

# Detecting Migrating Birds at Night

Jia-Bin Huang<sup>1</sup>, Rich Caruana<sup>2</sup>, Andrew Farnsworth<sup>3</sup>, Steve Kelling<sup>3</sup>, and Narendra Ahuja<sup>1</sup>  
<sup>1</sup>University of Illinois Urbana-Champaign, <sup>2</sup>Microsoft Research, <sup>3</sup>Cornell Lab of Ornithology  
 jbh Huang1@illinois.edu, rcaruana@microsoft.com, {af27, stk2}@cornell.edu, n-ahuja@illinois.edu

## Abstract

*Bird migration is a critical indicator of environmental health, biodiversity, and climate change. Existing techniques for monitoring bird migration are either expensive (e.g., satellite tracking), labor-intensive (e.g., moon watching), indirect and thus less accurate (e.g., weather radar), or intrusive (e.g., attaching geolocators on captured birds). In this paper, we present a vision-based system for detecting migrating birds in flight at night. Our system takes stereo videos of the night sky as inputs, detects multiple flying birds and estimates their orientations, speeds, and altitudes. The main challenge lies in detecting flying birds of unknown trajectories under high noise level due to the low-light environment. We address this problem by incorporating stereo constraints for rejecting physically implausible configurations and gathering evidence from two (or more) views. Specifically, we develop a robust stereo-based 3D line fitting algorithm for geometric verification and a deformable part response accumulation strategy for trajectory verification. We demonstrate the effectiveness of the proposed approach through quantitative evaluation of real videos of birds migrating at night collected with near-infrared cameras.*

## 1. Introduction

Bird migration is the regular seasonal, large-scale, often long-distance movement between breeding and wintering grounds. Many species of bird migrate. Migration behavior is a critical indicator for evaluating environmental health [24]. By identifying important stopover and wintering locations, one can take action to save these key locations to protect endangered species. Scientists use a variety of methods to monitor bird migration, including satellite tracking, weather radar, moon-watching, or attaching geolocators on captured birds. However, these methods are either expensive (e.g., satellite tracking), inaccurate because they are indirect (e.g., weather surveillance radars), labor-intensive and error-prone (e.g., moon-watching), or intrusive (e.g., geolocators). Moreover, these techniques only crudely estimate the bulk density of migrating birds aloft.

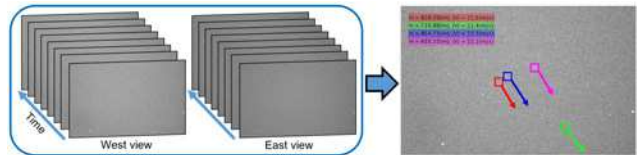


Figure 2. An example of automatic bird detection in stereo sequences. Our system takes stereo videos of the night sky as inputs, detects migrating birds in flight, and infers their orientation, speed, and altitude in very low SNR.

We propose to use a vision-based approach as a complementary sensing modality to build a bird migration monitoring system. By setting up stereo cameras facing up to the night sky, we can detect and track migrating birds in flight illuminated from below by light pollution in the recorded videos, as shown in Figure 2. Vision-based solutions offer several advantages over existing techniques. First, we can automatically and accurately count the number of individual birds aloft along with detailed trajectory estimation such as orientation, speed, and altitude. Such unprecedented accuracy in the reconstructed trajectories of individual birds may help re-evaluate migration, locomotion and navigation theories. Second, the estimated statistics could be used to calibrate other sensing modalities such as weather radar. Third, low-cost digital cameras allow us to build large-scale, distributed monitoring systems that cover broad areas.

There are three main challenges in developing a robust bird detection algorithm from videos. First, as migration usually occurs at night, the recorded videos inevitably contain substantial noise because of the low-light environment — the birds are generally invisible to the naked eye in the sky unless they pass in front of an illuminated object such as the moon. We illustrate this using sample frames from three video sequences in Figure 1. Second, depending on the species, migrating birds fly at altitudes ranging from several hundred feet to two miles. If the lens and camera provide an adequate field of view, the imaged bird may span only 1-2 pixels in a frame. This suggests that motion is the only reliable cue for detecting a bird. Third, efficient algorithms are required for large-scale deployment.

Several methods have been proposed to detect small ob-

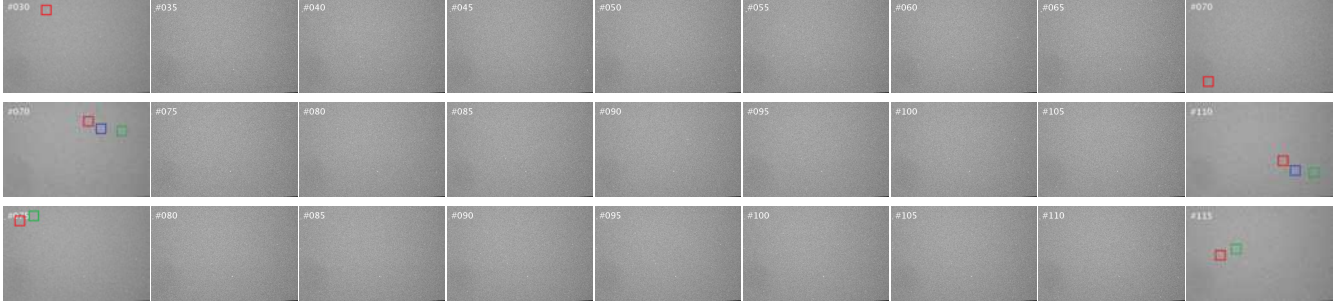
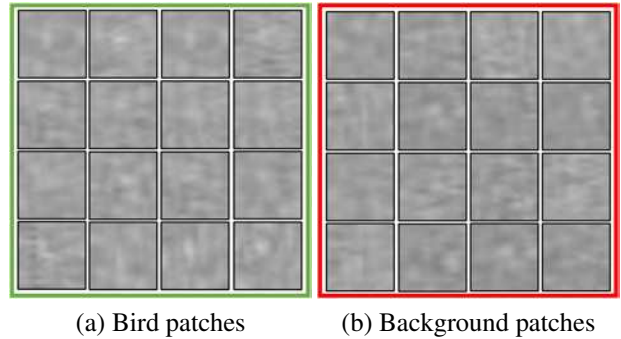


