Design and Development of a System for Aerial Video Survey of Large Marine Animals

Dr. Hari Kalva
Associate Professor
hari.kalva@fau.edu

Sagar Aghera
Graduate Student
saghera@fau.edu

Rafael Giusti
Graduate Student
rgiusti@fau.edu

Dr. Oge Marques
Associate Professor
omarques@fau.edu

Waazim Reza
Graduate Student
wreza@fau.edu

Asif Rahman
Graduate Student
arahman2@fau.edu

Florida Atlantic University - CEECS – Multimedia Lab
777 Glades Road. SE, Room 300
Boca Raton, FL 33431 PO BOX 3091
Phone: 561-297-3000

ABSTRACT

Population counts of large marine animals and associated tasks are currently performed using aerial surveys by human observers. Six hour-long survey sessions are common, and could lead to fatigue and visibility biases which affects the quality of the results. Survey quality can be improved implementing a video recording system that allows post-flight analysis intended to look for the presence of marine animals. In this paper we present the design and development of a system to record the video of the regions being monitored by human observers. One of the key challenges is to meet the requirements of the biologists and closely follow current aerial observation practices. The system is designed to meet the needs of biologists performing sea-turtle observation on the east coast of Florida. Nonetheless, the system is designed to be flexible to meet any other requirements, for example, different targets such as seals, dolphins, whales, or boats.

Keywords: Aerial surveys, image acquisition systems, image processing, video processing.

1. INTRODUCTION

This project is part of a larger ocean energy effort trying to harness energy from ocean currents by placing turbines deep under the ocean. Sea-turtles are currently facing major threats such as the destruction and alteration of their nesting and foraging habitats. The planned deployment of huge underwater turbines in the Florida Gulf Stream could affect the pattern of sea-turtle migration and reproduction. Therefore, a comprehensive study of the distribution of sea turtles is planned. The current state of the art is to have biologists fly on an aircraft and record the sightings of sea turtles and the species of the sea turtles found. The population counts are extrapolated using the distance sampling theory. The results based on the observations made by human observers during flights are error-prone and the counts can be inaccurate, due to visibility biases.

The goal of this project is to design and develop a system to record aerial video, analyze the video, and establish an accurate population count of the sea turtles. The video analysis can be used to determine an automatic count or to aid biologists to verify and correct their visual observations. The scope of this paper is limited to the design and development of an aerial video surveillance system. The system presented will be mounted under the wings of a Cessna 337 Skymaster (a four seat aircraft) flying 500 feet above the Atlantic Ocean to record the video of the sea turtles passing by. The operating environment, the requirements placed by biologists, and the need for robustness and reliability make this a very challenging task. This paper makes the following key contributions: 1) presents the challenges of designing a complete multimedia system, 2) shows how requirements and the operating environment affect design choices in multimedia systems, 3) provides a complete example of a multimedia system design. This work consists of developing a digital image based solution divided into the following stages: data collection, data analysis, and data browsing & dissemination. Indeed, the system is designed to meet the following global requirements:

- Design a system capable of emulating the task of a biologist conducting visual observations in order to fulfill the distance sampling methodology used in aerial surveys.
- Collect accurate and high quality data, including images as well as GPS tracking data of the survey.
- Analyze collected data to extract parameters and data needed as the input in distance sampling methods.
- Design a visual browsing tool for accessing collected data as well as analysis of results.

Providing a complete automated tool to the research community will help them improve their current procedures, by saving time and cost, and by improving the collected data in terms on robustness, reliability, repeatability, reusability and reproducibility.

2. Background

Aerial survey is a widely used technique based on distance sampling theory, which allows estimating species population and distribution (Buckland et al., 1993). It is very useful in marine environments where displacement of observers is compromised at least when comparing against ground environments. Marine animal surveys have been done by biologists before, and some of the recent work includes surveying of sea turtles in North Carolina (Chester et al., 1995) and Western Mediterranean (Cardona et al., 2005), as well as whales in Alaska (Hobbs et al., 2000). All of these studies implement a similar survey design with some difference depending on the target and environment conditions. The process is entirely manual where biologists make observation from aircraft, looking down though the aircraft window. These surveys are often repeated regularly to keep track of changing populations.

