Development of Active Safety Software of Road Freight Transport, Aimed at Improving Inter-City Road Safety, Based on Stereo Vision Technologies and Road Scene Analysis

V.E. Prun, V.V. Postnikov, R.N. Sadekov and D.L. Sholomov

Abstract The article considers the active safety system of road freight transport. The stereoscopic computer vision is the core of the system. The article also describes the major algorithms of active safety and the accuracy characteristics of algorithms' application.

Introduction

It is widely known that the major causes of road accidents (about 85%) are the mistakes and non-compliance with the traffic laws. Speeding, crossing into oncoming traffic, inattentive driving, sleeping during the driving, and other causes are the most frequent driving mistakes. One of the ways to improve the driving safety is to deploy active safety systems, aimed at timely notification of a driver of

V.E. Prun (✉) · V.V. Postnikov · D.L. Sholomov
JCS "Cognitive", Moscow, Russia
e-mail: v.prun@cognitive.ru

V.V. Postnikov
e-mail: pvv@cognitive.ru

D.L. Sholomov
e-mail: d.sholomov@cognitive.ru

R.N. Sadekov
MEI "Institute of Engineering Physics", Moscow, Russia
e-mail: sadekovlar@mail.ru

hazardous situations. Such systems are known as advanced driver-assistant system (ADAS) [1].

The potential of such systems is being intensively developed, primarily due to the technological equipment of the vehicles and the development of intelligent algorithms for the analysis of road scenarios. Many automobile manufacturers have already introduced for additional car features to facilitate the process of driving and to make it more secure and more comfortable. The laws enable the progress in the deployment of such systems, so that the introduction of active safety systems will be mandatory for new models of trucks from 1 Nov 2018 (EU Directive 3472012). Hence, the systems and their functional capabilities will be further evolving, which emphasizes the rationale for the research.

Related Works

Many automobile manufacturers as well as representatives of the research community are currently working in the field of development of 'advanced driver assistance systems'. The existing projects may differ in the techniques used: high-performance camera [2], radar, camera, and ultrasonic sensor [3], so in a list of solvable problems, e.g., speed bump detect [4], pedestrian detect [5], and vehicle detect [6].

At the Russian national level the 'advanced driver assistance systems' and the robotic transport systems are best developed by 'KB Avrora' [7] and the Central research and development automobile and engine institute NAMI [8].

The results presented in this paper are new both in terms of the techniques and the algorithms used, so they will be of great interest and used for both national and international developers.

General Description

The objective of the project is to create advanced driver assistance systems that provide the truck driver with efficient information on dangerous traffic situations according to the stereoscopic vision systems.

The introduction of active safety systems will increase the level of road safety, will reduce the number of road accidents, and will reduce the cost of ownership of the vehicle.

The following hardware configuration was used to implement the project. A laptop with Intel core i7 processor, 8 GB of RAM, and a video card GTX960 m (2 GB VRAM) served as an estimator for the stereo system. Two USB cameras (IDS) are combined with stereo rod with a 1.57 m base. The cameras are

synchronized via the synchronization cable in a master–slave mode. The degree of mal-synchronization is less than 5 ms. Geolocation sensor and strapdown inertial navigation system Xsens with integrated GPS antenna are used. Additional integration with sensors is also possible. Lidar is mounted on the front bumper and scans in a parallel to the road surface. Front radar is with narrow and mean angles, and infra-red range measuring system.

C++ with the library Open CV 3.1 [9] was the main development language. The complex is cross platform. The software is widely employing the parallel computing on multicore processors GPU with CUDA technology to provide the performance of ~ 10 frames per second.

Due to some limitation in scope, the paper deals only with the following most in-demand driver assistance systems: lane departure warning system, pedestrian protection system, forward collision warning system, and traffic sign recognition system.

Lane Departure Warning System

Lane departure warning system can work on roads with clearly visible lane markers at the speeds exceeding 50 km/h. The following sequential steps constitute the lane detection algorithm: identifying the road lanes, determining the parameters of lane markers, and selecting of road lanes.

First, the image I_t is filtered aimed to identify the lines which are road markers. The procedure is accomplished by using the difference of gaussians. This procedure converts the original color image (Fig. 1a) to a grayscale one.

Second, the fast Hough transform is used to identify the road lanes. The Hough transform is a numeric technique used in detecting the lines in an image. The fast Hough transform makes it possible to determine the common vanishing point of the lines in the image.

Third, the lines detected are checked to be parallel and further these lines are selected. The information from inertial sensors is employed at this stage, which

Fig. 1 The results of lane edges detection on the highway

allows to identify the horizon in the image (Fig. 1), as well as the number of road lanes.

