region.

and [32], and returns a 2D flow vector $\mathbf{w}(x, y) := (u(x, y), v(x, y))^T$ for each pixel (x, y) . The red dots in the second thumbnail point to the optical flow direction.

- 2) Since background pixels share a similar motion pattern, which is different from that of the foreground pixels, we can model the background flow using a 2D Gaussian distribution (see the middle thumbnail in Fig. 3). We apply the Minimum Covariance Determinant Estimator [33] to obtain the estimated mean $\bar{\mathbf{w}}$ and covariance of the distribution with a 75% inlier ratio and an iteration number of 100.
- 3) The Mahalanobis distance between every flow vector \mathbf{w} and the background mean $\bar{\mathbf{w}}$ is then calculated. For those distances that fall out of a flexible quantile of the chi-square distribution, we label their corresponding pixels as foreground using [34] which provides an adaptive selection of the quantile. The second thumbnail from the bottom of Fig. 3 shows a result of this labeling process.
- 4) We apply an Active Contour algorithm [35], [36] to the labeled binary image (as an initial mask) to locally generate a smooth contour of the bird (the last thumbnail in Fig. 3). We further filter out false detections via size and color consistency constraints. Therefore, a set of boundary points of the bird region on each frame is obtained.

B. Recognizing Salient Extremities

With bird contour extracted, we can search for the salient extremities on the contour. Salient extremities refer to the longest dimension of the body in 3D Euclidean coordinate. For birds, this is IWTD, which is defined as L_W in the 3D coordinate. Let us define bird body length in 3D as L_B . Correspondingly, the IWTD and bird body length in the image coordinate system are defined as l_W and l_B , respectively. l_B may appear longer than l_W due to perspective distortions and different bird flying poses (Fig. 4). Recognizing salient extremities in image frames is nontrivial because camera relative perspectives to the bird are unknown and may change from time to time. We cannot identify the salient extremities by simply looking for the longest distance on the bird contour 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 when the flying bird fully extends the wing ought to exist in the video. At the moment, the chance that we can find IWTD is high. 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. *With a single period, the probability that the IWTD is the longest distance on the bird contour at the*

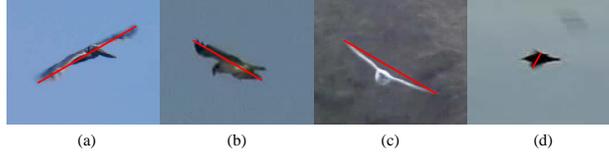


Fig. 4. Examples of fully extended bird body in different perspectives.

moment when the flying bird fully extends its wings is no less than $1 - \frac{2}{\pi} \arctan(\frac{L_B}{L_W})$ for an arbitrary camera perspective.

The proof of Lem. 1 is elaborated in Appendix A. We examined all the species listed in the ornithology reference book [37]. There are 71.0% of all the species that have the ratio L_W/L_B larger than 1.97. Their probability lower bound is 0.701 if we just simply search for the longest distance in bird contours across the single period. When more data are available, we have the following corollary.

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

This conclusion can be straightforwardly derived from Lem. 1. 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 [37], 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 0.777 successful rate.

Remark 1. *Corollary 1 shows a desirable trend that more data means high successful rate. It is worth noting that adjacent wingbeat periods usually have correlated perspectives. We find that most videos show that perspective change is quite significant during bird flying. Therefore, Corollary 1 still applies for periods providing sufficient time difference between periods. The only exception is when the bird flies around the camera on the same plane while maintaining the same distance to the camera, which is very rare.*

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 birds species.

Lem. 1 suggests that we can search IWTD across frames to find the frame when the bird fully extends its wing. Denote l_{ij} to be the Euclidean pixel distance between two boundary points i and j , and φ_{ij} to be the orientation of the vector pointing from i to j . For a particular frame t , we first extract the maximum value among l_{ij} 's, for all pairs of boundary points i and j in the frame. This distance value, as well as its corresponding angle, is regarded as an initial solution, denoted as $d_0(t)$ and $\eta_0(t)$:

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

where $m(t)$ is the index set of points of the bird boundary in frame t . $\eta_0(t)$ can be trivially computed when $d_0(t)$ is obtained. Fig. 1(a) shows examples of $d_0(t)$'s in red dashed lines for a 9-frame sequence.

Then, we consider a sequence of frames from $t - \Delta$ to $t + \Delta$, where Δ has a lower bound $\Delta \geq \frac{r}{2f_0} - \frac{1}{2}$ which ensures the sequence with frame rate r covers at least a period for the target species. The maximum value among $d_0(t - \Delta), \dots, d_0(t + \Delta)$ is extracted to be the IWTD for the moment that the bird fully extends its wing in the period.

2) *Recognizing IWTD series for the entire period:* When the bird does not fully extend its wing, then IWTD might not be the longest dimension in the bird contour. The bird thumbnails corresponding to time t and $t - 1$ in Fig. 1(a) illustrate the problem. To address the problem, we introduce wing spreading direction (WSD) as the direction of the line connecting the two wing tips. In the image space, WSD is represented by its tilting angle $\eta(t)$ w.r.t. horizontal axis at time/frame t . Basically, WSD describes the direction along which IWTD is extracted. For a single period, WSD can be viewed as a constant because the camera perspective change and bird translation movement are ignorable in the short wingbeat period. When we extract the maximum IWTD for the period, we can obtain its WSD correspondingly. This WSD can then be applied to other frames to search for IWTDs. In the example shown in Fig. 1(a), frame $t + 4$ has the maximum $d_0(t)$. Hence $\eta_0(t + 4)$ is the WSD in that period, and $\eta(t) = \eta_0(t + 4)$.

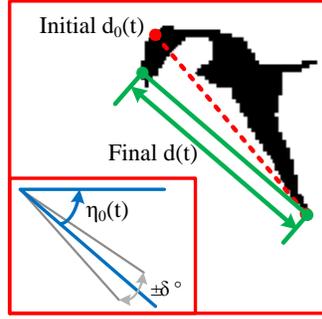


Fig. 5. Searching for IWTd using WSD $\eta_0(t)$. The initial $d_0(t)$, which found by searching for the longest distance between boundary points, is not the correct IWTd. This is corrected by searching for $d(t)$ in the δ -neighborhood of $\eta_0(t)$.