Figure 1. Detecting migrating birds from noisy image sequences. Each row shows a set of frames from a video sequence. From top to bottom, the sequences shown here have increasing levels of difficulty. Most of the bright spots in the images are stars. Color boxes indicate the birds in the first and the last frame of each sequence. Because of the low SNR and small size of high-flying birds (1-2 pixels), detection is very difficult, and often impossible, when looking at individual frames. It is only by detecting motion in the video stream that the human perceptual system can identify and track most birds. Similarly, the detection algorithm can only detect the more difficult high-flying birds by looking at the full video sequence and by simultaneously using stereo constraints from both cameras. Results are best viewed on a high-resolution display with adequate zoom level.

jects in image sequences under different problem contexts. In Automatic Target Recognition (ATR) [11, 36] the presence of the target is detected either using simple frame differencing, filter responses, or matching against a known template and then tracking over time. Similarly, in ridge detection in three-dimensional volumetric data (e.g., vessel extraction [28]), the ridge is often detected using a predefined set of oriented filters. The common drawback of these approaches is that the detection is mostly performed *locally*. These techniques are thus not directly applicable to our problem due to the extremely low SNRs in our case. Recent methods address this issue by designing filter banks to improve detection of faint signals [22, 23, 16, 17] or by searching a large set of curves [2, 32]. However, most of these algorithms, designed specifically for 2D images, are computationally infeasible for 3D image data. In multi-object tracking, several algorithms have been proposed to track objects in 3D using stereoscopy [35, 3].

In general, the problem of target tracking can be divided into four main categories: 1) Large objects in bright light (e.g., tracking cars, pedestrians, faces in daylight). 2) Small objects in bright light (e.g., meteor streaks in sky surveys, planes with lights at night or in daylight at a great distance, rockets/missiles that are bright in IR). 3) Large objects in dim light (e.g., people detection and tracking at night under surveillance illumination). 4) Small objects in low light (e.g., birds flying over 1 mile high at night illuminated by light pollution). Unfortunately, a direct application of existing techniques does not suffice for our problem (category 4). These techniques often pose tracking and trajectory reconstruction as independent problems of frame-level target localization and cross-frame and cross-view data-association. The target size and SNR in our case are so low that targets cannot be reliably detected in individual frames.

In this paper, we tackle this problem using a two-stage



(a) Bird patches (b) Background patches  
Figure 3. The difficulty of detection based on local image patches. (a) 16 cropped local image patch along a manually labeled bird trajectory. (b) 16 cropped random background patches. These patches are virtually indistinguishable by the naked eye.

robust model fitting approach. In contrast to prior work that aims at local detections in each frame, we aim at detecting using domain knowledge and global reasoning. Our fundamental assumption is that the migrating birds do not significantly change course and speed over short temporal spans (e.g., 5 seconds). We can thus cast the bird detection as finding curved 3D ridges in a spatiotemporal volume. The core detection algorithm consists of two main stages:

(1) *Geometric verification*: Given a large collection of noisy local detections, we extend the RANSAC-based 3D line fitting algorithm by explicitly incorporating stereo vision constraints. Specifically, we fit the model to both views *jointly*, which offers several advantages over a straightforward application of RANSAC independently in each view. First, the sample subset is used to determine the full bird model including altitude, speed, orientation, and position. Second, we can quickly reject a large number of physically implausible model hypotheses by checking the disparity, the temporal alignment, and extreme speed and altitude. Third, our model hypothesis allows us to exploit simultaneously

the detected foreground points from both view by compensating the disparity. We set a loose threshold for line fitting so that birds flying at time-varying speed or directions could also be detected. To the best of our knowledge, while RANSAC has been extensively applied to two-view robust correspondence problems (e.g., solving the fundamental matrix, homography), it is less explored in robust model fitting (e.g., fitting 3D lines in volumetric data) by incorporating multi-view inputs and constraints.

(2) *Trajectory verification*: In this step, we aim at verifying the presence of the bird using guidance from geometric verification. Given a small set of 3D line hypotheses, we integrate the signals along the direction of the coarse 3D trajectory while accounting for spatial uncertainties due to time-varying speed, direction, and altitude. This is technically realized using the generalized distance transform to efficiently search over all possible spatial deformations. The trajectory verification allows us to integrate all of the local responses along the predicted trajectory, resulting in a more discriminative signal for separating birds from noisy background night sky and ranking hypothesis. This step is critical for handling challenging low-SNR scenarios.

We make the following contributions in this paper:

1. We address a novel application domain using computer vision algorithms. The vision-based system provides a low-cost, accurate, and new sensing modality for monitoring and studying bird migration.
2. We propose a RANSAC-based 3D line fitting algorithm that explicitly incorporates stereo vision constraints. We demonstrate that such constraints are crucial for robust model fitting in very low SNRs.
3. We account for birds flying with time-varying speeds and directions using deformable part modeling. The trajectory verification step allows us to gather all the local responses along the predicted trajectory, resulting in a discriminative signal for separating birds from the noisy background night sky.