Then, one can determine the dependence of the brightness of pixels on the length of the horizon line. The dependence has several extrema, which correspond to abscissas of the intersection points of the lines detected with the horizon. Considering the data on the number of the road lanes, we select the brightest intersection points, which are further referred to as road marking lines.

This software was tested on a set of the 5-s-long video clips. Actual vehicle departure off the road was visually determined in a case when the wheel crossed 10 cm or more over the road marking line (in compliance with the European standard [10]). The total number of the lane-changing video clips equaled 122. The tests of the proposed algorithm software showed that the true positive index (TP) was 83% while the number of false negatives (FN) was 1.2%.

The LDWS on a truck is active only when the lane changing is not supported by the turn signal.

Pedestrian Protection System

The pedestrian protection system is active all the time since the turn on of the system. The pedestrian protection algorithm, which is the core of the system, is built on the decision trees with broad attribute space [11]. The LUV transformation, the Gradient magnitude, and the Histogram of gradient are used to estimate attributes.

By the LUV transformation, we mean color space CIE L*u*v*, which is calculated from an image in RGB format, according to the known empirical formulas given in [12]. This transformation allows to perform image norming, to allocate the most effective signal and to even the outliers. The gradient is calculated in each point of the image using a formula of difference along two directions by applying a mask.

The histogram of oriented gradients (HOG) is formed as a sum of gradient module in the directions. The histogram consists of a number of cells. The size of each cell is 6 x 6 pixels.

Therefore, the attributive description of X pedestrians is represented by a vector comprising three LUV transformation parameters, one M, and six HOG parameters.

The method of decision trees with a depth of two was used as the classification method. The attributive information and its threshold value are located in the nodes of trees (obtained in the training stage) which is used to classify the measurements into two categories—with or without a pedestrian ($\mathbf{Y} = \{-1, 1\}$) in the image, respectively.

The AdaBoost algorithm is employed to train the classifier. Further two samples are developed for the learning procedure: positive sample with images of pedestrians and negative one with other images. The 90 x 36 pixel-size images with pedestrians in different weather conditions (winter, summer, rain) were used as reference images. The training samples can be mathematically represented as pairs

$\{\mathbf{X}_1, \mathbf{Y}_1\}, \ldots, \{\mathbf{X}_m, \mathbf{Y}_m\}$, where m is a number of pairs of values used for the training purposes. The classifier F: $F(\mathbf{X}) \rightarrow \mathbf{Y}$ is involved in the process of training the AdaBoost algorithm.

To detect pedestrians in the image we use previously trained classifier, and the detector receives a fragment of the original 90×36 size image. The search for pedestrians throughout the image, as well as the removal of constraints on the size of the pedestrians, is carried out by the use of sliding window and image scaling technologies. The reduction of the first and second kind errors is carried out by the use of the information on the scene geometry. The results of the pedestrian detection are exemplified in the figure (Fig. 2).

The detector application in the image is accelerated by techniques [11]. The results of testing for field data have proved that the TP probability of the system is 78.7%, and the FN is 9.4%.

The driver's alert of a potential hazardous proximity with a pedestrian is based on the calculation of time T_m (time to collision) before the collision and comparing it with a certain threshold, based on the equation

$$T_m = -Z/V = \frac{\Delta t}{S - 1} > t_{add}, \tag{1}$$

where Z is the relative distance between the objects, V is the relative speed, $S = w_{t-1}/w_t$, $w_{t,t-1}$ is the width of the rectangular area in the image describing the pedestrian at time instants t and $t - 1$, respectively, Δt is the time interval, t_{add} is the tolerance when exceeded, and an 'alert' message is activated.

Equation (1) is true at cruise. Given the acceleration, which is a frequent case because of heavy braking, Eq. (1) may be as follows:

$$T = T_m \cdot \frac{1 - \sqrt{1 + 2 \cdot C}}{C}, \tag{2}$$

Fig. 2 Pedestrians, detected in the image (marked with rectangles), and the line of the local horizon (thickened line)

where $C = 1 + \frac{T_m^t - T_m^{t-1}}{\Delta t}$.

Forward Collision Warning System

The system is implemented by the use of a hybrid detector. The system operation is carried out at speeds exceeding 5 km /h. The proposed algorithm for pedestrian detection, which is trained on samples containing vehicles, is used as a first detector; the AdaBoost with Haar-like features [Haar-like features] is the second method of detection.

Haar-like features consist of rectangular areas within which the intensities of the pixels of the original image (in grayscale) are summed up with weights $\{-1, 1\}$ for the black- and white-colored rectangles, respectively.

Examples of Haar-like features for detecting vehicles moving behind are presented below (Fig. 3). The first figure focuses on detection of the vehicle bumper, and the second and third figures show the side arches of the vehicle.

The AdaBoost algorithm employs decision trees with a depth of one as the weak classifiers. The detectors are simultaneously run, and each of them gives solutions with a probability from 0 to 100%. The probability is statistically calculated, by counting the number of matches of the objects found in the original image, taken at different scales and by moving the sliding window. The objects found by the first and the second detectors are summed and further filtered in the final stage with the information on the scene geometry. Upon the application of filtering, we evaluate the possibility of finding the object's height h at a given point of the scene, by using information from a stereo system; the object is excluded in a case when the operation is not possible.

Haar-like featured detectors proved to be good at detecting far-away objects. It works on a wide scale pyramid and detects objects ranging in size from 30 to 300 pixels. The average time of such a detector in the FULLHD image equals 50 ms. The realtime performance of object detectors is obtained by using geometric

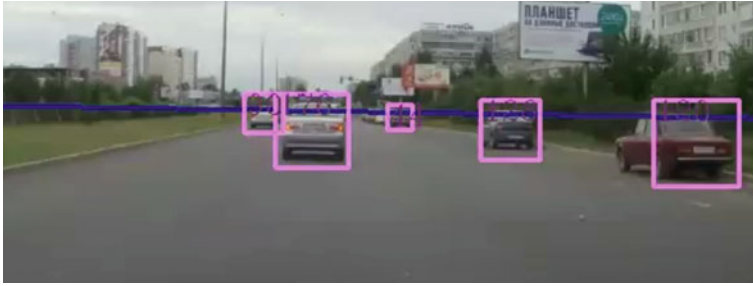**Fig. 3** Examples of Haar-like features used in the system

Fig. 4 Vehicles detected in the image (marked with *rectangles*) and the line of the local horizon (*dark thickened line*)

filtration technique, which is often reffered to as foveal vision module [13, 15]. The accuracy of detection is increased using image spatial normalization—forward rectification [14]. The quality of the detector operation is TP 86, 9% while the number of false negatives FN is 7% (Fig. 4).

Traffic Sign Recognition System

This system is designed to alert the driver of violations associated with the non-fulfillment of the traffic signs requirements.

A few detectors, each designed for the localization of particular category signs, are employed to detect road signs. The detectors used in the system are divided into two categories: first, 'bitmap' detectors, which calculate features on the fragment bitmap, and second, 'vector', which identify related components and analyze the latter by their shape and color distribution. The above division into categories is rather relative because the two types of detectors respond to both color and shape, but show their specific character.

Bitmap detectors use the Viola-Jones method, and the training is performed on a special color channel, calculated for a specific family of marks. For example, the channel $I = k_R R - k_G G - k_B B$ is used for limiting and warning signs, while the yellow-frame signs are calculated with $I = k_G G - k_B B$, and k_R, k_G, k_B are real coefficients. The Viola-Jones method uses features that are very similar to Haar-like features. Within the work done, we considered the features, aimed at recognition of rectangular, diamond-shaped, triangular, and circular forms. The system uses four bitmap detectors aimed at the detection of each class of marks and colors such as 'bitmap detector of rectangular blue marks'. These detectors are useful in a range of distances from 5 to 50 m on a wide-angle camera.

Vector detectors are used for an image, which is four times diminished to ensure the search speed. This detector is designed primarily for searching the signs at short distances (up to 30 m). The vector detector builds color masks, based on the

allocation of the required color in the HSV color space; then, it carries out morphology operations, defines related components, and selects objects in the image. It later filters components by selecting smooth-edged signs on the basis of the matrix filters and accepts the samples. The vector detector filters by finding the objects satisfying the predetermined geometrical shapes (circle, rectangle, triangle, and diamond) with dynamic programming methods.

The information on regions of interest intended to search for signs in places of their probable location can speed up the detectors' performance.

The vector detector is complemented by OCR module (Optical character recognition) for some groups of signs (such as 'speed limit'). This module also uses the analysis of related components; it removes decoys by calculating the width and the height ratios of the objects found, and it compiles and recognizes text fragments in the alphabet, which is characteristic for the analyzed subset of signs. The algorithms used were proposed by OCR Cognitive Cuneiform, Cognitive Forms, developed by the JSC 'Cognitive'.

The hypothesis, which includes the information on the detected region of an image, on a presumed sign group to which the latter belongs and on a priority, is a result of the detecting unit. The highest priority is given to the hypothesis, supported by both detectors. Coincidence is determined by calculating the degree of overlap of regions and comparing it with a threshold.

