Autonomous Ground-Based Tracking of Migrating Raptors using Vision

Sana Sarfraz*
Jack W. Langelaan†
The Pennsylvania State University, University Park, PA 16802, USA

The following paper presents the design and simulation of a system of ground-based stereo cameras to enable the autonomous tracking of migratory raptors in order to facilitate the study of their flight patterns. A number of stations consisting of a camera pair and a processor are set up along the ridge under observation. Each station computes an estimate of the position and velocity of the birds viewed by its cameras using an Unscented Kalman Filter. The estimates from each station are transmitted to a master computer that fuses all the independent estimates and transmits the resulting estimate back to all the stations. Results of Monte Carlo simulations show the convergence of the estimated error to the true error for estimates from one or more stations.

I. Introduction

Migration places extreme demands on birds asthey travel long distances.[1] The ability to conserve energy through the use of atmospheric lift is crucial for successful migration. The long ridges of the Appalachian Mountains in Pennsylvania, as shown in Figure 1 are a critical migration corridor for North American raptors and songbirds. Over 150 species use this migration corridor,[2] and some ridges are renowned for local concentrations of hawks and eagles. An estimated 800 Golden Eagles migrate through Pennsylvania each year, with many of them passing by Tussey Mountain (approximately 10km east of Penn State's University Park campus). Factors influencing when and where they stop to feed are not well understood, and may include internal conditions such as amount of body fat and external conditions such as local weather. By collecting data on individual behavior in response to internal and external conditions researchers will be able to build models of bird behavior and identify critical habitats along migration routes.

Figure 1. A satellite image of the Appalachian Mountains in Pennsylvania.

Current methods used for tracking bird migration involve either teams of human observers stationed at observation points along the predicted migration route[3] or tagging individual birds with a GPS receiver and data transmitter.[3,4] The human observers are able to provide a count of bird species, thus giving a big-picture view of migration, and GPS data gives a detailed view of the migration of a single bird. Cinetheodolites have been used to enable measurements of bird track and speed, but these are expensive and cumbersome.[5]

*Graduate Student, Department of Aerospace Engineering
†Assistant Professor, Department of Aerospace Engineering, Senior Member AIAA.

This paper describes a vision-based system for autonomously tracking birds as they fly past an observation site along the migratory route. In addition to migration studies, such a system can be used to examine changes in flight patterns caused by obstructions such as wind turbines. Wind power development and associated wind turbines are increasing throughout the U.S., and there is great interest in establishing the infrastructure to use wind as a renewable energy source in Pennsylvania. Wind turbines are often established along mountain ridges: in Pennsylvania these ridges are also prime migration routes. Biologists have expressed concerns that wind turbines can result in direct mortality and avoidance behavior along migration routes.[6,7]

The remainder of this paper is organized as follows: Section II presents previous work in the field of bird monitoring and tracking; Section III describes the bird tracking problem and the system designed to tackle it; Section IV presents the design of the estimator used to compute the state of the bird while addressing the problem of data association and track initiation; Section V tabulates the simulation parameters and presents the results obtained for a Monte Carlo simulation that uses the stated values.

II. Previous and Related Work

The most popular method for studying hawk migration has been counting of birds by human observers, followed by trapping and banding.[5] Counting is best for monitoring the population but it is not enough to understand the flight patterns of raptors. Tagging gives additional information about geographic origins and approximate flight paths but the low recapture rates of banded birds makes it impossible to reliably answer important questions. Radiotelemetry, radar and visual techniques have provided the best data for the study of flight patterns but they can be prohibitively expensive.[5] In the case of observing the effect of wind farms on the mortality of migrating hawks, periodic mortality surveys are carried out by teams of observers at the site specified. The mortality rate is observed by counting carcasses of birds at periodic intervals. The effect of scavenging by other predators also has to be accounted for in order to avoid the underestimation of mortality rates. The effects of location, weather and flight behavior on collision possibilities have previously been analyzed using generalized linear modelling.[7] Although a valid concern, previous studies indicate that avian collision mortality associated with wind plants is much lower than other sources of collision mortality. Using the annual avian collision mortality estimate of 200-500 million, it is estimated that wind turbines constitute 0.01 percent to 0.02 percent of the avian collision fatalities.[8] The primary reason is the low relative numbers of wind farms to other man made structures and the fact that many of the wind plants are located in areas with relatively low bird and raptor use.

An autonomous system to actively monitor the flight paths of raptors can add more insight to the available information without incurring the cost associated with having human observers and expensive equipment. Cameras are inexpensive sensors that impart a wealth of information pertaining to color, prominent features, position and motion of filmed objects. Stereo camera systems also allow triangulation to add depth information and provide measurements from various viewing angles to significantly lower uncertainty about the state of the subject under observation. Stereo vision has been used by Muñoz-Salinas and Auirre to track moving people using a Kalman filter.[9] The unique color information associated with each subject being tracked has been used by the aforementioned, as well as Bahadori and Iocchi, to perform data association and maintain the identities of people being tracked.[9,10] Color histogram based tracking has been made popular by Zivkovic and Krose,[11] but it is computationally expensive to extract color information in every frame. Using a Kalman filter for frame to frame tracking and color information over longer time intervals, to maintain data association, is more viable for real-time applications.

Distributed systems have received increasing attention due to the several advantages that they offer over centralized systems. These include: lowered computational dependence at any one node; the option of incorporating any number of measurements or variety of sensors into the tracking process; the ability to cover large areas of observation. The theory and implementation algorithms of Distributed Sensor Networks(DSN) for tracking multiple targets is presented by Bar-Shalom and Li.[12] Tracking of human subjects and mobile robots in an indoor smart environment using has been achieved by Karuppiah and Zhu through a distributed vision system with heterogenous sensors of various processing rates, synchronized and fused to achieve real-time tracking.[13]

The following paper has built upon previous tracking methods to simulate the outdoor, long-distance tracking of raptors using stereo camera systems and distributed processors, implemented using various forms of the Kalman Filter.