With WSD obtained, we can search for IWTds for the rest of frames. 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. 5). On the other hand, the actual WSD on each frame may be slightly different from WSD obtained from the maximum IWTd because the discrepancy caused by the discretization error of WSD due to the limited frame rate and by small changes in flying poses and camera perspectives exists. Therefore, $d(t)$ is obtained by searching a δ -neighborhood of the obtained WSD:

$$d(t) = \max_{|\varphi_{ij}(t) - \eta(t)| < \delta} l_{ij}(t) \quad (2)$$

where δ is a pre-set small threshold of angular difference. δ is selected to cover the aforementioned discrepancy. 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

Although $d(t)$ is obtained, it is unclear if $d(t)$ reflects the WF of the target species. We need to show that $d(t)$ shares the same periodic property of the wingbeat motion regardless of camera parameters. We begin with a kinematic model of the bird wing.

A. Kinematic Modeling of Bird Wings

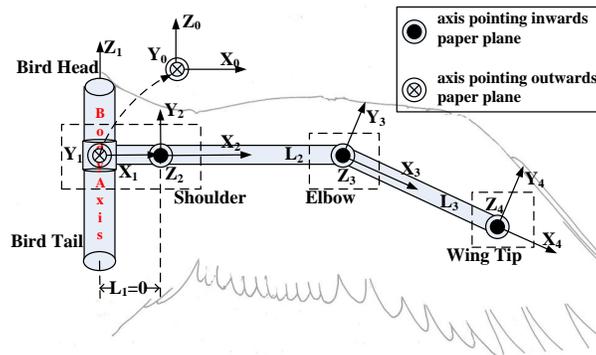


Fig. 6. A kinematic model of the right wing of a bird.

Following the steady-flight skeleton model in [38], we model a bird wing using three revolute joints (see Fig. 6). Frame 0 is the bird coordinate system (BCS) with its origin attached to the intersection point between the wing and the body axis of the bird and its Z -axis pointing to the direction of the bird head. Other frames are assigned by following Denavit-Hartenberg notations in [39], see Fig. 6.

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 [39] to transform coordinates from frame 4 to frame 0, we have

$$\mathbf{X}_{rw} = \begin{bmatrix} L_2 c\theta_1 c\theta_2 + L_3 c\theta_1 c(\theta_2 + \theta_3) \\ L_2 s\theta_1 c\theta_2 + L_3 s\theta_1 c(\theta_2 + \theta_3) \\ L_2 s\theta_2 + L_3 s(\theta_2 + \theta_3) \\ 1 \end{bmatrix}, \quad (3)$$

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

$$D = 2(L_2 c\theta_1 c\theta_2 + L_3 c\theta_1 c(\theta_2 + \theta_3)). \quad (4)$$

Now let us project D into the image coordinate. Since the distance from a flying bird to the camera is always significantly larger than the bird size, we can approximate the perspective projection using an affine camera model. Then, 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

$$d = 2(L_2 c\theta_1 c\theta_2 + L_3 c\theta_1 c(\theta_2 + \theta_3)) \|\mathbf{p}_1\|_2 = D \|\mathbf{p}_1\|_2, \quad (5)$$

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. Recall that the period length is τ_0 and the corresponding circular frequency is ω_0 in Section III. Pennycuick [28] shows that τ_0 and ω_0 are constants in steady flight. Liu et al. [38] show that all joint angle $\theta_i(t)$'s are periodic functions and can be expressed by a Fourier series,

$$\theta_i(t) = \alpha_i + \beta_i \sin(\omega_0 t + \phi_{i1}) + \gamma_i \sin(2\omega_0 t + \phi_{i2}), \quad (6)$$

where $\alpha_i, \beta_i, \gamma_i, \phi_{i1}$, and ϕ_{i2} are constants for $i = 1, 2, 3$. α_i 's are phases. Since we only care about the basic WF (ω_0), we drop the harmonic frequency component in the last component. This dropping does not change the WF peak in the FFT of $d(t)$ later. Thus, we can simplify (6) to the following,

$$\theta_i(t) = \alpha_i + \beta_i \sin(\omega_0 t + \phi_i). \quad (7)$$

Considering the geometric constraints and limits on wing joints, we know $\alpha_i \in [-\pi, \pi], \beta_i \in (0, \pi/2]$. We are interested in whether (5) reveals the same period length τ_0 . Before proving it, we have the following two lemmas about the period length of two intermediate functions that will be used to compute the period length of d later. The first intermediate function is defined as $f(t) = \cos(\alpha + \beta \sin(\omega t + \phi))$, where $\sin(\omega t + \phi)$ is a periodic function with period length $\tau = 2\pi/\omega$, $\beta \in (0, \pi/2]$. The first lemma is,

Lemma 2. *Function $f(t)$ is a periodic function with the following period length*

$$\tau_f = \begin{cases} \tau & \text{if } \alpha \neq k\pi \\ \frac{1}{2}\tau & \text{if } \alpha = k\pi \end{cases} \quad (8)$$

where $k \in \mathcal{Z}$, the integer set.

The proof of Lem. 2 is in Appendix B. Before we introduce the second lemma, let us define the following functions to simplify notations,

$$\begin{aligned} \Psi_c(\alpha, \beta, \phi) &= f(t) \\ \Psi_s(\alpha, \beta, \phi) &= \sin(\alpha + \beta \sin(\omega t + \phi)) \\ \tau &= 2\pi/\omega, \quad \alpha_{1\pm 2} = \alpha_1 \pm \alpha_2 \\ g(t) &= \Psi_c(\alpha_1, \beta_1, \phi_1) \Psi_c(\alpha_2, \beta_2, \phi_2), \end{aligned} \quad (9)$$

where $\beta_i \in (0, \pi/2]$. Then we have the following lemma,

Lemma 3. Function $g(t)$ is a periodic function with its period length $\tau_g = \tau$, except when Boolean function $\Gamma(\alpha_1, \alpha_2, \beta_1, \beta_2, \phi_1, \phi_2)$ is true where $\Gamma(\alpha_1, \alpha_2, \beta_1, \beta_2, \phi_1, \phi_2) = \Gamma_1 + \Gamma_2 + \Gamma_3$, ‘+’ is logical ‘OR’, and

$$\begin{aligned}\Gamma_1 &= (\alpha_{1+2} = k_1\pi) \cdot (\alpha_{1-2} = k_2\pi), \\ \Gamma_2 &= (\beta_1 = \beta_2) \cdot (\phi_1 = \phi_2 + (2k_1 + 1)\pi) \cdot (\alpha_{1-2} = k_2\pi), \\ \Gamma_3 &= (\beta_1 = \beta_2) \cdot (\phi_1 = \phi_2 + 2k_1\pi) \cdot (\alpha_{1+2} = k_2\pi),\end{aligned}$$

where ‘ \cdot ’ is logical ‘AND’ and $k_1, k_2 \in \mathcal{Z}$.

