Camera based vehicle detection and tracking using shadows and adaptive template matching

MARCO KRIPS†, ANDREAS TEUNER‡, JÖRG VELTEN†, and ANTON KUMMERT†

†University of Wuppertal
Communication Theory
Rainer-Gruenter-Strasse 21, 42119 Wuppertal
GERMANY

‡Delphi Delco Electronics Europe GmbH
Lise-Meitner-Strasse 14, 42119 Wuppertal
GERMANY

Abstract: - Enhancement of road traffic safety means that the human component as well as the technical components have to work at highest level. Together they build a complex system. Due to rising road traffic, mental and physical stress, driver’s concentration is deteriorating and is the main reason for traffic accidents. Improvement of the human factor is very difficult, but the introduction of driver assistance systems offers a wide range of applications that help to compensate or correct human faults. Typical accidents, caused by missing driver’s attention, are accidents during lane change. Drivers forget to assure that no other vehicle is alongside; especially they forget to check the blind spot. This paper presents our first approaches to detect vehicles in the blind spot of a driving car.

Key-Words: - Digital image processing, adaptive template matching, intelligent vehicle, driver assistance

1 Introduction

Enhancement of road traffic safety means that the human component as well as the technical components have to work at highest level. Together they build a complex system. Due to rising road traffic, mental and physical stress, driver’s concentration is deteriorating and is the main reason for traffic accidents. Improvement of the human factor is very difficult, but the introduction of driver assistance systems offers a wide range of applications that help to compensate or correct human faults. Typical accidents, caused by missing driver’s attention, are accidents during lane change. Drivers forget to assure that no other vehicle is alongside; especially they forget to check the blind spot. One of the basic assumptions for the presented method is the restriction to vehicles approaching from the rear. In this context the term vehicles includes cars and trucks. Typical road environment is German/European motorways. Moreover, in this first approach, daylight and good weather conditions can be assumed, i.e. no rain, snow, fog etc.

A cheap charge coupled device (CCD) grey level camera served as sensor and was mounted at the driver’s outside rear-view mirror to monitor the blind spot and the alongside lane. The goal of the system discussed in this paper can be characterized as follows. Approaching vehicles should be detected in time and tracked until they leave the blind spot. In the case of an intended lane change and a potential collision with another vehicle in the blind spot, the system should notify the driver of this potential hazard by acoustical or optical alarm.

2 Object hypothesis

Current research with respect to vision systems for vehicles scopes mainly the analysis of the area in front and behind the car (e.g. [1]). Frequently, the aim of these research works is an autonomous vehicle [2][3]. For driver assistance systems it is not necessary to reach all the objectives of an autonomous vehicle. Nevertheless, a robust system is needed that supports the driver in critical situations. The system presented here has not to identify the type of approaching vehicles, so it is unnecessary to differentiate a car from a truck.

Basically, the present case is a pattern detection problem, which has been solved for other applications in various ways. Thus, a huge variety of algorithms for object detection exists, e.g. neural network based, movement based, or shape based approaches. Usually neural network based approaches lead to a high computational effort, so that they are not considered here. A promising approach is the movement based method. There is already one commercial camera based driver assistance system available that uses optical flow [4]. But nevertheless optical flow computation is complex. Standard solutions in pattern recognition are shape based solutions, but it is impossible to realize vehicle detection by a standard template matching method. It is impracticable to have a complete database of all vehicles taking into account also optical distortion and the camera perspective. But even if it would exist, computations would take much time due to this large database. More general clues for vehicles have to be used. These clues must
be independent of the size, shape or colour of the objects, which have to be detected. One of these general features can be the shadow below the vehicle. In [5] and [6] it is shown that the shadow below a car can be used as a hypothesis for a potential vehicle.

Since the final system should be used in cars, the computational efforts have to be reduced to a minimum. Data reduction is an appropriate instrument. The definition of a region of interest (ROI) reduces the amount of pixels that have to be evaluated. The original size is 240(v)×320(h) pixels, whereas the ROI has a size of 129(v)×213(h) pixels (Fig. 1). If in the following the term frame is used, just the appropriate region of interest is meant.

2.1 Shadow hypothesis

Basic assumption of the shadow hypothesis implies that the grey values of pixels under a car are less than those of the paving outside the shadow. Information of the grey value range of the shadow pixels is needed to extract the shadow regions. Obviously, the lower limit of these grey values can be set to zero because in best case the shadow is black.

The upper limit cannot be set as easy as the lower one. The value is scene dependent and, therefore, can alter. Among other things the weather conditions, sunny or overcast weather, or the condition and the nature, respectively, of the paving have influence on this upper limit as well. Computation can be done by analyzing the free driving space (FDS)[6]. By free driving space the visible paving of the road is denoted. For the extraction of this part, the edge information of the image is determined by a Sobel operator. The FDS is extracted by scanning the image from the bottom to the first edge pixel. This bottom-up operation is done for every column of the image. As a result, the FDS is obtained (Fig. 2). It can be seen that edges caused by lane markings reduce the FDS due to the bottom-up scanning technique. Nevertheless, this method covers a sufficient number of pixels of the paving for doing the required computations.

