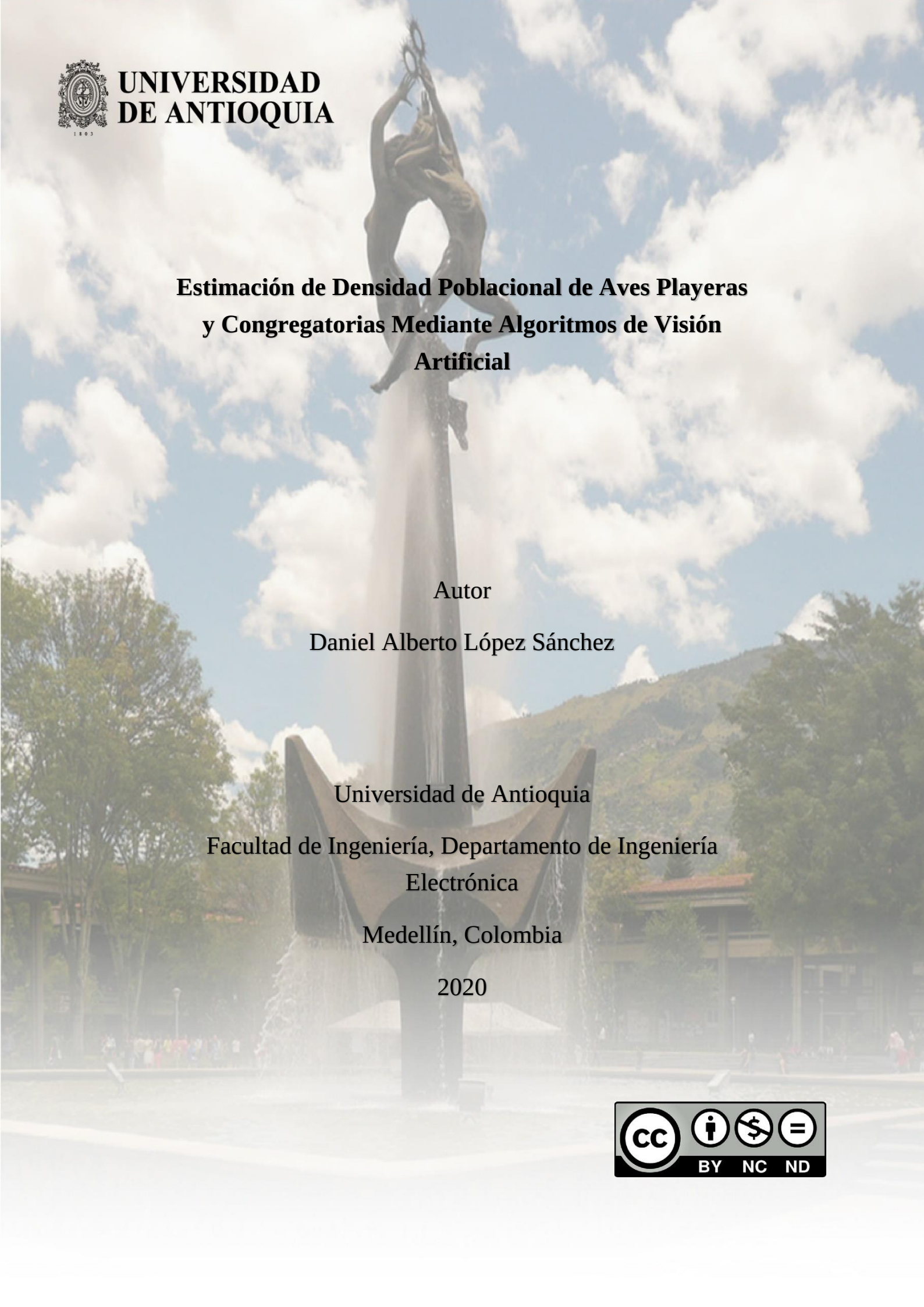

UNIVERSIDAD DE ANTIOQUIA

Estimación de Densidad Poblacional de Aves Playeras y Congregatorias Mediante Algoritmos de Visión Artificial

Autor

Daniel Alberto López Sánchez

Universidad de Antioquia

Facultad de Ingeniería, Departamento de Ingeniería Electrónica

Medellín, Colombia

2020

Estimación de Densidad Poblacional de Aves Playeras y Congregatorias Mediante Algoritmos de Visión Artificial

Daniel Alberto López Sánchez

Informe de proyecto de investigación como requisito para optar al título de:
Ingeniero Electrónico

Asesor

Claudia Victoria Isaza Narváez

Profesor Vinculado Universidad de Antioquia, PhD en sistemas automáticos

Co-Asesor:

Richard Johnston Gonzalez

Investigador Visitante - Instituto de Investigaciones Marinas y Costeras José Benito Vives de Andréis -INVEMAR- e Investigador Asociado - Asociación para el Estudio y conservación de las Aves acuáticas en Colombia-Caladris

Universidad de Antioquia
Facultad de Ingeniería, Departamento de Ingeniería Electrónica.
Medellín, Colombia
2020.

Resumen

En este proyecto de investigación se propone una metodología de bajo costo computacional, para la estimación de densidad de aves playeras y congregatorias en imágenes y vídeo, a través de algoritmos de visión artificial con un bajo costo computacional. El sistema se compone de tres metodologías: análisis de imagen, estabilización de vídeo y análisis de vídeo. En la metodología de análisis de imagen, se estima el número de aves a través de un algoritmo de regiones de interés, un clasificador basado en redes neuronales convolucionales, y la selección de un umbral para la detección. En la metodología de estabilización de vídeo, se propone una compensación de movimiento a través de la estimación, modelamiento y estabilización de movimiento en la grabación, para su posterior corrección. Por último, en la metodología de análisis de vídeo, se estima el número de aves a través de un algoritmo de substracción de fondo utilizando la mediana temporal de los frames que componen la grabación. Adicionalmente, se desarrolló una herramienta de software de código abierto y libre distribución, para estimar la densidad poblacional de aves playeras y congregatorias en imágenes y vídeo, usando las metodologías descritas previamente.

Palabras claves: Redes Neuronales Convolucionales, Region Proposals, Estabilización de vídeo, ornitólogos, Aves playeras.

Contenido

Lista de Figuras

Lista de Tablas

Introducción

El manejo de recursos naturales requiere del conocimiento de las condiciones en que se encuentran las poblaciones que constituyen los ecosistemas; de esta forma, es posible lograr la preservación de los acervos genéticos que son el banco fundamental de la biodiversidad [1]. Para estudiar la distribución de una especie o llevar cuenta de los cambios producidos en las poblaciones en un determinado periodo de tiempo, se han desarrollado diferentes métodos y técnicas, las cuales dependen, por ejemplo, de si se desea documentar la presencia de una especie o cuantificar su abundancia relativa [2]. Una de las técnicas que se destaca por su popularidad es el fototrampeo, en la cual se hace uso de cámaras trampa que son activadas por el movimiento de individuos o de acuerdo a una programación previa; de esta forma, es posible recolectar una gran cantidad de datos relacionados con la distribución de especies en una zona determinada [3].

Actualmente, el INVEMAR en trabajo conjunto con ornitólogos de la Universidad de Cornell y la Asociación Calidris, están tomando fotografías y realizando vídeos en Cunita, bocana de Santa Barbara Iscuandé-Nariño, la mayor área intermareal en Colombia, para obtener un mapa que permita medir el uso de hábitat por grupos de aves congregatorias y playeras. Con esta técnica, el volumen de datos obtenido es abundante, y los tiempos de análisis aumentan conforme crece el número de imágenes y grabaciones.

En las últimas décadas se han presentado avances en la automatización y conteo de varios grupos vertebrados e invertebrados, pero no se cuenta con una herramienta específica, que no requiera computadores de alto desempeño, para estimar la densidad de aves que se congregan. Con este proyecto de investigación, se desarrolló una herramienta de bajo costo computacional llamada TRINGA, la cual permite a los ornitólogos estimar de manera automática en un área de interés, el cambio de densidad de aves congregatorias y playeras en imágenes y vídeos. Esta propuesta permite agilizar los tiempos invertidos por el experto al analizar cada uno de los datos primarios de forma individual. Además de esto, se planteó un protocolo de grabación de vídeo que permite mejorar la calidad de las tomas, de tal manera que sean apropiadas para la identificación automática de las aves. El protocolo se presenta como anexo a este documento.

Objetivos

General

- Desarrollar una herramienta de bajo costo computacional, de código abierto y de libre distribución, que permita estimar la densidad poblacional de aves playeras y congregatorias en imágenes y vídeo.

Específicos

- Analizar algoritmos de tratamiento de imágenes, de bajo costo computacional, útiles para segmentar, en cada frame de un vídeo las zonas donde posiblemente hay presencia de aves.

- Desarrollar un segmentador que permita, de forma automática, extraer las zonas de las imágenes o en frames del vídeo, donde posiblemente hay presencia de aves.

- Desarrollar un algoritmo que permita cuantificar la densidad poblacional de aves en las imágenes o en frames de vídeo segmentados.

- Implementar un aplicativo de software que permita integrar mediante una interfaz gráfica, los módulos realizados para la estimación de densidad poblacional en tomas de grabaciones de vídeo

Marco Teórico

En esta sección se describen los conceptos y antecedentes que permitieron el desarrollo de las metodologías para la estimación y conteo de aves congregatorias y playeras mediante algoritmos de visión artificial.

Fototrampeo

El monitoreo del tamaño de las poblaciones de aves se ha convertido en una actividad usual de muchos ornitólogos e investigadores de la vida salvaje. Una de las técnicas usuales para el monitoreo de poblaciones de aves, se basa en el uso de cámaras trampa, localizadas en ciertas regiones de interés, en la cual se recolecta registros de imágenes o grabaciones de vídeo que dan cuenta de la actividad de las especies estudiadas, minimizando de esta manera el impacto de la presencia humana en el ambiente natural [4]. Sin embargo, hay una alta posibilidad de tener registros que no contienen información útil para los casos de estudio, presentándose tasas de hasta apenas un 1% de información útil (fotografías donde se presenten aves) en las sesiones llevadas a cabo [5]. Adicionalmente, el conteo manual de aves en fotografías puede ser bastante tedioso en registros donde el número de individuos es del orden de centenas, lo que incrementa considerablemente el tiempo de estudio empleado por los expertos.

Estabilización de vídeo por software

En algunas ocasiones los videos de cámaras de fototrampa pueden contener movimiento no deseado producto de las condiciones naturales del ambiente, por lo que se requiere de módulos de estabilización que realicen un preprocesamiento de la grabación para su posterior análisis. La estabilización de vídeo consiste en la eliminación del movimiento no deseado producido durante una sesión de grabación. Generalmente, el proceso consiste en tres fases: 1) Estimación del movimiento de la grabación, 2) Compensación del movimiento, y 3) Estabilización de vídeo [6]. En la primera fase, se calcula el movimiento sobre puntos característicos de interés en cada frame, obteniéndose una transformación afín que da cuenta de la trayectoria del vídeo. En la segunda fase, con ayuda de filtros suavizadores, es posible eliminar los cambios bruscos producidos en la trayectoria del vídeo, lográndose con esto una trayectoria más estable. Por último, en la tercera fase, una vez obtenida la nueva trayectoria, es posible aplicar la transformación afín a cada uno de los frames, obteniéndose de esta manera la estabilización del vídeo [7].

