

The Automatic Blind Spot Camera: A Vision-Based Active Alarm System

Kristof Van Beeck^(✉) and Toon Goedemé

EAVISE, KU Leuven, Technology Campus De Nayer, Sint-Katelijne-Waver, Belgium
{kristof.vanbeeck,toon.goedeme}@kuleuven.be

Abstract. In this paper we present a vision-based active safety system targeting the blind spot zone of trucks. Each year, these blind spot accidents are responsible for numerous fatalities and heavily injured. Existing commercial systems seem not to be able to cope with this problem completely. Therefore, we propose a vision-based safety system relying solely on the blind spot camera images. Our system is able to detect all vulnerable road users (VRUs) in the blind spot zone, and automatically generates an alarm towards the truck driver. This inherently is a challenging task. Indeed, such active safety system implicitly requires extremely high accuracy demands at very low latency. These two demands are contradictory, and thus very difficult to unite. However, our real-life experiments show that our proposed active alarm system achieves excellent accuracy results while meeting these stringent requirements.

Keywords: Computer vision · Real-time · Active safety systems

1 Introduction and Related Work

Each year the blind spot zone of trucks is responsible for an estimate of about 1300 casualties in Europe alone [9]. These accidents almost always occur in a similar fashion: the truck driver takes a right hand turn at an intersection and overlooks vulnerable road users (VRUs – e.g. pedestrians or bicyclists) which continue their way straight ahead. They have one of two distinctive causes: inattentiveness of the truck driver or the fact that these victims were located in a blind spot zone around the truck. Several such zones around the truck exist where the driver has only limited or no view. Evidently, the zone starting from the front right corner of the truck’s cabin which extends further to the right hand side of truck is the most crucial one. To cope with these blind spot zones, several commercial systems were developed. However, each of them has specific disadvantages, and as such none of them seem to handle this blind spot problem completely. These commercial systems can be subdivided into two main groups: *passive* and *active* systems. Passive safety systems still rely on the attentiveness of the truck driver himself. The most widely used passive systems are the – since 2003 obliged by law – blind spot mirrors. However, their introduction did not result in a decrease in the number of casualties [14]. A second passive safety

system is found in passive blind spot cameras. These systems are mounted in a robust manner and – using wide-angle lenses – display the blind spot zone on a monitor in the truck’s cabin when a right hand turn is signalled. However, these systems again rely on the attentiveness of the truck driver. Active safety systems automatically generate an auditive, visual or haptic alarm signal towards the truck driver (e.g. ultrasonic distance sensors) and as such avoid this disadvantage. However, existing active safety systems are unable to *interpret* the environment. No distinguishment between static objects (e.g. traffic signs) and VRUs is made, and often false positive alarms are generated. As such, the truck driver experiences this as annoying, and tends to disable the system. To overcome the aforementioned challenges, we present an active safety system relying solely on the input images from the blind spot camera. Using computer vision object detection methodologies, our system is able to efficiently detect VRUs in these challenging blind spot images, and automatically warns the truck driver of their presence. Such a system eliminates all disadvantages mentioned above: it is always adjusted correctly, is easily integratable in existing passive blind spot camera setups, does not rely on the attentiveness of the truck driver and implicitly provides some scene understanding. However, this is not an easy task. Indeed, these VRUs are multiclass (they consist of pedestrians, bicyclists, children and so on) and appear in very different viewpoints and poses. Furthermore excellent accuracy results need to be achieved for such a system to be usable in real-life scenarios. However, achieving high accuracy often comes at the cost of high computational complexity. This is unfeasible for our application: we aim to develop an active safety system which runs in real-time on low-cost embedded hardware. Additionally, we need to cope with the large viewpoint and lens distortion induced by traditional blind spot cameras. As a starting point, we employ the vanilla implementation of a VRU detection and tracking framework that we proposed in previous work [16], able to efficiently detect and track both pedestrians and bicyclists in these challenging blind spot camera images. However, currently this framework focuses on the aspect of VRU detection and tracking only. As such, the framework remains far away from a total active safety system. Therefore, in this paper we extend, polish and elevate this tracking-by-detection VRU framework into such an active alarm system. We present extensive real-life experiments and indicate that our final active alarm system meets the stringent requirements that such a system should achieve to be usable in practice.

