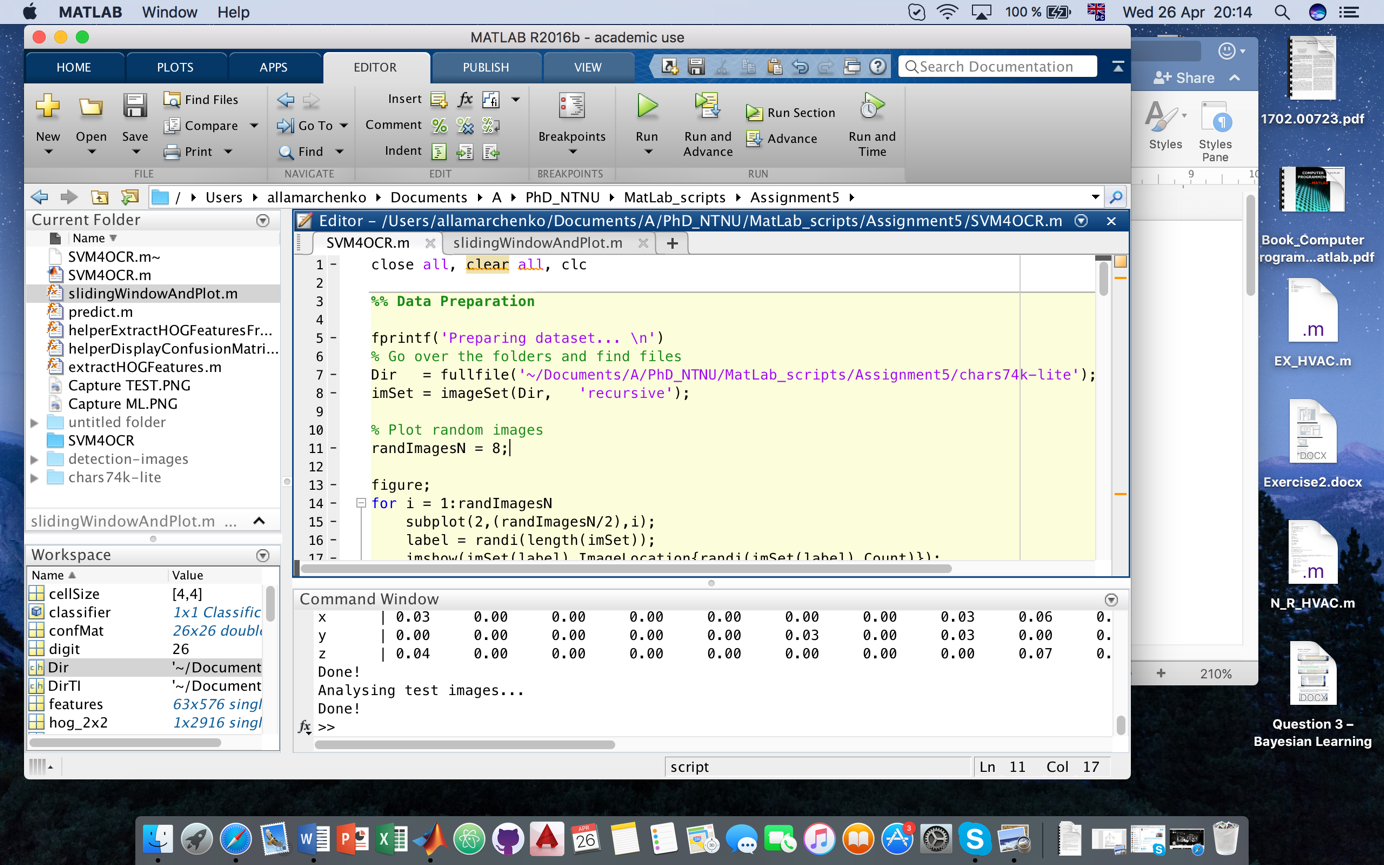
