EPFL

1

**Teacher : Prof. Pascal Fua**
**CS-442 Computer Vision - MA**
**02/07/2023**
**Duration : 90 minutes**

# Student 1

SCIPER : **999000**

**Do not turn the page before the start of the exam. This document is double-sided, has 12 pages, the last ones possibly blank. Do not unstaple.**



- Place your student card on your table.
- **A one page two-sided hand-written cheat-sheet** is allowed to be used during the exam.
- Using a **calculator** or any electronic device is not permitted during the exam.
- All questions have one or more correct answers.
- The grading scheme is such that random answering is discouraged:
  - Each answer of a multiple choice question is awarded +1 point if correct and −1 point if incorrect. If the **whole** question is left unanswered no points (positive nor negative) are awarded. Note that "correct" means that a true answer should be ticked and that a false one should be left unticked.

| | Correct answers: | Student's answers: | Grading: |
|---|---|---|---|
| a) | ■ | ■ | +1 |
| b) | ☐ | ■ | -1 |
| c) | ■ | ☐ | -1 |
| d) | ☐ | ☐ | +1 |

  - The scores for separate questions are **not clipped to 0**, that is, you can get negative score for a question.
- Use a **black or dark blue ballpen** and clearly erase with **correction fluid** if necessary.
- If a question is wrong, the teacher may decide to nullify it.

Respectez les consignes suivantes | Observe this guidelines | Beachten Sie bitte die unten stehenden Richtlinien

| choisir une réponse \| select an answer<br>Antwort auswählen | ne PAS choisir une réponse \| NOT select an answer<br>NICHT Antwort auswählen | Corriger une réponse \| Correct an answer<br>Antwort korrigieren |
|---|---|---|
| ☒ ☑ ■ | ☐ | ☐ |

ce qu'il ne faut **PAS** faire | what should **NOT** be done | was man **NICHT** tun sollte

Pour votre examen, imprimez de préférence les documents compilés à l'aide de auto-multiple-choice.

## First part: Multiple choice questions

For each question, mark the box corresponding to the correct answer. Each question has **at least one** correct answer.

**Question 1** Which of the following statements about methods for removing noise from images is(are) true?

- ☐ The Sobel filter is a common low-pass filter to eliminate high-frequency noise.
- ■ The Discrete Fourier Transform(DFT) can be utilized to suppress the high-frequency noise.
- ■ The Gaussian kernel and $X$-derivative $\frac{\partial f}{\partial x}$ (or $Y$-derivative $\frac{\partial f}{\partial y}$) can be merged into one operator due to the property of separability, which can accelerate computation.
- ■ One of the reasons that we need to remove noise from images when computing gradients is that differentiating emphasizes high-frequency noise.

**Question 2** Assume we have the following operators (kernels) for edge detection in 2D images. Which filter(s) is/are more robust to Gaussian noise?

- ■ $\frac{\partial I}{\partial y} = \begin{bmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}$
- ☐ $\frac{\partial I}{\partial y} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$
- ■ $\frac{\partial I}{\partial x} = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix}$
- ☐ $\frac{\partial I}{\partial x} = \begin{bmatrix} -1 & 0 & 1 \end{bmatrix}$

**Question 3** Which of the following statement(s) about edge detection is(are) true?

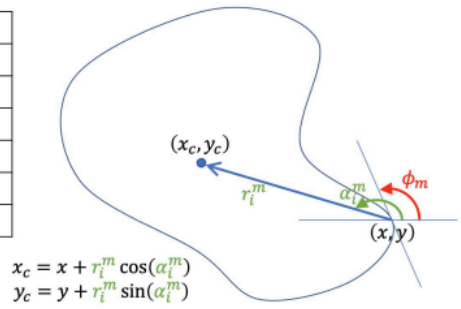
- ■ Gaussian smoothing is an important part of the Canny edge detector.
- ■ The gradient of an image is defined as the direction of the most rapid change in intensity.
- ☐ The computation of Gaussian smoothing can be more efficient than 1D-convolution.
- ■ The canny edge detector doesn't require training data.

**Question 4** We have seen several methods based on Machine Learning. Which of the following statements is(are) true?

- ■ U-Net is an Encoder-Decoder architecture, that is commonly used in computer vision tasks.
- ☐ The pooling layer will increase the output size.
- ■ The logistic loss cannot handle a classification problem with a complex non-linear boundary.

