Computer Vision (INFOMCV) - Exam Solutions (Brief!)

2014-2015

The answers here are not complete, but just to give an indication of the correct direction.

1a. (1) When you zoom in, the focal length (fx and fy) increases. These are the elements at position (1,1) and (2,2) in the intrinsics matrix. (2) When the resolution changes, both focal length and principal point change. The latter are at positions (1,3) and (2,3) of the intrinsics matrix.

1b. The CCD contains three types of sensors that are each sensitive to a different range of wavelengths. So when light of a specific color (wavelength) comes in, the three sensors produce a different activation. Since the three sensor types are typically placed very close together, we can assume they have recorded the same wavelength. By "looking up" which wavelength produces a certain activation, we can rework the color.

2a. (1) We additionally need a standard deviation per pixel, per color channel. (2) Instead of taking the absolute difference between each color channels of the current pixel and the background model, we scale the difference by the standard deviation. The threshold that we apply is then the number of standard deviations that a value in each color channel can deviate.

2b. Erosion is one post-processing technique that can be used to remove white speckles from the binary image. It looks at the four (or eight) neighboring pixels and switches the switches the current pixel off it at least one of the neighboring pixels is off.

3a. (1) Holes cannot be recovered, (2) The quality of the voxel model depends on the quality of the background subtraction, (3) The quality of the voxel model depends on the number of cameras, (4) The level of detail of the reconstruction is typically limited. Explanation of each is left to the reader.

3b. We calculate the XOR between frame 1 and frame 2. Pixels with a 1 value have changed, and we only have to revisit those. So we loop over all cameras, then loop per camera over all pixels with value 1. We find the corresponding voxel and check the projected pixel values in all cameras. We set the voxel to "on" is all projections are in the foreground, and "off" otherwise.

5a. (1) Detect scale-space extrema, (2) Detect keypoints, (3) Determine orientation, (4) Determine local descriptor

5b. Scale is determine by the octave and scale within the octave in which the maximum was found.

5c. The BCA: brightness constancy assumption. It assumes that a pixel under motion does not change in intensity.

5d. Horn-Schunck aims to address the aperture problem by introducing a smoothness term into the equation. The penalty of the estimation optical flow of a pixel therefore depends on the change in intensity and the amount of movement. Both terms are weighted. Horn-Schunck further proposed a global estimation approach.

7a. Overfitting

7b. We split our training into *k* folds. We then train *k* classifiers that each omit one different fold from the training data. We evaluate our trained classifier on the fold that was left out. We average the evaluation values over the *k* folds. When evaluating multiple values of a hyperparameters, we repeat this procedure for each value we consider and eventually choose the model (with corresponding hyperparameter value) of the best evaluated model.

8a. The input is a list of detections, each with a confidence score. We sort the detections on confidence score, in decreasing order. We then loop over all detections. For each detection, we consider those detections that are later in the list. If their intersection over union (IoU) is above a certain threshold, the detection lower in the list is remove. We continue until we are at the end of the list. A threshold needs to be determined in advance.

8b. Recall = TP / (TP + FN)

8c. One option would be the F1 score: F1 = 2 * (Precision * Recall) / (Precision + Recall)