Image & Video Understanding, written examination
Catalogue

Name: 
Mat.Nr.: 

Please answer all 6 questions. Each completely answered question will earn you 1 point. Please note that most of the questions consist of more than just one part.

Only complete answers addressing all parts can yield full points.

Grading scheme: 0-2.99: 5; 3-3.69: 4; 3.7-4.39: 3; 4.4-5.09: 2; 5.1-6: 1

Duration of the exam: 1½ hours

1. Please explain the functionality of the following receptive fields in the human visual system: On-Off, Off-On, simple, and complex cell. Which of these functions can be modeled by Gabor-filters (why + how?)?

2. Please explain the functionality of the following receptive fields in the human visual system: On-Off, Off-On, simple, and complex cell. Which of these functions can be modeled by 'scale-space derivatives' (why + how?)?

3. Please explain the functionality of the following receptive fields in the human visual system: On-Off, Off-On, simple, and complex cell. Which of these functions correspond with Gaussian and Laplacian pyramids (why + how?)?

4. What types of light sensitive cells in the retina do you know? What is the distribution of these cells over the field of view? What is the relationship between fovea and 6-neighborhood? What is the "blind spot"?

5. What types of light sensitive cells in the human retina do you know? In what part of the field of view do humans perceive color? What is the “blind spot”? Compare 4-, 6-, and 8-neighborhood – what are the advantages of 6-neighborhood? Explain the term “spatial resolution” of a digital raster image and its relation to the sampling theorem.

6. The role of cognitive psychology in vision research: Discuss mechanisms for perceptual grouping, optical illusions, bottom-up versus top-down processes, and the importance of 3D.

7. Please answer this question from the topical point of view of “stereo as a visual module”: In the processing of a visual stimulus (from the retina → visual cortex), how
is the perceived visual field propagated to the 2 hemispheres? Why? What is a "random dot stereogram" and what does it demonstrate? What is a “2-1/2D representation? How is depth reconstructed in the “canonical stereo configuration” (please draw a sketch and derive the relation between disparity and depth)?

8. Describe the neurophysiological propagation of visual stimuli from light sensitive receptors in the human retina to visual cortex. How is the perceived visual field propagated to the 2 hemispheres? What are the reasons for this propagation mechanism?

9. Explain the term “spatial frequency”. How can you implement a highpass, a lowpass, and a bandpass filter by using the Fourier spectrum?

10. What is an ‘image pyramid’? Explain the terms ‘reduction factor’, ‘reduction window’, and ‘reduction function’? What is a ‘separation plane’?

11. Explain the term ‘scale’ and relate it to the spatial resolution of a digital raster image. List and explain all the various possibilities to represent scale.

12. Scale-space: explain the representation. What are scale-space-blobs, scale-space-ridges, and scale-space-edges? Explain the concept of automatic selection of a ‘characteristic scale’.

13. Scale at the image level versus scale at the object level: Please give a (brief!) overview of the most important methods to represent spatial scale globally, i.e. for the whole image. Describe possible ways to represent scale locally, i.e., scale per image object?

14. a) Explain the Marr-paradigm (i.e. computational framework).
   b) Describe the various levels of representation according to Marr (primal sketch, 2-1/2D-sketch, 3D object representation). What do the terms ‘viewer-’ and ‘object-centered’ mean in this context?

15. Please sketch the system diagram for image understanding and explain the following components in detail: segmentation, image description, and grouping. What do the terms ‘token’ and ‘tokenset’ mean in this context?

16. From an image → image description: Explain the terms “token” and “tokenset”. What token types do you know, and how can these types be represented? Explain the representation of regions in “constellation tokens”.

17. Explain the representation of regions in “constellation tokens”. What are meaningful descriptors for the shape of a constellation token? Please give an example, too.

18. Explain the formal definition of segmentation. Which segmentation algorithms do you know (give just a brief overview)? Explain 'Graph Cut' in detail.
19. Explain the formal definition of segmentation. Which segmentation algorithms do you know (give just a brief overview)? Explain 'Level-Set' based segmentation in detail.

20. Explain the terms “token” and “tokenset”. What 2D models do you know? How can these models be used to find 2D structures in images/tokensets?

21. Please sketch the system diagram for image understanding and give a very brief explanation of all its components. Discuss the differences between 2D and 3D scene description?

22. Describe edge-based segmentation using the LoG edge detector. How can LoG be approximated by DoG? Describe also the important properties of this type of edge detector. What is the difference between DoG/LoG edge detector and DoG/LoG corner detector?

23. Explain the following four types of image objects: edge, line, corner, and blob. Draw grayvalue profiles for each of them. Discuss these profiles: How can the four types be detected, either as extrema or as zero crossing (of which derivative)? Explain the difference between the DoG/LoG edge detector and the DoG/LoG corner detector.

