The oral exam will last 30 minutes and will consist of one application question followed by two theoretical questions.

Please find below a **non-exhaustive** list of possible application questions.

The list of theoretical question is instead **exhaustive**, i.e., it contains all the topics that you should learn about the course.

Application questions

1. Summarize the building blocks of a visual odometry (VO) or visual SLAM (VSLAM) algorithm.
2. Augmented reality (AR) is a view of a physical scene augmented by computer-generated sensory inputs, such as data or graphics. Suppose you want to design an augmented reality system that super-imposes text labels to the image of real physical objects. Summarize the building blocks of an AR algorithm.
3. Suppose that your task is to reconstruct an object from different views. How do you proceed?
4. Building a panorama stitching application. Summarize the building blocks.
5. How would you design a mobile tourist app? The user points the phone in the direction of a landmark and the app displays tag with the name of it. How would you implement it?
6. Assume that we have several images downloaded from flicker showing the two towers of Grossmünster. Since such images were uploaded by different persons they will have different camera parameters (intrinsic and extrinsic), different lighting, different resolutions and so on. If you were supposed to create a 3D model of Grossmünster, what kind of approach would you use? Can you get a dense 3D model or it will be a sparse one? Please explain the pipeline that you propose for this scenario.
7. Assume that you move around a statue with a camera and take pictures in a way that the statue is not far from the camera and always completely visible in the image. If you were supposed to find out where the pictures were taken, what would you do with the images? What kind of approach would you use? Since the camera motion is around the statue, the images contain different parts of the statue. How do you deal with this problem?
8. Suppose that you have two robots exploring an environment, explain how the robots should localize themselves and each other with respect to the environment? What are the alternative solutions?

Theoretical questions

01 – Introduction

1. Provide a definition of Visual Odometry.
2. Explain the most important differences between VO, VSLAM and SFM.
3. Describe the needed assumptions for VO.
4. Illustrate its building blocks

02-03 – Image Formation

1. Explain what a blur circle is
2. Derive the thin lens equation and perform the pinhole approximation
3. Define vanishing points and lines
4. Prove that parallel lines intersect at vanishing points
5. Explain how to build an Ames room
6. Derive a relation between the field of view and the focal length
7. Explain and write the equations of the perspective projection, including lens distortion and world to camera projection.
8. Given an image and the associated camera pose, how would you superimpose a virtual object on the image (for example, a virtual cube). Describe the steps involved.
9. Normalized image coordinates and geometric explanation.
10. Describe the general PnP problem and derive the behavior of its solutions. What's the minimum number of points and what are the degenerate configurations?
11. Explain the working principle of the P3P algorithm. What are the algebraic trigonometric equations that it attempts to solve?
12. Explain and derive the DLT for both 3D objects or planar grids. What is the minimum number of point correspondences it requires (for 3D objects and for planar grids)?
13. Define central and non-central omnidirectional cameras.
14. What kind of mirrors ensure central projection.
15. What do we mean by normalized image coordinates on the unit sphere?

04 – Filtering and edge detection

1. Explain the differences between convolution and correlation
2. Explain the differences between a box filter and a Gaussian filter
3. Explain why one should increase the size of the kernel of a Gaussian filter if 2σ is close to the size of the kernel
4. Explain when we would need a median & bilateral filter
5. Explain how to handle boundary issues
6. Explain the working principle of edge detection with a $1D$ signal
7. Explain how noise does affect this procedure
8. Explain the differential property of convolution
9. Show how to compute the first derivative of an image intensity function along x and y
10. Explain why the Laplacian of Gaussian operator is useful
11. List the properties of smoothing and derivative filters
12. Illustrate the Canny edge detection algorithm
13. Explain what non-maxima suppression is and how it is implemented

