5 May 2022

Motion Tracking, MP7

James Wilkinson

The motion tracking algorithm implemented here is simple, and whilst effective, this assignment has highlighted some of the further improvements that can be made, and the challenges associated with motion tracking in general. The method employed here is such that an initial bounding box is initialised by the user, which defines the object that we intend to track. The values of the initial frame that lies within this box are stored. Upon viewing the next frame in the video, the algorithm performs a local exhaustive search by sampling the new frame in a range of 10 pixels on all sizes of the initial box from the prior frame. Within this sub-image, the original box and its values are slid over all possible positions, and a metric is used to calculate the similarity between the prior box and the new position on the new frame. The optimal value of this metric when computed for all possible positions of the box on the new frame yields the new position of the box.

In this assignment, similarity metrics were considered as the sum of squared differences, cross-correlation and normalised cross-correlation. The performance of NCC was the most effective method. The CC metric was the worst performer, and I believe its lack of normalisation is key to this poor performance. In all cases, the algorithm ran into increased problems when the subject turned her head (since the rotation alters not just the position of the image being tracked, but directly alters the shape being tracked). All algorithms performed well when the subject's face was closer (larger) to the camera and further away (smaller) from the camera. The NCC algorithm performed well when the man blocked the woman's face, where SSD failed slightly.

I also hypothesise that these methods are specifically prone to tracking the more extreme valued pixels within the initial defining box, especially in the CC case where there is no normalisation. In particular, this subject had a background window with a lot of light. When calculating the SSD, CC or NCC, these metrics will be most sensitive to the most extreme, outlier pixels. Thus, in this case I suspect the motion-tracking algorithm was aided in tracking the woman's head when her background was uniquely formed of the window, but hampered when she moved in front of other backgrounds.

Further steps would be to incorporate size-changes to the bounding box as well as x-y translations, which is a trivial extension in the current regime, by altering the size of the reference box by interpolating pixel values and performing an exhaustive search over size-changes as well as x and y positions. Rotation is another similar feature that can be explored in the same way, and would be relevant for the sections of the video where the woman tilts her head left and right.