16-720 HW3: Lucas-Kanade Tracking & Background Subtraction

Xiaoyu Bai (xiaoyub)

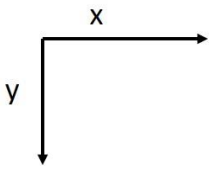
Q1.1

A^TA is the first order Tyler expansion approximation of Hessian matrix with respect to [u, v].

A^TA must be invertible or full rank, in other words, determinant of A^TA cannot be zero.

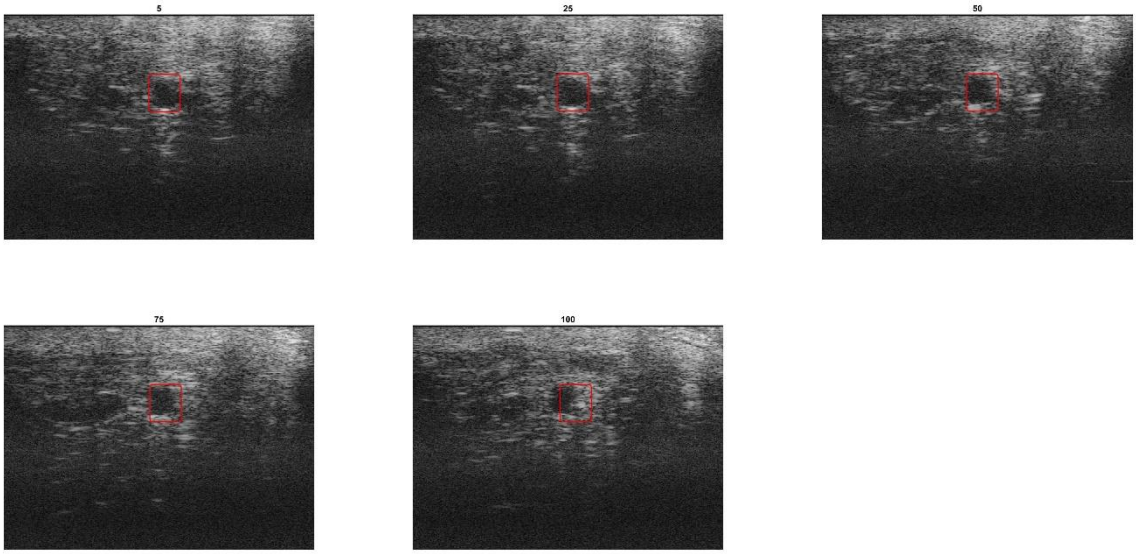
Q1.2

The following coordinate structure is assumed:

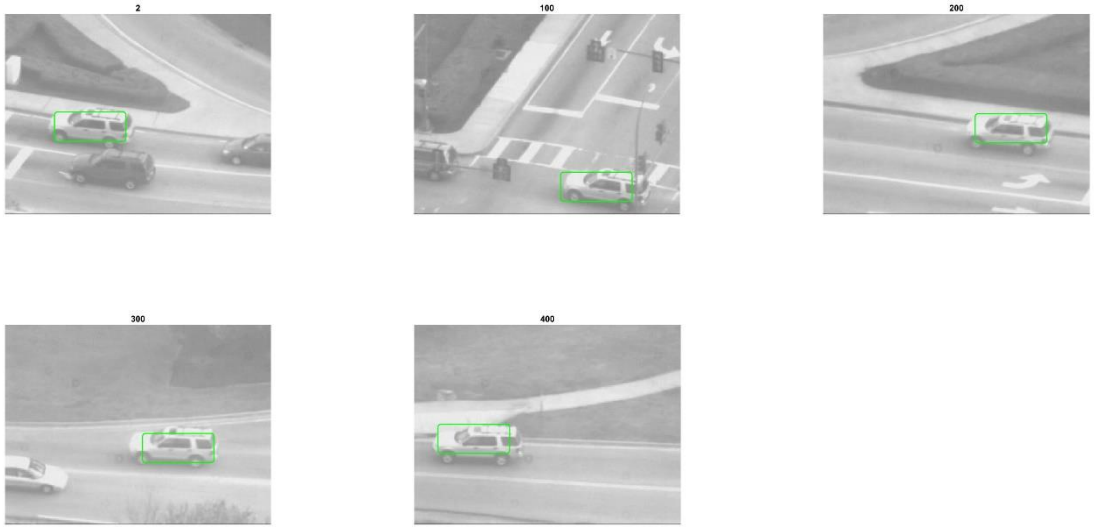


Q1.3

It is observed that once the target passes another object in front of it (e.g. the lamp in this case), the rectangle will have lags. This is due to larger amount of pixel intensity change in frame matching because of the obstacle interference. And as the error accumulates, the mismatch will become more distinguishable for later frames.



Q1.4



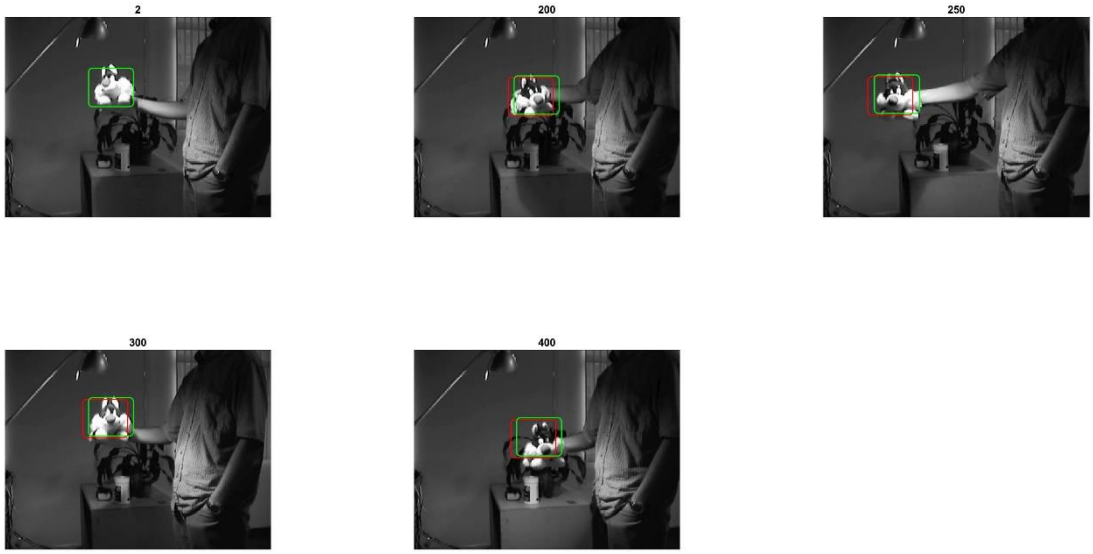
To implement with template correction, I simply add another threshold when updating [u, v]. Instead of using last frame as the template directly, we calculate the [u, v] from last frame. If L2 norm of [u, v] is greater than a threshold, then we use the first image as the template. Particularly here I set the threshold to be 5.5. This effectively is to set a binary regularization term to our transformation. If the transformation varies too much, then we use the first frame as template.

Q2.1

Since B_c is a set of orthogonal bases, we have $B_iB_j=0$ if i≠j and $B_iB_j=1$ if i=j. Therefore,

