X

(https://swayam.gov.in)

(https://swayam.gov.in/nc_details/NPTEL)

cs19b1027@iiitr.ac.in

Register for Certification exam (https://examform.nptel.ac.in/2022_04/exam_form/dashboard)

Week 3: Assignment 3

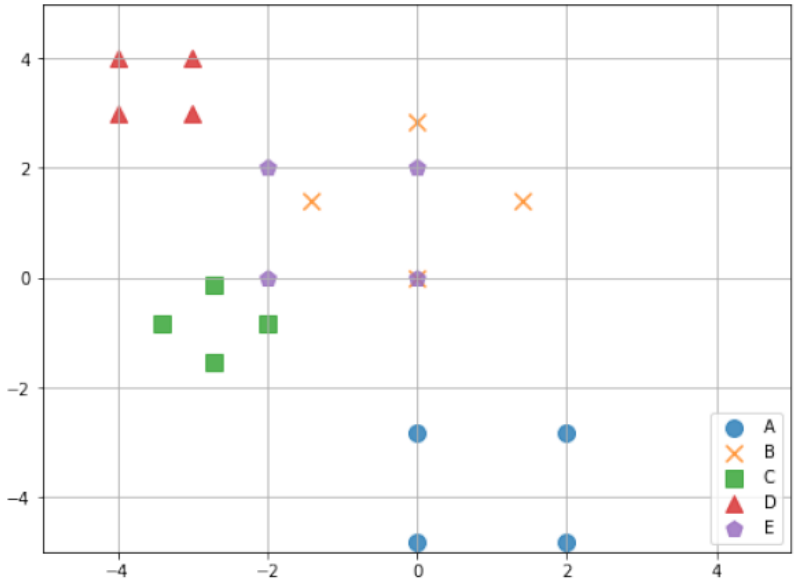
Assignment not submitted **Due date: 2022-08-17, 23:59 IST.**

Instructions:

- Starter code for Question 1 to Question 3 of this assignment is provided in DL4CV-Assignment-3-2022.ipynb. (https://drive.google.com/drive/folders/1EaGUL9wHND9Gjn7WHqv3YEEj4Pm8g6vS?usp=sharing)
- Use Python 3.x to run the notebook. As instructed in the notebook, write your code only in between the lines 'YOUR CODE STARTS HERE' and 'YOUR CODE ENDS HERE'.
- Do not change anything else in the code; if you do, the answers you are supposed to get at the end of this assignment might be wrong.
- Read documentation of each function carefully.
- All the best!

1) For this question, please see Question 1 in the iPython notebook(.ipynb file) provided alongside. Complete your implementation under the "YOUR CODE STARTS HERE" segment therein. Apply the following chain of transformations on the generated coordinates:

1. Rotation by 135 degrees.
2. Translation of x coordinate by -2 units and y coordinate by 2 units.
3. Scaling by 2 units.
4. Rotation by 135 degrees.



The following figure represents 5 different set of points(each in different color) plotted on a cartesian plane.
One of them is the correct output that will be produced when we apply the transformations mentioned above to the given input.
Which of the following is the correct output which is produced after applying these transformations?

- ○ A
- ○ B
- ○ C
- ◉ D
- ○ E

2) For this question, please see Question 2 in the iPython notebook(.ipynb file) provided alongside. Complete your implementation under the “YOUR CODE STARTS HERE” segment therein. What is the floor value of the mean? ***1 point***

- ○ 15

Practice: Week 3: Assignment 3 (Non Graded) (assessment?name=160)

Week 4 ()

Download Videos ()

Text Transcripts ()

Problem Solving Session ()

Books ()

- ○ 5
- ○ 17
- ◉ 9

3) For this question, please see Question 3 in the iPython notebook(.ipynb file) provided alongside. Complete your implementation under the “YOUR CODE STARTS HERE” segment therein. Follow the below mentioned steps: ***1 point***

Step 1. Given a folder of images, use SIFT to extract their feature descriptors.
Step 2. Perform k-means clustering on the extracted descriptors.
Step 3. Create histograms by taking help of the clusters calculated in the previous step.
Step 4. Estimate 5 nearest neighbours to a given image and add the distances of those 5 neighbours.

What is the range of the sum as calculated in step 4?

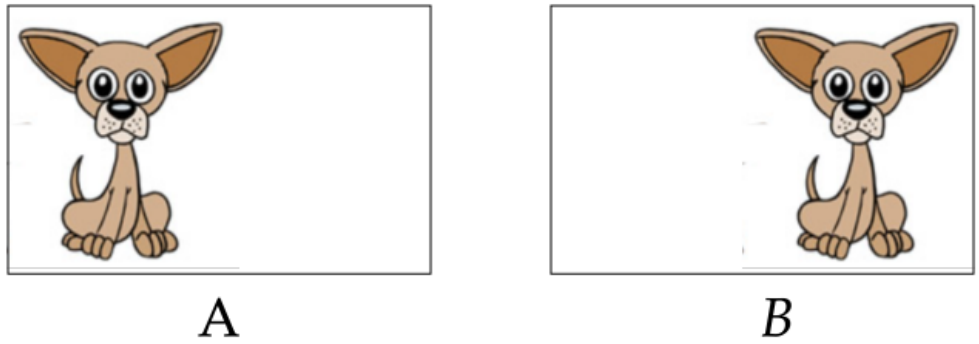
- ○ 0-200
- ○ 200-400
- ◉ 400-600
- ○ 600-800

4) Match the following: ***1 point***

1) Dense registration through optical flow	i) Assumption of rigid transformations
2) Wide baseline spatial matchingof same object	ii) Detect different instances
3) Rotation in x − y plane	iii) Two degrees of freedom
4) Bag-of-words	iv) Aperture problem
	v) One degree of freedom

- ○ 1→iv, 2→i, 3→iii, 4→ii
- ◉ 1→iv, 2→i, 3→v, 4→ii
- ○ 1→v, 2→ii, 3→iv, 4→i
- ○ 1→v, 2→i, 3→iii, 4→ii

5) Consider the two images A and B below. In A, a dog is present on the left part of the image and in B, the same dog appears on the right part of the image. These two images have _____ bag-of-words representations because bag-of-words representation _____ .



- (●) same, preserves spatial information
- ( ) different, preserves spatial information
- ( ) same, does not preserve spatial information
- ( ) different, does not preserve spatial information

6) Given a set of n images represented by a matrix $Z \in \mathbb{R}^{k \times n}$ and a query image q, which one of the following quantities gives the index of best matched image from the pool of n images to the query image q? Here, k is the dimensionality of image representations (e.g. bag-of-words). ***1 point***

- ( ) arg min $\{Z^T q\}$
- ( ) arg max $\{q^T Z\}$
- ( ) arg min $\{q^T Z\}$
- (●) arg max $\{Z^T q\}$

7) Match the following: ***1 point***

1) RANSAC	i) Used for image segmentation
2) VLAD	ii) Performance depends on number of inliers
3) Image retrieval	iii) Assumes nearby pixels have similar motion

4) Optical flow iv) Inverted file index
v) Vector representation of visual word

- ○ 1→ii, 2→v, 3→iii, 4→i
- ○ 1→iii, 2→iv, 3→v, 4→ii
- ○ 1→iii, 2→iv, 3→v, 4→i
- ◉ 1→ii, 2→v, 3→iv, 4→iii

8) In RANSAC, what is the probability that the algorithm never selects a whole set of inliers from 20 data points (which has 5 inliers in total) for $k = 2$ time steps? Here, $n = 2$ is the minimum number of samples to fit a model. ***1 point***

- ○ 0.2
- ○ 0.93
- ○ 0.87
- ○ 0

9) Which one of the following statements is false? ***1 point***

- ○ The difference between VLAD and BoW is that VLAD yields a residual vector
- ○ BoW does not allow partial matching
- ○ Aggregated Selective Match Kernel uses concepts from Hamming Embedding method and BoW
- ○ A codebook or a visual vocabulary consists of visual words

10) Let us consider RANSAC applied on 100 datapoints. Suppose you fit 10 lines *line*1 − *line*10 at each iteration time step t1 − t10. Number of outliers for each line is given by [*line*1 : 80, *line*2 : 90, *line*3 : 70, *line*4 : 60, *line*5 : 70, *line*6 : 40, *line*7 : 50, *line*8 : 80, *line*9 : 90, *line*10 : 60]. While applying the RANSAC algorithm, which line will be stored as solution at time step *t*5 and *t*10? ***1 point***

- ○ *t*5 : *line*5 , *t*10 : *line7*
- ○ *t*5 : *line*5 , *t*10 : *line10*
- ○ *t*5 : *line2* , *t*10 : *line9*
- ○ *t*5 : *line4* , *t*10 : *line6*

11) What will be the shape of output obtained after the element-wise encoding step in BOW and VLAD given that there are 2000 features and each descriptor is of 200 dimensions? Assume that there are 150 visual words.

- BOW:(2000, 150) , VLAD:(2000, 200)
- BOW:(2000, 150) , VLAD:(2000, 200, 150)
- BOW:(2000, 200) , VLAD:(2000, 200, 150)
- BOW:(2000, 200) , VLAD:(2000, 150)

12) Given that a geometric transformation requires 3 correspondences between an input image and the corrosponding transformed image, the degrees of freedom for that respective geometric transformation will be: ***1 point***

- 5
- 6
- 5 and 6
- 5 or 6

13) Consider 2 images $(X \text{ and } Y)$ where 2 binarized descriptors from each image are assigned to the same visual word. Given the following descriptors for each image, compute the score using Hamming embedding: ***1 point***

- First Binarized descriptor for image $X(b_{x1}) : [1, 1, 0, 0, 1]$
- Second Binarized descriptor for image $X(b_{x2}) : [0, 1, 1, 1, 1]$
- First Binarized descriptor for image $Y(b_{y1}) : [0, 0, 0, 1, 1]$
- Second Binarized descriptor for image $Y(b_{y2}) : [0, 1, 1, 1, 1]$

Consider the threshold $\tau = 2$.

- 1
- 2
- 3
- 4

14) Match the following: ***1 point***

1) Aggregated selective match kernels	i) Non-linear selective function
2) VLAD Matching	ii) Sum of residuals of same object
3) Efficient match kernels	iii) Non-threshold-based matching
4) Bag-of-words matching	iv) Cosine similarity

- ○ 1→i, 2→ii, 3→iii, 4→iv
- ○ 1→i, 2→iii, 3→ii, 4→iv
- ○ 1→iv, 2→iii, 3→ii, 4→i
- ○ 1→iv, 2→ii, 3→iii, 4→i

Consider the following sentence ”Artificial neural networks, usually simply called neural networks, are computing systems inspired by the biological neural networks that constitute animal brains.” The size of vocabulary required to apply a simple Bag-of-Words model is __a__ and the maximum frequency count obtained in the histogram after applying Bag-of-Words is __b__ .

15) a

0.5 points

16) b

0.5 points

You may submit any number of times before the due date. The final submission will be considered for grading.

Submit Answers