Possible Exam Questions

Author: Gianmarco Scarano

gianmarcoscarano@gmail.com

Table of contents

Amerini

Filtering:

- Given an image Im and a filter $f1$, show the intermediate passages and the resulting image g after applying the convolution operator between Im and $f1$. Use the following coordinates (3,3), (1,1), (2,4) with zero padding.

$$Im = \begin{bmatrix} 3 & 1 & 2 & 0 \\ 0 & 2 & 1 & 5 \\ 1 & 0 & 1 & 3 \\ 1 & 2 & 0 & 5 \end{bmatrix} \qquad f1 = \begin{bmatrix} -2 & 3 & 2 \\ 0 & -1 & -2 \\ 1 & -2 & 0 \end{bmatrix}$$

Image Pyramids:

- Show the differences between Laplacian/Gaussian Pyramid and illustrate some real-world applications.

Corner Detector:

- List the main steps of the Harris corner detector.
- Is the Harris corner detector robust with respect to uniform intensity changes in the image i.e. uniform intensity shifts and intensity scaling? Justify your answer.
- Is the Harris Corner detector robust with respect to rotation? Justify your answer.

Optical flow:

- Describe the optical flow algorithm. What is the aperture problem? What is the direction in the image along which optical flow cannot be reliably estimated?

Russo

Typical questions are three and each one is related on one part of the course.

The first question is about camera calibration, the second one is about epipolar geometry or structure from motion and third part is about Deep Learning.

Camera calibration:

- What's the difference between intrinsic and extrinsic parameters?
    - **Answer**: Underline the matrices and that intrinsic parameters are parameters built-in into the camera (Physical properties of the camera). Extrinsic parameters are referring to the pose of the camera, for example.
- Describe the 4 steps that project the 3D world point into a 2D image point.
- Describe the camera calibration algorithm, reporting what is the input data, which quantities can be estimated with it and how it works (starting and ending questions will be sufficient)
- What are the general characteristics of the P matrix?
- What quantities can you recover from P matrix?
- What quantities is the P matrix mapping?

Epipolar geometry:

- What are the epipoles e and e' in a 2-view geometry (definition), reporting also their characteristics, their main features, and the most important epipoles-related equations.
    - **Answer**: Definition of the epipoles, what are they and the equation $l' = e' \times x'$ (from which we can conclude that it is equal to Fx). As for the properties, check Slide 8 (Epipolar Geometry), page 17.
- What is the Fundamental Matrix? Which are the most important equations? The characteristics?
- Describe the issues behind the uniqueness of the F matrix.
- Where is Canonical form of P and P' is coming from?
    - **Answer**: It is coming from the fact that there are not unique P and P'.
- When can you use Essential Matrix instead of F matrix? What's behind the reconstruction of the Essential Matrix?
    - **Answer:** We are using it when using calibrated cameras. As for the second question, we can simply say that we can have 1 possible solution (out of 4) of calibrated reconstruction from E.

Fundamental Matrix Estimation:

- How can we estimate the Fundamental Matrix?
  - **Answer**: 8-point algorithm explaination.
- Is convenient to do anything to your data before applying the 8-point algorithm?
  - **Answer**: Yes, we can by normalize it by [-1, +1], RANSAC, etc.
- Why can't we just simply project back the two points x and x' in order to recover the 3D point?
  - **Answer**: PDF 10 (Fundamental Matrix estimation) - Slide 10
- Basic idea behind the structure from motion (and incremental structure from motion).

Deep learning:

- Why can I not apply a fully connected layer to an image in order to do Image Classification?
  - **Answer**: First of all is not enough and plus we must use a structured network in order to extract significant features (Convolution Layers). Also we are introducing a lot of trainable parameters.
- Why CNN is superior to hand-crafted feature extractor (SIFT, etc) + one last Fully Connected Layer for classification?
  - **Answer**: Those features from SIFT, etc. are fixed and they are also low-level features. While, in CNN, we are building an hierarchical set of features, customized in that specific task, able to recognize low-level features and high level features.
- What is an hierarchical set of features in a Deep Learning model? Do we need to take any special action to build it when employing a Deep Convolutional Neural Network?
  - **Answer**: The hierarchical set of features in a DL model, are the set of features automatically built by the model when training and it's called hierarchical since it's divided (while training) into low-level features (gradients, lines, textures), mid-level features and high-level features (pieces of object/faces etc) in order to minimize the loss. Also, the high-level ones are built on top of the lower-level ones (That's why it is an hierarchy).
- Why do we need activation function?
  - **Answer**: We need them otherwise our model would be linear.
- What is the learning rate?
  - **Answer**: Is a weight to the update step. The network weights, without this, would jump very quickly and the network would not converge.
- What is overfitting? Why it happens? What are the techniques to avoid it?
  - **Answer**: Definition of overfitting, the effects and as for techniques we can talk about Dropout, Early stopping, etc.
- Why do we work with stochastic gradient descent instead of the standard gradient descent?
  - **Answer**: We don't have enough memory computation in order to do such task as it would require a full pass on the whole dataset. Plus, the stochastic gradient descent would work better since it would take random batches of data, trying to reach global minima with less memory computation.

- Why a Convolutional Layer is better than a Fully Connected Layer for some Computer Vision tasks?
    - **Answer**: As explained in other answers, with a Convolutional Layer we can extract more informations and features w.r.t a Fully Connected Layer which returns us something simply linear (while in Convolution we have low-level features and high-level features).

Architectures:

- Why data augmentation is important? What are the main data augmentation algorithms?
- What do we mean with Transfer Learning technique?
- What is VGG and why it is good? What about ResNet and residual blocks?

Batch Normalization:

- What is batch normalization? What are the improvement given by it?
    - **Answer**: In general, normalized data is always preferred. However because of the combination of CNN + non-linearity, data is not normalized during training so we use this in order to have 0 mean. Also there is faster convergency, more accuracy and we can use a bigger learning rate.

Segmentation Detection:

- What is the task of segmentation/detection? What about object detection?
- What is the basic concept behind the architecture of U-Net?

GANs:

- What is the basic idea behind GAN?
- What can you do with a GAN and how?
    - **Answer**: Start from noise z, we can generate an image. Plus, we can do Domain adaptation and Image-to-image translation.