To efficiently generate an alarm, we employ object detection algorithms that enable the detection of vulnerable road users in the blind spot images. A vast amount of literature concerning pedestrian detection exists. Since a detailed discussion is out of the scope of this work, we refer the reader to [1, 6–8, 18] for an extensive overview on the evolution of different pedestrian detectors. Several works exist which perform pedestrian and bicycle detection specifically for traffic safety applications, and are thus related to our work. However, to the best of our knowledge, no publications exist which explicitly discuss accuracy and usability results at safety system level. Furthermore only forward-looking cameras [3, 13] or stationary cameras [15] are used. Indeed, traditionally only forward-looking



Fig. 1. Our final active alarm system. See http://youtu.be/0S-uEPA_R5w.

datasets are employed. The images in Fig. 1 display example frames of the blind spot safety application we target here. As seen, the specific viewpoint and lens distortion significantly increases the difficulty. The remainder of this paper is structured as follows. In Sect. 2 we briefly discuss the vanilla implementation of [16] which we use as our baseline. In Sect. 3 we then discuss our extension, and elevate this baseline framework into a complete active safety system. We present accuracy results at system level. Based on these results, we discuss the usability of our safety system for real-life scenarios in Sect. 4. We conclude this paper and discuss future work in Sect. 5.

2 Baseline Algorithm

As baseline framework we start from the VRU detection and tracking framework for blind spot camera images presented in [16]. An overview of this VRU detection and tracking framework is visualised in Fig. 2. In a nutshell, the framework works as follows. As seen in the images of Fig. 1, the vulnerable road users appear distorted, rotated and scaled based on the position in the image. This scene knowledge is exploited as follows. We assume that the blind spot camera is mounted at a fixed position w.r.t. the ground plane. In this case, the exact transformation only depends on the position in the image. During an offline calibration, this distortion is modelled as a perspective transformation. Thus, for each region of interest (ROI) in the input image the transformation due to the viewpoint distortion is locally modelled. Based on this information each ROI is rewarped to an upright, undistorted fixed-height image patch. Since the scale is fixed, only a single search scale needs to be evaluated for the detection models. Next, image features are extracted on this rewarped patch. Three different detection models are evaluated. As object detection methodology, the deformable part based models (DPM) are employed [10, 11]. Apart from the standard pedestrian model, an upper body model and one of three bicycle components (i.e. different viewpoint) – selected depending on the position in the image – are evaluated. This upper body model, combined with a bicycle model enables the efficient detection of bicyclists. For each of these three models a probability map is generated. These hypothesis maps are then combined into a single detection probability map for that image patch. Finally, to cope with missing detections these detection maps are integrated in a tracking-by-detection methodology. For this,

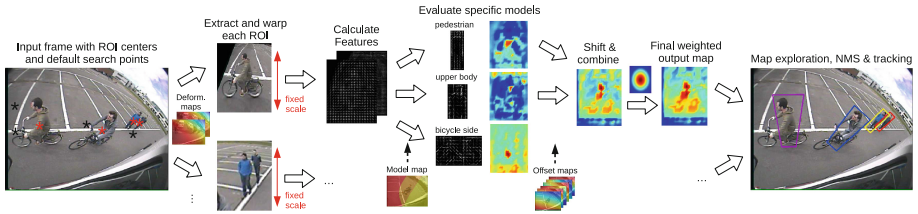


Fig. 2. An overview of the baseline blind spot VRU detection algorithm [16].

at strategic positions in the image initial search locations are defined (indicated with the black asterisks in Fig. 2). Each of these initial search locations is evaluated in each frame. If a VRU is found, a Kalman filter is instantiated based on a constant velocity motion model. In future frames, the next location is predicted and evaluated using the detection pipeline discussed above.

3 An Active Vision-Based Blind Spot Safety System

Our VRU detection and tracking framework from [16] is able to detect and track both pedestrians and bicyclists with high accuracy at reasonable processing speeds. Here we now elevate the tracking-by-detection methodology into an active safety system, and present extensive experiments as such. In its current form the existing framework has several caveats which need to be tackled first. Currently, the calculation time is non-deterministic. The tracking-by-detection framework relies on initial search coordinates which are defined at strategic positions in the image. When VRUs are detected at these positions, a new track is instantiated and evaluated in the consecutive frames. This approach thus implies that the processing speed depends on the number of tracks that are evaluated. Such non-deterministic behaviour is not suited for hard real-time applications where predictable *latency* and *processing speed* are of crucial importance. We must be able to guarantee that the system reacts within a constant time. Therefore, in Subsect. 3.1 we propose a methodology which tackles the non-deterministic behaviour of this framework. In Subsect. 3.2 we then elevate the tracking approach, convert it into a final active alarm system and present accuracy experiments as such.