These aerial surveys have similar characteristics such as airplane speed, altitude, number of observers, distance of observations, and length of surveys. For instance, turtles observation in North Carolina (Chester et al., 1995) and West Mediterranean (Cardona et al., 2005) have similar airplane speed and altitude around 75 knots and 500 ft respectively, while seals and whales observation are performed faster and higher, around 100 knots speed and 600 to 800 ft altitude. Regarding minimum weather conditions, all surveys use the Beaufort scale to monitor sea conditions and are carry out if the estimated value is less than 3 (gentle breeze). Other common characteristics found were the use of GPS coordinates to register sightings locations and the use of inclinometers to determine the location of sightings with maximum precision.

Some attempts to automate the process have been tried by recording surveys in videotape for manual post analysis as is in Belugas counting in Alaska (Hobbs et al., 2000). In that case, two cameras with different zoom area were mounted in the airplane. Initially, the videotape camera models used were Cannon Hi-8 814XL-S and Ricoh Hi-8 R800-H with 8x magnification for the zoomed camera. Later, video camera models were substituted with Sony Digital 8 DCR-TRV103 and Sony Digital Camcorder DSR-PD100A with 5x magnification. Improvements in the process were achieved by manually analyzing the whole recorded video and comparing against live counting during the trip. To the best of our knowledge, there is no other related work with the same broad scope as the current project.

3. System Requirements

The proposed system is not intended to replace or modify the procedure researchers follow to accomplish their aerial surveys; instead the system serves as a tool to improve the collection of the survey data. The survey we are modeling is designed based on distance sampling theory (Buckland et al., 1993) and is similar to the procedures used by many researchers in aerial surveys (Caughley, 1977), (Chester et al., 1995), (Cardona et al., 2005), (Elko and Holman, 2005) and (Hobbs et al., 2000). The proposed design is intended to develop a tool to emulate the human tasks when watching, counting, measuring, and registering data during each flight. In order to achieve such a goal, the proposed model has to meet the survey requirements while being aware of the most relevant environmental conditions. This section summarizes the requirements of the biologists for an aerial video surveillance system.

3.1 BIOLOGIST NEEDS

A primary requirement is to emulate human observation so that established sampling techniques can be used in determining populations. The aerial survey technique basically consists of flying an aircraft along a well defined path known as a transect line and recording the sea turtles found in an observation area known as the Strip at a given distance from the transect line. As shown in Figure 1, the strip is 150 meters wide and is 150 meters away from the transect line. In the proposed system, cameras in the airplane must follow the transect lines and must record the target objects found on both sides of the transect line inside the strip area. To obtain accurate population estimate using the distance sampling theory, it is also very important to record as accurate as possible the perpendicular distance from the object found to the transect line as well as the location of the observer. Since biologists also record the species of the sea turtles and their observed location, the proposed should have very high resolution cameras and a GPS to record the position. Another key requirement is that the system should operate with any human intervention. Biologists on board do not have the time or resources to configure or monitor the system. The system should record upon power-up and adjust automatically to changes in environmental conditions.

