Heartbeat of a Nest: Using imagers as biological sensors

Shaun Ahmadian, Teresa Ko, Sharon Coe, Michael Hamilton, Mohammad Rahimi, Stefano Soatto, Deborah Estrin

ssahmadian@gmail.com, tko@cs.ucla.edu, sharon.coe@ucr.edu, mhamilton@jamesreserve.edu, mhr@cens.ucla.edu, soatto@cs.ucla.edu, destrin@cs.ucla.edu

Abstract

We present a scalable end-to-end system for vision-based monitoring of a biological phenomenon. Our system enables automated analysis of thousands of images, where manual processing would be infeasible. We automate the analysis of raw imaging data using statistics that are tailored to the task of interest, the study of avian behavior during nesting cycles. The system uses simple image statistics (features) as the low-level representation to be fed to generic classifiers and final inferences exploit the temporal and spatial consistencies. Our testbed achieves bird detection accuracy of 82%, and egg counting accuracy of 84%, allowing inference of avian nesting stage with accuracy within a day. Our results demonstrate the challenges and potential of using imagers as biological sensors. An exploration of system performance under varying image resolution and frame rate suggest that an *in situ* adaptive vision system is technically feasible.

1 Introduction

In this paper we present the design and deployment of a vision system of embedded wireless and wired imagers, and automated server-based image processing to monitor avian behavior during a nesting cycle. Birds are important indicators of the health of ecosystems. Therefore the ability to measure bird nesting patterns accurately and in a scalable manner is broadly relevant to ecosystem studies, including responses to climate change and land use. Currently, avian biologists personally inspect nesting locations and *visually* log the stage of the nest for future analysis. What they hope to find are trends and differences in behavior that ultimately influence reproductive success. Some indicator variables include the number of eggs that are laid and eventually hatch and the occupancy of the nest over the different nesting stages. By visiting the sites to collects this information, biologists can incorporate domain knowledge to filter unwanted data and detect important events in real time. However, not only is this time consuming but it limits the number of observations a biologist can collect. Moreover, observations are typically limited to the behavior of birds outside their nests "...because [they] lacked the capability to peer inside the private lives of birds"[1] for long durations for fear of disturbing the birds.

Imagers can play an integral part in this data gathering process. First, imagers can be placed *in situ* at locations of interest with minimal disturbance to the phenomenon being observed. Secondly, through advances in low-power image technology, wireless imagers enable access to locations otherwise inaccessible to wires. Thirdly, current vision techniques can be applied to the massive sequences of collected images to automatically infer statistics of interest for the end-user. Lastly, and most importantly, control over the imagers can allow for constraints to be placed on the collected images, more specifically the type of objects to expect and the scale (size in image) at which they should be detected. In turn these assumptions can be incorporated into the vision inferences making it easier to automate. And this is only possible by considering the inference problem as an end-to-end system.

Beyond deploying the system and collecting data, the biggest challenges to realizing such a biological tool are in automating the vision inference process. The sample images from a nesting season in Figure 1 demonstrates some of the challenges of using computer vision for inference. Environmental factors such as lighting and bird behavior, and non-environmental factors such as sensor sensitivity and placement, introduce noise that have to be overcome. Consequently, part of the vision algorithm includes the extraction of low-level *features* (i.e. statistics) from the images that are minimally impacted by the presence of such variability, yet discriminating enough for accurate classification. Furthermore, each of the multiple inferences biologists make requires its own separately tuned algorithm.

We developed and deployed an end-to-end system consisting of *in situ* image capture combined with server side image analysis techniques, resulting in a powerful and scalable experimental tool for biologists. The system samples the environment at a constant rate and delivers images to a central repository for the entire duration of the nesting season. Once there, different vision algorithms are applied to the image sequences in order to extract features. Then spe-

[1] private communication Prof. John Rotenberry, UC Riverside

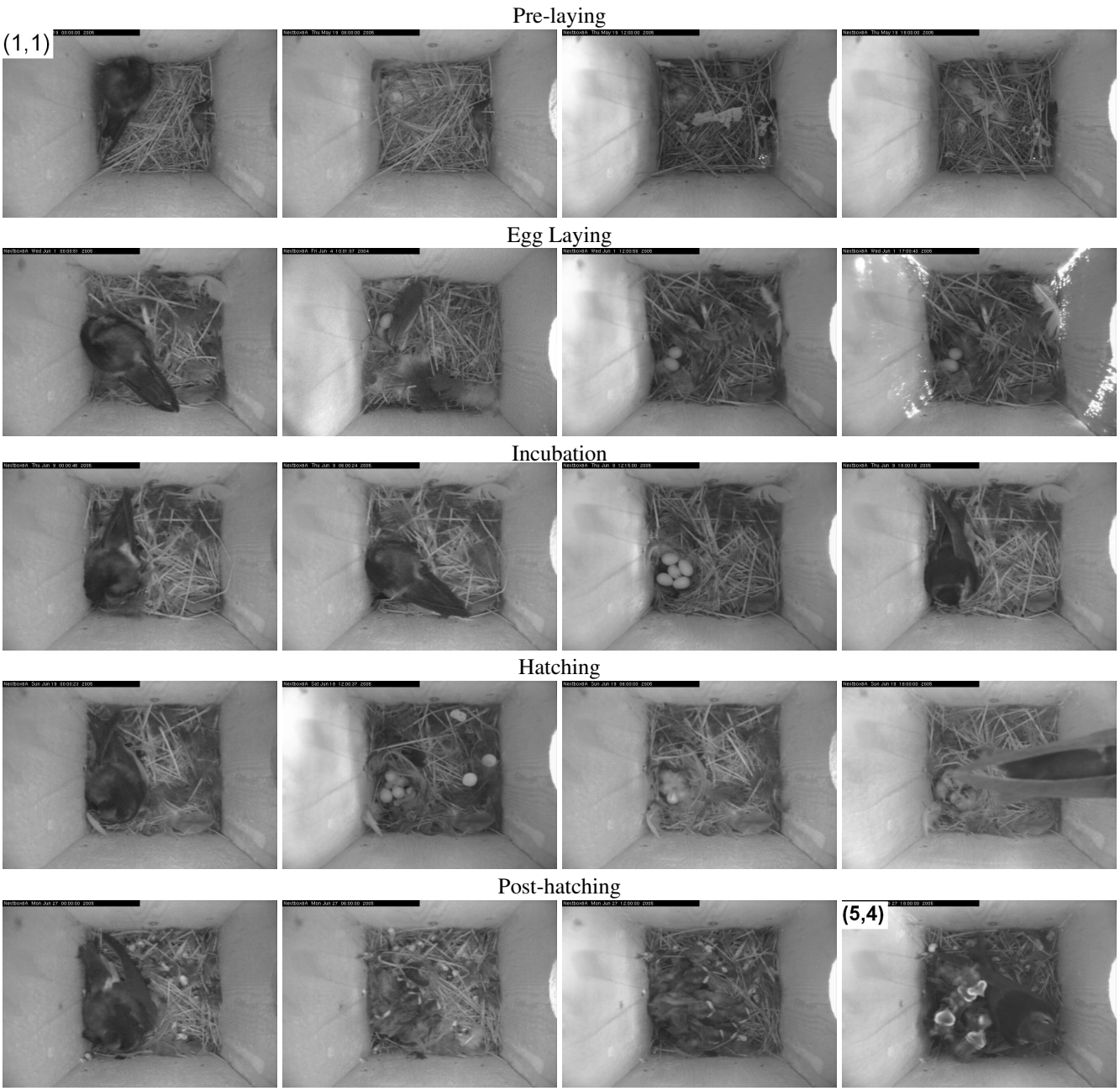

Figure 1. Images from various stages of the nesting season taken from a wired imager as described in Section 2.1. First thing to note, the system can be designed to provide constraints, a *top view* of the *box interior* from a *fixed* location, which can help reduce the search space of the problem. Even so, noise from environmental factors make this challenging. Local image statistics (intensity, gradient orientation, histograms) are often similar for the birds or hatchlings and the background nest, see for instance image (3,1). Eggs are occasionally occluded by feathers (2,2), or by straw (4,2), and objects appear that are similar to eggs (4,2) which make the count non-trivial even for trained human experts. And there are wide variation in intensity profile depending on the time of the day, including sharp highlights (2,4) from direct sunlight.

cially tailored classifiers infer biologically significant behavior, in particular presence/absence of the bird, number of laid eggs, and the nesting stage transitions. Therefore, massive (thousands of) image sequences can be translated to time-series of various statistics. These time series traces can be visualized to highlight spatial and temporal trends across a single or multiple nest boxes, and aggregate statistics derived to quantify observed trends.

We briefly summarize the systems implications of our work for other vision systems. First, systems can be designed to minimize the impact of errors from vision inferences by taking advantage of domain knowledge and system constraints. For example, we leverage the spatio-temporal consistency that exists in our application, uncommon in many vision problems. Second, image quality, and therefore the final inference, is significantly affected by aperture size, pixel-filters, and light sensitivity of the imager, not only resolution. Third, vision systems can and should use image resolution as a knob to control resource consumption, similar to how sampling is adjusted in other monitoring systems. These implications are discussed in greater length throughout the paper.

