DSE312 Computer Vision Internal Assessment

Coding Assignment (70 marks)

Submission date: 4th November 2022, 11:59PM (before midnight)

1. Given the image “touching_grayscale”, (pls convert to grayscale first), implement a clustering method of your choice to separate the objects (that are touching). Use appropriate methods to show your result images before clustering and after to display the results. Next, use the Otsu’s thresholding and connected components. Display your results with the outer boundary and centroid of each object (plotted) to show your results. (10)

2. Self-study+assignment question: Local Feature Matching using Harris Corners and SIFT description and SURF.

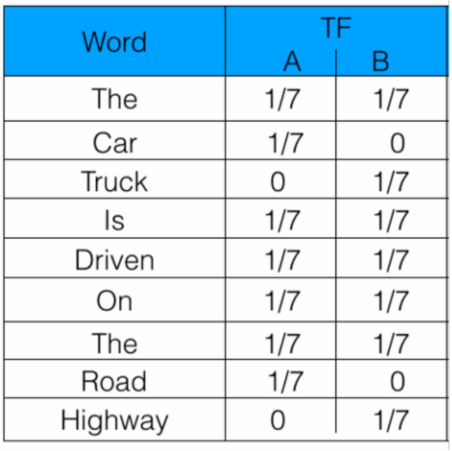
i) Write the pseudocode for the SIFT descriptor and few lines in which scenarios such a descriptor will be useful? (5)

ii) Just like the SIFT descriptor that we already discussed in the class, there is another descriptor called the SURF descriptor (in-built function typically in all your python/matlab). Briefly explain the steps that are built into these descriptors in your own words.

Pick seemingly two different views of any monument building in the world (find 2 images at 2 different view). Apply the Harris corner to it and SURF detectors on them. Also comment on how the methods work on any “repetitive structures” within the images (i.e whether such points are unique features or not and can be helpful for finding similarities)? (10)

3. For the datasets available at https://www.kaggle.com/datasets/ciplab/real-and-fake-face-detection
   i) implement the viola-jones appropriate to detect faces (using a pre-trained algorithm). Explain your steps in brief (your approach) and general steps in the algorithm.
   ii) Take any face picture from the dataset and apply the LBP method to it and display the results.(5)

4. For the following, write your own computer code to compute Term frequency–Inverse document frequency (TFIDF). Display the results with a histogram. (5)



Word	TF	
	A	B
The	1/7	1/7
Car	1/7	0
Truck	0	1/7
Is	1/7	1/7
Driven	1/7	1/7
On	1/7	1/7
The	1/7	1/7
Road	1/7	0
Highway	0	1/7

5. Use any real world image sequences (eg: moving vehicle, people walking) with at least 5 frames and implement optic flow (whole image: dense flow) and LK-feature tracking (sparse flow, computed only on features). (10)

6. On the database: https://www.kaggle.com/datasets/ciplab/real-and-fake-face-detection , Apply HoG descriptor. Use the HoG features to classify face versus non-faces using SVM (binary classification). Evaluate the results. (5)

7. MNIST is a commonly used dataset in the field of neural networks. Download the set of training images and labels from http://yann.lecun.com/exdb/mnist/.

i) Extract the HoG features that are also useful for Optical Character Recognition (OCR) of the training set and use the test set for testing. You need to use a multi-class model here. Evaluate the results.

ii) Start with the MINST dataset and build a CNN based classifier. Use any pre-trained model and build your code for digit classification using transfer learning.

Fit the model above with the training dataset (decide how much dataset you want to split into training & test (eg: 80:20), set epochs, minibatch size and learning rate.Find the right hyperparameters to have high accuracy. Show the result using confusion matrix.

If the size is large, A subset of the Modified National Institute of Standards and Technology database (MNIST) dataset can be used here. Comment very briefly on the results with HoG feature based classification versus CNN results. (20)

For any classification questions:

- Use a labeled data set with images of the target objects.
- Partition the data set into a training and a test set.
- Train classifier using features extracted from the training.
- Test classifier using features extracted from the test set.
- Pre-trained models require images to be in specific dimensions. Please read up the requirements and then proceed to build your code.

Please name all your submission files with ‘yourname and rollnumber’.