A Low False Negative Filter for Detecting Rare Bird Species from Short Video Segments using a Probable Observation Data Set-based EKF Method [*]

Dezhen Song and **Yiliang Xu**
Department of Computer Science and Engineering,
Texas A&M University, College Station, TX 77843
{dzsong, ylxu}@cse.tamu.edu

Abstract

We report a new filter for assisting the search for rare bird species. Since a rare bird only appears in front of the camera with very low occurrence (e.g. less than ten times per year) for very short duration (e.g. less than a fraction of a second), our algorithm must have very low false negative rate. We verify the bird body axis information with the known bird flying dynamics from the short video segment. Since a regular extended Kalman filter (EKF) cannot converge due to high measurement error and limited data, we develop a novel Probable Observation Data Set (PODS)-based EKF method. The new PODS-EKF searches the measurement error range for all probable observation data that ensures the convergence of the corresponding EKF in short time frame. The algorithm has been extensively tested in experiments. The results show that the algorithm achieves 95.0% area under ROC curve in physical experiment with close to zero false negative rate.

Introduction

Our group focuses on developing autonomous observatories to assist nature scientists to search rare birds. In our recent project, a camera was installed in the middle of a forest in Bayou DeView in eastern Arkansas, running 24 hours a day, to assist the ornithologists to search for the thought-to-be-extinct ivory-billed woodpecker (IBWO) (see Fig. 1).

Three critical conditions must be met for the searching task. First, a rare bird only appears in front of the camera with very low occurrence (e.g. less than ten times per year) for very short duration (e.g. less than a fraction of a second), our algorithm must have very low false negative rate. Second, since the final verification has to be performed by human experts, it is necessary to reduce the huge data volume to a manageable size, which also means that the filter can tolerate a less ideal false positive rate. Third, the system must be easy to setup in the forest. Due to power and communication constraints, a single camera is preferred because it does not require the precise calibration and synchronization as dislocated stereo rigs would for distant birds.

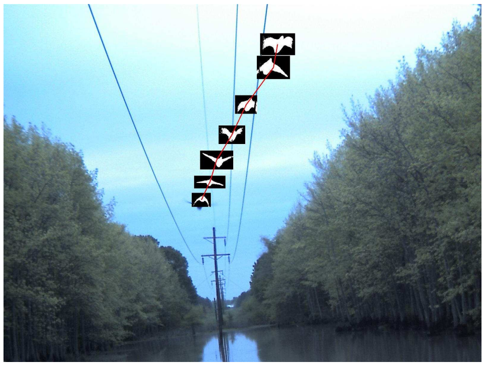


Figure 1: An example of a short video sequence of a flying bird that is captured in Bayou DeView in eastern Arkansas. It superimposes the segmented bird images from consecutive video frames on the top of the background frame.

Fig. 1 shows the input of the problem is a short segmented motion sequence of an object. The output of the problem is to determine whether the motion sequence is caused by a targeted bird species. We verify the bird body axis information with the known bird flying dynamics. Since a regular extended Kalman filter (EKF) cannot converge due to the high measurement error and the limited observation data due to the high flying speed of the bird (e.g. the sample bird sequence in Fig. 1 only contains seven data points), we develop a probable observation data set (PODS)-based EKF and an approximate computation scheme. The new PODS-EKF searches the measurement error range for all probable observation data that ensure the convergence of the corresponding EKF. The filtering is based on whether the PODS is non-empty and the corresponding velocity is within the known bird flying velocity profile. We show that the PODS-EKF theoretically ensures a zero false negative rate. We have evaluated the filtering algorithm using both the simulated data and field test data.

Related Work

Animal detection and recognition using video images has been a active research direction. Most of the existing approaches build appearance models of animals by silhou-

[*] This work is supported in part by NSF CAREER program under IIS-0643298 and MRI program under CNS-0923203, in part by Arkansas Electric Cooperatives Corp., and in part by U.S. Fish and Wildlife Service.

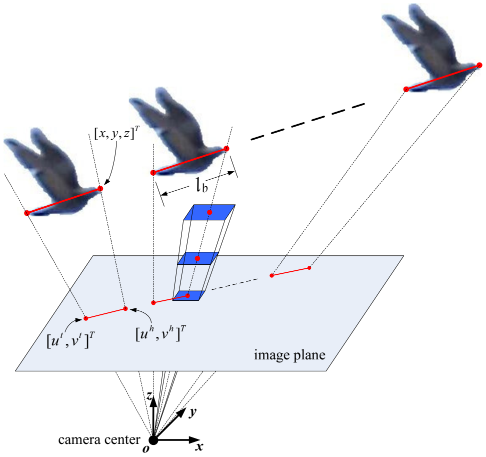


Figure 2: An illustration of bird filtering. The motion sequence of the bird can be used to extract a set of moving line segments that correspond to the body axis of the bird. The segmentation error of the end of body axis is uniformed distributed in the u-v image plane and can be represented as an inverse pyramid when the error range is back-projected from the camera center to the 3D space.

ette/contour (Sebastian, Klein, and Kimia 2004), 2D kinematic chains of rectangular segments (Ramanan, Forsyth, and Barnard 2006) etc. A known set of animal images are used to train and test the model using techniques such as clustering (Ramanan, Forsyth, and Barnard 2006), template matching (Sebastian, Klein, and Kimia 2004) etc. Different from this class of techniques, a large learning set is unavailable for our applications such as detecting rare birds.