Our contributions are two fold. This is the first paper to the best of our knowledge to present the deployment of an end-to-end vision system, covering all steps from the continuous capture of images to inference of aggregated statistics. Secondly, we show that we can provide "useful" inferences rather than raw image sequences through the coupling of system constraints, such as temporal consistency, and vision techniques. Consequently, our system detects the presence/absence of a bird correctly 82% of the time, provides the correct egg count 84% of the time, and detects stage transitions to within a day.

2 System Overview

Our system is an end-to-end image sensing system consisting of *in situ* image sensors, a combination of wired and wireless transmission, image archiving and visualization, and an integrated image processing and data analysis system running on the server. In this section, we provide an overview of our experimental setup and a more detailed view of the deployed system components. Also the challenges that we encountered in deploying those systems are shared. The heart of the paper is in Section 3 where we present the application and evaluation of vision techniques, and show how the inferred statistics can be useful from a biological standpoint. In Section 4, we explore the implications of our data analysis on the system. We discuss supporting work and their implications to our work in Section 5. We conclude in Section 6 and share ideas for future works.

2.1 Deployment of Imagers

The active experimentation season for avian studies is initiated by warming temperatures in early April when birds start moving to higher altitudes in the James San Jacinto Mountain Reserve for nesting and breeding. Based on current trends in the area the season lasts anywhere from 2-3 months until the end of July. There are currently 40 human-made nest boxes across the reserve area (Figure 2) which are regularly visited by the biologist during the avian breeding season to log important information such as the seasonal occupancy of the nests, species of bird, or timing of the major

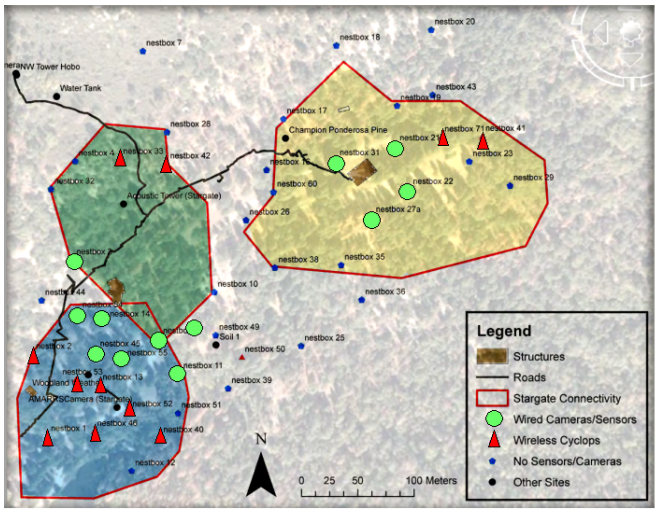

Figure 2. Map of James San Jacinto Mountain Reserve annotated with wired (large circles) and wireless(triangles) instrumented nest boxes. The solid points are additional nest box locations that have yet to be instrumented. The wireless connectivity clouds show that it is possible to reach a major portion of these uninstrumented boxes as well as any new sites.

breeding stages. A subset of the nest boxes have been instrumented with wired video cameras that periodically take images and get recorded into a database.

The wired system has high quality imagers that provide a very accurate representation of the birds activity throughout the season. However, the wired system consumes significant power, requires expensive infrastructure, and has limited geographical coverage. The latter severely hinders expansion of the wired camera system particularly in light of minimum intra-nest distances to avoid avian territorial conflicts. Hence, only 13 nest boxes that are close to available data and power facilities are instrumented with this system (installed since summer of 2001).

To explore avian breeding in a wider geographical area, the system was augmented with battery-powered wireless Cyclops cameras. The advantage of a wireless system is spatial coverage and minimum infrastructure requirements. However, there are trade-offs in that the wireless system does not have the image quality and image rate capabilities of wired imagers and requires batteries to be replaced intermittently. In the next two subsections we provide a brief overview of both wired and wireless system and discuss some of the important system challenges. The automated image analysis techniques described in Section 3 are designed to be applicable to both wired and wireless platform types, however the specific algorithms tested here were tuned to the imagers used in the wired nest boxes because there were larger numbers of them deployed and therefore occupied with birds at the time of this study. We note that for Spring of 2007 season, we have deployed a larger set of wireless nest boxes and hope to have sufficient occupancy to provide a much richer wireless dataset.

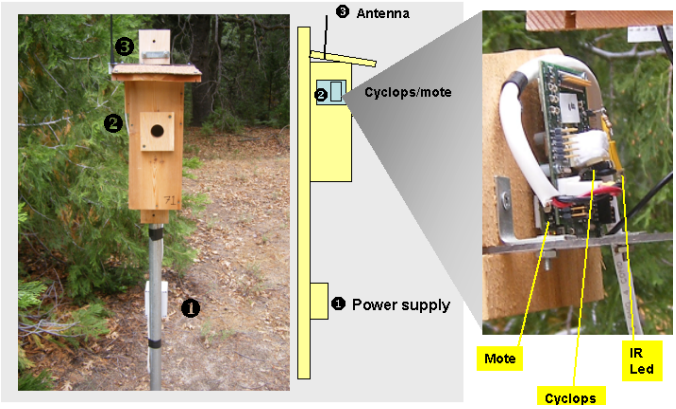

Figure 3. Typical wireless instrumentation of a nest box. In the interior near the top, a camera sensor node is located on a transparent tray looking down with minimal intrusion into the bird's nesting activities. Infrared LEDs provide a more consistent illumination of the box especially at night without disturbing the bird.

2.2 Existing Wired Image System

The existing wired camera network consists of miniature video CCD cameras, high performance video servers that networks up to four cameras and a database for recording the images. The camera uses a high quality VGA (640×480) black & white Sony CCD sensor with analog output that continuously acquires images at 10fps. It is a small module that operates at very low light intensity settings (down to 0.0003lux). In turn this provides images with high signal to noise ratio in the dim lighting condition of the nest. The drawback for such a high performance (i.e. image resolution, image rate and high SNR) is a high power consumption of 1.2W. Hence, each camera is powered by a 12V supply that is located in the infrastructure facility through cables protected by conduits buried under ground. To capture images from within the nest, the camera is placed on the top side of the nest box on a transparent tray looking down onto the nest (Figure 3). An infrared LED which consumes 24mW continuously illuminates the nest to enable high quality image acquisition in the dark nest environment. Although the infrared light source is not at the center wavelength of the sensitivity of the camera, it has been chosen to minimizes any disturbance on the normal bird activity.

The wired nodes are divided into four groups each connected to a high performance video server. Video servers which are located in two of the facilities in the reserve, each support 4 analog input video channels and are accessible for configuration and programming through the Internet. A PhP script that runs on the video servers collect images from the camera nodes and places them into a database at 15 minute intervals. The video servers consume significant power, hence are directly connected to the reserve's power line. Since deploying each wired camera sensor is a labor intensive task that involves trenching ground and laying out conduits and power and data cables, only the 12 nest boxes located closest to the reserve facilities are instrumented with the wired system, thereby severely limiting the systems geographical coverage (Figure 2).

2.3 The Wireless Imaging System

We introduced our first wireless imagers in the spring of 2006. The wireless system uses pairs of Mote/Cyclops [15] as sensor nodes[2], Stargate micro-servers [18] for network management and image acquisition scheduling, and a database for archiving and managing the image data.

Each Cyclops sensor node uses a Mica2 mote for radio communication and a Cyclops for image acquisition. Cyclops has a CIF (320×280) color Agilent CMOS image sensor which consumes 22mW and is capable of acquiring images at a minimum of 5Lux scene luminescence. To enable image acquisition under the nests' dim light conditions, we removed the internal infrared filter of the CMOS sensors and used an infrared LED to illuminate the box during the image capturing period. The Cyclops sensor has various power-saving features but we only exploited a few of them to minimize the risk of reliability problems in this first deployment; the implemented features included radio duty-cycling to minimize radio power-consumption, image sensor scheduling to take still images and minimize CMOS sensor power-consumption, and infrared flashing only during image acquisition period. As configured, the deployed Cyclops sensor nodes each consume 45mW during image transmission and 28mW during idle mode. To provide power to the Cyclops sensor, we used four D-size alkaline batteries to sustain the sensor nodes for an average of 15 days capturing images at 15 minute interval before renewing the batteries. The deployed system has *not* been optimized for energy consumption to date, and based on our system experience, we expect the energy consumption to be significantly reduced for future tuned deployments.

Micro-servers, which are Stargate nodes running Linux, have a broadband radio and a mote radio to communicate with Cyclops sensor nodes for image configuration, image acquisition and reading of the battery level. A script on the micro-servers send periodic battery reading, image configuration and image acquisition messages to each Cyclops sensor and retrieves the results. The micro-servers, which are accessible via Internet, continuously compress the acquired images and upload them to a server for archiving in a database. We have a web front-end that allows the user to browse the gallery of acquired images.

Figure 2 illustrates the map of our wireless system deployment in the reserve, which consists of three areas, each equipped with a micro-server serving a total of 11 Cyclops sensors. Our wireless system was operational from late March until early June and provided on average between 4000-10,000 images depending on connectivity. Overall, 4 of the wireless instrumented nest boxes were regularly visited by a bird during the season and 3 of them eventually inhabited.