Regiones de Interés

Las imágenes producto de la recolección de fotografías de cámaras de fototrampa pueden contener diversos elementos que pueden ser o no de interés para el experto. La detección y localización de objetos de interés en el área de visión por computadora es una de las tareas que más retos impone en el campo. Actualmente, se han propuesto diversas técnicas que permiten localizar regiones de interés en una imagen. Una de ellas es el algoritmo *Selective Search* [8], el cual, a partir de un proceso de segmentación iterativo y asignación de regiones, consigue proponer sectores específicos de la imagen que pueden corresponder a un objeto determinado de interés. Sin embargo, la principal desventaja de esta técnica reside en su alto costo computacional, y en el alto número de regiones propuestas para una imagen [9]. Algunas propuestas alternas han buscado integrar un proceso de aprendizaje a partir de características de bajo y alto nivel de la imagen a partir de redes neuronales convolucionales, con las cuales es posible predecir la posición de los objetos de interés a la vez que se indica la clase y la probabilidad de confianza,

consiguiéndose mejoras hasta de un 10% en *mean Average Presicion* en comparación con el algoritmo *Selective Search* [10]. Otras propuestas se basan en la detección de regiones de interés a partir de algoritmos de substracción de fondo y detección de valores de intensidad atípicos.

- **Detección de puntos atípicos para substracción de fondo**

La detección de puntos atípicos en una señal permite dar cuenta de aquellos valores que se desvían de la distribución natural de la misma. Algunos algoritmos de substracción de fondo basan su análisis en la detección de niveles de intensidad atípicos en la imagen para detectar los objetos de interés. Otros, por el contrario, realizan un modelado del fondo a partir de técnicas de clustering como K-means, mezcla de gaussianas, kNN, entre otros; de esta manera, es posible substraer el fondo y extraer el objeto de interés, obteniéndose segmentaciones de objetos de interés altamente precisas [14-15].

- **Redes Neuronales Convolucionales (CNN)**

En algunas ocasiones los algoritmos de regiones de interés proponen sectores que pueden no ser de interés para el experto, por lo que se requiere de un clasificador que permita distinguir aquellos sectores que si lo son. En el área de aprendizaje profundo (*Deep Learning*), una de las redes más populares que ha tenido gran acogida en la última década, son las redes neuronales convolucionales, que, a partir de convoluciones entre matrices, muestreos, y funciones de activación no lineales, permiten generar características útiles para el proceso de detección, clasificación y localización de objetos en una imagen [11-12]. Entre las arquitecturas que destacan por su bajo costo computacional se encuentra las redes tipo MobileNets, las cuales, realizan el proceso de convolución en dos etapas, llamadas originalmente como Depth-Wise and Point-Wise Convolution, lográndose con esto, una disminución en el número de cálculos de hasta 9 veces menos que en las arquitecturas convencionales (AlexNet, GoogleNet, Inception V3) [13]. Para el proceso de entrenamiento y reconocimiento de patrones en la red, se requiere del uso de una base de datos fuente, que permita ajustar cada uno de los parámetros de la arquitectura a valores específicos, los cuales, determinaran la calidad de la detección, localización y/o clasificación de la red.

Materiales y Métodos

Base de datos

Para el desarrollo de los algoritmos de análisis de imágenes, estabilización de vídeo y análisis de vídeo, se contó con una base de datos de 1064 imágenes, de las cuales se construyó una base de datos de entrenamiento con 1734 segmentos correspondientes a aves, y 1768 sectores de fondo, tomados en 6 zonas diferentes de Iscuandée, Colombia. En los segmentos de aves se incluyeron diferentes especies que llegan a esa zona. Además de esto, se contó con 206 vídeos, de los cuales 100 presentaban movimiento, y 106 presentaban leve o nulo movimiento.

A continuación, se describirá la metodología de cada uno de los algoritmos desarrollados e implementados de análisis de imagen, estabilización de vídeo y análisis de vídeo.

Metodología algoritmo de análisis de imágenes:

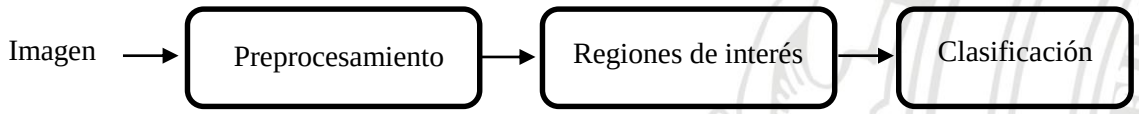

Figura 1. *Metodología propuesta para realizar el análisis de imágenes con el objetivo de identificar el número presente de aves en cada imagen*

El módulo de análisis de imágenes consta de 3 pasos fundamentales: preprocesamiento, extracción de regiones de interés y clasificación de las regiones de interés.

- **Preprocesamiento**

  Mediante el preprocesamiento es posible eliminar el ruido de alta frecuencia presente en la imagen. Para este módulo, se convirtió la imagen a escala de grises; se usó un filtro de media en una ventana de 5x5 en cascada con un filtro gaussiano, con una ventana de 5x5, y desviación estándar 0,6.

- **Regiones de interés**

  La selección de regiones de interés en la imagen se realizó a través de un proceso base de segmentación local de la imagen.

  - **Segmentación**

    La segmentación de la imagen se realizó a través de una ventana deslizante de 256x256, con un paso de 128. Cada uno de los sectores locales es procesado a través de la diferencia absoluta del sector con su media (se realizó durante un ciclo de 10 veces); posteriormente se hizo un proceso de umbralización de la imagen (con umbral mayor a 20 de valor de intensidad en escala de grises); se realizó un proceso de dilatación morfológica con un elemento estructurante circular de 7x7, y el resultado final, es

incluido en una imagen objetivo de las mismas dimensiones de la imagen original en la posición correspondiente del sector local.

- **Selección de regiones**
  Una vez segmentada la imagen, se procedió a binarizarla, para de esta manera detectar las coordenadas de la caja que rodea a cada objeto (*bounding box*), obteniéndose de esta manera las regiones de interés.

- **Clasificación**

Una vez obtenida las *bounding box* de cada sector de interés, se procedió a recortar cada uno de los sectores de interés de la imagen original en el espacio de colores RGB. Para la clasificación biclase (ave, no ave), se entrenaron 4 algoritmos de redes neuronales convolucionales basados en la arquitectura MobileNet, denominados como MobileNetV2, MobileNetV16, MobileNetV18, MobileNetV20, los cuales, presentaban diferencias en cuanto al número de épocas de entrenamiento, número de capas adicionales y adición de datos (a partir de inversión vertical y rotación de las imágenes originales de la base de datos). En la Tabla 1 se indican los valores usados para cada uno de los parámetros mencionados.

Tabla 1. *Parámetros de entrenamiento de clasificadores entrenados*

Modelo	Épocas	Capas extra	Adición de datos
MobileNetV2	20	2, 1024 neuronas cada una	No
MobileNetV16	20	2, 1000 neuronas cada una	No
MobileNetV18	30	2, 1000 neuronas cada una	No
MobileNetV20	20	2, 1000 neuronas cada una	Si

Posterior a esto, el conteo final del número de aves para una determinada fotografía es igual al número de sectores que fueron catalogados como aves. Para considerar que el segmento es ave, inicialmente se propuso que el umbral mínimo de certeza entregado por la red fuera de 0.5 pero en la implementación en el software TRINGA se dejó este umbral como decisión del usuario.

Metodología algoritmo de estabilización de vídeo

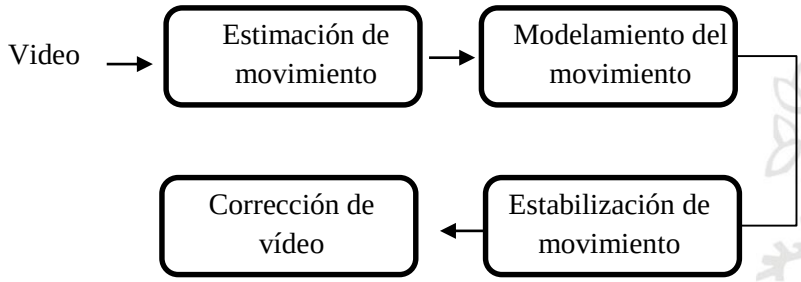

Figura 2. *Metodología propuesta para la estabilización de vídeo con el fin de eliminar el movimiento no deseado en las grabaciones.*

Para el desarrollo del módulo de estabilización de vídeo, se convirtió cada fotograma a escala de grises, realizando un preprocesamiento a través de un filtro gaussiano en una ventana de 5x5 con desviación estándar 0.6. El módulo de estabilización de vídeo se compone de 4 pasos fundamentales: Estimación de movimiento, Modelamiento del movimiento, Estabilización de movimiento y corrección de vídeo.

- **Estimación de movimiento**

A cada frame, se le estiman puntos característicos de interés. Posteriormente se calculó el flujo óptico para esos puntos con ayuda del frame subsiguiente.

- **Modelamiento del movimiento**

Con la información aportada por el flujo óptico, se procedió a estimar la transformación euclidiana que permite estimar el movimiento del frame actual con respecto al frame subsiguiente (el movimiento se descompuso en traslaciones a lo largo del eje x e y, y la rotación obtenida entre los frames). Posterior a esto, la trayectoria a lo largo de cada dirección, y el ángulo de rotación, fue calculada como la suma acumulada de los valores x e y, y de rotación, de cada uno de los frames con sus subsiguientes, obteniéndose de esta manera una curva en cada dirección y ángulo.

- **Estabilización de movimiento**

Para estabilizar el movimiento, se procedió a suavizar cada una de las curvas de movimiento a lo largo de cada dirección y ángulo

- **Corrección de vídeo**

Una vez estabilizada la trayectoria, y corregido los valores de movimiento de cada frame, se procede a estabilizar el vídeo a través de una transformación geométrica de los frames con su correspondiente valor de desplazamiento en cada dirección (x e y), y su ángulo de rotación corregidos (estabilizados). De esta manera, es posible volver a reproducir el vídeo de forma estabilizada.

Metodología de análisis de vídeo

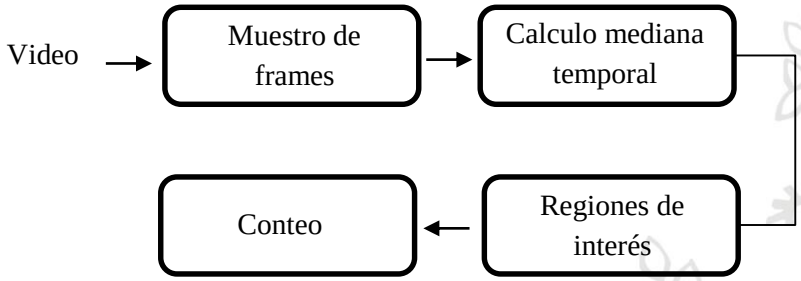

Figura 3. *Metodología propuesta para realizar el análisis de vídeo con el objetivo de identificar el número de aves presente.*

El módulo de análisis de vídeo consta de 4 pasos: Muestreo de frames, cálculo de la mediana temporal del vídeo, algoritmo de regiones de interés y conteo de aves.

- **Muestreo de frames**

  De cada vídeo, se extran 10 frames en escala de grises a una rata de cada 15 frames consecutivos. Posterior a esto, a cada uno de los frames muestreados, se les aplicó un preprocesamiento a través de un filtro de media en una ventana de 5x5.

- **Calculo mediana temporal**

  Una vez realizado el muestreo del vídeo y realizado el preprocesamiento, se procedió a calcular la mediana de las muestras (mediana temporal).

- **Regiones de interés**

  Para la selección de regiones de interés, se realizó una segmentación local de cada uno de los frames en una ventana de 200x200, con stride 100. La segmentación local, se realizó a partir de la resta de cada uno de los frames con la imagen producto del cálculo de la mediana temporal. Posterior a esto, se realizó un proceso de binarización a un nivel de intensidad mayor a 10, luego se procedió a realizar una dilatación con un elemento estructurante circular de 3x3, y se procedió a calcular las *bounding box* de las regiones resultantes en cada frame.

- **Conteo**

  El conteo final reportado corresponde a la mediana del número de bounding boxes de cada frame.

Desarrollo del software TRINGA.

