# Testing and Validation

### SVM Training Error – Precision and Recal

As explained in [12], in classification tasks, the terms **true positives**, **true negatives**, **false positives**, and **false negatives** compare the results of the classifier under test with the actual ground truth. The terms *positive* and *negative* refer to the classifier's prediction (sometimes known as the *observation*), and the terms *true* and *false* refer to whether that prediction corresponds is conforming to the ground truth (sometimes known as the *expectation*). This is illustrated by the table below:

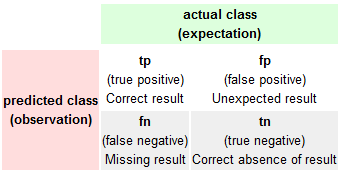


Figure 6..., Source: [6],

Precision and recall are then defined as:

,

The video which we worked with was 40 minutes long, on which we have performed different runs of the training set aquisition process whithin the first 15 minutes, varying the region of interest thus aquiring training images of different poses and size. Within these 15 minutes we managed to store appromatelly 1000 positive samples and 4 times as many negative exmamples. For our classifier training we have varied the number of positive and negative examples around 1000 or less for each different runs. The next table shows the svm training properties and results for the best svm which we obtained and working with:

|  |  |
| --- | --- |
| Positive Training Samples | 999 |
| Negative Training Samples | 999 |
| Support Vectors | 863 |
| Precision | 98.51% |
| Recal | 99.20% |
| Accuracy | 98.85% (1975 correct, 23 incorrect, 1998 total) |

Figure 6... , SVM training results

Next we tested the classifier on two test sets, which were composed by images that were not part of the SVM training process and which were extracted within a different data set extraction process so that the test set could exibit different poses and size then the training set. Thus we had to spare around 400 positive samples and have choosen to test 1000 of our available negative samples. The accuracy on the positive test set was 94.74% - 378 samples were correctly classified while only 21 were incorreclty classified from a total of 399 positive test samples, thus the precision/recal on this test set was 100% / 94.74%. Also, out of the 1000 negative test samples 973 were correccly classifier while 27 were missclassified resunting in 97.30% accuracy and the precision/recal was 100% / 97.30%. We were expecting that the result to be this good because our observed scene is simple and observed from a high viewing poitn and although we tested the classifier on samples extractacted from another region of interest, these samples were not expositing that much variantion in size and pose to other extractions.

### Tracker Accuracy with LBP

In such an application is hard to evaluate the tracker in an automatic way as we do not have the ground truth to compare results with for our video. Thus the only method that remains in evaluating the tracker is by human obsersevation by evaluating in the save time the ground truth and the output that the tracker is giving. In the next figure we are showing an approximate comparison of the ground with the tracker results at different time steps in the video. On the horizontal axis is the time flow, and the vertical axis we record the number of vehicles that have passed though the frames as the ground truth and also the tracker estimation for comparison. We hope to estiblish if the tracker underestimates or overestimates the true number of passing vehicles.

From the below curves, figure ... , we notice that the tracker does not overscore the tracks, and its was to be expected vehicles exibiting unusual positions and collision will not have enough detections in order for it to be properly scored. As it happens low detected vechicles will initiate a tracks but under lack of detections it will be negativelly scored and removed.

### Tracker Accuracy with NCC

The NCC Tracker suffers the same problems as the LBP tracker. The NCC method of deciding if a detection can be merged with a track is too restrictive. The NCC matching method is directly related to the robustness of the Detector. Normaly the Detector will not provide detections in which vehicles are perfectly centered in the detection bounding rectangle, but will vary around the true position of the vehicle. As it should have been expected, while NCC produces good results, it is not robust to pose variantion and variations in detections relative locality to the actual vehicle.

Figure 6... Ground truth – LPB Tracker comparison

Figure 6..., Ground Truth – NCC Tracker comparison

In the next table we summarize the performance of the two tracking methods. The GT row stands for the ground truth vehicle count recorded at different time instances. The LBP row contains the number of vehicles that the LBP Tracker has recorded until that time instance. Similarly the NCC row contains at each column position the number of vehicles that it has scored as positive. We can observe that local bit patterns perform a better job than the normalized cross correlation, due to its tolerance to local region similarity.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| GT | 15 | 21 | 45 | 63 | 100 | 178 | 250 | 300 | 348 |
| LBP | 13 | 19 | 42 | 58 | 94 | 167 | 238 | 284 | 326 |
| NCC | 13 | 20 | 41 | 55 | 90 | 157 | 221 | 265 | 307 |
| Error  LBP | 1.3% | 9% | 6% | 7% | 6% | 6.1% | 5.3% | 5.3% | 6.3% |
| Error  NCC | 1.3% | 4% | 8% | 12% | 10% | 11% | 11% | 11% | 11% |

Figure 6… Comparison of cumulated tracking errors for LPB and NCC.

### Examples of tracking errors

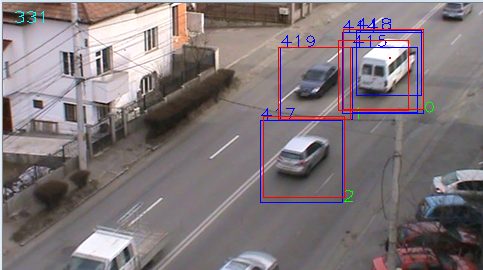


Figure 6… Multiple detections for the same vehicle

Figure 6... is an example of a vehicle having multiple detections. The cause for this error is the strictness in matching a detection with an existing track. Both the NCC and the LBP can experience this, but we metion that la later exposes more tollerance to these cases. As a vehicle moves through the scene,changing pose and scale, the detector migh provide detections centered differently around the vehicle and worst, the size of the detection bounding box is signifficantly different. The distance of the new detection bounding box around the vehicle ant the model stored in the track for the car can be enough to determine not the merge the detection with the track, and thus start a new track. It is also true that we generally noticed these variations in location and size for medium sized vehicles with which the classifier was not properly trained.

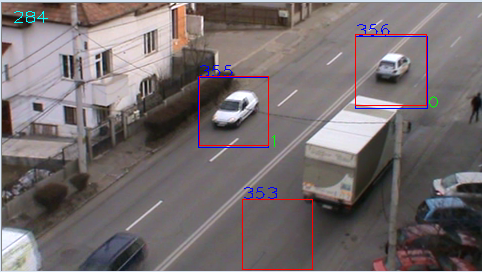


Figure 6..., Tracker has drifted from the tracked vehicle

The figure above is an example of a track that has drifted from the true path of the vehicle. The source of this problem is due to the lack of detections for a initialized track. With no detections arriving from the detector, it is the job of the Median Flow Tracker and the Kalman filter estimator to predict the position of the vehicle. It can happen that the Median Flow Filter has not enough points to track and fail and thus the prediction is left to the Kalman Filter which is not updated and presumes constant velocity in the movement which is not always the case. In these situations the drifted track will get negativelly scored each frame until it is removed.