III. The Bird Tracking Problem

A. Tracking Scenario

THE PROBLEM CONSIDERED here is estimating the positions and speeds of birds as they pass the ground-based observation platforms. For a single bird the vector of states to be estimated includes components of position x^o, y^o and z^o and velocity $\dot{x}^o$, $\dot{y}^o$ and $\dot{z}^o$ expressed in the local NED frame:

$$\mathbf{x}^o_{b,i} = \begin{bmatrix} x^o_{b,i} & y^o_{b,i} & z^o_{b,i} & \dot{x}^o_{b,i} & \dot{y}^o_{b,i} & \dot{z}^o_{b,i} \end{bmatrix}^T \tag{1}$$

Available measurements are bearings obtained from an array of cameras placed at known positions and orientation. This is shown schematically in Figure 2.

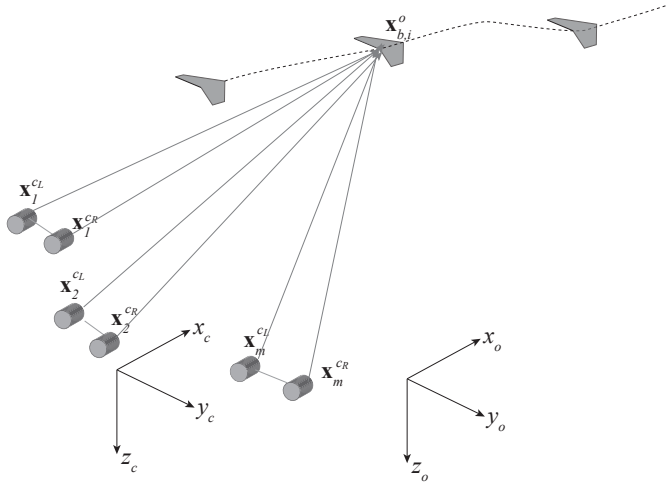


Figure 2. Schematic of the bird tracking problem.

The flight path of a bird, b, is shown. The state of the bird, $\mathbf{x}^o_{b,i}$, must be estimated at every instant of time, i. The coordinates $(x_o y_o z_o)$ represent position in global NED coordinate system. The system of sensors consists of M camera stations. Each station has a stereo system of M_n cameras. Every camera in the system is considered to have its own local coordinate system, with the x axis pointing outwards from the lens along the optical axis of the camera and the y and z determined by the pan or tilt of the camera about its origin. Each camera in the stereo system has its own origin and rotation angles. The coordinates $(x_c y_c z_c)$ represent the position of the bird as seen in the local coordinates of the camera in which it is viewed. Available measurements are bearings obtained from the array of cameras placed at known positions and orientation.

Computing position estimates from bearing measurements involves trigonometric functions. The tracking problem is therefore a non-linear one. The Kalman Filter, which is the optimal linear filter, will not be able to provide a solution within the desirable limits of uncertainty. A tracking method that can produce fairly accurate estimates despite the non-linearity of the system equations must be employed to achieve the results required. Tracking closely spaced, distant objects with a limited sensors can be difficult due to the high degree of uncertainty introduced by Dilution of Precision. Having symmetrically spread out sensors over large distances can counter this problem since the ambiguity in opposite directions cancel out. Implementation details, however, dictate that the distance between cameras in a stereo pair, or the baseline, is not more than a few meters. It is difficult to transmit data to a central station over a large baseline. Having a short baseline increases the Dilution of Precision (DOP), especially when the bird is further away. However, merely spreading cameras to obtain widely dispersed bearing measurements introduces difficulties related to transmitting data over long distance. The designed system must therefore address the implementation issues presented by the terrain of observation site and the limitations of the hardware available to achieve a sensor distribution that can counter DOP to obtain reasonable estimates of the bird state.

B. System Description

The requirement for a spatially spread out and robust system has been met by designing a Distributed System which has several processors, each connected to a stereo camera pair, independently tracking the birds and a central processing unit that fuses all available estimates to compute a more accurate global estimate of the states.

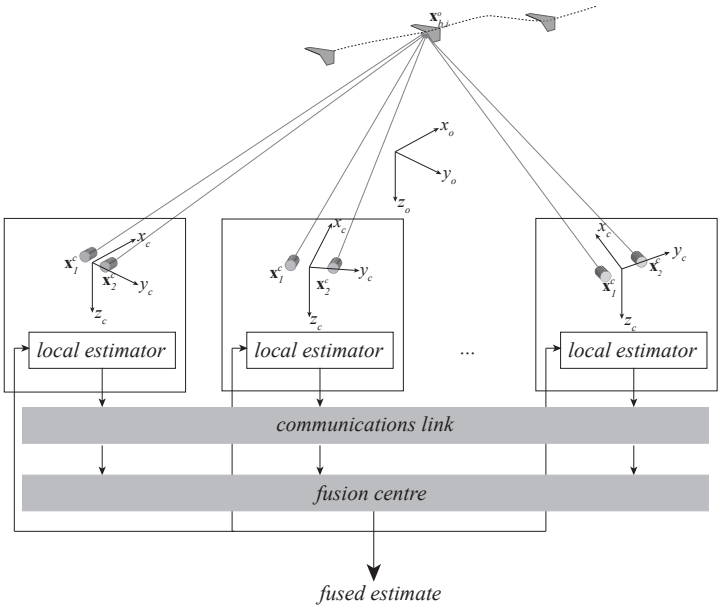


Figure 3. Schematic of the Distributed Hierarchical Estimation system.

A set of stations, each consisting of a camera pair, a processor and a two-way communication link form the components of the distributed system to be employed for estimating the position and velocity of the birds in view. The cameras are spread out along the ridge with overlapping fields of view to obtain multiple, simultaneous measurements of a bird. The DOP is therefore reduced due to the geometry of the array of cameras. Although a Sigma Point Kalman Filter may be used to fuse the available measurements at a central processor to compute one estimate, this approach makes the operation of one processor critical to the performance of the entire system. Therefore, a hierarchical approach is proposed for this project as shown in Figure 3, wherein bearings from pairs of cameras are fused using an Unscented Kalman Filter (UKF)[14,15] to compute independent estimates at numerous stations. The prediction step is driven by bird kinematics (Equation 2) and the measurement update step is computed using bearings from the cameras (Equation 19). Since the state estimates for different birds are uncorrelated a separate UKF is initiated for each bird as it enters the field of view.