In our deployment, the quality of the image sensors had a significant impact on the performance of our automated image analysis. Although, quality is usually equated to the resolution of the images, in our experience there were other factors that contributed to the quality loss. Important parameters include: *aperture size*, *light sensitivity*, and *pixel-filters*

[2]We call a pair of Mote/Cyclops as Cyclops sensor through the rest of the paper.

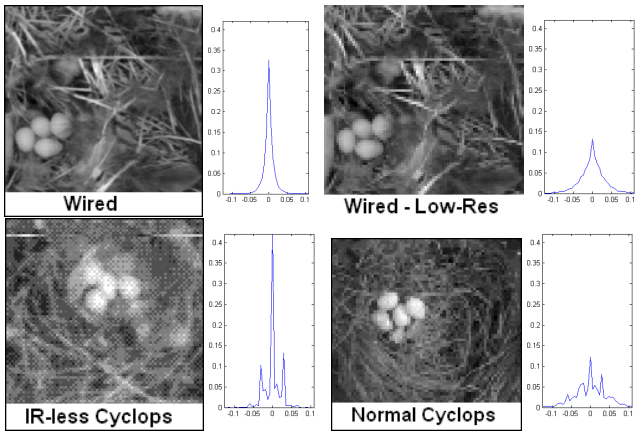

Figure 4. Even though the two images are of the same resolution, the image quality of the Cyclops camera is inferior to the wired cameras. The image quality of the Cyclops with IR better approximates the wired cameras suggesting that future results will improve.

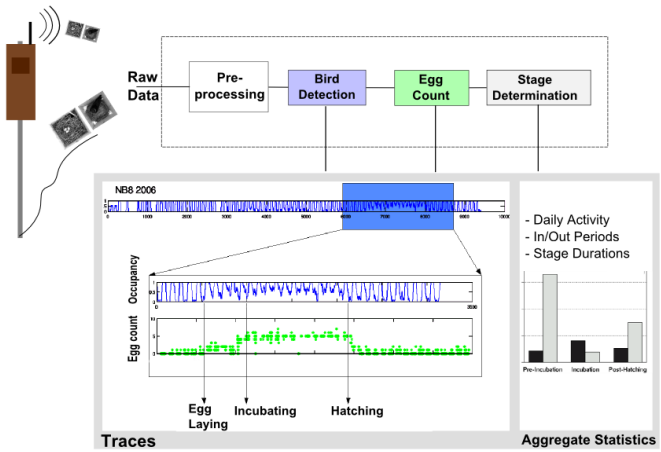

Figure 5. Components of the vision inference system Data collected by the wired and wireless imagers are inputs into the system, time-series traces of inferred behavior and aggregated statistics are outputs.

of the camera. The former two affect the minimum illumination required of the nest. Since the CCD Sony sensor in the wired system is capable of operating in very low light intensity settings, it provides images with more spatial detail. The pixel-filter determines the frequency response of the pixels in the imager. The Cyclops is a color sensor with filters designed for the visible spectrum of light. Typically one of these filters is an IR filter placed near the aperture of the imager to act as low-pass filter. For our deployment, we had to remove this filter to capture images under infrared lighting. This had two major impacts on the eventual quality of our images. First the pixels of the camera are more prone to noise when operating at non-nominal frequencies (IR range). Secondly, since the imager is comprised of red, blue, and green pixel-filters, each color is modulated by different amounts, exacerbating the already noisy readings. Together, these resulted in low-contrast images with grid-like noise patterns.

To make explicit the differences in image quality, Figure 4 shows representative images collected from the wired and Cyclops camera with their respective gradient histograms. Since gradients are used in the algorithms we exploit in our system, we look at how the distribution of gradient magnitudes are affected by different image sensors under various conditions. The distribution of the gradient from an image captured by a wired camera is typical of a natural image. The gradient histogram of a down-sampled image becomes a smoothed out version of the original image showing a similar distribution of peaked low gradients and smooth drop off. A Cyclops under nominal conditions (with IR filter and normal lighting conditions) that outputs an image with similar resolution does produces a slightly poorer quality image as seen by the noise in gradient histogram. However, the removal of the IR filter from the Cyclops to take images under infrared illumination results in a very distorted gradient histogram. As the evaluations in Section 3.2 show, these alterations greatly impacted the accuracy of the inferences made on the wireless data set. For this reason, we hope to resolve these issues in future deployments by replacing the current CMOS sensor with a monochrome sensor with higher sensitivity.

3 Inference Architecture

We now turn our attention to inferring high-level information from images collected *in situ*. The scientists' requirements provide a natural organization of the system: Raw images are first processed by a *"bird detection module;"* next, images not containing birds are analyzed to *detect the presence and infer the number of eggs;* finally, the egg counts are fitted to a probabilistic finite-state machine to determine *nesting stages and transitions* among them.

A challenge in image analysis is the wide variability of the data with respect to "nuisance factors" that are of no interest to the task, but that nevertheless greatly affect the data. For instance, we wish to detect the presence of a bird regardless of its position, orientation, color, texture profile, ambient illumination etc. And yet, these nuisance factors result in macroscopic changes in the image that far exceed those due to the onset of the event of interest, say the appearance of an egg (Figure 1).

We address these challenges by exploiting local image statistics (a.k.a. "features") that are designed to be *insensitive* to illumination changes. These are located at automatically-detected *interest points*, that makes them insensitive to location and scale changes. Examples of interest points include location of edges and corners; features can include mean/variance of the intensity values within a subregion of the image. Both interest point and local features can be computed using only local gradient operations. A collection (or "bag") of such features is then fed to classifiers tailored to the problem. Additionaly, a hidden Markov model demonstrates the way *temporal consistency* of the nesting cycle can be exploited.

3.1 Evaluation Setup

Due to avian occupancy rates (20-30%) in the region and connectivity issues, only images from one full season and two half seasons were obtained from the wireless imagers. Consequently a major portion of the data analysis was done on wired images. To compensate for this, all wired images

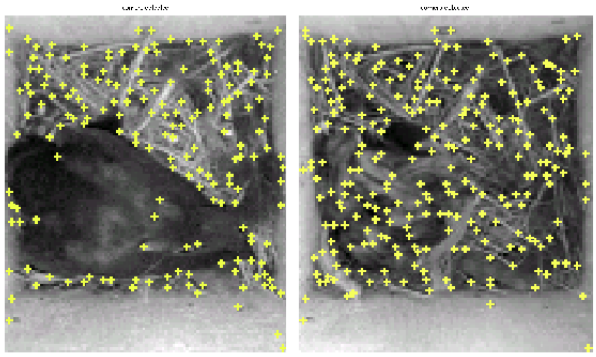

Figure 6. Corners detected by algorithm with bird (left) and without (right).

were down-sampled to wireless image size (128x128) in the data analysis sections. The exploration in Section 2.3 suggests this is reasonable given that by using a camera module more favorable to the conditions experienced in the deployment, the quality would come close to that of down-sampled wired images.

For the purpose of evaluation, we employed 10 full seasons of manually labeled image data of nesting activity (1 wireless, 9 wired). Despite the infrared capability of the sensor, daytime data was the primary focus since most avian activity takes place between the hours of 5am and 8pm. Daytime data are challenging because changes in illumination, nest appearance and motion of the bird are most severe. The image sequences were taken from the moment a complete nest was constructed to a week after the eggs had hatched. These were chosen because they can be reliably detected automatically, as we will illustrate. Images taken during the latter stages of young development were not analyzed since size and color of the young makes it difficult to distinguish from the adults even for a trained human.

3.2 Bird Detection

Our first inference goal is to determine the presence of a bird in the nest. The key idea is to identify a *simple discriminating feature*. This is an image statistic that is insensitive enough to nuisance factors and yet discriminative enough[3] to enable subsequent stages of processing. Feature design is a fine art where domain knowledge plays an important role. Here we present simple strategies that are general enough to be useful in a wide range of image analysis scenarios.

3.2.1 Pre-processing

At the outset, because the cameras are static, images are cropped to reduce the region of interest to the bottom portion of the nest. Next the cropped image is normalized with respect to a patch from a remote region of the nest box that we identified before hand as being disturbance free.

3.2.2 Interest point detection

In order to avoid processing large regions of the image that are approximately homogeneous, we restrict our attention to regions commonly called *interest points*, or *key points, salient points*, or *corners*. These are regions where

[3]In the sense of achieving sufficient classification performance when fed to a simple classifier.

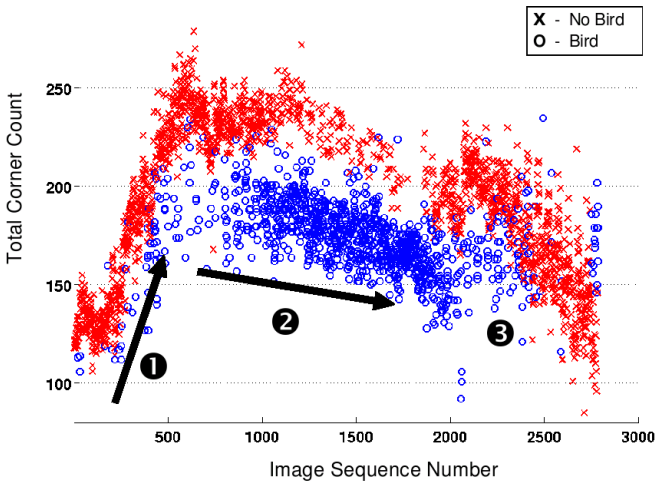

Figure 7. Total corner point counts over sequence of images for one nest box. We are able to view the features extracted from images in such a way since our images have consistency over time and space.