The proof of Lem. 3 is in Appendix C. Now we are ready to compute the period length of $D(t)$. Let τ_d be the period length of $D(t)$, we have the following theorem.

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.

Proof: We need to analyze the period length of $D(t)$ in (4). Eq. (4) has two periodic components: the first part is $c\theta_1c\theta_2$ and the second part is $c\theta_1c(\theta_2 + \theta_3)$. Denote τ_{d1} and τ_{d2} to be the period lengths of the first and the second parts, respectively. For $c\theta_1c\theta_2$, we can apply the two lemmas and obtain the following,

$$\tau_{d1} = \begin{cases} \frac{1}{2}\tau_0 & \text{If } \Gamma(\alpha_1, \alpha_2, \beta_1, \beta_2, \phi_1, \phi_2) \text{ is true,} \\ \tau_0 & \text{otherwise.} \end{cases}$$

For $c\theta_1c(\theta_2 + \theta_3)$, let us define the following variables,

$$\begin{aligned}\kappa_5 &= \beta_2c(\phi_2) + \beta_3c(\phi_3); & \kappa_6 &= \beta_2s(\phi_2) + \beta_3s(\phi_3); \\ \kappa_{56} &= \sqrt{\kappa_5^2 + \kappa_6^2}; & \phi_{\kappa_{56}} &= \arctan(\kappa_6/\kappa_5).\end{aligned}$$

Now we can again apply the two lemmas and have

$$\tau_{d2} = \begin{cases} \frac{1}{2}\tau_0 & \text{If } \Gamma(\alpha_1, \alpha_{2+3}, \beta_1, \kappa_{56}, \phi_1, \phi_{\kappa_{56}}) \text{ is true,} \\ \tau_0 & \text{otherwise.} \end{cases}$$

In steady flight, we know $\beta_3 \neq 0$ because the elbow joint does not fix at an angle. Therefore τ_d should be the least common multiple of τ_{d1} and τ_{d2} . Because $\beta_1 = \beta_2$ and $\beta_1 = \sqrt{\kappa_5^2 + \kappa_6^2}$ do not happen simultaneously, Theorem 1 is proved. \blacksquare

Remark 2. For a fixed camera w.r.t the bird, the projective matrix does not change. Therefore, $\|\mathbf{p}_1\|_2$ remains constant and $d(t)$ have the same period length as that of $D(t)$ based on (5).

Remark 3. 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 those 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 [7] 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 a person tracks a bird at a distance. Most birds have a WF significantly higher than 1 Hz. Therefore, we use a high pass filter to filter out the signal below 1 Hz to avoid noise introduced by bird gliding and 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 (6). Fig. 1(c) shows the extracted WF and the frequency distribution of the signal from video in Fig. 1(a).

VI. SPECIES PREDICTION

With the extracted WF, we can predict species of the bird. However, noise exists in the measurement. We need to know how accurate our measurement is. We first perform a variance-based error analysis before the actual species detection with trustable measurements.

Step 1: Error Bound Analysis: Due to the discreteness in frequency domain, the extracted WF has an error bound f_e . The true WF in the video is modeled as a uniform random variable in $(f_0 - f_e, f_0 + f_e)$, which is a conservative approximation.

Let the total number of frames to be N , recall that frame rate of the video is r , then the frequency interval after FFT is r/N and $f_e = \frac{r}{2N}$ is the half length of the interval. Thus, f_e can be expressed by the following equation,

$$f_e = \frac{r}{2N}. \quad (10)$$

N is rounded up to a power of 2, if it is not. Eq. (10) is quite intuitive. For a video clip with a fixed frame rate, the more frames we have, the smaller error we can get. Since the extracted WF is uniformly distributed within the error bound, the variance of the extracted WF is

$$\text{Var}(f_0) = \frac{1}{12}((f_0 + f_e) - (f_0 - f_e))^2 = \frac{1}{3}f_e^2. \quad (11)$$

For a known species s , its reference WF from the aforementioned prior knowledge has a variance of σ_s^2 . We believe that a measured WF is reliable only if its variance is less than that of the reference. Hence we establish the error bound for reliable measurements:

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 using trustable measurements. Since more frames means a smaller f_e , the least number of frames for a fixed rate video can be calculated inversely. For a 30 fps video, 100 frames approximately result in a measurement variance of 0.1 Hz, 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 $\{\mu_s, \sigma_s\}$ would be

$$L(\mu_s, \sigma_s | f_0) = \frac{1}{\sqrt{2\pi\sigma_s^2}} e^{-\frac{(f_0 - \mu_s)^2}{2\sigma_s^2}}. \quad (12)$$

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

$$L'(\mu_s, \sigma_s | f_0, f_e) = \int_{f_0 - f_e}^{f_0 + f_e} \frac{1}{2f_e} L(\mu_s, \sigma_s | f) df. \quad (13)$$

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

$$L'(\mu_s, \sigma_s | f_0, f_e) = \frac{1}{2f_e} [G(\frac{f_0 + f_e - \mu_s}{\sigma_s \sqrt{2}}) - G(\frac{f_0 - f_e - \mu_s}{\sigma_s \sqrt{2}})]. \quad (14)$$

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 or same WF distributions, and it is not desired to miss many false negative predictions for application.

VII. EXPERIMENTS

We have implemented the frequency-based salient extremity bird filtering algorithm using Matlab on a PC laptop running Microsoft Windows 7 operating system. In our experiment, WSD searching range δ is set to 5° .



Fig. 7. Samples of birds test videos.