Para el desarrollo de la interfaz gráfica del software TRINGA se usó el *framework* de desarrollo de aplicaciones web ***ElectronJS,*** el cual cuenta con librerías y paquetes que permiten ejecutar scripts del lenguaje de programación Python (el cual fue escogido para codificar los algoritmos propuestos). En la Figura 4 se presenta la ventana de inicio de la interfaz gráfica de usuario del software TRINGA.

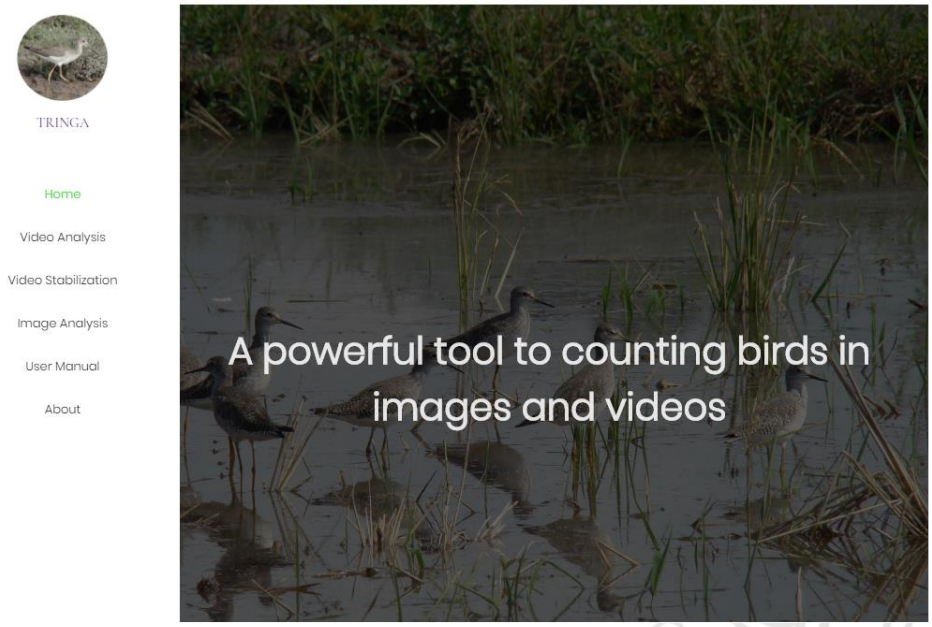

Figura 4. *Ventana de inicio de la interfaz de usuario TRINGA.*

A continuación, se describe cada uno de los módulos presentes en el software TRINGA.

- **Módulo de análisis de vídeo**

  El módulo de análisis de vídeo se compone de 3 pasos: 1) en el primer paso se indica dónde está el directorio de los vídeos que componen la sesión, 2) en el segundo paso se elige el área de interés específica donde operará el algoritmo de estabilización de vídeo, y 3) en el tercer paso, se ejecuta el algoritmo y se genera un archivo Excel con dos columnas, la primera con el nombre de cada uno de los vídeos de la sesión y la segunda con el conteo del número de aves para cada uno de los vídeos.

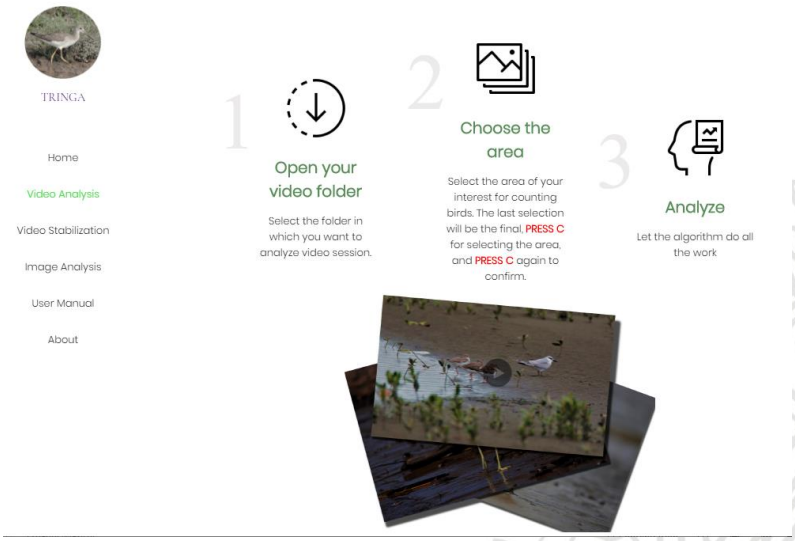

Figura 5. *Pasos a seguir para el análisis de vídeo en el aplicativo de software*

- **Módulo de estabilización de vídeo**

El módulo de estabilización de vídeo se compone de 3 pasos: 1) en el primer paso se índica dónde está el directorio de los vídeos de una sesión que se desea estabilizar, 2) en el segundo paso se indica el directorio donde se desean almacenar los vídeos estabilizados del directorio indicado, y 3) en el tercer paso se ejecuta el algoritmo de estabilización de vídeo.

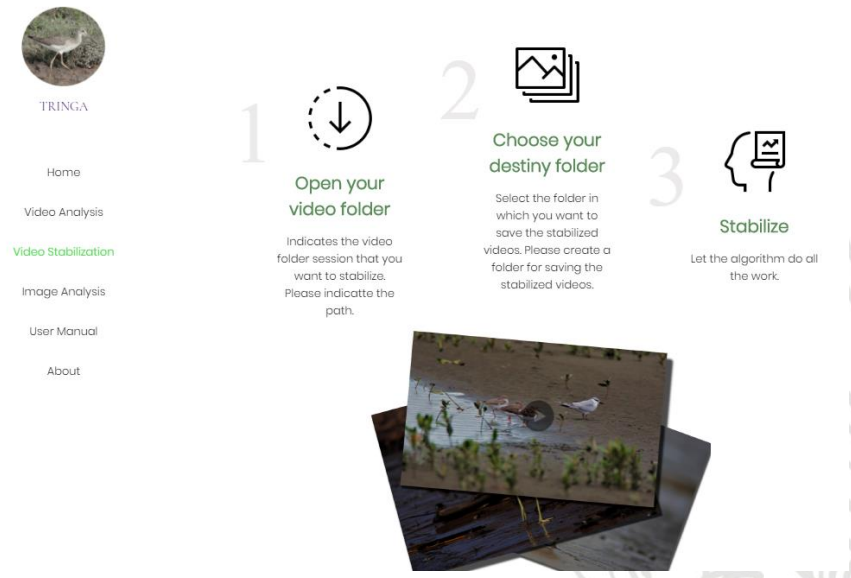

Figura 6. *Pasos a seguir para la estabilización de video en el aplicativo de software*

- **Módulo de análisis de Imagen**

el módulo de análisis de imagen se compone de 4 pasos: 1) en el primero se indica el directorio donde se encuentran las imágenes de una sesión, 2) en el segundo se indica el área de interés de la imagen sobre la cual operará el módulo de análisis de imagen, 3) en el tercero se ejecuta el módulo de análisis de imagen con el algoritmo de clasificación seleccionado (este se selecciona dando clic

en el incono de engranaje), y 4) en el cuarto, se elige un valor de umbral basado en la imagen con mayor detecciones, y se exporta un archivo Excel con dos columnas: una con el nombre de la fotografía de cada sesión y la otra con el conteo del número de aves predicho.

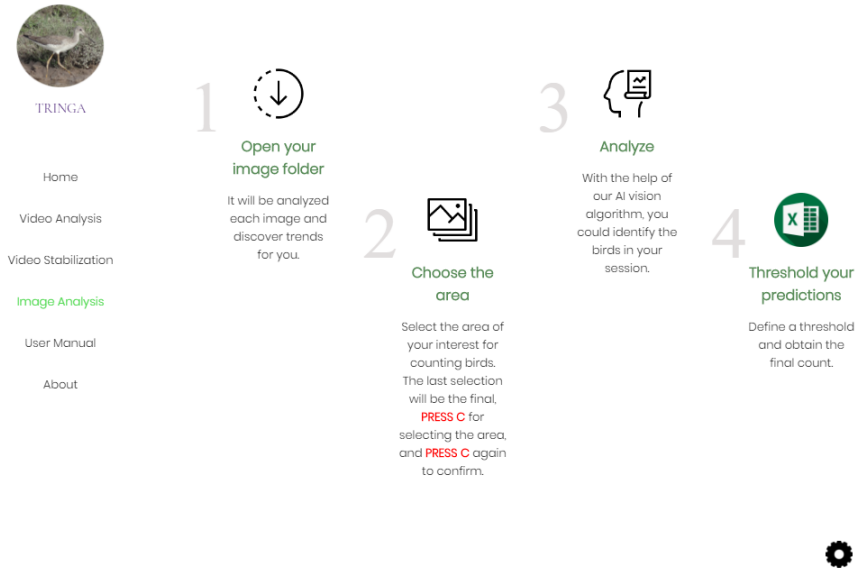

Figura 7. *Pasos a seguir para el análisis de imágenes en el aplicativo de software*

- **Manual de usuario**

En el manual de usuario se especifican los requerimientos mínimos de hardware para que el software se ejecute con éxito, al igual que los términos de uso, los modelos, agradecimientos, autores y algunas recomendaciones generales para la ejecución de los algoritmos. Se cuenta también con vídeos guía de los módulos de análisis de vídeo, estabilización de vídeo y análisis de imagen.

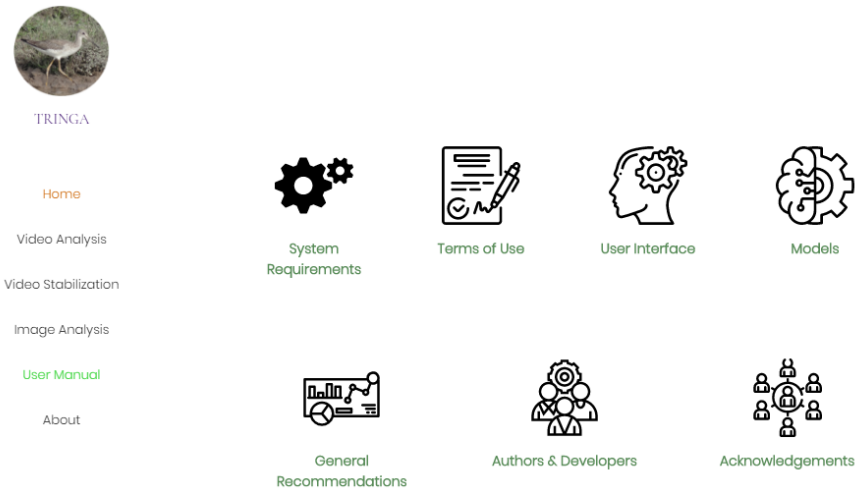

Figura 8. *Manual de usuario con sus diferentes opciones del aplicativo de software*

El software se encuentra disponible en el siguiente link:

https://drive.google.com/drive/folders/1VW-P5jtT3wiGiiSXJGLuZj3yp1ODSzM2?usp=sharing

Resultados y Análisis de Resultados

Con el fin de analizar el desempeño del software se realizaron pruebas con imágenes de la base de datos inicial y con nuevas sesiones de fotografía y videos de la zona de interés. A continuación, se describen las métricas empleadas para medir el desempeño y luego se presentan y analizan los resultados de estas pruebas.