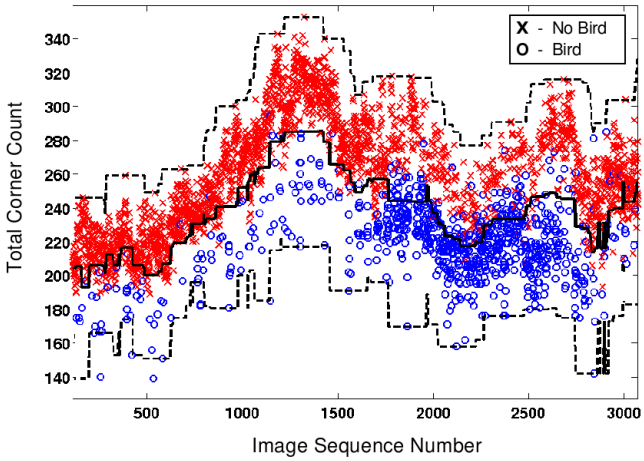

Figure 9. Total corner point counts over image sequences. The top and bottom dashed lines are the min and max values of the 4 day window respectively. The solid middle line represents the average of the min and max values. It also turns out to be a good determinant of the boundary between bird and no bird images.

the image exhibits large gradients in two independent directions, and are detected as extrema of scale-space using established techniques dating back to [9].

A side benefit of this computation is that birds often exhibit smooth feather coats, and therefore trigger few interest points, whereas the nest and its high-contrast texture triggers many, as is evident in Fig. 6. Hence the density of interest points can be interpreted as a texture descriptor, which in turn can be used to easily detect the presence of a bird. For instance, Fig. 7 shows the total number of interest points per image as a function of time: Early on in the season the bird is building the nest (label 1) and one can observe a sharp increase in the number of interest points in the image. These level off as the nest is completed and the bird proceeds to lay eggs and incubate them (label 2). After hatching the young

Stages	NB8 2004	NB8 2005	NB8 2006	NB55 2005	NB71 2006
Pre-Incubation	88.0%	65.1%	88.8%	87.8%	54.0%
Incubation	84.2%	83.2%	90.9%	95.6%	60.1%
Post-Hatching	85.9%	80.9%	74.2%	89.9%	52.0%

Stages	NB31 2004	NB31 2005	NB31 2006	NB54 2005	NB55 2006
Pre-Incubation	73.8%	61.3%	78.4%	93.4%	84.2%
Incubation	90.2%	81.8%	90.1%	92.3%	93.6%
Post-Hatching	60.2%	77.4%	67.5%	95.7%	73.0%

Figure 8. Performance when applying bird detection algorithm to Violet-Green Swallow (top) and Western Bluebird (bottom) nest boxes

start to cover more and more of the nesting material as they grow (label 3) hence decreasing the interest point count.

3.2.3 Classification

In order to classify images as having a 'bird' or 'no bird,' we use the fact that interest point density is a good discriminator for the task of detecting the presence of the bird as can be seen in Fig. 7. A cross indicates absence of bird and a circle presence. Since the classes are roughly separable, but not linearly so, one could unleash a variety of supervised classification techniques such as kernel methods [17]. However, if we take advantage of the *temporal consistency* seen in the counts a simpler solution is possible. A sliding window can be used to adaptively set a sufficient threshold to support the subsequent levels of processing. In particular, a simple midpoint of min and max values over a sequence of 200 frames provides enough support to set a time-varying threshold that achieves consistent classification throughout the sequence.

3.2.4 Evaluation

To evaluate the algorithms ability to detect bird presence, ten seasons worth of images were used; first to obtain interest points and then classify according to the min-max average. For interest point detection, an implementation of the so-called Harris-Stephens corner detector was used over each image sequence [6] to generate a time series of the total counts. These all followed the qualitative trends described in the previous section and allowed detecting the completion of the nest-building phase to within one day. The end point of the image sequences were selected a week after egg hatching had occurred. To determine the appropriate threshold for each sequence of interest point counts, a min-max filter was applied over a 4 day window. For each point the min and max interest count over the window was determined and the average of the two was used as the threshold. The performance of this simple strategy is illustrated in Figure 9. A mixture model of two Gaussians was also applied, but due to the skewed distribution of points at certain stages in the season, the algorithm would fail to find an appropriate model, and at other times similar results to the min-max filter were obtained.

The accuracy results of the algorithm can be found in Table 8. The results are separated according to species as well as nesting stages. Generally the nest boxes with Violet-Green Swallows (top table)performed better and were more consistent across nesting stages than Western Bluebird boxes (bottom table). This can be attributed to the way swallows construct their nests; the makeup and construction of their nests are simpler and have a shallower cup allowing the interest point detection to pick up an even distribution of interest points, while for Western Bluebird it is more patchy.

A second observation is that the bird detection algorithm is most accurate during the incubation phase. During this phase the nest is most stable with minimal disturbance from the birds. In pre-incubation, nest material can be added or altered – a perfect example being NB8 2005. In this image sequence, the bird brought in feather-like nesting material that caused the number of interest points in the image to decrease enough so that they would be wrongly labeled as having a bird.

The worst performing of all was NB71, which also happened to be instrumented with the wireless nodes. Its poor performance can be attributed to the image quality as described in Section 3.3 that made it difficult to discriminate the presence/absence of a bird even to a human observer. We are currently upgrading our wireless system to use a more appropriate camera module, for which the results can reasonably be expected to be on par with the wired cameras.

3.3 Egg Count

Determining the number of eggs in a nest box is investigated for two reasons. First, the biologist would like to know how many eggs are laid. Second, the number of eggs informs the biologist of the stage the nest box is in. Unlike the bird detection problem, here we have a multi-class decision, with an a-priori unknown number of classes. To solve this multi-class decision, we pose a binary decision at a given sub-image centered around different location in the image and at varying scales. We have multiple stages in which each stage rule out many candidate sub-images , so that latter stages may run more efficiently. The resulting number of positive decisions is used as the count of eggs in a given image. What makes this task easier is the fact that we are not interested in the number of eggs in a particular image, but rather the number of eggs laid in a nest box. As can be seen in Fig. 1, a single image is often insufficient to arrive at a reliable egg count even for a human observer. Therefore, each image will provide only what is called a *weak classifier*. These classifiers are then combined into a *HMM* [14], exploiting temporal consistency to improve the overall reliability despite the modest performance of classifiers based on individual images.

Our strategy is outlined in Fig. 10. First, we filter out images in which birds have been detected, exploiting the redundancy that arises from having fine temporal sampling. Then, an image is broken down into candidate location and scale combinations that are tested for eggs using a variety of heuristics that we describe next. In the following we assume a perfect bird detection algorithm so that the analysis can reflect the performance of the egg detector alone. As in the bird detection algorithm, we exploit the fact that the camera

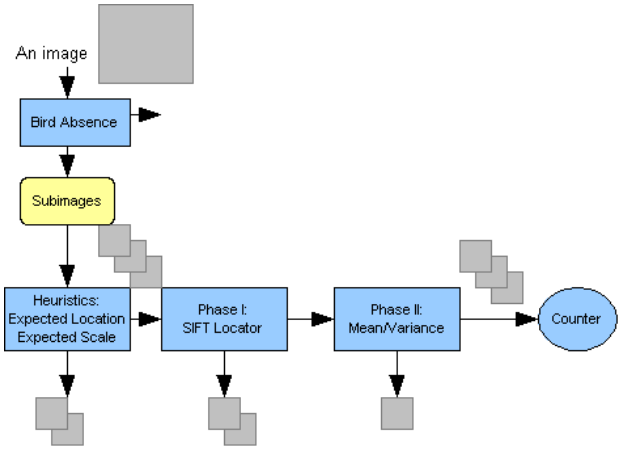

Figure 10. Decomposed egg counter. First images with birds are ignored. We further simplify the task by disregarding locations and scales where eggs cannot be due to the choice of imager location and orientation relative to nest box. SIFT Detector is used to further reduce the possible location/scale candidates for eggs. Finally, a linear classifier is used to separate eggs and non-egg images on the remaining candidates. The number of positively classified candidates is the final output of the egg counter.

Figure 11. The SIFT features found for an example image of 5 eggs. Blue marks indicate SIFT features corresponding to eggs.

in a fixed position relative to the nest. This allows us to neglect the periphery of the image as it portrays the walls of the nest box.

3.3.1 SIFT Detector

In the remaining region, we search for candidate egg location using a "blob detector" at multiple locations and scales, based on scale-space image processing following [9]. This is popularly known as "SIFT detector" where SIFT stands for scale-invariant feature transform [10].