## 2. Related Work

**Bird migration monitoring techniques.** Scientists use methods such as weather radar [15, 33, 19] and acoustic sensors [31, 8, 18] to monitor migrating birds [9, 1, 13]. Radar networks can provide wide area coverage over 1000's of kilometers, but radar reflectivity data is difficult to interpret and requires careful calibration as the data contain many biological (birds, bats, and insects) and meteorological phenomena. Calibration often is based on a traditional method for counting migrating birds: the use of a telescope to count birds as they pass across the full moon. Although moon-watching [29, 25] can provide direct visual bird counts, it is labor-intensive, error-prone (e.g., when multiple birds fly across), and only covers a very small portion of the night sky (the moon is about 0.5 deg wide in the

sky). In contrast, our vision-based approach can accurately detect birds, infer their orientations, speeds, altitudes, and cover a large portion of the sky — a 5-10 degree FOV covers 250 to 1000 $\times$  larger area than the moon.

**Small target detection in image sequences.** Detecting and tracking small targets in infrared image sequences is a long-standing problem in computer vision with numerous military applications. These methods typically rely on detecting the small targets locally, e.g., using frame-differencing [12], max-mean/max-median filter [14], top-hat transformation [5], or directional filters [4]. Local detections are then linked over time using sequential hypothesis testing or motion models such as Kalman, particle filters, or global optimization approaches [3, 35]. As our videos contain a substantial amount of noise, local detections are not reliable (as shown in Figure 3). Unlike previous approaches that aim at getting correct *local* detections, we leverage top-down models with *global* reasoning for robust detection.

The work most related to our work is that of Balzerini et al. [7], which uses stereo vision to reconstruct 3D positions of individual birds to study the collective behavior of flocks of birds during the day. Our problem differs from theirs because many birds migrate *at night*. The challenge thus lies in how to detect birds in very low SNRs reliably. We can perform detection only by doing detection and tracking simultaneously so that detection is enabled by additional constraints coming from tracking, and vice versa.

**Ridge detection in three-dimensions.** We can view our problem as ridge detection in three-dimensional volumetric data (i.e., spatiotemporal volume). Ridge detection techniques often detect ridges using a pre-defined set of oriented filters at multiple scales. However, the local filters are not optimal for detecting faint signals in low SNR settings. Recent efforts include designing image representation for facilitating faint signal detection [22, 23, 17] or detecting faint curved edges in images [2, 32].

**Geometric model fitting.** Our work is related to classical parametric shape fitting techniques in computer vision such as RANSAC [21] and generalized Hough transform [6]. In our problem context, Hough transform would need to construct a 5-D parameter space, making the memory cost prohibitively high. Our method uses a RANSAC-based algorithm to perform line fitting in 3D point clouds (2D space + 1D time). The novelty lies in that we propose a sample selection approach for generating hypothetical inliers by leveraging the stereo vision constraints.

## 3. Overview

Figure 4 illustrates the three main steps for detecting migrating birds in flight. Given a pair of stereo videos, we

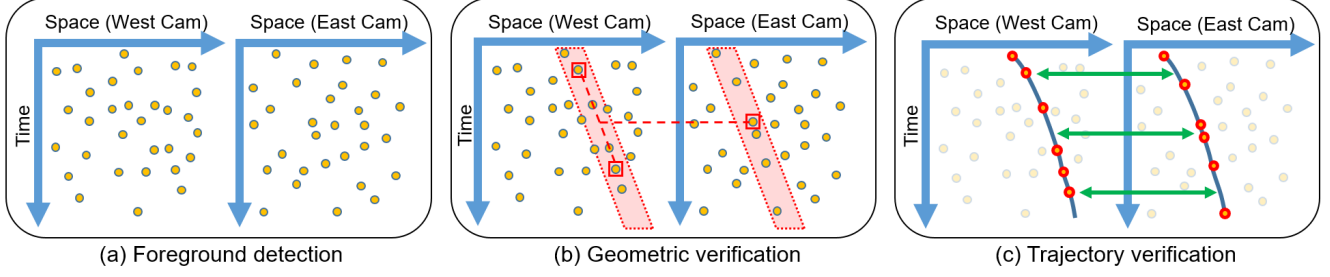


Figure 4. Overview of the bird detection algorithm. Our algorithm consists of three main modules: (a) Foreground detection: using statistical background modeling for moving object detection. (b) Geometric verification: RANSAC-based line fitting with stereo vision constraints. The three red boxes indicate the selected hypothetical inliers. This strategy naturally handles disparity estimation and offer computational efficiency by rejecting a large number of physically implausible configurations. (c) Trajectory verification: with the coarse 3D line fitting, we integrate weak signals along the predicted trajectory for both videos to verify if there is a bird. To account for birds flying at time-varying speed and directions, we interpret the motion compensated local image patch as a “part” of an object and use the generalized distance transform [20] for handling such spatial uncertainty. We detect the birds by thresholding the final response map.

first use classical statistical background modeling to detect foreground candidates (Section 4.3). As shown in Figure 4, the substantial number of outliers obscure the hidden curved line. Second, we use a RANSAC-based 3D line fitting algorithm to generate and verify hypotheses (Section 4.4). We propose a sampling strategy that explicitly incorporates stereo vision constraints. Such constraints are powerful because it allows us to reject a large portion of physically implausible configurations, and thereby offers computational efficiency when a large number of random samples are required due to the unusually high outliers ratio. We use a coarse threshold to maintain high recall in detection. Third, we use trajectory verification (Section 4.4) to integrate the faint signals along the predicted trajectory from geometric verification while accounting for spatial uncertainties. Unlike RANSAC-based detection methods that use *sparse* detection data (i.e., 3D point clouds), we exploit *dense* information across the spatiotemporal volume. Through gathering local evidence across a long temporal span, we get a clean and discriminative signal that allows us to separate birds from the noisy background with high precision.

## 4. Stereo-based Bird Detection

In this section, we describe the proposed method in detail. We first present the local bird trajectory model by assuming a weak perspective camera model. We then briefly describe pre-processing steps for rectifying videos of the night sky by registering stars, followed by the core detection algorithm: (1) foreground detection, (2) geometry verification, and (3) trajectory verification.

