Queen Mary
University of London

Introduction to Computer Vision

Coursework

Submission 2

Your name **SHELLY SRIVASTAVA**

Student number **190385633**

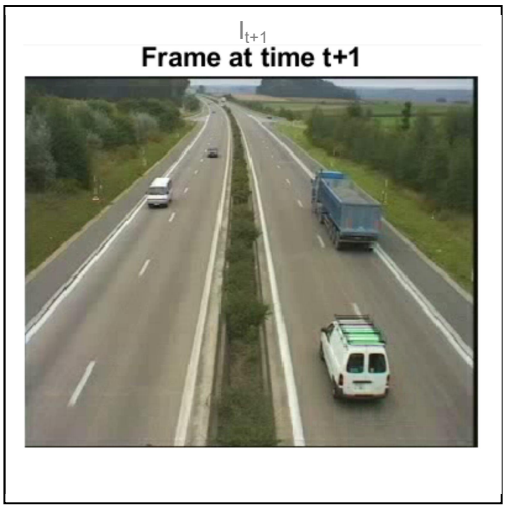
Question 4(a)

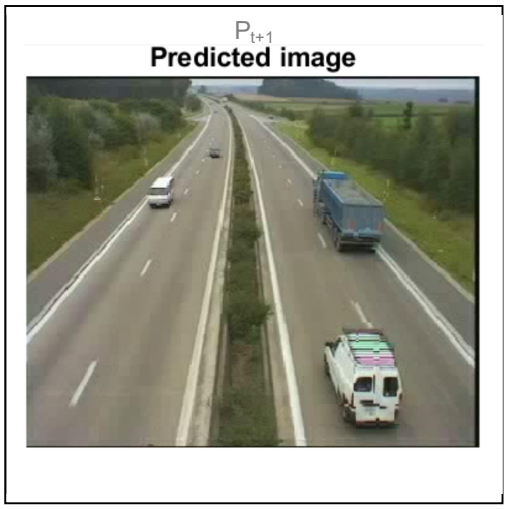
I_t

I_{t+1}

Motion field of I_{t+1}

Question 4(b)





Your comments

The motion vectors obtained are used to create the predicted image. The predicted image is distorted because of some error.

It can be seen that the white van is not recreated perfectly in the predicted image. This is due to the errors that crept in while creating the motion field. The movement of all blocks were not registered. Hence, when the new image is created, the white van actually appears longer.

Question 4(c)

P_{t+1}
Block size = 4x4
Predicted image

P_{t+1}
Block size = 8x8
Predicted image

P_{t+1}
Block size = 16x16
Predicted image

Your comments:

The predicted image with block size 4x4 has more distortion than compared to the image generated with block 8x8. This is due to the fact that the image has been divided into more moving block, resulting in more motion vectors. Hence the predicted image has more distortion to due to error generated while calculating the motion fields.

The predicted image with block size same as the search window has no distortion but is also the same image as the frame at time t. Thus searching the same location of the block to minimise the error, does not yield any motion vectors. Hence, the generated image is exactly the same as the one in the previous frame.

Question 4(d)

P_{t+1}
Window size = 8x8
Predicted image

P_{t+1}
Window size = 16x16
Predicted image

P_{t+1}
Window size = 32x32
Predicted image

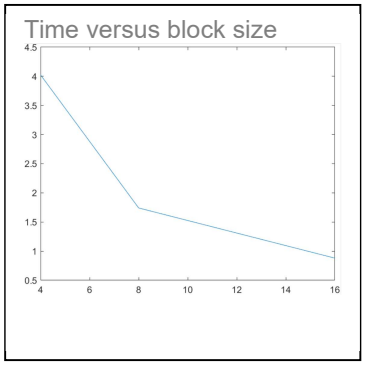
Your comments:

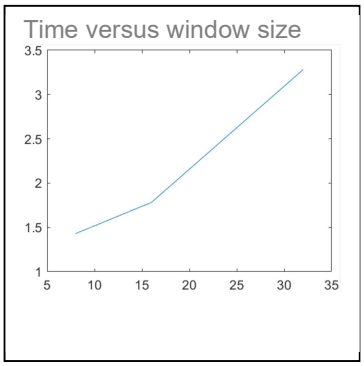
With window size and block size being the same i.e. 8x8, the predicted image doesn't resemble the frame at time t+1. As only one location was searched for the block and hence motion vectors were not generated. The predicted image is the same as the frame at time t.

Increasing the window size helps in detecting motion of objects with increased velocity. The predicted image with window size as 16 is less distorted than the one with window size 32.

Question 4(e)

Plot graphs:





Your comments:

Time versus block size:
The graph above as a downward trend which is justifiable as when the block size increases in the fixed search window, lesser number of comparisons are made. Thus, the time to generate the blocks decrease rapidly.