Métricas de rendimiento

En el área de inteligencia artificial suelen encontrarse diferentes tipos de métricas que permiten caracterizar el rendimiento de un clasificador: error cuadrático medio (MSE), mediana del error absoluto (mAE), f1 score, precisión, sensibilidad, entre otras. Estas métricas pueden calcularse a partir de la salida esperada (Y) la salida real ($\hat{Y}$) y la matriz de confusión, en la cual se contabilizan cuatro condiciones de clasificación: Verdaderos Positivos (TP: aquellos sectores que se clasificaron como aves y son aves), Falsos Positivos (FP: aquellos sectores que se clasificaron como aves y no son aves), Verdaderos Negativos (TN: aquellos sectores que se clasificaron como no ave y no son aves), y Falsos Negativos (FN: aquellos sectores que se clasificaron como no aves y son aves).

$$MSE = \frac{1}{N}\sum_{i}\left(\hat{Y} - Y\right)^2$$

$$mAE = mediana(\hat{Y} - Y)$$

$$Sensibilidad = \frac{TP}{TP + FN}$$

$$Precisión = \frac{TP}{TP + FP}$$

$$F1\ Score = \frac{2 * Sensibilidad * Precisión}{Sensibilidad + Precisión}$$

Para el problema que se trató en este trabajo, la métrica MSE permite estimar de forma cuantitativa la cercanía entre el conteo predicho y el conteo esperado; la métrica mAE permite estimar el error en el conteo (conteo real $\pm$ error); la métrica sensibilidad permite estimar que tan bueno es el algoritmo para reconocer los sectores que son aves como aves; la métrica de precisión permite estimar que tan bueno es el algoritmo para discriminar entre ave y no ave; y por último, la métrica F1 score permite sintetizar los resultados de sensibilidad y precisión a partir de la media armónica de estos.

A continuación, se presentan los resultados de la implementación de las metodologías de análisis de imagen y vídeo, con sus respectivas métricas de rendimiento:

Análisis de imagen

1. Clasificadores:

En la Tabla 2 se presentan los resultados para las métricas de sensibilidad y precisión de cada uno de los clasificadores entrenados cuando se evalúa las métricas sobre el conjunto de entrenamiento, cada clasificador busca diferenciar si el segmento entregado es un ave o no. Puede observarse, como para el clasificador MobileNet V2 la precisión es de aproximadamente 97%, en contraste con los otros clasificadores, cuya precisión es mayor al 99%. Además de esto, puede observarse como la sensibilidad para el clasificador MobileNet V18 es del 97,35%, en contraste con los demás clasificadores, cuyo valor supera el 99%.

Tabla 2. *Resultados Clasificador Ave - No Ave*

Modelos	Sensibilidad	Precisión
MobileNet V2	99,94%	97,09%
MobileNet V16	99,07%	99,19%
MobileNet V18	97,35%	99,53%
MobileNet V20	99,71%	99,94%

2. Resultados algoritmo análisis de imagen:

El módulo de análisis de imagen se analizó bajo dos condiciones con imágenes de prueba:

1) Análisis sin umbral: en el software, se configuró previamente para cada uno de los clasificadores, que para que un sector sea clasificado como ave, debía superar una probabilidad de detección de ser ave mayor al 50%; de lo contrario, era considerado como no ave.
2) Análisis con umbral: en el software, el usuario tenía la posibilidad de incrementar el umbral de decisión de cada uno de los clasificadores (una vez aplicada la condición sin umbral), basado en la fotografía de la sesión de grabación de imágenes con más detecciones. De esta manera, aquellos sectores que estaban por debajo del umbral escogido por el usuario se consideraban como no aves, y de lo contrario, se consideraban como aves.

a. Error cuadrático medio y mediana del Error Absoluto sin umbral:

Para el análisis del error cuadrático medio, se analizó el conteo del número de aves en 6 sesiones (en estas sesiones se segmentaron sectores de aves y de fondo para construir la base de datos),

catalogadas como Muestra 1, Muestra 2, Muestra 3, Muestra 4, Muestra 5 y Muestra 6, con cada uno de los clasificadores entrenados, y se calculó el error cuadrático medio (MSE) y la mediana del error absoluto (mAE).

En las Figura 9 y 10 se puede observar el conteo esperado para cada una de las fotografías en cada una de las zonas de entrenamiento para cada clasificador, y los resultados de MSE (amplificado por un factor de 10) y mAE para cada uno de los clasificadores en cada una de las zonas de entrenamiento. En la Figura 9 para Muestra 1, 3, 4, 5, y 6 se destaca un número de aves mucho menor que el presente en Muestra 2 (línea continua en azul); además de esto, se observa para Muestra 1 y Muestra 4 que el número de aves presente en la sesión es casi nulo. En la Figura 10 para los clasificadores MobileNet V16 y V18, se observan valores de MSE y mAE bajos, lo que indica que son aquellos que mejor rendimiento presentan.

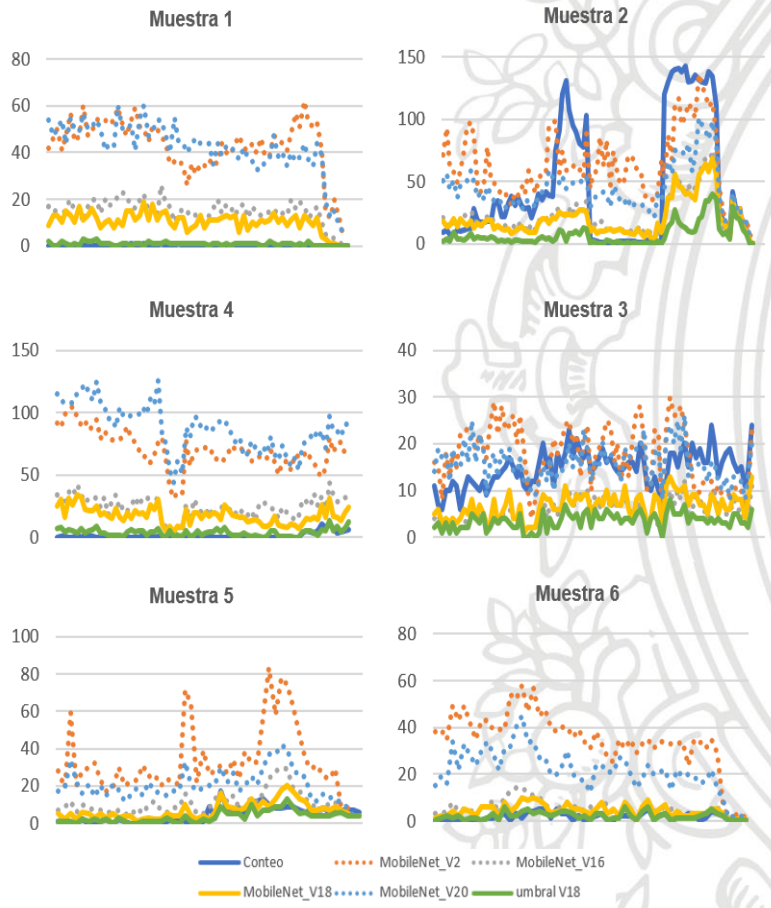

Figura 9. *Conteo esperado y conteo predicho para cada uno de los clasificadores sobre cada una de las zonas de entrenamiento. El eje vertical corresponde al número de aves presente en cada sesión, y en eje horizontal a cada una de las fotografías que componen la sesión*

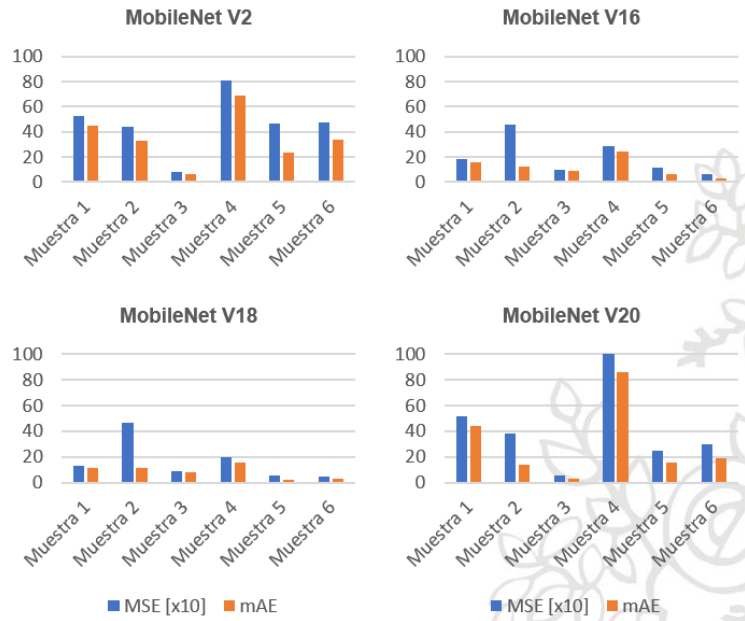

Figura 10. *Métricas de MSE (amplificado por un factor de 10) y mAE para cada clasificador en cada una de las sesiones analizadas.*

Muestra 1

Analizando la Figura 9 se identifica que en la mayor parte de las fotografías un conteo de aves predicho mayor a cero, presentándose especialmente para los clasificadores MobileNet V2 y V20 un alto índice de falsos positivos. Además de esto, se puede observar en el Figura 10, que los clasificadores con mejor desempeño fueron MobileNet V16 y V18, presentando valores de MSE de 1,86 y 1,32 respectivamente, y valores de mAE de 16 y 11 respectivamente.

En la Figura 11 puede apreciarse que los falsos positivos en esta zona para el clasificador MobileNet V18, se dieron en sectores donde existía presencia de hojas o sectores regulares que pueden mimetizar son similares a las posiciones que adoptan las aves en este tipo de entornos

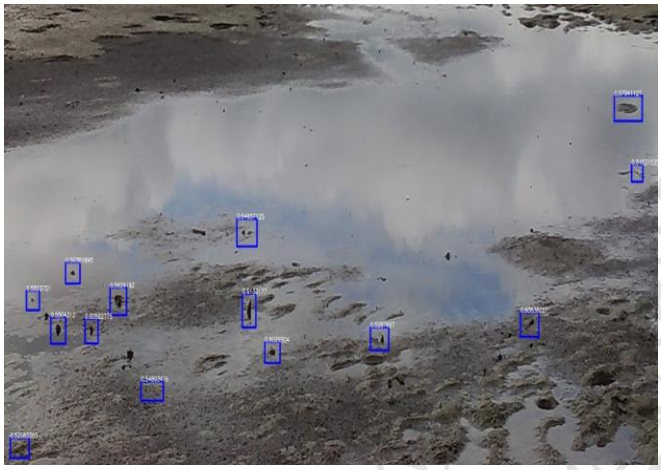

Figura 11. *Falsos positivos Muestra 1 utilizando el clasificador MobileNet V18.*

Muestra 2:

La Figura 9 corresponde a un ejemplo donde existe un gran número de aves (del orden de 100 o más), esta situación se presenta en diferentes fotografías que componen la sesión (Muestra2). Se puede observar como para los clasificadores MobileNet V2 y V20, que, aunque el conteo predicho en los picos es más acorde al conteo esperado, en fotografías donde no existe una presencia significativa de aves se presenta un gran número de falsos positivos. Por otro lado, los clasificadores MobileNet V16 y V18 presentan valores de conteo predicho similares entre sí, siendo estos más robustos en su conteo en zonas donde no se espera un número significativo de aves, pero presentando a su vez un error grande en los picos. Todos los clasificadores presentan un valor de MSE similar entré sí (del orden de 4); sin embargo, los valores de mínimos de mAE se presentan para los clasificadores MobileNet V16 y V18 (12,5 y 11 respectivamente). Esto se puede ver en la Figura 10.

En la Figura 12 puede apreciarse las detecciones para el clasificador MobileNet V18. En especial, se tiene que para zonas en donde las aves están más alejadas de la línea de vista de la cámara se presentan problemas en la detección, al igual que cuando estas se mimetizan en el entorno.