The SIFT detector is based on maximal response of a Laplacian of Gaussian filter, approximated for efficient computation with a difference-of-Gaussians. Such filters look like elongated, oriented "blobs," with scale and orientation determined by the covariance parameters. The location of the maxima, together with the associated scale automatically detected by the algorithm are displayed in Fig. 11. In a cascade of classifiers approach it is important that each weak classifier has few missed detections, although it can have a large number of false alarms [4]: The precision/recall curves shown in Table 1 show precisely that. When the SIFT detector misses an egg, it is usually because it is (partially) occluded by nesting material, feathers, or other eggs, as shown in Fig. 1. In Table 1, 'Image Recall' gives the percent of images where *all* eggs are detected, providing an upper bound on the final performance.

nest box	Precision	Recall	'Image Recall'
NB8 2004	5.8%	96.03%	97.26%
NB8 2005	4.5%	91.63%	94.88%
NB8 2006	4.8%	97.86%	98.55%

Table 1. This table indicates that SIFT has high egg recall and low precision. 'Image Recall' refers to the percent of images where *all* eggs are detected. This number tells us the upper bound on our final performance.

3.3.2 Mean/Variance

Unfortunately, one of the most common approaches currently used to this end in Computer Vision for object detection and recognition in clutter fails in our context. The so-called "bags of features" approach [11] is based on computing a SIFT descriptor, or SIFT key, at each location found by the SIFT detector just described. The SIFT descriptor is a quantized histogram of gradient orientation that is designed to be insensitive to illumination changes. Collecting a number of SIFT keys into a set (a "bag"), regardless of their position and scale, provides insensitivity to geometric transformations and object locations. Finally, robust comparisons allow for considerable robustness to clutter, for many SIFT keys will fall outside the object of interest.

Unfortunately, because the egg has approximately uniform intensity, the SIFT descriptor turns out to characterize the distribution of nesting material around it, rather than the egg itself. A support vector machine (SVM) [23] trained on one nest box using manual labeling of SIFT keys and tested on the others performs essentially at chance level (53.43% correct). An additional difficulty is that hatchlings look very much like eggs at low resolution and the presence of egg shells is a confounding factor (Fig. 1).

For these reasons, we abandon the SIFT descriptor altogether, and just retain the SIFT detector as the building block of our weak classifiers. The simplest discriminating statistics in this context are just the *scale* at which SIFT detection occurred (eggs have a characteristics range of sizes), and the *intensity mean and variance* inside the region detected by SIFT. The image is first normalized for contrast and scaled globally to reduce the effects of illumination changes. A simple linear classifier (Fisher linear discriminant) based on these features was trained on one nest box and tested on the others, with an average performance of 72.34% correct. Performance across nest box can be found in Table 2, and specific counts for a particular nest box is shown in Figure 13.

The number of eggs usually increases monotonically during the egg-laying phase, and decreases rapidly during hatching, but otherwise stay the same. To enforce this temporal consistency, these weak classifiers are fed into an Hidden Markov Model, allowing significant improvement from

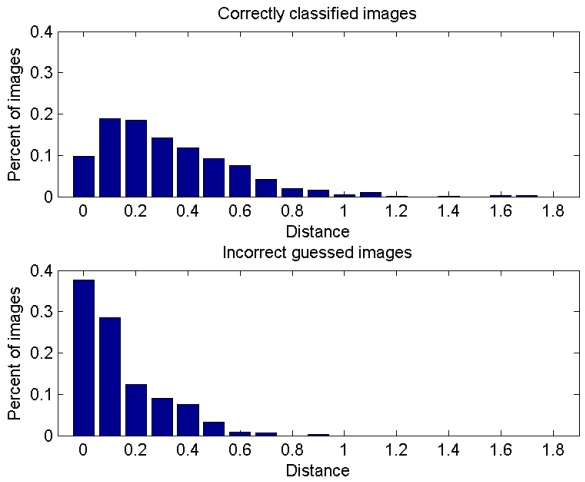

Figure 12. Histograms showing the distribution of the distance from the boundary on correctly classified images and incorrectly classified images. Most incorrectly classified images have samples which are close to the boundary.

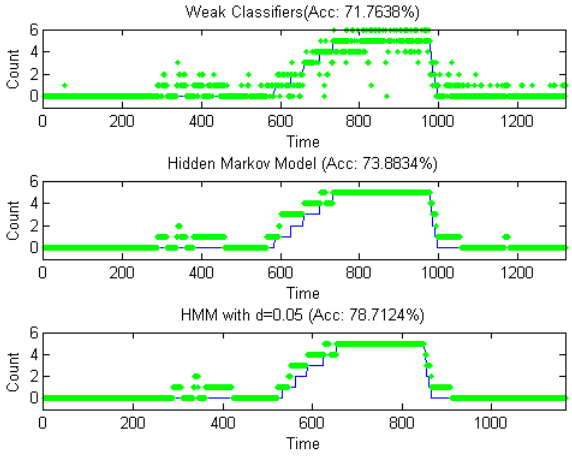

Figure 13. Time-series plot of predicted egg counts for a particular nest box. This corresponds with the accuracy reported in Table 2 for nest box NB8 2006.

nest box	Weak	HMM	HMMv2
NB8 2004	71.76%	86.67%	91.42%
NB8 2005	73.12%	78.93%	82.24%
NB8 2006	72.13%	73.88%	78.71%

Table 2. The accuracy of the egg counter. First shown is the accuracy of individual images (Weak). Accuracy improves when enforcing temporal consistency (HMM) and even further, when rejecting weak classifications (HMMv2).

the performance of the weak classifiers. The HMMs were trained on data from 2 nest boxes and tested on the remaining one. The states of the HMM were the true egg count and the observations were the estimated egg counts. In this case, the average accuracy was 79.83%. As before, the performance and specifics can be found in Table 2 and Figure 13.

Recognizing that there are some images that are harder to classify than others, we hypothesize that there is a correlation between the distance the features are from the decision boundary and the chance of correct classification. Figure 12 shows that there is indeed a correlation. We use this distance to suggest the confidence we have in our decision, and reject images in which we have low confidence. The average accuracy is improved to 84.12%, when rejected images within 0.5 of the decision boundary. Performance and specifics are reported in Table 2 and Figure 13, as before.

3.4 Nesting Stage Determination

Given an accurate egg count, it is possible to infer the nesting stage of the bird. Eggs are only present during the egg-laying, incubation, and hatching stages, with an increase in eggs during egg-laying, a constant number during incubation, and a decrease during hatching. Similar to the previous section, an HMM is used to reduce the temporal inconsistencies of the weak egg counter classifier. Rather than improving the overall egg count, this time an HMM is trained to infer the nesting stage. The states of the HMM are the nesting stages, while the observations are the number of eggs detected by the egg counter. An HMM will find the most likely stage transitions for a *sequence* of egg count observations given the temporal constraint that one stage must follow logically to the next. This can also be thought of as an attempt to minimize the amount of contradictions that the weak egg count classifier gives. Stages of short duration have a disadvantage because it does not have many observations to accurately represent the stage. Errors close to any transition is very costly.

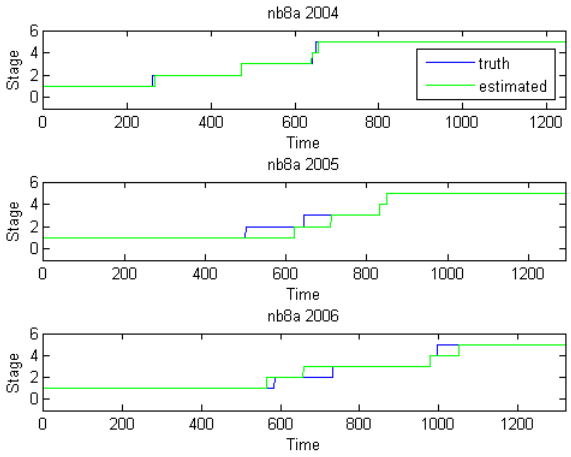

Figure 14. Resulting estimation of nestbox stages for various nest boxes.

We use the results from the weak classifiers rather than the smoothed HMM version or the HMM with removed images. These resulted in inferior results, perhaps because specific biases are reinforced for each particular training set. Using the weak classifier, the resulting estimated time and duration of nesting stages closely follow the true stages of the nesting cycle, as shown in Figure 14. The results are given for the case where no images are rejected. With requiring increasing

