HW7 Report

Problem Description

In this homework, we are implementing template-matching based target tracking and we are doing it to track the head of a girl over 500 frames. This happens by initializing a frame or many frames that will be used as reference(s) later for the matching methods. After that, some search method will be applied to look for potential frames to compare to our reference, and then for every frame that is chosen by the search method, we apply some image matching method that will be used to compare the frame we are on and the reference(s) chosen. It is important to choose the frames and the reference to be of the same shape.

Proposed Algorithm

In this section, we will explain the choice for the different parts of the algorithm described above:

- For the initialization, we create a bounding box manually by adding the coordinates in the python file and we choose the first image as reference. Hence, I added the bounding box for the head of the first frame As an additional change to the algorithm to solve the optional part, I also choose 4 bounding boxes for 4 other pictures as reference with all 5 frames having the same size, and then we will use them simultaneously to solve our problem. This will be described later. The 4 other bounding boxes were used to be representative of most frames, and they cover the back of the head and the side of the head since the first frame covers the front of the head.
- For the search method, I started by just looking at every frame possible of the size of chosen references and applying the matching algorithm to every frame. After that, I also applied another method that takes only the frames whose left corner x and y are within L distance of the previous output's x and y respectively. This is what is described in the PDF as local exhaustive search. This reduces the number of frames checked drastically and improves the time complexity of the algorithm without affecting the performance, because the head is not moving quickly and it is very unlikely that it will be more than 20 pixels away for x and y of the previous output frame. The later method is the one seen in the python code.
- For the matching method, we cover the methods suggested in the HW description, and these are sum of squared differences (SSD), cross-correlation (CC), and normalized cross-correlation (NCC). Given the method in the function, the search algorithm goes through the potential frames, and for every frame, we apply the given method by checking with the reference frame. This is done for every image and after going through all the frames and getting the scores for the frame, we choose the frame with the best score. The best score would be the lowest for SSD, and the highest for the correlation.
  To apply the algorithm to work for the optional mark, we use 5 reference boxes, and for every frame of the search method, we compare to all 5 of the reference boxes and

choose the best score and store it. For every image, we choose the best score out of the ones we stored. This will cover more cases like the back of head and the side of the head, which are not seen in the first frame and can't be covered from just one frame. The chosen score will be the closest to one of the frames that fits best with the image. The output includes a folder where only 1 reference was used (the first image) in folder "./one_ref", and the one where 5 references were used in the folder "./five_ref". In the 2 folder, we have 3 videos corresponding to the 3 methods that were mentioned in the paragraph above.

Results Analysis

In terms of initialization, I tried changing the main reference between many options: the first image with the clear face, other frames with the less clear and smaller face, the side of the face, the back of the head … The best results from a qualitative perspective seemed to be the one with the clear face in the first image. The results are seen in the folder "./one_ref".
As a reminder, In terms of search methods, we used 2 methods: the first one being covering all the frames, and the second one only covering certain frames within a certain limit. The 1st method is not included in the results because the results of the 1st method are very similar to the 2nd when I chose L to be equal to be 15, and it takes much longer to run with the 1st method. Also, based on the experimentation, the 2nd method helps stabilize the output by having a more restricted range, and when the other person comes in, the algorithm is still focused on the girl. All of the results use the 2nd search method.
In terms of how many references are chosen, the experiments with the 5 references give better results than the one with 1 reference as the 5 references cover more than just the face of the girl, but also the different perspective in which the head can be seen from. This gives the algorithm more options to match the frames with and gives better results. The only downside is the longer runtime as we have to calculate to normalized cross-correlation five times, but I believe the improvement in the results is worth it.
Comparing the matching methods, the worst results seem to be the results from the basic cross-correlation and that was expected because it will favor the frames that have higher pixel values to have higher product which is the main component of the cross-correlation, and this will translate into choosing the brighter frames instead of the actually relevant ones and will disrupt the output of the algorithm. The solution is obviously normalizing, which gives the best results as it captures the similarity between the reference and the frame being inspected while not favoring the higher value pixels through the normalization. The SSD method gives decent results but the results are not as good as the NCC. This is maybe due to the fact that SSD is much more sensitive to outliers and a small area that is different than the reference(s) might make the algorithm choose less logical frames that have a closer distribution to the reference.

Optional

The optional part was solved by using the 5 reference boxes and using NCC for matching method and the result can be seen in the folder “./optional” and it is obvious even when the girl is occluded by the boy, the bounding box is still centered on the girl and it doesn’t shift.