A Comparison of Background Subtraction Algorithms for Detecting Avian Nesting Events in Uncontrolled Outdoor Video

Kyle Goehner, Travis Desell
Department of Computer Science
University of North Dakota

Rebecca Eckroad, Leila Mohsenian, Paul Burr, Nicholas Caswell, Alicia Andes, and Susan Ellis-Felege
Department of Biology
University of North Dakota

September 1, 2015  eScience 2015, München, Germany
What is Wildlife@Home?

• A *citizen science* project that combines both crowd sourcing and volunteer computing.

• Users volunteer their brain power by observing videos and reporting observations.

• Users volunteer their computer power by downloading videos and performing.

• A scientific web portal to robustly analyze and compare results from users, experts and the computer vision techniques.
Between 2012 and now, Dr. Ellis-Felege has gathered over 85,000 hours of avian nesting video from the following species:

1. Sharp-tailed grouse (*Tympanuchus phasianellus*), an important game bird and wildlife health indicator species.
2. Piping plovers (*Charadrius melodus*), a federally listed threatened species.
3. Interior least terns (*Sternula antillarum*), a federally listed endangered species.

A recent collaboration with Ducks Unlimited added another 15,000 hours of Blue Winged Teal (*Anas discors*) nesting video.

We have also recently received over 2 million motion sensor camera images and ~100,000 aerial images taken by UAVs from a new Hudson Bay project.
The three species (Grouse, Plover and Tern) investigated in this work are ground nesting birds.

Sharp-tailed grouse nest in the dense grass (top left). Nests were monitored in areas of high oil development, moderate oil development and no oil development (protected state land).

Piping plover and interior least tern are shore nesting species (top right). Nests were monitored along the Missouri River in North Dakota.
Most grouse video is sleeping birds and grass blowing in the wind. But occasionally, interesting things happen.
Piping plover and tern video is more interesting, with active biparental involvement and less obscuring vegetation.
There are many challenges:

1. Dramatically changing weather conditions
2. Dawn/Day/Dusk/Night lighting conditions
3. Model species (sharp tailed grouse and piping plover) and some predators have cryptic coloration (camouflage).
4. Moving vegetation and insects can cause false negatives.
5. Lower quality video due to limitations on cameras.

Harnessing Citizen Science

Volunteer computing, where people volunteer their computers to different computing projects, has emerged as a viable and significant source of computing power being successfully used to perform research in scientific applications ranging from astronomy [5, 63, 84] to biology [7, 435, 45, 03], chemistry [443, 49], and physics [443, 0] to climate modeling [58] as well as many other fields of enquiry. Berkeley’s Open Infrastructure for Network Computing (BOINC) [9, 0] is the most widely deployed volunteer computing framework, in part due to its open source code and easy extension. As of April 5, over 793,333 volunteered computers are participating in BOINC and contributing over 91408 petaFLOPS ($10^{15}$ floating point operations per second of computing power) [49], more powerful than the world’s second fastest supercomputer [87, 49].

On the other hand, crowd sourcing, where people volunteer their brain power, has been successfully used by citizen science projects to tackle problems requiring human feedback. GalaxyZoo [9, 8] has had great success in using volunteers to classify galaxies in images from the Sloan Digital Sky Survey [6]; and PlanetHunters [96] has been used to identify planet candidates in the N’S’ Kepler public release data. However, these focus on volunteers doing identification and classification of images, not video.
We have been information about the video through a crowd sourcing interface, and a similar interface used by research assistants in biology.
Background Subtraction Methods
Mixture of Gaussians

MOG describes the probability of a pixel belonging to the background as a sum of Gaussians:

\[
f_X(X|\Phi) = \sum_{k=1}^{K} P(k) \cdot f_{X|k}(X|k, \theta_k)
\]

Where \(P(k)\) is the probability of the surface \(k\) appearing in the pixel view, and \(f_{X|k}\) is the Gaussian distribution for surface \(k\) with Phi being the set of theta input parameters for the Gaussian distributions describing each feature.

\(P(k), u_k, \text{and } \theta_k\) can be estimated with running averages calculated at each frame, and \(f_{X|k}\) can be estimated by a boolean value which is true for a pixel value if it is within 2.5 standard deviations of the mean.
Vibe stores the history of 20 previous pixel values, and compares new values to this pixel history.

If a pixel is within some threshold of any pixel within this stored model, it is classified as background.

The background model is updated stochastically, with each new pixel value having a $1/16$ chance to replace one of the 20 stored pixel values selected at random. If a replacement is done, there is an additional $1/16$ chance of also updating one randomly selected neighborhood pixel's previous values.
Pixel-Based Adaptive Segmentation (PBAS)

PBAS is an extended version of ViBe which adjusts the threshold for selecting a pixel as background dynamically.