24. What is an 'edgel', what is a 'ridgel'? How can they be detected? Explain a simple 2D grouping algorithm to group edgels to contours and ridgels to ridges.

25. Describe (briefly) what you know about 2D models, explain the main representation methods in the context of bottom-up vs. top-down grouping.

26. 'Recognition' and 'reconstruction school', 2D and 3D scene description, bottom-up vs. top-down strategies in image understanding: Explain the differences, and the advantages and disadvantages of the various approaches.

27. Generic vs. specific object recognition: Describe the main differences. Describe important guidelines in building a good database (training and test images) for generic object recognition. What does the term 'weak supervision' mean in this context?

28. Explain the 'bag of keypoints' representation for generic object recognition and discuss advantages and disadvantages of this method. Is this approach better suited for image categorization or for object localization? Why?

29. Explain the 'constellation' model for generic object representation and discuss advantages and disadvantages of this method. Is this approach better suited for image categorization or for object localization? Why?

30. Explain the 'boundary-fragment' model for generic object representation and discuss advantages and disadvantages of this method. Is this approach better suited for image categorization or for object localization? Why?
31. Compare generative and discriminative models for generic object recognition. What is actually being modeled by which approach? Which approach is better suited for image categorization, which one for object localization? Why?

32. Describe (briefly!) the point of view of “object categorization as a problem of classification”. How would you approach it?

33. Describe (briefly!) the point of view of “object categorization as a problem of learning”. What learning algorithms do you consider well-suited? Why?

34. Describe (briefly!) the point of view of “object categorization as an issue of datasets”. Discuss important design principles for such datasets. Address potential issues related to training and testing. What does the term “weak supervision” denote?

35. Object categorization as a problem of representation: Give a brief overview of object representations. Discuss advantages and disadvantages of these representations.

36. How can you evaluate the quality of recognition/detection of an object categorization algorithm? Explain the terms “positive detection rate” and “false detection rate”. What does the “receiver-operator-characteristic (ROC)” tell you? Why do you need “recall-precision-curves (RPC)?

37. ‘Saliency, corners, keypoints’: Give an overview of methods to detect ‘salient’ points in images. What property does ‘affine covariance’ add to corner detectors and why is this important?

38. What is the fundamental principle behind corner detection based on autocorrelation? What is the ‘second moment matrix’? Please explain the ‘Harris corner detector’?

39. What is the basic idea of corner detection using the Hessian determinant? Please explain the ‘DoG keypoint detector’?

40. Give an overview of approaches to describe local patches for object recognition (‘local descriptors’). What are the main requirements for such descriptors in generic object recognition (categorization)?

41. What is a 2-1/2D representation? How can it be produced (which visual 'shape from X' modules do you know, including a brief description of each type of module)?

42. Object recognition: What are ‘geons’ and how can they be used to recognize objects in images? Describe advantages and disadvantages of geons compared to other representations for object recognition.
43. 2D models to represent shape: Explain the Hough transform for lines, and the “boundary fragment model” for the generic description of shape of a category.

44. Spatial scale: Describe briefly the various representations of scale in image pyramids, Gabor filter banks, and in “scale space”. Relate these representations to the various kinds of receptive fields in the human visual system.

45. What is a “bag of keypoints” representation? Give an overview on keypoint detectors, and keypoint descriptors.

46. Representations for object categorization: Describe (briefly!) “geons”, “bag of keypoints”, and the “boundary fragment model”. Compare the three kinds of representations and discuss advantages and disadvantages.

**Linear Filtering:**

Q

A) Briefly describe how the Fourier Transform can be used as an image processing tool. Which basis functions are used to decompose an input image? What is a single point in the Fourier domain image representing?

B) For the following images, assign the labels: (a) input image in the spatial domain, (b) Fourier transform, (c) Filter, (d) spatial domain image after filtering.

D) Describe how filtering can be implemented in the spatial and in the Fourier domain. Which techniques can be applied to handle the image borders during filtering?

E) In what situation will the output of convolution and correlation for an arbitrary image yield the same result?

Q

A) Describe the concept of linear filtering by application of a filtermask to the image. For which tasks can image filtering be useful?
B) Fill in the empty table (below-right) with the resulting image obtained after convolution of the original image (bottom-left image) with the Sobel filter:

\[
\begin{bmatrix}
1 & 0 & -1 \\
2 & 0 & -2 \\
1 & 0 & -1 \\
\end{bmatrix}
\]

The given Sobel filter represents an approximation of the first derivative in the horizontal dimension. Assume that the image is zero padded at the borders. The origin is located at the top-left corner with coordinates \([0, 0]\).