A. Dataset

The *prior knowledge* (extended version of Table I) from [29] 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 YouTube¹ and Internet Bird Collection². All videos are recorded by moving cameras. The collected dataset contains 18 video clips of different flying birds, covering 6 species in [29]. 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 ITWDs 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 difficult set of data to use because there many overlapping WFs among species. Samples of testing videos are shown in Fig. 7. We can see that the set includes various flying poses and viewing perspectives.

video	$f_0 \pm f_e$ (Hz)	$\mu \pm \sigma$ (Hz)	video	$f_0 \pm f_e$ (Hz)	$\mu \pm \sigma$ (Hz)	video	$f_0 \pm f_e$ (Hz)	$\mu \pm \sigma$ (Hz)
1	3.5156 ± 0.0488	3.61 ± 0.207	7	5.4687 ± 0.1953	4.58 ± 0.183	13	3.3984 ± 0.0586	3.31 ± 0.149
2	3.395 ± 0.0585	3.61 ± 0.207	8	3.3203 ± 0.0488	3.05 ± 0.129	14	3.2813 ± 0.0586	3.31 ± 0.149
3	3.0469 ± 0.1172	3.01 ± 0.109	9	2.2266 ± 0.0586	2.24 ± 0.050	15	3.125 ± 0.0977	3.31 ± 0.149
4	2.8125 ± 0.1172	3.01 ± 0.109	10	2.2266 ± 0.0586	2.24 ± 0.050	16	3.0273 ± 0.0244	3.31 ± 0.149
5	2.9297 ± 0.0977	3.01 ± 0.109	11	2.1368 ± 0.0486	2.24 ± 0.050	17	3.3203 ± 0.0244	3.31 ± 0.149
6	4.6875 ± 0.0977	4.58 ± 0.183	12	2.4585 ± 0.0585	2.24 ± 0.050	18	3.125 ± 0.1953	3.31 ± 0.149

TABLE II

COMPARISON OF THE EXTRACTED WF AND ITS ERROR BOUND ($f_0 \pm f_e$) AND THE CORRESPONDING PRIOR KNOWLEDGE ($\mu \pm \sigma$) FOR ALL VIDEOS.

B. 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. Table II in next page shows a comparison between the extracted WFs and the corresponding ground truth from [29]. It is clear that the extract WFs are very close to the ground truth, and the difference between them does not exceed 0.3 Hz for all test videos, except for video 7. The results show that the system is capable of extracting WFs from different camera perspectives. It shows that WF is a stable signature for the species recognition.

¹<http://www.youtube.com/>

²<http://ibc.lynxeds.com/>

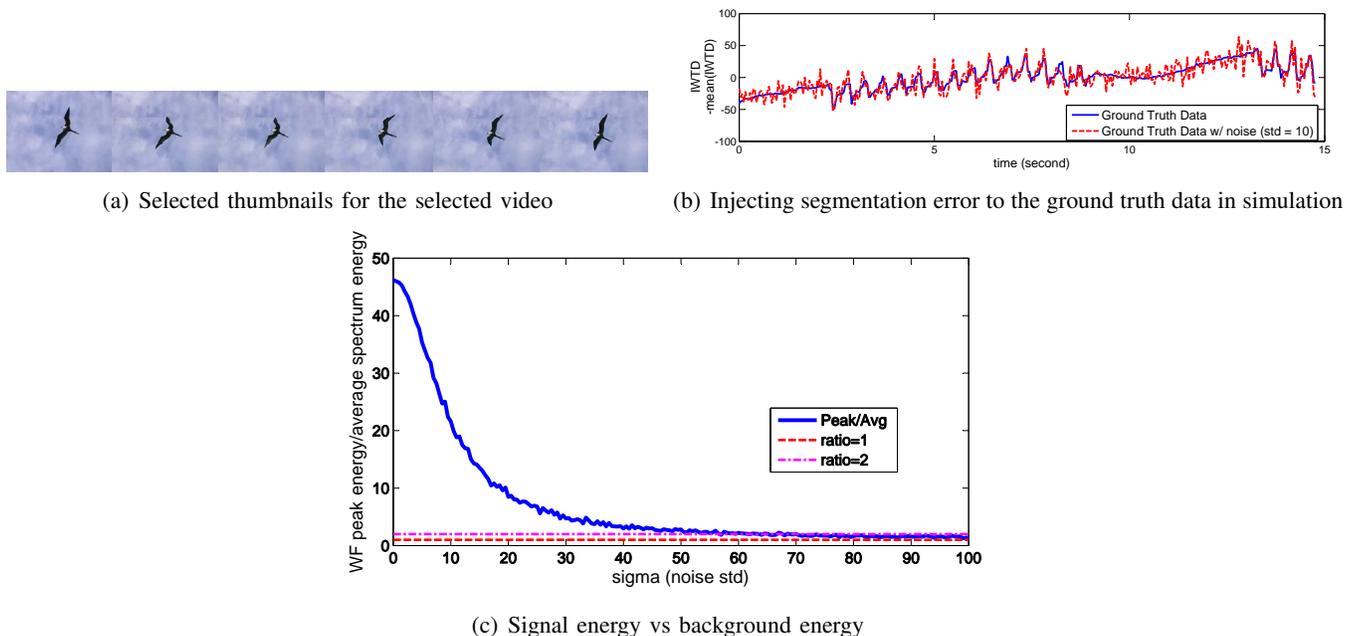


Fig. 8. Simulation results on testing the robustness of frequency analysis with respect to segmentation errors. (a) A sample wingbeat period in the video (b) Blue solid curve: the ground true $d(t)$. Red dotted curve: $d(t)$ after adding Gaussian noise with zero mean and a standard deviation of 10 pixels. (c) 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.

C. Robustness to segmentation error

Since our method relies on the extraction of pixel distance, the temporal feature is inevitably affected by the foreground segmentation error at the bird wing tip. This error happens when image resolution is low or motion blur appears. The error influences the accuracy of pixel distance $d(t)$. We use simulation to evaluate on how segmentation errors affect WF results. Consider the segmentation error at a wing tip to follow a zero-mean Gaussian distribution, the Euclidean distance between wing tips follows Gaussian distribution as well. The simulation is conducted on a real signal from test video 11 (Fig. 8(a)), where we manually annotated the wing tip positions in every image in the video. A sequence of $d(t)$ is therefore calculated upon the annotation and treated as a low noise ground truth signal as illustrated as the blue solid curve in Fig. 8(b). 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 of $d(t)$ is 108.82. Different levels of Gaussian noise is added to the signal. The red dotted curve in Fig. 8(b) show the simulated signal when the standard deviation is 10. We gradually increase the noise standard deviation and measure the ratio between the WF peak energy and the average spectrum energy (Fig. 8(c)). It is shown that with noise standard derivation varies from 0 pixel to 100 pixels, the WF energy is still larger then average spectrum energy. While in our experiment 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.

