Frisbee Detection and Tracking in Ultimate Frisbee Footage

1st Noah Walle
School of Mechatronics Engineering
University of Canterbury
Christchurch, New Zealand
nwa470@uclive.ac.nz

2nd Richard Green
School of Computer Engineering
University of Canterbury
Christchurch, New Zealand
richard.green@canterbury.ac.nz

***Abstract*—This paper proposes a method for detecting and tracking frisbees in video footage of ultimate frisbee games. The approach combines traditional image processing techniques with object tracking algorithms and a custom-trained deep learning model. Preprocessing techniques such as Gaussian blurring, Canny edge detection, blurring, and thresholding are employed to enhance feature visibility in video frames. Classical tracking methods are evaluated, with the Channel and Spatial Reliability Tracker (CSRT) demonstrating higher robustness than Kernelized Correlation Filters (KCF) in handling moderate occlusion and motion. To further improve detection accuracy, a convolutional neural network is trained on a dataset of 6,750 manually annotated images, achieving a confidence threshold of 65% for frisbee classification. These results show that the proposed approach can support automated analysis of ultimate frisbee footage and enable applications in sports analytics.**

***Index Terms*—Object Tracking, Ultimate Frisbee, CSRT Tracker, Custom Dataset, Small Object Detection, Real-Time Processing**

I. Introduction

Automated detection and tracking of flying discs in ultimate frisbee presents unique challenges due to the frisbee's high-speed motion, variable orientation, and frequent occlusion by players. Accurate localization of the frisbee is essential for trajectory analysis, game-play event recognition, and performance evaluation. However, the small size and intermittent visibility of the object make this task non-trivial, especially in unconstrained environments such as full-field video footage.

This paper proposes a method that combines handcrafted feature extraction and tracking with learned representation-based classification to achieve accurate frisbee localization over time. The method is designed for real-world sports footage where lighting, background, and scale vary significantly. The proposed approach aims to bridge classical image analysis techniques and modern deep learning pipelines to improve robustness and scalability.

II. Background

Object detection and tracking are crucial techniques in modern video analysis systems, particularly in sports analytics where object speed and occlusion frequently challenge detection robustness. In recent years, deep learning-based methods have significantly outperformed classical techniques in detection accuracy and adaptability across diverse scenes [1], [2].

A. Object Detection in Sports

Traditional object detection methods, including background subtraction and optical flow, have been employed to track players and objects in sports settings. However, these techniques often struggle with complex scenarios involving occlusions and varying lighting conditions. The advent of deep learning has introduced more robust solutions, with models like YOLO (You Only Look Once) and SSD (Single Shot MultiBox Detector) offering real-time object detection capabilities. These models have been successfully applied to sports like soccer and basketball, enabling real-time analytics and strategy formulation.

In the context of ultimate frisbee, the application of these models presents unique challenges. The frisbee's small size, high velocity, and frequent occlusions by players make it a difficult object to detect consistently. Moreover, the lack of large, annotated datasets specific to ultimate frisbee further complicates the training and evaluation of deep learning models in this domain.

Traditional image processing methods such as Gaussian blurring and Canny edge detection were initially used for motion-based segmentation [3], [4]. While useful for preprocessing and highlighting edges, these methods lack semantic understanding and suffer under variable lighting and cluttered backgrounds [5].

The Kalman filter and its variants have historically been employed for object tracking due to their low computational cost [6]. However, their performance degrades in the presence of nonlinear motion and abrupt trajectory changes [7]. More advanced trackers like KCF (Kernelized Correlation Filters) improve robustness through efficient correlation operations [8], but tend to struggle with scale variation and occlusion [9]. CSRT (Discriminative Correlation Filter with Channel and Spatial Reliability) offers enhanced adaptability by incorporating spatial reliability maps, making it more suitable for fast-moving and partially occluded targets [10].