C) Compute the gradient magnitude at pixels \([0, 0]\), \([3, 3]\) and \([4, 5]\) in the left image (the image pixels marked in bold) by using the sobel filter from above as well as the 90° rotated version of the filter for the vertical direction.

D) Which techniques can be applied to handle the image borders during filtering?

E) In what situation will the output of convolution and correlation for an arbitrary image yield the same result?

Q

A) For which tasks can filtering an image be useful? Describe what is meant by (i) a linear filter, and (ii) a separable filter. Which of these properties are satisfied by (i) a Gaussian filter, and (ii) a median filter?

B) Which techniques can be applied to handle the image borders during filtering?

C) By considering the Fourier transforms of a box filter and Gaussian filter, explain why an image that has been box filtered is more likely to have high frequency content than one that has been Gaussian filtered.

Q
A) For which tasks can filtering an image be useful? Describe what is meant by (i) a linear filter, and (ii) a separable filter. Which of these properties is satisfied by (i) a Gaussian filter, and (ii) a median filter?

B) Why does consecutive smoothing (convolution) with a series of Gaussians, each with a particular standard deviation (scale) $\sigma_i$, yield the same result as a single convolution with a Gaussian of variance $\sum_i \sigma_i^2$?

Q

A) Which techniques can be applied to handle the image borders during filtering?

B) In what situation will the output of convolution and correlation for an arbitrary image yield the same result?

C) By considering the Fourier transforms of a box filter and Gaussian filter, explain why an image that has been box filtered is more likely to have high frequency artifacts than one that has been Gaussian filtered.

Q

A) Describe what is meant by (i) a linear filter, and (ii) a separable filter. Which of these properties is satisfied by (i) a Gaussian filter, and (ii) a median filter? For which tasks can image filtering be useful?

E) Why does consecutive smoothing (convolution) with a series of Gaussians, each with a particular standard deviation (scale) $\sigma_i$, yield the same result as a single convolution with a Gaussian of variance $\sum_i \sigma_i^2$?

B) Which techniques can be applied to handle the image borders during filtering?

C) In what situation will the output of convolution and correlation for an arbitrary image yield the same result?

D) By considering the Fourier transforms of a box filter and Gaussian filter, explain why an image that has been box filtered is more likely to have high frequency artifacts than one that has been Gaussian filtered.
Convolutional Neural Networks:

Q

A) Describe the layers of a Convolutional Neural Network architecture:

- Convolutional Layer
- Pooling Layer
- Fully-Connected Layer

Which operations do these components perform and what is the purpose of each component?

B) Which algorithm is used to train these networks and how does training work in principle?

Which techniques can be applied to reduce overfitting?

Q

A) What is the difference between a convolutional-based representation and a patch-based representation? What is the main advantage of the convolutional representation?

B) Describe the core components of a Convolutional Neural Network architecture. Which operations do these components perform and what is the purpose of each component?

C) Which algorithm is used to train these networks and how does training work in principle?

D) Which techniques can be applied to reduce overfitting?

E) What is the purpose of pre-training a Convolutional Neural Network? Give an outline of how you would apply the transfer learning scheme.

Q

A) Describe how the training of convolutional neural networks works in principle. Which algorithm is used to train these networks?

B) Which techniques can be applied to reduce overfitting during training?

C) Explain how a convolutional network can be applied for video recognition tasks.

Q

Consider a convolutional neural network architecture defined by a list of Layers that transform the image volume into an output volume (e.g. holding the class scores), with each layer performing a differentiable function.

A) Which distinct types of Layers do you know? Elaborate the operation that each layer performs.

B) Which layers hold parameters that are learned by the network?

C) Which techniques are important for the training of very deep networks?
Q

A) Describe how the training of convolutional neural networks works in principle. Which algorithm is used to train these networks?

B) Which techniques can be applied to reduce overfitting during training?

C) What is the purpose of pre-training a convolutional neural network? Give an outline of how you would apply the transfer learning scheme.

Q

For learning extremely deep features batch normalization and residual connections have been very useful.

A) Describe the batch normalization algorithm and the intuition behind it for example by using a non-linearity of your choice (e.g. sigmoid function).

B) Illustrate the concept of residual learning based on residual units for optimization of extremely deep networks.

Q

A) Briefly describe how transfer learning on an auxiliary task (e.g. pre-training for ImageNet classification) can be utilized to solve a target task of interest.

B) Looking at the recent progress in object detection, what are the advantages of “Faster-RCNN” over “Fast-RCNN” over “R-CNN”.

Q

A) Describe how the problem of object detection has been approached in recent convolutional network approaches.

B) Describe how the training procedure of a “Faster-RCNN” object detector works in practice.

Q

A) Describe two different approaches for visually debugging convolutional neural networks.

B) What characteristics can be observed when successively looking deeper into a convolutional representation that has been trained on an image classification dataset?