The local estimates are transmitted to a central processor that fuses all available estimates and transmits a global estimate back to all stations, which is used by each station to correct its prediction for the consequent time step. Although this approach still relies on the central processor for fusing the estimates, its operation is no longer critical to the estimation process. The load on the processor and communication link is reduced, and in the event of their failure, track is still maintained at each station.

C. Models for Measurements and Bird Flight

A constant velocity model will be used to model bird flight. The choice of states results in a linear model for bird kinematics:

$$\dot{\mathbf{x}}^o_{b,i} = \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{x}^o_{b,i} + \begin{bmatrix} \mathbf{0} \\ \mathbf{I} \end{bmatrix} \mathbf{v} \tag{2}$$

where $\mathbf{I}$ is the 3×3 identity matrix and $\mathbf{v}$ is zero-mean Gaussian random noise.

Rather than using direct stereo vision techniques for computing range to a target (i.e. computing range based on disparity between the left and right cameras), the bearing from each camera to every bird is computed and treated as an independent measurement. The bearings obtained by the cameras at a station are fused using a Sigma Point Kalman filter. This method has the advantage of being easily scalable to adding more measurements and adaptable to the number of cameras at the station in which a bird is in view.

A pinhole camera model defines the projection of a vector onto the image plane as

$$\boldsymbol{\gamma} = \frac{f}{x^c} \begin{bmatrix} y^c \\ z^c \end{bmatrix} \tag{3}$$

where f is the focal length and $\mathbf{x}^c = \begin{bmatrix} x^c & y^c & z^c \end{bmatrix}^T$ is the vector expressed in the camera frame. The focal length f can be normalized without loss of generality.

The orientation of the camera frame for camera m, C_m ,with respect to the inertial frame O is assumed to be known and is given by a rotation matrix $\mathbf{T}^c_m$. Each camera is assumed to have its optical axis aligned with the camera frame's $\hat{\mathbf{x}}^c$ axis and is offset from the origin of the camera frame by a vector $\Delta\mathbf{x}^c_m = \begin{bmatrix} \Delta x^c_m & \Delta y^c_m & \Delta z^c_m \end{bmatrix}^T$. A bearing measurement to the i^{th} bird at camera m is therefore

$$\boldsymbol{\gamma}_{m,i} = \frac{1}{x^c_{m,i}} \begin{bmatrix} y^c_{m,i} \\ z^c_{m,i} \end{bmatrix} \tag{4}$$

where

$$\begin{bmatrix} x^c_{m,i} \\ y^c_{m,i} \\ z^c_{m,i} \end{bmatrix} = \mathbf{T}^c_m \begin{bmatrix} x^o_i - x_o \\ y^o_i - y_o \\ z^o_i - z_o \end{bmatrix} + \begin{bmatrix} \Delta x^c_m \\ \Delta y^c_m \\ \Delta z^c_m \end{bmatrix} \tag{5}$$

The measurement vector $\mathbf{z}$ is formed by concatenating bearings from each of the M cameras in the array:

$$\mathbf{z} = \begin{bmatrix} \boldsymbol{\gamma}_1 & \boldsymbol{\gamma}_2 & \cdots & \boldsymbol{\gamma}_M \end{bmatrix}^T \tag{6}$$

IV. Estimator Design

A. Non-Linear and Distributed Estimation

The Unscented Kalman Filter (UKF) employs the propagation of particles through the system in order to deal with its non-linearities.[16] The particles used by the UKF are obtained deterministically and not randomly because the UKF is based on the assumption that the particles distributed through it have a Gaussian distribution. For an n state estimation, 2n+1 particles, or Sigma Points, are sampled symmetrically about the mean. These Sigma Points have a distribution with mean and covariance of a random vector $\mathbf{x} = \mathcal{N}(\hat{\mathbf{x}}, \mathbf{P})$. The set of sigma points is given by:

$$\mathbf{X} = \begin{bmatrix} \bar{\mathbf{x}} & \bar{\mathbf{x}} + \eta\sqrt{\mathbf{P}} & \bar{\mathbf{x}} - \eta\sqrt{\mathbf{P}} \end{bmatrix} \tag{7}$$

where η is a scale factor that determines the spread of the Sigma Points about the mean and $\sqrt{\mathbf{P}}$ is an orthogonal matrix square root of the covariance of the distribution. The Sigma Points are propagated through the prediction and correction steps of the Kalman Filter and the mean and covariance of these points give the state estimate, $\hat{\mathbf{x}}$, and the associated covariance, $\mathbf{P}$.

Local estimates are computed at each station using the UKF described above. All local estimates are transmitted to a global fusion center to compute a more accurate global estimate. It is important to realize that the estimates from each station are not entirely independent and that they share the same process noise. The fusing algorithm must account for the shared noise at the processors in order to avoid overconfidence in the estimated state caused by adding the same information more than once. A global state fusion algorithm based on the information form of the Kalman Filter is employed, which accounts for the correlation of local estimates.[12,17] The local estimates computed by the UKF are gaussian and retain this distribution despite

the non-linearity of system equations they are propagated through. A linear gaussian estimation procedure may therefore be assumed. The fusion equations for hierarchical estimation in a linear gaussian case are presented in the compilation by Bar-Shalom.[12] An unbiased estimate is achieved by subtracting the predicted estimate and covariance of each station from their corrected values, to account for the correlation between node estimates, before adding them to the global predictions of covariance and estimates. The notation of the equations presented below is modified from the source for convenience. The global, a posteriori estimate, ${}^{F}\hat{\mathbf{x}}_{k|k}$, and the associated covariance ${}^{F}\mathbf{P}_{k|k}$ are similarly achieved by fusing the local estimates, ${}^{L}\hat{\mathbf{x}}_{i,k|k}$ and the covariance associated with each of them, ${}^{L}\mathbf{P}_{i,k|k}$ as follows[12] :

$$ {}^{F}\mathbf{P}^{-1}_{k|k} = {}^{F}\mathbf{P}^{-1}_{k|k-1} + \sum_{i=1}^{n} ({}^{L}\mathbf{P}^{-1}_{i,k|k} - {}^{L}\mathbf{P}^{-1}_{i,k|k-1}) \tag{8} $$

$$ {}^{F}\mathbf{P}^{-1}_{k|k}\,{}^{F}\hat{\mathbf{x}}_{k|k} = {}^{F}\mathbf{P}^{-1}_{k|k-1}\,{}^{F}\hat{\mathbf{x}}_{k|k-1} + \sum_{i=1}^{n} ({}^{L}\mathbf{P}^{-1}_{i,k|k}\,{}^{L}\hat{\mathbf{x}}_{i,k|k} - {}^{L}\mathbf{P}^{-1}_{i,k|k-1}\,{}^{L}\hat{\mathbf{x}}_{i,k|k-1}) \tag{9} $$