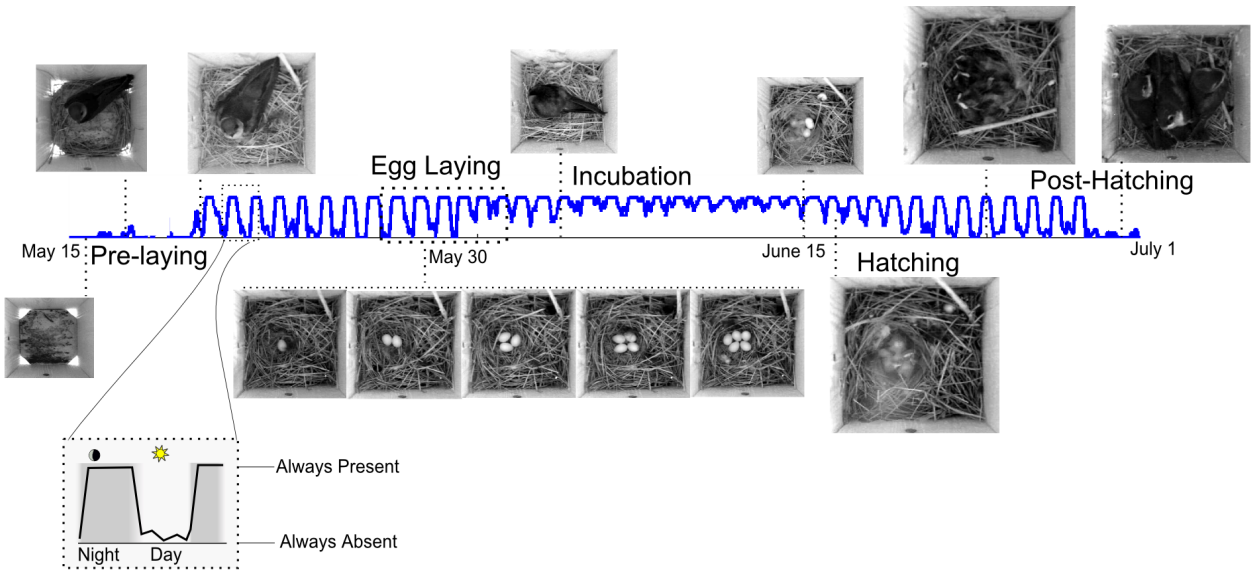

Figure 16. Results of the automated analysis of presence/absence from a sequence of 5000 images has been smoothed and plotted. A visual effect of this representation is that of a "heartbeat" signal as seen on most EKG monitors.

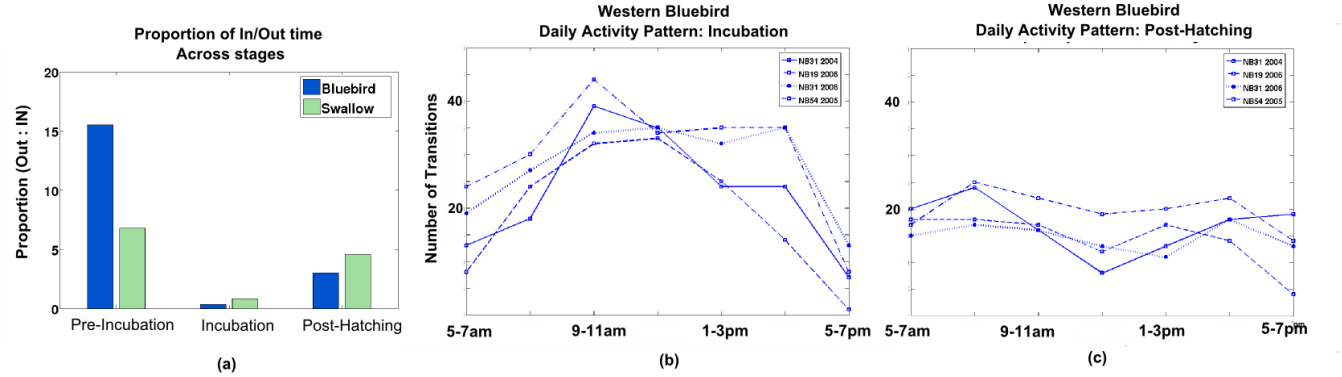

Figure 17. Various aggregate statistics used by biologist when investigating avian nesting behavior. (a) Proportion of times the bird is out of the box versus in (b) & (c) daytime activity levels for incubating and post-hatching phases, respectively.

distances from the margin for classification of an image, the transition error decrease. The specific error rates are shown in Figure 15.

3.5 Putting it Altogether

The continuous measurement and automatic inference over the entire duration of the nesting season can provide invaluable insight into avian nesting behavior. First a visualization of the data in time-series format is presented to demonstrate it's usefulness to extract "macroscopic" patterns from "microscopic" *in situ* measurements. Next some aggregate statistics are derived to quantify these observed patterns along with biological interpretation. Herein, we focus on the data implications of our vision system to show what is possible rather not an in-depth statistical analysis.

Typical avian nesting cycle follow a deterministic series of stages (before, egg laying, incubation, hatching, post-hatching) that starts with the selection of a nesting site and ends with the fledging of young (Figure 1 and 16). These stages have well defined boundaries that we exploit in our system. 'Egg laying' begins when the first egg is laid. Once the final egg is laid, 'incubation' begins. 'Hatching' typically occurs a few weeks after the onset of incubation and usually lasts one or more days . In the early parts of 'post-hatching', the mother will spend the night brooding to keep the hatchlings warm until they develop feathers to retain heat. Finally, fledging occurs when the young are mature enough to leave the nest.

Figure 16 is a visualization of a full seasons worth of avian nesting activity. Here the predicted presence/absence from one of the nests has been smoothed to provide temporal consistency - the higher the curve the greater proportion of time the bird has been observed. Images from various stages of the nesting season have been juxtaposed with the smoothed trace to provide context. In this time-series representation of the images behavioral patterns start to emerge. A prominent feature in the trace is the diurnal cycles corresponding to the night and day time presence of the bird in the box. Before and during egg laying the bird is either al-

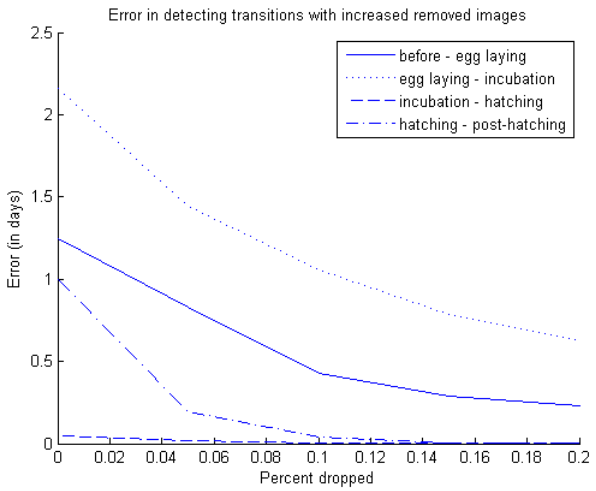

Figure 15. The average error in delimiting nest box stages using a Hidden Markov Model with increasing number of dropped images. Images are dropped according to the distance from the decision boundary as described in Section 3.3.2.

most never present or only present during the night. With the laying of the last egg and onset of incubation, the presence of the bird inside the box dramatically increases, almost equaling that of night time. This intensive period of incubation is followed by a gradual decrease in the presence of the bird after hatching. In this post-hatching phase the mother spends the night brooding to keep the hatchlings warm until they develop feathers to retain heat after which she only visits the box to feed them. Because of how ease with which wireless systems can be deployed (relative to wired systems), this type of analysis can be extended to multiple nest boxes for inter-species and intra-species behavioral patterns, as illustrated in the Appendix.

Besides being able to explore behavioral patterns on a scale not possible before, aggregate statistics from the inferences lend quantifiable measures of these trends. Consider Figure 17 where the presence/absence data from the deployment has been used to derive the proportion of times the bird was absent as compared to being in the box. The proportions were calculated for daytime hours and have been averaged for each species per nesting stage. These proportions can determine parental nest attentiveness across stages and its affects on reproductive success. Early in the season before eggs have been laid, the bird spends most of its time outside and only occupies the box in short spurts to construct a nest. During incubation we see that the mother spends more time inside the box incubating the eggs ("on bout") than outside ("off bout"). The proportion to which the mother is on versus off are of great interest to biologist since it determines the mother's energy investment in terms of how long she is willing to incubate versus other activities such foraging for food, which primarily take place outside the box. The general trends in these proportions are what a biologist would expect since during incubation the mother invests a large proportion of her time incubating, as apposed to pre-incubation and post-hatching during which more time is spent outside the box on activities such as foraging and socializing.

Figure 17 has used the same data but to derive activity levels throughout the day. Once again since most birds are inactive during the night, the observations here are geared towards daylight hours, specifically 5am-7pm. The measure of activity graphed here is the total number of in/out transitions of the bird during a particular nesting stage over 2 hour periods. In (b) there is a noticeable peak formed in the late morning hours, indicating that during those hours of the day the bird is more active, most likely foraging for food and frequently visiting the box during the peak temperature hours. This is expected since birds are generally more active during the morning hours and less so in the afternoon and late evening. In (c) we see a flatter activity curve, suggesting that the visits the mother makes to feed the young are fairly evenly distributed throughout the day. Unfortunately, this is not entirely the case since during the post-hatching stage, visits to the box are short lived (less than 2 minutes), and hence the system is greatly under-sampling. As the last example demonstrates, the deployed system due to under-sampling effects is not completely reliable. Nevertheless, our analysis is encouraging in the face of these downfalls since the observed global patterns and trends in the traces and aggregate statistics is in agreement with current understanding of avian behavior [2], [5].

Another major benefit of these time-series traces is that it allows comparisons with other classes of temporally evolving processes. Micro and macro-climate conditions that have been sampled for similarly long periods of time, can be compared alongside the inferred image statistics to find correlations. Also new biological questions can be formulated that take advantage of these continuous, long-lived observations.

4 Discussion

The discussion here focuses on the opportunities available to an *in situ* imaging systems such as we described in this paper. Data analysis has shown that more efficient allocation of resources (power) is possible. In particular, we discuss the resolution with which the images are captured and subsequently transmitted and the frame rate used. Additionally, we discuss the feasibility of pushing computation to the device.

