# AER1517 Winter 2024

# Lab3 - Georeferencing Using UAV Payload Data

Chris Anderson, Kevin Hu, Yifeng Han

## Introduction

This lab focuses on utilizing the bottom-facing camera of a UAV to detect circular targets on the ground, a fundamental task in georeferencing and spatial analysis. This lab explores the nuances of converting 2D image data into precise 3D world coordinates. By utilizing the Parrot AR.Drone 2.0's camera capabilities, alongside data from the Vicon motion capture system, the lab aims to pinpoint the locations of specific targets within a defined area.

The core challenge lies in the accurate georeferencing of these targets, requiring a sophisticated image processing algorithm that can translate pixel coordinates to georeferenced coordinates based on the UAV's position and orientation. The provided dataset, essential for this task, contains the UAV's pose data—its position and orientation across various timestamps, captured by the Vicon system. This data is crucial for mapping image pixels to real-world coordinates, facilitating the accurate identification of target locations on the ground. The lab encourages the application of real-time image processing, supported by tools like the OpenCV library, to enhance learning and practical experience in UAV-based data analysis.

## Methodology

The methodology outlined in the provided Python script represents our approach to georeferencing circular targets detected in UAV imagery. First, we remove the camera distortion from the images using OpenCV's undistort function. This is useful because we no longer have to deal with camera distortion within the problem. The process commences with image preprocessing, where images are converted to grayscale and then thresholded to isolate white rectangles based on pixel color intensity. This simplification is crucial for the subsequent contour detection step, which employs OpenCV's capabilities to identify shapes within the thresholded image. Contours are sorted by area, as a maximum of two targets of interest lie within their corresponding contour areas in one scene. Contours with the top two largest areas are taken for further processing.

The contoured images are blurred to reduce noise, followed by circle detection, achieved through Hough Circle Transformation. This specificity ensures that subsequent centroid calculations focus exclusively on the intended circular targets. The conversion of pixel coordinates to world coordinates is a critical next step, leveraging the UAV's pose data (position and orientation) and the camera's intrinsic parameters. A transformation matrix, derived from these parameters, maps the 2D pixel coordinates to accurate 3D world coordinates, facilitating precise target geolocation.

For the pixel to world coordinate transformation, we leverage the pixel to world coordinate equation given below.

$$S \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K \begin{bmatrix} R & 1 & t \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

Equation 1: Pixel from World Coordinates Equation

It uses the camera intrinsic matrix K, the modified camera extrinsic matrix [R | t], and a scaling factor S. Note that the camera extrinsic changes over time in this problem since the camera is attached to a moving UAV. To get the camera extrinsic matrices at each time step we first require the poses of the UAV in 4x4 transformation matrix form. We convert the quaternion orientations of the UAV to 3x3 rotation matrix form and create 4x4 transformation matrices from the rotation matrix and the Cartesian position coordinates of the UAV at each time step. Then we transform the pose of the UAV at each time step with the body to camera frame transform by doing matrix multiplication. We then have the camera extrinsic matrix at each time step and have the modified version by just removing the bottom row.

$$K^{-1}\begin{bmatrix}u\\v\\1\end{bmatrix} = \frac{1}{S}\begin{bmatrix}R & 1 & t\end{bmatrix}\begin{bmatrix}x\\y\\z\\1\end{bmatrix} \qquad K^{-1}\begin{bmatrix}u\\v\\1\end{bmatrix} = \frac{1}{S}(R\begin{bmatrix}x\\y\\z\end{bmatrix} + t)$$

$$K^{-1}\begin{bmatrix}u\\v\\1\end{bmatrix} = R\begin{bmatrix}\frac{x}{S}\\\frac{y}{S}\\0\end{bmatrix} + \frac{t}{S} \qquad K^{-1}\begin{bmatrix}u\\v\\1\end{bmatrix} = R\begin{bmatrix}\frac{x}{S}\\\frac{y}{S}\\0\end{bmatrix} + \frac{t}{S} \quad a = \frac{x}{S}, b = \frac{y}{S}, c = \frac{1}{S}$$