The fused covariance and estimates thus obtained are transmitted back to all the stations and updated to obtain a priori estimated in the next time step.

$$ {}^{L}\mathbf{P}_{i,k|k-1} = {}^{F}\mathbf{P}_{k|k-1} \tag{10} $$

$$ {}^{L}\hat{\mathbf{x}}_{i,k|k-1} = {}^{F}\hat{\mathbf{x}}_{k|k-1} \tag{11} $$

Transmitting the fused estimates back to all stations ensures that there is no drift between stations over time and all stations benefit from the information from all other station at each time step, which leads to an improved estimate in the next time step.

B. Data association

Data association can be a difficult issue in many tracking problems, especially when attempting to fuse bearing only data. Although the frequency of bird passage is low enough that in general we can expect only one bird to be in the field of view, we will address the problem of data association. In addition to allowing operation when multiple birds are in the field of view, it will also increased robustness to clutter.

The data association is preformed at both camera-to-camera and frame-to-frame levels. Data is associated between stereo pairs at each station by checking epipolar constraints between images seen in both cameras.[18] If the epipolar constraint, given by Equation 12, is satisfied, the camera coordinates $\mathbf{x}_{c_R}$ and $\mathbf{x}_{c_L}$ correspond to the same bird .

$$ \left(\mathbf{x}^{T}_{c_R}\mathbf{T}\right) \cdot (\mathbf{d}_c \times \mathbf{x}_{c_L}) = 0 \tag{12} $$

$\mathbf{T}$ is the rotation matrix of one camera with respect to the other.

In this particular case, the rotation matrix of each camera is computed with respect to the global coordinate system. The epipolar constraint used is therefore given by Equation 13.

$$ \left(\mathbf{T}_L\mathbf{d}_c + \mathbf{T}_L\mathbf{T}_R\mathbf{x}_{c_R} - \mathbf{d}_c\right)^{T} \cdot (\mathbf{d}_c \times \mathbf{x}_{c_L}) = 0 \tag{13} $$

$\mathbf{T}_L$ and $\mathbf{T}_R$ are the rotation matrices for the left and right cameras respectively and $\mathbf{d}_c$ is the 3D displacement between the stereo pair.

For frame-to-frame data association, a standard gated nearest neighbor approach is used. Association of corresponding camera coordinates from one frame to the next is based on the Mahalanobis distance between the best estimate computed in the previous time step,$\mathbf{x}_i$, and the prediction of the state,$\bar{\mathbf{x}}_j$, computed from the motion update of $\mathbf{x}_i$:

$$ d_{ij} = \left(\mathbf{x}_i - \bar{\mathbf{x}}_j\right)^{T} \mathbf{P}^{-1}_{jj} \left(\mathbf{x}_i - \bar{\mathbf{x}}_j\right) \tag{14} $$

The matrix $\mathbf{P}_{jj}$ is an arbitrary covariance associated with the predicted state $\bar{\mathbf{x}}_j$.

It is also possible that significant information to assist in data association will be available from the image itself (e.g. from an intensity histogram of the pixels identified as belonging to a bird). However, this feature has not been incorporated yet.

C. Track Initiation

Track initiation can be a thorny problem, especially when range data is uncertain (which is the case for short-baseline multiple camera systems). A closely related problem, feature initialization in Simultaneous Localization and Mapping (SLAM), has the same difficulties. Feature initialization in bearings-only SLAM is especially difficult, and has been the subject of a large amount of research. "Delayed approaches" collect several bearings over time and fuse them to compute an initial estimate of landmark position.[19–21] "Undelayed approaches" represent the conical probability distribution associated with a bearing measurement as a series of Gaussians which are then pruned as more measurements become available.[22–24] These are essentially multiple hypothesis filters. The main issue which these methods attempt to address is ensuring that the initial landmark position is "Gaussian" enough be incorporated into a Kalman filter (e.g. EKF or UKF) without causing stability problems.

Here we have used an undelayed approach based on the Particle Filter.[25] Since an undelayed approach is adopted, only one set of measurements is used and there is no time update. Initializing the state of the bird based on the triangulation of available bearings is a nonlinear problem and will not result in a sufficiently Gaussian estimate. A copious set of $3D$ particles that cover all possible locations of the bird, can be processed by the Particle Filter to compute a Gaussian estimate of the initial position of the bird. The initial set of particles are simulated along the entire range, R_{min} to R_{max}, the first available azimuth, ξ , and elevation, θ to a bird entering the field of view. The 3D position, $\mathbf{x}_p$,of any one particle in this normal distribution is given as:

$$\mathbf{x}_p = \begin{bmatrix} r_p\cos(\theta_p)\cos(\xi_p) \\ r_p\cos(\theta_p)\sin(\xi_p) \\ r_p\sin(\theta_n) \end{bmatrix} \tag{15}$$

where

$$r_p = U(R_{min}, R_{max}) \tag{16}$$

$$\theta_p = N(\theta, v_m^2) \tag{17}$$

and

$$\xi_p = N(\xi, v_m^2) \tag{18}$$

v_m being the measurement noise of the camera from which the bearing is obtained.

The set of particles thus generated have a Gaussian distribution across the conic section of the solid angle formed by the deviations about the true bearing to the bird, on account of the measurement noise v_m. This is shown schematically in Figure 4.

