Computer Vision (INFOMCV) - Exam

2014-2015, Utrecht University

January 27, 2015
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 (10 points):**

   (a) In the camera intrinsics matrix, which elements (give their name and position in the matrix) are affected, and how, when (1) you zoom in; and (2) you change the resolution of your image. Motivate your answer.

   (b) Explain how color information is obtained in an imaging sensor (CCD).

2. **Background subtraction (10 points):** Background subtraction takes the (absolute) difference between a background frame (background model) and an image frame, and determines foreground pixels by applying a threshold on the result. Typically, differences in pixel color are larger for brighter pixels (high pixel values) compared to darker pixels (low pixel values). A single threshold for the whole image is therefore not always optimal. As an alternative, we can also model each pixel as a Gaussian and then apply the same threshold for each pixel.

   (a) Explain (1) what needs to change in the construction of the background model; and (2) how it is determined whether a pixel in the image frame is foreground. Motivate how these steps lead to a better solution for dealing with high/low pixel values.

   (b) Name and describe one post-processing technique that can be applied to remove potential (speckle) noise from a binary foreground image.

3. **Silhouette-based volume reconstruction (11 points):** We can construct voxel models from foreground silhouettes extracted in multiple views.

   (a) Explain at least three important limitations of this technique.

   (b) Instead of creating a new voxel model in each frame using the silhouette-based volume reconstruction algorithm, the algorithm can be made more efficient by only considering changes from one frame to the next. Assume that a voxel model in frame 1 is constructed, and that the foreground images for frames 1 and 2 for all views are available. Explain (in pseudocode or in text) the algorithm that constructs the voxel model in frame 2, based on the voxel model in frame 1.

4. **Particle filters (14 points):** Assume we have a video of a street, filmed from a window on the first floor. We will be tracking people using particle filters. For this example, we will use one particle filter per person, i.e. the state of the particle filter corresponds to the state of one person.

   (a) What is a suitable state of the particle filter? Motivate your answer.

   (b) To arrive at a good state estimate, we need to weigh the particles. Assume that we have a color model of the person that we are tracking. The model is described as a color histogram that has 16 equally-spaced bins of (only) the hue channel in HSV color space. Explain how we can obtain the weight of a particle using the color model.

   (c) Mention three ways to determine the final state estimate given the weights of all particles. For each, mention one advantage and one disadvantage.

5. **Image descriptors (16 points):**

   (a) What are the four steps involved in calculating SIFT descriptors. Explain each step in one sentence.

   (b) SIFT descriptors are invariant to scale. How is the scale of a SIFT descriptor determined?

   (c) Histograms of Oriented Gradients (HOG) consider the gradient of pixels in the bounding box. An alternative image descriptor is the Histogram of Flow (HOF). Instead of gradient direction and magnitude, it makes use of the flow field, calculated using an optical flow algorithm. What is the main assumption underlying optical flow algorithms?

   (d) Explain how the Horn-Schunck algorithm works, conceptually.

6. **Support vector machines (14 points):**

   (a) A linear support vector machine (lSVM) is a binary classifier. What does that mean?

   (b) During the training of an lSVM, "a hyperplane is estimated such that the margin between the two classes is maximized". Explain (1) what is a hyperplane, and give the equation; and (2) explain what the margin is. For (2), use the concepts "hyperplane" and "support vector".

   (c) Assume we have set of (image descriptor, label) pairs. Each image descriptor is a 3-dimensional vector containing the red, green and blue values of a color (expressed in RGB color space). Half of the labels has the "red" label, with image descriptors corresponding to different shades of red. The other half of the pairs is labeled "yellow", with image descriptors corresponding to shades of yellow. The colors partly overlap. We train an lSVM on this set. Explain, conceptually, what the resulting hyperplane corresponds to.

   (d) Imagine that we change the ratio of the samples to 99% "red" and 1% "yellow" and then train an lSVM. What happens, conceptually, to the hyperplane (especially when C is larger)?

7. **Image classification (10 points):**

   (a) When training a classifier, there is the risk that we tune it to the training set, instead of learning a classifier that generalizes over the classes. How is this called?

   (b) Often, we can solve it by using a different parameter setting. For SVMs, there is the C parameter that can be tuned. Explain how we can determine the optimal value of C in a sound way using cross-validation. Explain clearly how data is divided, processed and how the best parameter is selected.

8. **Object detection (15 points):** If we run a detector over all windows in an image, we typically get many detections that differ slightly in their position and scale. We can use non-maximum suppression as a subsequent step.

   (a) Explain how this algorithm works. What is the input, what are the processing steps and what is the output?

   (b) Give the equation to calculate the recall of a detector on an image. Use the concepts "true/false positive/negative".

   (c) Give (1) the name; and (2) the equation of a single-number measure that takes into account both precision (P) and recall (R).

That's it.