Computer Vision (INFOMCV) Exam 2018 Solutions

Note that this exam was NOT an open book exam.

1. C
2. See KC1. The intrinsics stay the same under rotation. $\boldsymbol{R}$ and $\boldsymbol{t}$ both change, see slide 14.
3. It is possible if the optical flow assumptions are met: pixels under motion do not change intensity and no pixels leave/enter the image. In this case, we can propagate the foreground from the first frame to the next. It is then also required that the foreground is sufficiently textured.
4. Instead of requiring that all cameras see the voxel as foreground, you could allow one to be off and predict background. In the specific case of having a hole in a silhouette, you could fill interior contours in large foreground regions.
5. From k-means, we obtain the 100 cluster centers. All pixels that are closest to bin *i* are put in bin *i*. For a new image, we check for each pixel the closest cluster. This is similar to the assignment step of EM.
6. B, D
7. (1) Orientation is determined from the gradient orientation at the keypoint location
   (2) Scale is determined from the octave at which the keypoint was detected
8. D
9. Negative examples are obtained by randomly cropping regions from the training image that do not overlap with the player bounding boxes. For hard negative mining, we evaluate the object detector on the training data. We then apply non-maximum suppression. Those detections that do not overlap with the player bounding box, and do not overlap (much) with the negative samples are selected as additional negative samples.
10. Accuracy is the number of true positives + number of true negatives as a fraction of all positives and negatives. With object detection, there are two issues. First, many window will overlap if no non-maximum suppression is used. Second, there will be many true negatives, which will severely bias the accuracy. For example, if there is one object and it hasn't been found, accuracy can still be high when there are a sufficient number of negative regions classified as negative.
11. 30x30x6
12. A
13. B
14. You could use a stack of RGB frames. The input would then be wxhx(25x3) for 25 frames with 3 color channels. Our you could use a two-stream CNN with input one RGB image (wxhx3) and a stack of optical flow images, e.g. wxhx(25x2) where the two optical flow channels correspond to horizontal and vertical displacement, respectively. The output of the network would be 3x1, with the neurons corresponding to "nothing", "left to right" and "right to left".