$$\begin{aligned}K_1^{-1}u + K_2^{-1}v + K_3^{-1} &= R_1a + R_2b + t_1c\\ K_4^{-1}u + K_5^{-1}v + K_6^{-1} &= R_4a + R_5b + t_2c\\ K_7^{-1}u + K_8^{-1}v + K_9^{-1} &= R_7a + R_8b + t_3c\end{aligned} \qquad \begin{bmatrix}R_1 & R_2 & t_1\\R_4 & R_5 & t_2\\R_7 & R_8 & t_3\end{bmatrix}\begin{bmatrix}a\\b\\c\end{bmatrix} = K^{-1}\begin{bmatrix}u\\v\\1\end{bmatrix}$$

$$A = \begin{bmatrix}R_1 & R_2 & t_1\\R_4 & R_5 & t_2\\R_7 & R_8 & t_3\end{bmatrix} \qquad \begin{bmatrix}a\\b\\c\end{bmatrix} = A^{-1}K^{-1}\begin{bmatrix}u\\v\\1\end{bmatrix} \qquad S = \frac{1}{c}, x = \frac{a}{c}, y = \frac{b}{c}, z = 0$$

Figure 1: Step-by-step derivation of calculating landmark world coordinates from pixel coordinates

The step-by-step derivation of calculating the landmark world coordinates from pixel coordinates is presented in the figure above from right to left, starting from the Pixel to World Coordinates Equation. The derivation leverages the fact that the z coordinate is assumed to be zero in this problem, so a system of equations in the form Ax = b where A is invertible can be formed to find x, y, and the scaling factor S which are unknown. Otherwise, the world coordinates would not be possible to determine.

Initially, all landmark positions were calculated for every frame that had detections and plotted. To enhance the accuracy of the detected target locations, the script applied K-means clustering. K-means clustering groups the estimated target positions, effectively distinguishing individual targets amidst noise and multiple detections. However, it was concluded that the estimated landmark locations were not accurate as they were widely dispersed across the target area and it was difficult to determine accurate clusters. The estimated landmark locations sometimes appeared in the center of the groups of landmark locations rather than on top of them.

Another method was then used to capture the time correlation of the landmarks. The method split all roughly continuous frames of detections into landmark clusters. Roughly continuous means a gap of 5 or less frames between detections is allowed, and it would still be part of the same cluster. Due to the maximum number of detections at the same time being two, the clusters were subdivided into clusters that only had singular detections per frame and clusters that had at least one non-singular detection in

one of the frames in the cluster. The logic is that the non-singular detections indicate that there are two separate landmarks and two separate estimates that need to be performed. For the singular clusters, the landmark centroids of each cluster were found, and the weight or number of landmark contributions of each cluster was recorded. For the multi-detection clusters, two separate lists were created for two separate landmarks for each cluster. The first list contained all the landmarks until the first multi-detection was reached. Then the two new landmarks compared their distances to the landmark centroid of the first list and whichever one was closer to the landmark centroid went to the first list and the other one went to the second list. The process would continue until all the landmarks for the cluster were completed. The centroid of the two lists would then be computed for each cluster separately. The weights of each centroid were recorded as well. All the computed centroids were then compiled into a singular list alongside a list of weights. The new method dramatically reduced the number of centroid landmarks. Weighted K-means clustering was applied to the centroids landmarks to obtain the final landmark locations.

This comprehensive approach integrates computer vision techniques for target detection, mathematical transformations for coordinate conversion, and optimization techniques for refining target locations. The methodology exemplifies the effective use of UAV imagery for precise target geolocation, leveraging the UAV's pose data and the camera's intrinsic parameters to achieve accurate georeferencing of detected targets, essential for applications like environmental monitoring and search and rescue missions. The full code is provided in the Appendix.

## Results

The execution of the Python script for georeferencing circular targets from UAV imagery resulted in accurately locating these targets through a transformation from pixel to world coordinates. The application of K-means clustering led to precise adjustments in the landmark coordinates, significantly enhancing georeferencing accuracy. The scatter plots below show the methodology's effectiveness.

