University of Arkansas – CSCE Department
Capstone I – Final Proposal – Fall 2021

Automatic Drone Tracking

Student Team:

Andre Fuentes, Parker Weber, Zachery Gansz, Byron Denham, Corbett Stephens

Champion Advisor:

Prof. Khoa Luu

Abstract
The aim of this project is to bring facial tracking technology to drones. This will be accomplished by the development of a computer vision algorithm for facial detection that will interface with and control the drone. The drone's camera will communicate with a Haar cascade object detection model to gather information about the presence and location of a human face and make decisions about how the drone should be maneuvered to locate and track the face. The development of the technology in this project will have benefits for surveillance, tracking, and the film industry.

1.0 Problem

The integration of facial detection and tracking technology with drones will help give further practical use to existing ideas and technologies in the area of artificial intelligence. This combining of the two areas will help to automate some of the uses drones already have. One example of this is for surveillance. Because of their mobility, drones are already useful for surveillance, but the integration of facial tracking will improve the ability to locate and track criminals. Drones used for this type of tracking would no longer need to be controlled manually and could instead operate automatically. Another area it would be useful for is the film industry. This technology can automate the tracking of actors and actresses in shots taken with drones. One final benefit would be for presenting in a classroom setting. Drones with this facial tracking ability would be able to follow a presenter as they move around during a presentation.

2.0 Objective

The objective of this project is to use computer vision technology to give a drone the ability to recognize and track a human face. The drone will be able to maintain a defined distance from the target's face in order to ensure proper tracking. The drone will accomplish this by using the camera included on it. The drone will respond to and use the Tello API in order to maneuver itself. The developed facial recognition software will work alongside an expert system to interpret the still frames of the drone camera's video and issue responses. The final deliverables will include the facial recognition software, the drone control system (the expert system), and the drone with the attached camera.

3.0 Background

3.1 Key Concepts

A drone is an aircraft without any human crew on board. The drone that is used in this project is a small DJI Tello drone. This drone is a Consumer Off the Shelf (COTS) drone. There is another COTS drone that is used as a backup in case of any damages or hardware malfunctions in the DJI Tello drone. The backup drone is a larger Parrot drone. The decision to move forward with the DJI Tello rather than the Parrot stemmed from two main reasons. The first reason is that the DJI Tello drone has newer hardware, software and it is in better condition.

The second reason is that the DJI Tello is smaller with less powerful propellers. Powerful propellers can cause serious harm to objects and people if used improperly. The small body size and propellers of the DJI Tello makes flight testing much safer and easier in a small lab setting. These drones are powered by rechargeable lithium polymer batteries and are made for producing aerial footage. The standard drone is controlled via a radio controller that gives the pilot the ability to remotely maneuver the aircraft within a specified range. These drones come stock with a camera. The camera is used as input data for an algorithm to detect a human's face. The algorithm that is used to detect a human's face is a Haar cascade. The subsequent section will go into more detail about Haar cascades. The DJI Tello drone does not have adequate hardware to run the computation necessary to identify faces on its own. Thus, the drone will need to transmit its camera feed to a computer that is more capable. The algorithm will be able to identify a human face within a video feed processing 20 frames per second. A high frame per second will reduce the latency between the human face's movement and the drone's movement. The DJI Tello drone has an API available that will allow us to view the camera's image and send flight commands remotely via python code. An expert system written in Python will guide the drones flight and keep the human face in center frame. Putting all of these pieces together, the drone will have the capabilities to detect a human face and maneuver itself autonomously in order to keep the face in center frame.

3.2 Related Work

Detectors are a widely discussed topic in machine learning. To be more specific, the detectors being referred to are object detectors that use Haar cascade classifiers. Haar cascade classifiers are a machine learning based approach where a cascade function is trained from a large amount of positive and negative images. The positive and negative images consist of images with faces and without faces respectively. Haar features are used to determine features of an object. An object like the face is assigned Haar features that represent the most relevant features of the face. The Haar features are composed of white and black zones. The arrangement of the white and black zones can determine edge features, line features, and four-rectangle features. For an image, a window of pixels is examined in which a feature is applied, and a single value is obtained by subtracting the sum of the pixels under the white zone from the sum of pixels under the black zone. The integral image method is used to provide accuracy across multiple iterations without the cost of efficiency. The integral method works by using a singular value that represents the sum of a large region. The closer the value is to one, the more likely it is that a Haar feature has been detected. To speed up the process Adaboost (adaptive boosting) is used to select the best features among the calculated features. Adaboost uses several weak classifiers to create one strong classifier. It works by adjusting weights of correctly classified and misclassified items until the error is of satisfactory degree. Once a certain degree of error is achieved, a strong classifier can be made. The concept of Cascade of Classifiers can be used to make the process even more efficient. The features are grouped into different stages of classifiers where they are applied one-by-one rather than applying every single feature to a window every time [1]. This improves the efficiency specifically if a window fails in the first stage. In this case, the remaining stages are not tested. If a window passes each stage, then it is considered an object region such as a face region. The Haar cascade classifier, trained on detecting faces, gives a high efficiency method for accurately detecting faces in a video.