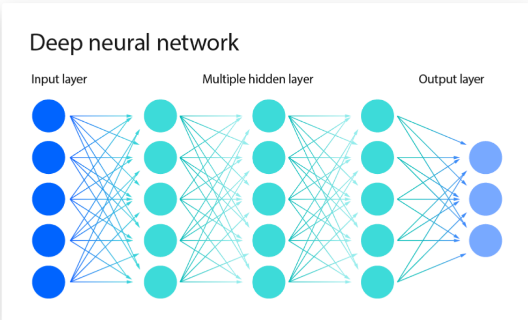

Fig. 1: A deep neural network architecture consisting of an input layer, multiple hidden layers, and an output layer. Each layer transforms the input through a series of weighted connections and nonlinear activations to extract increasingly abstract features [16].

B. Tracking Techniques

Object tracking in sports has evolved from simple centroid tracking to more sophisticated algorithms like the Kalman Filter, Mean Shift, and correlation-based trackers. Among these, the Discriminative Correlation Filter (DCF) based trackers, such as the Kernelized Correlation Filter (KCF) and the Channel and Spatial Reliability Tracker (CSRT), have shown promise due to their balance between accuracy and computational efficiency.

However, these trackers are not without limitations. KCF, for instance, assumes a fixed object scale and may struggle with scale variations and occlusions. CSRT improves upon KCF by incorporating spatial reliability maps, enhancing its robustness to occlusions and scale changes. Yet, both trackers can falter in scenarios involving rapid object movements and complex backgrounds, common in ultimate frisbee matches.

C. Deep Learning Approaches

Recent advances in object detection have been driven by convolutional neural networks (CNNs), particularly the YOLO (You Only Look Once) family [11]–[13]. YOLOv5 and YOLOv8 offer real-time performance with reasonable accuracy, but small object detection — like frisbees — remains challenging due to their size relative to the frame and limited temporal context [14]. Refer to Figure 1 for relevant CNN architecture.

Transfer learning, where models pre-trained on large datasets are fine-tuned on specific tasks, has also been employed to mitigate the scarcity of annotated data in niche sports like ultimate frisbee.

Despite these advancements, challenges persist. Deep learning models require substantial computational resources and large amounts of annotated data for effective training. In ultimate frisbee, the limited availability of high-quality, annotated datasets hampers the development of robust models. Furthermore, the variability in camera angles, lighting conditions, and player uniforms across different games adds to the complexity of model generalization.

Evaluating detection models typically involves metrics such as precision, recall, and mean Average Precision (mAP), which collectively offer insight into the detector's ability to minimize false positives and negatives [18], [19]. However, many works focus solely on detection performance without integrating or validating tracking capabilities, limiting their practical application in real-time systems [20].

Object detection and tracking have become integral components of sports analytics, enabling detailed performance assessments and strategic insights. These studies highlight a gap in models designed specifically for small, fast-moving sports equipment like frisbees, motivating an integrated approach that combines classical image processing, modern tracking algorithms, and deep learning for detection and tracking. While significant progress has been made in popular sports such as soccer and basketball, ultimate frisbee remains relatively underexplored in this domain. The sport's unique dynamics—characterized by rapid disc movements, frequent occlusions, and variable environmental conditions—pose distinct challenges that necessitate specialized approaches.

D. Existing Datasets and Tools

Dataset availability is another crucial factor influencing performance. COCO and ImageNet datasets, while comprehensive, offer limited annotations for niche objects such as frisbees in dynamic sports contexts [15]. Custom datasets, when annotated with domain-specific objects and bounding boxes, can significantly improve model specialization [17].

Efforts have been made to create datasets specific to ultimate frisbee. For instance, Roboflow hosts an open-source ultimate frisbee dataset comprising annotated images of players and frisbees. While such datasets are valuable resources, they often lack the diversity and scale required for training deep learning models capable of generalizing across various game scenarios.

Tools like Label Studio and LabelImg have facilitated the annotation process, enabling researchers to create custom datasets. However, the manual annotation of frames is time-consuming and prone to inconsistencies, especially when dealing with high-frame-rate videos and multiple objects.