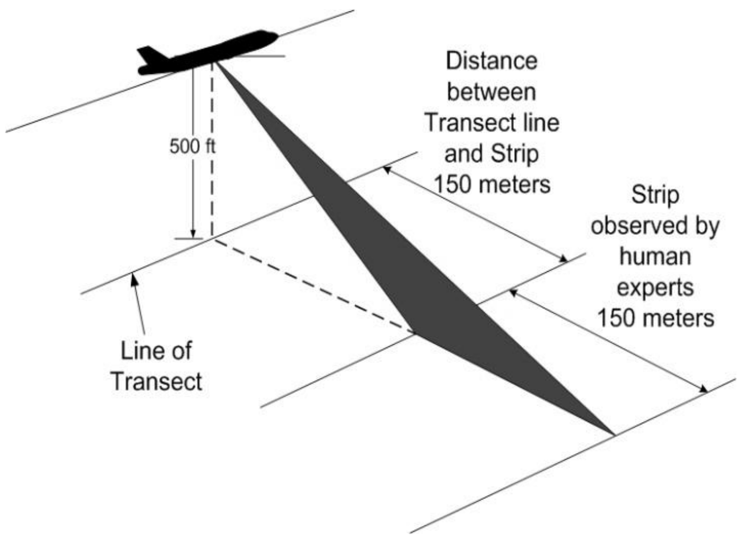

Figure 1: Transect line and observation strip

3.1.1 SURVEY PATH

In the current work, the survey is designed very similar to previous surveys (Chester et al., 1995), which follow the recommendations from the distance sampling theory applied to sea turtle targets in the marine environment. The directions of transect lines are west-to-east and east-to-west covering the area limited from Fort Lauderdale to Bahamas. Two successive transect lines are separated by 1 Km as is shown in figure 2.

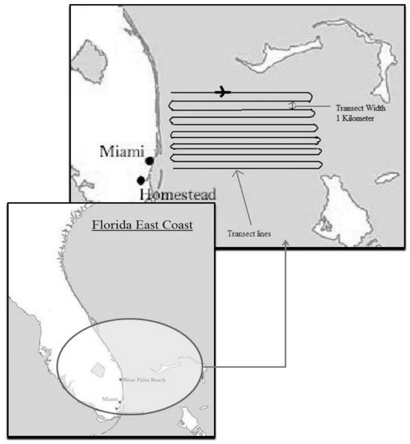

Figure 2: Survey area and transect lines

The strip line, where the turtle are going to be counted, is 150 m wide and is separated 150 m from the transect line. The altitude of the flight is 500 ft and the speed is 90 knots. Each survey trip is expected to last 6 approximately hours. There will be monthly surveys over a one year period. The data collected by biologists during these surveys will form the basis for determining the ground truth. This ground truth is then used in analyzing the performance of the analysis algorithms.

3.2 Environmental Challenges

The quality of the images collected will depends on many factors that can be classified as physical and optical factors. The first class includes airplane vibration, wind, waves, altitude and speed. The optical effects are represented as sunlight, reflection from ocean surface, and brightness variations. All these factors together define the correct selection of cameras suitable to handle each of these problems.

Environmental conditions affect directly both the physical and optical factors and subsequently the quality of the images recorded. Designing for the environmental conditions becomes the most important challenge since selecting an appropriate type of cameras and other settings in the system, the system can be adapted to any expected variation in environmental conditions. Ocean conditions can change during the trip and the system has to automatically adjust to these variations without human intervention.

4. System Design

The system can be split in three main stages: data collecting, processing and browsing. The first stage takes care of capturing image data as accurately as possible. The first stage occurs on the aircraft while the biologists are conducting the surveys. Upon return from the survey flight, the main database is updated with the collected data. The second stage consists on analyzing the data by accessing the raw frames and extracting the information required for later usage. The final step is browsing, which is intended for use by biologists and general public to access the survey data. Figure 3 shows a high level diagram of the system.