D. 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 Table III. It is clear that our algorithm significantly out perform the random guess. At a list length of 3, it reaches a hit rate of 0.6111 while that of a random guess can only reach 0.0938. Mean Reciprocal Rank (MRR) of the 18 testing videos is 0.4401. The rank of the correct species in the output list for each test video is showed in the RoCS row of Table IV. Though

List length	1	3	5	10
Hit rate	0.1667	0.6111	0.7222	1

TABLE III
HIT RATE COMPARISON AT DIFFERENT LIST LENGTHS.

the extracted WF is close to ground truth, the rank of correct species may not be very high, such as in video 4 and 12, because some species have close ground truth WFs and are not easy to be distinguished using WFs. This is not a significant issue in practice because other information such as habitat location can be used to distinguish birds in the ranked list.

E. 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 Table IV. The loss rate varies from 0 to 30%. Even for the video with most data lost (video 18), the correct species still can be ranked in the top three positions.

VIII. CONCLUSION AND FUTURE WORK

We developed a bird species filtering method that takes videos from cameras with unknown parameters as input and outputs likelihood of candidate species. The method can extract the time series of salient extremities, which is inter-wing tip distance (IWTP), from the videos without assuming knowledge on camera motion and perspective changes. We also derived the probability that the salient extremity can be recognized in the 2D image frame for an arbitrary relative perspective between the camera and the bird. With the exception of ignorable degenerated cases, we also proved that the periodicity of the wingbeat in the image is the same as that in the 3D space regardless camera parameters. This allowed us to apply Fast Fourier Transformation to the observed IWTD series to obtain wingbeat frequency. We also proposed a species prediction metric using likelihood ratios. We have implemented the algorithm and tested it in experiments using 18 video clips against 32 candidate bird species. Experimental results validated our design and analysis.

In the future, we will develop recognition methods using other features such as flying speed, color and shape in combination with frequency signatures to achieve more precise prediction. Note that the method also has the potential to be applied to other animals with frequency characteristics.

ACKNOWLEDGMENTS

Thanks for K. Goldberg, R. Volz and J. Yi for insightful discussions. Thanks for C. Kim, Y. Lu, H. Li, and H. Ge 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, August 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, "Estimating 3d trajectories of periodic motions from stationary monocular views," in *ECCV*, Marseille, France, October 2008.
- [6] —, "View-invariant analysis of periodic motion," in *IEEE/RSJ Intl. Conf. Intelligent Robots and Systems*, St. Louis, Missouri, USA, October 2009.
- [7] —, "3d reconstruction of periodic motion from a single view," *IJCV*, vol. 90, no. 1, pp. 28–44, October 2010.

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

TABLE IV
RoCS FOR TESTING VIDEOS WITH DIFFERENT MLR.

- [8] D. H. A. Elgamma and L. Davis, "Non-parametric model for background subtraction," in *ECCV*, 2000.
- [9] Y. Ran, I. Weiss, Q. Zheng, and L. Davis, "Pedestrian detection via periodic motion analysis," *IJCV*, vol. 71, no. 2, pp. 143–160, 2007.
- [10] C. Cohen, L. Conway, and D. Koditschek, "Dynamical system representation, generation and recognition of basic oscillatory motion gestures," in *Proc. Intl. Conf. Automatic Face and Gesture Recognition*, 1996.
- [11] S. Seitz and C. Dyer, "View-invariant analysis of cyclic motion," *IJCV*, vol. 25, no. 3, pp. 231–251, 1997.
- [12] P. Tsai, M. Shah, K. Keiter, and T. Kasparis, "Cyclic motion detection for motion based recognition," *Pattern Recognition*, vol. 27, no. 12, pp. 1591–1603, 1994.
- [13] I. Laptev, S. Belongie, P. Pérez, and J. Wills, "Periodic motion detection and segmentation via approximate sequence alignment," in *ICCV*, October 2005.
- [14] 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.
- [15] 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.
- [16] A. Bobick and J. Davis, "The recognition of human movement using temporal templates," *TPAMI*, vol. 23, no. 3, pp. 257 – 267, 2001.
- [17] R. Polana and R. Nelson, "Detection and recognition of periodic, nonrigid motion," *IJCV*, vol. 23, no. 3, pp. 261–282, 1997.
- [18] R. Cutler and L. Davis, "Robust real-time periodic motion detection, analysis, and applications," *TPAMI*, vol. 22, no. 8, pp. 781 – 796, 2000.
- [19] L. Gammaitoni, E. Menichella-saetta, S. Santucci, and F. Marchesoni, "Extraction of periodic signals from a noise background," vol. 142, no. 2,3, Dec. 1989.
- [20] T. Yoshizawa, S. Hirobayashi, and T. Misawa, "Noise reduction for periodic signals using high-resolution frequency analysis," 2011.
- [21] 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.
- [22] 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.
- [23] 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.
- [24] 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 *Special Track on Physically Grounded AI(PGAI), the 24th AAAI Conference on Artificial Intelligence (AAAI-10)*, Atlanta, Georgia, USA, July 2010.
- [25] —, "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.
- [26] S. Blackman, "Multiple hypothesis tracking for multiple target tracking," vol. 19, no. 1, Jan. 2004.
- [27] A. Jerri, "The shannon sampling theorem - its various extensions and applications: A tutorial review," *The Proceedings of IEEE*, vol. 65, no. 11, 1977.
- [28] C. Pennycuik, "Wingbeat frequency of birds in steady cruising flight: New data and improved predictions," *The Journal of Experimental Biology*, vol. 199, pp. 1613–1618, 1996.
- [29] —, "Predicting wingbeat frequency and wavelength of birds," *The Journal of Experimental Biology*, vol. 150, pp. 171–185, 1990.
- [30] C. Liu, "Beyond pixels: Exploring new representations and applications for motion analysis," Ph.D. dissertation, Massachusetts Institute of Technology, May 2009.
- [31] T. Brox, A. Bruhn, N. Papenbergh, and J. Weickert, "High accuracy optical flow estimation based on a theory for warping," in *ECCV*, 2004.
- [32] A. Bruhn, J. Weickert, and C. Schnörr, "Lucas/kanade meets horn/schunck: combining local and global optical flow methods," *IJCV*, vol. 61, no. 3, p. 211231, 2005.
- [33] P. J. Rousseeuw and A. M. Leroy, *Robust Regression and Outlier Detection*. New York: Wiley, 1987.
- [34] P. Filzmoser, R. G. Garrett, and C. Reimann, "Multivariate outlier detection in exploration geochemistry," *Computer and Geosciences*, vol. 31, pp. 579–587, 2005.
- [35] T. F. Chan and L. A. Vese, "Active contours without edges," *TIP*, vol. 10, no. 2, pp. 266–277, 2001.
- [36] S. Lankton, "Sparse field methods - technical report," Georgia Institute of Technology, Tech. Rep., April 2009.