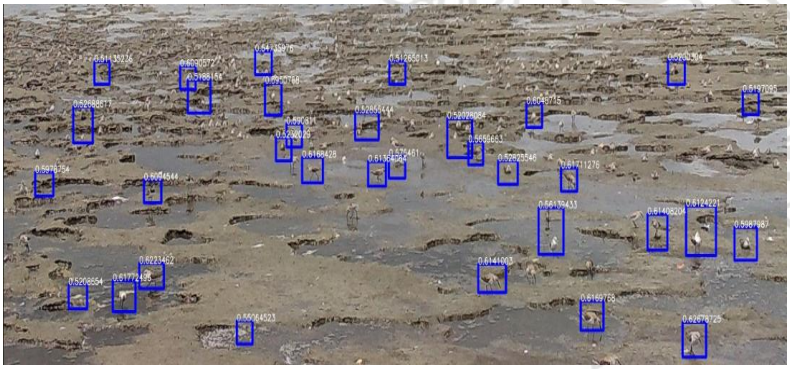

Figura 12. *Detecciones Muestra 2 utilizando el clasificador MobileNet V18.*

Muestra 3:

En la Figura 9 (muestra 3) se evidencia como los conteos predichos para los clasificadores MobileNet V2 y V20 son más cercanos al conteo esperado, en contraste con los clasificadores MobileNet V16 y V18. En la Figura 10 se identifica que los valores de mAE para cada uno de los clasificadores MobileNet V2, V16, V18 y V20 son de 6, 9, 8 y 3 respectivamente, siendo el más bajo el que se obtiene con el clasificador MobileNet V20.

En la Figura 13 se puede observar las detecciones realizadas para una fotografía de la sesión para el clasificador MobileNet V18. Se puede observar para este entorno un comportamiento similar al de la Ilustración 2, en donde se presenta congregaciones de aves lejos de la línea de vista de la cámara, mimetismo, y aves con un tamaño pequeño.

Figura 13. *Detecciones en Muestra 3 utilizando el clasificador MobileNet V18.*

Muestra 4:

En la Figura 9 se observa para esta sesión un comportamiento similar al de la sesión Muestra 1, donde la cantidad de aves presente en las fotografías es casi nula. Se observa, además, como los clasificadores MobileNet V2 y V20 presentan un alto índice de falsos positivos, y los clasificadores con mejor desempeño logran ser MobileNet V16 y V18. En la Figura 10, se tiene para los clasificadores MobileNet V16 y V18, valores de MSE de 3 y 2 aproximadamente, y valores de mAE de 24 y 16 respectivamente.

En la Figura 14 se puede ver que gran parte de los falsos positivos obtenidos se dan en zonas donde se presentan pantanos distribuidos a lo largo del escenario, y sombras con formas similares a las alas de aves.

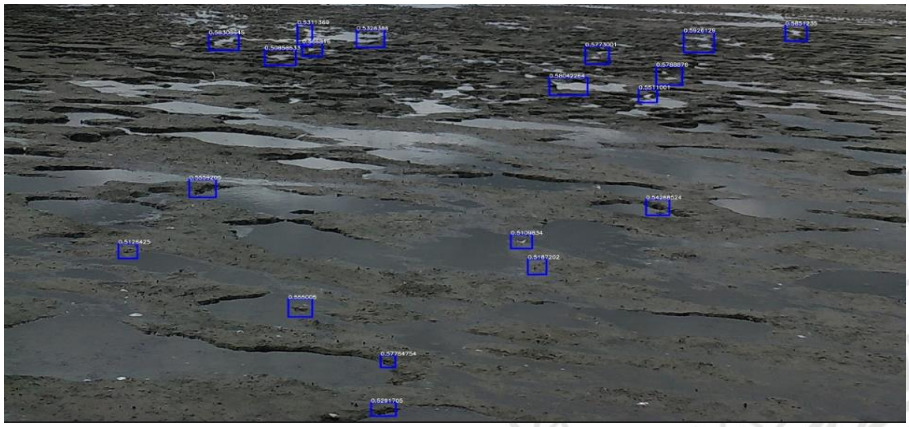

Figura 14. *Falsos positivos Muestra 4 utilizando el clasificador MobileNet V18.*

Muestra 5:

En la Figura 9 se destaca que para esta sesión el número de aves esperado no superaba un valor de 20. Además de esto, en la Figura 10, se puede observar que los clasificadores con menor índice de MSE y mAE son MobileNet V16 y V18, llegándose a obtener un valo de mAE de 2 para el clasificador MobileNet V18.

En la Figura 15 se puede observar un fenómeno interesante en la detección, en el cual, cuando las aves se congregan, se tienen detecciones en conjunto que son catalogadas como aves. Además de esto, se puede observar que fue posible detectar aquellas aves que se encontraban en vuelo.

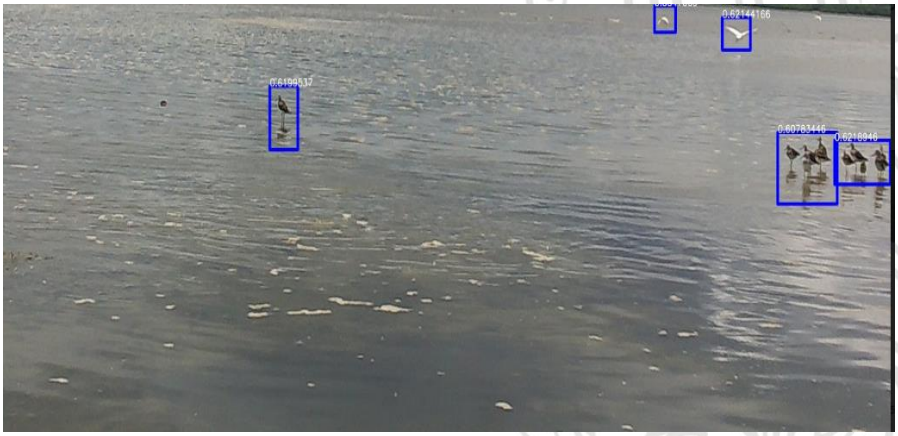

Figura 15. *Detecciones en Muestra 5 utilizando el clasificador MobileNet V18.*

Muestra 6:

En la Figura 9 se puede observar que para esta sesión el conteo esperado cualitativamente es cercano al conteo predicho por los clasificadores MobileNet V16 y V18. En la Figura 10, se destaca que los valores de MSE para y mAE para los clasificadores MobileNet V2 y V2 presentan valores relativamente altos en

comparación con los clasificadores MobileNet V16 y V18, cuyos valores de MSE son 0,65 y 0,46, y de mAE son de 3, respectivamente.
En la Figura 16 se puede observar las detecciones realizadas para una fotografía de la sesión para el clasificador MobileNet V18. Se puede observar como que se logra detectar con éxito aquellas aves que presentan un contraste con un entorno y tienen un tamaño definido. Sin embargo, existen aves que por su tamaño y mimetismo dificultan el reconocimiento en la imagen.

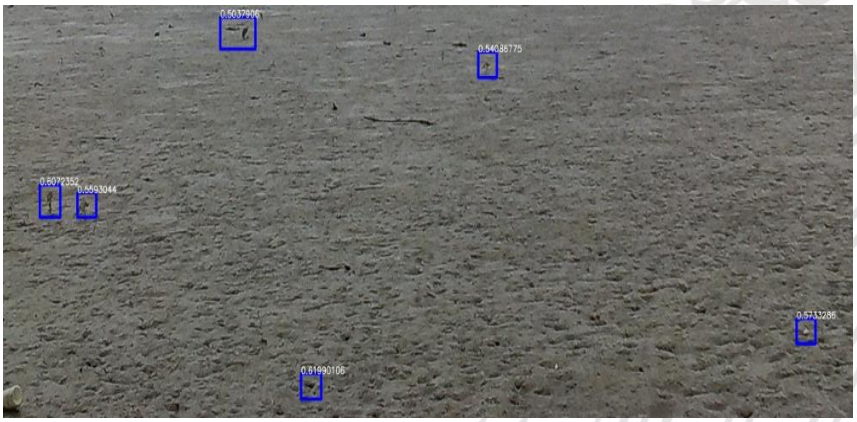

Figura 16. *Detecciones en Muestra 6 utilizando el clasificador MobileNet V18.*

En la Tabla 3 se muestra de forma sintética el promedio de las métricas de MSE y mAE para cada uno de los clasificadores. Se puede observar, que el clasificador que mejor desempeño obtuvo fue el MobileNet V18 (al MSE ser bajo, se tiene que el conteo esperado es cercano al conteo predicho para cada una de las fotografías, y al tener un mAE bajo, se tiene que el error de conteo es menor), presentando en promedio un error de 9 detecciones.

Tabla 3. *Métricas de rendimiento promedio para cada clasificador.*

Modelo	MSE	mAE
MobileNet V2	4,68	34,83
MobileNet V16	2,00	11,75
MobileNet V18	1,66	8,50
MobileNet V20	4,19	30,33

b. Error cuadrático medio y mediana del Error Absoluto con umbral:

Basados en los resultados anteriores, se observa que el software cuando se deja un nivel de certeza del 50% como umbral permite detectar las aves, pero encuentra un número importante de falsos positivos. Teniendo en cuenta que la salida del clasificador (red convolucional)

permite tener un nivel de certeza con el que se asegura si el segmento es ave o no, y que el contexto en cada zona puede cambiar, se incluyó la posibilidad de que el usuario seleccione el nivel de exigencia (umbral) para decidir si un segmento es ave o no (Figura 17). El usuario, a partir de una imagen que el algoritmo le presenta decide el umbral que quiere incluir. El software le muestra al usuario la imagen con mayor número de aves detectadas en la sesión que esté analizando (sesión: carpeta elegida por el usuario con múltiples fotografías) y sobre la imagen se adiciona el nivel de certeza que entrega el algoritmo para decir si un segmento es ave. Así el usuario puede, a través de este ejemplo, estimar el umbral más adecuado para la zona y objetivos propios del estudio.

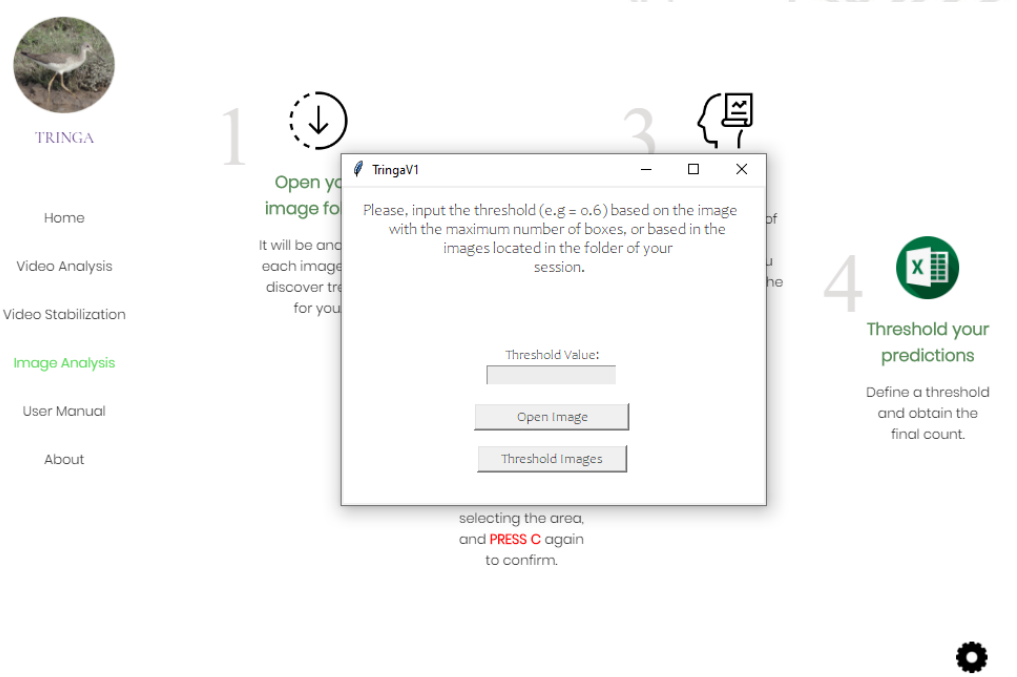

Figura 17. *Selección de umbral en el software TRINGA*

En la Figura 18 y 19, se observa como al implementar la opción de selección del umbral existe una disminución notable en el error de las sesiones Muestra 1, 4, 5 y 6 (aumentando el rendimiento del algoritmo). Por otro lado, se observa que existe un incremento del MSE y mAE (lo que ocasiona un decremento del rendimiento) para las sesiones Muestra 2 y 3 (en estas sesiones el algoritmo presenta un rendimiento limitado debido a los picos en el número de aves y la lejanía de estas de la línea de vista de la cámara), aumentando el mAE de 11 a 30 para Muestra 2.