Periodic motion detection (Ran et al. 2007; Briassouli and Ahuja 2007) assumes objects, such as animals, with periodic motion pattern and applies time-frequency analysis (Briassouli and Ahuja 2007) or image sequence alignment (Laptev et al. 2005) to capture the periodicity. Applications of periodic motion detection have been found to vehicles, humans, and even canines. However, recognizing birds is different because a bird flying pattern combines both gliding and wing-flapping and the periodic motion assumption does not hold.

Recently, the 3D structure inference using monocular vision has drawn increasing research attention. Ribnick et al. (Ribnick, Atev, and Papanikolopoulos 2009) propose an algorithm for estimating the 3D parabolic trajectories of projectiles in monocular views. Saxena et al. (Saxena, Sun, and Ng 2008) propose a learning algorithm that estimates 3D structures of a static scene based on a single still image. The work models the scene with sets of planes using Markov Random Field (MRF) and trains the model based on depth cues such as texture variations & gradients, color, haze, and defocus etc. Different from these works, our approach deals with a highly dynamic object (i.e., the bird) and its trajectory is not necessarily parabolic.

Visual tracking has been a active research area. Various techniques have been proposed and a recent survey can be found in (Yilmaz, Javed, and Shah 2006). The fundamental technique we used is the extended Kalman filer. (Extended) Kalman filter (Spinello, Triebel, and Siegwart 2008) and their variations can be viewed as model-based filtering methods that verify the prior known dynamic models of the objects. It has seen a wide range of applications in object tracking and recognition such as vehicles, pedestrians etc. However, there is no existing work regarding how to detect a flying bird. Most existing works assume rigid objects and do not worry about the convergence of Kalman filter because an ample amount of observation data is available. Unfortunately, those conditions do not hold for the filtering of a highly-dynamic and high-speed flying bird.

Our group has developed systems and algorithms (Song 2009; Song, van der Stappen, and Goldberg 2006; Song and Goldberg 2007) for networked robotic cameras for a variety of applications such as construction monitoring, distance learning, panorama construction, and nature observation. Our previous work (Song et al. 2008) details how to build an autonomous nature observation system using motion detection. We learn that mere motion detection cannot save the biologists from the herculean task of image sorting, which inspires this work.

Problem Description

Our system is a monocular vision system with a narrow field of view (FOV). The position of objects with respect to the camera is based on a 3D Cartesian camera coordinate system (CCS) with its origin at the camera center as shown in Fig. 2. The x-axis and y-axis of the CCS are parallel to the u-axis and the v-axis of the image coordinate system, respectively.

From the knowledge provided by ornithologists, we know that a flying bird is usually an adult bird. A bird does not change its size once reaching its adulthood. Birds of the same species share a similar size and flying speed range. This biological information allows us to distinguish the targeted species from other moving objects.

Assumptions

To establish the bird detection problem, we also have the following assumptions,

- A fixed and pre-calibrated camera is used. With a calibrated camera and without loss of generality, we can always transform camera intrinsic parameter matrix K_c to diag$(f, f, 1)$, where f is the focal length of the camera in units of pixel side length. ICS must have its origin located on the principal axis (z axis) of CCS. Hence we have perspective project matrix $P_c = [K_c|\mathbf{0}_{3\times 1}]$.
- There is only one bird in the image sequence. If there are multiple birds in the scene, we assume each individual bird sequence has been isolated out using multiple object tracking techniques (Yilmaz, Javed, and Shah 2006).
- The bird is flying along a straight line with a constant speed when captured by the camera. This assumption usually holds considering a fast flying bird can only stay in the FOV for less than a second.

Inputs and Output

The input of the problem is a sequence of n images which contain a moving object of any type. Each frame is timestamped. Based on the information from ornithologists, we know the body length l_b and the flying speed range $\mathcal{V} = [v_{min}, v_{max}]$ of the targeted bird species. The output is to determine if the motion sequence is caused by the targeted bird species or not.

Modeling a Flying Bird

To develop a bird filter, the key is to extract the bird flying dynamic information from the segmented bird motion sequence and associate the information with the known flying models and the prior information regarding the targeted species. As detailed in (Song et al. 2008), we segment the moving object from its background and obtain a motion sequence. The most important bird flying information is the body axis. Based on the information provided by ornithologists and our data, we know that a flying bird maintains a fixed body length during flight and birds from the same species share a similar size. Let us define the bird body line segment in the image frame as

$$\mathbf{z} = [u^h, v^h, u^t, v^t]^T, \tag{1}$$

where (u^h, v^h) is the head position and (u^t, v^t) is the tail position. From $\mathbf{z}$, we can compute the body axis orientation $\theta = \text{atan2}(u^h - u^t, v^h - v^t)$, and the body axis length $l = \sqrt{(u^h - u^t)^2 + (v^h - v^t)^2}$. Note that l is different from l_b. l is the projection of l_b on the image plane and is in unit of pixel. Since the bird body axis is almost parallel to the bird flying trajectory as in Fig. 2, we can extract the bird body axis using approaches such as Hough transform, detailed in (Song et al. 2008).

