|  |  |  |  |
| --- | --- | --- | --- |
| Question | Q1 | Q2 | Q3 |
| Marks | 6 | 6 | 8 |
| Total | 20 | | |

**Full Name (as in course registration): A0108517H**

**Date: 14/03/2020**

**Course:** **Spatial Reasoning from Sensor Data**

**Question 1.** For each question, select the **single** most appropriate answer, and **provide your justification** (no longer than two sentences).

a). Which of the followings is the **most** significant drawback of *bag of words* (BOW) models?

|  |  |
| --- | --- |
| 1. They don't capture spatial layout information of the image. | 1. They are only suitable for image classification. |
| 1. Creating BOW features is time-consuming. | 1. They require a huge dataset of images to build BOW models. |

Answer: A. Bag of words only look at frequency of image/words and assumes independence between them, ignoring the spatial relationship between information in images.

b). What is the *scale-invariant feature transform* (SIFT) descriptor **least** robust against?

|  |  |
| --- | --- |
| 1. Global illumination change (e.g., brighter or darker) | 1. Rotation (e.g., rotate clockwise or counter clockwise) |
| 1. Scaling (e.g., zoom in or zoom out) | 1. Occlusion (e.g., occluded car plate due to another car) |

Answer: A. It’s based on histogram of gradients wrt neighbouring pixels and global illumination drastic changes may distort relative importance of initial key points extracted.

**Question 2.** You are engaged by a robotic vision company to develop a 3D vision based object retrieval system. Table 1 illustrates the input query and the top six retrieved results of the system. The retrieved results are displayed in the descending order that the most left result has the highest similarity score. Suppose that there are totally 3 sofa objects in the database. Evaluate the **average precision** (AP) performance for this system using this single query input. Show your calculations to justify your answer.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Input (sofa) | C:\Users\isstj\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\4F4D5CD.tmp | | | | | |
| Top results | C:\Users\isstj\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\CC7FF05F.tmp | C:\Users\isstj\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\72683663.tmp | C:\Users\isstj\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\18B938C5.tmp | C:\Users\isstj\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\C518E169.tmp | C:\Users\isstj\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\DEAC1B.tmp | C:\Users\isstj\AppData\Local\Microsoft\Windows\INetCache\Content.MSO\ABDAB7E1.tmp |
| Label | Sofa | Bed | Sofa | Chair | Toilet | Sofa |

Table 1. The input and top six retrieved results of the point cloud based object retrieval system.

Precision = Returned / Relevant = Tp / (Tp + Fp) = 3 / 6 = 1/ 2

Out of 6 topmost items returned, 3 are sofas, similar to input image label which is also a sofa.

**Question 3.** The *Housing Development Board* (HDB) in Singapore has planned that all future public housing blocks will be designed with solar-ready roofs. Inspecting these solar panels for possible defects can be a tedious process and even pose a safety risk to inspectors. It is potential to explore the *unmanned aerial vehicle* (UAV) inspection system using computer vision techniques. You are asked to **design an intelligent inspection system** that uses what you have learned in this class to perform inspection task for solar panels. The proposed system is required to have following two functionalities: (a) localization (inferring the location of the UAV), you can suggest what sensors should be used for this task; (b) automatic segmentation of the defect solar panel regions in the image captured by the camera mounted at the UAV. You can draw a flow chart to describe your proposed system.



If present, need to check if panel is damaged or defective

Infrared camera detects presence of solar panel

Ranking of search results obtained query indexing

Feature Indexing for search: Hierarchal Clustering followed by constructing vocabulary tree.Purpose is to map high dimensional descriptors to token/words by quantizing feature space.

Automatic Segmentation using Fully Convolutional Neural Networks, using stride convolutional plus pooling

Feature extraction using SIFT and mapping with help of cosine similarity

Feature Encoding (creation of dictionary): Vector of Locally Aggregated Descriptors