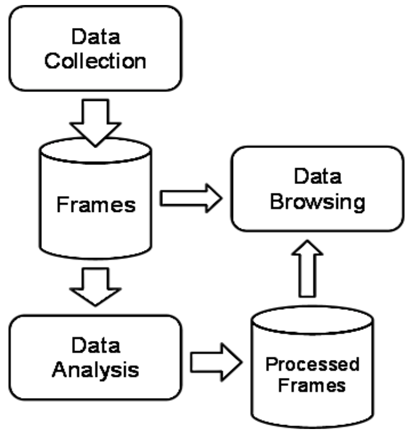

Figure 3: Aerial survey system

4.1 SYSTEM MODELING

Figure.1 shows the field of view from the airplane and the area to be covered by camera(s). Factors such as the height of the airplane, the strip location, and width and the size of the target –sea turtles – determine the data acquisition systems. As per survey design, 150 meters strip on either side of aircraft which is 150 meters away from transect line has to be recorded by the cameras. The size of target varies from 1 foot long loggerhead turtles to 3.3–6.6 ft leatherback turtles. The cameras should be able to capture with reasonable granularity, a 1 square foot object from the area which is 150 meters wide. Long focal length lenses with a high resolution camera can be very handy for capturing spotting 1 square feet object but it won't be able to cover the entire 150 meter strip. One Camera with long focal length lenses is not enough as the area to be covered is very large and density of image needed is very high to spot out a turtle from the image. In order to meet the requirements, a two-camera configuration with medium focal length lenses was considered. In this configuration, each camera focuses on the 90 meter strip with about 30 meter overlap between the cameras and together they cover 150 meter wide strip of interest. Overlap in field of view of two cameras was taken into account to avoid missing some area of interest in case of vibration of plane or other external disturbances to camera. Table 1 shows the number of pixels/feet for the 90 meter wide strip using different resolution cameras. For object detection algorithms to work effectively, we need minimum of 10 pixels per feet. Hence cameras with 2.5K resolution or higher are ideal for our application.

Table 1: Pixel distribution for different resolution cameras

Width	Avg. Pixels/ft for 2k res. camera	Avg. Pixels/ft for 2.5K res. camera	Avg. Pixels/ft for 3K res. camera	Avg. Pixels/ft for 4K res. camera
90m	8	11	13	16

4.2 DATA COLLECTION

Data collection stage is designed to take place independently in the aircraft, with the simplest operation procedures possible. The system consists of cameras mounted outside the aircraft and connected to a recording console placed inside the aircraft. A diagram of the design is shown in figure 4. A PC console is in charge of collecting frames as well as GPS data coming from attached cameras and GPS devices. All these devices are plugged into power supply available in the airplane (an inverter connected to a DC power source).

Cameras are stream frames continuously during the 6 hour trip at a rate of 15 fps to the console PC. The program will run continuously without interruption starting when the console boots, so the only operation needed is powering on the PC console.

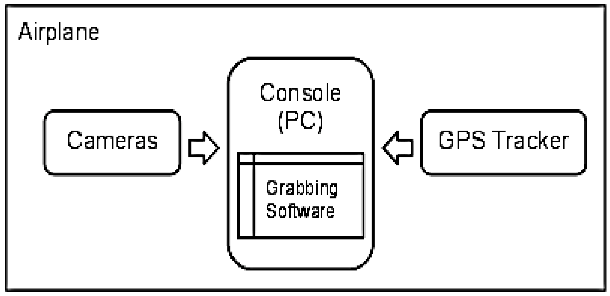

Figure 4: Collecting Stage

4.2.1 DSLR VS. INDUSTRIAL CAMERAS

Selection of camera is not a trivial task. Initially DSLR camera from Nikon and Canon brands were considered for their high resolution and high speed performance. There are some special requirements for our application for which DSLR cameras are not suitable. These requirements include:

- Continuous operation for 6 hours – DSLRs are not designed for continuous operation over 6 hours. With long continuous use the sensors are known to get hot and cause camera malfunction.
- Interface with PC for saving stored images in real time – DSLRs have minimal support for saving data directly to hard disk.
- Electronic shutter – with long continuous use, the mechanical shutters in DSLRs are known to break
- Camera weight – with the cameras mounted outside the aircraft and the aircraft moving at 120 miles per hour, minimizing the weight of the cameras is critical to minimize vibration and use light weight mounts.