Each particle has a unique 3D position. The measurement vector, $\mathbf{z}_p$, is computed for every particle in the distribution. It is formed by concatenating bearings from all but one of M cameras in which the bird is viewed. The bearings used to form the initial probability distribution are not included to avoid the bias from counting it twice. $\mathbf{z}_p$ is given as follows:

$$\mathbf{z}_p = \begin{bmatrix} \boldsymbol{\gamma}_1 & \boldsymbol{\gamma}_2 & \dots & \boldsymbol{\gamma}_{M-1} \end{bmatrix}^T \tag{19}$$

A weighting of the distribution can now be carried out based on how closely the measurement vector, $\mathbf{z}_p$, to each particle matches the actual measurement vector, $\mathbf{z}$, for the $M-1$ cameras. The weight of each particle, w_p, is inversely proportional to its exponential distance from the mean in the multivariate normal distribution of mean $\mathbf{z}$ and covariance $\boldsymbol{\Sigma}_m$. w_p is computed using the following equation:

$$w_p = \exp[-0.5\boldsymbol{\Delta}_z^T\boldsymbol{\Sigma}_m^{-1}\boldsymbol{\Delta}_z] \tag{20}$$

where

$$\boldsymbol{\Delta}_z = \mathbf{z} - \mathbf{z}_p \tag{21}$$

and

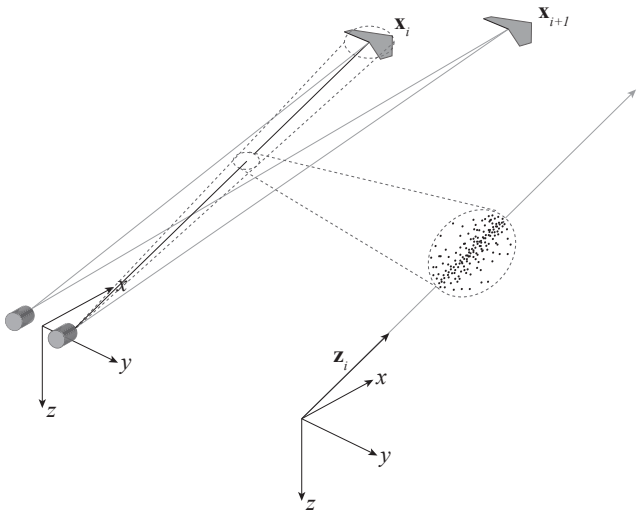


Figure 4. Track initiation. Bearings to a bird are treated as rays originating from camera frame origin to the bird. The uncertainty in the bearing measurement is treated as a zero-mean Gaussian, creating a probability cone. Each particle in the distribution has a weight associated with it, which is inversely proportional to the deviation of the bearing to that particle from the measured bearing to the bird. The bird's initial state is computed as the weighted mean of particles.

$$\mathbf{\Sigma}_m = v_m^2 \mathbf{I} \tag{22}$$

$\mathbf{I}$ being an $M-1 \times M-1$ identity matrix.

Having computed weights, w_p, to each particle, the initial estimated position, $\hat{\mathbf{x}}_{p_0}$ and the covariance, $\mathbf{P}_{p_0}$ associated with this estimate may be computed. For this application, the weighted mean and weighted covariance of the particle distribution are used for $\hat{\mathbf{x}}_{p_0}$ and $\mathbf{P}_{p_0}$ respectively.

In order to form a complete estimate of the initial state, $\hat{\mathbf{x}}_0$, and its covariance, $\mathbf{P}_0$, NED components of the bird's velocity also need to be estimated. Having adopted an undelayed approach for initialization, it is not possible to observe the dynamics of the bird. However, since the range of the velocity of the bird is a lot more limited than that of position, arbitrary flight speeds are picked for the velocity components of the state vector based on documentation provided by Kerlinger,[5] part of which is shown in Table 1. Note that best glide is defined by the maximum lift to drag ratio achieved by the bird.[5] A reasonable covariance is also selected based on the range of values observed and the estimated velocity vector and covariance matrix are concatenated with $\hat{\mathbf{x}}_{p_0}$ and $\mathbf{P}_{p_0}$ to obtain $\hat{\mathbf{x}}_0$ and $\mathbf{P}_0$. The speed in the West-East direction is initialized higher than the other components of velocity since it is assumed that the birds are flying East. These initial values are bounded and gaussian estimates of the initial position and can be used in gaussian filters like the UKF without risking divergence.

Note that the track initialization as described above is only carried out when a bird is first viewed by any one of the cameras in the distributed system. If a bird is initialized on being viewed by one set of stations, consequent stations can initialize their track, once the same bird comes into view, with the fused global estimate transmitted to them.

V. Simulation Results

THIS SECTION DISCUSSES the results obtained to asses the tracking of the true bird position using the methods described above. A Monte-Carlo simulation consisting of a number of varying flight paths is run to test various aspects of the estimation process. The simulation is initiated with the bird at any point along the edge of the Field of View (FOV) of the first camera in the array of stations. The distance of the bird from the camera array is set up to be between 10 to 100 times the baseline of the stereo pairs, d, in the

Table 1. Summary of Aerodynamic Performance of Raptors

Species	Air Speed at Best Glide (mps)	Sink Rate at Best Glide (mps)	Cruising Speed (mps)
Sharp-shinned Hawk	10.5	1.2	22.5
Broad-winged Hawk	11.6	1.1	24.2
Lanner Falcon	10-14	1.0	-
Red-Tailed Hawk	14.5	1.6	23.9
Osprey	11	0.9	24.9
Black Vulture	13.9	1.2	16.8
White-backed Vulture	13.5	1.1	16-20+
Andean Condor	NA	NA	15.0

system. The simulation is carried out with $M = 2$ camera stations, each consisting of $M_n = 2$ cameras. The simulation parameters used for the results presented are listed in Table 2.

