UNIVERSITY OF GHANA

SCHOOL OF ENGINEERING SCIENCES

COLLEGE OF BASIC AND APPLIED SCIENCES

DEPARTMENTOF COMPUTER ENGINEERING

FIRST SEMESTER 2024/2025 ACADEMIC YEAR

COURSE CODE: CPEN 419 – COMPUTER VISION

COURSE INSTRUCTOR: DR. NII LONGDON SOWAH

PROJECT 1

VOT 2022 ROBUST SCALE ADAPTIVE MEAN SHIFT TRACKING PROJECT CHALLENGE REPORT

GROUP 4

SAMUEL ADAMS – 10949889

FIADOR KELVIN – 10967291

HAARIS ALI WAQAS – 10947015

Table of Contents

1. ABSTRACT

In this report, we detail our process of setting up and evaluating the ASMS_VOT tracker—originally proposed in "Robust Scale-adaptive Mean-Shift for Tracking"—on a real-time camera feed. Due to the absence of a publicly annotated dataset, we used a single hand as the tracking target, with a manually defined initial bounding box provided in region.txt and a dynamically generated list of frame paths in images.txt. Although quantitative performance metrics could not be computed without ground truth data, our qualitative analysis shows that the tracker reliably follows the target, indicated by a single blue bounding box overlay. We also discuss encountered issues and how they were resolved.

2. INTRODUCTION

The ASMS_VOT tracker is a robust tracking algorithm that applies a scale-adaptive mean-shift method to follow objects across consecutive frames. Its design was motivated by the challenges posed in the 2022 VOT challenge, and it has since been adopted for various real-time tracking tasks. In our evaluation, we ran the tracker using our own live camera feed, targeting the tracking of a group member’s hand. This report outlines our methodology, input file preparation, execution details, encountered issues, and our qualitative observations.

3. METHODOLOGY

3.1 Overview of the ASMS_VOT Tracker

The tracker requires two main input files:

- **images.txt:** A file containing absolute paths to each frame in sequential order.
- **region.txt:** A file specifying the initial bounding box in the first frame. The expected format is a single line with comma-separated values:
  x, y, width, height

Upon initialization, the tracker reads the initial region from region.txt and sequentially processes the frames listed in images.txt. For every frame, it predicts a new bounding box that is written to output.txt using the same format. During visualization, these predicted bounding boxes are drawn on each frame using a blue overlay.

3.2 Preparation of Input Files

3.2.1 Creating images.txt

Since we used a real-time camera feed, we dynamically captured and saved frames during testing. The paths to these frames were then sequentially stored in images.txt. Each line in the file contained an absolute path, ensuring that the tracker could locate and process the frames in order.

3.2.2 Creating region.txt

For initialization, we manually defined the starting position of the target (a group member's hand) by drawing a bounding box around it. The coordinates were saved in region.txt with the format:

x, y, width, height

During early experiments, an issue arose when the file contained extra newline characters or formatting errors. We corrected this by ensuring that only the necessary values were present on a single line.

3.3 Execution

With the input files in place, we executed the ASMS_VOT tracker. The tracker processed the frames sequentially, generating predicted bounding boxes that were output to output.txt. These predictions were overlaid on the frames in blue to visually assess tracking performance.

4. EXPERIMENTAL OBSERVATIONS

4.1 Qualitative Tracking Performance

Without an annotated dataset containing ground truth bounding boxes, our evaluation relied solely on visual inspection. The following observations summarize our findings:

- **Tracking Stability:** The blue bounding box was observed to follow the hand consistently across frames. There were minor fluctuations in the position and size of the bounding box due to natural hand motion, but overall tracking was robust.
- **Adaptability:** The scale-adaptive nature of the tracker allowed it to adjust the size of the bounding box as the hand moved closer or farther from the camera.

- **Real-Time Operation:** The tracker operated in real time, providing immediate visual feedback that was valuable for debugging and iterative improvements.

4.2 Challenges and Resolutions

4.2.1 Input File Formatting Errors

- **Error:** “Error loading initial region in file region.txt!”
- **Cause:** The initial region file was either misformatted or contained extraneous characters.
- **Solution:** We revised region.txt to ensure it consisted solely of a single line with the format x,y,width,height, eliminating any extra spaces or newlines.

4.2.2 Incorrect File Paths in images.txt

- **Error:** “Error loading image file images.txt!”
- **Cause:** The file paths did not match the actual locations of the image frames due to dataset differences.
- **Solution:** We verified each path and regenerated images.txt to match the dynamically captured frames.

