

L9.2 Modern SLAM for AD Quiz ANS

1. A pinhole camera has $f = 800$ px and principal point $c_x = 640$ px. A point at $X = 2$ m, $Z = 10$ m (camera frame) projects to which horizontal pixel u ?

- A) $u = 640$
- B) $u = 160$
- C) $u = 800$
- D) $u = 1,440$

ANS: C) $u = 800$

2. Why can a single (monocular) camera not recover metric scale?

- A) Scaling the whole scene and the camera translation by the same factor produces identical images
- B) Monocular cameras have no lens distortion model
- C) The focal length is always unknown in a monocular camera
- D) Feature matching fails when only one camera is used

ANS: A) Scaling the whole scene and the camera translation by the same factor produces identical images

3. What do ORB features consist of, and how are descriptors compared?

- A) Gradient histograms of 128 floats, compared with the Euclidean distance
- B) Raw pixel patches, compared with the cross-correlation of the whole image
- C) Learned 3-D point descriptors, compared with the Mahalanobis distance
- D) Oriented FAST corners with rotated BRIEF binary descriptors, compared with the Hamming distance

ANS: D) Oriented FAST corners with rotated BRIEF binary descriptors, compared with the Hamming distance

4. What does the ratio test do in feature matching?

- A) It removes all features whose descriptor has more than half of its bits set
- B) It rejects a match if the best descriptor distance is not clearly smaller than the second best (ambiguous match)
- C) It computes the ratio between the image width and height to rescale the features
- D) It keeps only the match with the largest descriptor distance

ANS: B) It rejects a match if the best descriptor distance is not clearly smaller than the second best (ambiguous match)

5. RANSAC with sample size $s = 3$ and inlier ratio $w = 0.5$ needs about how many iterations for 99 % confidence? ($N = \log(1 - p) / \log(1 - w^s)$)

- A) About 35
- B) About 3
- C) About 8
- D) About 1,200

ANS: A) About 35

6. What does the epipolar constraint $\mathbf{x}_2^T \mathbf{E} \mathbf{x}_1 = 0$ with $\mathbf{E} = [\mathbf{t}] \times \mathbf{R}$ allow us to estimate?

- A) The absolute position of the camera in the world frame, with metric scale
- B) The depth of every pixel of the first image
- C) The relative rotation $\mathbf{R}$ and the direction of the translation $\mathbf{t}$ (but not its length) between two views
- D) The intrinsic parameters of the camera (f , c_x , c_y)

ANS: C) The relative rotation $\mathbf{R}$ and the direction of the translation $\mathbf{t}$ (but not its length) between two views

7. In ORB-SLAM tracking, what problem does PnP (Perspective-n-Point) solve?

- A) The 3-D position of a point from two known camera poses
- B) The camera pose from known 3-D map points and their 2-D pixel observations
- C) The loop-closure candidate with the most similar bag of words
- D) The scale of a monocular map from the IMU

ANS: B) The camera pose from known 3-D map points and their 2-D pixel observations

8. What does bundle adjustment minimize?

- A) The Hamming distance between all pairs of ORB descriptors
- B) The time between consecutive camera frames
- C) The number of keyframes stored in the map
- D) The (robust) sum of squared reprojection errors over camera poses and 3-D map points

ANS: D) The (robust) sum of squared reprojection errors over camera poses and 3-D map points

9. Which set describes the main parts of ORB-SLAM3?

- A) Tracking, local mapping, and loop closing / map merging threads, plus an Atlas of multiple maps
- B) A detector, a tracker, and a planner
- C) Odometry at 10 Hz, mapping at 1 Hz, and no loop closure
- D) A camera front-end, an EKF back-end with N landmarks, and particles

ANS: A) Tracking, local mapping, and loop closing / map merging threads, plus an Atlas of multiple maps

10. What does adding an IMU to visual SLAM mainly provide?

- A) Dense color for every map point
- B) Removal of the need for camera calibration
- C) Metric scale and gravity direction, and the ability to bridge short visual failures such as blur
- D) Immunity to GNSS spoofing

ANS: C) Metric scale and gravity direction, and the ability to bridge short visual failures such as blur

