## Introduction

Sudden scene changes, such as those due to camera movement like panning introduces difficulties in object tracking as it is equivalent to a sudden increase in velocity of all objects in the frame, not to mention introducing noise and possibly causing objects to exit the frame.

This report discusses a possible method to allow for the continuous tracking of objects across scene changes by means of stitching together camera frames to create a static “global frame” on which detection and tracking are performed.

## Methodology

### Pre-processing

Camera lenses found in all manners of photographic equipment such as security cameras, mobile phones and imaging drones result in optical aberrations such as optical distortion. Distortion due the nature of the lens is known as radial distortion, and results in objects located at different radial distances from the optical centre of the lens to be mapped differently to the image space. This leads to points in the image frame having different motions during camera pans, which makes it difficult to correctly identify the motion of the scene. The distortions thus need to corrected for in order to correctly map the image frame onto the global frame.

Radial distortion and tangential distortion arising from the lens being non-parallel to the imaging plane can be corrected with the radial distortion coefficients, the tangential distortion coefficients, , the focal lengths of the lens , and the optical centre of the lens, . These parameters can be found through camera calibration or from a lookup table.



Figure 1 Raw frame from video



Figure 2 Undistorted frame after correcting for distortion

Optical of certain features is then performed in order to determine the motion of the frame to determine how to move the camera frame within the global frame. To do so, roughly 300 points are chosen using the Shi-Tomasi corner detector with a minimum quality level of 0.01 and distance of 10. These points are then tracked using the Lucas-Kanade method with pyramids. Points which are occluded are discarded, leaving two 2-D point sets from the frame its preceding frame which are used to estimate the affine transformation between the point sets.



Figure 3 Feature locations in the previous locations indicated with red circles with their corresponding positions indicated by green circles

The estimated transformation matrix is given in the form:

Where is the rotation, is the scaling factor and and are the translations in the and axes respectively.

Performing successive summations on the translations, the bounds of the global frame can be obtained. From there the image frames can be written onto the global frame, with the motion of the camera frame compensated by translating it within the global frame. This allows the introduction of a sudden velocity change in the frame to be overcome, allowing for stationary objects to be easily tracked across scenes.



Figure 4 Global frame obtained by translating and stitching camera frames together

## Results



Figure Tracking results for a 1920x1080 camera frame with a scene change

The above plot shows the tracks of two drones relative to a 1920x1080 camera frame. The colour of the track corresponds to the time in frames, with purple representing an earlier frame and increasing wavelengths representing later frames, with red representing frames towards the end of the video. The camera panning occurs during the yellow frames.

Track 59 belongs to a single drone which is moving to the right as the camera follows it to the right. Note how its position in the frame moves leftwards during the yellow frames due to the camera motion being faster than the motion of the drone.

Tracks 32 and 90 belong to the stationary drone. Notice how track 32 moves leftwards at a high velocity and overshoots the position of the drone in the second scene. This overshoot, combined with the fact that the stationary drone was invisible to the detection script during the pan due to the increased noise and proximity to the horizon resulted in it being dropped. Track 90 is the stationary drone which is picked up again in the second scene.