05-06 – Point feature detection, descriptor, and matching

1. Explain what is template matching and how it is implemented
2. Explain what are the limitations of template matching. Can you use it to recognize cars.
3. Illustrate the similarity metrics SSD, SAD, NCC, and Census transform
4. What is the intuitive explanation behind SSD and NCC
5. Explain what are good features to track. In particular, can you explain what are corners and blobs together with their pros and cons. How is their localization accuracy?
6. Explain the Harris corner detector. In particular:

   a. Use the Moravec definition of corner, edge and flat region.
   b. Show how to get the second moment matrix from the definition of SSD and first order approximation (show that this is a quadratic expression) and what is the intrinsic interpretation of the second moment matrix using an ellipse?
   c. What is the M matrix like for an edge, for a flat region, for an axis-aligned 90-degree corner and for a non-axis—aligned 90-degree corner?
   d. What do the eigenvalues of M reveal?
   e. Can you compare Harris detection with Shi-Tomasi detection?
   f. Can you explain whether the Harris detector is invariant to illumination or scale changes? Is it invariant to view point changes?
   g. What is the repeatability of the Harris detector after rescaling by a factor of 2?
1. How does automatic scale selection work?
2. What are the good and the bad properties that a function for automatic scale selection should have or not have?
3. How can we implement scale invariant detection efficiently? (show that we can do this by resampling the image vs rescaling the kernel).
4. What is a feature descriptor? (patch of intensity value vs histogram of oriented gradients). How do we match descriptors?
5. How is the keypoint detection done in SIFT and how does this differ from Harris?
6. How does SIFT achieve orientation invariance?
7. How is the SIFT descriptor built?
8. What is the repeatability of the SIFT detector after a rescaling of 2? And for a 50 degrees' viewpoint change?
9. Illustrate the 1st to 2nd closest ratio of SIFT detection: what's the intuitive reasoning behind it? Where does the 0.8 factor come from?
10. How does the FAST detector work? What are its pros and cons compared with Harris?

07 – Stereo Vision

1. Can you relate Structure from Motion to 3D reconstruction? What's their difference?
2. Can you define disparity in both the simplified and the general case?
3. Can you provide a mathematical expression of depth as a function of the baseline, the disparity and the focal length?
4. Can you apply error propagation to derive an expression for depth uncertainty and express it as a function of depth? How can we improve the uncertainty?
5. Can you analyze the effects of a large/small baseline? Can you illustrate it with a sketch?
6. What is the closest depth that a stereo camera can measure?
7. Are you able to show mathematically how to compute the intersection of two lines (linearly and non-linearly)?
8. What is the geometric interpretation of the linear and non-linear approaches and what error do they minimize?
9. Are you able to provide a definition of epipole, epipolar line and epipolar plane?
10. Are you able to draw the epipolar lines for two converging cameras, for a forward motion situation, and for a side-moving camera?
11. Are you able to define stereo rectification and to derive mathematically the rectifying homographies?
12. How is the disparity map computed?

13. How can we establish stereo correspondences with subpixel accuracy?
14. Describe one or more simple ways to reject outliers in stereo correspondences.
15. Is stereo vision the only way of estimating depth information? If not, are you able to list alternative options? (make link to other lectures of course)