Detecting faces from video is a hard task, and there are many variables that come into play in non-lab settings. In lab settings, you can achieve very consistent and optimal lighting of the individual's face. This helps the model detect the face, but in the real world things become much more challenging. Some challenges that can arise are low light settings, over exposure, different skin colors, poor camera quality, complex backgrounds. We may also have issues with the distance from the camera and the orientation [2]. To overcome these issues the system will utilize a high quality camera with automated focus and exposure settings to extract the most accurate information from the scene. The Haar cascade algorithm will be trained on the Flickr-Faces-HQ dataset because it has more variation in terms of age, ethnicity, quality, lighting and background than any other dataset currently available. The variation will improve the drones ability to generalize to many more faces and settings.

The famous company DJI, based in China, has had object tracking in their drones since 2016. They produce the best COTS drone systems that autonomously track objects in motion. This system functions extraordinarily well, but it does come at some cost. The Phantom 4 Pro V2.0, for example, costs around $2,000 for the base model. The premium price tag is due to the thorough research and development put into the software along with the extra sensors of the drone. The DJI drone uses GPS and a forward and downward vision system that uses many sensors to achieve this task [3]. Our implementation will be a low cost tracking solution that will utilize one camera and a flight controller to achieve facial tracking.

4.0 Design

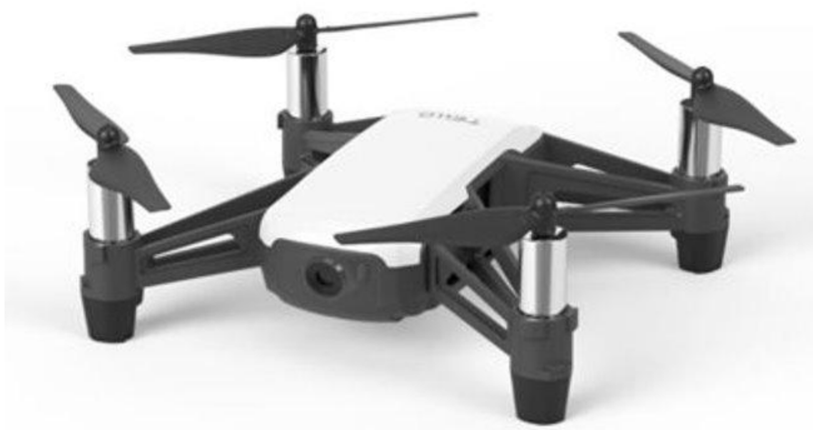
4.1 Requirements and/or Use Cases and/or Design Goals

- High Quality Camera - to provide a high resolution image, as well as function in low light scenarios and be able to clearly depict a person's face from roughly 5 to 10 feet away.
- Drone Control System - the backend must be able to properly interpret data given from the facial recognition software in order to have the correct move commands given.
- Drone Stability - the drone must be able to provide smooth and consistent movement.
- Facial Recognition Software - the system must be able to correctly identify human faces of all races and overcome some facial obstructions such as glasses.
- Environment adaptability - the drone must be able to simultaneously track a human face while also avoiding environmental collision/disruption.
- Processing Power - enough computing power to issue at least 20 commands per second, as the system will only be receiving about 20 frames per second of information.