To determine whether the motion information is caused by the targeted species, we need to establish a bird flying model in the image frame. Let $\mathbf{p} = [x, y, z]^T$ denote the head position of the bird body axis and $\mathbf{v} = [\dot{x}, \dot{y}, \dot{z}]^T$ denote its velocity in the CCS. Since the bird flies along a straight line with a constant velocity, we have

$$\dot{\mathbf{x}} = [\dot{\mathbf{p}}, \dot{\mathbf{v}}]^T = [\dot{x}, \dot{y}, \dot{z}, 0, 0, 0]^T = [\mathbf{v}, \mathbf{0}]^T, \tag{2}$$

where the state variable $\mathbf{x} = [\mathbf{p}, \mathbf{v}]^T \in \mathbb{R}^6$ describes the position and velocity of the bird head. Define $\mathbf{p}_{tail} = [x^t, y^t, z^t]^T$ as the position of the bird tail and we have $\mathbf{p}_{tail} = [x - \dot{x}l_b/\|\mathbf{v}\|, y - \dot{y}l_b/\|\mathbf{v}\|, z - \dot{z}l_b/\|\mathbf{v}\|]^T$.

As illustrated in Fig. 2, the relationship between the measurement data $\mathbf{z}$ defined in (1) and the corresponding state $\mathbf{x}$ can be described using the pin-hole camera model,

$$\mathbf{z} = \begin{bmatrix} fx/z \\ fy/z \\ fx^t/z^t \\ fy^t/z^t \end{bmatrix} = \begin{bmatrix} fx/z \\ fy/z \\ f\frac{x\|\mathbf{v}\| - l_b\dot{x}}{z\|\mathbf{v}\| - l_b\dot{z}} \\ f\frac{y\|\mathbf{v}\| - l_b\dot{y}}{z\|\mathbf{v}\| - l_b\dot{z}} \end{bmatrix} + \mathbf{w} := h(\mathbf{x}) + \mathbf{w}, \tag{3}$$

where $h(\cdot)$ is usually called measurement function and $\mathbf{w}$ represents the measurement noise.

Probable Observation Data Set-based EKF

Extended Kalman Filter

The nonlinear dynamic system described by (3) captures the prior known information regarding the targeted species. If the motion is caused by the targeted species, then the bird body axis information should follow the nonlinear dynamic system described by (3), which can be validated using an EKF to track the states of the moving object. Eqs. (2) and (3) can be re-written in a discrete-time form,

$$\mathbf{x}(k+1) = A(k+1)\mathbf{x}(k) + \mathbf{q}(k), \tag{4a}$$

$$\mathbf{z}(k) = h(\mathbf{x}(k)) + \mathbf{w}(k), \tag{4b}$$

where $\mathbf{q}(k) \in \mathbb{R}^6$ and $\mathbf{w}(k) \in \mathbb{R}^4$ represent the white Gaussian transition and measurement noises at time k with covariance matrix $Q(k) \in \mathbb{R}^{6\times6}$ and $W(k) \in \mathbb{R}^{4\times4}$, respectively, $\mathbf{q}(k) \sim \mathcal{N}(0, Q(k))$, $\mathbf{w}(k) \sim \mathcal{N}(0, W(k))$, and $A(k+1)$ is the state transition matrix at time $k+1$,

$$A(k+1) = \begin{bmatrix} \mathbf{I}_{3\times3} & \Delta T(k+1|k)\mathbf{I}_{3\times3} \\ \mathbf{0}_{3\times3} & \mathbf{I}_{3\times3} \end{bmatrix},$$

where $\Delta T(k+1|k)$ is the time interval between time k and time $k+1$. We define $P \in \mathbb{R}^{6\times6}$ as the covariance matrix for the state variable $\mathbf{x}$. The EKF for the system in (4) can be implemented as a state prediction step $\hat{\mathbf{x}}(k|k-1)$, $\hat{P}(k|k-1)$ and a measurement correction step $\hat{\mathbf{x}}(k|k)$, $\hat{P}(k|k)$ recursively as follows,

$$\hat{\mathbf{x}}(k|k-1) = A(k)\hat{\mathbf{x}}(k-1|k-1), \tag{5a}$$

$$\hat{P}(k|k-1) = A(k)\hat{P}(k-1|k-1)A^T(k) + Q(k), \tag{5b}$$

$$K(k) = \frac{\hat{P}(k|k-1)H^T(k)}{H(k)\hat{P}(k|k-1)H^T(k) + W(k)}, \tag{5c}$$

$$\hat{\mathbf{x}}(k|k) = \hat{\mathbf{x}}(k|k-1) + K(k)(\mathbf{z}(k) - h(\hat{\mathbf{x}}(k|k-1))), \tag{5d}$$

$$\hat{P}(k|k) = (\mathbf{I}_{6\times6} - K(k)H(k))\hat{P}(k|k-1), \tag{5e}$$

where $K(k) \in \mathbb{R}^{6\times4}$ is the "Kalman gain" and $H(k) \in \mathbb{R}^{4\times6}$ is the Jacobian matrix of $h(\cdot)$ in (3) with respect to $\mathbf{x}$.