3.1 Deterministic Calculation Time

We first define a blind spot zone in the image in which all pedestrians and bicyclists ought to be detected. Determining this zone correctly is of crucial importance for the effectiveness of our final alarm system. A strong correlation exists between the size of this detection zone and the latency of our final detection system. Most accidents occur when the truck makes a right turn without noticing pedestrians or bicyclists that continue their way straight ahead. Research indicates that it takes a worst-case reaction time of about 1.5s for a



Fig. 3. Blind spot zone. **Fig. 4.** 4×3 search grid. **Fig. 5.** 5×4 search grid.
(Color figure online)

truck driver to react when confronted with an event and undertake the effective break action [4]. Thus, an early detection of the VRUs is crucial. For this, a large detection zone is needed, which ideally starts far behind the truck itself. In such scenarios, e.g. fast moving bicyclists are detected early and enough time remains for the truck driver to interpret the alarm signals and undertake a corresponding action. However, such a large detection zone requires significant calculation power. Indeed, the size of this blind spot detection zone essentially determines a trade-off between the latency of our alarm system and the required computation power. To perform our consecutive validation experiments, we constructed a detection zone denoted in red in Fig. 3 (on the ground plane). All VRUs which enter this zone (approximately 6.60 m by 2.60 m) ought to be detected as soon as possible. The vanilla implementation performs the transformation based on the centroids of the VRUs. Therefore, we define the slightly larger and higher positioned detection zone displayed in green.

We now determine fixed search points within this previously defined (green) blind spot zone. At each point we perform the exact same approach as in [16]. Note that we still employ the tracking-by-detecting approach to cope with missing detections and to increase the robustness. However, we do not utilise the predicted Kalman future locations as input to our warping framework. These future locations are now only used to match detections from the previous frames in future frames. This is needed to cope with missing detections (e.g. in between two search points). Evidently, since the number of search points now is fixed, the calculation time becomes deterministic. We positioned these search points on a linear grid distributed in the above mentioned green detection zone. Evidently, the number of grid points determines a trade-off between the computational complexity and the accuracy. To determine this trade-off we evaluated the performance of five different grids, ranging from dense to fine: a 3×3 grid, a 4×3 grid, a 4×4 grid, a 5×4 grid and finally a 5×5 grid. As an example, two of these grids are visualised in Figs. 4 and 5. However, due to the non-linear distortion and specific viewpoint a linear grid does not represent the most optimal distribution. Therefore we developed an algorithm which splits up the blind spot zone in segments of optimal sizes, taking into account the rotation and scale robustness of our detector. For this we first we evaluated the invariance of our deformable part detectors with respect to both rotation and scale variation. These results are visualised in Figs. 6 and 7. As seen, slight variations on both

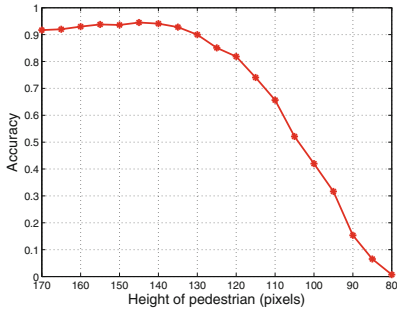


Fig. 6. Acc. of DPM vs. height.

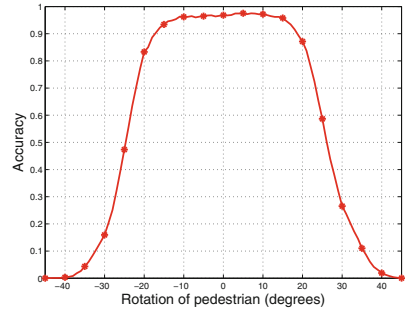


Fig. 7. Acc. of DPM vs. rotation.