4.2.3 Ground Truth Misalignment in Evaluation

- **Observation:** Initially, only the first line of region.txt was used by the tracker, which resulted in a static ground truth annotation across frames.
- **Resolution:** To better reflect the intended ground truth, we augmented region.txt by adding subsequent lines, which improved the visualization overlay, though quantitative evaluation remained unfeasible without a fully annotated dataset.

5. VISUALIZATION OF RESULTS

5.1 Qualitative Results

Below are example screenshots from our experiment. The blue bounding box represents the tracker's output overlay on the video frames:

- **Figure 1:** Initial frame with manually set bounding box

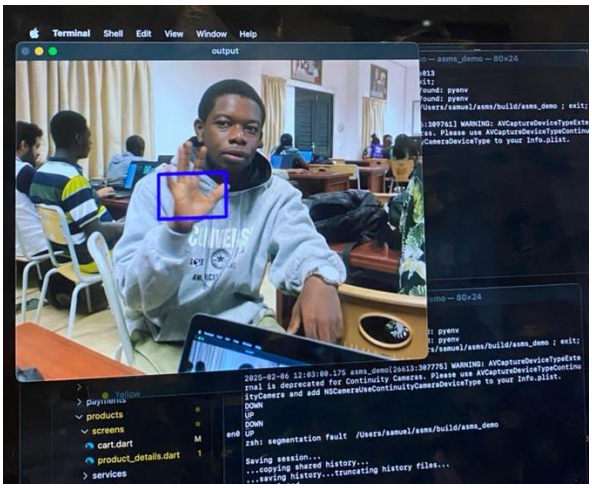

Fig. 1

- **Figure 2:** Mid-tracking frame showing the blue bounding box following the hand

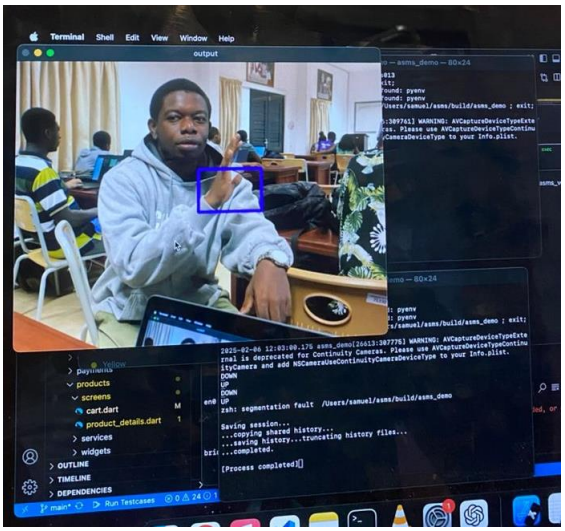

Fig. 2

- **Figure 3:** Final frame demonstrating sustained tracking performance

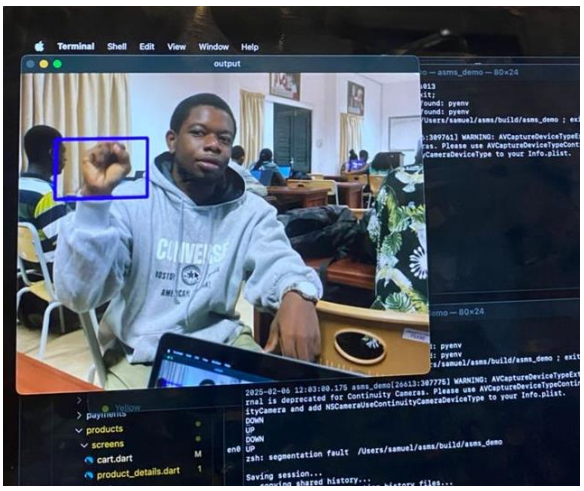

Fig. 3

5.2 Discussion

Even in the absence of numerical metrics, the visual results indicate that the ASMS_VOT tracker is capable of effectively following a moving target in real time. The slight misalignments and variations in the bounding box reflect natural tracking challenges (e.g., occlusions or rapid motion) and suggest areas for potential improvement in post-processing or additional adaptation.

6. CONCLUSION

The ASMS_VOT tracker, when applied to our real-time hand tracking scenario, demonstrated robust performance despite the absence of a fully annotated ground truth dataset. The tracker's ability to dynamically adjust to scale and consistently follow the target was confirmed via qualitative analysis. While quantitative evaluation was not feasible in our setup, our observations and visualizations support the viability of the tracker for similar real-world applications. Future work could include collecting annotated datasets to enable numerical performance metrics and further refine tracking accuracy.

7. REFERENCES

1. Vojir, T., Noskova, J., & Matas, J. "Robust Scale-adaptive Mean-Shift for Tracking."