- [37] J. Dunn and J. Alderfer, *Field Guide to the Birds of Eastern North America*. National Geographic, Washington D. C., 2008.
- [38] T. Liu, K. Kuykendoll, R. Rhew, and S. Jones, "Avian wing geometry and kinematics," *AIAA Journal*, vol. 44, no. 5, May 2006.
- [39] J. Craig, *Introduction to Robotics Mechanics and Control (Third Edition)*. Pearson Education, Upper Saddle River, New Jersey, 2005.

APPENDIX A
PROOF OF LEM. 1

Proof: When the wingspan reaches its maximum in steady flight, the wing spreading direction (WSD) is

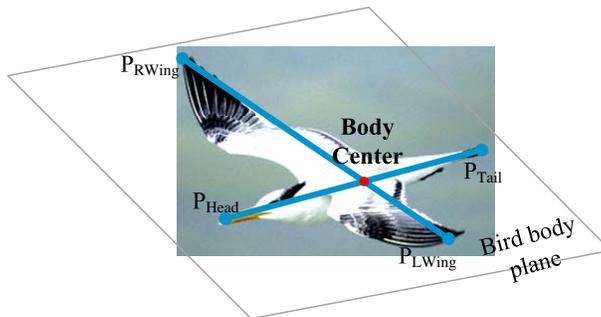


Fig. 9. Bird body plane and wing and body stick model.

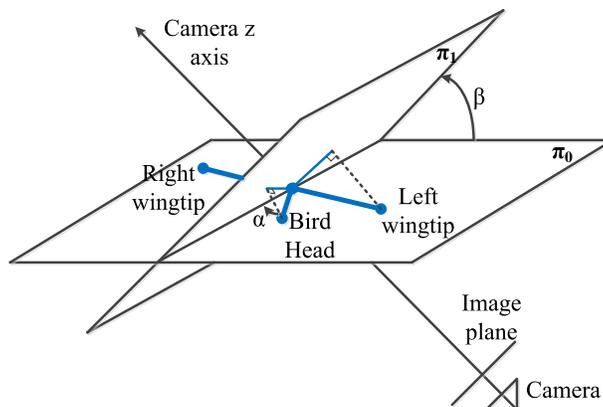


Fig. 10. Analysis on the angle of camera optical axis.

perpendicular to the bird body axis. Model the bird skeleton by a cross (Fig. 9) with two orthogonal bars. The two bars determine a bird body plane. Recall that the perspective camera follows a pinhole camera model and the bird is in far field of the camera view. The relationship between a 3D point $P = [X, Y, Z]^T$ and its 2D projection $p = [x, y]^T$ in the image follows $x = fX/Z$ and $y = fY/Z$ where f is the camera focal length. Notations in figures are defined as follows:

- $P_{BC} = [X_{BC}, Y_{BC}, Z_{BC}]^T$ is the 3D coordinate of bird body center.
- P_{LWing} and P_{RWing} are bird left and right wing tips in 3D, respectively.
- P_{Head} and P_{Tail} are bird head and tail end points in 3D, respectively.
- α is the angle of bird flying trajectory w.r.t. the image plane (see Fig. 10).
- θ is the angle of bird body center projection ray w.r.t. the camera optical axis (z axis).
- L_W is the length of IWTD in 3D, while L_B is the length of bird body in 3D.
- l_W is the length of IWTD on image, while l_B is the length of bird body on image.

Since the bird is in steady flight, its body plane is horizontal. The camera has a tilting angle β w.r.t. the horizontal plane. We first analyze how the camera's tilting angle affects the probability of successful recognition of salient extremities. Consider $\theta = 0$ (the analysis is similar when $\theta \neq 0$). Plane π_0 represents the bird body plane in Fig. 10,

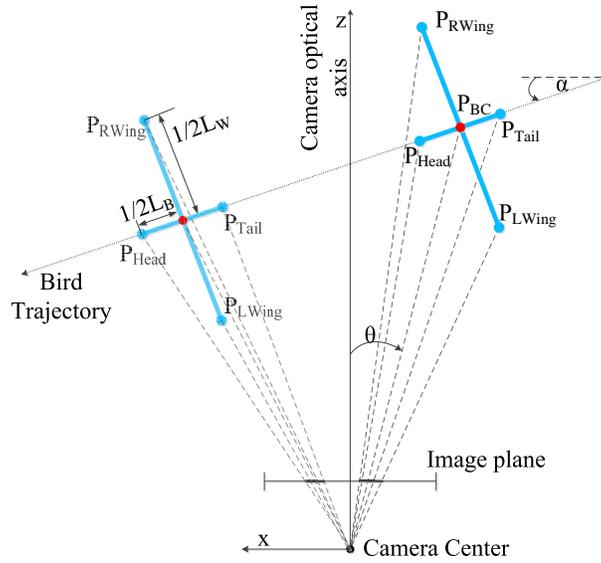


Fig. 11. Illustration of projecting bird onto image plane.

while π_1 is parallel to the image plane and intersects π_0 at the bird body center. From Fig. 10, then the projection of bird body length L'_B and IWTD L'_W on π_1 are:

$$L'_B = L_B \sqrt{\cos^2 \alpha + \sin^2 \alpha \cos^2 \beta} \quad (15)$$

$$L'_W = L_W \sqrt{\sin^2 \alpha + \cos^2 \alpha \cos^2 \beta} \quad (16)$$

By geometry similarity, the ratio between l_w and l_B can be approximated by L'_W/L'_B (since the bird is in far view). To ensure $l_w/l_B > 1$, it must satisfy the following:

$$\tan \alpha > \sqrt{\frac{1 - (L_W/L_B)^2 \cos^2 \beta}{(L_W/L_B)^2 - \cos^2 \beta}}. \quad (17)$$

As $|\beta|$ grows larger from 0 to 90°, the threshold becomes larger and the region of bird trajectory orientation α for successful recognition becomes smaller. When $\beta = 90^\circ$ (that is the bird plane is perpendicular to the image plane), the probability of success reaches the minimum.