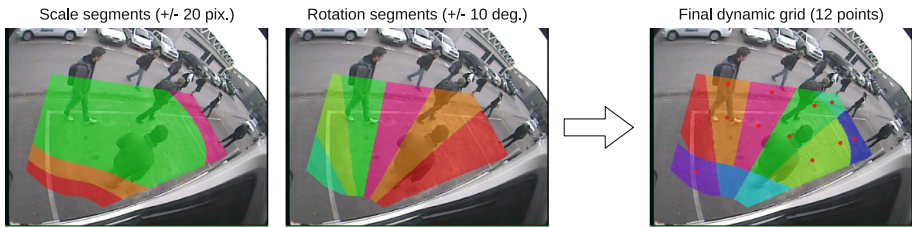


Fig. 8. The construction of the dynamic search grid.

the exact height and rotation cause negligible loss in accuracy. Based on these results, the delineated blind spot zone is segmented and search points are determined automatically as visualised in Fig. 8. This results in a grid of 12 points, which we coined the *dynamic grid*. We executed thorough experiments concerning both accuracy and speed when using these detection grids. To validate our algorithms we recorded real-life experiments using a commercially available blind spot camera (Orlaco 115°) and a genuine truck (Volvo FM12). For this, several dangerous blind spot scenarios were simulated, including both pedestrians and bicyclists. Our final test set consists of about 5500 blind spot image frames, in which about 8000 VRUs were labelled. Since we now only detect VRUs in this delineated blind spot zone, we evidently discard all annotations outside this zone. About 42% of all annotations are maintained. Figure 9 displays these accuracy results for all grids, and the original tracking-by-detection accuracy. The dashed black line indicates the accuracy of the VRU tracking-by-detection implementation from [16] which relies on initial search coordinates without the use of our blind spot zone. The full black line displays the accuracy for this original implementation when only taking into account detections and annotations in the blind spot zone as discussed above. As seen, a significant gain in accuracy is achieved. When only detecting the VRUs in the delineated blind spot zone, the framework achieves an average precision of 91.92%. This is due to multiple reasons. Annotations far outside the blind spot zone are difficult to detect, since they are very small. Furthermore, several different annotators were involved.

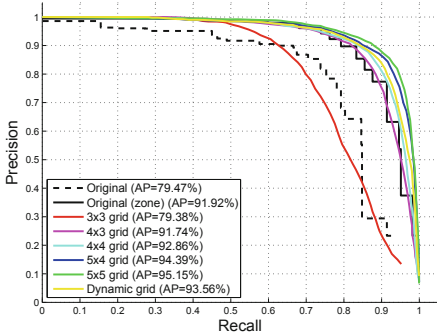


Fig. 9. Acc. when using fixed grids.

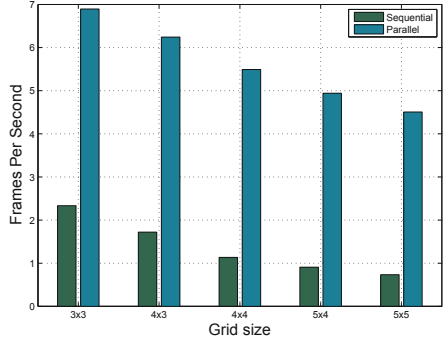


Fig. 10. Speed when using fixed grids.

This induced that often the exact location behind the truck, where the annotations start, significantly diverges between different sets. Due to the position of the initial search points in the tracking-by-detection framework, sometimes it was unfeasible to detect specific pedestrian or VRUs early enough. As seen, the 3×3 grid evidently is not dense enough. Similar accuracy performance as to the original tracking-by-detection framework is achieved with a grid of 4×3 points. Apart from the deterministic calculation time, these grids have another significant advantage over the standard tracking-by-detection framework. There, a VRU that is being tracked might be lost if for multiple frames in a row no detection is found (e.g. due to occlusions). No new search point is predicted in such cases. When using this default grid, tracks are much more easily recovered. Furthermore, we also performed experiments with additional datasets which include children. Our experiments indicated that, apart from adults, our approach is able to efficiently detect children without the need for additional search scales. This is due to the perspective transformation: when enough search points are utilised the probability that a child incidentally is retransformed to the fixed adult scale increases. Note that the accuracy results of Fig. 9 are still on single VRU detection capacity, the accuracy of the overarching blind spot alarming system is to be discussed in the next subsection.

Figure 10 displays the processing speed when using the fixed detection grids in the delineated blind spot zone. The evaluation is performed on an Intel Xeon E5 CPU at 3.1 GHz. The framework is mainly implemented in Matlab with time-consuming parts in both C and OpenCV. We tested both a sequential and a parallel implementation of our framework. Parallelisation was simply obtained by evaluating each grid point on a parallel CPU core. When using for example a 4×3 grid (an identical size to the dynamic grid), we achieve a parallel processing speed of 6.2 frames per second (FPS). However, even when evaluated in parallel, increasing the size of the detection grid lowers the detection speed. Note that we still employ Matlab which implies that multi-threaded processing and the data transfer between different threads is far from optimal.