In summary, while significant progress has been made in object detection and tracking within sports analytics, ultimate frisbee presents unique challenges that are yet to be fully addressed. The sport's dynamic nature, coupled with the frisbee's small size and rapid movement, necessitates specialized approaches and the development of comprehensive, annotated datasets. Future research should focus on creating robust models that can handle the complexities of ultimate frisbee, leveraging advancements in deep learning and data annotation tools.

III. Proposed Method

The proposed approach will utilize three key techniques to efficiently track players and the frisbee within the scope of this term project:

The CSRT (Channel and Spatial Reliability Tracker) uses a discriminative correlation filter to guess the location and appearance of the object-of-interest. The filter learns what the object looks like by making use of positive and negative training samples that update the filter iteratively during tracking. The tracker gets important features from these samples including Histogram of Oriented Gradients (HOG) and Scale-Invariant Feature Transform (SIFT).

Optical flow techniques, such as the Lucas-Kanade method, help track the movement of players and the frisbee across frames by estimating motion patterns. This method is particularly useful for identifying movement direction and velocity, which aids in tracking the frisbee's trajectory.

By applying background subtraction techniques, moving objects can be effectively segmented from the static field. This assists in isolating players and the frisbee from irrelevant background elements, improving tracking efficiency. A frisbee is particularly useful for thresholding operations due to its distinct white colour.

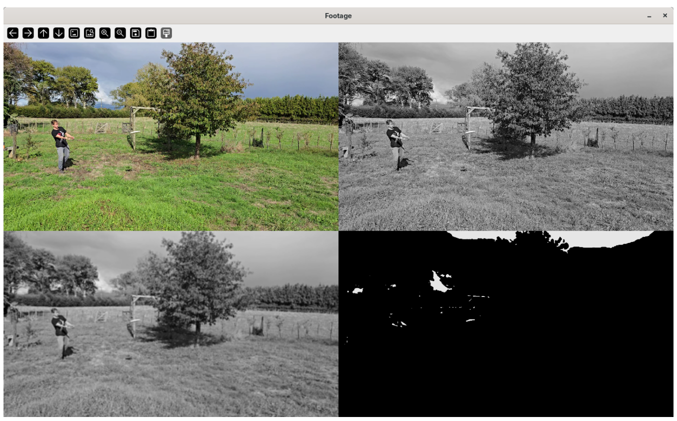

Fig. 2: Examples of the operations performed to make tracking easier: Original (top left), grayscale (top right), blur (bottom left), threshold (bottom right)

Deep learning uses AI to detect patterns in image datasets to try and determine what pixels make an object. Trained models can be used to classify objects, which is the goal of this paper.

The proposed method is composed of three primary components: image preprocessing, object tracking, and frisbee classification using a deep learning model.

A. Image Preprocessing

Each video frame undergoes preprocessing to reduce noise and isolate potential frisbee candidates. Gaussian blurring is applied to smooth textures and remove high-frequency noise. A binary threshold is applied to create a mask of high-contrast regions, which can correspond to the frisbee's edge features, especially in contrast with the playing field and sky. Frisbees are particularly easy to threshold due to their typically white colour.

B. Tracking

Two trackers are evaluated: KCF and CSRT. The KCF tracker, based on fast Fourier domain operations, shows limited performance due to its sensitivity to scale variation and motion blur. The CSRT tracker, which incorporates adaptive spatial reliability maps and channel-wise feature analysis, yields more stable tracking across short sequences. The CSRT tracker is selected as the primary tracking component based on visual stability and bounding box persistence during play.

C. Deep Learning-Based Detection

To support accurate identification of frisbees, a CNN is trained on a custom dataset of 6,750 labelled images extracted from ultimate frisbee games. The dataset is curated to include variations in lighting, scale, and frisbee orientation. The model is trained using a standard cross-entropy loss and evaluated on a held-out validation set. The trained model is then applied frame-by-frame to detect the frisbee and complement the output of the tracking module. Detections above 65% confidence are used to update tracking information and refine frisbee localization.

