Examination questions, subject PGR001
3D Computer vision

Václav Hlaváč
Czech Technical University in Prague
Czech Institute of Informatics, Robotics and Cybernetics
160 00 Prague 6, Jugoslávských partyzánů 3, Czech Republic
http://people.ciirc.cvut.cz/hlavac, vaclav.hlavac@cvut.cz

December 23, 2022

The written test consists of six questions. Answers to each of the questions is assessed by five points maximally. Questions will be randomly selected from the following list. Some of questions may be modified or replaced by other questions. The list of questions may be extended or modified gradually.

---

1. What is the difference between image analysis (also computer vision) on one side and computer graphics on the other side? Give two examples, which illustrate the difference.
2. Interpretation (understanding) of the image can be expressed mathematically using a formal languages theory as *mapping*: *observed image data* $\rightarrow$ *model of a theory*. The model of the theory matches to a specific world, in which the "theory" is valid. A single "theory" may match to more than one world. Interpretation can be understood also as the *mapping*: *syntax* $\rightarrow$ *semantics*. Semantics represents knowledge about a specific world. We understand usually in image analysis that images contain certain objects.

   Give two practical examples of image analysis tasks, in which interpretation is used. With each of the examples, be specific how is the semantics utilized.
3. Signal analysis and low level image processing does not usually interpret analyzed image data. Explain (preferably using mathematical formalism), what the interpretation is. What gain does the interpretation provide on one side and what constraints does it bring on the other side?
4. Why is it difficult to "understand" general 3D scenes in computer vision? Give several reasons and outline them by a short comment. (I mentioned six reasons in the lecture.)
5. Local and global processing.
   - Discuss briefly the difference between the local and global approach to image analysis. Mention advantages and disadvantages of both approaches.
   - Give two examples od local operations and comment it briefly.
   - Give two examples od global operations and comment it briefly
6. Explain the notion of the image (continuous image function) $f(x, y)$ or $f(x, y, t)$. What do parameters x, y, t correspond to? Give several examples of 'real life' images captured using different physical principles.
7. What is the image quantization? How and in which device is the quantization performed? Explain the principle. How many quantization levels does a young healthy person distinguishes in a grey level image? What does the person sees in the image, which has less quantization levels than needed?

8. Consider digitization of a 2D image. The distance between equidistant samples is set according to Shannon sampling theorem similarly as in the 1D signal case. In a 2D image case, it is necessary to solve one more issue (relation) besides finding the distance of samples. Which issue is it? How is it solved typically. Which advantages and disadvantages these solutions have? Let me note that I do not talk about intensity quantization here.

9. The contiguity relation between two pixels of a binary image (there is a path between them) defines the image (set) into equivalence classes. Which three properties must such relation fulfil in order to be equivalence? Verify these three properties for the continuity relation in 2D binary image.

10. (a) Define (i) the region and (ii) the convex region in a 2D image. Draw the example of the convex and non-convex region.
    (b) Define the convex hull.
    (c) Draw the convex hull for the example of the non-convex set, which you used in the item (a) of this question.

11. Write the definition formula of the Shannon (also information) entropy. Explain variables in the formula. What is the Shannon entropy used for in general? Consider a gray scale image. Give at least two applications of the Shannon entropy in digital image processing.

12. When capturing an image of a 3D world using a camera, the involved geometry is represented by a pin-hole camera model (also called perspective projection). 3D point (x, y, z) projects into the image plane as (x', y'). Draw a corresponding figure (a planar one suffices, do not draw, e.g., y-coordinate). Assume that you know 3D coordinates (x, y, z), the focal length f, i.e., the distance of the image plane from the center of projection. Derive $x' = \ldots$.

13. What is the main role of the lens in a (photo)-camera? Describe the role of the lens from the physical point of view informally.

14. The lens of a camera is often explained by a simplified geometric optics model (theory) in practice. What are the preconditions enabling to use this model?. I note that the more complex model is the wave optics model.

15. Compare properties of a pin-hole camera and an (lens) objective composed of individual lenses.

16. Explain what is the natural vignetting. Is the natural vignetting more pronounced for wide-angle lenses or for tele-lenses. Explain (derive in the better case), what is the cause of natural vignetting.

17. Explain what is the radial distortion of a lens. How the radial distortion manifests itself. Can it be corrected? If so, how.