3.2 Final Safety System

We now elevate the tracking-by-detection framework towards an autonomous active safety system. We step away from individual bounding boxes for each VRU, and generate an alarm if one or more VRUs are detected in the blind spot zone delineated as discussed above. Figure 1 shows how our final alarm system currently displays the detection of VRUs in the blind spot zone. For each frame we determine if an alarm needs to be generated. For validation we thus classify each frame as a *true positive* (TP), *false positive* (FP), *false negative* (FN) and *true negative* (TN). Thus, for each frame, we validate the effectiveness of our system. This is far from ideal, since no temporal information is taken into account. Furthermore, such evaluation metric is pessimistic when evaluating an alarm system such as in our application. Take for example the scenario where multiple true positive frames in a row occur when one or more VRU(s) enter the blind spot zone at the end of the truck. In such cases the truck driver is warned. Now suppose – due to e.g. missed detections – a few consecutive frames are classified as *false negatives*. While not optimal, in real-world scenarios this is less of a concern since the truck driver was already warned. In such cases only a short interruption of the (auditory) alarm signal would be noticed. However, the evaluation results presented here fail to take these considerations into account. To further improve the accuracy of our system, we evaluated the integration of temporal smoothing as follows. We aim to reject single (or short periods of) false positives. For this, we perform *majority voting* on a window sliding over the temporal frame per frame detection results. The exact size of this window (number of frames, N) is used as a parameter in our accuracy experiments. Figure 11 displays the accuracy results of our final active alarm system (black line), as compared to the accuracy of the original tracking-by-detection framework when only VRUs are accounted in the delineated blind spot zone (as discussed above, indicated with the dashed black line). Our active alarm system achieves an average precision of 97.26 %. We observe a significant accuracy improvement over the vanilla tracking-by-detection framework where the accuracy is defined by taking into account each individual VRU track. This improvement is explained by the fact that the exact accuracy conditions are now shifted towards the system level. Take for example two pedestrians walking side by side in the blind spot zone, where one pedestrian is (partially) occluded by the other. If in such case our tracking framework only detects the non-occluded pedestrian, and fails to detect the occluded pedestrian a false negative is counted resulting in a lower recall. Regarding our alarm system, finding only the non-occluded pedestrian in the blind spot zone is sufficient, since this already generates an alarm (and thus the frame is regarded as being a *true positive*). A similar observation for false positives exist. As mentioned, to further increase the detection accuracy we employ a sliding majority voting over a window of N frames. For this, we simply take the most occurring classification in the window of the N previous frames as a final decision for that frame. Figure 12 displays the precision recall curves of our alarm system for increasing sizes of this majority window (zoomed in on the top right corner). The black curve indicates our original algorithm ($N = 1$).

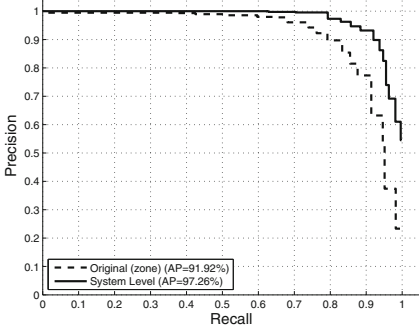


Fig. 11. Acc. of our final alarm system.

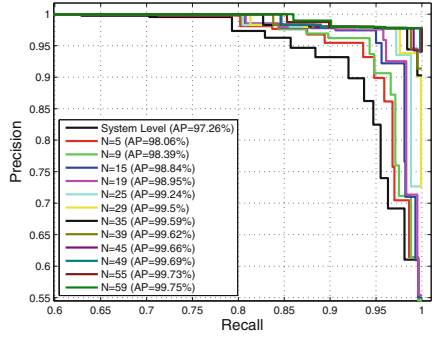


Fig. 12. Acc. for different sizes of N .

An increase in size of this window evidently increases the accuracy. The latency introduced by our majority voting scheme equals $\frac{N+1}{2}$ frames on average, 1 frame best-case and N frames worst-case. Take for example $N = 29$ frames. In such cases, our final alarm system achieves an average precision of about 99.5%. At a frame rate of 15 FPS (and taken into account real-time detection performance), an alarm is generated after worst-case two seconds.

4 Discussion

The acceptability and usability of an active alarm system as presented in the previous sections depends on many considerations. These range from reliability, predictiveness, false error rate and so on. An active blind spot detection system consists of at least two main components: the detection of the VRUs in the blind spot zone, and communicating this information to the truck driver. Although this latter component is of extreme importance towards the development of a commercial system, in this paper we focused on the first component: detecting the VRUs in an efficient manner. According to [12], two main technical criteria exist: the system should be able to perform *good* VRU detection and the system should be able to give an alarm system *in time* such that enough time remains for the truck driver to take action. We translated both criteria in three distinctive requirements: the *throughput*, *latency* and *accuracy* of the system. Here, we now discuss the required specifications and discuss the usability of our final alarm system in real-life situations.

Throughput. The throughput, defined in the number of FPS, is easily quantifiable. Evidently, to be used in hard real-time scenarios our alarm system should be able to classify each detection frame at least as fast as the rate at which new frames need to be analysed. Typical commercial blind spot cameras achieve a frame rate of 15 FPS. The detection speed of our final active alarm system depends on the number of grid points. At e.g. a 3×3 detection grid, a processing speed of about 7 FPS is achieved. Our final alarm system fails to meet

this requirement. However, the final multiclass detection framework (used as a baseline in our final active alarm system) served as a proof-of-concept Matlab implementation. Plenty of room for speed optimisation exists. Indeed, in [5] we proposed a highly optimised hybrid CPU/GPU implementation of the VRU detection system presented in [17], and proofed that real-time performance is achieved. Increasing the throughput of our final alarm system is only a matter of a more efficient implementation, e.g. no algorithmic redevelopment is needed.

Latency. The latency is defined as the time delay between the moment that VRUs enter the blind spot zone, and the moment that the system was able to raise an alarm. In the most extreme case, the truck driver should be warned before at least his reaction time (worst case 1.5 s) added with the time needed to perform a stopping maneuver in advance. Quantifying the time needed to perform this stopping maneuver is difficult, since it depends on several factors such as weather conditions, the combined mass of the HGV and its truckload. Due to the methodology of our active alarm system presented above, the maximally allowed latency depends on several factors. In our current final alarm system we defined that the zone in which VRU detection needs to be performed starts about 6.60 m behind the front of the truck’s cabin. However, this is only a design parameter and as such can be increased – at the cost of higher computational complexity. Indeed, an increase of this detection zone allows for a larger latency. Furthermore, the allowed latency of our detection system is correlated with the relative speed difference between the VRUs and the truck. For pedestrians, this is of limited concern. However, for bicyclists this needs to be taken into account. Additionally, when the *majority voting* scheme presented above is used, the latency is correlated with the accuracy. When increasing the number of frames (N) of the sliding majority window, the latency increases. The optimal value of this parameter depends on how the system is employed. Evidently, when no majority voting is used (i.e. $N = 1$) an immediate decision is taken for each individual frame. The latency thus equals the detection time for a single frame. If e.g. a detection speed of 15 FPS is achieved, a latency of only 67 ms exist. However, this ideal scenario is based on offline processing of the image frames, and ignores the latency introduced by e.g. the frame grabber when capturing the image frames from the camera. For the remainder of this section we assume that this latter time is negligible. When majority voting is used to increase the accuracy, an increase in latency occurs. As an example, suppose $N = 5$. The (worst case) latency between a bicyclist who enters the frame at the rear of the truck, and the generation of an alarm now takes 333 ms (at 15 FPS and real-time detection performance). The relative speed difference between the bicyclist and the size of the detection zone needs to be taken into account to determine if this latency is small enough. Suppose a bicyclist with a velocity of 20 km/h approaches a truck taking a right hand turn at 10 km/h. In such scenario, the time between entering the delineated detection zone and reaching the front right corner of the truck’s cabin equals about 2.4 s. Thus, slightly more than 2 s remain for the truck driver to react in this particular situation. As noticed, determining the minimal latency that such an alarm system requires is a difficult task. At

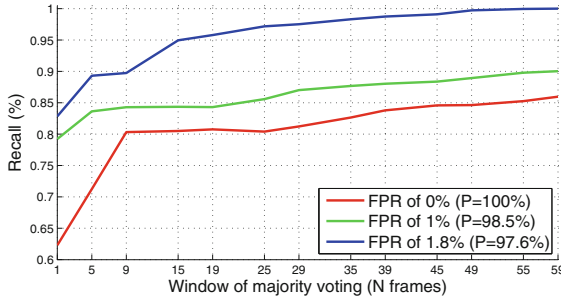


Fig. 13. The recall of our alarm system for three fixed values of the FPR.

best, when achieving real-time performance, our alarm system achieves a detection latency of 67 ms. In such cases, we achieve an average precision of 97.26 % on our test set. If needed, the detection accuracy could further be increased at the cost of an increase in latency. If such an increase in latency is not allowed, the VRU detection zone could be extended at the cost of extra computational complexity.

Accuracy. The system should be able to perform *good* VRU detection. As such, the false alarm rate and miss rate should be minimal. Exact quantitative figures of these false positive rates (FPR) are difficult to find and not consistent in the literature. For example, research on vehicle-based pedestrian collision warning systems state a maximum false alarm rate of 2 % and a miss rate of 1 % [2] need to be achieved, whereas [12] indicates a false alarm rate of up to 5 % is allowed. In Fig. 13 we plot the recall of our final alarm system in function of N for three fixed values of the false positive rate (and thus precision): 0 % (perfect precision), 1 % ($P=98.5\%$) and 1.8 % ($P=97.6\%$). This last FPR rate approximately equals the allowed rate in the literature. If we allow for a false positive alarm rate of 1.8 %, our final alarm system achieves for e.g. $N = 5$ a recall of 89.3 % (at a precision of 97.6 %). Thus, when allowing for a false positive rate of 2 %, about 90 % of the time VRUs are in the blind spot zone, an alarm is generated. For increasing N this raises quickly to about 98 % at 1.8 % missed detections ($N = 29$). A perfect recall is achieved for $N = 55$. However, defining the usability of a detection system solely on a specific value of the false positive rate is not optimal. For example, research indicates that a correlation exists between the acceptance of false alarms and the predictiveness of an alarm system [12]. The false alarms are easier accepted if the circumstances in which they occur are predictable. A safety system of which is known that false alarms occur during e.g. rainy conditions is easier accepted as compared to a safety system which produces a similar amount of false alarms at random moments. Furthermore, a fixed value for the false positive rate implies a specific operating point (i.e. detection threshold) of our system. At low detection thresholds most VRUs are detected, at the cost of relatively many false alarms (i.e. high recall at low precision). At high detection thresholds only instances that have a high probability of being a VRU

are detected (i.e. low recall at high precision. Defining the exact operating point of our active alarm system (and thus the trade-off between precision and recall) remains open for discussion. This highly depends on *how* the system is used. If utilised as autonomous safety system (e.g. as an autonomous emergency braking system), a perfect accuracy is needed. Currently, our active alarm system fails to achieve perfect performance. However, if used as a decision support safety system for the driver, it offers a significant advantage if a high recall is achieved at (almost) perfect precision, which is the case. For such scenarios the accuracy and usability of our alarm system is excellent. Take for example the precision recall curve in Fig. 12, for $N = 9$. If set for a perfect precision (i.e. no false alarms), a recall of 80.3% is achieved. Thus, in 80% of the time VRUs are present in the blind spot zone, our system generates an alarm. If a low false positive rate is allowed, the recall further increases. For example, at a recall of 85.4%, a false positive rate of only 1% is achieved. At these settings, about 34 frames of our entire dataset were classified as false positives, on a total of about 5500 frames (i.e. 0.62%). When measured in time units, about 27s for each driven hour a false alarm is generated. Keep in mind that these false positive frames are not distributed randomly over the entire dataset, since such single frame false positives are easily filtered out. The remaining false positives occur where an object in our dataset is misclassified as being a VRU for multiple frames in a row. After further examination, for these specific settings the main cause of the false positive frames was due to the false detections where, due to shadows – with some imagination – an upper body-like appearance is seen. Such false positives are easily eliminated when validated with e.g. an additional ultrasonic distance sensor. Furthermore, currently we assume that our alarm system continuously performs detection. However, commercial alarm systems are only activated at low speeds (using GPS) or when the driver signals a right hand turn.

5 Conclusions and Future Work

In this paper we presented a vision-based active alarm safety system guarding the blind spot zone of trucks. Our alarm system manages to efficiently detect vulnerable road users that are present in the blind spot zone, and actively generates an alarm. We showed that our system is able to meet the stringent requirements that such an active safety system requires, when used as a decision support active alarm system. At perfect precision, a recall rate of more than 80% is achieved. At low false positive rates and slightly higher latency, our alarm system reaches recall values of up to three nines five. These results are promising when keeping a commercial system in mind. However, we note that further large-scale tests are needed to draw final conclusions with respect to the usability of our active alarm system. Furthermore, the inclusion of additional sensors (such as long wave infrared and ultrasonic distance sensors) allow to further boost the accuracy.

References

1. Benenson, R., Omran, M., Hosang, J., Schiele, B.: Ten years of pedestrian detection, What have we learned? In: Agapito, L., Bronstein, M.M., Rother, C. (eds.) ECCV 2014. LNCS, vol. 8926, pp. 613–627. Springer, Heidelberg (2015). doi:[10.1007/978-3-319-16181-5_47](https://doi.org/10.1007/978-3-319-16181-5_47)
2. Chan, C.Y., Bu, F., Shladover, S.: Experimental vehicle platform for pedestrian detection. California PATH Program, Institute of Transportation Studies, University of California at Berkeley (2006)
3. Cho, H., Rybski, P.E., Zhang, W.: Vision-based bicycle detection and tracking using a deformable part model and an EKF algorithm. In: Proceedings of the 13th IEEE International Conference on Intelligent Transportation Systems (2010)
4. Coley, G., Wesley, A., Reed, N., Parry, I.: Driver reaction times to familiar but unexpected events (2008)
5. De Smedt*, F., Van Beeck*, K., Tuytelaars, T., Goedemé, T.: Pedestrian detection at warp speed: exceeding 500 detections per second (* indicates equal contribution). In: Proceedings of the Conference on Computer Vision and Pattern Recognition Workshops, Portland, Oregon (2013)
6. Dollár, P., Wojek, C., Schiele, B., Perona, P.: Pedestrian detection: a benchmark. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2009)
7. Dollar, P., Wojek, C., Schiele, B., Perona, P.: Pedestrian detection: an evaluation of the state of the art. *IEEE Trans. PAMI* **34**(4), 743–761 (2012)
8. Enzweiler, M., Gavrila, D.M.: Monocular pedestrian detection: survey and experiments. *IEEE Trans. PAMI* **31**(12), 2179–2195 (2009)
9. EU: Commission of the European Communities, European Road Safety Action Programme: mid-term review, 22 february 2006
10. Felzenszwalb, P., McAllester, D., Ramanan, D.: A discriminatively trained, multi-scale, deformable part model. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2008)
11. Felzenszwalb, P.F., Girshick, R.B., McAllester, D., Ramanan, D.: Object detection with discriminatively trained part based models. *IEEE Trans. PAMI* **32**(9), 1627–1645 (2010)
12. Hoedemaeker, D., Doumen, M., De Goede, M., Hogema, J., Brouwer, R., Wennemers, A.: Modelopzet voor dodehoek detectie en signalerings systemen (2010)
13. Jung, H., Ehara, Y., Tan, J.K., Kim, H., Ishikawa, S.: Applying MSC-HOG feature to the detection of a human on a bicycle. In: Proceedings of the 12th International Conference on Control, Automation and Systems (2012)
14. Knight, I.: A study of the implementation of dir. 2007/38/EC on the retrofitting of blind spot mirrors to HGVs (2011)
15. Takahashi, K., Kuriya, Y., Morie, T.: Bicycle detection using pedaling movement by spatiotemporal gabor filtering. In: Proceedings of the IEEE Region 10 Conference (2010)
16. Van Beeck, K., Goedemé, T.: Efficient multiclass object detection: detecting pedestrians and bicyclists in a truck’s blind spot camera. In: Proceedings of the 3rd IAPR Asian Conference on Pattern Recognition, Kuala Lumpur, Malaysia (2015)

17. Van Beeck, K., Goedemé, T., Tuytelaars, T.: Real-time vision-based pedestrian detection in a trucks blind spot zone using a warping window approach. In: Ferrier, J.-L., Bernard, A., Gusikhin, O., Madani, K. (eds.) *Informatics in Control, Automation and Robotics*. LNEE, vol. 283, pp. 251–264. Springer, Heidelberg (2014). doi:[10.1007/978-3-319-03500-0_16](https://doi.org/10.1007/978-3-319-03500-0_16)
18. Zhang, S., Benenson, R., Omran, M., Hosang, J., Schiele, B.: How far are we from solving pedestrian detection?. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2016)