IV. Results

The machine used for the project was a 2015 Intel MacBook Air running Arch Linux. The CPU was a Intel i5-5250U (4) running at 2.7GHz and the GPU was an Intel HD Graphics 6000. The IDE used for the project was VSCodium, an Open Source binary of VS Code, and the code was written either in bash or Python. Source footage was drawn from films of ultimate frisbee games sourced from public footage on the internet. Links to the relevant games can be found in the Appendix. Other footage was shot on a Samsung Galaxy S20 that has an Ultra High Definition (UHD) camera with a resolution 7680 x 4320, also known as 8K, and a frame-rate of 24 fps. OpenCV version 4.11.0 was used.

Experiments were conducted on high-definition video footage from ultimate frisbee matches recorded in outdoor conditions. The evaluation considered detection accuracy, tracking consistency, and robustness to occlusion.

The CSRT tracker maintains stable bounding boxes across moderate occlusion events and sudden changes in motion. However, in frames with rapid frisbee movement or full occlusion by players, continuity is interrupted. In some cases, the tracker found success using a highly processed image to track. The tracking box could then be projected back onto the original unprocessed image. A frame showing the tracking pipeline is shown in Figure 3.

In deep learning models, the success of a trained model is usually measured by three key metrics; precision, recall, and mean average precision. The precision metric tells us the percentage of the time the model finds the object it is looking for correctly. The formula to determine precision is seen in Equation 1.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \tag{1}$$

Recall is the relationship between the amount of actual objects in the image that are being searched for, and the

Fig. 3: Frame from the tracking pipeline using the CSRT algorithm. The green bounding box highlights the predicted location of the frisbee during mid-flight. The tracker maintains object identity despite background clutter and motion.

amount that it found. The formula also takes into account false negatives. Recall is worked out using Equation 2.

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \tag{2}$$

Intersection over Union, also known as IoU, measures how well the predicted model overlaps with the ground truth. Equation 3 describes how it is calculated. The metric is used as a threshold value for the next key metric, mean average precision.

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}} \tag{3}$$

Mean average precision, also known as mAP, is used mostly when multiple classes are involved. Since the project's immediate aim was only to track the frisbee, only one class was required. Most models are already adept at identifying people, so custom models for this were not used. Mean average precision sums the average precision values for each class and takes the mean. The mean average precision formula is shown in Equation 4.

$$\text{mAP@0.5} = \frac{1}{N}\sum_{i=1}^{N} AP_i \tag{4}$$

Quantitative results show that the combined method detects the frisbee with a precision of 0.65 and recall of 0.36 over annotated sequences. Frames where the frisbee is fully visible and unobstructed show higher classification confidence and tracking stability. These results show that the proposed approach can successfully identify and localize frisbees in complex game-play footage, demonstrating the value of hybrid methods for object tracking in sports video. However, there are still frequent scenarios where the trained model has difficulty classifying the frisbee, indicating that there is room for improvement in this area.

TABLE I: Quantitative evaluation of frisbee detection and tracking.

Metric	**Value**
Precision	0.659
Recall	0.364
mAP@50	0.425

Over 6000 images were hand-annotated to build a custom deep learning model that would be able to classify frisbees in footage from games. These images were then split into 70% training data, 20% validation, and 10% test data. Figure 4, 5, 6, 7, 8, and 9 on the following page show the hand-annotated ground truth images on the left alongside the respective model predictions on the right. The images were selected from different game-play scenarios where the frisbee takes up a large, medium, and small part of the frame.

The combination of classical image analysis and learned detection provides a balance between speed and robustness. While traditional tracking algorithms offer frame-to-frame consistency, their limitations under rapid motion are mitigated by the deep learning model's capacity for re-identification. This is especially useful in situations where the camera angle changes and re-identification is almost always required.

The use of a domain-specific dataset was critical to improving detection accuracy, as general-purpose object detectors fail to distinguish frisbees from similarly shaped background elements. The modular nature of the pipeline also allows for future integration of temporal smoothing or predictive tracking to reduce frame-level inconsistencies.