Table 2. Simulation Parameters

Number of Runs	100
Number of Stations	2
Distance between Stations	800 meters
Cameras per station	2
Distance between Cameras	14.14 meters
Distance of Birds from Stations	150 to 1500 meters

The initialization of the bird involves making use of the first set of information to formulate a viable initial estimate of the bird. It is a heuristic process that depends on the strategic formulation of an initial set of particles and using a weighting procedure to compute initial estimate and the covariance associated with it. Since the uncertainty along the bearing is higher than that across, the uncertainty ellipse obtained from the initialization step is skinny and has a high condition number. Intuitively, an ill conditioned covariance can cause the estimator to diverge exponentially. Figure 5 shows a scatter plot of the true error at the final time step of the run that the estimator converges to vs. the condition number of the covariance associated with the initial estimate at the first time step. Note that condition number is either on the order of 10^5 or 10^2 which can be attributed to the distance of the bird from the station at the time of initialization. Contrary to intuition, the eventual error in the estimate at the final time step does not exceed 5 meters for the highest of condition numbers at the time of initialization. This indicates that even with ill conditioned initial estimates the UKF still converges to a reasonably accurate estimate.

The path of a bird flying through the field of view for five different runs is shown in Figure 6. It can be seen that the estimated position of the bird closely matches the true position and the covariance ellipse reduces with time. Initially, the ellipse is strongly elongated along the bearing from the camera to the bird: this is caused by the relatively small baseline between the cameras which causes dilution of precision along the bearing. A closer view of the bird at a later time is shown in Figure 6. The uncertainty ellipse seen here is well conditioned and less than 2 meters wide.

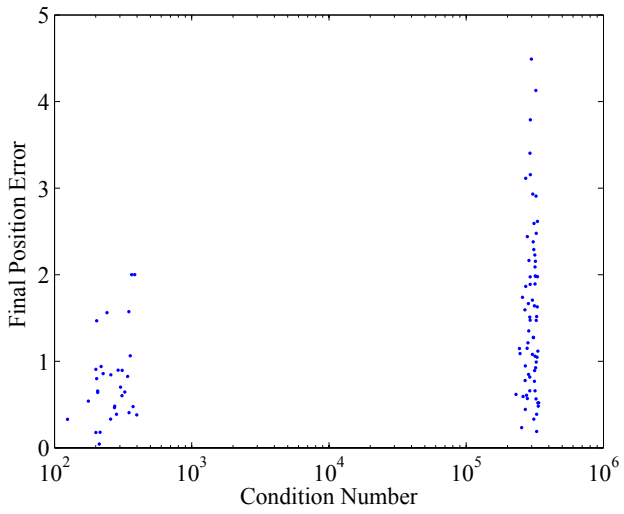


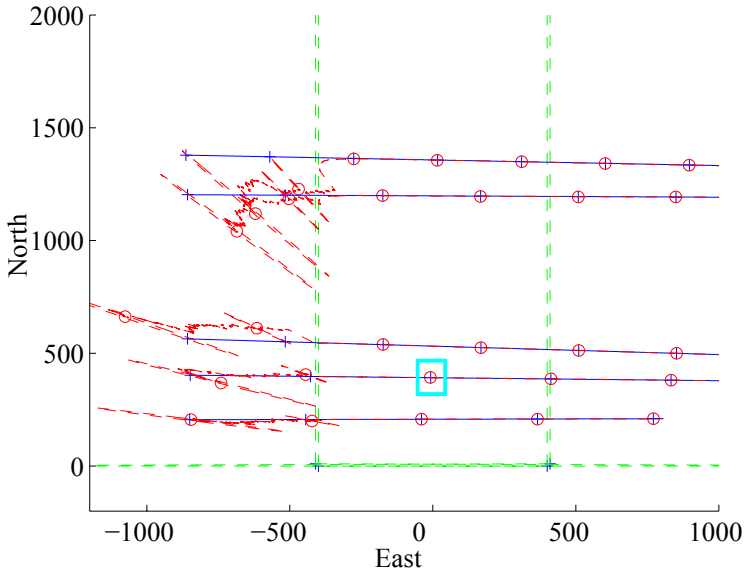
Figure 5. Monte Carlo simulation results of initialization parameters.

To assess estimate consistency we compare the mean of the estimated error variance (i.e. $\sqrt{\text{Tr}(\mathbf{P})}$) with the mean of the 2-norm of the estimate error (i.e. $\sqrt{(\mathbf{x}-\hat{\mathbf{x}})^T(\mathbf{x}-\hat{\mathbf{x}})}$). Monte-Carlo results for a 100 run simulation are presented in the remaining section. Results in Figure 7 show the performance of a centralized system with one estimate and covariance computed from measurements from all stations. Results in Figure 7 show the performance of the proposed distributed system that fuses independent estimates from all stations to compute a global estimate and covariance of the bird's state at each time step.

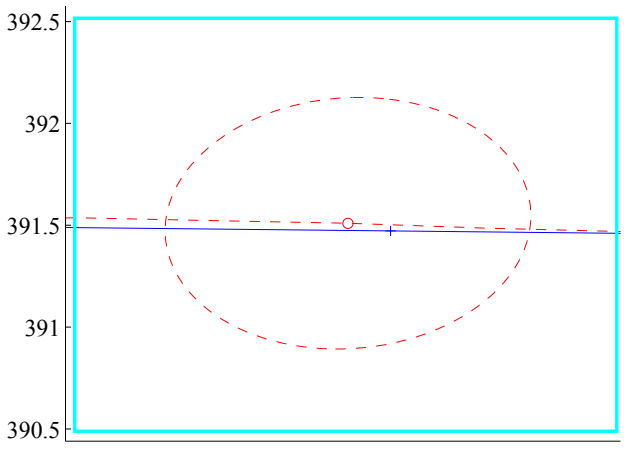
The vertical axis shows true and estimated error on a logarithmic scale. The horizontal axis shows time normalized with respect to the total time a bird was within the field of view of both cameras. The true and predicted errors match each other consistently through the entire time that the birds are in view. The estimate error dips sharply when the birds come into view at both stations and the number of available measurements or estimates double. Note that the results obtained by using a distributed system are identical to those obtained by the centralized system. Also, recall that the bird's velocity is assumed to vary by a random walk, and this uncertainty propagates into the position estimate. The average uncertainty in bird position once seen at two stations is 1 meter: note that the uncertainty in an individual bird's position will depend strongly on its distance from the cameras and the number of cameras where the bird is in view.