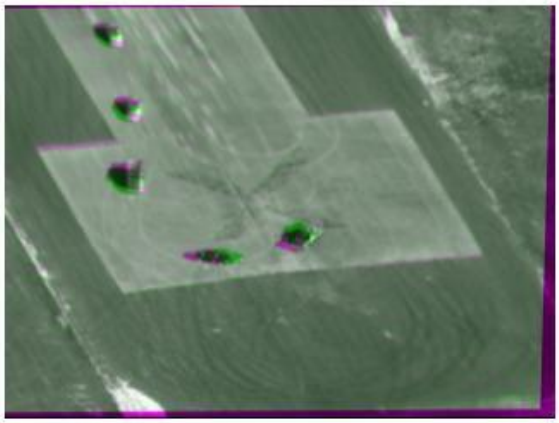
$$w_c = B_c^T \cdot (I_{t+1} - I_t)$$

Q2.2 & 2.3



Q3.1

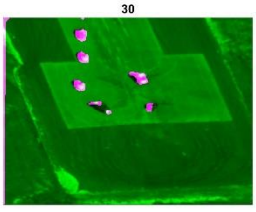
In this section, instead of the previous inverse compositional L-K method, I implement the original L-K method. It is more straight-forward to consider, because the steepest descent needs to be calculated over the overlapped area which changes at each iteration. In this way, Hessian matrix (which is $SD^{T}xSD$) needs to be calculated for each iteration. Therefore, compared with the previous implementation, the computational cost is heavier. Each calculation for transformation matrix *M* takes about 0.3 second. And the accuracy appears acceptable to me. Below is a fused image of two pictures which are 4 frames next to each other. Most of the region matches pretty well except for the moving vehicle. In detail, I set L2 norm of delta P to be the criteria and the threshold is set to be 1e-3.

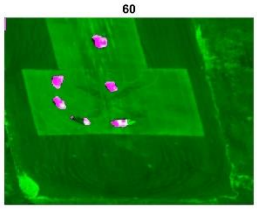


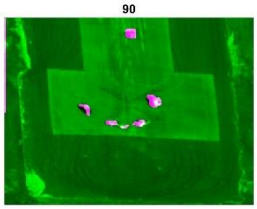
Q3.2

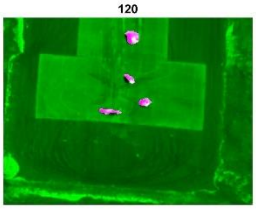
Demonstrated below, we show the comparison between one frame and the mask calculated from itself and its next time point neighbor. In detail, the threshold of judging one pixel intensity is 0 or 1 is 0.1 here.

Q3.3

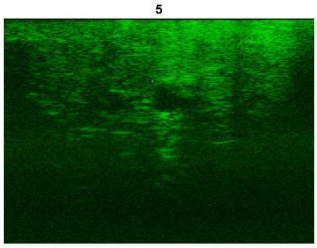


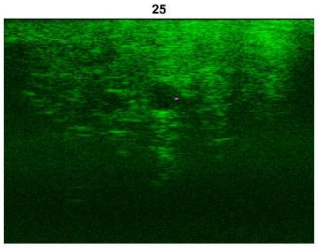


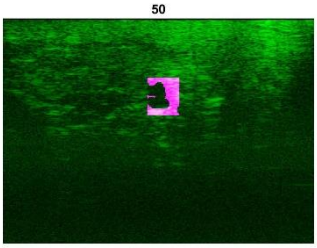


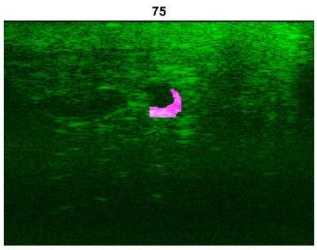


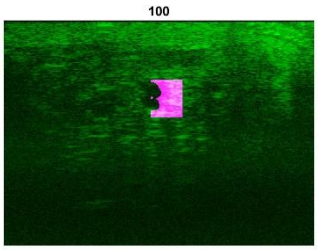
At the end, there are a few images that covers the grass nearby as moving object. According to my observation, this error is because the boundary of the grass is not as uniform as it is in the middle. So this algorithm mainly focuses on moving objects in front of uniform background.











In the test of ultra-sound video, the detection performance is not as good from my observation. The reason I attribute is still the same: non-uniform background.