Industrial cameras are designed for use in special application (like our application) where user can programmatically control camera parameters and can work in rugged environmental settings for long period of time. Hence option of using DSLR camera was ruled out.

Industrial Cameras fall into two main categories namely,

- Cameras with CCD image sensors
- Cameras with CMOS image sensors

4.2.2 CCD VS. CMOS IMAGE SENSORS

CCD (charge coupled device) and CMOS (complementary metal oxide semiconductor) image sensors are two major technologies for capturing images digitally. Each one has unique strengths and weaknesses giving advantages in some applications and disadvantages in others. The choice of one over other is mainly influences by the application. Comparing both sensors based on the cost of chip manufacturing does not give any significant difference in price points. CMOS image sensors offer more integration which results into more functions but at the cost of image resolution. CCD sensors on other hand offer higher performance based on highest image quality at the expense of system size (at sensor level). We chose CCD sensor over CMOS as our application demands higher image quality.

4.2.3 GLOBAL SHUTTERS VS. ROLLING SHUTTERS

CCD sensor cameras uses electronic shutters which consist of two types of shutters, namely Global shutters and Rolling shutters. In sensors with rolling shutter, all pixels in one row of the imager collect light during exactly the same period of time, but the time light collection starts and ends is slightly different for each row. This creates rolling or skew effect which distorts the image with moving objects (as moving plane in our application). While in sensors with Global Shutter, all pixels in imager are exposed simultaneously for same amount of time and hence don't have rolling or skew effect on the image.

Hence CCD camera with global shutter meets all the requirements for this application. We chose AVT GC2450C Industrial Camera. Following are the key features of AVT GC2450C camera:

- 5 Megapixels (2448x2050)
- Fast – 15 frames per second
- Progressive Scan – Global Shutter
- Gigabit Ethernet interface
- Raw Image format: Bayer8 and Bayer16

4.2.4 SELECTION OF LENS:

Based on the geometry shown in figure 1 and the sensor size of camera, lenses need to be selected. We got C-mount Pentax fixed focal length lenses, which goes with the AVT GC2450C Industrial Camera. We got 35mm and 50mm focal length lenses to cover the width of the strip with two cameras as defined in Geometry modeling. This configuration gives successive images to have 30% overlap when recorded at 5 frames per second.

Circular polarizing filters are used to remove glares and reflections of sunlight from waves. Polarizing filters are important for post processing stages for sea animal identification as waves can become obstacles in object detection algorithms.

4.2.5 GLOBAL POSITION SYSTEM (GPS)

GPS is also an important part of Data Collection system to tag images with proper location co-ordinates which are necessary to cross verifying turtles found by biologists and also for data post processing stages. We chose Garmin GPS N96, which is Aviation GPS which has WAAS capable receiver. WAAS (Wide Area Augmentation System) is a system of satellites and ground stations that provide GPS signal corrections, giving a better position accuracy. WAAS capable receiver is considered the best for open land and marine applications.

4.2.6 SOFTWARE DESIGN

The camera software is designed to be robust and handle every event that could possibly disrupt the functioning of the system such as software or camera crashes, communication losses and operating system crashes. The camera software works on the concept of threading which helps in handling multiple cameras with single code/application (see Figure 5). Whenever the camera is plugged in, a new thread is created which runs some functions like CameraSetup, CameraOpen and CamCapture to start streaming and capturing frames with user defined settings from configuration file. The software is designed in such a way that it can handle multiple cameras by creating different threads which work parallel. Whenever the camera is unplugged it closes the thread smoothly without disturbing other online threads. The figure 5 shows a software sequence diagram for the camera software.