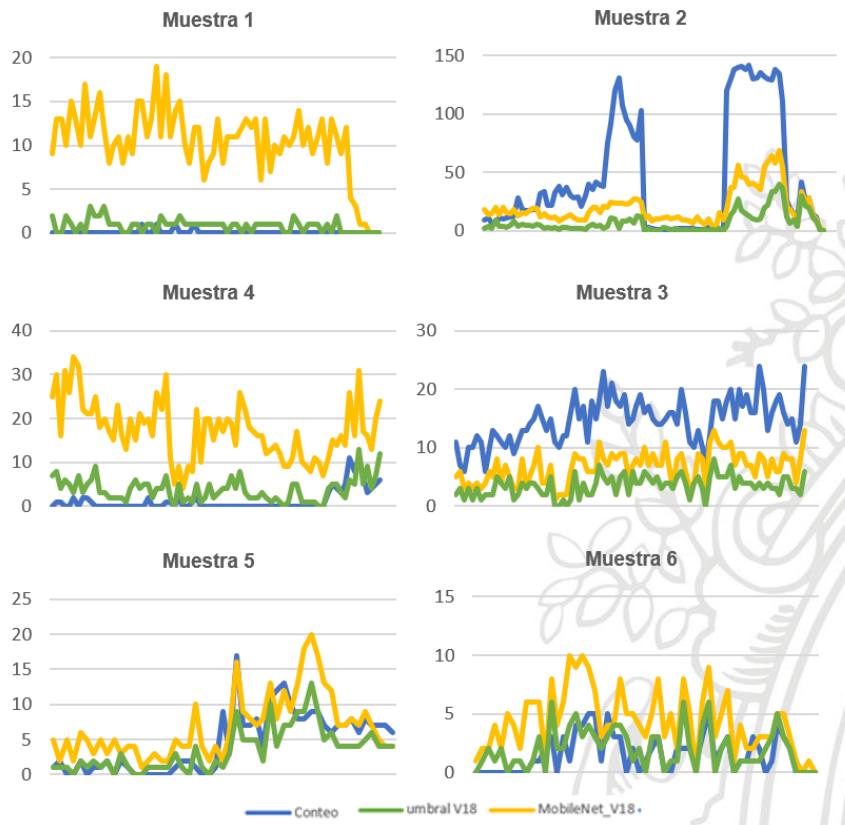

Figura 18. *Conteo esperado y conteo predicho para cada uno de los clasificadores sobre cada una de las zonas de entrenamiento para el clasificador MobileNet V18 con y sin umbral. El eje vertical corresponde al número de aves presente en cada sesión, y en eje horizontal a cada una de las fotografías que componen la sesión.*

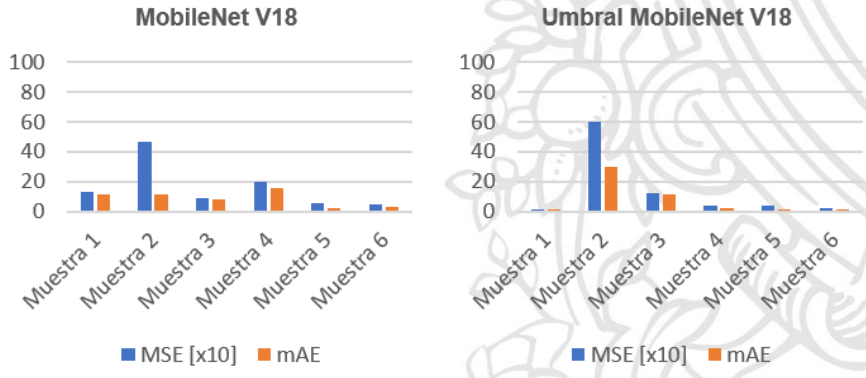

Figura 19. *Comparación de MSE y mAE con y sin umbral para el clasificador MobileNet V18.*

En la Tabla 4 se contrasta el promedio de las métricas de MSE y mAE para el clasificador MobileNet V18. Se observa cómo hay una disminución de 0,26 en el valor de MSE y de 0.83 en mAE al introducir el umbral, lo que incrementa el rendimiento del clasificador en la detección de aves en las fotografías de cada sesión.

Tabla 4. Métricas de rendimiento promedio para el clasificador V18 con y sin umbral.

Modelo	MSE	mAE
MobileNet V18	1,66	8,50
MobileNet V18 Umbral	1,40	7,67

c. Métricas de sensibilidad y precisión para cada uno de los clasificadores con y sin umbral:

Con el fin de evaluar el desempeño global del software y la influencia del umbral, se analizaron 19 sesiones adicionales no incluidas en la base de datos inicial, ubicadas en zonas con diferentes características. Para cada sesión se eligieron 4 umbrales denominados Um1, Um2, Um3 Um4 ordenados de menor a mayor, y se contrastó con el comportamiento del clasificador sin umbral (SU). Basados en los resultados de sensibilidad y precisión, se agruparon las zonas en 3 grupos: zonas de difícil conteo de aves, zonas de alta precisión en el conteo de aves, y zonas de buen rendimiento en el conteo de aves.

El agrupamiento por zonas se justifica debido a que existen zonas en donde no existe presencia de aves en las imágenes, y no es posible calcular con certeza la sensibilidad, zonas donde no existe presencia de aves y no se realizan detecciones falsas, y no es posible calcular con certeza la precisión, zonas donde se tiene un alto acierto en la detección de aves, y zonas donde se tiene un bajo acierto en la detección de aves.

Para aquellas fotografías donde no hay presencia de aves, y no se detectaba aves, se reporta un valor de sensibilidad 1, y para aquellas donde no había aves, y no se detectan falsos positivos como aves, se incluyó un valor de precisión 1. Esto se hizo teniendo en cuenta que las medidas de desempeño empleadas presentan valores indefinidos para situaciones donde no se espera identificar ningún ave.

Zonas de baja detección y bajo desempeño en el conteo:

En la Figura 20 se puede observar el promedio de las métricas de sensibilidad y de precisión de cada uno de los clasificadores sin umbral y con los diferentes umbrales para aquellas carpetas en donde la detección no se considera adecuada.

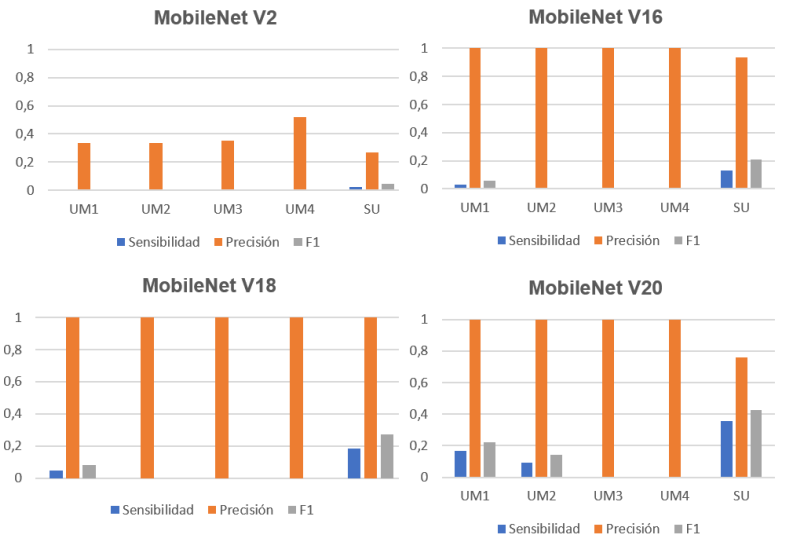

Figura 20. *Métricas de precisión y sensibilidad en zonas de baja detección.*

En la Figura 21 se puede observar aquellas zonas de baja detección y bajo desempeño. Estas zonas se caracterizan porque las aves están distribuidas a lo lejos de la línea de vista principal de la cámara, se mimetizan con su entorno, o no existe suficiente contraste, lo cual dificulta su identificación. Para este tipo de zonas, el segmentador MobileNet V20 es el que mejor desempeño presenta; sin embargo, es importante que exista un nivel de contraste adecuado en la zona y un ambiente uniforme, ya que, en ocasiones, la precisión del clasificador se ve afectada por falsos positivos (ej: pantano, rocas, agua, entre otros).

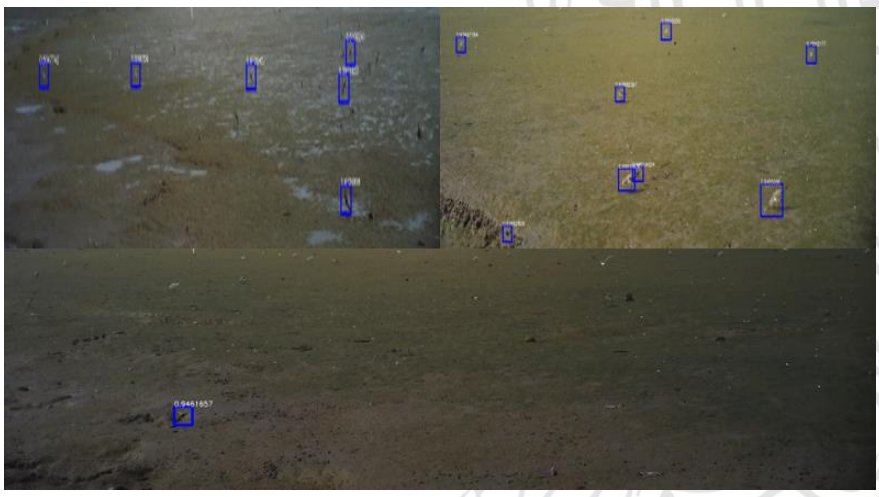

Figura 21. *Zonas de difícil baja detección y bajo desempeño.*

En la Figura 20 se observa que los valores de sensibilidad para cada uno de los clasificadores baja a medida que se aplica los diferentes umbrales: esto es debido principalmente a que el algoritmo no detecta a priori el número de aves adecuado, y a medida que se filtra mediante umbralización, se pierden detecciones. Además de esto,

se puede observar para los clasificadores MobileNet V16 y V20, como la precisión aumenta a medida que se aplica la umbralización; sin embargo, con la disminución de la sensibilidad, disminuye el F1 Score para estos casos. Se observa además de esto, como el clasificador MobileNet V18, para cada uno de los casos en estas zonas, tuvo un comportamiento robusto ante la precisión con y sin umbral, pero, de igual manera, a medida que se aplica los diferentes umbrales, las métricas de sensibilidad y de F1 Score bajan rápidamente.

Zonas de alta precisión en detección y adecuado conteo

En la Figura 22 se observa como los valores de precisión para los clasificadores MobileNet V16, V18 y V20 permanecen constante con y sin el uso de los diferentes umbrales. Aquellas sesiones que tenían pocas aves aumentaron su sensibilidad a medida que se aplicó el umbral (se destaca en este punto que la umbralización tiene un efecto positivo en el rendimiento de los clasificadores). Además de esto, se puede observar como el clasificador MobileNet V2 contó con una alta sensibilidad con y sin umbral; sin embargo, las métricas de precisión son altamente bajas (esto se debe a que se detecta un número alto de falsos positivos).

