Computer Vision Assignment 3

Muhammad Alfi Ramadhan - 21/472839/PA/20345

Implement into a program how to identify an adult falling using tracking.

Google Colab Link

I. Introduction

In computer vision, visual tracking is a computer vision task that involves estimating the target location of an object in a video sequence. It aims to predict the future state of the target object, including, but not limited to, its position, velocity and acceleration. When compared to detection, tracking can be much faster as when you are tracking an object that has already been detected in the previous frame, a lot can be learnt about the appearance of it, hence it can be used to predict its location in the next frame.

II. Problem

In this report, we will be applying visual tracking into a program for detecting an adult person falling, by considering the sort of features that will be taken into consideration and what exactly constitutes a fall. This can be beneficial for surveillance purposes especially in emergency situations where elderly people who have fallen, can be properly detected by surveillance and provided much-needed help as soon as possible.

III. Proposed Method

In the case of detecting a person falling to the ground, there are various methods of visual tracking that can be implemented as viable solutions. Some of these methods include template matching, Kalman filter and pose estimation. To determine the appropriate method to use, it is important to decide what qualifies as a fall. In this paper, we heavily take into account the centroid and the bounding box of the person. The idea is that if the centroid of the person drastically changes in position and rapidly moves downward for a certain amount of time, or that the bounding box of a person transitions from a vertical position to a horizontal one, we may infer that the person likely has experienced a fall.

IV. Methodology

A. Pose Estimation and Video Input

The program in this report utilizes OpenCV and Mediapipe libraries in order to perform video processing as well as pose estimation. A pose confidence threshold is defined for detecting poses. The input is a video file, its path is specified for processing, and each frame of the video will be read by a video capture object. As an appropriate example, the video that has been chosen as the input shows a common motion of a person falling down from a standing position. The pose estimation model is applied to each frame of the video in order to detect the person's landmarks, providing pose information.

B. Fall Detection Parameters

Based on the proposed method from the previous sections, several parameters have been chosen and initialized in order to explicitly identify a fall. The parameters are:

- prev_centroid_y: this parameter represents the y-coordinate of the centroid in the previous frame, which initially is set to None.
- fall_frames: this parameter represents a counter that keeps track of consecutive frames that indicate a fall.
- fall_threshold: this parameter represents the minimum number of frames required to classify a fall. This is set to 10 frames.
- centroid_drop_threshold: this parameter represents the minimum threshold for a significant vertical drop in the centroid position. This is set to 0.2.

C. Bounding Box and Centroid Calculation

Whenever the program detects a pose in each frame, it will continue to calculate the bounding box and the centroid of the person in the video. The coordinates of the bounding box is calculated by finding the minimum and maximum x and y values among the detected landmarks, and the width and height of the bounding box are calculated using the frame's dimensions. The centroid of the person is calculated simply as the midpoint or center of the bounding box.

D. Centroid Drop and Fall Detection

The program proceeds to calculate the vertical drop in the centroid's position by subtracting its position in the current frame from its position in the previous frame (prev_centroid_y - centroid_y). If its vertical drop exceeds the centroid_drop_threshold (0.2), it qualifies as a significant drop. The fall_frames, which is the number of consecutive frames that indicate a fall, will be incremented by 1 whenever a significant drop in centroid position is captured, else, it resets to 0.

E. Determining a Fall

The program will check if a fall is detected on each frame based on either of these two conditions:

- fall_frames is greater than or equal to fall_threshold, which suggests a consistent drop over consecutive frames
- The width of the bounding box is greater than its height, which indicates a falling posture

V. Results

After running the program, the output consists of each frame of the video accompanied by the relevant parameters that are taken into consideration when detecting a fall. Below are several notable frames of the output.

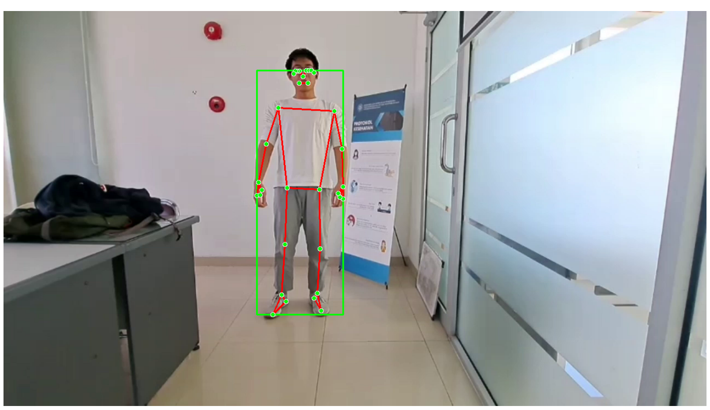

Figure 1. Frame 1 - Starting Frame

- centroid drop: 0
- fall frames: 0
- fall threshold: 10
- bounding box width: 157.6412582397461
- bounding box height: 444.6368479728699
- no fall detected

This frame shows the initial position of the person, which is a standing position, resulting in a vertical bounding box.