This is done using another set of 20 values, however in this case these are the minimal decision distance (minimum distance between an updated pixel and the previous 20 pixels). The average of these 20 minimum decision values is used to calculate the threshold, $R(x_i)$, which increases/decreases by a user defined scale whenever it is above or below that average.
ViBe and PBAS were modified and compared to MOG for this work:

1. They were made 2nd frame ready - the initial 20 previous pixels were selected at random from the first image.

2. An open/close filter was added to reduce foreground detection noise. This essentially smoothes the image, aiding in the reduction of video artifacts.

3. The convex hull of any connected foreground features used as foreground mask. This increases the selected foreground area, as in many cases the head and other parts of the bird are foreground while the rest of the bird matches the background too well due to cryptic coloration.
Motion Detection for Avian Nesting Video

With these additions, the foreground mask needed to be converted to a measure of the probability of an event of interest occurring.

The count of foreground pixels is used as a time series of data points, which is smoothed by an exponential moving average:

\[ m_t = \alpha \cdot x_t + (1 - \alpha) \cdot m_{t-1} \]

Where \( m_t \) is the mean at time unit \( t \), \( x_t \) is the number of foreground pixels at time \( t \), and alpha is the learning rate.

If at any time, \( x_t \) is greater than the three standard deviations from the time series mean, \( m_t \), then that frame is flagged as having an event.
Motion Detection for Avian Nesting Video

This can then be used to generate time series which can be compared to crowd sourced and expert video observations:
Results
Experiments

MOG, as well as our modified ViBe and PBAS were run over 105 tern and plover videos (77.05 total hours), and 109 sharptailed grouse videos (205.39 total hours).

Video lengths range from 30 minutes to 2 hours, and each algorithm ran at \(~10\) frames per second.

Results were gathered using a Mac Pro with 12 logical cores, and took approximately 48 hours.
Detecting Interesting Events

The above shows two time series of detected foreground pixels. Red arrows at the bottom show the beginning and end of scientist observed events.

The algorithms were described as correctly detecting an event if it fell between the start and end time of a user observed event.
Detecting Interesting Events

### TABLE III
**Algorithm Accuracy vs Expert Scientists on Grouse Nests**

<table>
<thead>
<tr>
<th>Event Type</th>
<th>Event Count</th>
<th>MOG</th>
<th>ViBe</th>
<th>PBAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not In Video</td>
<td>284</td>
<td>274</td>
<td>258</td>
<td>270</td>
</tr>
<tr>
<td>Eggshell Removal</td>
<td>6</td>
<td>4</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>In Video</td>
<td>130</td>
<td>128</td>
<td>129</td>
<td>129</td>
</tr>
<tr>
<td>Predator</td>
<td>6</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Unspecified</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Attack</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Physical Inspection</td>
<td>60</td>
<td>52</td>
<td>56</td>
<td>56</td>
</tr>
<tr>
<td>Observation</td>
<td>44</td>
<td>41</td>
<td>39</td>
<td>41</td>
</tr>
<tr>
<td>On Nest</td>
<td>216</td>
<td>196</td>
<td>174</td>
<td>178</td>
</tr>
<tr>
<td>Off Nest</td>
<td>492</td>
<td>470</td>
<td>439</td>
<td>461</td>
</tr>
</tbody>
</table>

### TABLE I
**Algorithm Accuracy vs Expert Scientists on Tern and Plover Nests**

<table>
<thead>
<tr>
<th>Event Type</th>
<th>Event Count</th>
<th>MOG</th>
<th>ViBe</th>
<th>PBAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preen</td>
<td>180</td>
<td>170</td>
<td>138</td>
<td>147</td>
</tr>
<tr>
<td>Scratch</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Not In Video</td>
<td>732</td>
<td>632</td>
<td>578</td>
<td>607</td>
</tr>
<tr>
<td>Nest Exchange</td>
<td>22</td>
<td>16</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>Foraging</td>
<td>82</td>
<td>71</td>
<td>52</td>
<td>56</td>
</tr>
<tr>
<td>Adult-to-Adult Feed</td>
<td>20</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Nest Defense</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Predator</td>
<td>12</td>
<td>10</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>Non-Predator Animal</td>
<td>22</td>
<td>16</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>Unspecified</td>
<td>350</td>
<td>93</td>
<td>66</td>
<td>78</td>
</tr>
<tr>
<td>On Nest</td>
<td>932</td>
<td>665</td>
<td>582</td>
<td>608</td>
</tr>
<tr>
<td>Off Nest</td>
<td>2312</td>
<td>1960</td>
<td>1775</td>
<td>1876</td>
</tr>
</tbody>
</table>