**Question 5** Which of the following statement(s) about Hough Transform is(are) true?

| $\phi$ | $R(\phi_i)$ |
|---|---|
| $\phi_1$ | $(r_1^1, \alpha_1^1), (r_2^1, \alpha_2^1), \ldots\ldots (r_{n1}^1, \alpha_{n1}^1)$ |
| $\phi_2$ | $(r_1^2, \alpha_1^2), (r_2^2, \alpha_2^2), \ldots\ldots (r_{n2}^2, \alpha_{n2}^2)$ |
| .. | ..... |
| $\phi_m$ | $(r_1^m, \alpha_1^m), (r_2^m, \alpha_2^m), \ldots\ldots (r_i^m, \alpha_i^m) \ldots\ldots (r_{nm}^m, \alpha_{nm}^m)$ |
| .. | ..... |
| $\phi_M$ | $(r_1^M, \alpha_1^M), (r_2^M, \alpha_2^M), \ldots\ldots (r_{nM}^M, \alpha_{nM}^M)$ |




- ☐ In the generalized Hough transform algorithm, an accumulator for the displacement vectors $(r_i^j, \alpha_i^j)$ is used.
- ■ The Hough transform is a voting scheme to find a matched template shape given the locations of edge points with their gradient direction and magnitude.
- ☐ Scaling the template doesn't change the R-table in the generalized Hough transform.
- ■ As the computational cost grows exponentially with the number of model parameters, the generalized Hough transform only works with simple shapes that can be defined with a small number of parameters.

**Question 6** In the human vision system, which of the following statements is/are correct:

- ■ In general, eye lenses keep the focal plane near the retina.
- ☐ The reaction time of human vision is at the nano second level.
- ☐ Cones sense color and rods sense depth.
- ■ Ganglion cells serve the role of filter banks.

**Question 7** Which of the following statements is(are) correct regarding the pinhole camera:

- ■ If the hole size is decreased, the image becomes dimmer.
- ☐ The projection performed by a pinhole camera is an orthogonal projection.
- ■ The process of estimating the parameters of a camera is often referred to as camera calibration.
- ☐ If the hole size is very large, the image will be blurred because of diffraction.

**Question 8** In monocular camera calibration, which of the following statements is(are) correct:

- ☐ After calibration, we can project a 2D pixel back to the corresponding 3D point.
- ☐ The internal parameters include focal length, aperture, and skew.
- ■ The external parameters include rotation and translation.
- ■ After calibration, we can project a 3D point to the image plane via the projection matrix.
- ■ Homogeneous coordinates help formulate the projection as a linear operation.

**Question 9** When 3D scenes are projected to images via a calibrated pinhole camera, which of the following statements is(are) correct:

- ■ Projected parallel lines meet at vanishing points.
- ■ Assume that two 3D objects move with the same speed from left to right, the image projection of the distant one moves slower than that of the closer one.
- ☐ The camera needs to be recalibrated for each new 3D scene even if its internal parameters do not change.
- ☐ There is at least one vanishing point inside an image.

**Question 10** For a physical camera which captures 8-bit RGB images, which of the following statements is(are) correct:

- ☐ The image pixels can represent up to $3 \times 2^8$ colors.
- ☐ If we increase the aperture, the depth of field increases.
- ■ If we increase the focal length, the depth of field decreases.
- ■ The captured image can get darker towards its edges. This is called vignetting.
- ■ Pixel value is connected to the electrical charges activated by the photons hitting the sensor array.

**Question 11** Which of the following segmentation methods relies on a graph to partition an image into segments by minimizing a cost function?

- ☐ Region Growing
- ☐ Histogram Splitting
- ☐ K-means Clustering
- ■ ST Min-Cuts

**Question 12** In region growing segmentation, a neighboring pixel $p$ is added to a region $R$ if it satisfies which condition? Denote $I(p)$ the intensity of pixel $p$, $I(q)$ the intensity of a pixel $q$ already in the region.

- ■ $p = \text{argmin}_{p'} \min_{q \in R} |I(p') - I(q)|$
- ☐ $\forall q \in R, I(p) \geq I(q)$
- ☐ $\forall q \in R, I(p) < I(q)$
- ☐ $p = \text{argmin}_{p'} \max_{q \in R} |I(p') - I(q)|$

**Question 13** Histogram splitting involves finding a threshold $T$ that separates the different objects in the image, using the histogram of pixel values $H$. For simplicity, let us assume the image is gray-scale, and we are segmenting a relatively small and bright object in front of a darker background. The histograms represents all possible pixel values $H = (H_i)_{i \in [0,255]}$ and $H_i$ denotes the number of pixels with intensity $i$. Which of the following strategies is/are a reasonable choice to pick $T$?

- ■ Visualize the histogram $H$ and manually set $T$.
- ■ Choose the $T$ that minimizes: $\text{Variance}(H_{i<T})^2 + \text{Variance}(H_{i>T})^2$.
- ☐ $T = \max_i(H_i)$
- ☐ Choose $T$ to be the median of $H$.