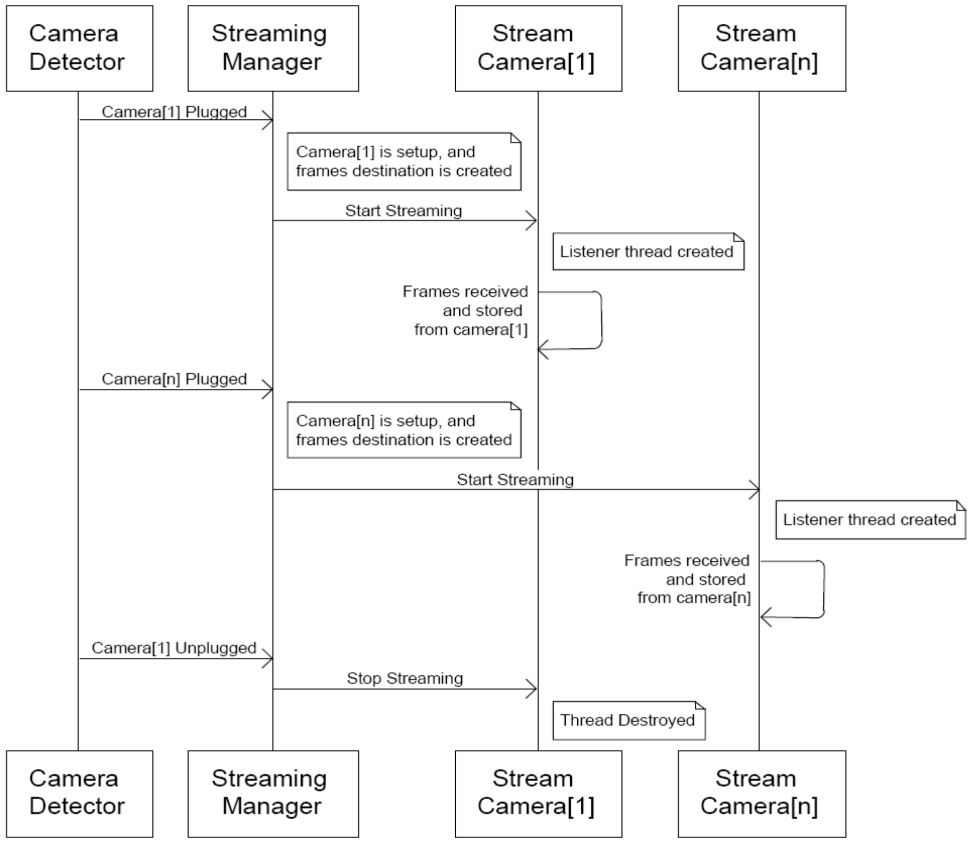

Figure 5: Software sequence diagram

The data collected by the software is of supreme importance and the safe recovery of data during system crashes has to be ensured. For this reason, a simple directory structure and file naming convention was proposed which also makes querying and retrieving data in the post processing stage easier. The survey will include Timestamp readings for folders as well as images for specific Camera. Images captured by a camera are saved with its corresponding frame counter and timestamp, which distinguishes with other images. Also the three level hierarchy of the directory structure ensures that the images do not overwrite if the system has restarted after a crash.

4.2.7 RECORDING CONSOLE WITH STORAGE

A recording console is integrated with the cameras and GPS devices in order to synchronize and coordinate them and collect data from these units. As the survey conducted is a data intensive process with no importance for the processing needs, the sole purpose of a recording console is to record and store the camera and GPS.

The key requirement for storage devices are reliability and recording speed. Reliable data storage is a mandatory requirement as the trip cannot be repeated. Solid state disk drive is a good choice considering all the requirements stated above. However, for the 2 TB disks we need, these disks are extremely expensive compared to the standard magnetic drives. Since we need an inexpensive solution that provides consistent and reliable data, solid state drives are voted out. Instead two regular SATA, RAID 0 hard disk drives with a total capacity of 2 TB are selected which is fast and has enough space for all the data that are captured for the six hour trip. Table 2 shows the storage space requirements for different file formats for 2 Camera configurations.