Mean value, \( \mu \), and standard deviation, \( \sigma \), of the paving pixels are determined for this subset of the ROI. The upper limit of grey values of the shadow must be below the computed average. An appropriate threshold, \( T_{u\text{Limit}} \), for the upper bound is given by

\[
T_{u\text{Limit}} = \mu - 3\sigma ,
\]

where \( \sigma \) is the standard deviation and \( \mu \) denotes the mean value of the FDS pixels [6]. We are only interested in the change of brightness from bright to dark (if scanning bottom-up). Hence, we do not need to apply a standard edge detection method like Sobel operator. Furthermore, for the detection of vehicles far behind the own car, it can be assumed that the shadows are horizontally oriented. This assumption is not valid for vehicles’ shadows near to the camera due to the optical distortion and the camera perspective. Extraction of these horizontal edges can be achieved by the following method

\[
I_{h\text{Edge}} = \left\{ \left( I_{shift} - I \right) \geq T_{shift} \right\} \cap \left\{ I \leq T_{u\text{Limit}} \right\} ,
\]

where the resulting binary edge image is denoted by \( I_{h\text{Edge}} \), the original image is given by \( I \) and the one row to top shifted image by \( I_{shift} \). \( T_{shift} \) and \( T_{u\text{Limit}} \) are the thresholds for vertical shift (\( T_{shift} = 10 \)) and for the upper grey value limit of shadows, respectively. The threshold for vertical shift has been determined experimentally. The AND-operation can be described as the set union of horizontal edges and dark regions in the image. Results are given in Fig. 3.

Further improvements of the edge image concerning significant edges for vehicle detection, can be reached by using size constraints. As only shadows of far away vehicles should be regarded these shadows are characterized by a maximum and a minimum horizontal length.
Fig. 3. The ROIs of two different frames of a sequence are shown in the first column. Partial and final results of object hypothesis are presented in the next three columns. From left to right: after horizontal edge detection, improved results using length constraints, final results including entropy evaluation. Possible targets are denoted by a white rectangle.

If this is taken into account and if the very small and long edges are removed, the edge images can be improved (Fig. 3). Generally, the small edges behave like noise and the large edges are caused by lane markings and crash barriers. Both can be widely eliminated by the described methods.

2.2 Entropy verification

The shadow hypothesis method still leads to several possible targets. Unfortunately many of them do not belong to any vehicle. Therefore the number of targets has to be reduced by post-processing. This can be done by analyzing the entropy of the part of the image right above the pretended shadow. The region of a vehicle should have a higher entropy value than a background region [7]. As region of interest (ROI_{entropy}) a rectangular area above the position of the edge is chosen. The width \( w_e \) is depending on the length of the horizontal edge, the height \( h_e \) is 80% of the width. This region should cover a potential vehicle approximately.

For ROI_{entropy} the vertical and horizontal entropy is calculated. If both values are larger than a predefined threshold, the region is classified as a possible target/vehicle. In Fig. 3 results of positive entropy verifications are shown by white rectangles. But also the entropy constraint cannot provide targets restricted to vehicles only, since there are also background regions with high entropy.

2.3 Time and spatial constraints

For further reduction of misclassified targets a time and a spatial constraint is considered. As the movement of a vehicle in a short period, here 3 consecutive frames, can be assumed as linear and the speed as constant, the position of a possible target can move from one frame to the other only by some pixel.

Furthermore, the conditions for the before described methods can be assumed as to be constant during this short time period as well, hence the detection results should lead to similar results. Consequently, in each of these frames a possible target position is calculated and the distance between these positions should be small. If this is true, the verified flag of this possible target is set and post processing can be started for this target. Otherwise, if the time and spatial constraint is not met for three consecutive frames the possible target is dismissed. By this procedure the number of false targets can be reduced significantly, as objects in the background mostly occur as a kind of noise, i.e. only in one frame or maybe in two consecutive frames. Moreover, targets in the background move faster, as the relative velocity between them and the own car is greater than the one between two vehicles moving into the same direction. So, the change of object’s position is significantly larger and movement’s direction is of no interest.
3 Adaptive template matching

As the object hypothesis presented before does not lead to perfect vehicle detection, post processing is required. Therefore, an adaptive template matching (AdTM) method is used. AdTM is characterized by a self adjusted template and a matching score.