18. Explain what is depth of focus (or depth of field) for the optical lens. Which parameter (usually controllable) parameter of the lens allows to change it?

19. Imagine capturing of a 3D scene. The elementary surface patch in the scene reflects certain radiation L into a camera. The irradiation E on the light sensitive camera chip in the pixel x, y is directly proportional to the image function $f(x, y)$ value ($\approx$ brightness). On which elementary patch and light sources properties does the value of $f(x, y)$ (brightness, color) depends for the chosen x, y? Formulate the preconditions of this radiometric model and sketch the corresponding figure.

20. Consider the simplest radiometric model of image formation, i.e., a point light source at infinity, an elementary surface with Lambert (fully diffuse) reflectivity. With knowledge of the directions to the light source, to the observer and the local orientation of the considered surface area, given, e.g., by the normal vector, computer graphics can calculate the brightness (color) of the elementary area. The inverse problem, which would be useful in computer vision, remains unsolvable.
    (a) Why?
    (b) What are the consequences?
    (c) How computer vision deals with the unsolvability of an inverse problem?

21. Explain the concept “bidirectional distribution function"of an image abbreviated as BRDF. Why and for what is BRDF good for?

22. - What reflectance properties has the Lambertian surface?
    - What is the simplification by a Lambertian reflectance model used for? Give two examples at least, please.

23. What solves the irradiance equation in radiometry? Formulate the task and the basic thought of its derivation. The formulas are not necessary. (It might help you if you draw a sketch and denote variables in it.)

24. Characterize the image preprocessing. What constitute its input and output? What is the image preprocessing good for? List three examples of image preprocessing methods. Names of methods suffice.

25. Characterize 2D convolution. For what is 2D convolution used in digital image processing?

26. Homogeneous coordinates are often used when expressing geometric transformation of an image. Explain what are the homogeneous coordinates. What advantage homogeneous coordinates bring when expressing affine geometric transformation. (Hint: Recall the language describing a page named PostScript).

27. Consider an affine planar geometric transform (comprising the scale change, rotation, translation and sheer) given by the equation

$$\begin{aligned} x' &= a_0 + a_1 x + a_2 y \,, \\ y' &= b_0 + b_1 x + b_2 y \,. \end{aligned} \tag{1}$$

    (a) How many control points do you need to know at least if you like to calculate coefficients of the affine transformation in (1)?
    (b) More control points are used often practically, which corresponds to the overdetermined system of equations (1). Why the redundant number of control points is used?
    (c) What is the name of the method used commonly to solve the overdetermined system of equations?

28. When using geometric transforms for discrete images there is need to approximate the value of the image function $f(x, y)$. Why? List two approximation methods at least (suggestion: draw a figure, provide a formula ...).

29. Explain concepts in the image domain and their relations: edge, edge element (abbrev. edgel) and boundary of a region.

30. What excellent features of edge element is used in 3D computer vision in 3D reconstruction (SfM, shape from motion)?

31. In 3D computer vision, points of interest or areas of interest are widely used in the SfM (shape from motion) task. Why? How? For what?

32. Explain the principle of detecting points of interest (e.g., Harris corners).

33. Formulate the 2D image segmentation task. What constitutes the input and what the output? Name two different segmentation methods and explain briefly the algorithm they use.

34. In the 2D image segmentation context, explain concepts: region and object. Select a specific segmentation method and write what is the relation between the region and the object in this particular method.

35. Segmentation is based on semantics of a specific task, i.e. on the ability to explore the prior knowledge leading to the image interpretation. Explain on the example how semantics simplifies image analysis.

36. 3D computer vision. Formulate a task (or several tasks). What is the input, what is the output of 3D computer vision tasks?

37. David Marr (1945-1980, author of a new model of image visual perception, influential in the field of computational neuroscience) formulated the task and the theory of 3D computer vision, which allowed to investigated them by computational methods. Explain basic ideas of the theory. Specify the task, which the theory attempts to solve. Rely on the concepts: input iconic image, primal sketch, two and a half dimensional sketch, 3D representation related to the object.

38. Marr's model of understanding the visual image of a 3D scene has remained the only widely accepted theory since 1980s, although its applicability is very limited. Marr's stimulus led to the development of methods collectively called Shape from X. List at least five such methods.

39. One of the widely used results of Marr's model of understanding the visual image of a 3D scene is the shape from shading, also the shape from motion (SfM). Formulate what task SfM solves.