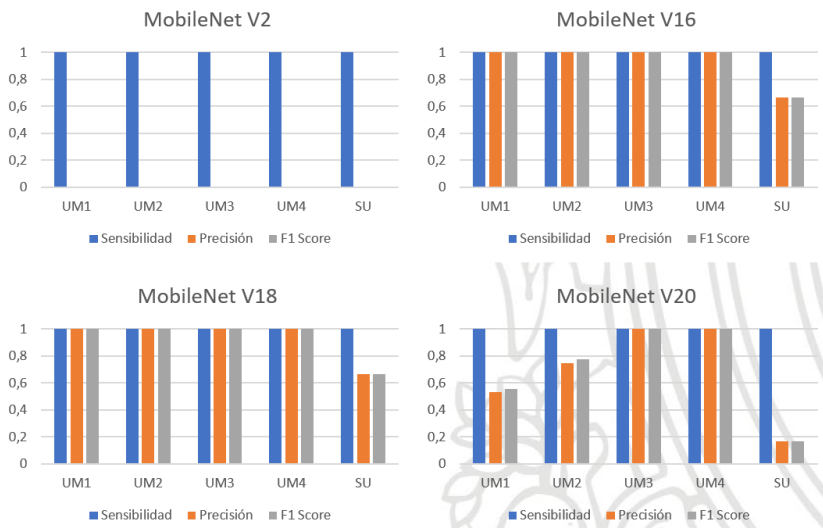

Figura 22. *Métricas de precisión y sensibilidad en zonas de alta precisión en detección y adecuado conteo.*

En la Figura 23 se observan las zonas de alta precisión en detección y adecuado conteo. Estas zonas se caracterizan por presentar un entorno uniforme, sin mucho pantano, y con poco mimetismo de objetos que pueden confundirse con la posición que adoptan las aves en su entorno. Para este tipo de entornos se recomienda usar los clasificadores MobileNet V16, V18 y V20.

Figura 23. *Zonas de alta precisión en detección y adecuado conteo*

Zonas de buen rendimiento en detección y adecuado conteo

En la Figura 24 se nota como el efecto de aplicar el umbral permite aumentar los valores de sensibilidad, y en consecuencia de F1 Score. Además de esto, se observa para los clasificadores MobileNet V16, V18 y V20 como los valores de sensibilidad aumentan por encima del 80%, y los valores de F1 Score para los clasificadores MobileNet V16 y V18 son cercanos al 60%.

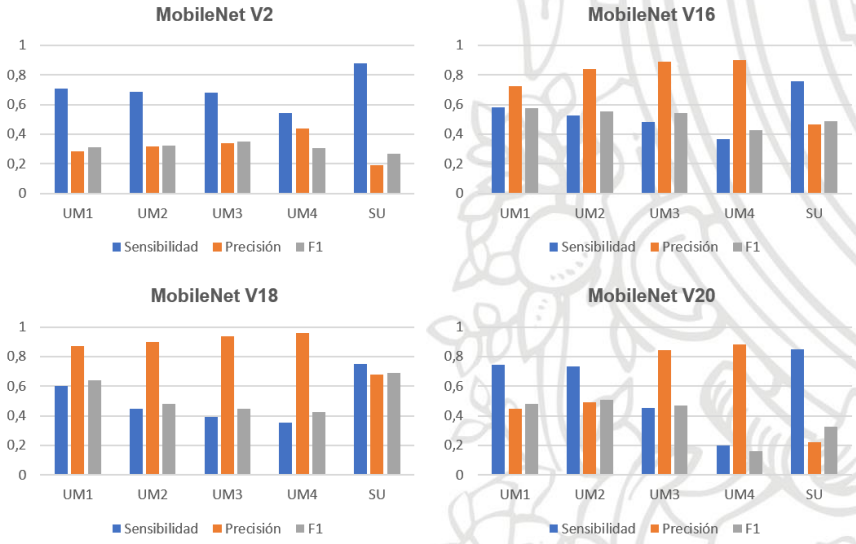

Figura 24. *Métricas de precisión y sensibilidad en zonas de buen rendimiento en detección y adecuado conteo.*

En la Figura 25 se puede observar las zonas de buen rendimiento para detección y conteo. Estas zonas se caracterizan por tener una distribución uniforme del entorno: no existe alto ruido de fondo (pantanos, rocas, huecos que mimeticen las posiciones de las aves). Además de esto, las aves se ubican en la línea de vista más cercana de la cámara, y existe un buen nivel de contraste entre el fondo y las aves. Para estas zonas, se recomienda usar los clasificadores MobileNet V16 y V18, ya que son los que exhiben un F Score cercano o superior al 60%.

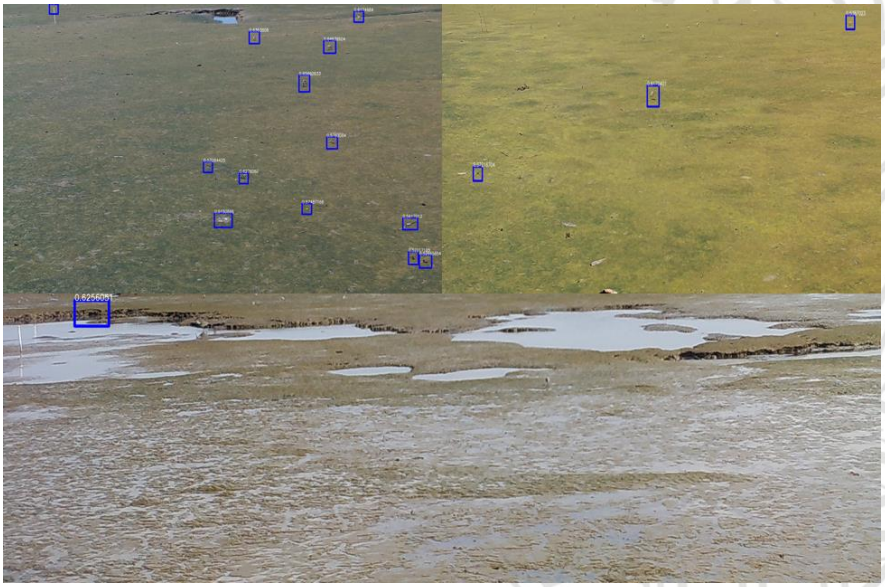

Figura 25. *Zonas de buen rendimiento en detección y adecuado conteo.*

Análisis de Vídeo

Para el análisis de las métricas de MSE y mAE se analizaron 100 vídeos con movimiento, y 106 vídeos con leve o nulo movimiento, y se comparó la influencia que tenía el módulo de estabilización en el conteo del número de aves.

1. **Error cuadrático medio y mediana del error absoluto sin estabilización:**

   Videos con movimiento:

   En la Figura 22 se puede observar a priori, que se presenta un alto error en el conteo en la mayoría de los vídeos, con falsos positivos de hasta 140 detecciones.

   En la Tabla 5 se observan las métricas de MSE y mAE del módulo de análisis de vídeo en vídeos con movimiento sin estabilización. Se puede observar que el valor de mAE es del 42, lo cual indica que se tiene un alto índice de detecciones falsas en el conteo del número de aves en cada vídeo.

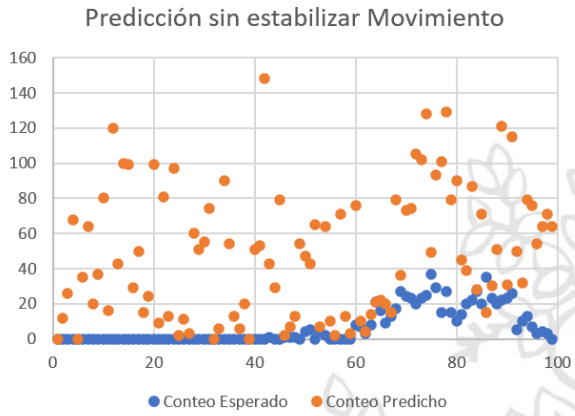

Figura 26. *Conteo esperado (en azul) vs conteo predicho (en naranja) para los vídeos con movimiento. Eje vertical: número de aves, eje horizontal: índice del vídeo de la sesión.*

Tabla 5*. Métricas de rendimiento promedio MSE y mAE del módulo de análisis de vídeo en vídeos con movimiento sin estabilizar.*

Vídeos	MSE	mAE
Sin Estabilizar	5,46	42

Se observa entonces, para videos en movimiento, que con el módulo de análisis de vídeo se tiene un alto índice de falsos positivos. En general, no se recomienda el uso del módulo de análisis de vídeo si los vídeos tienen movimiento.

Vídeos con leve o nulo movimiento:

En la Figura 27 se observa el conteo esperado y el conteo predicho por el algoritmo de análisis de vídeo de los vídeos con leve o nulo movimiento. Se destaca que se presenta un conteo predicho altamente similar al conteo esperado.

En la Tabla 6 se observan las métricas de MSE y mAE del módulo de análisis de vídeo en vídeos con leve o nulo movimiento sin estabilización. Se encuentra que el valor de mAE es de 2, lo cual indica que el error es bajo. Se recomienda entonces, para este tipo de vídeos, usar el módulo de análisis de vídeo para estimar el número de aves en las grabaciones.

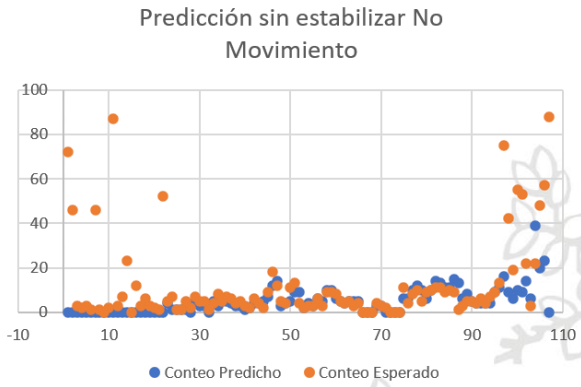

***Figura 27**. Conteo esperado (en azul) vs conteo predicho (en naranja) para los vídeos con leve o nulo movimiento. Eje vertical: número de aves, eje horizontal: índice del vídeo de la sesión.*

***Tabla 6.** Métricas de rendimiento promedio MSE y mAE del módulo de análisis de vídeo en vídeos con leve o nulo movimiento sin estabilizar.*

Vídeo	MSE	mAE
Sin Estabilizar	1,88	2

2. Error cuadrático medio y mediana del error absoluto utilizando el módulo de estabilización:

A partir de los resultados anteriores, se observa que el movimiento en la grabación afecta de manera importante el desempeño del módulo de análisis de vídeo. Por tanto, se tomaron los vídeos que presentaban movimiento en su grabación, y se pasaron a través del módulo de estabilización de video (de esta manera es posible comparar el efecto de los vídeos están estables con aquellos que no lo están). Además de esto, se analizó el efecto que tenía el módulo de estabilización en los vídeos que presentaban leve o nulo movimiento. A continuación, se reportan los resultados para cada caso:

Videos con movimiento:

En la Figura 28 se presenta el conteo esperado y el conteo predicho por el algoritmo de análisis de vídeo de los vídeos con movimiento cuando se estabilizan. Se puede observar cómo, en contraste con la Figura 26, se logra mejorar notablemente el conteo predicho en cada uno de los vídeos, obteniéndose un error menor en el conteo predicho para cada uno de los vídeos.