11. In LOAM, the smoothness value c of a LiDAR point is computed from its neighbors on the scan line. How is it used?

- A) A large c marks a planar feature; a small c marks an edge feature
- B) A large c marks an edge feature (matched to lines); a small c marks a planar feature (matched to planes)
- C) c is the reflectivity of the point and is used for place recognition
- D) c is the time stamp of the point and is used to deskew the scan

ANS: B) A large c marks an edge feature (matched to lines); a small c marks a planar feature (matched to planes)

12. What is the idea of the dual-rate design of LOAM?

- A) Two LiDARs running at different frequencies, fused with a Kalman filter
- B) Two separate maps, one for edges and one for planes, merged at the end
- C) A camera at 30 Hz and a LiDAR at 10 Hz, synchronized by the IMU
- D) A fast, coarse scan-to-scan odometry (about 10 Hz) plus a slower, more accurate scan-to-map mapping step

ANS: D) A fast, coarse scan-to-scan odometry (about 10 Hz) plus a slower, more accurate scan-to-map mapping step

13. Which description matches LIO-SAM?

- A) A pure EKF with N landmarks and range-bearing measurements
- B) A bag-of-words visual place recognition method
- C) A factor graph with IMU preintegration, LiDAR odometry, optional GNSS, and loop-closure factors, solved incrementally with iSAM2
- D) A particle filter that samples the LiDAR scan

ANS: C) A factor graph with IMU preintegration, LiDAR odometry, optional GNSS, and loop-closure factors, solved incrementally with iSAM2

14. Why does scan matching fail in a long, straight corridor?

- A) All plane normals are perpendicular to the corridor axis, so $H = \sum n n^T$ has a zero eigenvalue along the corridor
- B) The LiDAR cannot measure distances to walls
- C) The walls reflect too much light and saturate the sensor
- D) ICP always needs at least 1,000 points to converge

ANS: A) All plane normals are perpendicular to the corridor axis, so $H = \sum n n^T$ has a zero eigenvalue along the corridor

15. A lane is 3.7 m wide and the car is 1.9 m wide. If 3σ of the lateral localization error must stay within the one-sided margin, what is the largest σ ?

- A) About 0.9 m
- B) About 0.3 m
- C) About 1.8 m
- D) About 0.05 m

ANS: B) About 0.3 m

16. Which recipe do Autoware and Apollo both follow for localization in a prior map?

- A) Pure visual SLAM with a monocular camera and no prior map
- B) GNSS only, because RTK is always available
- C) A neural network that outputs the pose directly, with no filter
- D) IMU/odometry propagation, LiDAR scan matching against the map, and GNSS correction in a Kalman filter

ANS: D) IMU/odometry propagation, LiDAR scan matching against the map, and GNSS correction in a Kalman filter

17. An attacker adds a GNSS bias that grows by 0.12 m per step. Why does a per-step innovation (χ^2) gate barely notice it?

- A) Each step's disagreement is within the noise, so the filter absorbs the bias and follows the spoofer; a windowed check is needed
- B) The χ^2 test is only valid for LiDAR measurements
- C) The gate threshold is always larger than 100 m
- D) The attacker removes the innovation from the filter's memory

ANS: A) Each step's disagreement is within the noise, so the filter absorbs the bias and follows the spoofer; a windowed check is needed

18. What is the main lesson of the FusionRipper attack on multi-sensor fusion (Shen et al., 2020)?

- A) Multi-sensor fusion makes GPS spoofing impossible
- B) The attack only works if the car has no IMU
- C) Fusing several sensors is necessary but not sufficient: the fusion logic itself can be exploited by GPS spoofing to make the pose drift
- D) The attack requires physical access to the LiDAR

ANS: C) Fusing several sensors is necessary but not sufficient: the fusion logic itself can be exploited by GPS spoofing to make the pose drift

19. What does SLAMspooftarget, and what did it achieve in real-world tests?

- A) The camera exposure; it blinded the car for one second
- B) LiDAR-based scan matching in localization; position errors of 4.2 m or more against three popular algorithms
- C) The HD map file on disk; it deleted lane markings
- D) The wheel encoders; it changed the odometry scale by 1 %

ANS: B) LiDAR-based scan matching in localization; position errors of 4.2 m or more against three popular algorithms

20. A localization pose arrives 80 ms late while the car drives at 25 m/s. How far does the car travel in that time?

- A) 0.2 m
- B) 0.32 m
- C) 2.0 m
- D) 20 m

ANS: C) 2.0 m