Recall that $\hat{\mathbf{x}}(k|k) = [\hat{\mathbf{p}}(k|k), \hat{\mathbf{v}}(k|k)]^T$. For the n-image motion sequence, the predicted $\hat{\mathbf{x}}(n|n)$ contains the bird velocity $\hat{\mathbf{v}}(n|n)$. The decision of accepting (indicated as "1") or rejecting (indicated as "0") the moving object as a member of the targeted species is defined as the following indicator function,

$$I(\mathbf{Z}^{1:n}) = \begin{cases} 1 & \text{if } \|\hat{\mathbf{v}}(n|n)\| \in \mathcal{V} \text{ and EKF converges,} \\ 0 & \text{otherwise,} \end{cases} \tag{6}$$

where $\mathbf{Z}^{1:n} = \{\mathbf{z}(1), \mathbf{z}(2), ..., \mathbf{z}(n)\}$ is the set of body axes across n frames. $\mathbf{Z}^{1:n}$ is also referred to as the observed data. Eq. (6) basically states that the moving object is a member of the targeted species if the EKF converges to the desired absolute velocity range $\mathcal{V}$.

EKF Convergence

As indicated in (6), automatically determining whether the EKF converges or not is necessary. Define the estimated state set as $\mathbf{X}^{1:n} = \{\hat{\mathbf{x}}(1|1), \hat{\mathbf{x}}(2|2), ..., \hat{\mathbf{x}}(n|n)\}$. Since velocity convergence implies position convergence and $\hat{\mathbf{v}}(k|k)$ converges means $\|\hat{\mathbf{v}}(k|k) - \hat{\mathbf{v}}(k-1|k-1)\| \to 0$, we determine the convergence of the EKF by inspecting

$$\varepsilon(\mathbf{X}^{1:n}) = \sum_{k=2}^{n} \omega(k)\|\hat{\mathbf{v}}(k|k) - \hat{\mathbf{v}}(k-1|k-1)\|,$$

where $\omega(k) > 0$ is the weighting factor at time k. $\omega(k)$ is a monotonically-increasing function of k, which gives more weight to later states. $\omega(k)$ is usually pre-generated using simulated random inputs across the entire possible parameter range without measurement error (i.e. $W(k) = \mathbf{0}_{4\times4}$). Setting $W(k) = \mathbf{0}_{4\times4}$ is to ensure EKF convergency, which will be explained later in the paper. Denote $\|\hat{\mathbf{v}}\|$ as the speed of the bird known in each trial of simulation. We repeat the EKF with randomized inputs for over 10^6 times to observe the quantity of $\frac{\|\hat{\mathbf{v}}\|}{\|\hat{\mathbf{v}}(k|k)-\hat{\mathbf{v}}(k-1|k-1)\|}$, which is the inverse of the relative speed change at time k. We choose the weighting factor as $\omega(k) = E\left(\frac{\|\hat{\mathbf{v}}\|}{\|\hat{\mathbf{v}}(k|k)-\hat{\mathbf{v}}(k-1|k-1)\|}\right)$, where function $E(\cdot)$ computes the expected value over all simulation trials for the targeted species. When the EKF converges, $\|\hat{\mathbf{v}}(k|k) - \hat{\mathbf{v}}(k-1|k-1)\|$ appears as a decreasing function of k after a few initial steps. Correspondingly, $\omega(k)$ is an increasing function of k. If $\|\hat{\mathbf{v}}(k|k) - \hat{\mathbf{v}}(k-1|k-1)\| \to 0$, then $\varepsilon(\mathbf{X}^{1:n})$ is smaller than that of the case $\|\hat{\mathbf{v}}(k|k) - \hat{\mathbf{v}}(k-1|k-1)\| \nrightarrow 0$. Therefore, to determine the EKF convergence, we employ a threshold δ on $\varepsilon(\mathbf{X}^{1:n})$ and introduce a new indicator variable,

$$I_{\text{EKF}}(\mathbf{X}^{1:n}) = \begin{cases} 1 \text{ (converge) if } \ \varepsilon(\mathbf{X}^{1:n}) < \delta, \\ 0 \text{ otherwise.} \end{cases} \tag{7}$$

Then the decision-making in (6) is re-written as,

$$I(\mathbf{Z}^{1:n}) = \begin{cases} 1 \ \text{ if } \|\hat{\mathbf{v}}(n|n)\| \in \mathcal{V} \text{ and } I_{\text{EKF}}(\mathbf{X}^{1:n}) = 1, \\ 0 \ \text{ otherwise.} \end{cases} \tag{8}$$

The underlying condition for (8) to be an effective bird filtering mechanism is that if observation $\mathbf{Z}^{1:n}$ is caused by the targeted bird species then the convergence of the EKF has to be guaranteed. Unfortunately, the condition usually does not hold due to two reasons: n is small and the measurement noise $\mathbf{w}(k)$ is too big. Due to the fact that the bird flies very fast, the bird usually stays in the camera FOV for less than 1 second and thus $n < 11$ for most cases in our experiments. The measurement noise covariance matrix $W(k)$ is directly determined by the image segmentation error. Even at 1 pixel level, its relative range is 4% for a bird body length of 25 pixels. For the nonlinear deterministic discrete time system in (4), the large $W(k)$ means the EKF either fails to converge or converges very slowly according to (Boutayeb, Rafaralahy, and Darouach 1997). The performance of the bird filter would be close to that of a random guess if the simple EKF implementation is used, which makes it useless.

Probable Observation Data Set-based EKF Method

Since simply applying EKF cannot address the bird filtering problem, a new approach is required. Let us first assume there is no measurement noise (i.e. $W(k) = \mathbf{0}_{4\times4}$) and no state transition noise $Q(k) = \mathbf{0}_{6\times6}$. At each time k, the EKF in (5) is a system of equations with four inputs, which is the dimensionality of $\mathbf{z}$, and six outputs, which is the dimensionality of $\mathbf{x}$. We also know that matrix A introduces two constraints: the constant speed and the linear trajectory. Therefore, the equation system can be solved within one step. The convergence of the EKF is not a problem when there is no noise and the bird does not fly in a degenerated trajectory (i.e. flying along the camera optical axis). Although usually $Q(k) \neq \mathbf{0}_{6\times6}$, the state transition noise $\mathbf{q}(k)$ is often very small, which leads to the following lemma,

Lemma 1. *The EKF described in (5) converges when* $W(k) = \mathbf{0}_{4\times4}$.

Proof. We skip the proof because our system in (4) is a linear time-invariant discrete time system with a nonlinear observer. The convergence of its EKF can be proved by the results in (Boutayeb, Rafaralahy, and Darouach 1997). □

At first glance, this result is useless because we cannot get rid of the measurement noise. However, this result opens the door to a new approach. Define the observation data without measurement error as $\mathbf{Z}^{1:n*} = [\mathbf{z}^*(1), \mathbf{z}^*(2), ..., \mathbf{z}^*(n)]^T$. Although we do not have $\mathbf{Z}^{1:n*}$, we know it is within the segmentation error range of $\mathbf{Z}^{1:n}$. For the k-th image, the measurement data is $\mathbf{z}(k) = [u^h(k), v^h(k), u^t(k), v^t(k)]^T$. Define the error-free measurement data at time k as $\mathbf{z}^*(k) = [u^{h*}(k), v^{h*}(k), u^{t*}(k), v^{t*}(k)]^T$. Let us define the segmentation error is within τ pixels. The value of τ is usually small since the camera is set with short shuttle time and small iris. Define $S_1(k) = [u^h(k) \pm \tau]$, $S_2(k) = [v^h(k) \pm \tau]$, $S_3(k) = [u^t(k) \pm \tau]$, $S_4(k) = [v^t(k) \pm \tau]$, and the segmentation error range at time k as $\mathbb{S}(k)$. Hence, $\mathbf{z}^*(k) \in S_1(k) \times S_2(k) \times S_3(k) \times S_4(k) = \mathbb{S}(k)$.

Definition 1. *Define the probable observation data set (PODS)* $\mathbb{Z}^{1:n}$ *as the set of observation data* $\mathbf{Z}^{1:n}$ *that satisfies the condition for the EKF convergence as in (7),* $\mathbb{Z}^{1:n} = \{\mathbf{Z}^{1:n} | \mathbf{z}(k) \in \mathbb{S}(k), k = 1, ..., n,$ *and* $\varepsilon(\mathbf{X}^{1:n}) < \delta\}$.

Obviously $\mathbf{Z}^{1:n*} \in \mathbb{Z}^{1:n}$. Each $\mathbf{Z}^{1:n} \in \mathbb{Z}^{1:n}$ is likely to be $\mathbf{Z}^{1:n*}$ and hence it is named as the probable observation data. For a given PODS $\mathbb{Z}^{1:n}$, there is a corresponding estimated state set $\mathbb{X}^{1:n}$, which contains a set of all possible estimated velocities at time n, which is defined as $\mathbb{V} = \{\|\hat{\mathbf{v}}(n|n)\| \text{ such that } \mathbf{X}^{1:n} \in \mathbb{X}^{1:n}\}$. Then the decision making in (8) can be written as,

$$I(\mathbf{Z}^{1:n}) = \begin{cases} 1 \ \text{ if } \ \mathbb{V} \cap \mathcal{V} \neq \emptyset \text{ and } \mathbb{Z}^{1:n} \neq \emptyset, \\ 0 \ \text{ otherwise.} \end{cases}$$

Hence we have the following lemma,

Lemma 2. *If the non-degenerated observation data* $\mathbf{Z}^{1:n}$ *is triggered by a bird of the targeted species, then* $I(\mathbf{Z}^{1:n}) = 1$.

Proof. Since $\mathbf{Z}^{1:n}$ is triggered by the targeted species, its corresponding $\mathbf{Z}^{1:n*}$ ensures the convergence of the EKF according to Lemma 1. Define $\mathbf{X}^{1:n*}$ as the corresponding estimated states for $\mathbf{Z}^{1:n*}$. Hence $\varepsilon(\mathbf{X}^{1:n*}) < \delta \to \mathbb{Z}^{1:n} \neq \emptyset$, because $\mathbf{Z}^{1:n*} \in \mathbb{Z}^{1:n}$. Following our naming convention, $\hat{\mathbf{v}}^*(n|n)$ is the velocity component of $X^*(n|n) \in \mathbf{X}^{1:n*}$. Since the observation data is not degenerated, $\|\hat{\mathbf{v}}^*(n|n)\| \in \mathcal{V}$. We also know $\|\hat{\mathbf{v}}^*(n|n)\| \in \mathbb{V}$ by definition, $\mathbb{V} \cap \mathcal{V} \neq \emptyset$ holds. Since both conditions are satisfied, $I(\mathbf{Z}^{1:n}) = 1$. □

Lemma 2 ensures that the PODS-EKF method theoretically has a zero false negative rate, which is a very desirable property for the purpose of searching for rare bird species.

Approximate Computation for PODS-EKF

Computing the PODS $\mathbb{Z}^{1:n}$ is nontrivial. Note that we actually do not need $\mathbb{Z}^{1:n}$ because all we need to know is whether the conditions $\mathbb{V} \cap \mathcal{V} \neq \emptyset$ and $\mathbb{Z}^{1:n} \neq \emptyset$ hold or not. This allows an approximation method. For a given observation $\mathbf{Z}^{1:n}$, we define the following optimization problem,

$$\widetilde{\mathbf{Z}}^{1:n} = \arg \min_{z(k)\in\mathbb{S}(k);k=1,\ldots,n} \varepsilon(\mathbf{X}^{1:n}), \tag{9}$$

where $\widetilde{\mathbf{Z}}^{1:n}$ is the optimal solution to the minimization problem above. Actually, (9) is a typical nonlinear optimization problem with the error range $z(k) \in \mathbb{S}(k); k = 1, \ldots, n$ and the EKF in (5) as constraints. Define $\widetilde{\mathbf{X}}^{1:n} = \{\widetilde{\mathbf{x}}(1|1), \widetilde{\mathbf{x}}(2|2), \ldots, \widetilde{\mathbf{x}}(n|n)\}$ as the estimated states corresponding to $\widetilde{\mathbf{Z}}^{1:n}$. We have the following lemma,

Lemma 3. $\varepsilon(\widetilde{\mathbf{X}}^{1:n}) > \delta \iff \mathbb{Z}^{1:n} = \emptyset$.

Proof. Since (9) is a minimization problem, $\widetilde{\mathbf{X}}^{1:n}$ yields the minimal $\varepsilon(\mathbf{X}^{1:n})$, namely, $\varepsilon(\widetilde{\mathbf{X}}^{1:n}) > \delta \iff \varepsilon(\mathbf{X}^{1:n}) > \delta, \forall \mathbf{X}^{1:n} \in \mathbb{X}^{1:n} \iff \mathbb{Z}^{1:n} = \emptyset$. □

It is worth mentioning that this method is an approximation in computation because the nonlinear programming solver often falls in a local minimum. Now we want to determine whether $\mathbb{V} \cap \mathcal{V} \neq \emptyset$. If we view the EKF output $\hat{\mathbf{v}}(n|n)$ as a function of $\mathbf{Z}^{1:n}$, it is continuous and differentiable with respect to each entry in $\mathbf{Z}^{1:n}$. Since $\mathbb{Z}^{1:n}$ is actually very small, the variance of the speed in the set $\mathbb{V}$ is very small. Instead of comparing $\mathbb{V}$ to $\mathcal{V}$, we select a value in $\mathbb{V}$ to check if it is in $\mathcal{V}$. Define $\widetilde{\mathbf{v}}(n|n)$ as the velocity component of $\widetilde{\mathbf{x}}(n|n) \in \widetilde{\mathbf{X}}^{1:n}$. The chosen value is the $\|\widetilde{\mathbf{v}}(n|n)\|$ because it is readily available. Therefore, the approximation is $\|\widetilde{\mathbf{v}}(n|n)\| \in \mathcal{V} \iff \mathbb{V} \cap \mathcal{V} \neq \emptyset$. Due to the approximation, the zero false negative rate cannot be guaranteed. However, the false negative rate is still very low under the approximation as shown in the experiments.