More Efficient Allocation of Resources

Sampling Rate: In some ways, the current sampling rate (every 15 minutes) is unnecessarily high to correctly delimit nest box stages, which typically span multiple days. Reducing the sampling rate could free up some of the resources of the imaging sensor for additional processing or increase the longevity of the system. However, because of the noise in the egg counter, as shown in Figure 13, the performance of the HMM will inevitably become more sensitive to the errors as the sampling rate is decreased. As shown in Figure 18, in general, the fluctuations in error increase as sampling rate decreases. Each sampling interval was tested 100 times with random phase sifts, and the average is presented here. In fact, sampling intervals of 3 hours or greater have a non-negligable probability of having no observations during incubation and hatching. Because of high bird presence to absence ratio during this time, it is difficult to capture an image of the eggs, when sampling at that low rate. Depending

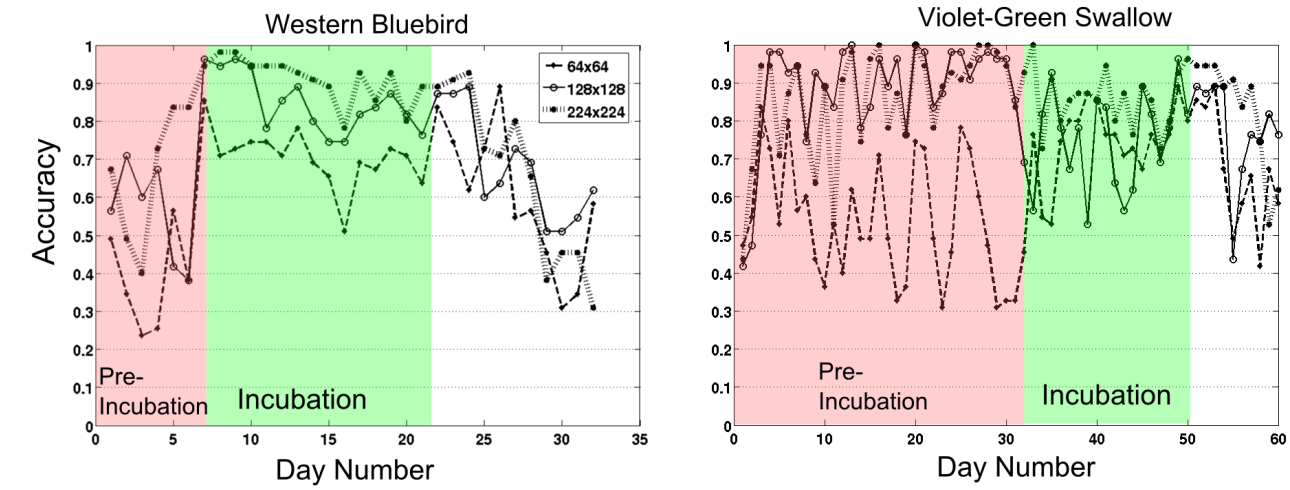

Figure 19. Accuracy per day over time (days) for different resolution images of NB31 during 2004 season (left) and NB8 during 2006 (right)

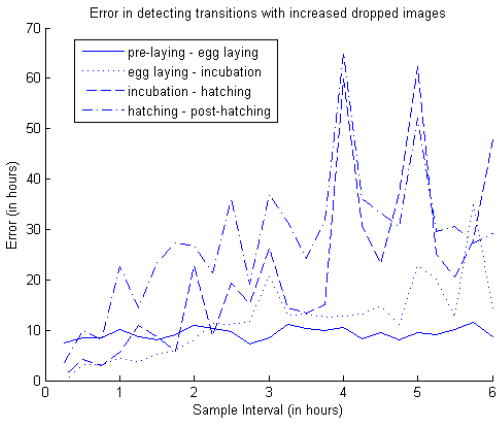

Figure 18. Accuracy of stage determination when the nest box is sampled at lower frequencies.

on the acceptable error range of the particular application, it is possible to vary the minimum frame rate during different stages of the nesting cycle to better capture the transitions.

Figure 18 also highlight another aspect of the sampling rate. Errors tend to peak at sampling interval that fall on the hour. This is particularly noticeable when the sampling interval is high. Taking a closer look at the resulting sample sequence, it appears that the bird has a very consistent habit of appearing and disappearing in the nest to the hour. This results in large variations in sampling points when sampling at hour intervals. Careful consideration on sampling rate adjustments need to capture the underlying process of bird presence and absence.

To properly measure the occupancy of the box a faster frame rate than every 15min is necessary, as mentioned in Section 3.4. However, the diurnal patterns seen in Figure 16 indicate that lower frame rates would be sufficient during the night since the bird is either always there or not. Therefore, besides "macro" events such as nesting stages, "micro" events such as day-night transitions can trigger frame rate changes. One can imagine that by re-allocating the extra samples that were taken at night to sample more during the day, the node could last just as long while more accurately sampling the phenomenon.

Resolution: A closer analysis of our bird detection algorithm versus resolution suggests that a higher-resolution image is not always necessary nor directly translate to higher-accuracy. Tests were performed for each nest box at three different resolutions: 64x64, 120x120, 225x225. Figure 19 are results from two of the boxes. The graphs show accuracy of the algorithm per day over the course of the season. In the graphs there are times when all three resolutions perform equally well, as well as times when one or two of them outperform the rest. For nest box 31 of 2004 (left side of Figure 19), early on in the season and during incubation both 120x120 and 225x225 images perform equally well and outperform the lowest resolution (64x64). However in the post-hatching stage, all three algorithms provide the same level of accuracy, most likely due to the growing young in the box that render higher-resolutions ineffective.

Further analysis of the accuracy results suggest that the type of species can play a role in how the algorithm performs over time. For the Violet-Green Swallow box (right side of Figure 19), the 120x120 and 225x225 images perform equally well during the early parts of the season as was found for Bluebird box, but all three resolutions perform equally well during incubation. This goes back to the fact that swallows build simpler nests and during incubation when very little changes occur all three resolutions provide good separation between the 'bird' and 'no bird' images. For these reasons a combination of nesting stages and species type could be appropriate cues for when different resolutions are appropriate.

Local Processing

Both bird and egg detection algorithms mentioned in this paper are based on gradient operations and convolutions. Since these are pixel level operations, a programmable logic such as a CPLD or FPGA can be used to perform them in parallel at low energy cost. In addition, application specific

optimizations can be applied to the algorithms to boost speed and reduce power consumption. A case in point being the egg counting algorithm that relies on the detection of interest points (potential sites of eggs) by searching across a large set of possible scales. However, the nest box setup where imagers are placed a fixed distance from the object of interest places a constraint on the distance of eggs. This allows the SIFT algorithm to reduce its search space to a much smaller range of scales in which the egg can be found. Spatial consistency can further reduce the search space, since eggs are typically grouped together in in a small region of the nest. Upon detecting the first egg the algorithm can focus on areas of the image in the neighborhood of previously detected egg points. By performing these algorithms, we are better

5 Related Work

There is a growing number of deployment research that explore using a network of wireless sensors to study environmental phenomena. In particular, previous deployments at James Reserve [19, 3] , Great Duck Island [20] and Redwood forest [21] exploit a network of wireless sensors to monitor a range of environmental parameters such as temperature or humidity at very low rates. These research have made valuable contribution in applying sensor network technology to study environmental phenomena. They propose a system architecture with similar components including *in situ* sensors in the field, micro-servers acting as bridges between low rate wireless links and high-speed wired infrastructure, and a database archival unit. They also provide a thorough analysis of the performance of the wireless system including power consumption and lifetime, network connectivity and data yield, and the application interpretation of the data. For instance, [21] illustrates a through analysis of temporal and spatial characteristic of the environmental data that is collected by 40 nodes a in 44 days life of redwood tree.

A second class of deployment research involves monitoring relatively high data rate phenomena with two important applications: seismic activity and structural monitoring. For instance, recent work by [13] and [22] study application of a wireless network of sensors to monitor the health of a structure such as buildings or bridges and to monitor seismic activity of an active volcano. Since in these applications there are additional stress on the resources of a network, most of the emphasis is on investigating the performance of the system such as reliability and robustness in terms of data delivery, accurate timing of the signals, and data transfer performance. From the data collection and analysis side, both systems look at techniques such as event-based triggers to reduce the quantity of data quality.

Several factors distinguishes our work from previous deployment research which we discuss in this paper. First, we exploit vision in a long-lived network and provide unique insights into the challenges of deploying imagers *in situ*. Unlike environmental or seismic sensors, different deployment settings of the image sensors may have significant impact on quality of acquired images and ultimately on the performance of the automated data analysis system. Second, unlike other sensing modalities that readily provide interpretable data, images require an integrated approach for image processing and data analysisThird, we discuss performance properties of our image processing routines under image *resolution* and *frame rate*, important factors influencing power consumption in a vision system. Additionally , our collected data suggests that "micro" and "macro"-level events could act as triggers to adjust these system parameters to more accurately sample the environment and efficiently allocate resources.