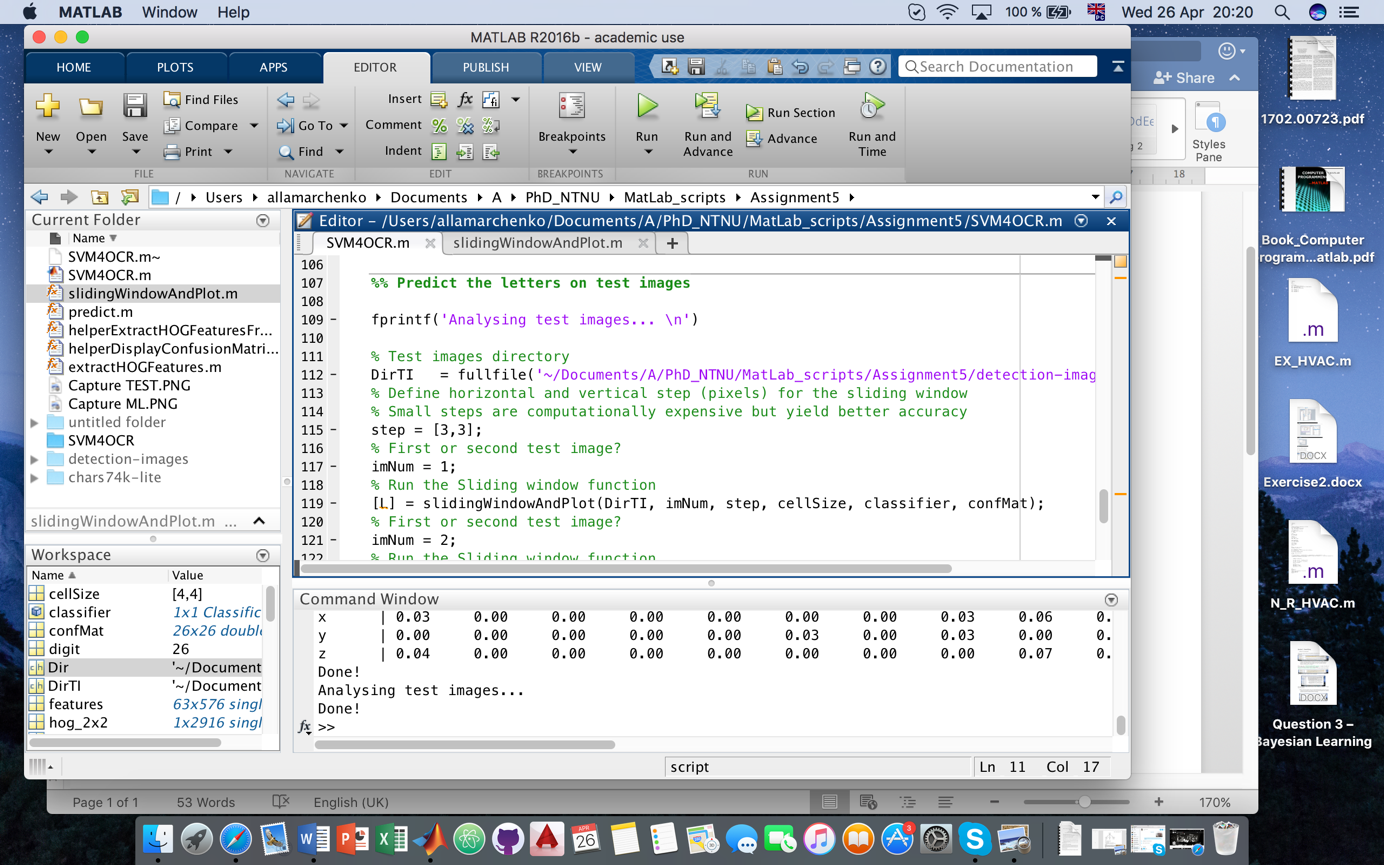
How to run script in MatLab

