Laura Machlab
MP7 Report
Computer Vision
May 4, 2023

Task

The task of this MP is to create a function that can track the face of an individual in a video. Specifically, in the video given we were to put a bounding box around the face of the girl in all frames. To do this, we use three different image-matching methods: sum of squared difference, cross-correlation, and normalized cross-correlation.

Algorithm

The algorithm starts by manually setting the bounding box of the first image of the original video. After this, it steps through each following image from the video. At each step, it does an exhaustive search of the possible bounding boxes for the next image and for each of these possible bounidng boxes, it compares each pixel to the corresponding one of the previous frame's bounding box. Which image-matching method is selected – sum of squared difference, cross-correlation, or normalized cross-correlation – determines the metric used for the comparison between bounding boxes. When using the sum of squared difference to select the next bounding box, the difference between the two is minimized. When using cross-correlation and normalized cross-correlation, the bounding box with the highest correlation to the previous one is chosen. After finding the best bounding box location for the next image, it becomes the previous image, and this process is repeated for the next image. Once the bounding box location has been found and added to the image for all images, they are stitched together to create the output video.

Results

The output for the test was three different videos. The first is the sum of squared difference video (SSD_video.mp4). In this video, the boundary box location is accurate until the girl turns around and only the back of her head is visible to video. The second is the cross correlation video (CC_video.mp4). In this video, the location of the boundary box is very inconsistent, and it often is over the areas of the frame with the lightest pixels. I think this is because maximizing the cross correlation value would filter for maximizing the total values between the compared areas, and lighter pixels have the highest values. This video performed the worst of the three. The third is the normalized cross correlation video (NCC_video.mp4). This video performed the best out of the three, and the box was most consistently in the right place – even when the boy in the video put his face in front of the girls'.