Time versus window size:
The second graph has an upward trend. As the search window increases, the number of comparisons increases for the exhaustive search. Thus, it takes more time.

Question 5(a)

Original frames:

Reference frame

Selected frame 1

Selected frame 2

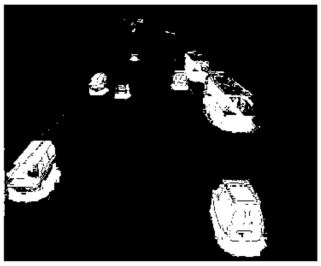
Frame differencing:

without threshold

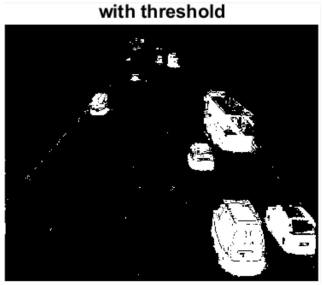
without threshold

Threshold results:

with threshold



with threshold



Your comments:

With the first frame as reference, the image difference shows all the cars in the first frame as well as the latter frames. For selected frame 1, the total number of cars visible is 8, i.e. 4 from the first frame and 4 from the second frame.

A binary image is produced after thresholding. Here a constant threshold is used. Thresholding helps in removing the noise in the image which can be due to external factors.

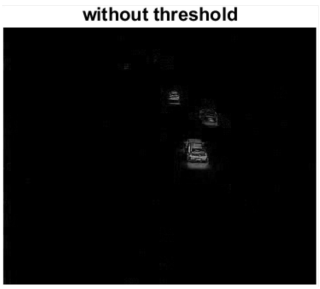
Question 5(b)

Original frame:

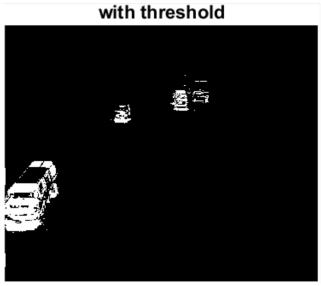
Selected frame 1

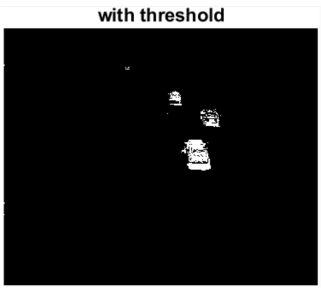
Selected frame 2

Frame differencing:



Threshold results:





Your comments for 5a,5b:

Frame differencing with reference to the previous frame highlights only the objects that have moved since the last frame. In the given video clip, only the moving cars are visible. For frame 1, the number of moving objects are 4, which can be clearly seen in the binary image.

It can be seen that there is still some noise as the exact outline of the moving cars are not generated. One can still see the shadow of the cars in both the images (with and without threshold). Also, the parts of the car are not highlighted. Instead of having a uniform white region, the resulted region has lots of holes in it. Usually these holes are removed before processing further with the help of dilation and erosion. This technique helps in obtaining a uniform region.

Question 5(c)

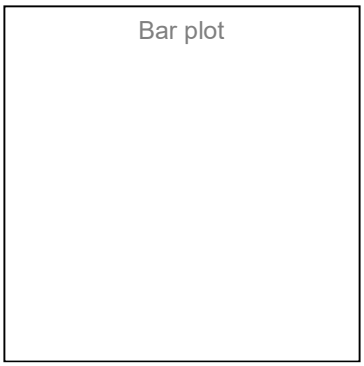
Generated background

Your comments:

The generated background has been developed by using the mode method. The pixel value at the selected is the mode of all the values of the pixel in the video frames. Thus moving objects don't contribute to the background image.

Moreover, an object placed on the background after a period of time can become the part of the background if it doesn't move. This means the background gets updated with time.

Question 5(d)

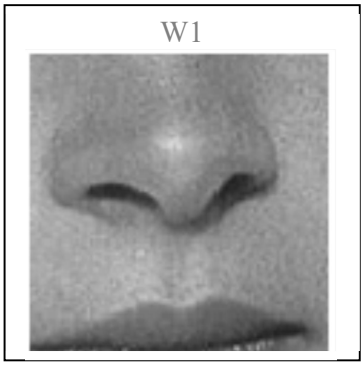


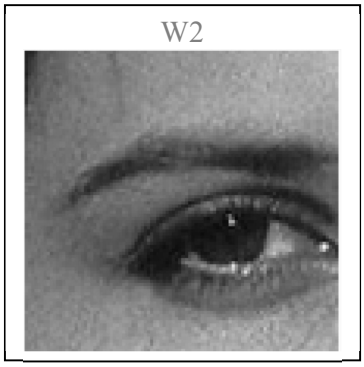
Your comments:

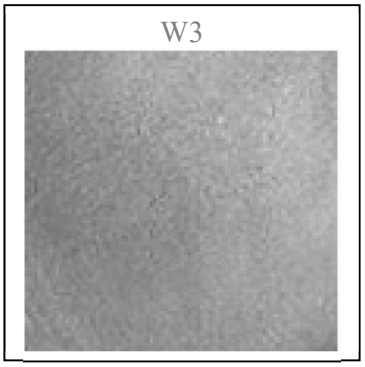
The method used was:
1. get the frame difference between the background image and the frame of the video (background image is used as the reference image)
2. threshold the image
3. fill the holes of the image to get a uniform region
i. dilate the image with a 3x3 structuring element
ii. erode the dilated image with the same structuring element
4. Count the number of uniform regions (not implemented)

Question 6(a)

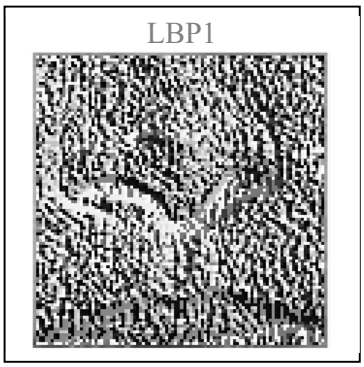
Three non-consecutive windows

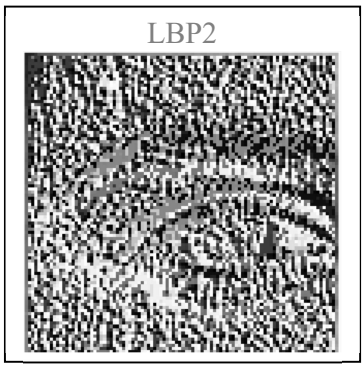


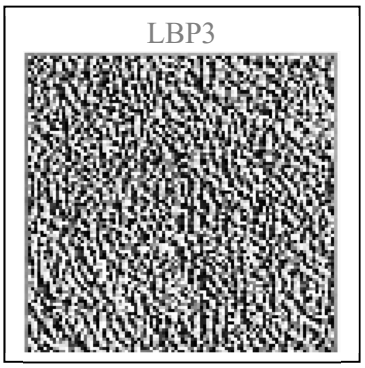




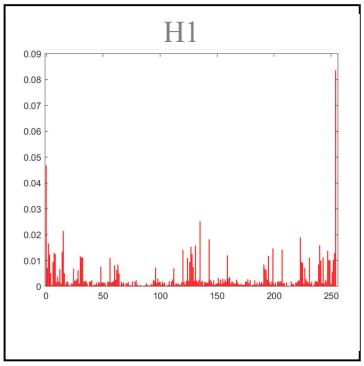
LBP of windows

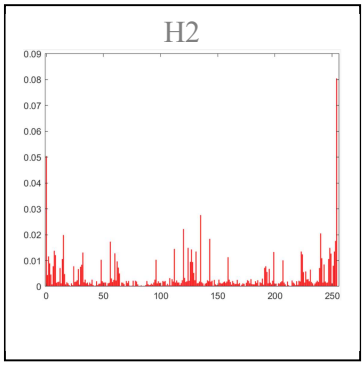


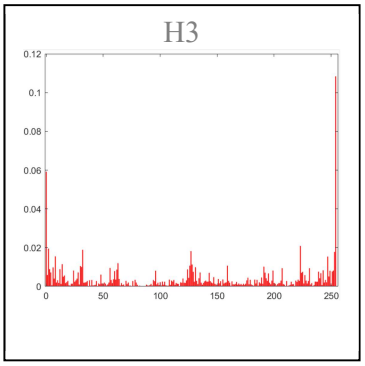




Histograms of LBPs

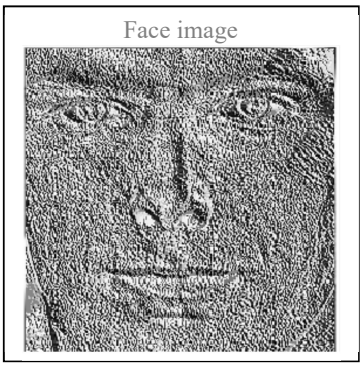


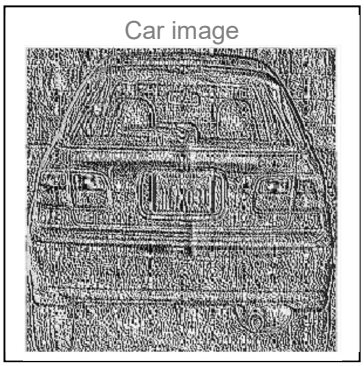




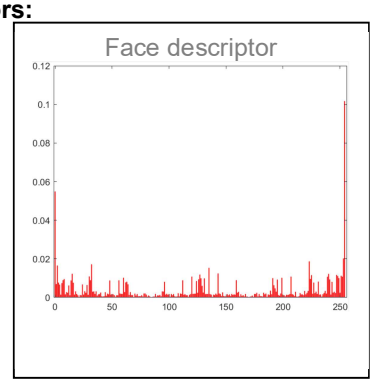
Question 6(b)