Table 2: Storage requirements for different file format for a 6-hours survey

Frames Per Second	**6-Hours Survey Storage (GBytes)**				
	Raw	Compressed - JPEG			
	tiff (100)	high (100)	med-high (75)	med (50)	low (25)
3	1800	300	30	18	13
5	3000	500	50	30	22
15	9000	1600	160	90	67

To accommodate the aspects of vibration, physical space constraints and automated operation of the recording console inside the aircraft, many solutions were considered. Custom assembled embedded devices like Stealth Computer's rugged mini PC were very attractive by the specification listing. However, these mini PCs are prone to heating and not expandable. Instead of over provisioning of the system, we planned to setup up a system with the bare essentials and then upgrade the system if required after every flight trip. This brings us to use a highly powerful workstation, Dell Precision, with Intel Core 2 Quad CPU, which is twice as fast and half as expensive as the mini PCs.

4.3 Camera mounting equipment

The aircraft to be used for aerial survey is Cessna m-337 Skymaster which will fly at speed of 100-110 knots at 500-550 feet height. Only mounts available are bomb racks and wing mounts. Gyro stabilizer is one of the options for Camera mounting. Gyro stabilizers are used inside airplane and helicopters and works well to dampen vibrations. These stabilizers cost several thousand dollars and increase the cost of the system. Additionally, all gyro stabilizers on market are designed for operating inside the aircraft and we need to mount cameras outside the Airplane to get clear view and proper angle rather than seeing through window glass. Hence wing mount was one of the options to mount camera. Figure 6 shows wing mount on the Airplane. We will be using wing mounts to mount the cameras. To damp vibration of the Airplane we will use rubber damping with mini tripod stand, which will be clamped with wing mounts. This will make sure Cameras are properly placed at angle to capture the region of interest. These mounts together with very short exposure time in bright daylight are expected to eliminate the effects of vibration on the recorded images

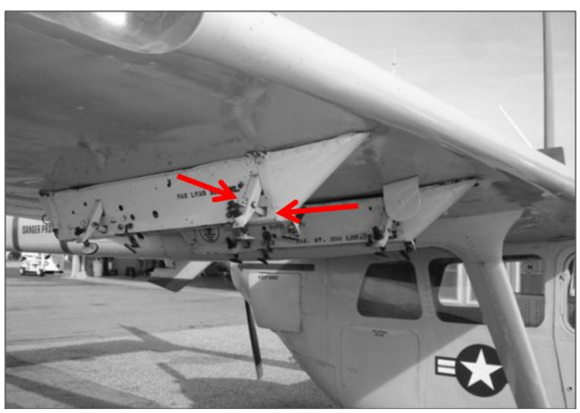

Figure 6: Wing mounts for mounting the cameras

4.4 Data analysis

The main task of this stage is to extract all details possible from each frame, which includes detecting sea turtles and counting them. The analysis and turtle detection tasks are outside the scope of this paper and not discussed in detail.

The quality of the images affected by the variation of environmental conditions is going to determine the capabilities of the analysis stage. Images are expected to be compromised by sunlight reflection, brightness variation, presence of shades, clouds reflection, waves, airplane speed, and vibration. Although, all this challenges can be improved by computational techniques, an improper setting in the collection stage can make frames unusable.

In the current work, the open source library, OpenCV, is the image processing and computer vision library chosen to implement the analysis algorithms. A stand-alone application written in C++ is going to read collected frames and process them with the OpenCV API running the designed algorithm, which consists on the following steps: Pre-processing for brightness, color, contrast, saturation and blurring adjustments; segmentation for background/foreground and objects detection; feature extraction; and classification based on target characteristics like shape, color and texture. After processing frames, the output is going to be metadata linked to frames files. The descriptor data will contain objects found as well as details like size, type, color and shape of each finding.