**Question 14** Interactive segmentation methods involve user input to guide the segmentation process. Which of the following statements about interactive segmentation is/are true?

- ■ In ST Min-Cut, the user can refine the boundaries of segmented regions by providing additional foreground/background pixels.
- ■ In K-Means segmentation, the user can improve the segmentation by correctively setting the number of clusters.
- ☐ Using a U-Net for segmentation, the user can specify prompt points in low-dimensionality space.
- ■ User interaction can help reduce ambiguity in segmentation.

**Question 15** When attempting to use the following reflectance map equation for shape-from-shading reconstruction:

$$I = albedo * \mathbf{N} \cdot \mathbf{L}, \tag{1}$$

where $\mathbf{N}$ denotes the surface normal and $\mathbf{L}$ denotes the light source direction, which of the following actions can help resolve the bas-relief ambiguity?

- ☐ Increasing the resolution of the captured image.
- ■ Explicitly modeling the distance to light source in the equation.
- ☐ Adding an integrability term as regularization.
- ☐ Adding a smoothness term as regularization.

**Question 16** When using the relectance map equation to do 3D reconstruction with shape-from-shading, which of the following statements is/are true?

- ■ Having multiple light sources that can be selectively turned on and off is helpful.
- ☐ In general, neither of the smoothness term and integrability term may be minimized to zero at the global minimum.
- ■ Secondary illumination must be minimized in the scene to be reconstructed.
- ■ There are more unknowns per pixel than there are measurements when writing the equations.

**Question 17** In an art gallery, you see a beautiful marble statue. You decide to use shape-from-shading to reconstruct its 3D shape from a single image, but always fail with your reflectance map method. What is/are the possible reason(s)?

- ■ There are shadows on the surface of statue.
- ■ There is indirect lighting around the art gallery.
- ■ There are specularities on the surface of the statue.
- ☐ The surface is bumpy.

**Question 18** Which of the following statements regarding shape-from-shading methods is/are true?

- ☐ Shape-from-shading methods require only a single image as input, but they recover high-frequency details less precisely than shape-from-stereo methods.
- ☐ Shape-from-shading cannot deal with surfaces with specularities.
- ■ Shape-from-shading is an ill-posed inverse problem.
- ■ Shape-from-shading can be used in conjunction with shape-from-stereo for better performance.

**Question 19** While trying to extract shape of an object using silhouettes, we capture multiple images of it. However, we could not extract the shape accurately. What are the possible reasons for this failure:

- ■ The images are not captured from enough angles.
- ■ The object has concavities on its surface.
- ■ The cameras are ill-calibrated.
- ☐ The object has a convex shape.

**Question 20** Shape from motion is preferable to shape from stereo and shape from contours in the following scenarios:

- ■ Capturing the shape of large scale scenes and objects.
- ■ Camera poses are not precisely known.
- ☐ Capturing convex shaped objects.
- ☐ Capturing microscopic details on an object.

**Question 21** Which of the following is (are) true about making use of the DFT (Discrete Fourier Transform) in computer vision?

- ☐ The DFT is useful for separating objects having the same shape but different colors.
- ■ It is possible to train a convolutional neural network on the DFT of images rather than using raw pixel values.
- ■ The DFT of some homogeneous structural textures is characteristic, and can be used to easily detect and classify them.
- ☐ The DFT is an abstract mathematical transformation on an image and cannot be visualized.

**Question 22** This question was removed as it was found to be ambiguous.

**Question 23** Which of the following is (are) true about recovering shape from texture?

- ☐ Recovering shape from texture is possible for any input image.
- ■ Strong assumptions such as texture homogeneity are necessary for it to work.
- ☐ It cannot be done using machine learning based methods.
- ■ Classical approaches to recovering shape from structured textures rely on projective geometry.

**Question 24** Which of the following is (are) **false** about textures?

- ☐ Statistical approaches can be used to help characterize textures that have no clear structure.
- ☐ Texels are the smallest texture unit, and repeating texels creates texture.
- ■ A fixed bank of 12 Gabor filters called the Gabor primitives are commonly used to process textures.
- ■ Textures are a local measure computed pixel-wise, i.e. independently for each pixel.

**Question 25** What is the influence of the window size in shape from stereo?

- ☐ A small window will often cover the silhouette of an object, leading to improved results.
- ☐ A large window will often cover the silhouette of an object, leading to improved results.
- ■ A small window delivers an improved precision when correct.
- ■ A large window is robust to noise in the input images.

