

Formalized Feature Vector Generation Algorithm for Object Detection

Pre-Computation of Training

1. Load all positive images with mirroring.
2. Convert them to 8-bit grayscale.
3. DoG band pass filter to avoid aliasing.
4. Normalize image resolution to a fixed resolution.
5. Detect Keypoints (FAST, SURF)
6. Keypoint vectors are fed into a descriptor (FREAK, BRISK, SURF) to receive descriptors.
7. Takes all of those keypoints from every image ever seen (called bag of features) and clusters them into a fixed number of clusters using K-means with Euclidean distance for SURF, and Hamming distance for BRISK, FREAK.
8. The keypoints of the clusters (centroids) become the code words that we use to describe images.
9. Save these keypoints in a YAML file to save multiple hours on subsequent test runs.

Training

10. Given an image I, we perform the same preprocessing as before (Steps 1-4)
11. We then take the descriptor computed from FREAK/SURF
12. Match each keypoint generated from image I to the closest feature in the cluster model we created earlier. (FLANN or BruteForce Matcher in OpenCV)
13. Make a histogram of the number of each visual words found in the image and this histogram. (A vector (whose length is Size of Dictionary). That vector is our feature vector.

Prediction

14. The prediction is done the same way as the training, except instead of passing it to the svm.train method we pass it to the svm.predict method. The input should be the histogram feature vector. The return response is the class of that image.

Avoiding overfitting

15. Over fitting happens when there is too much data and the accuracy of the predictions go down.
16. Make sure the number of bins in the histogram is correct and the number of words in the vocabulary is optimal. Overfitting will reduce accuracy by about 20%, where incorrect bins reduce accuracy from 4-6%. Overall, fine tuning these parameters can bring a bad classifier of 65% accuracy to 91%

Results

Freak NU_RBF receives 91% accuracy on 1000+ images, about 85% accuracy at 250 images. 10 images gives around 60 percent. 3000 words in vocabulary is optimal.