Experiments

We have implemented the PODS-EKF algorithm and tested the algorithm on both the simulated data and the real data from field experiments. The computer used in the test is a desktop PC with a Pentium(R) D 3.20GHz CPU and 2GB RAM. The PC runs Windows XP. The PODS-EKF has been implemented using Matlab 7. We choose Arecont Vision 3100 high resolution networked video cameras as the imaging devices. The camera runs at 11 fps with a resolution of 3 Mega-pixel per frame. The lens for the camera is a Tamron auto iris vari-focus lens with a focal length range of 10-40 mm. We adjust the lens to ensure a 20° horizontal FOV.

Simulation

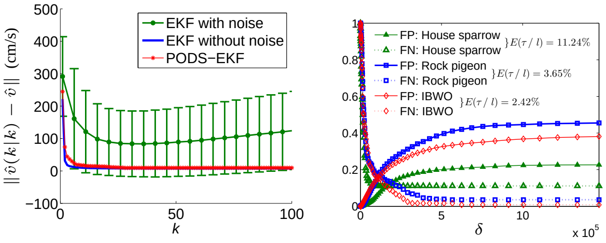
We first test our PODS-EKF using the simulated inputs. An intermediate step is to analyze the convergence of the EKF. We generate 10^6 random 3D bird trajectories with constant velocity, which intersect with the conic camera FOV. The trajectory segment intersecting with the camera FOV projects back to the image plane as the visual measurements.

Table 1: Species used in the experiments

Species	l_b (cm)	$\mathcal{V}$ (km/h)
House Sparrow	15 [1]	[29, 40] [2]
Rock pigeon	33 [3]	[24, 56] [4]
Ivory-billed woodpecker	48 [5]	[32, 64] [6]

We simulate three types of birds: house sparrows, rock pigeons, and IBWOs (Table 1). The three species are small, medium, and large in size, respectively. Their speeds range from 24 to 64 km/h and cover the range of the most of existing bird species. The initial state $\hat{\mathbf{x}}(0|0)$ is estimated by solving the linear equation system in (3) with the first at least 3 observations and an additional constraint/guess $\|\hat{\mathbf{v}}(0|0)\| = E(\mathcal{V})$. Details are omitted for space constraint.

Fig. 3(a) shows the EKF convergence for rock pigeon in different configurations by tracking the signal $\|\hat{\mathbf{v}}(k|k) - \hat{\mathbf{v}}\|$, where $\hat{\mathbf{v}}$ is the true bird velocity known in simulation. It is shown that without image noise, the EKF nicely converges as Lemma 1 points out. With the image noise ($\tau = 1$ pixel), the EKF diverges with both increasing signal mean and variance. The PODS-EKF on the other hand, ensures the EKF to converge very close to the noise-free case. This validates the foundation of our PODS-EKF method.



(a) (b)

Figure 3: (a) Convergence for different EKF configurations based on simulated rock pigeon data. (b) False positive (FP) and false negative (FN) rates with different δ in simulation.

Now we are ready to analyze the performance of PODS-EKF. We generate a set of random inputs to mimic three birds as in Table 1. We set a speed range from 15 to 85 km/h with an incremental step of 5 km/h and a bird size range from 10 to 60 cm with an incremental step of 2 cm. We set the segmentation error range $\tau = 1$ pixel. For each setting of the input data, 20 trials are carried out. The average computation time for each trial is 5.6 seconds. Fig. 3(b) demonstrates how the rates of false positive (FP) and false negative (FN) change according to δ. After $\delta > 1.0 \times 10^6$, the FN rates can be reasonably controlled to be less than 10%, 4% and 1%, for house sparrow, rock pigeon and IBWO, respectively. This confirms that the approximation computation is reasonable. The reason PODS-EKF works worst for house

[1] http://en.wikipedia.org/wiki/House_Sparrow.

[2] http://www.garden-birds.co.uk/information/flight.htm

[3] http://www.allaboutbirds.org/guide/Rock_Pigeon/lifehistory

[4] http://www.ct.gov/DEP/cwp/view.asp?A=2723&Q=326076

[5] http://animals.nationalgeographic.com/animals/birds/ivory-billed-woodpecker.html

[6] http://news.mongabay.com/2007/0217-ibw.html

sparrow is that with the same FOV in the simulation, the smallest house sparrow leads to the highest noise-signal ratio, indicated as $E(\tau/l)$ in Fig. 3(b). Our PODS-EKF is not biased for particular bird. To cope with small birds, we can increase the focal length to reduce $E(\tau/l)$. The FP rates of the PODS-EKF converge to 23%, 45% and 38%, respectively, which are a little high. However, considering that we are comparing the targeted bird with birds similar in size and speed, this result is not surprising. In fact, the algorithm should behave better in real tests where the noise from the moving objects has much larger range in both size and speed. Furthermore, the monocular system has its problem in detecting objects with their trajectories close to the camera optical axis, which also contributes to the high FP rate.

