Instructions:

- Attempt all questions.
- Show intermediate steps to receive partial credit.
- All the best!

Problem 1 (5 points)

Write your name and PID on top of each page.

Problem 2 (10 points)

(a) For the smoothing operation to construct an octave in SIFT, why is a Gaussian filter used instead of any other type of smoothing filter? [2 points]

(b) Briefly state which steps allow the SIFT descriptor to achieve scale invariance. [2 points]

(c) You wish to acquire images of a precious artifact (which has some texture on it) at a museum, in order to reconstruct it with SFM. Assume you may cover the object sufficiently by moving the camera in a circle around it. See Figure 1.

![Figure 1: The imaging setup for Problem 2(c).](image-url)

Ticket prices for entrance are $5, $10, $30, $50, $100 and $200. You may acquire a maximum of $k$ images when entering with a ticket priced at $\$k$. You plan to use SIFT for correspondence. In order to spend the least money and still have a reasonable chance of success at reconstruction, which ticket should you purchase? Explain your reasoning. [2 points]
(d) Instead of SIFT, you wish to use the basic Lucas-Kanade optical flow for matching. How much should you expect to pay at the entrance? [2 points]

(e) What can you do to still use optical flow, but keep the cost lower than part (d)? [2 points]

Problem 3  (10 points)

(a) Consider a Siamese network for patch similarity, trained to output 0 for matching patches and 1 for patches that do not match. You are given two images with $n$ interest points each. If using the output of the similarity network to determine correspondence, how many forward passes (evaluations) are needed at test time? [2 points]

(b) You wish to be invariant to in-plane rotations and use a spatial transformer for it. What would be the minimal representation of the transformation parameters that the localization net must predict? Hint: How many degrees of freedom for an in-plane $2 \times 2$ rotation? [2 points]

(c) For the above rotation, write the transformation that the grid generator would apply to determine points where the input feature to the spatial transformer must be sampled? [2 points]

(d) For this and the next question, assume you may use an approximate nearest neighbors method. You now use an intermediate feature from the above patch similarity network for correspondence. How many forward passes would be needed to establish correspondence at test time? [2 points]

(e) What is a drawback of having used an intermediate feature from the network in part (d) above? How could you change the training objective to overcome it? [2 points]

Problem 4  (10 points)

You wish to implement a learning-based SFM system. You are given a structure module, $S$, that outputs the per-pixel depth map given a single image as input. You are also given a motion module, $M$, that estimates the relative pose given two images as input.

You do not have ground truth for either structure or motion, but can assume that photoconsistency holds (that is, image of the same 3D point in two views will have the same intensity). Assume that the scene is entirely static, while only the camera moves. Also assume that occlusions are negligible, so every scene point of interest may be considered visible in all the cameras.

(a) Draw a block diagram for how you would design the SFM system to be trained using two consecutive images $I_1$ and $I_2$ as input. Write an expression for the training objective. [4 points]

(b) Draw a block diagram for how you would design the SFM system to be trained using three consecutive images $I_1$, $I_2$ and $I_3$ as input. Write an expression for the training objective. [6 points]
Problem 5  (15 points)

(a) Given corresponding points \( x_i \in \mathbb{R}^2 \) in image 1 and \( x'_i \in \mathbb{R}^2 \) in image 2, for \( i = 1, \ldots, n \). What is the minimum \( n \) for any method to be able to estimate the fundamental matrix? [2 points]

(b) Suppose you are making a real-time system. Is the above \( n \) the number of points you might use in a RANSAC method to estimate the fundamental matrix? Why or why not? [2 points]

(c) If you are already provided the intrinsic calibration matrix for the two cameras, how many correspondences would you use within RANSAC to compute the relative pose? [2 points]

(d) On what factors does the required number of trials for RANSAC depend? [2 points]

(e) You are now given a set of correspondences \( X \). Let \( \text{modelSize} \) be the number of correspondences from part (c) above. You want to implement RANSAC which gives you the correct answer with a 99% probability, but using as few iterations as possible. You expect your data to have no more than 50% outliers, but do not know exactly how many outliers. Replace the \( \text{??A}, \text{??B}, \ldots, \text{??G} \) in the code snippet below in order to do RANSAC with minimum number of iterations. In your answer sheet, simply write down the corresponding answer for each of \( \text{??A}, \text{??B}, \ldots, \text{??G} \). [7 points]

Inputs: \( X \), modelSize

\[
\text{numCorrespondences} = \text{getNumCorrespondences}(X); \quad \text{// Assume implemented}
\]
\[
\text{probabilityOfCorrectness} = \text{??A};
\]
\[
\text{maxOutlierRatio} = \text{??B};
\]
\[
\text{MAXITER} = \text{??C};
\]
\[
\text{iter} = 0;
\]
\[
\text{while} \quad (\text{iter} < \text{MAXITER})
\]
\[
\quad \text{iter} = \text{iter} + 1;
\quad \text{randomSet} = \text{pickRandomSet}(\text{modelSize}, X); \quad \text{//Assume this is implemented}
\quad \text{model} = \text{fitModel}(\text{randomSet}); \quad \text{//Assume this is implemented}
\quad \text{numInliers} = \text{findNumberOfInliers}(\text{model}, X); \quad \text{//Assume this is implemented}
\quad \text{outlierRatio} = \text{??D};
\quad \text{if} \quad (\text{??E})
\quad \quad \text{maxOutlierRatio} = \text{??F};
\quad \quad \text{MAXITER} = \text{??G};
\quad \}
\]

---

3
Problem 6   (14 points)

You want to build a self-driving bus to operate within the UCSD campus, that uses structure from motion (SFM) to localize its position. It has two constraints:

- **Constraint 1:** To keep costs low, the only measurement available is a stream of images acquired through a single camera.
- **Constraint 2:** The bus must ferry students quickly between classes, so it always takes the shortest path between two points.

(a) You observe that the estimated trajectory was accurate at first, but after some time, the errors progressively increase due to scale drift. Explain the reason for scale drift in SFM. [3 points]

(b) To fix the scale drift, you get a dataset of geo-tagged images of the UCSD campus. That is, for every image, you have information on the precise location where it was acquired. You also managed to write code that implements an image retrieval system. Explain how you might use this dataset and code to resolve scale drift in your SFM system as your bus drives across the UCSD campus (the two constraints above must still be satisfied). [4 points]

(c) You decide to remove the first constraint and mount a second camera on the bus. It has the exact same intrinsics as the first camera and is mounted such that there is no relative rotation with respect to the first one, but just a translational distance of 1m in the $x$-direction. Suggest how you may resolve the scale drift using this setup. [4 points]

(d) What is the essential matrix between the two cameras in the above setup? [3 points]

Problem 7   (6 points)

(a) State one challenge of metric learning using the triplet loss on a large-scale dataset and how it may be overcome in practice. [2 points]

(b) State one opportunity and one challenge for large-scale face recognition, in contrast to generic object recognition, arising due to the fact that faces have a special structure. [2 points]

(c) Explain the trade-off between the identification and verification objectives for learning features in large-scale face recognition. [2 points]