### 4.1. Bird trajectory modeling

To model the coarse bird trajectory in a video, we make the following two assumptions. First, we assume affine camera models because the migrating birds in flight are reasonably far away from the camera (with altitudes ranging

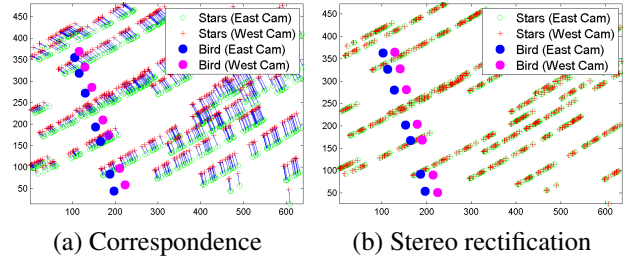


Figure 5. Stereo image rectification using star registration. (a) Star detection from a video frame, (b) Correspondence, (c) Stereo image rectification.

from several hundred feet to two miles) compared with the size of a bird. Second, we assume that birds fly at relatively constant speed, orientation, and altitude during a short time-frame (e.g., 5 seconds).

Denote the three-dimensional position in space of a bird at time  $t$  as  $\mathbf{P}_t = [X_t, Y_t, Z_t]^T$ , we can express the imaged position of the bird  $\mathbf{p}_t = [x_t, y_t]^T$  as  $\mathbf{p}_t = \mathbf{M}[\mathbf{P}_t^T, 1]^T$ , where  $\mathbf{M} \in \mathbb{R}^{2 \times 4}$  is the camera projection matrix. Using the constant speed, orientation, and altitude assumptions, we simplify the 3D position  $\mathbf{P}_t$  as  $\mathbf{P}_t = \mathbf{P}_0 + t[V_x, V_y, 0]^T$ , where  $\mathbf{P}_0$  indicates the position at time  $t = 0$ , and  $V_x, V_y$  are the physical speeds in space. We can write down the imaged position  $\mathbf{p}_t = \mathbf{p}_0 + t[v_x, v_y]^T$ , where  $v_x, v_y$  are the speed in the image space. We can thus view this *idealized* bird trajectory as a thin, straight ridge in the spatio-temporal video cube.

### 4.2. Stereo image rectification

Our system uses stereo vision to determine the altitude of a flying bird from correspondence. To simplify the 2D correspondence search to 1D, we first rectify the images from two views so that all epipolar lines are parallel to the horizontal axis. We follow the standard procedure for stereo image rectification: (1) finding a set of correspondences in the stereo pair of videos, (2) estimate the fundamental matrix



[37], and (3) compute the rectifying matrices using [30].

For night sky images, we cannot apply the commonly used local interest point and feature descriptor matching approaches to establish correspondences. Fortunately, mother nature provides stars as markers. The two cameras are setup to capture roughly the same patch of the sky, so we exploit the imaged star positions for image registration. For each video, we first apply a moving average filter over the temporal axis to suppress the background noise. We then apply a single-scale 2D Laplacian of the Gaussian (LoG) to locate bright blob structures. After thresholding the LoG filter response and non-maximum suppression, we obtain a set of star positions (i.e., 2D point cloud) for each video.

With the detected star positions, we use the Iterative Closest Point (ICP) algorithm [10] with an affine motion model to find the transformation and inlier matches. However, as the stars are infinitely far away from the camera, the correspondences from stars gives rise to a degenerated case in fundamental matrix estimation. To eliminate this degeneracy, we manually label the position of a flying bird in several frames. We only need to do this manual labeling *once* because we assume the cameras remain fixed through the videos. It is possible to use the proposed automatic flying bird detection to perform self-calibration (e.g., for cases where the stereo camera setup cannot remain fixed over time), but we leave that for future work.

We show in Figure 5(a) the detected stars in two views (Red and Green) the correspondence from ICP in Blue line. Figure 5(b) shows the rectified positions for the stars and the manually labeled bird. The stars from two views align accurately (as they are infinitely far) and the labeled birds fall on horizontal lines. Note that the results shown here contain star positions over 20 mins. The purpose of using this “star trail” is to provide additional accuracy for registration.

### 4.3. Foreground detection

In this step, we look for *local* evidences for detecting flying birds. As imaged flying birds appear brighter than the surrounding background (illuminated from below by light pollution), the imaged bird trajectory can be seen as an intensity ridge in the video sequence. The problem of bird detection could be naturally cast as a *ridge detection* task in a 3D spatiotemporal video cube. Ridge (and valley) detection have been extensively studied in computer vision and image analysis with typical applications for detecting road in aerial images and for detecting blood vessels in 2D retinal or 3D magnetic resonance images. These methods often rely on designing filters that respond to locally linear intensity features followed by linking processes. However, these methods cannot directly be applied to our problem. As our videos have very low SNR, achieving accurate local detection would require evaluating a large collection of oriented filters with large kernel sizes, and thus would not scale well

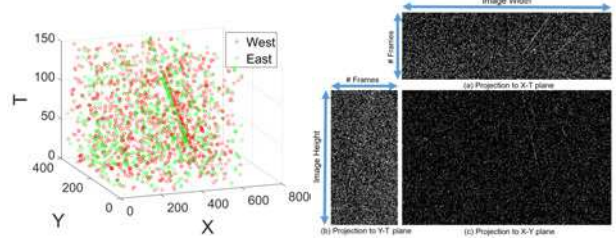


Figure 6. Foreground detection. (a) Sample foreground detection plots. Flying birds in a video appear like curved lines in the spatio-temporal volume. In this scattered plot, there are three curved lines. (b) Projection of foreground detection onto X-Y, X-T, and Y-T planes.

with large-scale video datasets.

