Time: 15 minutes. Max marks: 10

Name: **Roll No.:**

Instructions:
- Do not plagiarize. Do not assist your classmates in plagiarism.
- Attempt all questions that you can.
- A statement is true if *all of it* is *always* true. MCQs may have multiple correct answers.
- In the unlikely case a question is not clear, discuss it with an invigilator. Please ensure that you clearly include any assumptions you make, even after clarification from the invigilator.

1. ($6 \times 1 = 6$ points) Answer True or False and provide the justification.

   (a) The Harris corner detector does not need the gradient images to identify the keypoints.
   Solution: False. The gradient images are needed for computing the auto-correlation (or covariance) matrix, which is also the matrix of coefficients of the error term.

   (b) The SIFT feature points can be robustly matched across images with varying illumination and viewpoints because they use keypoint detection at different scales.
   Solution: False. The robustness for matching largely arises from the descriptors, which are dependent on the distributino of the gradient orientation rather than the magnitude (which is more dependent on the intensity).

   (c) The normalized 8-point algorithm for Fundamental Matrix estimation requires a Euclidean transformation of the detected points.
   Solution: False. The normalization is implemented as a similarity transformation. Strictly speaking, it is restricted to only a scale & shift operation. However, since scaling is necessary, a Euclidean transformation of Isometry is not sufficient.

   (d) Unlike the 8-point algorithm for Fundamental Matrix estimation, homography estimation using DLT and RANSAC does not need any normalization.
   Solution: False. DLT for homography estimation also formulates the problem as $\mathbf{Ah} = 0$, where the rows of $\mathbf{A}$ have elements that have single pixel coordinates (x_i or y_i) as well as product terms like ($x_i y_i'$), which can lead to a poorly conditioned coefficient matrix. Thus, in order to get reasonable estimates of $\mathbf{h}$, normalization is necessary.

   (e) Harris corner detection is invariant to all changes in intensity.
   Solution: False. Harris corner is invariant to constant change (addition/subtraction) in the intensity, but only partially invariant to scaling of the intensity term. This results from the fact that the detected corners are a function of the gradient magnitude in a local region.

   (f) The SIFT descriptor uses the distribution (histogram) of the gradient magnitude around the keypoint to robustly find correspondences.
   Solution: False. The SIFT descriptor uses the distribution of gradient orientations, which results in robustness to illumination variations in the scene, as well as rotations of the image.

2. ($4 \times 1 = 4$ points) Check ***all*** the correct answers. Partial grading if a subset of the correct answers are provided. Zero if *any* provided answer is incorrect.

   (a) Say you are fitting a line in 2D by minimizing the mean squared error, and are given 100 noisy points along the $y = x + \eta$, with $\eta \sim N(\mu = 0, \sigma^2 = 0.01)$, where η is normally distributed with zero mean and variance of 0.01. There are a few outlier points lying on the x-axis between (10,0) and (15,0). What would be the slope (m) of the fitted line?
   (A) $m = \tan(\pi/4)$
   (B) $0 < m < \tan(\pi/4)$
   (C) $m > \tan(\pi/4)$
   (D) $m < 0$
   Solution: (B). Without the outliers, we would expect the fitted line to be close to $y = x$, i.e.,

have a slope of $\tan(\pi/4)$. However, with the outliers on the x-axis, we'd expect the least squares fit to be biased towards minimizing the distance w.r.t. the outliers, thereby leading to line passing through the space between the outliers and inliers, yielding a line with a slope less then $\tan(pi/4)$ but greater than 0.

(b) My student is trying to fit a 2D plane in a 3D LIDAR point cloud. While he wants to use RANSAC, he argues that instead of the usual strategy, we should pick a *slightly larger than minimal* subset of points. Why is his strategy not a good idea?
(A) RANSAC will need more sampling iterations to find a better model hypothesis than usual.
(B) RANSAC will need a more accurate threshold to find a better model hypothesis than usual.
(C) RANSAC will generate worse model hypotheses when picking an *all inlier* subset that is larger than the minimal subset size.
(D) RANSAC will just not converge.
Solution: (A). When we pick a larger than minimal subset for generating the model hypothesis, the probability of picking an *all inlier* subset drops lower than usual. However, when we do pick an all inlier subset, a larger than minimal subset usually leads to better hypothesis as our estimates are least-squares based, which would mitigate biases due to noise in the sampled points in the subset.

(c) For estimating a planar homography from images that are captured under severe lighting and viewpoint (both orientation and distance) changes, which of the following approaches would be the most effective?
(A) Image smoothing → Harris corner detection → SSD[1] (of intensity) based matching → RANSAC with DLT → Least squares fit using only inliers.
(B) SIFT detection → 1-NN[2]/2-NN ratio-based matching → RANSAC with DLT → Least squares fit using only inliers
(C) Eliminating outliers with RANSAC → SIFT detection → SSD based matching of SIFT keypoints → Least squares fit using only inliers
(D) I will use the raw images and input them into a ResNet-101 and predict the inliers before the least-squares estimation.
Solution: (B). SIFT detection takes care of any smoothing etc. necessary for the keypoing detection process. It is also the preferred choice due to severe lighting, viewpoint (including distance, i.e., scale) changes. Secondly, the 1-NN/2-NN ratio score works better than the SSD-based scoring for keypoint matching. RANSAC and DLT are typically applied at the third step to filter out outliers. The inliers are then used for better estimation using least squares. The (B) pipeline is simply better than (A) for keypoint matching before the RANSAC step. With more initial inliers, RANSAC is more likely to find the right model hypothesis and the corresponding set of inliers. (C) is absurd, since it talks about eliminating outliers using RANSAC before even detecting the keypoints. (D) is also absurd because the ResNet-101 model is a classification model. In the absence of any other information, assuming that it will predict inlier points is again, absurd.

(d) Suppose we were to compute the gradient vector for a 3D image (e.g., an MRI or CT image) where each element is a *voxel* (volume element indexed by (i, j, k)) as opposed to a *pixel* (picture element) in a 2D image. What would be the degrees of freedom for such a gradient vector's orientation?
(A) 0 (B) 1 (C) 2 (D) 3 (E) None of the above.
Solution: (C). The gradient vector will be a 3D vector and it's orientation can be given by a a unit vector in 3D, which has 2 degrees of freedom.

[1] sum of squared differences
[2] nearest neighbor