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

2015-2016

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

1a. (1) The extrinsics matrix models the transformation from 3D world coordinates to 3D camera coordinates. The intrinsics matrix models the projection from 3D camera view to 2D image plane. (2) Intrinsics matrix in KC01, slide “Intrinsic parameters 8”, extrinsics in KC01 slide “Extrinsic parameters”. Focal length (in x and y direction) is the distance to the image plane, s is the skew that determines the angle between x and y axes, (x0, y0) is the principal point to which the origin projects. R is the rotation of the camera in the world, t is the translation of the camera in the world.

1b. All crossings are in the same plane and at the same distance which means that their coordinates are easy to provide; The tiles are dark and light, which means the crossings can be estimated more reliably, improving the quality of the calibration.

2a. The algorithm is described in KC07, slide “Look-up table 2”

2b. The look-up table speeds up the process of linking voxels to pixels, which we have to do for each frame in a video.

2c. Stereo vision is one approach. For this, two cameras need to be placed side-by-side with similar orientation and a known distance between them.

3a. (1) Clusters have to be known in advance, (2) algorithm can get stuck in a local minimum, (3) no structure is taken into account.

3b. (1) assignment and update. (2) The assignment step labels all points to the closest cluster. It makes sure that each point is correctly labeled. The update step moves the cluster center to the mean of all assigned points, making sure that the space is better covered. Both steps are needed for convergence.

5a. (1) the point should have high contrast and (2) the point should not be on an edge

5b. Orientation: determined from the gradient of the point. Scale: determined from the octave and scale within the octave on which the maximum was found. Location: determined from the spatial location of the maximum.

6a. A flow field represents the flow (u(x), v(x)) for each location (x,y). So it stores the 2D vector when pixel (x,y) move to in the next frame.

6b. Sparse optical flow is only calculated at interest points (e.g., obtained from SIFT keypoints or Harris corner detection). Dense optical flow is calculated at each pixel.

6c. We can use image pyramids. The technique is used in a coarse-to-fine setting. Specific explanation is left to the reader but mention how an initial estimate propagates to a finer level of detail.

7a. Input is a set of images. We calculate local descriptors in each image. We then cluster those descriptors using a pre-defined number of clusters. The output are the cluster centers.

8a. One approach could be to see if training without the training instance improves over training with the instance. In this case, one could argue that the training instance is noisy. Algorithm is left for the reader.

8b. The validation is a set that has not been seen during training. It's used as a proxy for the test set, to understand what performance could be expected. When evaluation different hyperparameters, you repeatedly train on the training set and evaluate on the validation set. The performance then functions as a criterion to select the best hyperparameter value.

8c. Typically, different precision/recall values are obtained by changing a threshold that leads to more or less positive classifications. Usually, having more positive classifications entails that the quality of each positive classification is, on average, lower. This means that, for increasing numbers of positive classifications, the probability of a false positive increases. So the recall increases (more positive classifications overall) but the precision decreases (more false positive classifications).