All extracted data will be available in a database for access from the browsing stage, avoiding the reprocessing of large amount of data.

4.5 Data browsing and dissemination

This is the stage where the system data is shown to the end users. The goal is to offer a web based interactive tool with the following features to browse the collected data trying to keep the quickest response possible.

4.5.1 Web-based User Interface

The raw frame data as well as the post-processed data will be stored with reference to the frame in a MySQL database. Utilizing this database, the information will be displayed to the users in a web-based application interface. The application will be developed in a combination of PHP and MySQL for backend querying and retrieval of data, along with cutting-edge HTML5 and JavaScript for the front-end interface.

We decided to go with a web-based application for the client end interface due to the fact it is very quick and simple for the end user to access the application from a variety of systems. The nature of the project also involves large and numerous media files, which include uncompressed high-resolution images and encoded high-resolution videos. By utilizing a web-based application users can connect from anywhere and browse the data using a compatible web browser, which acts as a “thin-client”.

4.5.2 HTML5 Video Playback

There are many ways to allow video playback on the web browser, but most solutions are dated and have compatibility and performance issues or require the use of 3rd party plugins for playback. HTML5 was designed to answer the need for a unified way to deliver media-rich content to users in an efficient way. HTML5 also provides us a way to interact with the media on the presentation layer using JavaScript and CSS to have more control over how we want to display the data to the viewer. This is useful for building a more intuitive interface to browse the video using a specialized timeline viewer that highlights frames and scenes of interest either derived from the post-processing algorithms, or from the end-user’s annotations.

4.5.3 Querying and Annotation

A key aspect to the application front end is the ability to annotate and apply metadata to frames and/or scenes. This provides a multitude of ways that the data browser can mark and subsequently, re-visit frames or scenes of interest. The front end is designed with this in mind and provides easy and intuitive ways to annotate scenes or frames. All of this data is relationally mapped in the MySQL database and can be queried upon by the user(s) of the system.

4.5.4 Use of popular tools

By overlaying GPS coordinates that are associated in the metadata of the frames, it is possible to overview the transect lines as well as the detected objects (sea turtles) on an interactive earth map using the Google Earth APIs.

This provides a more visual and intuitive summary of the findings of the image analysis system and also helps map objects of interest over geographic space.

5. Concluding Remarks

A system capable of capturing and analyzing images and a platform for browsing and querying is designed and setup. The system fulfills the challenge of running without any human intervention in a completely automated mode. Initial experiments to test the feasibility of the solution were conducted and it proved to be very solid. Currently, a real test in an airplane fly has not been performed yet and test is pending regulatory approval required before recording endangered species. The current tests have been with video of the ocean taken from a drive by car. Once the first flight is conducted and the first footage collected, the designed system will be refined for a proper implementation based on the results and findings. With the progress of the current project, we expect to gain even more learning from this valuable experience and share all the knowledge we achieved to help future researchers who need to face similar challenges.

References

Buckland, S.T. et al. (1993). ‘Distance Sampling, Estimating Abundance of Biological Populations’

Caughley, G. (1977), 'Sampling in Aerial Survey', The Journal of Wildlife Management 41(4), 605--615.

Cardona, L. et al. (2005), 'Western Mediterranean immature loggerhead turtles: habitat use in spring and summer assessed through satellite tracking and aerial surveys', Marine Biology 147(3), 583-591.

Chester, A. et al.(1995), 'Aerial surveys for sea turtles in North Carolina inshore waters', Fishery Bulletin 93(2), 254-261.

Elko, N. A. et al. (2005), 'Quantifying the Rapid Evolution of a Nourishment Project with Video Imagery', Journal of Coastal Research 21(4), 633-645.

Hobbs, R.C., et al. (2000).' Beluga, Delphinapterus leucas, group sizes in Cook Inlet, Alaska, based on observer counts and aerial video'. Marine Fisheries Review. 62(3):46-59.

Authorization and Disclaimer