For efficiency, we rely on top-down knowledge and global reasoning for detecting dim flying birds and resort to a simple statistical background modeling approach for local foreground pixel detection. Specifically, we build a per-pixel Gaussian model and compute the response of a pixel by measuring the intensity deviation from the learned model. We detect foreground pixels by thresholding the local responses. We estimate the parameters of the per-pixel Gaussian model (mean and variance) online using a pre-defined learning rate. Note that while other more sophisticated background modeling and subtraction techniques are available, we did not observe substantial improvement. Figure 6 shows the three-dimensional (X-Y-T) scattered plot of the foreground detection on West and East camera on a video with a flock of three birds. Figure 6(b) shows the projections of the 3D point cloud onto X-Y, Y-T, and X-T planes, respectively. We could visually spot the three flying birds. The challenge, however, lies in how to handle the high outliers ratio.

### 4.4. Geometric verification

Our coarse bird model (i.e., a straight line in a 3D video cube) consists of 5 parameters, including an initial spatial position in the image plane (2D), constant motion vectors (2D), and disparity from stereo vision (1D). The goal of geometric verification is to fit coarse bird models to the 3D point clouds with a significant portion of outliers from the foreground detection step. The most widely used robust fitting algorithms are (1) Generalized Hough Transform (GHT) and (2) RANSAC. We choose to perform geometric verification using RANSAC because of the demanding memory complexity in GHT for estimating 5D models.

A straightforward approach would be using RANSAC-based 3D line fitting method independently for each video and then solve the disparity by matching fitted lines in two views *after* the models in each video are found. However, such an approach does not exploit the available physical constraints presented in the stereo videos. For example, the

two corresponding 3D lines in the stereo pair should be parallel, having the same y-coordinate at all frames, and with positive disparity values. To incorporate these constraints, we propose a stereo-based 3D line fitting algorithm. Specifically, of the detected foreground points from the stereo pair, we select random subsets of three detected points to estimate the bird model, where two points are drawn from one video, and one point is drawn from the other video.

Figure 4(b) illustrates the three-point hypothetical inlier. The proposed three-point subset sampling strategy offers several advantages. First, we can fully determine the 5D bird model using the selected three points. Second, we can quickly reject a large collection of model hypotheses that are not physically plausible by checking the disparity and temporal alignment. Third, as we also have disparity in the estimated model, we can simultaneously exploit the detected foreground points from both videos by compensating for the disparity.

We follow the standard RANSAC algorithm and count the number of inliers (number of foreground points fall inside the 3D tube). We then apply the J-linkage clustering algorithm [34] to group repeatedly sampled hypothesis. Once we have the grouped model hypothesis, we perform least squares fitting using all the inlier foreground points from both videos to compute a more reliable bird model estimation. We solve this refinement step iteratively. Given an estimated disparity, we can solve the orientation using Singular Value Decomposition. In turn, we fix the orientation and update the disparity using least-square fitting.

#### 4.5. Trajectory verification

While geometric verification can efficiently detect flying birds by exploiting the stereo vision constraints, we observe a high false positive rate due to inevitable noisy foreground detections. We address this issue by integrating signals along the bird’s trajectory. Unlike geometric verification that fit models to *sparse* foreground candidates, trajectory verification exploits *dense* information across the entire video cube.

One way to achieve this is to use the corresponding matched filter that computes the average local response along its trajectory. However, the actual bird trajectory may not be a perfect 3D straight line in the video because the bird may not fly along the same direction or maintain constant speed and altitude. Simply using the estimated coarse bird model to filter the videos is clearly sub-optimal as spatial misalignments lead to blurry accumulated response.

We address this problem by allowing the bird trajectory to be “deformed” as illustrated in Figure 7. We interpret the bird response at the predict position using the coarse model at a frame as the local response for a “part”. The detection of the bird can then be cast as the detection of a deformable part model. Specifically, we evaluate the score of a small

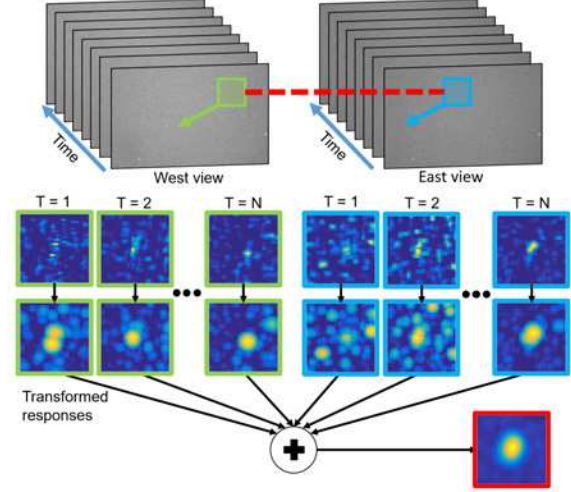


Figure 7. Trajectory verification. Given a 3D line model, we gather the spatial patches along the coarse trajectory from  $T = 1$  (when the bird enters the frame) to  $T = N$  (when the bird leaves the frame). These local responses are noisy and misalignment due to time-varying speed and directions. We transform the responses to account for spatial uncertainty.

window (e.g.,  $15 \times 15$ ) as

$$\text{score}(x, y) = \sum_{t=1}^{N_t} \max_{dx, dy} [R_t(x_t + dx, y_t + dy) - \alpha(dx^2 + dy^2)],$$

where  $R_t$  is the response map for foreground object,  $(x_t, y_t)$  is the predicted position at time  $t$  using the hypothesized 3D line from the geometric verification step, and  $\alpha$  is the weights for allowing different levels of spatial deformation. As we also have the disparity estimation, we aggregate the scores from two views. We use the Generalized Distance Transform [20] to efficiently search over all possible deformations through time. These transformed responses can then be added together and ranked for verification.