40. Explain what projective space is.

41. What are homogeneous coordinates? What is the main advantage of using homogeneous coordinates when using them for projection?

42. What is homography? What are the two most common uses of homography in computer vision? Formulate these two tasks and write down what they are used for.

43. 2D homography maps a plane to a plane.

$$\alpha \begin{bmatrix} u' \\ v' \\ 1 \end{bmatrix} = H \begin{bmatrix} u \\ v \\ 1 \end{bmatrix}, \quad \alpha \neq 0\,,$$

where H – $[3 \times 3]$ is the homography matrix. Explain how you can calculate homography matrix entries $h_{11}, \ldots, h_{33}$ knowing corresponding points in two images related by homography. Describe the idea. Do not derive. Formulas are not needed.

44. What is the cross-ratio in projective transformation? Express it with a picture. What is the cross ratio in computer vision used for?

45. The projective transformation can be expressed in matrix form as M having three-row and four columns. What is the input and output of the projective transformation expressed in this way. Write a relationship.

46. The projective matrix M can be factorized (i.e., expressed as the product of partial matrices).

$$\mathbf{u} \simeq M\,\mathbf{X} = K \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & -R\,\mathbf{t} \\ \mathbf{0}^\top & 1 \end{bmatrix} \mathbf{X}\,.$$

Name the sub-matrices and write what they are good for. What is the size of matrix K? What special form does the matrix K has?

47. What is the matrix of internal calibration parameters K? Write it and explain its parameters.

48. The external calibration parameters are described by the rotation matrix R and the vector $\mathbf{t}$. Explain what these parameters are for. How are external calibration parameters determined?

49. Formulate the task of calibrating one camera in a known 3D scene. Explain how the projective matrix M is calculated. How many least points in the correspondence do you need to calculate the M matrix? Why?

50. Consider the task of calibrating a single camera in a known 3D scene. You need at least six points in the correspondence to calculate the projective matrix M. Usually many more points are used. Why? Briefly explain the procedure used to estimate parameters of the matrix M.

51. Suppose we have already calculated the projective matrix M when calibrating one camera. Now we need to decompose the matrix M to obtain the matrix of internal calibration parameters K and the external calibration parameters, i.e. the rotation matrix R and the translation vector $\mathbf{t}$. Indicate how you do this decomposition?

52. Consider two perspective cameras. What is epipolar constraint? Shat is its significance? Express with a picture.
53. How is epipolar constraint expressed mathematically? (hint: bilinear transformation) Write.
54. Formulate the role of correspondence in the analysis of (intensity) images. What is the role of correspondence looking for? What is its input? What is its output?
55. What are correspondence search algorithms used for (intensity) image analysis and 3D computer vision?
56. The role of correspondence of (intensity) images in 3D vision tasks is in principle ill-conditioned. Write down three possible causes for this ill-conditionality.
57. What algorithms are used to search for correspondences between two intensity images? Sketch the principle of two qualitatively different algorithms. The first algorithm assumes nothing more than that some of the images overlap. The second algorithm will use certain (discussed in the subject) knowledge to simplify the correspondence search algorithm.
58. The task of searching for correspondence can be facilitated by additional technical means. List the method names you can think of. Describe one of the methods in more detail, but still briefly.
59. What sensors used in robotics, such as in self-driving cars, while measuring depth in the image and working on the principle of energy flight time?
60. Explain briefly the principle of operation of radar.
61. Explain briefly the principle of operation of lidar.
62. Briefly explain the principle of obtaining a depth map using triangulation-based sensors. Describe the principle (a) of the laser plane sensor; (b) two stereo cameras and (c) structured light sensors.
63. What sensors used in robotics, such as self-driving cars, measuring depth in the image, work on the principle of triangulation?
64. The result of obtaining a depth map, for example with a stereo camera, lidar or radar, is a point cloud in the observer's coordinate system. The depth map is a point cloud with (x, y, z) coordinates. What data do we lack for such a depth map so that we can continue to work with it and describe a 3D surface?
65. The next step in moving from 3D point clouds to a 3D surface is to join the cloud points into a point network, such as a triangular network. Explain why it's done? What is the principle of procedure?
66. The Iterative Closest Points (ICP) algorithm is used to register overlapping point clouds (or depth maps). Explain the assumptions of its use, the basic idea, and the properties. What prerequisites must the data meet in order to use the ICP algorithm?