The above charts show how well each algorithm matched up to events classified by project scientists (the paper also includes comparisons to our citizen scientists). All the algorithms performed well detecting events, with MOG detecting the most.
Detecting Interesting Events

TABLE V
Algorithm Accuracy with Consensus vs Expert Scientists on Tern and Plover Nests

<table>
<thead>
<tr>
<th>Event Type</th>
<th>Event Count</th>
<th>Any Alg</th>
<th>All Alg</th>
<th>MOG &amp; ViBe</th>
<th>MOG &amp; PBAS</th>
<th>ViBe &amp; PBAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preen</td>
<td>180</td>
<td>174</td>
<td>137</td>
<td>138</td>
<td>143</td>
<td>137</td>
</tr>
<tr>
<td>Scratch</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Not In Video</td>
<td>732</td>
<td>635</td>
<td>576</td>
<td>576</td>
<td>606</td>
<td>576</td>
</tr>
<tr>
<td>Nest Exchange</td>
<td>22</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>Foraging</td>
<td>82</td>
<td>73</td>
<td>51</td>
<td>52</td>
<td>54</td>
<td>51</td>
</tr>
<tr>
<td>Adult-to-Adult Feed</td>
<td>20</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Human</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Nest Defense</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Predator</td>
<td>12</td>
<td>11</td>
<td>6</td>
<td>6</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>Non-Predator Animal</td>
<td>22</td>
<td>19</td>
<td>12</td>
<td>12</td>
<td>14</td>
<td>13</td>
</tr>
<tr>
<td>Unspecified</td>
<td>350</td>
<td>94</td>
<td>66</td>
<td>66</td>
<td>77</td>
<td>66</td>
</tr>
<tr>
<td>On Nest</td>
<td>932</td>
<td>669</td>
<td>572</td>
<td>580</td>
<td>606</td>
<td>572</td>
</tr>
<tr>
<td>Off Nest</td>
<td>2312</td>
<td>1974</td>
<td>1763</td>
<td>1769</td>
<td>1868</td>
<td>1763</td>
</tr>
</tbody>
</table>

The above chart shows results for combining the different algorithms. Having a consensus from multiple algorithms tended to lower event detection.
Analysis of False Positives

An analysis of false positives was provided. A false positive was measured as the number of events classified during a user classified Not In Video event.

The grouse video, which has significant amounts of high wind and moving vegetation had far more false positives (as to be expected). On the other hand it also had a very high standard deviation - suggesting that for videos without high wind and moving vegetation the background subtraction performed well.

Plover and Tern video had significantly less false positives, however the standard deviation was high, suggesting that for some videos (high wind or light fluctuations) these algorithms performed poorly.

While MOG detected the most events, it also had significantly more false positives.
Effectiveness of Background Subtraction

The modified PBAS and ViBe both performed well in detecting events, while MOG had rates of false positives that were too high to be effective.

While PBAS and ViBe were highly effective for a large number of video, there still remains a challenging subset of video with high wind and/or frequent lighting changes which will require more advanced techniques.
We have used Wildlife@Home's volunteered computers to run the motion detection methods over all the collected video. Results have been incorporated as a timeline into the user interface. Users can click on the timeline to skip ahead to areas of interest.
What's Next?

The motion detection methods used, especially ViBe and PBAS work well on "easy" segments of the video.

New methods need to be developed to handle the challenging sections of video with rapidly changing light conditions and/or windy rapidly moving vegetation. Potential ideas: convolutional neural networks, Retinex to normalize brightness.

Expanding crowd sourcing to imagery from UAVs and motion sensing cameras taken in North Dakota and the Hudson Bay, Canada.
Reproducibility

All the videos and observations used in this work have been made available in the first Wildlife@Home data release:


And all Wildlife@Home source code is freely available on GitHub:

https://github.com/travisdesell/wildlife_at_home
Acknowledgements

Wildlife@Home is currently being supported by NSF award no. 1319700 through the Division of Intelligent Information Systems's Information Integration and Informatics (III) program.

Wildlife@Home has been generously supported by a collaborative research award and new faculty SEED grant from UND's Office of Research Development and Compliance. The project's video streaming server is hosted by UND's Computational Research Center and the volunteer computing server is hosted by UND's Scientific Computing Center. DNA@Home is under partial support from a Basic Sciences SEED Grant.

North Dakota Game and Fish has provided financial support for field logistics to collect sharp-tailed grouse videos.

The US Geological Survey has provided financial support for camera equipment, video storage, and field assistance to collect data for the piping plover and interior least tern.

And of course all our volunteers.
Thanks!

Questions?

http://people.cs.und.edu/~tdesell/

http://csgrid.org

tdesell@cs.und.edu