## 5. Experimental Results

### 5.1. Implementation details

In foreground detection, we classify a pixel as a foreground if its intensity is greater than the mean background intensity by 2.75 standard deviations. In geometric verification, we keep model hypotheses with at least 5 inliers and reject the rest. In trajectory verification, we use  $15 \times 15$  windows and set the spatial deformation parameter  $\alpha = 0.5$ . We fix these parameters throughout all experiments.

We process a video in a mini-batch mode, by dividing a long video sequence into a set of overlapping five-second sequences with a one-second interval. We detect birds in each video clip and cluster these detections in the nearby clips to generate our final results. In a video with frame rate 30 fps, we have in total 150 frames. For processing one

Table 1. Quantitative performance

Method	Precision	Recall
Geometric verification only	6.08%	83.10%
Geometric and Trajectory verification	97.30%	83.72%

5 second video clip, our MATLAB implementation takes 7 seconds on a PC with 1.9GHz and 8 GB memory. The data and source code are available on the project website <sup>1</sup>.

## 5.2. Evaluation on real videos

To evaluate the proposed method, we have developed a prototype stereo video system to capture videos of migrating birds at night. In what follows, we present the data collection steps and our results on real videos.

**Data collection** We use two low-light near-IR mono industrial VGA cameras to capture the stereo video. We chose the cameras because of their superior low-light sensitivity (10k–100k  $\times$  more sensitive than consumer video cameras). The cameras have a spatial resolution of  $640 \times 480$  pixels. We use a pair of 50 mm lenses and set the two cameras on tripods facing the sky with a two-meter baseline. We captured hours of stereo video on different nights and selected a 40-minutes long video from Spring migration for testing.

**Quantitative results** To evaluate performance, we developed a Graphical User Interface to allow experts to annotate the birds flying across video frames. In total, 86 birds were found in the video. A majority of the birds head North  $+ - 20$  degrees. Among the 86 annotated birds, our method detects 74 of them, with two false positives and 12 missed detections. In Table 1, we show the quantitative performance of our algorithm. When using geometric verification only, we achieve 83.10% in recall. However, precision is very poor, with only 6.08%. Coupled with trajectory verification, precision rises above 95% with 83% recall. The automatic system detects 9 birds missed by the experts.

We further evaluate the relative contributions from (1) fusing information from two views and (2) using the deformable part model to account for the inevitable spatial uncertainty when using real videos of birds migrating at night. Specifically, we report the precision and recall values using the four variants. *One View*: use only the video from the West camera. *Two Views*: use both West and East videos. *Without deformation*: did not transform the scores in each local image patch. *With deformation*: use the generalized distance transform to allow spatial uncertainty.

To make the contribution of each term clear, Figure 8 shows the precision and recall of these four variants using a version of the system that does no post-processing to reduce



Figure 8. Precision and recall of four variants of the proposed trajectory verification approach on real videos.

false detections. In cases of integrating signals along the estimated trajectory (from geometric verification) in one view, both the precision and recall improve when we account for the spatial deformation. When using two views *without* accounting for the spatial deformation, we found that the recall drops significantly. We attribute the performance degradation to the imperfect disparity estimation between the two views. Integrating scores from two views without taking the spatial uncertainty into account, the results suggest that the algorithm may not be able to accumulate the weak signals due to the misalignment, and, therefore, fails to detect dim birds. Overall, the best performance is achieved by taking advantage of the stereo constraint while also allowing for deformation to account for spatial uncertainty.

**Qualitative results** In Figure 9, we show detection results in a variety of scenarios to demonstrate the effectiveness of the proposed approach. For example, our method can detect birds flying at altitudes ranging from 200 meters to more than 2,500 meters as well as at different directions and speeds. We can also handle multiple birds flying across the video frame. Unlike existing techniques that can only detect the presence of the birds, the direct visual analysis provides detailed measurements about the trajectory of *individual* birds. We believe such information may provide valuable insights about the behavior of migrating flocks.

## 5.3. Discussion

**Limitations** One potential problem and limitation in evaluating the performance on real videos is that the groundtruth annotations are not available, and the human visual system may not be able to detect very dim, high-flying birds from the video. In the future, we plan to investigate a multi-modal solution (e.g., vision-based, acoustics-based, and weather radar) toward this problem. Figure 10 shows a few of the limitations of our method. First, as our foreground detection is based on a statistical background mod-

<sup>1</sup><https://sites.google.com/site/jbhuang0604>



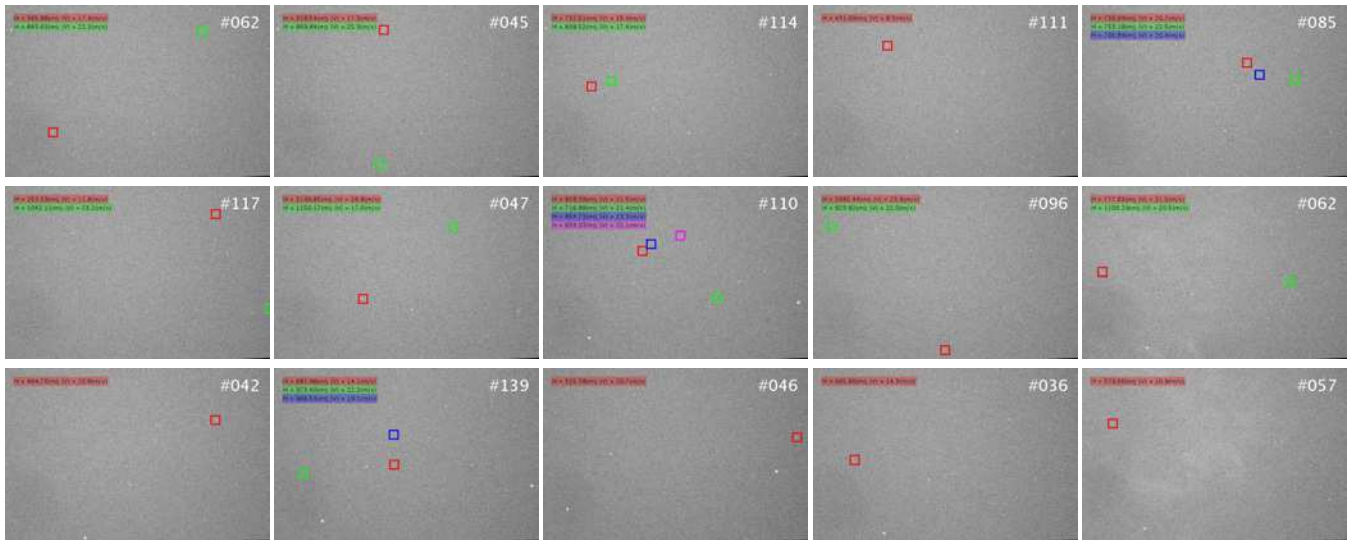
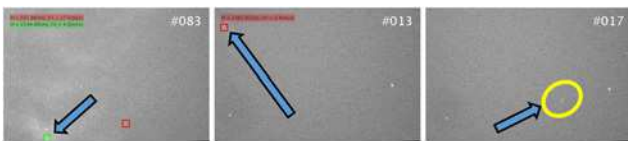


Figure 9. Detection results on real videos. Our system can handle diverse scenarios, e.g., single, multiple birds, birds flying parallel with each other, or birds flying at very different altitudes.



(a) FP (cloud) (b) FP (noises) (c) TN (insect)

Figure 10. Interesting cases: (a): a false positive detection due to a moving cloud. (b) a false positive detection due to noise. (c) a true negative — the moving blob is an insect. Our system uses the estimated altitude to avoid confusion with high-flying objects (e.g., above 3000 meters) such as satellites or planes and low-flying objects (e.g., under 50 meters) such as insects.

eling approach, we are not able to handle dynamic background or sudden illumination changes. For example, in Figure 10(a), our method falsely detect the moving cloud as a bird. Second, even with the use of stereo-based constraints for rejecting physically implausible detections (e.g., Figure 10(b)), our method may sometimes produce false positives due to the substantial noise in the video. One potential solution is to use three or more cameras covering the same patch of the night sky. Our framework could be extended to multi-camera settings to further improve the detection performance. Third, in Figure 10(c) we show that our method is robust to other types of flying objects. The altitude estimation provides important cues for separating migrating birds from high-flying objects (satellites or airplanes) and low-flying objects (insects).

## 6. Conclusions

We presented the first stereo-vision-based approach for monitoring migrating birds at night. From a pair of stereo videos, we perform stereo image rectification by detecting

and registering stars. The core bird detection algorithm then consists of three main steps. First, we use a statistical background modeling for foreground detection for each video. This produces a noisy three-dimensional point cloud. Second, we propose a novel RANSAC-based 3D line fitting that explicitly takes into account stereo vision constraints. Third, we apply deformable part modeling for handling the spatial uncertainty of birds due to time-varying speed and orientation. Through evaluation on real videos captured from a physical setup, we demonstrate the effectiveness of the proposed method. We believe the new capabilities will make significant impact on computational ecology.

While our work address a particular application, the approach for detecting and tracking multiple small targets in 3D volumic data with very low SNR using multiple cameras is general and potentially can be applied to many other important problems. In this work, we show how to leverage the underlying physical constraints and domain knowledge to achieve physically plausible detection that otherwise would not be feasible due to the high level of noise.

## Acknowledgments

This work is supported in part by Office of Naval Research N00014-12-1-0259, and Air Force Office of Scientific Research AF FA8750-13-2-0008, National Science Foundation grant IIS-1125098, and the Leon Levy Foundation. We want to thank Peter Gural, David Samuels, and the CAMS (Cameras for Allsky Meteor Surveillance) Project for their generous help in the initial stages of this project (the 1st images of birds flying at night came from their meteor cameras), and Danil Radchenko and Konstantin Kobylinsky at Sensoray for modifying their frame-grabber drivers for this project.