1. Make sure that all scripts are in the same folder.
2. In line 7, put the path to the folder with image library. For this specific task, it will be: “chars74k-lite”. (see pic. 1)



Picture 1: Writing path to the image library.

1. In line 112, put the path to the folder with images that need to be recognised. (see pic. 2)



Picture 2: Writing path to the images for recognition.

1. You may change parameters for pixel set with respect to X and Y in line 115 pic. 3.

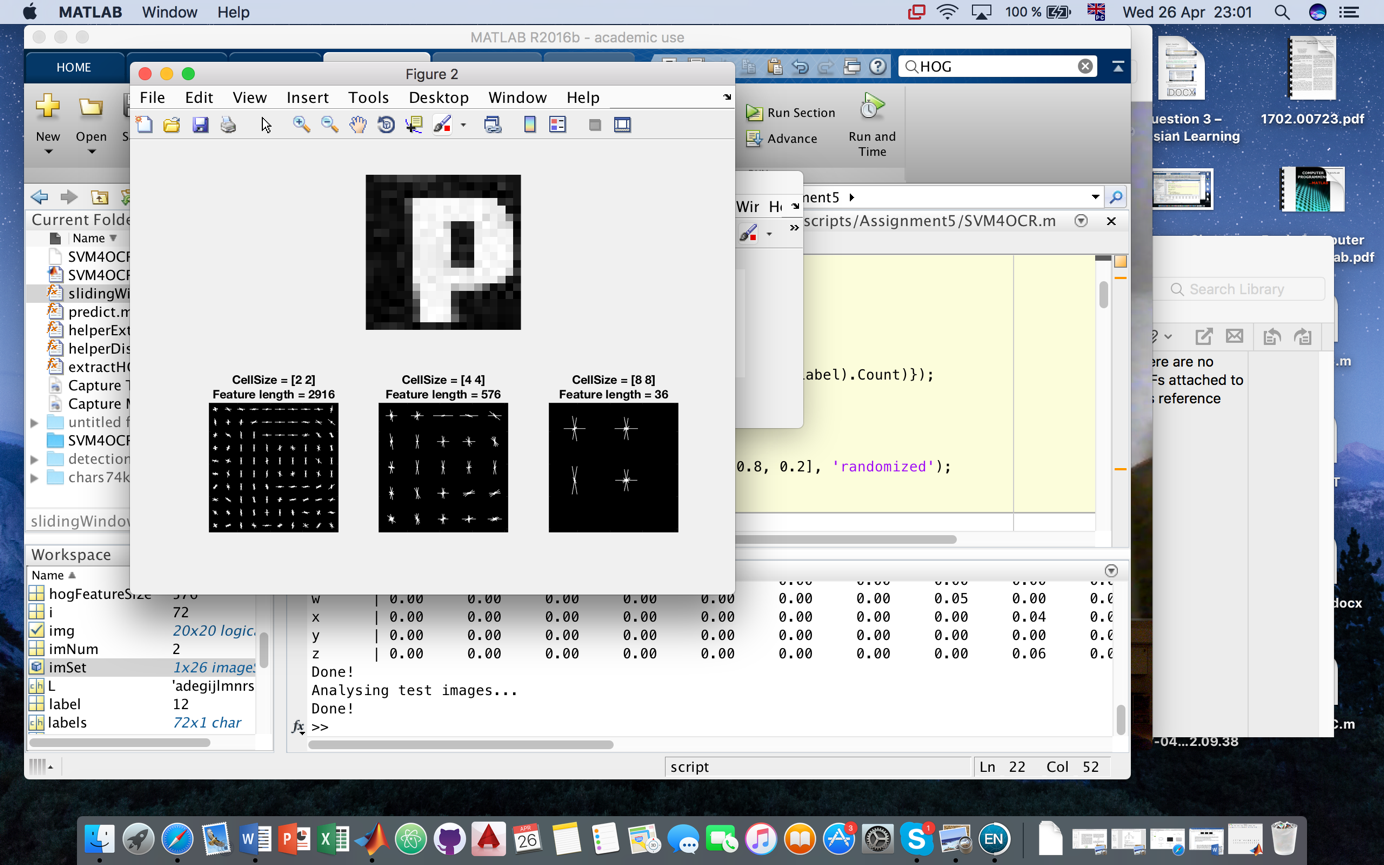


Picture 3: Change pixel step for sliding window

Support Vector Machines with Histogram of Oriented Gradients for Optical Character Recognition.

Support Vector Machines (SVM) can be successfully applied for such problems as: classification, clustering and regression [1]. SVMs belongs to class of the algorithms that uses kernel substitution and they are well suited for data mining [2].