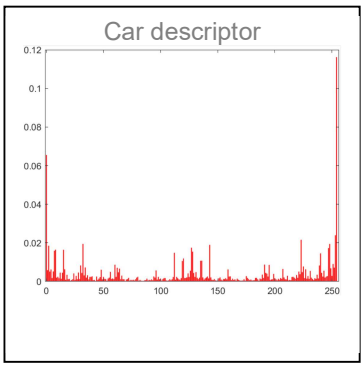
Two example images:





Descriptors:





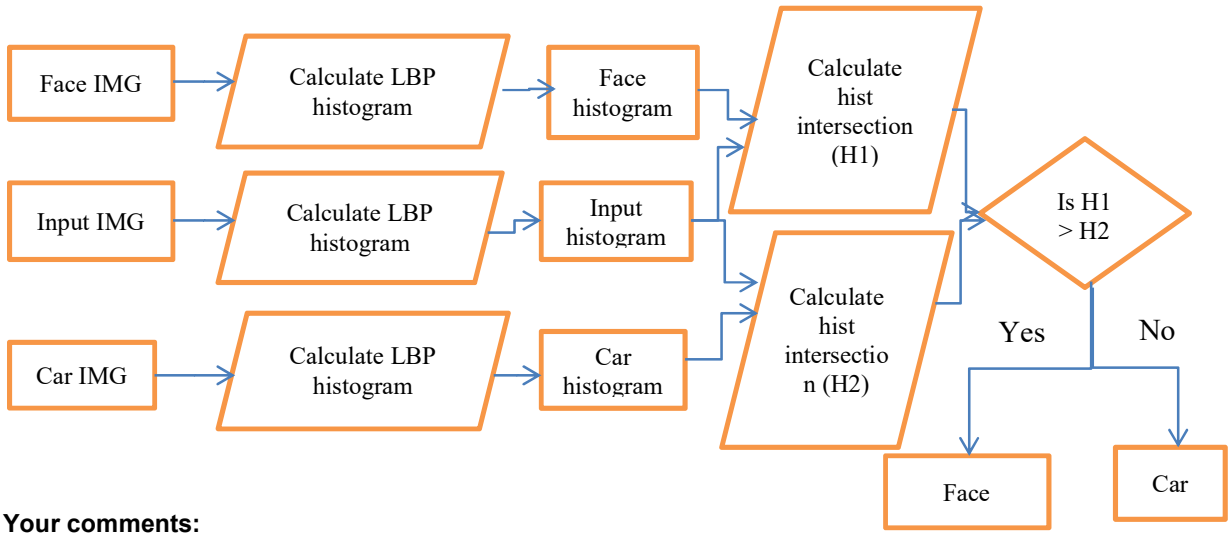
Your comments:

LBP generated with fixed window sizes has more variations in the histograms with peaks at different points, as can be seen in the histograms H1 and H2. With a plane region as the cheeks, the histogram H3 has less peaks in the middle range and a larger peak at the extremes with more than 0.06 and 0.1 for 0 and 255 respectively. Whereas, the histogram H1 and H2 have extremes with less than 0.05 and 0.09.

Similarly, in the case of face and car descriptor, the car have more features than face.

Question 6(c)

Block diagram of classification process



Your comments:

The classifier has two base images (car-1 and face-1) which is used to classify any input image. The LBP histogram is first calculated for the base images. Then the LBP histogram is calculated for the input image as well. The histogram intersection is calculated for face-1 and input image and car-1 and input image. If the intersection value of face-1 and input image is more, the then input image is labelled as a face, else it is labelled as a car.

The histogram intersection tells the measure of similarity between the two images.

Question 6(d)

Your comments:

Decreasing the window size and performing the classification again on each non overlapping window still classifies the overall image into the correct category. It has a few miss hits with respect to particular window.

This can be due to some similarities of smoothness across the face and car body.

Question 6(e)

Your comments:

When the classification window is increased, the windows are classified into the respective classes with fewer miss hits and even the overall image is classified into the right category.

Question 6(f)

Your comments

For dynamic texture analysis, a method called volume local binary patterns (VLBP) is used. It is an extension of the LBP operator. It combines the motion and appearance together.