Computer Vision (INFOMCV) - Exam

2015-2016, Utrecht University

January 26, 2016
Duration: 13.30 - 16.30

Instructions:

1. *Do not look at or read the questions before you are asked to do so.*
2. *Write your name and student number on every separate answer sheet.*
3. *Write your answers as complete as possible. However, adding irrelevant information will decrease your score. You may use examples to make your statements clearer.*
4. *Ensure that your handwriting is readable.*
5. *You are allowed to leave the room anytime after 14.00, by first handing in both your answer and question sheets.*
6. *During the exam, you are not allowed to speak with other students, use your phone or additional materials. Cheating or other misconduct will result in a fail for the course and will be reported to the exam commission.*

Good luck!

1. **Image formation (12 points)**

   (a) (1) What do the intrinsic and extrinsic camera matrices model and (2) what are their elements? For (2), give a short description for each element. You get full points if you also write out the matrices.

   (b) Explain why a chessboard is often used for calibration. (1) Mention at least two characteristics, and (2) explain how each is used in the calibration process.

2. **Silhouette-based volume reconstruction (12 points):** A lookup table can be created for silhouette-based volume reconstruction.

   (a) Describe the algorithm for creating the lookup table in pseudo-code. Mention inputs, outputs and processing steps explicitly.

   (b) Explain why we need the lookup table.

   (c) With the silhouette-based volume reconstruction algorithm, the volume of non-convex objects (objects with "holes") cannot be estimated well. (1) Mention and describe a technique that can be used to estimate the depth of "holes". (2) Explain how cameras should ideally be placed for this technique to work well.

3. **Clustering: (13 points):** The K-means algorithm can be used to find clusters in data.

   (a) Describe two main drawbacks of the K-means algorithm.

   (b) Describe (1) the two iterative steps of the algorithm and (2) explain the significance of each.

   (c) What can we do to cluster the data in Figure 1 using K-means and obtain two clusters each representing a class? (So one cluster that contains all crosses, and one that contains all circles.)

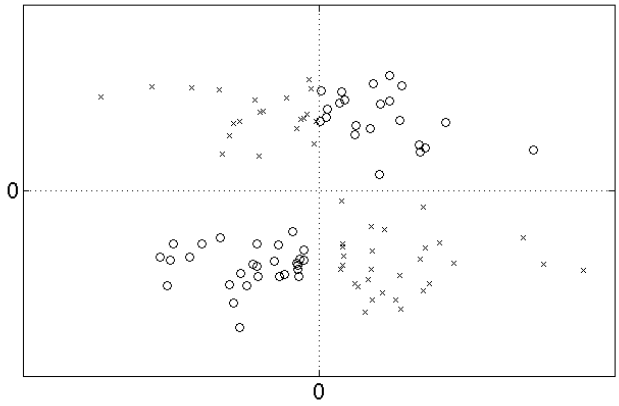

Figure 1: Data distribution for two classes: crosses and circles.

4. **Particle filtering (14 points)**

   (a) In a particle filtering framework, (1) mention three ways to determine the final state estimate given the weights of the particles. (2) For each, mention one advantage and one disadvantage.

   (b) Particle filters can be used to track people in 2D from silhouettes that are obtained using background subtraction. Explain the state of the particle filter, i.e., what does each dimension in the state correspond to.

   (c) How could the weight of a particle be calculated from its state and a silhouette image? Include at least a description how you could deal with scale (people closer or further away from the camera) and which distance function you use.

5. **Scale-invariant feature transform (12 points):** Given an image, the first phase of SIFT is to generate keypoint candidates.

   (a) In principle, this phase results in a large pool of candidates. Based on which two criteria is this pool reduced? Briefly describe the criteria.

   (b) Each keypoint has a location, an orientation and a scale. Explain, for each of these three, how they are determined.

6. **Optical flow (10 points):** When we have two (subsequent) frames, we can calculate the optical flow field between them.

   (a) Explain what a flow field is.

   (b) What is the difference between sparse and dense optical flow?

   (c) Which technique can we use to overcome the limitation of the Lucas-Kanade algorithm that we can only estimate motion in a small window? Explain how we can use this technique. You can give pseudo-code, or explain the processing steps.

7. **Object recognition (14 points):** One of the methods in object recognition is bags of visual words. In the method, there are two steps in the training process: (1) generating visual words, (2) calculating the classifiers.

   (a) For step (1), explain how the process works. Explain the inputs, processing and outputs.

   (b) There are several classifiers, of which we consider k-nearest-neighbor (kNN) and linear Support Vector Machines (lSVM). What is the most important advantage of lSVM over kNN?

   (c) Condensed kNN is an algorithm that aims to reduce the number of training samples. Explain how the algorithm works. Make sure you explain each symbol in your answer.

8. **Training and testing (13 points)**

   (a) There can be noise in a training set. If we assume that the labels of our training data are correct, this noise is in the feature vectors (image descriptors). Explain a method or algorithm to detect noisy training samples. Present the processing steps, use pseudo-code if needed.

   (b) When using cross-validation, what is the role of the validation (not test) set? Explain how it is used in the training/testing pipeline.

   (c) Explain why there is typically a trade-off between the precision and recall of an object detector.

That's it.