Histogram of Oriented Gradients (HOG) was used for human body detection but now it is also being used for computer vision and pattern recognition [3]. HOG divides images into small square cells and computes histogram of gradient directions (see pic. 4). From time to time HOG has been criticized as computationally expensive [4].

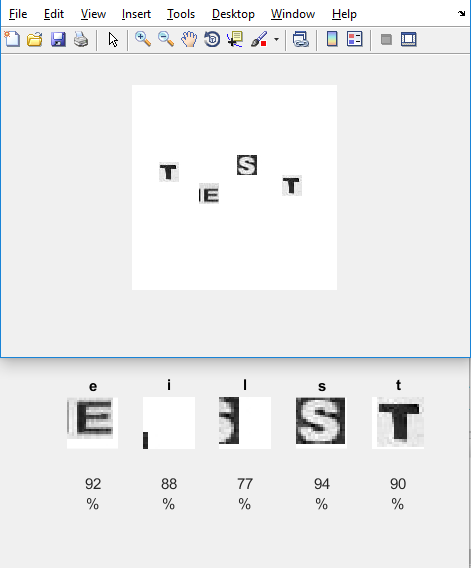


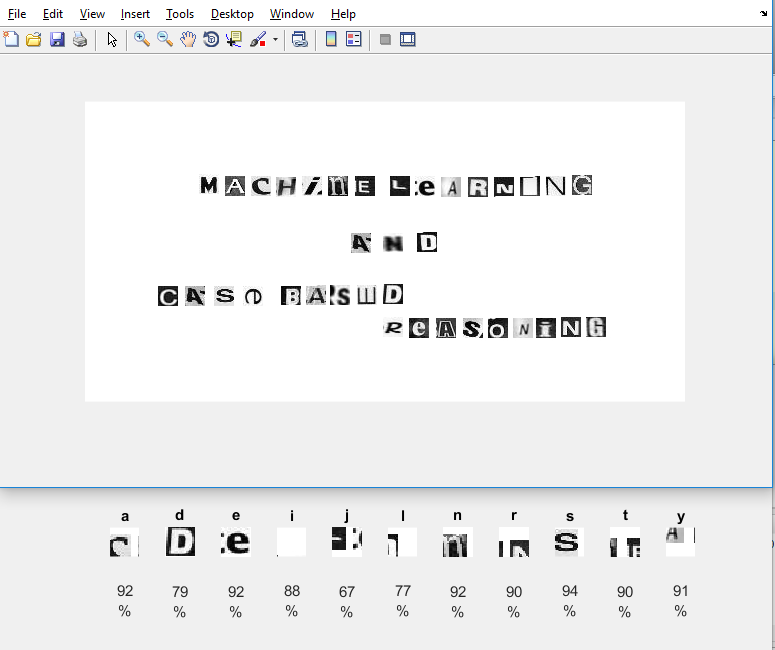
Picture 4: HOG features performed in MatLab for OCR of the text.

For OCR task were given large set of data stored in separated 26 folders, where each folder contains image of the letter from alphabet in jpg format. Script runs through them and process images in order to create an array for further image processing.

80% of the data were used to train algorithm and 20% left for test.

Evaluation of the classifier is performed with use of the images from testing set. This step generates the confusion matrix.





With respect to achieved results from image recognition can be done following conclusions:

Some letters are recognized poorly due to that fact that in our training data set were pictures with white letters on black background and vice versa.

Matrix of confusion presented that letter “f” is often confused with “i”, “t”; “o”, “q” and “r” also has degree to be recognized with mistake.

1. Murty, M. and R. Raghava, *Support Vector Machines and Perceptrons: Learning, Optimization, Classification, and Application to Social Networks*. 2016: Springer.

2. Bennett, K.P. and C. Campbell, *Support vector machines: hype or hallelujah?* ACM SIGKDD Explorations Newsletter, 2000. **2**(2): p. 1-13.

3. Ebrahimzadeh, R. and M. Jampour, *Efficient handwritten digit recognition based on histogram of oriented gradients and svm.* International Journal of Computer Applications, 2014. **104**(9).

4. Lee, K., et al. *Human detection using histogram of oriented gradients and human body ratio estimation*. in *Computer Science and Information Technology (ICCSIT), 2010 3rd IEEE International Conference on*. 2010. IEEE.