Table 1: Final Calculated Landmark Positions

| Landmark ID | X Coordinate | Y Coordinate | Z Coordinate (Expected to be 0) |
|---|---|---|---|
| 1 | -0.45005703 | 1.6464428 | 0 |
| 2 | 0.35637736 | -0.52204213 | 0 |
| 3 | 0.53502736 | 0.4492091 | 0 |
| 4 | -0.46044608 | -0.89620348 | 0 |
| 5 | -1.28310889 | 0.44288282 | 0 |
| 6 | 1.11390984 | 0.8820557 | 0 |

The tables provide a clear and concise summary of the calculated landmark positions. Precision is critical for various applications such as mapping, navigation, and spatial analysis, showcasing the methodology's potential in improving target geolocation accuracy for environmental monitoring, search and rescue missions, and more. The results underscore the script's success in achieving precise georeferencing of detected targets.

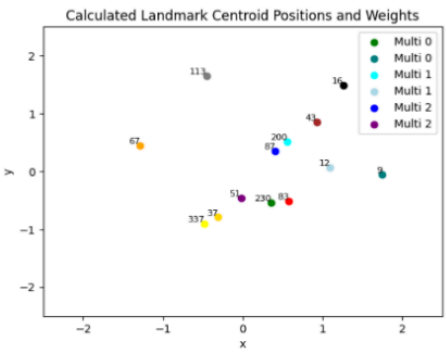


Figure 2: Calculated Landmark Centroid Positions and Corresponding Weights

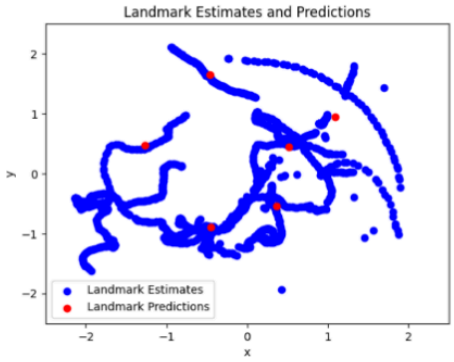


Figure 3: All Landmarks and Landmark Predictions

The visualization through scatter plots and the summarized data and shown in Figure 2 and 3. This precision is pivotal for applications requiring accurate geolocational data and validates the potential of the developed methodology for enhancing the accuracy of target geolocation in various scenarios, including environmental monitoring and search and rescue missions. The clear improvement in the accuracy of target positions post-optimization highlights the script's success in processing UAV imagery to achieve precise georeferencing of detected targets.

## Discussions

Discussion on Methodology Effectiveness

The methodology demonstrated a relatively high level of accuracy in converting 2D image data into precise 3D world coordinates, a task complicated by the inherent challenges in distinguishing targets from varied backgrounds and conditions. The use of grayscale conversion, thresholding, contour detection, and Hough Circle Transformation in the image preprocessing phase effectively isolated the circular targets for further analysis. This success highlights the potential of integrating computer vision techniques with UAV technology for spatial analysis tasks.

Implications of K-means Clustering

The application of K-means clustering technique played a pivotal role in enhancing the accuracy of the georeferenced targets. K-means clustering effectively managed to distinguish individual targets amidst potential noise and multiple detections, which is crucial for processing images with multiple or overlapping targets. This approach not only reduced the overall error in the model but also aligned the estimated positions of the landmarks closer to their real-world locations.

Potential Applications and Future Work

Despite the success, there is room for improvement and expansion. Future work could explore the integration of additional data sources, such as satellite imagery or other sensor data, to enhance the robustness and accuracy of the georeferencing process. Exploring real-time processing capabilities could further extend the methodology's applicability in dynamic environments where rapid data analysis is critical.

## Conclusions

The results and methodology discussed herein demonstrate an advancement in UAV-based georeferencing tasks, validated by the selection target positions determined by K-means clustering. The result of this lab highlights the potential of using sophisticated image processing algorithms alongside

UAV technology for precise spatial analysis. The methodologies developed in this lab can be further improved to more efficient and accurate geospatial data collection and analysis.

## References


1. Liu, Hugh, “AER1517: Development of Autonomous Unmanned Aerial Systems”, Course Lectures, Winter 2024, University of Toronto
2. Liu, Hugh, “Lab 3: Georeferencing Using UAV Payload Data”, AER1517 Course Notes, Winter 2024, University of Toronto
3. Nex, F., & Remondino, F. (2014). UAV for 3D mapping applications: A review. Applied Geomatics, 6(1), 1-15.
4. Colomina, I., & Molina, P. (2014). Unmanned aerial systems for photogrammetry and remote sensing: A review. ISPRS Journal of Photogrammetry and Remote Sensing, 92, 79-97.