08-09– Multiple view geometry 2 and 3

1. What's the minimum number of correspondences required for calibrated SFM and why?
2. Are you able to derive the epipolar constraint?
3. Are you able to define the essential matrix?
4. Are you able to derive the 8-point algorithm?
5. How many rotation-translation combinations can the essential matrix be decomposed into?
6. Are you able to provide a geometrical interpretation of the epipolar constraint?
7. Are you able to describe the relation between the essential and the fundamental matrix?
8. Why is it important to normalize the point coordinates in the 8-point algorithm?
9. Describe one or more possible ways to achieve this normalization.
10. Are you able to describe the normalized 8-point algorithm?
11. Are you able to provide quality metrics for the essential matrix estimation?
12. Why do we need RANSAC?
13. What is the theoretical maximum number of combinations to explore?
14. After how many iterations can RANSAC be stopped to guarantee a given success probability?
15. What is the trend of RANSAC vs. iterations, vs. the fraction of outliers, vs. the number of points to estimate the model?
16. How do we apply RANSAC to the 8-point algorithm, DLT, P3P?
17. How can we reduce the number of RANSAC iterations for the SFM problem? (1- and 2-point RANSAC)
18. Bundle Adjustment and Pose Graph Optimization. Mathematical expressions and illustrations. Pros and cons.
19. Are you able to describe hierarchical and sequential SFM for monocular VO?
20. What are keyframes? Why do we need them and how can we select them?
21. Are you able to define loop closure detection? Why do we need loops?
22. Are you able to provide a list of the most popular open source VO and VSLAM algorithms?
23. Are you able to describe the differences between feature-based methods and direct methods?
24. Sparse vs semi-dense vs dense. What are their pros and cons?
25. General definition of VO and comparison with respect VSLAM and SFM. Definition of loop closure detection (why do we need loops?). How can we detect loop closures? (make link to other lectures).

10 – Multiple view geometry 4 (benchmarking visual SLAM) not covered in class, so it won't be asked.

11 – Tracking

1. Are you able to illustrate tracking with block matching?
2. Are you able to explain the underlying assumptions behind differential methods, derive their mathematical expression and the meaning of the M matrix?
3. When is this matrix invertible and when not?
4. What is the aperture problem and how can we overcome it?
5. What is optical flow?

6. Can you list pros and cons of block-based vs. differential methods for tracking?
7. Are you able to describe the working principle of KLT?
8. What functional does KLT minimize? (proof won't be asked, only the first two slides titled "derivation of the Lucas-Kanade algorithm")
9. What is the Hessian matrix and for which warping function does it coincide to that used for point tracking?
10. Can you list Lukas-Kanade failure cases and how to overcome them?
11. How do we get the initial guess?
12. Illustrate alternative tracking using point features.

12a – Dense 3D Reconstruction

1. Are you able to describe the multi-view stereo working principle? (aggregated photometric error)
2. What are the differences in the behavior of the aggregated photometric error for corners, flat regions, and edges?
3. What is the disparity space image (DSI) and how is it built in practice?
4. How do we extract the depth from the DSI?
5. How do we enforce smoothness (regularization) and how do we incorporate depth discontinuities (mathematical expressions)?
6. What happens if we increase lambda (the regularization term)? What if lambda is 0? And if lambda is too big?
7. What is the optimal baseline for multi-view stereo?
8. What are the advantages of GPUs?

12b – Place Recognition

1. What is an inverted file index?
2. What is a visual word?
3. How does K-means clustering work?
4. Why do we need hierarchical clustering?
5. Explain and illustrate image retrieval using Bag of Words.
6. Discussion on place recognition: what are the open challenges and what solutions have been proposed?

12c – Deep Learning – Won't be asked at the exam

13 – Visual inertial fusion

1. Are you able to answer the following questions?
2. Why is it recommended to use an IMU for Visual Odometry?
3. Why not just an IMU, without a camera?
4. How does a MEMS IMU work?
5. What is the drift of an industrial IMU?
6. What is the IMU measurement model?
7. What causes the bias in an IMU?
8. How do we model the bias?
9. How do we integrate the acceleration to get the position (formula)?
10. What is the definition of loosely coupled and tightly coupled visual inertial fusions?

11. How can we use non-linear optimization-based approaches to solve for visual inertial fusion?
12. Can you write down the cost function of smoothing methods and illustrate its meaning?

14 – Event-based Vision

1. Are you able to answer the following questions?
2. What is a DVS and how does it work?
3. What are its pros and cons vs. standard cameras?
4. Can we apply standard camera calibration techniques?
5. How can we compute optical flow with a DVS?
6. Could you intuitively explain why we can reconstruct the intensity?
7. What is the generative model of a DVS (formula)? Can you derive its 1st order approximation?
8. What is a DAVIS sensor?
9. What is the focus maximization framework and how does it work? What is its advantage compared with the generative model?
10. How can we get color events?