## Second part, open questions

Answer in the empty space below each question. Your answer should be carefully justified, and all the steps of your argument should be discussed in details.
**Leave the check-boxes empty, they are used for grading.**

### Shape From Stereo

Your friend Amy has recently completed a computer vision course and is excited to build her own stereo vision system with the goal of reconstructing the 3D point cloud of her room. She has purchased two cameras and wants to design a camera stand to mount them. Amy is unsure how to position the cameras for optimal performance. She has asked for your guidance on how to properly place the two cameras in her stereo system.

**Question 26:** *This question is worth 6 points.*

☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ■6

How should Amy place the cameras, and why is this positioning important? Please list at least three considerations and explain the reasons for each.

**The cameras should be mounted horizontally (1pt); The cameras should be aligned with the same orientation (1pt); A proper baseline distance should be chosen based on the expected range of distances to the objects (1pt). The epipolar lines would be tilted (1pt); The epipolar lines would not be parallel (converge or diverge) (1pt); Narrow baseline for close objects, and wide baseline for distant objects (1pt) (other reasonable answers such as robustness to matching noise would be also correct).**



**Question 27:** *This question is worth 2 points.*

☐0 ☐1 ■2

Formulate the process of 3D reconstruction from two images captured by Amy's stereo system mathematically, assuming the cameras are properly positioned and the intrinsics are known.
*Hint: The process involves converting pixel disparity to 3D points.*

**Disparity to depth:** $z = f\frac{B}{d}$ **(1pt); Depth to 3D point:** $X = zK^{-1}\begin{bmatrix} u \\ v \\ 1 \end{bmatrix}$ **(1pt)**

Modify this to show an additional image, which would appear as figure 1 b:
A QR Code is a type of two-dimensional barcode. Your task is to devise an algorithm to read it from real-world photos.
Consider the following challenges that may be encountered in real-world photos:


Figure 1: Left: A photo of QR Code. Right: The recovered QR grid.

(a) The QR Code can be rotated, skewed, or captured from different angles.

(b) Portions of the QR Code may be corrupted. Note that QR Codes support error correction, allowing recovery of information even if some parts are lost or unreadable.

(c) The grayscale can vary not only between different images but also within the same image.

You will have to describe how would you implement 3 steps of the algorithm. Outline them in detail. Be precise: when referencing a method from the lecture, specify the inputs, outputs, the exact variant of the method, and any adjustments you propose.

**Question 29:** *This question is worth 6 points.*

☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ■6

Step 1. Detect the square markers in the corners of the QR Code.

**There are several possible answers to this question.**
**One possibility is using the Generalized Hough transform with with an image of the corner marker as the template. In this case, you should explain how you can use the R-table to deal with different rotations/scales, or simply iterate the transform with different rotated/scaled versions of the template.**
**Another one is using the line Hough transform, but how you go from lines to detecting corners should be clearly explained.**
**Similarly, deriving a a parametric Hough transform for a square could also work, still taking scale and angle into account.**
**Normalized cross-correlation with the corner template while trying different scales/rotations is also acceptable.**
**Other answers are also possible. However, using methods that require manual input or deep learning (since it requires annotations) is not an acceptable answer.**

**Question 30:** *This question is worth 6 points.*

Step 2. Divide the image into sets of pixels, each set representing a small white or black square in the QR code as seen on the right of Figure 1.

**For this part, we need to be able to extract the image coordinates of each QR-code square.**
**One answer is first deriving a transformation to transform the QR code into a perfect square. This is possible, since we know the positions of the corners from the previous step. Then, a regular grid splitting of the QR code is sufficient.**
**A grid can also be made without 'fixing' the QR code, with careful use of line geometry and linear interpolation. This solution would need to be explained clearly.**
**Using a line Hough transform to detect lines and then lengthening and connecting them while checking for consistency is also possible.**
**Other, less straightforward answers are also possible. Applying segmentation algorithms here does not answer this question, however.**



**Question 31:** *This question is worth 6 points.*

Step 3. Populate the final binary matrix with the content of the QR Code.

**This last part deals with determining whether the squares are black or white.**
**Ideally, we want to use a local adaptive thresholding approach, so that we can deal with both inter-image *and* intra-image lighting variations. One answer could be splitting the QR code into patches large enough to contain several squares, and then using the patch median as the black-white threshold for that patch.**
**Using a global threshold such as the per-image mean/median is not perfect and is penalized, as it cannot deal with local lighting variations. Using a completely fixed threshold (such as 0.5/127) is worse, as it won't be able to deal with images from different sources that are too bright or too dark.**
**Less direct approaches using things such as K-means or graph-cut are also acceptable, even though inefficient.**

Pour votre examen, imprimez de préférence les documents compilés à l'aide de auto-multiple-choice.

+1/11/50+



Pour votre examen, imprimez de préférence les documents compilés à l'aide de auto-multiple-choice.