There are several techniques used to detect and/or identify objects. The most popular algorithms do not rely on heuristics and use responses of generic filters to characterize objects. The hope is that the filter responses are unique to the object and therefore a statistical model can be learned to separate this object from all others. Examples of these algorithms include SIFT [10], Viola-Jones's Cascaded AdaBoost [24], and Mutch-Lowe [11]. On the other end of the spectrum are algorithms which rely on assumptions on the object and the non-object instances in the data set we would like to analyze. Assumptions on object size and intensity difference to background allow for the use of background differencing or blob detection.

Some vision architecture have been proposed [7, 12], that focus primarily on the collection mechanisms of the visual data. Our end-to-end system, besides collecting images its primary focus is on the vision based inference and the implications that has on the system.

6 Conclusion

Automated analysis makes useful the large scale continuous sampling of environmental phenomenon using imagers. We demonstrated the automation of vision inference as it applies to avian breeding behavior. We leveraged our end-to-end system to impose constraints in the way that images were collected so that the subsequent vision analysis was feasible. In particular, fixed location and orientation of the imager were used to select appropriate vision algorithms and temporal consistency in the image sequences exploited to reduce inference error. This frees the biologist from having to visit nesting locations to gather statistics about avian breeding behavior, a tremendously time consuming task.

More generally we have demonstrated how images can be used as biological sensors and have developed techniques and a systems approach that should be applicable to other biological phenomena in agriculture, ecology, and perhaps even human spaces. Future work in progress includes improving the camera module on our wireless nodes, creating a data logger version for locations that are out of continuous wireless reach, and extending the camera module so that it can be placed off the board itself.

7 References

[1] Aghajan, H., I. Downes, and L. Baghaei, "Development of a Mote for Wireless Image Sensor Networks," *COGIS*, 2006.

[2] C.R. Brown, A.M. Knott, and E.J. Damrose. 1992. "Violet-green Swallow." In the *Birds of North America*, No.14 (A. Poole, P. Stettenheim, and F.Gill, Eds.). The Birds of North America, Inc., Philadelphia, PA.

[3] A. Cerpa , J. Elson , M. Hamilton , J. Zhao , D. Estrin , L. Girod, "Habitat monitoring: application driver for wireless communications technology." In *Workshop on Data communication in Latin America and the Caribbean*, April 2001.

[4] Y. Freund, and R.E. Schapire,"A decision-theoretic generalization of on-line learning and an application to boosting,," *Journal of Computer and System Sciences*, 55(1):119–139, August 1997.

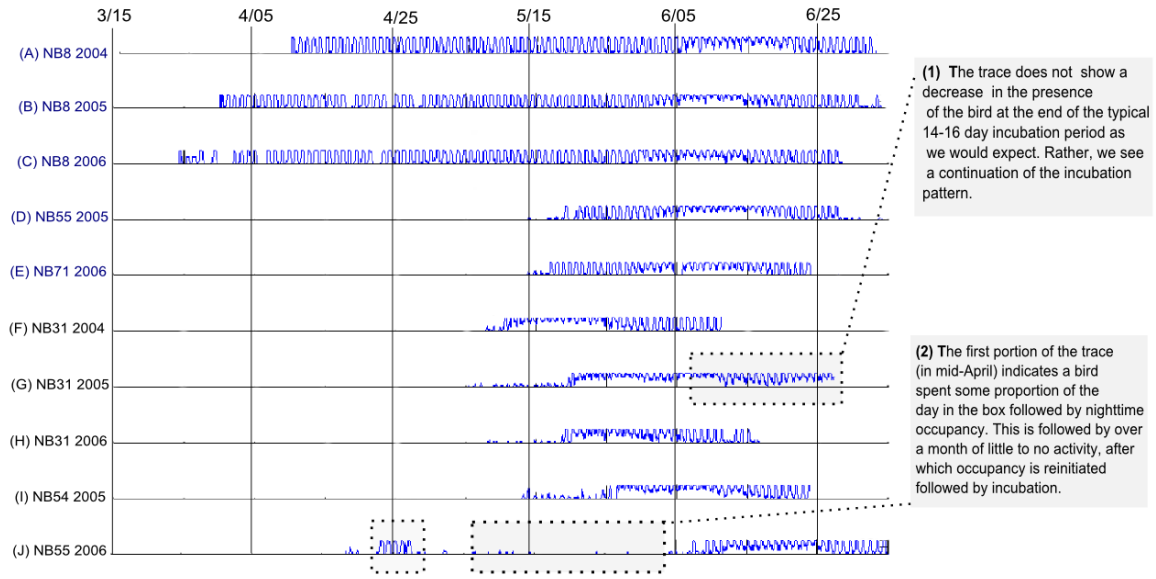

Figure 20. Traces from all ten seasons from different nest box locations. The first 5 nest boxes are of Violet-Green Swallows and the bottom 5 are of Western Bluebirds.

[5] J.A. Guinan, P.A. Gowaty, and E.K. Eltzroth. 2000. "Western Bluebird." In the *Birds of North America*, No.510 (A.Poole and F.Gille, eds). The Birds of North America, Inc., Philadelphia, PA.

[6] C. Harris and M. Stephens. "A combined corner and edge detector." Proceedings of the *4th Alvey Vision Conference*: pages 147–151, 1988.

[7] P. Kulkarni, D. Ganesan, P. Shenoy, Q. Lu, "SensEye: A Multi-tier Camera Sensor Network." In *Proceedings of ACM Multimedia*, 2005.

[8] Lin and Chang, "LIBSVM:A Library for Support Vector Machines," http://www.csie.ntu.edu.tw/ cjlin/libsvm/.

[9] D.G. Lowe, "Distinctive image features from scale-invariant keypoints," International Journal of Computer Vision, 60, 2, pp. 91-110, 2004.

[10] David G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, 60, 2 (2004), pp. 91-110.

[11] Mutch, J. and D.G. Lowe, "Multiclass Object Recognition with Sparse, Localized Features," In *Proceedings of Computer Vision and Pattern Recognition*, 2006.

[12] S. Nath, A. Deshpande, Y. Ke, P.B. Gibbons, B. Karp, S. Seshan, "IrisNet: An Architecture for Internet-scale Sensing Services." In *Proceedings of the 29th VLDB Conference*, 2003.

[13] J. Paek, O. Gnawali, K. Jang, D.N. Ramesh Govindan, J. Caffrey, M. Wahbeh, and S. Masri. "A Programmable Wireless Sensing System for Structural Monitoring." *4th World Conference on Structural Control and Monitoring (4WCSCM)*, July 2006.

[14] L.R. Rabiner, "Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," Proceedings of the *IEEE*, 77 (2), p. 257286, February 1989.

[15] M. Rahimi, R. Baer, O. Iroezi, J. Garcia, J. Warrior, D. Estrin, and M. Srivastava, "Cyclops: In Situ Image Sensing and Interpretation." *Third ACM Conference on Embedded Networked Sensor Systems*, 2005.

[16] A. Rowe, C. Rosenberg, and I. Nourbakhsh, "A Low Cost Embedded Color Vision System," In *Proceedings of IROS 2002*, 2002.

[17] S. Shoelkopf, C.J.C. Burges, A.J. Smola, "Advances in Kernel Methods," MIT Press, 1998.

[18] Stargate, http://staging.xbow.com/Products/productsdetails.aspx?sid=85.

[19] R. Szewczyk, E. Osterweil, J. Polastre, M. Hamilton, A. Mainwaring, D. Estrin, "Wireless sensor networks: Habitat monitoring with sensor networks." Communications of the ACM, June 2004.

[20] R. Szewczyk, J. Polastre, A. M. Mainwaring, and D. E. Culler, "Lessons from a Sensor Network Expedition." In *Proceedings of the First IEEE European Workshop on Wireless Sensor Networks and Applications*, 2004.

[21] G. Tolle, J. Polastre, R. Szewczyk, D. Culler, N. Turner, K. Tu, S. Burgess, T. Dawson, P. Buonadonna, D. Gay, and W. Hong. "A macroscope in the redwoods." In *Proceedings of the 3rd international Conference on Embedded Networked Sensor Systems*, SenSys '05.

[22] G. Werner-Allen, K. Lorincz, J. Johnson, J. Lees, and M. Welsh,"Fidelity and yield in a volcano monitoring sensor network." In *Proc. 7th USENIX Symposium on Operating Systems Design and Implementation (OSDI)*, November 2006.

[23] V. N. Vapnik, "The Nature of Statistical Learning Theory." Springer, 1995.

[24] P. Viola and M. Jones, "Robust Real-Time Face Detection," International Journal of Computer Vision, 2004.

Appendix A : Trace Study

Here we explore how the same traces described in Section 3.5 can be used to analyze multiple nest boxes. Here we look at a special case in which the observed general patterns can be used to detect "anomalies".In visualizing the time-series traces, these patterns can readily be picked-up as described by the gray callout boxes in Figure 20. By investigating the images, possible explanations can be found:

(1) After further scrutiny of the images, it was determined that only 1 of the 4 eggs in this nest hatched but that the female continued to incubate the remaining 3 eggs. This observation, although rare, is consistent with [5] which notes that the female may "incubate for weeks beyond normal incubation period" after a failed clutch. (2) The images indicated that another species, the Mountain Chickadee, had built a nest initially, followed by a Western Bluebird, a rare event that provides insight into inter-species interactions.