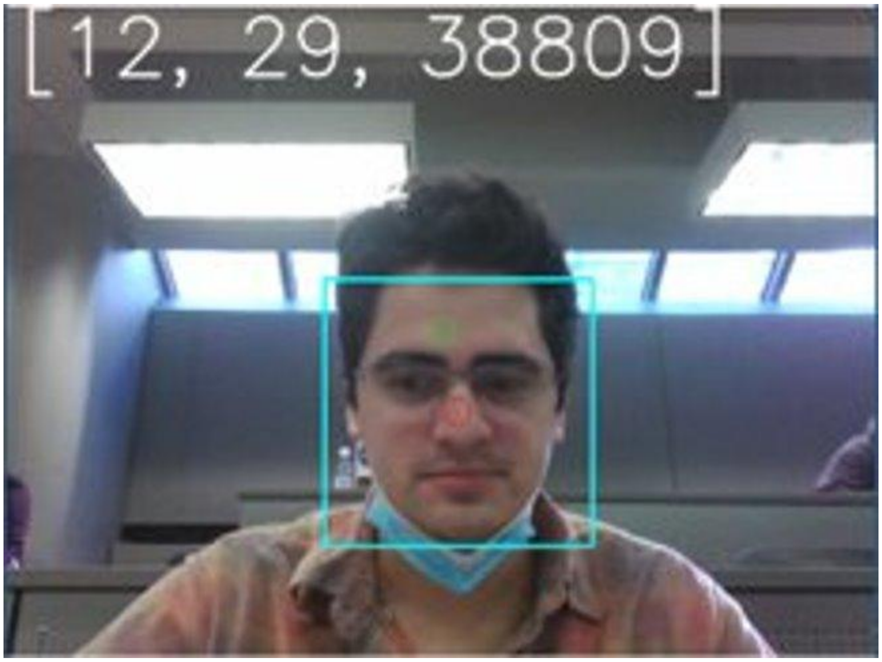
4.2 High Level Architecture

The system will contain two main parts - a drone, and an artificial intelligence backend. The two will communicate such that the drone sends real-time information to the backend, and the backend interprets this real-time information, processes it, and then instructs the drone on its next actions. Together, these two pieces will enable the creation of a system that has capabilities beyond normal, on-board drone systems.

The inputs for the design are a drone and a camera. The drone will use the Tello API found online to help enable enhanced functionality. Code and apps will be developed using this API to instruct the drone on how to move. This information and list of instructions will come from the back-end. The camera located on the drone will record video and send it to the back-end system for processing in real-time. This establishes the foundation of the relationship between the moving parts of the system (the drone and camera components) and the stationary parts of the system (the artificial intelligence backend located on the desktop computer).



The back-end to the system will be a desktop computer that contains a Haar cascade classifier. This part will also contain the API that will help it interact with the drone. Real-time information received from the drone will be saved and processed. The computer vision algorithm, Haar cascade, and an expert system will be utilized in order to both control the drone's actions through API calls and to process the received video frames, frame-by-frame, to create real-time tracking of a human subject's face. The expert system will interpret the information provided to it from the computer vision algorithm and make decisions on how the drone should align itself in the air. The result of these decisions will move the drone in order to properly track its target and maintain the ability to utilize facial recognition. These methods for both providing the drone its instructions and processing/saving incoming information will be taught in tandem using a training database developed and optimized during the implementation phase. The back-end of the system (the desktop computer that contains the implemented Haar cascade classifier and expert system) will remain stationary, with its primary purpose being, as described above, processing and providing direction for the moving parts of the system.



The outputs of the design are real-time movement instructions for the drone and saved information from previous runs. The finished product should be able to instruct a drone to properly track a human subject in order to maintain a line of sight on their face with a camera (attached to the drone). The system will perform a form of facial recognition in order to automatically continue this tracking process. Data obtained from this process will be saved on the back-end part of the system for recordkeeping and improvement purposes.

4.3 Risks

Risk	Risk Reduction
Injury by drone	Limit how close the drone can get to someone
Radio Frequency Interference	Add Radio frequency filters/relocate equipment

4.4 Tasks

Tasks	Personnel
Finish final proposal	Everyone
Read Haar cascade documentation	Andre, Byron, Corbett
Read Tello drone documentation	Parker, Zachery
Develop a process to allow for fluent connectivity between drone and computer	Parker, Zachery
Define input/output of face detection model and expert system	Everyone
Develop a process to stream video from drone camera to computer	Parker, Zachery
Develop a process to detect human face using face detection model	Andre, Byron, Corbett
Develop a program to detect data files from drone	Andre, Byron, Corbett
Create an expert system that reads data files and issues response to drone	Andre, Byron, Corbett
Create the set of potential responses to the drone (how it will move in response to human movement)	Parker, Zachery
Create set of tests for a live test	Everyone
Live test the system	Everyone
Debug the system	Everyone
Create presentation for the project	Everyone
Present the project	Everyone