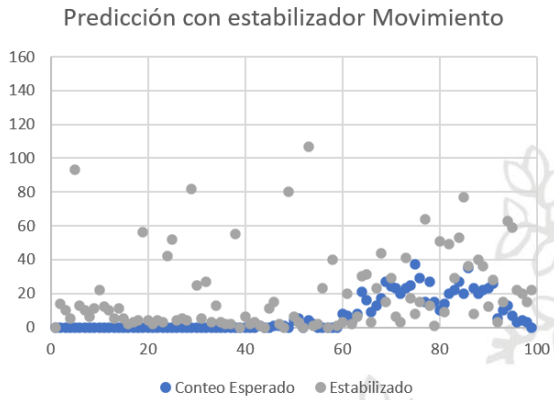

***Figura 28**. Conteo esperado (en azul) vs conteo predicho (en naranja) para los vídeos con movimiento. Eje vertical: número de aves, eje horizontal: índice del vídeo de la sesión.*

En la Tabla 7 se observan las métricas de MSE y mAE del módulo de análisis de vídeo en vídeos con movimiento cuando se estabilizan con el módulo de estabilización. Se puede observar una disminución del mAE de 42 a 10, y del MSE de 5,4593 a 2,6563; lo cual representa una mejora significativa, e índica, que, con el uso del módulo de estabilización, es posible disminuir el sesgo en el conteo.

***Tabla* 7.** *Métricas de rendimiento promedio MSE y mAE del módulo de análisis de vídeo en vídeos con movimiento con estabilizar.*

Vídeo	MSE	mAE
Con Estabilización	2,66	10

Se recomienda que, cuando los vídeos de la sesión presenten movimiento se use el módulo de estabilización, ya que existe disminución significativa en el error de conteo para cada uno de los vídeos que componen la sesión.

Vídeos con leve o nulo movimiento:

En la Figura 29 se puede observar el conteo esperado y el conteo predicho por el algoritmo de análisis de vídeo de los vídeos con leve o nulo movimiento. Se puede observar que, en comparación con la Figura 27, existe un incremento de puntos atípicos, en el que el sesgo del conteo aumenta.

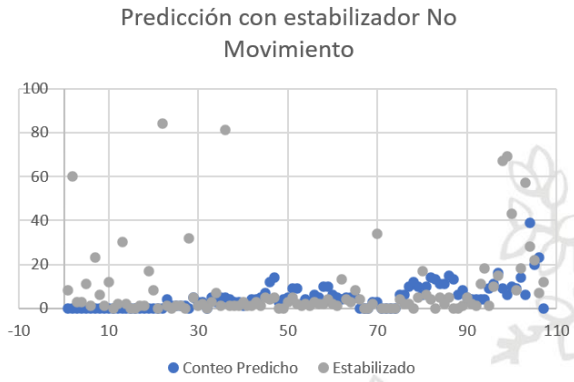

Figura 29. *Conteo esperado (en azul) vs conteo predicho (en naranja) para los vídeos con leve o nulo movimiento. Eje vertical: número de aves, eje horizontal: índice del vídeo de la sesión.*

En la Tabla 8 se observan las métricas de MSE y mAE del módulo de análisis de vídeo en vídeos con leve o nulo movimiento con estabilización. Se puede observar que, aunque el valor de MSE logra disminuir de 1,8828 a 1,7453, el valor de mAE aumenta de 2 a 3. Sin embargo, el incremento en el mAE no es significativo, lo cual indica que con el uso del módulo de estabilización no se afecta en gran medida los vídeos que inicialmente tenían leve o nulo movimiento.

Tabla 8. *Métricas de rendimiento promedio MSE y mAE del módulo de análisis de vídeo en vídeos con leve o nulo movimiento sin estabilizar.*

Vídeo	**MSE**	**mAE**
Sin Estabilizar	1,76	3

Se observa entonces, que cuando los vídeos que presentan leve o nulo movimiento se pasan a través del módulo de estabilización, no se presentan alteraciones significativas en el conteo de aves. En general, si las grabaciones contienen leve o nulo movimiento, se recomienda usar el módulo de análisis de video directamente.

Conclusiones

- Con este proyecto de investigación se propuso una herramienta de software de bajo costo computacional que permite estimar la densidad poblacional de aves en fotografías y vídeo de manera automática, para de esta forma dar un soporte a los expertos que deben analizar la información primaria, la cual puede contener un alto volumen de datos.
- Para el análisis de imágenes en escenarios donde las aves se encuentran a lo lejos de la línea de vista principal de la cámara, o existe mimetismo, no se recomienda en general el uso del software TRINGA, debido, a que los diferentes algoritmos no son robustos ante este tipo de análisis. En general, se recomienda el uso de los clasificadores MobileNet V16, V18 y V20, para la estimación y conteo, siendo el clasificador MobileNet V20 el que más bondades presenta ante escenarios donde no existe un buen contraste entre las aves y el fondo, y los clasificadores MobileNet V16 y V18, clasificadores con una buena precisión y sensibilidad en escenarios donde si lo hay.
- El software se recomienda para analizar imágenes donde el enfoque se haya realizado hacia la zona de interés, buscando que las aves que se van a contar queden en la línea principal de vista de la cámara. Es importante que se analicen imágenes con buen contraste, donde se evite reflejos dados por el agua, pequeños montículos de arena, o aves lejos de la línea principal de vista de la cámara.
- Para el análisis de vídeo en escenarios donde las grabaciones presentan movimiento, no se recomienda el uso del módulo de análisis de vídeo de forma directa, sino que se recomienda estabilizarlos a través del módulo de estabilización de vídeo. Además de esto, se observó que para aquellos vídeos en los que se presenta leve o nulo movimiento, el módulo de estabilización de video no introduce un ruido significativo que afecte su posterior análisis. Sin embargo, en si el movimiento es bajo es mejor no aplicar el módulo de estabilización.
- Para el análisis del módulo de estabilización de vídeo en escenarios donde las grabaciones no presentan movimiento, se recomienda el uso del módulo de análisis de vídeo sin estabilización. Sin embargo, debe notarse que el módulo de análisis de vídeo no tiene una etapa de clasificación de las regiones de interés que detecta, por lo que se recomienda en general que los escenarios de grabación solo contengan aves en su línea de vista principal.

Trabajo a Futuro

- Sería interesante implementar un algoritmo de aprendizaje por refuerzo para imágenes y vídeo, que permita actualizarse en línea. De esta forma se cree que es posible, aumentar la precisión y sensibilidad en las detecciones y conteo. Adicionalmente, se propone implementar una interfaz de usuario en servidor, que permita a los ornitólogos e investigadores operar el algoritmo en red, para de esta manera operar en plataformas de escritorio y móviles.

Bibliografía

[1] C. I. Selem-Salas, M. C. MacSwiney y S. H. Betancourt, “Aves y Mamiferos”, Técnicas de muestreo para manejadores de recursos naturales, pag. 351, 2004.

[2] J. M. Wunderle, “Métodos para contar aves terrestres del Caribe”, 1994.

[3] L. Pulido, “Metodologia de identificacion y clasificacion automatica de mamiferos en imagenes obtenidas mediante camaras trampa”, 2018.

[4]: Caravaggi, A., Banks, P. B., Burton, A. C., Finlay, C. M., Haswell, P. M., Hayward, M. W., ... & Wood, M. D. A review of camera trapping for conservation behaviour research. Remote Sensing in Ecology and Conservation, 3(3), 109-122, 2017.

[5]: A. Diaz-Pulido y E. P. Garrido, “Densidad de ocelotes (Leopardus pardalis) en los llanos colombianos”, Mastozoologıa neotropical, vol. 18, n.o 1, p´ags. 63-71, 2011.

[6]: S. Kulkarni, D. Bormane y S. Nalbalwar, “Video stabilization using feature point matching”, en Journal of Physics: Conference Series, IOP Publishing, vol. 787, p´ag. 012 017, 2017.

[7]: A. Lim, B. Ramesh, Y. Yang, C. Xiang, Z. Gao y F. Lin, “Real-time optical flow-based video stabilization for unmanned aerial vehicles”, Journal of Real-Time Image Processing, vol. 16, n.o 6, pags. 1975-1985, 2019.

[8]: J. R. Uijlings, K. E. Van De Sande, T. Gevers y A. W. Smeulders, “Selective search for object recognition”, International journal of computer vision, vol. 104, n.o 2, pags. 154-171, 2013.

[9]: J. H. Bappy y A. K. Roy-Chowdhury, “CNN based region proposals for efficient object detection”, IEEE International Conference on Image Processing (ICIP), IEEE, 2016, pags. 3658-3662, 2016.

[10]: S. Ren, K. He, R. Girshick y J. Sun, “Faster r-cnn: Towards real-time object detection with region proposal networks”, en Advances in neural information processing systems, pags. 91-99, 2015.

[11]: P. Christiansen, L. N. Nielsen, K. A. Steen, R. N. Jorgensen y H. Karstoft, “DeepAnomaly: Combining background subtraction and deep learning for detecting obstacles and anomalies in an agricultural field”, Sensors, vol. 16, n.o 11, pag. 1904, 2016.

[12]: S. Albawi, T. A. Mohammed y S. Al-Zawi, “Understanding of a convolutional neural network”, en International Conference on Engineering and Technology (ICET), IEEE, pags. 1-6, 2017

[13]: A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto y H. A. Mobilenets, “Efficient convolutional neural networks for mobile vision applications”, arXiv preprint ArXiv:1704.0486, 2017.

[14]: Zhu, Y. F, Background subtraction and color clustering based moving objects detection. In 2009 International Conference on Information Engineering and Computer Science (pp. 1-5). IEEE, 2009.

[15]: Panda, D. K., & Meher, S. Detection of moving objects using fuzzy color difference histogram based background subtraction. IEEE Signal Processing Letters, 23(1), 45-49, 2015.

Anexos

Recomendaciones para toma de vídeos y fotografías

- Programar la cámara para que cada que se vaya a tomar se realicen 6 fotografías seguidas, de esta manera se obtendría más información del fondo que podría ser procesada por el algoritmo.

- Evitar el movimiento de la cámara al máximo, ya que esto dificulta el modelamiento del fondo por parte del algoritmo. Este aspecto es muy importante para tener un conteo aproximado.

- En lo posible, ubicar la cámara en una posición donde el sol no llegue directo, ya que esto distorsiona la toma de imágenes

- Apuntar la cámara ligeramente hacia abajo (pendiente definir ángulo). Esto permite captar el área inmediatamente cerca de la cámara y evita un horizonte extenso (que no puede analizarse-

- Sesión de fotos se definirá como todas las fotografías/video que se capturan en una misma zona, en un mismo punto, durante una cantidad de tiempo determinada. (Eg: Malpelo, cámara 1, grabación continua desde las 6 am hasta las 6 pm; Malpelo, cámara 2, toma de fotos cada hora durante 3 días).

- Programar la hora de inicio de toma de fotografías/video para un minuto después de que se haya instalado la cámara. El objetivo es disminuir la cantidad de imágenes /tiempo de video que no corresponde al experimento.

- Antes de retirarse del sitio donde se instaló la cámara verificar que si haya encendido.

- Para escenarios donde las aves presentan alta oclusión con el ambiente, sustituir la toma de imágenes por vídeo corto (apx 5 seg). Igualmente, para escenarios donde las aves tienen colores muy similares con el ambiente donde se encuentran (eg: Iscuande), o donde las aves se visualicen muy pequeñas en comparación con su ambiente.

- Para la toma de vídeo, se recomienda un vídeo corto (no más de 5 segundos), ya que se pretende calcular el número de aves presentes en el intervalo analizado.

- Para realizar videos, es importante tener entre 33 y 34 frames por segundo de la cámara. Si se tienen menos frames se debe tomar videos mayores a 5 segundos.

Recomendaciones para el conteo manual

Revisar la hora precisa (centésima de segundo) e id de la imagen. Aunque la cámara está programada para cierto lapso de tiempo, algunas veces puede haber demoras o adelantos en el mecanismo de disparo generando imágenes dentro del mismo minuto (09:17:01 y 09:17:59). Tener en cuenta a lo hora de llenar la hoja de datos.