In the following proof, we analyze this worst scenario only, because it gives a lower bound of the probability of successful recognition of salient extremities. Fig. 11 shows the top view of the setting in our analysis, where the image plane is perpendicular to the paper plane, and bird body plane is parallel to the paper plane. The bird trajectory is assumed to be a straight line in a short time period.

By the projection relationship between 3D and 2D points, we have

$$l_B = \left| \frac{f(X_{BC} + 1/2L_B \cos \alpha)}{Z_{BC} - 1/2L_B \sin \alpha} - \frac{f(X_{BC} - 1/2L_B \cos \alpha)}{Z_{BC} + 1/2L_B \sin \alpha} \right| \quad (18)$$

$$l_W = \left| \frac{f(X_{BC} + 1/2L_W \sin \alpha)}{Z_{BC} + 1/2L_B \cos \alpha} - \frac{f(X_{BC} - 1/2L_B \sin \alpha)}{Z_{BC} - 1/2L_B \cos \alpha} \right| \quad (19)$$

That is

$$l_B = \left| \frac{fL_B(X_{BC} \sin \alpha + Z_{BC} \cos \alpha)}{Z_{BC}^2 - (1/2L_B)^2 \sin^2 \alpha} \right| \quad (20)$$

$$l_W = \left| \frac{fL_W(X_{BC} \cos \alpha - Z_{BC} \sin \alpha)}{Z_{BC}^2 - (1/2L_W)^2 \cos^2 \alpha} \right| \quad (21)$$

Since the bird is in far field of the camera view, $L_W, L_B \ll Z_{BC}$, we ignore the second term in the denominator. Therefore, the ratio between l_W and l_B is

$$\begin{aligned} \frac{l_W}{l_B} &\approx \frac{L_W}{L_B} \left| \frac{X_{BC} \cos \alpha - Z_{BC} \sin \alpha}{X_{BC} \sin \alpha + Z_{BC} \cos \alpha} \right| \\ &= \frac{L_W}{L_B} \left| \frac{\tan \theta - \tan \alpha}{\tan \theta \tan \alpha + 1} \right| \\ &= \frac{L_W}{L_B} |\tan(\alpha - \theta)| \end{aligned} \quad (22)$$

To successfully recognize the salient extremity as the IWTD, the ratio l_W/l_B should be greater than 1. Thus, we have

$$|\tan(\alpha - \theta)| \geq \frac{L_B}{L_W} \quad (23)$$

Consider α is uniformly distributed on $(-\pi/2, \pi/2)$, and θ is uniformly distributed on $(-\Theta_h, \Theta_h)$ where $0 < 2\Theta_h < \pi$ is the horizontal field of view of the camera. Let $\beta = (\alpha - \theta)$, then β follows the triangle distribution, with the probability density function

$$f_\beta = \begin{cases} \frac{\beta + \Theta_h}{2\Theta_h\pi} + \frac{1}{4\Theta_h} & \text{if } -\frac{\pi}{2} - \Theta_h \leq \beta < -\frac{\pi}{2} + \Theta_h \\ \frac{1}{\pi} & \text{if } -\frac{\pi}{2} + \Theta_h \leq \beta < \frac{\pi}{2} - \Theta_h \\ \frac{\frac{\pi}{2} - \beta + \Theta_h}{2\Theta_h\pi} + \frac{1}{4\Theta_h} & \text{if } \frac{\pi}{2} - \Theta_h \leq \beta \leq \frac{\pi}{2} + \Theta_h \\ 0 & \text{otherwise} \end{cases} \quad (24)$$

Define the indicator variable \mathbf{I}_β as

$$\mathbf{I}_\beta = \begin{cases} 1 & \text{if } l_W \geq l_B \\ 0 & \text{otherwise} \end{cases} \quad (25)$$

Then, given a ratio L_B/L_W , the probability of successful recognition of salient extremities in a wingbeat period is the integral

$$Pr\{\mathbf{I}_\beta = 1 | \frac{L_B}{L_W}\} = \int_{-\infty}^{+\infty} \mathbf{I}_\beta f_\beta d\beta \quad (26)$$

$$= \int_{|\tan \beta| \geq \frac{L_B}{L_W}} f_\beta d\beta \quad (27)$$

$$= \int_{|\tan \beta| \geq \frac{L_B}{L_W}, \beta \geq 0} f_\beta d\beta + \int_{|\tan \beta| \geq \frac{L_B}{L_W}, \beta < 0} f_\beta d\beta \quad (28)$$

Since the absolute tangent function is symmetric, the two parts in (28) are equal. Therefore, we only consider the integral on $\beta \geq 0$ as follows

$$Pr\{\mathbf{I}_\beta = 1 | \frac{L_B}{L_W}\} = 2 \int_{|\tan \beta| \geq \frac{L_B}{L_W}, \beta \geq 0} f_\beta d\beta \quad (29)$$

$$= 2 \int_{\arctan(\frac{L_B}{L_W})}^{\pi - \arctan(\frac{L_B}{L_W})} f_\beta d\beta \quad (30)$$

On the positive axis of β , the triangle distribution density changes at the point $\beta = \frac{\pi}{2} - \Theta_h$. Therefore, it is necessary to compare $\arctan(\frac{L_B}{L_W})$ with $\frac{\pi}{2} - \Theta_h$, in order to calculate the probability. We have two cases here:

Case 1: if $\arctan(\frac{L_B}{L_W}) \leq \frac{\pi}{2} - \Theta_h$, then (30) can be unfolded to

$$Pr\{\mathbf{I}_\beta = 1 | \frac{L_B}{L_W}\} \quad (31)$$

$$= 2 \int_{\arctan(\frac{L_B}{L_W})}^{\frac{\pi}{2} - \Theta_h} \frac{1}{\pi} d\beta + 2 \int_{\frac{\pi}{2} - \Theta_h}^{\frac{\pi}{2} + \Theta_h} \frac{-\beta + \Theta_h}{2\Theta_h\pi} + \frac{1}{4\Theta_h} d\beta \quad (32)$$

$$= 1 - \frac{2}{\pi} \arctan\left(\frac{L_B}{L_W}\right) \quad (33)$$

Case 2: if $\arctan(\frac{L_B}{L_W}) > \frac{\pi}{2} - \Theta_h$, then (30) is

$$Pr\{\mathbf{I}_\beta = 1 | \frac{L_B}{L_W}\} \quad (34)$$

$$= 2 \int_{\arctan(\frac{L_B}{L_W})}^{\pi - \arctan(\frac{L_B}{L_W})} \frac{-\beta + \Theta_h}{2\Theta_h\pi} + \frac{1}{4\Theta_h} d\beta \quad (35)$$

$$= 1 - \frac{2}{\pi} \arctan(\frac{L_B}{L_W}) \quad (36)$$

The two cases result in the same probability equation that is independent of Θ_h . The larger the ratio L_W/L_B is, the higher the successful probability can reach. Therefore, Lem. 1 is true. ■

APPENDIX B PROOF OF LEM. 2

Proof: The function $f(t)$ repeats at least every τ time because

$$f(t + \tau) = \cos(\alpha + \beta \sin(\omega t + \omega\tau + \phi)) \quad (37)$$

$$= \cos(\alpha + \beta \sin(\omega t + 2\pi + \phi)) \quad (38)$$

$$= \cos(\alpha + \beta \sin(\omega t + \phi)) = f(t). \quad (39)$$

Suppose the period length of $f(t)$ is τ_f , it is trivial that $0 < \tau_f \leq \tau$. We also have the following equation for all t .

$$\cos(\alpha + \beta \sin(\omega t + \phi)) = \cos(\alpha + \beta \sin(\omega(t + \tau_f) + \phi)) \quad (40)$$

Considering $\beta \in (0, \pi/2]$, if (40) is true, then either of the following cases must be true:

$$\begin{aligned} \alpha + \beta \sin(\omega t + \phi) = \\ \begin{cases} \text{case 1: } \alpha + \beta \sin(\omega t + \omega\tau_f + \phi), \\ \text{case 2: } -\alpha - \beta \sin(\omega t + \omega\tau_f + \phi) + 2k\pi, \end{cases} \end{aligned} \quad (41)$$

where $k \in \mathcal{Z}$. When the first case in (41) happens, we have

$$\sin(\omega t + \phi) = \sin(\omega t + \omega\tau_f + \phi) \quad (42)$$

Then there are two solutions for (42):

$$\omega t + \phi = \omega t + \omega\tau_f + \phi + 2k'\pi, \quad (43)$$

and

$$\omega t + \phi = \pi - \omega t - \omega\tau_f - \phi + 2k'\pi, \quad (44)$$

where $k' \in \mathcal{Z}$. However, Eq. (44) cannot be true for all t because all parameters except t are constants. Therefore, only (43) is true and it becomes,

$$\tau_f = k'\tau. \quad (45)$$

Since $0 < \tau_f \leq \tau$, k' can only have the value of 1. Therefore, $\tau_f = \tau$, for all α in this case.

Similarly, if the second case in (41) happens, we can prove

$$\tau_f = \frac{(2k' - 1)}{2}\tau = \frac{1}{2}\tau, \text{ and } \alpha = k\pi \quad (46)$$

Combining (45) and (46), Lem. 2 is proved. ■

APPENDIX C
PROOF OF LEM. 3

Proof: Let us decompose $g(t)$,

$$\begin{aligned}
g(t) &= c\alpha_1c\alpha_2\Psi_c(0, \beta_1, \phi_1)\Psi_c(0, \beta_2, \phi_2) \\
&\quad -s\alpha_1c\alpha_2\Psi_s(0, \beta_1, \phi_1)\Psi_c(0, \beta_2, \phi_2) \\
&\quad -c\alpha_1s\alpha_2\Psi_c(0, \beta_1, \phi_1)\Psi_s(0, \beta_2, \phi_2) \\
&\quad +s\alpha_1s\alpha_2\Psi_s(0, \beta_1, \phi_1)\Psi_s(0, \beta_2, \phi_2)
\end{aligned} \tag{47}$$

and define the following intermediate variables,

$$\begin{aligned}
\kappa_1 &= \beta_1c(\phi_1) + \beta_2c(\phi_2); & \kappa_2 &= \beta_1s(\phi_1) + \beta_2s(\phi_2); \\
\kappa_3 &= \beta_1c(\phi_1) - \beta_2c(\phi_2); & \kappa_4 &= \beta_1s(\phi_1) - \beta_2s(\phi_2); \\
\kappa_{12} &= \sqrt{\kappa_1^2 + \kappa_2^2}; & \kappa_{34} &= \sqrt{\kappa_3^2 + \kappa_4^2}; \\
\phi_{\kappa_{12}} &= \arctan(\kappa_2/\kappa_1); & \phi_{\kappa_{34}} &= \arctan(\kappa_4/\kappa_3);
\end{aligned}$$

Then, (47) can be transformed to

$$\frac{1}{4}(\Psi_c(\alpha_{1+2}, \kappa_{12}, \phi_{\kappa_{12}}) + \Psi_c(\alpha_{1-2}, \kappa_{34}, \phi_{\kappa_{34}})) \tag{48}$$

We have the following cases:

- $\kappa_{12} = 0$: This happens if and only if $\beta_1 = \beta_2$ and $\phi_1 = \phi_2 + (2k_1 + 1)\pi$. Then, from Lem. 2, $\tau_g = \tau$ unless $\alpha_{1-2} = k_2\pi$.
- $\kappa_{34} = 0$: This happens if and only if $\beta_1 = \beta_2$ and $\phi_1 = \phi_2 + 2k_1\pi$. Then $\tau_g = \tau$ unless $\alpha_{1+2} = k_2\pi$.
- Otherwise, $\kappa_{12} \neq 0$ and $\kappa_{34} \neq 0$ Then, the first component of (48) has period length of $\tau/2$ only when $\alpha_{1+2} = k_1\pi$. The second component has period length $\tau/2$ only when $\alpha_{1-2} = k_2\pi$. Therefore, $g(t)$ has period length $\tau_g = \tau$ unless $\alpha_{1+2} = k_1\pi$ and $\alpha_{1-2} = k_2\pi$.

Therefore, Lem. 3 is proved. ■