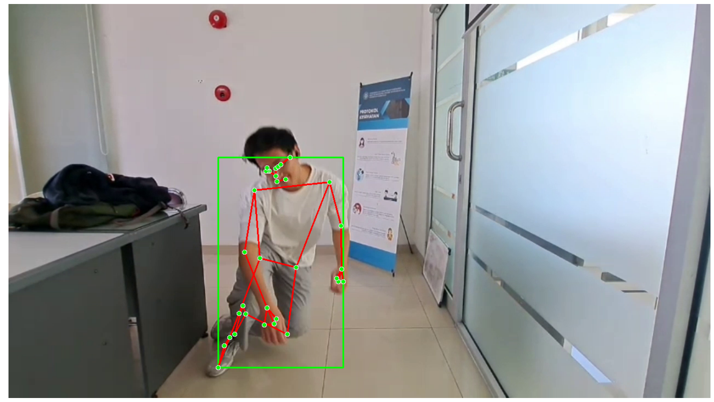

Figure 2. Frame 67 - First Frame Indicating Fall

- centroid drop: 13.516509294509888
- fall frames: 10
- fall threshold: 10
- bounding box width: 228.0581283569336
- bounding box height: 383.8580131530762
- **Fall detected!**

This is the first frame that had been detected as a fall. It can be seen that the fall frames had equalized the fall threshold of 10. This was sufficient in indicating a fall in motion.

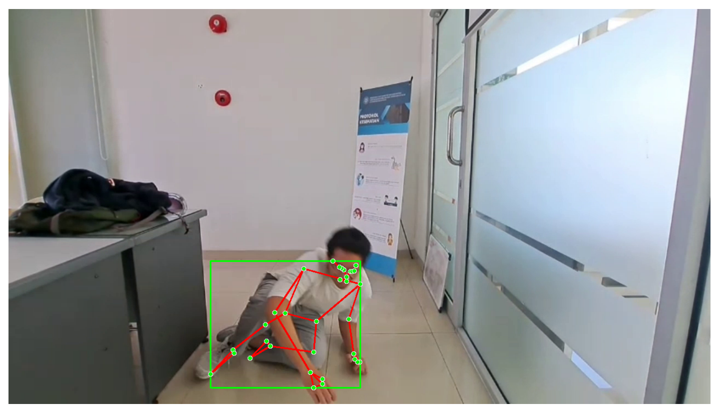

Figure 3. Frame 73 - Horizontal Bounding Box

- centroid drop: 9.757280170917511
- fall frames: 16
- fall threshold: 10
- bounding box width: 273.6729431152344
- bounding box height: 231.88937187194824
- **Fall detected!**

This is the first frame in which the width of the bounding box exceeded its height. The fall frames in this frame was also larger than its threshold. This frame showed a clear fall, and the program had successfully detected it as one.

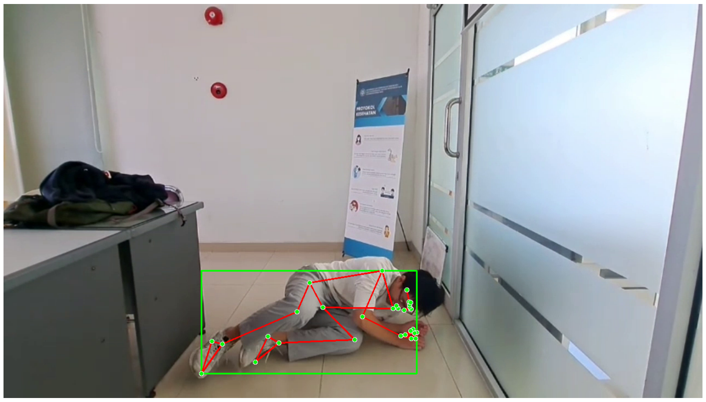

Figure 4. Frame 108 - Final Frame

- centroid drop: -1.7020130157470703
- fall frames: 34
- fall threshold: 10
- bounding box width: 391.7637634277344
- bounding box height: 187.39534378051758
- **Fall detected!**

This is the final frame of the video. The person's entire body is in contact with the ground and the bounding box is considerably horizontal in shape, clearly indicating that the person has fully fallen to the ground.

VI. Conclusion

Overall, the program was successful in detecting a fall on a video of a person falling, by taking into account the person's significant drop in centroid and their change in bounding box from vertical to horizontal, which are characteristics that heavily come into play in tracking a person falling down to the ground. However, further improvements can certainly be made. This can be achieved by increasing the variety of falling videos including different camera angles, falling motions and people. In addition to that, a larger variety of features rather than just the centroid and bounding box can be incorporated into the program for more accurate results. Moreover, differentiating between postures that indicate an ongoing fall and ones that indicate that the person has already fallen, can massively improve the program's specificity.