Data association is computed to determine the relation between the measurements obtained from the camera and the state estimates being computed by the tracking system. Data association is carried out between stereo camera pairs and from frame to frame. In a simulated system, which already has an inherent order to the measurements, bearing measurements are shuffled to test the performance of the data association procedure.

To assess estimate consistency we compare the mean of the estimated error variance, $\sqrt{\text{Tr}(\mathbf{P})}$, with the mean of the 2-norm of the estimate error, $\sqrt{(\mathbf{x}-\hat{\mathbf{x}})^T(\mathbf{x}-\hat{\mathbf{x}})}$. Results in Figure 8 compare the performance of a system with explicitly computed data association to that of a system where the order of the measurements remains unchanged, and data association is inherent. It can be seen that the result in both cases is identical. Hence the data association procedure fulfills its task of relating measurement of bearings to the bird with the state estimate associated with it.

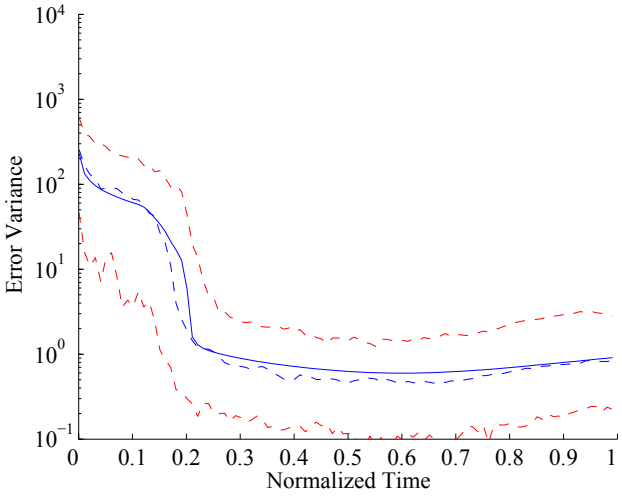


(a) True and estimated paths of raptors

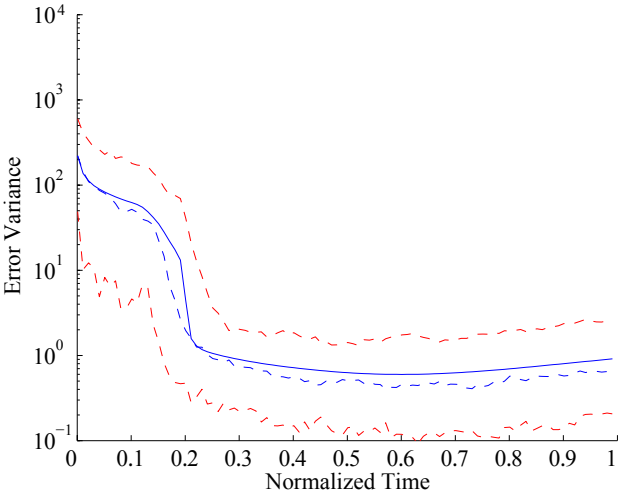


(b) Inset

Figure 6. Tracking performance, projected onto the plane z=0. Green dashed lines show field of view of the cameras, blue + show true bird positions, red dotted ellipsoids show 3-sigma position covariance and are centered on the estimated bird position.The lower plot shows the true and estimated positions of the bird and the associated uncertainty, at a later time in the flight path of the bird, after having obtained bearing measurements from both stations.

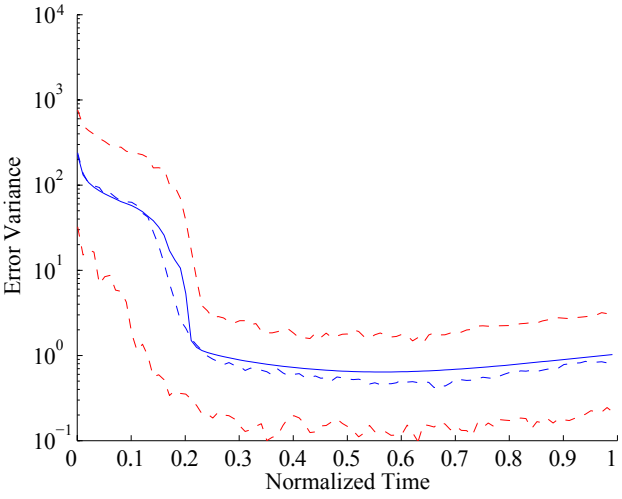


(a) Distributed System

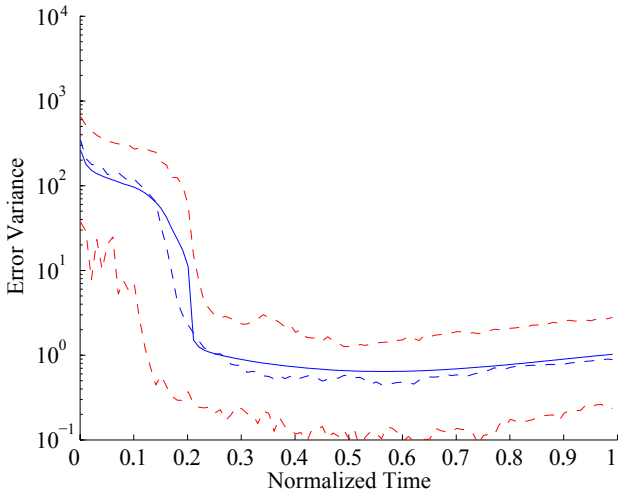


(b) Centralized System

Figure 7. Monte Carlo simulation results for a distributed and centralized systems. The dashed red lines show the maximum and minimum values of the 2-norm of the true estimate error, the dashed blue line shows the mean value of the 2-norm of the true estimate error and the solid blue line is the mean value of the estimated error variance.