## References

- [1] T. Alerstam, A. Hedenström, and S. Åkesson. Long-distance migration: evolution and determinants. *Oikos*, 103(2):247–260, 2003. [3](#)
- [2] S. Alpert, M. Galun, B. Nadler, and R. Basri. Detecting faint curved edges in noisy images. In *ECCV*. 2010. [2](#), [3](#)
- [3] A. Attanasi, A. Cavagna, L. Del Castello, I. Giardina, A. Jelic, S. Melillo, L. Parisi, F. Pellacini, E. Shen, E. Silvestri, et al. Greta-a novel global and recursive tracking algorithm in three dimensions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(12):2451–2463, 2015. [2](#), [3](#)
- [4] T.-W. Bae, F. Zhang, and I.-S. Kweon. Edge directional 2d lms filter for infrared small target detection. *Infrared Physics & Technology*, 55(1):137–145, 2012. [3](#)
- [5] X. Bai and F. Zhou. Analysis of new top-hat transformation and the application for infrared dim small target detection. *Pattern Recognition*, 43(6):2145–2156, 2010. [3](#)
- [6] D. H. Ballard. Generalizing the hough transform to detect arbitrary shapes. *Pattern recognition*, 13(2):111–122, 1981. [3](#)
- [7] M. Ballerini et al. Interaction ruling animal collective behavior depends on topological rather than metric distance: Evidence from a field study. *Proceedings of the National Academy of Sciences*, 105(4):1232–1237, 2008. [3](#)
- [8] R. Bardeli, D. Wolff, F. Kurth, M. Koch, K.-H. Tauchert, and K.-H. Frommolt. Detecting bird sounds in a complex acoustic environment and application to bioacoustic monitoring. *Pattern Recognition Letters*, 31(12):1524–1534, 2010. [3](#)
- [9] P. Berthold. *Bird migration: a general survey*, volume 12. Oxford University Press, 2001. [3](#)
- [10] P. BESL and N. MCKAY. A method for registration of 3-d shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):239–256, 1992. [5](#)
- [11] B. Bhanu. Automatic target recognition: State of the art survey. *IEEE Transactions on Aerospace and Electronic Systems*, (4):364–379, 1986. [2](#)
- [12] S. D. Blostein and T. S. Huang. Detecting small, moving objects in image sequences using sequential hypothesis testing. *IEEE Transactions on Signal Processing*, 39(7):1611–1629, 1991. [3](#)
- [13] E. S. Bridge, K. Thorup, M. S. Bowlin, P. B. Chilson, R. H. Diehl, R. W. Fléron, P. Hartl, K. Roland, J. F. Kelly, W. D. Robinson, et al. Technology on the move: recent and forthcoming innovations for tracking migratory birds. *BioScience*, 61(9):689–698, 2011. [3](#)
- [14] S. D. Deshpande, H. E. Meng, R. Venkateswarlu, and P. Chan. Max-mean and max-median filters for detection of small targets. In *SPIE's International Symposium on Optical Science, Engineering, and Instrumentation*, 1999. [3](#)
- [15] A. M. Dokter, F. Liechti, H. Stark, L. Delobbe, P. Tabary, and I. Holleman. Bird migration flight altitudes studied by a network of operational weather radars. *Journal of The Royal Society Interface*, 8(54):30–43, 2011. [3](#)
- [16] D. L. Donoho and X. Huo. Beamlets and multiscale image analysis. *Multiscale and Multiresolution Methods: Theory and Applications*, 20:149, 2002. [2](#)
- [17] D. L. Donoho and O. Levi. Fast x-ray and beamlet transforms for three-dimensional data. *Modern signal processing*, 46, 2002. [2](#), [3](#)
- [18] W. R. Evans and K. V. Rosenberg. Acoustic monitoring of night-migrating birds: a progress report. *Strategies for bird conservation: The Partners in Flight planning process*, pages 1–5, 2000. [3](#)
- [19] A. Farnsworth, B. M. Van Doren, W. M. Hochachka, D. Sheldon, K. Winner, J. Irvine, J. Geervarghese, and S. Kelling. A characterization of autumn nocturnal migration detected by weather surveillance radars in the northeastern us. *Ecological Applications*, 2015. [3](#)
- [20] P. Felzenszwalb and D. Huttenlocher. Distance transforms of sampled functions. Technical report, Cornell University, 2004. [4](#), [6](#)
- [21] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981. [3](#)
- [22] M. Galun, R. Basri, and A. Brandt. Multiscale edge detection and fiber enhancement using differences of oriented means. In *ICCV*, 2007. [2](#), [3](#)
- [23] K. Guo and D. Labate. Optimally sparse multidimensional representation using shearlets. *SIAM journal on mathematical analysis*, 39(1):298–318, 2007. [2](#), [3](#)
- [24] H. Higuchi. Bird migration and the conservation of the global environment. *Journal of Ornithology*, 153(1):3–14, 2012. [1](#)
- [25] K. G. Horton, W. G. Shriver, and J. J. Buler. A comparison of traffic estimates of nocturnal flying animals using radar, thermal imaging, and acoustic recording. 2015. [3](#)
- [26] J.-B. Huang, S. B. Kang, N. Ahuja, and J. Kopf. Image completion using planar structure guidance. *ACM Trans. Gr.*, 33(4):129, 2014.
- [27] J.-B. Huang, A. Singh, and N. Ahuja. Single image super-resolution from transformed self-exemplars. In *CVPR*, 2015.
- [28] C. Kirbas and F. Quek. A review of vessel extraction techniques and algorithms. *ACM Computing Surveys*, 36(2):81–121, 2004. [2](#)
- [29] F. Liechti, B. Bruderer, and H. Paproth. Quantification of nocturnal bird migration by moonwatching: Comparison with radar and infrared observations (cuantificación de la migración nocturna de aves observando la luna: Comparación con observaciones de radar e infrarrojos). *Journal of Field Ornithology*, pages 457–468, 1995. [3](#)
- [30] C. Loop and Z. Zhang. Computing rectifying homographies for stereo vision. In *CVPR*, 1999. [5](#)
- [31] T. A. Marques, L. Thomas, S. W. Martin, D. K. Mellinger, J. A. Ward, D. J. Moretti, D. Harris, and P. L. Tyack. Estimating animal population density using passive acoustics. *Biological Reviews*, 2012. [3](#)
- [32] N. Ofir, M. Galun, B. Nadler, and R. Basri. Fast detection of curved edges at low snr. *arXiv:1505.06600*, 2015. [2](#), [3](#)
- [33] D. Sheldon, A. Farnsworth, J. Irvine, B. Van Doren, K. Webb, T. G. Dietterich, and S. Kelling. Approximate bayesian inference for reconstructing velocities of migrating birds from weather radar. In *AAAI Conference on Artificial Intelligence*, 2013. [3](#)
- [34] R. Toldo and A. Fusiello. Robust multiple structures estimation with j-linkage. In *ECCV*, 2008. [6](#)
- [35] Z. Wu and M. Betke. Global optimization for coupled detection and data association in multiple object tracking. *CVIU*, 143:25–37, 2016. [2](#), [3](#)
- [36] W. Zhang, M. Cong, and L. Wang. Algorithms for optical weak small targets detection and tracking: review. In *International Conference on Neural Networks and Signal Processing*, 2003. [2](#)
- [37] Z. Zhang. Determining the epipolar geometry and its uncertainty: A review. *IJCV*, 27(2):161–195, 1998. [5](#)