The initial template $P_0$ is extracted from the ROI entropy used for entropy calculation. So, the size of $P_0$ depends on the size of the horizontal shadow edge that has been extracted. Reduction of processing time implies reduction of the image size that has to be scanned. Therefore, a new region of interest ROI AdTM is extracted from the following frame. Dimension of ROI AdTM is defined by $w_{AdTM}$ and $h_{AdTM}$, which are depending on $w_e$ and $h_e$, respectively.

$$w_{AdTM} = 3 \cdot w_e \quad \text{and} \quad h_{AdTM} = 1.6 \cdot h_e \ . \quad (3)$$

The position of this rectangular area is around the position of the best match, except for the first frame, where the position is based on location of ROI entropy (Fig. 4). (Here, first frame denotes the first input of the AdTM and not the first frame of a sequence.)

The position of the template in the following frames is determined by template matching. The maximum of the normalized cross correlation [8] defines the position of the template/object in the current frame. Since the size of the approaching cars is growing, the template $P_0$ will not fit for complete tracking until the car will leave to the front. The value of normalized cross correlation will decrease. The template has to be changed, since conformance of before generated template image and current image is insufficient. Adjustment of template size to the size of the car is not recommended because the data amount and computing effort, respectively, will increase dramatically.

Keeping the template size constant implies that a new template has to be chosen on occasion. Therefore, a new template $P_n$ is generated whenever the maximum value of normalized cross correlation is less than a fixed threshold. $P_n$ is taken from the actual frame, with the position of the actual match as centre. The fixed size of these generated templates contains just 255 (15×17) pixels.

Additionally, an AdTM score is computed. The general assumption for this score can be characterized as follows. Vehicles entering the blind spot from the rear are characterized by their movement. Basically a potential vehicle moves from upper left corner to lower right corner of the image. Therefore, the horizontal and the vertical positions of a target in two successive frames are used to compute the direction of movement for both spatial components separately. Thus, a horizontal change is rated positive, if movement is from left to right, a vertical movement has a positive classification, if it is from top to bottom. From these two ratings the AdTM score is deducted. The score (initial value is 0) is reduced by 1 if any of the two ratings is negative, otherwise the score is increased by one. Evaluation of the AdTM score allows exclusion of further false targets, if the score is less than -2, because this proves that the target is moving to a wrong direction, e.g. parts of the background or vehicles on the opposite lane. In Fig. 5 such a false target tracking is shown. After three positive object hypothesis tests the target is confirmed (black rectangle in frame 6) and AdTM is started. The next two frames show the results of AdTM. A black square marks the position of the pretended vehicle. But due to the decreasing AdTM score the target is rejected after two further frames (Fig. 5, frame 8).

Furthermore, the AdTM score can handle vehicles that will temporary slowdown. During an existing tracking procedure for several frames the
velocity reduction of the approaching vehicle has no influence, since the AdTM score is considerably greater than 0 and only reduced by -1 for every frame in which the vehicle is slowing down. But, if slow down of the approaching vehicle will happen right after starting AdTM and will prolong for at least three frames, the target will be dismissed certainly. But the object hypothesis described before will classify the vehicle once again as a possible target, so that AdTM will restart. AdTM will stop either the score is less than -2 or the vehicle will leave the blind spot to the front.

4 Results
In Fig. 6 the results of a sequence, containing an approaching car, are shown. There are 6 consecutive frames given, starting from the left image in the first row (frame 1) to the second image of the second row (frame 6). Detected edges by the before described methods are marked by white lines in the grey level images. It can be seen that in the first frame the size of the detected shadow edge is undersized to be considered as a vehicle. Furthermore, many further short horizontal edges can be found, whose length are undersized as well. Incipient with frame 2 the object hypothesis provides a possible target in each following frame, denoted by the white rectangle. In frame 5 it is confirmed by the time and spatial constraint. The white rectangle is turned to black. In frame 6 the vehicle is detected by the object hypothesis as well as the AdTM, denoted by the black square. This type of double coverage is marked by a light grey rectangle and the black square. The last two images in Fig. 6 are taken from the same sequence. It can be seen that the AdTM provides a tracking of the car until it intrudes the blind spot area and moreover until the car leaves this area (not shown explicitly in Fig. 6). Besides, there is no detection of the shadow based object hypothesis, since the vehicle’s shadow provides hardly any horizontal changeover from lighter to darker grey. Due to the optical distortion and the camera perspective, the orientation is going from horizontal to diagonal during the vehicle is approaching.

5 Conclusion
The presented adaptive template matching (AdTM) provides an excellent tracking method. Despite the small size of the template, results show a robust tracking on the target. Due to the small template size and the adequately chosen region of interest, computational effort can be contained. Although results of this first approach are promising, further work is necessary. So, a detection process for motorcycles has to be included and different weather conditions have to be considered. Moreover, vehicles, which enter the blind spot from the front due to slowing down, have to be covered by the detection system. Also evaluation of the position of the target in terms of the blind spot area has to be implemented and an alarm strategy has to be defined for dangerous situations. Last but not least, system integration and validation of the algorithms in extensive test series are necessary.

References:


