Vehicle counting

Grundlagen, Anwendungen und Fortschritte der digitalen Videoüberwachung
183.284, 2009SS

Johannes Perl

Peter Bodesinsky

1. Previous Work

Various approaches can be found in literature for solving the problem of vehicle counting, detection and tracking. It can be more focused on background subtraction and motion, or rather on object or gradient detection. The selection of methods is also task dependent, for example if the speed and trajectory of a vehicle should be measured to prevent accidents [1] or if a virtual loop is used for counting the vehicles.

Separating the foreground from the background is a very important task for motion based approaches and there have been efforts, for improving motion based segmentation. This can be achieved by removing the noise caused by illumination changes or shadows. It is based on an illumination invariant color model and a statistical model for removing the shadows [2].

For our task, we chose an object detection based approach, which relies on gradients to describe the texture and appearance of a vehicle. There have been previous works dealing with similar problems. In [3] the local gradient maxima are used for estimating the position of vehicles/ROIs (hypothesis generation), these hypotheses are verified in a second step by using AdaBoost classification. It is stated that the method worked fast and with high reliability for vehicles that are not too far away. Another Method is proposed in [4] were the gradients are used as representation of a vehicle. The gradients are tracked and grouped afterwards based on similarity measures like velocity difference and spatial distance.

We decided to use HOG descriptors to take the gradients into account and to use a SVM for classification, this approach also worked well in a study for the detection of humans [5]. In this study, the HOG descriptors are considered as a good method to discriminate the human form from the background and to describe their shape by the distribution of the local gradients. The stated advantages of this method are that the keypoint based, sparse features, are not that reliable compared to the dense grid of the HOG descriptor. Furthermore it is invariant against local transformations or rotations, if they are smaller than the histogram bin size. In the first step the gradients are computed with a centered 1D mask, subsequently the orientations are accumulated in a histogram for each cell, where each pixel contributes a weighted vote for a certain bin. After this the cells are block-normalized and a linear SVM is applied for classification.

In our opinion, this approach can also be applied in a similar way for other objects and therefore can be used for vehicle detection.

2. Design and implementation

2.1. Overview

Our System is implemented in matlab and consists of parts for reading the data, calculating the HOG descriptors, training of the SVM classifier and classification. We define 3 ROIs, which are independently used for training and classification with the SVM. A vehicle inside a ROI is counted if it is present (classified as vehicle) for a time period of more than 10 frames. The System is trained with examples from the file Road_Easy_indeo52.avi and tested with all frames from Road_Hard_indeo52.avi, the whole procedure can be started by executing “vc.m”.

The main steps are performed in “vc.m”. At first the training data for the ROIs is specified (cFr,ncFr) and a SVM classificator is trained for each of them. After this the test frames are processed and a classification for all ROIs in the current frame is carried out and the results are shown graphically (Figure 3).

2.2. Input data

Video data from the i-ILDS image library is used for training and classification (Road_Hard_indeo52.avi and Road_Easy_indeo52.avi), to read it platform independently and to have easy access to individual frames, we converted the videodata to a sequence of png images.

The function “read_pngs(ctime,folder, frames)” reads the frames in the specified subfolder, were “ctime” donates the time periods that should be read e.g. [0,37;0,42;1,31;1,38]. A random set of frames can be read by passing the argument vector “frames”.

2.3. HOG descriptors

The HOG descriptors are calculated for a certain ROI, this is performed by the function “hog”. The gradient magnitude (GMAG) and the gradient orientation (GOR) are calculated for the whole image patch (ROI). After this step, the image is subdivided into cells (fig. 1). For each individual cell, the histogram of oriented gradients is calculated by accumulating all pixels, which have a GOR that falls in a certain range (specified by numBins, which is currently set to 9). The histograms for all cells are stored in a vector and returned. We also tried block-normalization but this did not increase the quality of the classification.

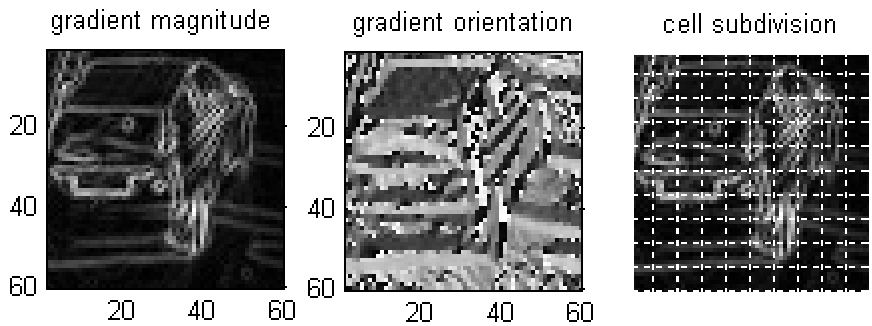

Figure 1: Graphical example of HOG calculation

2.4. SVM training

The training of the SVM takes place in “trainSVM.m”, which reads the training frames for a ROI. The classificator is trained with HOG descriptors calculated for a set of images containing cars (cFr) and another set that does not contain cars (ncFr). After this process a SVM structure is returned and can be used for classification tests in the ROI. Figure 2 shows an example of the training data used for the left lower region of the video.

Figure 2: Training data for left lower ROI

2.5. Classification

After the individual training for all 3 ROIs the classification is started (vc.m). Each frame of the test video is processed sequentially. The HOG descriptor is calculated for the current ROI in the test frame and is classified, if it is classified as a car, the counter “carTime” will be increased. A vehicle is counted, if it is classified positive for more than ten frames (carTime > 10). The detection is also shown graphically by a matlab plot, if a vehicle is detected, the color of the bounding box (for a certain ROI) changes to red (Figure 3).

3. Results

3.1. Evaluation

The evaluation of object detection systems is in general a difficult task, which can consume as much time as developing a detection method. Since every method has its own parameters and a different way of detecting objects, it has to be evaluated on an individual basis. Hence the assessment of the performance is an important task there has been some effort in developing a guideline of how to “characterize performance” [8] or in creating a set of unbiased metrics [9].

However for the evaluation of our algorithm we will use our own scheme. The performance assessment will take place on two levels:

Level 1 - evaluates the underlying object detection on the basis of single frames

Level 2 - quantifies the performance of counting of the vehicles

The sequence which is evaluated was created by training and testing our algorithm with frames out of the i-LIDS dataset. For creating the used Support Vector Machine (SVM) the Parked Vehicle Detection (Easy) video was used, for testing purpose the Parked Vehicle Detection (Hard) video was used.

Figure 3 shows a screenshot of the evaluation video. Our vehicle detection system was applied to three different sub regions marked by the rectangles. These different regions were chosen to see how robust the detection is for different background condition.

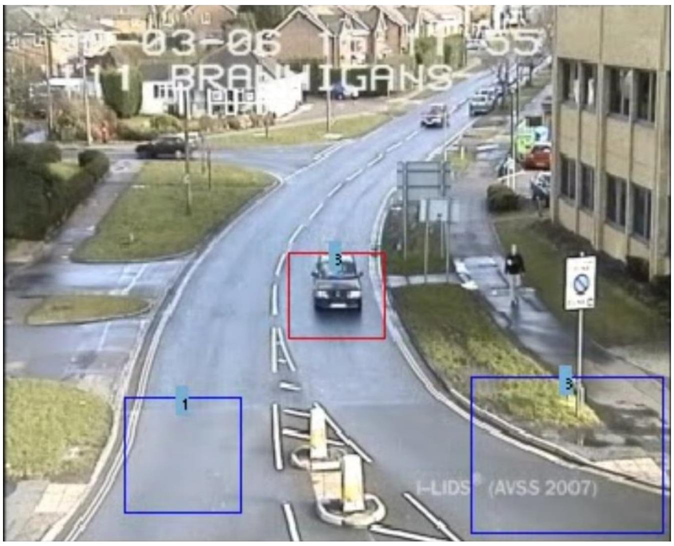

Figure 3: Screenshot of the evaluation video

3.1.1. Level 2 – Object detection on frame level

Object detection on frame level is very difficult to be evaluated for our system. Since we classify a whole area to contain a vehicle, we first would have to define what it means that an area contains a vehicle.

Instead we define a vehicle classified correctly on frame level if the vehicle is detected continuously from the first detection till the vehicle has passed the area. This means that there is no frame in-between where the vehicle is not detected.

Table 1 shows the evaluation on the basis of the frame level.

Table 1: Performance evaluation on frame level

	Region 1 (left)		Region 2 (right upper)		Region 3 (right lower)	
Correct	23	96 %	9	38 %	21	91 %
Incorrect	1	4 %	15	62 %	2	9 %
Total	24	100 %	24	100 %	23	100 %

The performance evaluation on the frame level shows good results for region 1 and 3. Occlusions by multiple cars within the area cause significant problems in vehicle detection in region 2.

3.1.2. Level 1 – Vehicle counting

Table 2 shows how well the algorithm performs on counting the vehicles. Therefore we must first define what counting a vehicle correctly means. A vehicle is classified as correctly counted if the number of counted vehicles is increased by one if the vehicle has left the measuring area. Incorrectly classified vehicles might be vehicles not counted or merged with other vehicles.

Table 2: Performance evaluation of the vehicle counting

	Region 1 (left)		Region 2 (right upper)		Region 3 (right lower)	
Correct	23	96 %	20	83 %	23	100 %
Incorrect	1	4 %	4	17 %	0	0 %
Total	24	100 %	24	100 %	23	100 %

Out of Table 2 we can see that counting in region 1 and region 3 does perform quite well for this short sequence. Region 1 does not count one vehicle, which is only due to the high velocity the vehicle is passing the area.

The problems in region 2 are due to occlusions by multiple cars within the area. As you can see the counting is highly dependent on the correct vehicle detection on the frame level.

4. Discussion

In this section we will discuss the strengths and weaknesses of our vehicle counting method in detail and give some possible improvements which can increase the performance.

4.1. Strengths

One of the strengths of our vehicle counting method is the simplicity of the HOG descriptor. It can be calculated quickly and delivers robust object detection, which can be used for the vehicle counting. Table 3 shows the number of images for every region our system has been trained with. You can see that the number of training data images is small.

Table 3: Dimension of training sets for the SVM

	Number of training images containing	
	vehicles	no vehicles
Region 1	24	22
Region 2	35	42
Region 3	33	35

According to our tests shadows do not cause problems, as they do with some motion detection methods. This is due to SVM training with no vehicle images containing shadows as it can be seen in Figure 2.

Apparently the HOG descriptors of pedestrians are discriminative from those of vehicles. So people crossing the street through our region of interest do not cause any false detection.

4.2. Weaknesses

Like a lot of other classification systems our vehicle counter has problems with occlusions if the region of interest contains multiple vehicles. The problem here is that it is not quite clear how occlusions can be overcome with our method.

Figure 4 shows the problem with two consecutive frames. The first frame is classified as containing a vehicle, the seconds as not containing a vehicle and the third image is the difference greyscale image between the first two. You can see that the difference between the two images is really small, but this difference results in one case as detection and in the other as no detection.

Figure 4: Object detection on occlusions by multiple cars

Another problem might be the illumination invariance of our approach, which we could not test with the available videos. According to [5] illumination invariance can be achieved by block normalization of the HOG, which we could not confirm, since the normalization led to decreased object detection in the test frames.

4.3. Possible performance improvements

Since occlusions by multiple vehicles cause counting problem, if possible the region of interest should be chosen smartly. It should be an area where occlusions rarely occur or vehicles do not tend to stop or linger for some time, since this might also cause problems.

However this is not possible for a lot of scenarios in reality so there must be other possibilities to improve performance.

Increasing the number of training images might lead to more robust vehicle detection.

The training data set could also include a third class, being trained with multiple vehicles within the ROI. If the ROI is then classified as containing multiple automobiles, this could trigger finer graded object detection.

Another possibility might be overlapping blocks of the HOG descriptor as it was described in [5].

A different approach could also be overlapping ROIs, which combined may be able to reduce occlusion problems.

5. References

[1] Chateau, T.; Malaterre, L.; Trassoudaine, L., “Vehicle Trajectories Evaluation by Static Video Sensors”, IEEE Intelligent Transportation Systems Conference, September 2006

[2] Wang, K.; Yao Q.; Qiao X.; Tang S.; Wang F., “Moving Object Refining in Traffic Monitoring Applications”, IEEE Intelligent Transportation Systems Conference, October 2007

[3] Khammari, A.; Nashashibi, F.; Abramson, Y.; Laurgeau C., “Vehicle detection combining gradient analysis and AdaBoost classification”, IEEE Conference on Intelligent Transportation Systems, September 2005

[4] Yang Z., Meng H., Wei Y., Zhang H., Wang X., “Tracking Ground Vehicles in Heavy-traffic Video by Grouping Tracks of Vehicle Corners”, October 2007

[5] Dalal N.; Triggs B., “Histograms of Oriented Gradients for Human Detection”, INRIA Rhone-Alps

[6] Adrian F. Clark, Christine Clark, “Performance Characterization in Computer Vision - A Tutorial”

[7] Bashir, F.; Porikli, F., "Performance Evaluation of Object Detection and Tracking Systems", IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS), June 2006 (PETS 2006)