V. Conclusion

This paper proposes a hybrid method for detecting and tracking frisbees in ultimate frisbee video footage using a combination of edge-based preprocessing, classical tracking algorithms, and a deep learning-based classification model. The method demonstrates consistent performance in varied game conditions and demonstrates the effectiveness of combining handcrafted and learned techniques.

A. Future Research

The CNN-based detection model will complement the CSRT tracker by re-initializing bounding boxes when confidence is high, particularly in frames where the tracker diverges or fails.

Future work will also explore integrating temporal attention mechanisms and recurrent models to improve performance under full occlusion. Expanding the dataset to include diverse weather conditions, player uniforms, and camera angles would improve generalizability. Additionally, combining multiple object detections (players and frisbees) may enable more comprehensive game-play analysis and event detection pipelines.

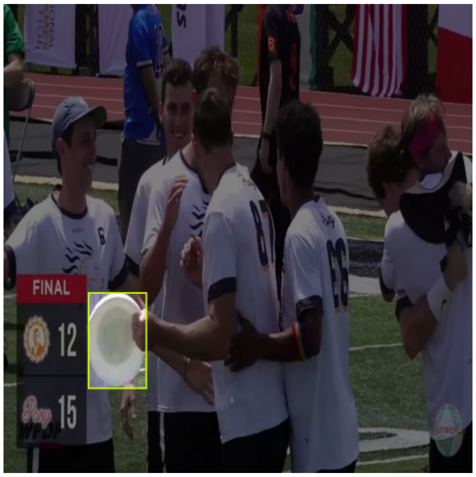

Fig. 4: Ground truth image 1

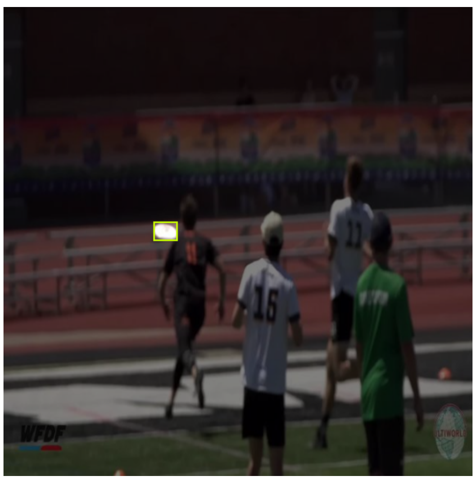

Fig. 5: Ground truth image 2

Fig. 6: Ground truth image 3

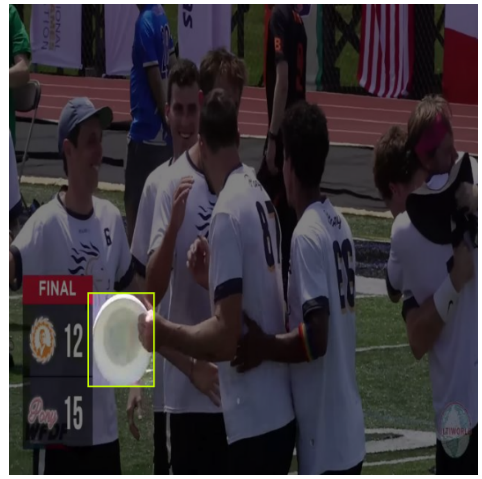

Fig. 7: Model prediction 1

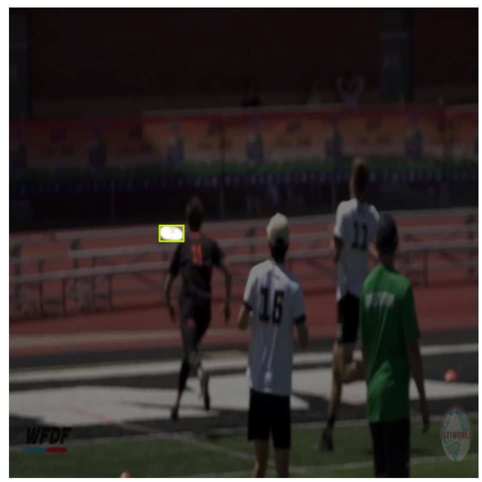

Fig. 8: Model prediction 2

Fig. 9: Model prediction 3

VI. References

[1] T.-Y. Lin et al., ”Feature Pyramid Networks for Object Detection,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2117–2125.