4.5 Schedule

Tasks	Dates
Finish final proposal	11/29
Read Haar cascade documentation	11/30-12/6
Read Tello drone documentation	11/30-12/6
Winter Break	12/17-1/18
Develop a process to allow for fluent connectivity between drone and computer	1/18-1/24
Define input/output of face detection model and expert system	1/24-1/31
Develop a process to stream video from drone camera to computer	1/31-2/7
Develop a process to detect human face using face detection model	2/7-2/21
Develop a program to detect data files from drone	2/21-2/28
Create an expert system that reads data files and issues response to drone	2/28-3/14
Create the set of potential responses to the drone (how it will move in response to human movement)	3/14-3/28
Create set of tests for a live test	3/28-4/6
Live test the system	4/6-4/7
Debug the system	4/7-4/14
Create presentation for the project	4/14-4/21
Present the project	TBD

4.6 Deliverables

- Design Document: A description of each component, including the code for facial recognition, the drones API calls, and the drone.
- Facial Recognition Software: The Python program used to take still-image data from a real-time video and identify human faces, informing the user of region of interest (ROI) coordinates and ROI area relative to the camera.
- Drone Control Software (Expert System): The code that contains the Tello API calls in order to correctly direct the drones movement based on incoming data.
- Drone and camera - The drone and attached camera that were used to provide input information for the Facial Recognition Software.
- Final Report - The final report for the project that includes detailed design architecture, implementation, and technologies used to complete the project.

5.0 Key Personnel

Andre Fuentes – Fuentes is a senior Computer Science major in the Computer Science and Computer Engineering Department at the University of Arkansas. He has completed Artificial Intelligence, Data Mining, Algorithms and Software Engineering. He also works in the Computer Vision Lab under Dr. Ngan Le as a research assistant. Fuentes will be responsible for contributing to the development of the AI model that detects human faces.

Parker Weber – Weber is a senior Computer Engineering and German major in both the College of Engineering and Fullbright College at the University of Arkansas. He has completed Software Engineering, Computer Organization, System Synthesis and Modeling, and Operating Systems. He has worked with full-stack development in order to control full simulation runs of prototype hardware systems. Weber will be responsible for contributing to the development of the drone’s signal sending/receiving capabilities.

Zachery Gansz - Gansz is a senior Computer Engineering major in the Computer Science and Computer Engineering Department at the University of Arkansas. He has completed Computer Organization, System Synthesis and Modeling, Operating Systems, Software Engineering, and GPU Programming. Gansz will be responsible for contributing to the development of the drone’s signal sending/receiving capabilities.

Corbett Stephens – Stephens is a senior Computer Science major in the Computer Science and Computer Engineering Department at the University of Arkansas. He has completed Algorithms, Cloud Computing and Security, Computer Networks, and Software Engineering. Stephens will be responsible for contributing to the development of the AI model that detects human faces and calculates adjustments to update the drone’s positioning.

Byron Denham - Denham is a senior Computer Science major in the Computer Science and Computer Engineering Department at the University of Arkansas. He has completed Artificial Intelligence, Data Mining, Software Engineering, and Algorithms. He has also been part of the summer Application Development internship at J.B. Hunt. Denham will be responsible for contributing to the development of the AI model that detects human faces.

Dr. Khoa Luu (Champion) - is currently an Assistant Professor and the Director of Computer Vision and Image Understanding (CVIU) Lab in the Department of Computer Science and Computer Engineering at University of Arkansas, Fayetteville. He is an Associate Editor of IEEE Access Journal. He is teaching Computer Vision, Image Processing and Introduction to Artificial Intelligence courses in CSCE Department at University of Arkansas, Fayetteville. His research interests focus on various topics, including Biometrics, Face Recognition, Tracking, Human Behavior Understanding, Image and Video Processing, Deep Learning and Quantum Machine Learning. He has received four patents and two best paper awards and co-authored 100+ papers in conferences, technical reports, and journals.

6.0 Facilities and Equipment

This project will require a DJI Tello drone, a lab space for flying the drone, and a computer for software development. We already have an available drone that we can use. The lab space that we are using is the Computer Vision and Image Understanding (CVIU) lab in JBHT room 447. There are computers in the lab that can be used for software development or personal computers can be used. At the moment, there is not any software that needs to be purchased.

7.0 References

[1] Schapire, Robert E. "*Explaining AdaBoost"*. Princeton University, https://www.cs.princeton.edu/~schapire/papers/explaining-adaboost.pdf

[2] Kumar, Ashu & Kaur, Amandeep & Kumar, Munish. (2019). Face Detection Techniques: A Review. Artificial Intelligence Review. 52. 10.1007/s10462-018-9650-2

[3] DJI Mavic Pro Specs https://www.dji.com/mavic/info

[4] "Cascade Classifier." *OpenCV*, https://docs.opencv.org/3.4/db/d28/tutorial_cascade_classifier.html