(a) Inherent Association



(b) Computed Association

Figure 8. Monte Carlo simulation results for systems with inherent and computed data association. The dashed red lines show the maximum and minimum values of the 2-norm of the true estimate error, the dashed blue line shows the mean value of the 2-norm of the true estimate error and the solid blue line is the mean value of the estimated error variance.

VI. Conclusion

A distributed system with stereo cameras for sensors was used to solve the inherently non-linear and noise ridden problem of tracking multiple objects at long distances using bearing measurements. The simulation of the system, which accounts for noise in the process as well as the sensors, and has realistic estimates of bird positions, flight speed and possible camera positions, shows that it is possible to track migratory raptors using this approach with a fair amount of accuracy. Further, the computation time of this tracking system is fast enough to allow it to track multiple birds in real-time. This is essential to the success of this procedure in studying the migratory paths of birds. The problems of data association as well as initialization have been addressed and favorable results have been obtained for the ad hoc approach adopted for the initialization problem and the association of the measurement to the state estimate it pertains to. Future work involves the field testing of the proposed system. The use of color data for the purpose of data association or even species recognition would greatly improve the performance of the system and make the application more versatile. To create a truly autonomous system to study the behavioral patters of migratory species, learning algorithms like Hidden Markov Models could also be implemented to allow this system to track, recognize as well as learn the migratory patterns of various avian species.

References

[1]Moore, F. R., editor, *Stopover ecology of Nearctic-Neotropical landbird migrants: habitat relations and conservation implications.*, No. 20 in Studies in Avian Biology, Cooper Ornithological Society, 496 Calle San Pablo, Camarillo, California, 2000.

[2]Goodrich, L., "Unpublished Data," .

[3]"HawkWatch International," *http://www.hawkwatch.org/home/index.php*.

[4]"Southern Cross Peregrine Project," *http://www.frg.org/SC_PEFA.htm*.

[5]Kerlinger, P., *Flight Strategies of Migrating Hawks*, University of Chicago Press, 1989.

[6]Orloff, S. and Flannery, A., "Wind Turbine Effects on Avian Activity, Habitat Use, and Mortality in Altamont Pass and Solan County Wind Resource Areas (1989-1991), Final Report," Tech. rep., Planning Departments of Alameda, Contra Costa and Solano Counties and the California Energy Commision, Sacrameto, CA, 1992.

[7]Barrios, L. and Rodriguez, A., "Behavioural and Environmental Correlates of Soaring-Bird Mortality at On-Shore Wind Turbines," *Journal of Applied Ecology*, Vol. 41, No. 1, February 2004, pp. 72–81.

[8]Wallace Erickson, G. J., "Avian Collisions with Wind Turbines: A Summary of Existing Studies and Comparisons to Other Sources of Avian Collision Mortality in the United States," Tech. rep., Washington,D.C, 2001.

[9]Rafael Muñoz-Salinas, Eugenio Auirre, M. G.-S., "People Detection and Tracking using Stereo Vision and color," *Image and Vision Computing*, 2007.

[10]S. Bahadori, L. Iocchi, G. L. D. N. and Scozzafava, L., "Real-Time People Localization and Tracking Through Fixed Stereo Vision," *Lecture Notes in Computer Science*, Vol. 3533/2005, 2005.

[11]Zivkovic, Z. and Krose, B., "An EM-Like Algorithm for Color-Histogram-Based Object Tracking," *cvpr*, Vol. 01, 2004, pp. 798–803.

[12]Bar-Shalom, Y., Li, X. R., and Kirubarajan, T., *Estimation with Applications to Tracking and Navigiation*, Wiley Interscience, 2001.

[13]Karuppiah, D. R., Zhu, Z., Shenoy, P., and Riseman, E. M., "A Fault-Tolerant Distributed Vision System Architecture for Object Tracking in a Smart Room," *Lecture Notes in Computer Science*, Vol. 2095, 2001.

[14]Julier, S., Uhlmann, J., and Durrant-Whyte, H. F., "A New Method for the Nonlinear Transformation of Means and Covariances in Filters and Estimators," *IEEE Transactions on Automatic Control*, Vol. 45, No. 3, March 2000, pp. 477–482.

[15]van der Merwe, R. and Wan, E., "The Square-Root Unscented Kalman Filter for State and Parameter-Estimation," *IEEE International Conference on Acoustics, Speech and Signal Processing*, IEEE, Salt Lake City, UT, 2001.

[16]Wan, E. and Van Der Merwe, R., "The unscented Kalman filter for nonlinear estimation," *Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000. AS-SPCC. The IEEE 2000*, 2000, pp. 153–158.

[17]Hashemipour, H., Roy, S., and Laub, A., "Decentralized Structures for Parallel Kalman Filtering," *IEEE Transactions on Automatic Control*, Vol. 33, January 1988, pp. 88–94.

[18]Forsyth and Ponce.

[19]Bailey, T., "Constrained Initialisation for Bearing-Only SLAM," *IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, Taipei, Taiwan, 2003.

[20]Fitzgibbons, T. and Nebot, E., "Bearing Only SLAM using Colour-based Feature Tracking," *2002 Australasian Conference on Robotics and Automation*, Auckland, New Zealand, 2002.

[21]Montesano, L., Gaspar, J., Santos-Victor, J., and Montano, L., "Fusing Vision-Based Bearing Measurements and Motion to Localize Pairs of Robots," *International Conference on Robotics and Automation*, Barcelona, Spain, 2005.

[22]Davison, A. J., "Real-Time Simultaneous Localisation and Mapping with a Single Camera," *International Conference on Computer Vision*, Nice, France, October 2003.

[23]i Ortega, J. S., Lemaire, T., Decy, M., Lacroix, S., and Monin, A., "Delayed vs Undelayed Landmark Initialization for Bearing Only SLAM," *Workshop on Simulateous Localisation and Mapping, International Conference on Robotics and Automation*, Barcelona, Spain, 2005.

[24]Kwok, N. M. and Dissanayake, G., "An Efficient Multiple Hypothesis Filter for Bearing-Only SLAM," *Interational Conference on Intelligent Robots and Systems*, Sendai, Japan, 2004.
[25]Doucet, A., "On Sequential Monte Carlo Sampling Methods for Bayesian Filtering," 1998.