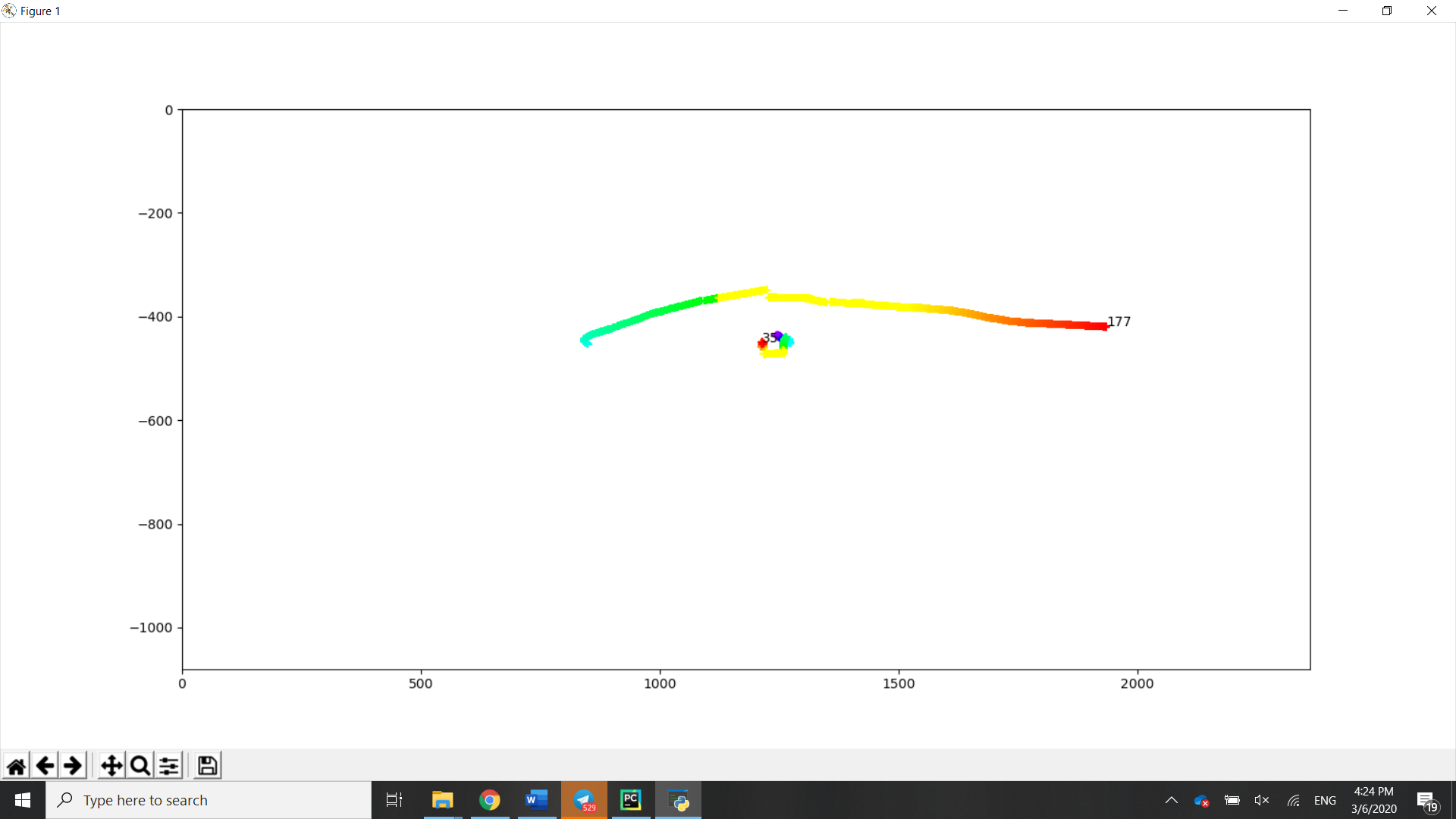


Figure 6 Tracking results of 2362 by 1080 global frame

The above plot shows the track of the same two drones in the global frame. Again, the camera panning occurs in the yellow frames.

Track 177 corresponds to the drone which is moving to the right. Notice how the rightwards motion is accurately reflected in the global coordinates.

Track 35 corresponds to the stationary drone. Not only is its stationary behaviour correctly reflected in the global coordinates, the track did not get dropped during the pan. Due to the position of the stationary drone being relatively constant in the global frame, the track was able to be reassigned to the detection of the drone after the pan, allowing its position to be described with a single track across both scenes.

Another thing to note is the increased track numbers. As the detection and tracking script was optimised for the 1920x1080 frame without correcting for distortion, the wider global frame with some stitching defects experienced more noise, which resulted in a greater number of dropped tracks attached to noise. Despite this, the overall performance of the tracking script on the global scene exceeded that of the 1920x1080 camera frame.

## Discussion

The results show that introducing a global frame is an effective way of correcting the relative motion of objects in the frame due to translation of the camera frame due to movement of the camera. It also shows that decreases the difficulties encountered in tracking objects across scene changes, allowing them to retain the same track before and after the scene transition.

However, this method comes with several significant drawbacks. The first is that it requires stationary objects in the form of the ground, trees or buildings to use as tracking points to calculate the optical flow of the scene. In the absence of such points, the objects to be tracked, which are often moving will instead be used to calculate the optical flow of the scene, which can result in nonsensical transformations. This problem is compounded by the fact that these objects often introduce noise, which introduce sections of the scene which may result in occlusion. This problem can be overcome by limiting the size of the scene dedicated to marker objects, or in the tracking stage by implementing techniques to allow reassigning of tracks after occlusion as demonstrated in this example.

The second drawback is that the construction of the global scene requires a pre-recorded video input, which limits its application to real-time tracking. To overcome this constraint, the techniques used in the construction of a global frame, i.e. identifying the transitional motion of the frame can be instead used to translate the camera origin to the global origin, creating a global coordinate system in which the tracks can be taken with reference to. By transforming the detections from the camera frame to the global frame, the relative motion of objects in the frame can be compensated for in real-time.

Finally, as this method uses 2-D point sets to determine an affine transformation, it is only applicable to rotational, scaling and translational transformations on the scene. Camera movements that result in perspective transformations would require a different approach which may not be able to be represented in a 2-D video and would require a different approach. Using a global coordinate system combined with information about the camera’s movement space might be able to overcome this constraint but it is a separate topic.