Physical Experiments

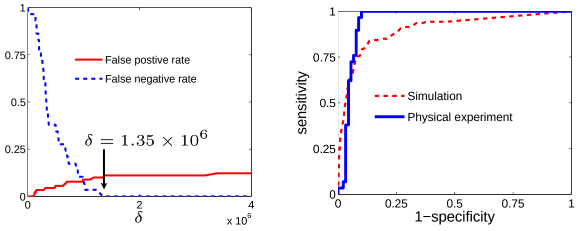
We have conducted a field experiment of detecting flying rock pigeons. With a camera setup in the forest in Bayou DeView in eastern Arkansas, we have captured 119 events with $n \geq 8$ for each motion sequence. 29 of the sequences are rock pigeons while the other 90 are not pigeons, which are image sequences of typical environment noises such as falling leaves, flying insects, and other bird species. Fig. 4(a) shows how the FN and FP rates change according to difference δ. The convergence threshold is set as $\delta = 1.35 \times 10^6$. It is shown that our algorithm can achieve extremely low FN rate ($0/29 = 0\%$). This is very important for the purpose of finding rare bird species. The FP rate is $9/90 = 10\%$, which is better than that of the simulation results. This is due to the fact that it is much easier for the algorithm to distinguish the targeted species from noises such as flying insects and falling leaves in real experiment rather than from similar birds as in simulation above.

Fig. 4(b) illustrates the ROC curves from both the simulation and physical experiment for the rock pigeon. The convergence threshold range is $[4.6 \times 10^3, 1.5 \times 10^6]$ for simulation and $[1.8 \times 10^4, 3.3 \times 10^6]$ for experiment. The areas under the ROC curve are 91.5% and 95.0% for simulation and physical experiment, respectively, which again shows the algorithm performs better in physical experiments.

We apply the algorithm to our IBWO search field data (Oct. 2006 - Oct. 2007). After initial motion detection filtering as in (Song et al. 2008), we reduce the total video data of 29.41 TB to 27.42 GB, which is still prohibitively huge for human experts. After applying the PODS-EKF, we eventually reduce the data volume to 146.7 MB (about 960 images) with close to zero false negative. The overall reduction rate is 99.9995%.

Conclusion

We reported our development of a bird filtering algorithm. We developed a novel Probable Observation Data Set (PODS)-based EKF method to ensure the convergence of the EKF under insufficient and noisy measurement data. The algorithm has been extensively tested using both simulated inputs and physical experiments. The results showed that the PODS-EKF bird filter has reduced the video data by 99.9995% with close to zero false negative and 95.0% area under the ROC curve in physical experiments.



(a) (b)

Figure 4: (a) Physical experiment for rock pigeons. (b) The ROC curves for both the simulation and the physical experiment. The corresponding areas under the ROC curves are 91.5% and 95.0%, respectively.

References

Boutayeb, M.; Rafaralahy, H.; and Darouach, M. 1997. Convergence analysis of the extended kalman filter used as an observer for nonlinear deterministic discrete-time systems. *IEEE Transactions on Automatic Control* 42(4):581–586.

Briassouli, A., and Ahuja, N. 2007. Extraction and analysis of multiple periodic motions in video sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29(7):1244–1261.

Laptev, I.; Belongie, S. J.; Pérez, P.; and Wills, J. 2005. Periodic motion detection and segmentation via approximate sequence alignment. In *IEEE International Conference on Computer Vision (ICCV)*, 816–823.

Ramanan, D.; Forsyth, D.; and Barnard, K. 2006. Building models of animals from video. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28(8):1319–1334.

Ran, Y.; Weiss, I.; Zheng, Q.; and Davis, L. S. 2007. Pedestrian detection via periodic motion analysis. *International Journal of Computer Vision (IJCV)* 71(2):143–160.

Ribnick, E.; Atev, S.; and Papanikolopoulos, N. P. 2009. Estimating 3d positions and velocities of projectiles from monocular views. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31(5):938–944.

Saxena, A.; Sun, M.; and Ng, A. 2008. Make3d: Depth perception from a single still image. In *Proc. of The AAAI Conference on Artificial Intelligence*, 1571–1576.

Sebastian, T.; Klein, P.; and Kimia, B. 2004. Recognition of shapes by editing their shock graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26(5):550–571.

Song, D., and Goldberg, K. 2007. Approximate algorithms for a collaboratively controlled robotic camera. *IEEE Transactions on Robotics* 23(5):1061–1070.

Song, D.; Qin, N.; Xu, Y.; Kim, C. Y.; Luneau, D.; and Goldberg, K. 2008. System and algorithms for an autonomous observatory assisting the search for the ivory-billed woodpecker. In *IEEE International Conference on Automation Science and Engineering*.

Song, D.; van der Stappen, F.; and Goldberg, K. 2006. Exact algorithms for single frame selection on multi-axis satellites. *IEEE Transactions on Automation Science and Engineering* 3(1):16–28.

Song, D. 2009. *Sharing a Vision: Systems and Algorithms for Collaboratively-Teleoperated Robotic Cameras*. Springer.

Spinello, L.; Triebel, R.; and Siegwart, R. 2008. Multimodal people detection and tracking in crowded scenes. In *Proc. of The AAAI Conference on Artificial Intelligence*.

Yilmaz, A.; Javed, O.; and Shah, M. 2006. Object tracking: A survey. *ACM Computing Surveys* 38(4):1–45.