[2] J. Redmon and A. Farhadi, ”YOLOv3: An Incremental Improvement,” arXiv preprint arXiv:1804.02767, 2018.

[3] W. Liu et al., ”SSD: Single Shot MultiBox Detector,” in European Conference on Computer Vision, 2016, pp. 21–37.

[4] M. Danelljan et al., ”Discriminative Scale Space Tracking,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 8, pp. 1561–1575, 2017.

[5] R. Tao et al., ”Siamese Instance Search for Tracking,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1420–1429.

[6] J. M. McKee and A. Revalla, ”Computer Vision-Driven Ultimate Frisbee Tracking and Analytics,” Stanford University, 2019.

[7] J. Link et al., ”Wearable Sensors for Activity Recognition in Ultimate Frisbee Using Convolutional Neural Networks and Transfer Learning,” Sensors, vol. 22, no. 7, p. 2560, 2022.

[8] ”Frisbee Tracking Dataset,” Roboflow, [Online]. Available: https://universe.roboflow.com/tracking-f1jov/frisbee-tracking. [Accessed: May 4, 2025].

[9] ”Ultimate Frisbee Object Detection Dataset,” Roboflow, [Online]. Available: https://universe.roboflow.com/computer-vision-soccer-project/ultimate-frisbee-qbihl. [Accessed: May 4, 2025].

[10] ”Ultimetrics - Analyzing Ultimate Frisbee Film using Computer Vision,” Carleton College, [Online]. Available: https://cs.carleton.edu/cs_comps/2324/ultimetrics/final-results/index.html. [Accessed: May 4, 2025].

[11] B. Eberhard et al., ”A Machine Learning Approach to Player Value and Decision Making in Professional Ultimate Frisbee,” MIT Sloan Sports Analytics Conference, 2025.

[12] ”Position Detection in Ultimate Frisbee Using Drones,” [Online]. Available: https://monarch.qucosa.de/api/qucosa

[13] ”Tracking Data Recorded with a Drone from an Ultimate Frisbee Game,” ResearchGate, [Online]. Available: https://www.researchgate.net/figure/Tracking-data-recorded-with-a-drone-from-an-Ultimate-Frisbee-game-during-the-pull-and-7s_fig8_359279938. [Accessed: May 4, 2025].

[14] ”Statto - Professional Ultimate Frisbee Stats,” [Online]. Available: https://statto.app/. [Accessed: May 4, 2025].

[15] H.-W. Huang et al., ”Iterative Scale-Up ExpansionIoU and Deep Features Association for Multi-Object Tracking in Sports,” arXiv preprint arXiv:2306.13074, 2023

[16] IBM, ”AI vs. Machine Learning vs. Deep Learning vs. Neural Networks,” IBM Think. [Online]. Available: https://www.ibm.com/think/topics/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks. [Accessed: May 4, 2025]

VII. Source Footage

Footage that was not filmed personally was obtained from the following sources.

https://www.youtube.com/watch?v=fyhb7uiw5lc
https://www.youtube.com/watch?v=jCCQGsqVPrA
https://www.youtube.com/watch?v=RQAgQboRr7U
https://www.youtube.com/watch?v=GriJrTDW5hg
https://www.youtube.com/watch?v=_NAOG3swQag
https://www.youtube.com/watch?v=BASMs-6eHxM
https://www.youtube.com/watch?v=iWS7BwH5jD0
https://www.youtube.com/watch?v=YAR10-nhO84
https://www.youtube.com/watch?v=Kjpo2crdJ2g