Neural network approach is further used to classify road signs. Neural network classifier is based on a fully related neural network with two hidden layers, trained to operate with large amounts of data. A three-channel color image of a road sign in the space of BGR ($16*16*3 = 768$ neurons for the first level of networks) is fed at the input of the network. The output layer contains 57 neurons, corresponding to the desired classes of signs.

We used 50 thousand signs, marked on the video frames to train the network. Real and synthesized images of signs should be used while training the second level of networks. The synthesized images were generated taking into account possible
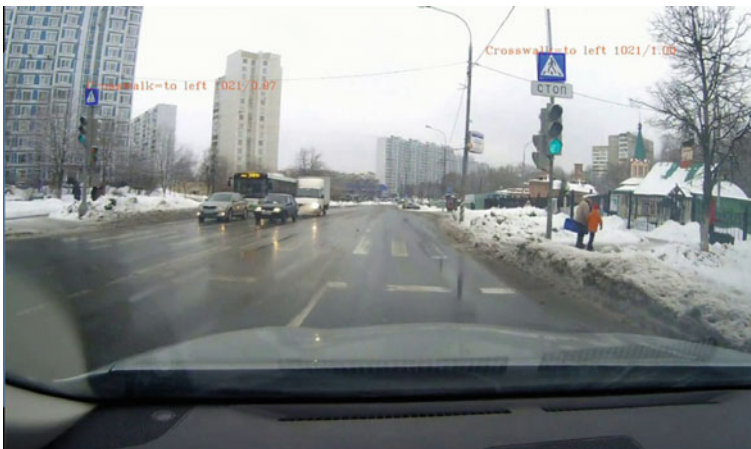

Fig. 5 Traffic sign recognition system (sign 'pedestrian crossing', the probability of detection = 1.0 for the *right* sign and = 0.97 for the sign on the *left*)

projective distortions, rotations, displacements, and other noises, typical of real images. The training sample set of networks included 'failure' classes, which do not contain fragments of road signs. Without taking into account the work of detectors, an independent assessment of the accuracy of the network was 99%. An example of the traffic sign recognition system is shown in the figure (Fig. 5).

The results of the complex testing of the whole sign processing subsystem (which includes a set of detectors and classifiers) for the advanced driver assistance system showed the following characteristics of TP, which is 80–85.4% and FN $\sim$ 6.2–10% (excluding stereo matching procedure).

Conclusions

The paper considers the active safety system of road freight transport. The main element of the software complex is a stereo camera. The paper describes the operation of the core systems of the software complex. There have been developed training and testing systems with the image database as a result of the project execution. The performance of all systems involved is of a frequency not lower than 10 Hz.

References

1. Mobilyeye/Main page (Online resource). Available from http://www.mobileye.com. Accessed on: 02.07.2016
2. Mihir, M., Swami, P.: High performance front camera ADAS applications on TI's TDA3X Platform. In: IEEE 22nd International Conference on High Performance Computing, pp. 1–8 (2015)
3. Khan, J., Elxsi, T.: Using ADAS sensors in implementation of novel automotive features for increased safety and guidance. In: 3rd International Conference on Signal Processing and Integrated Networks, pp. 753–758 (2016)
4. Devapriya, W., Nelson, C. Kennedy Babu, Srihari, T.: Advance driver assistance system (ADAS)—speed bump detection. In: International Conference on Computational Intelligence and Computing Research, pp. 1–6 (2015)
5. Geronimo, D., Lopez, A.M., Sappa, A.D.: Survey of pedestrian detection for advanced driver assistance systems. IEEE Trans. Pattern Anal. Mach. Intell., 1239–1258 (2010)
6. Lee, K.J., Kim, C., Lee, K.-R.: A 502-GOPS and 0.984-mW dual-mode intelligent ADAS SoC with real-time semiglobal matching and intention prediction for smart automotive black box system. IEEE J. Solid-State Circuits, 1–12 (2016)
7. KB 'Avrora'/Main page (Online resource). Available from http://www.kb-avrora.ru/projects/avtopilot-na-baze-gazel-biznes.html. Accessed on: 02.07.2016

8. Central Research and Development Automobile and Engine Institute NAMI Project Lada Kalina/Main page (Online resource). Available from http://nami.ru/projects/drone. Accessed on: 02.07.2016

9. Open CV/Main page (Online resource). Available from http://opencv.org/. Accessed on: 02.07.2016

10. EUR-Lex. Acessto European Union Law/Main page (Online resource). Available from http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32012R0351. Accessed on: 02.07.2016

11. Dollar, P., Perona, P.: The Fastest Pedestrian Detector in the West (2010)

12. CIELUV/Main page (Online resource). Available from https://en.wikipedia.org/wiki/CIELUV. Accessed on: 02.07.2016

13. Postnikov, V., Krohina, D., Prun, V.: Road shape recognition based on scene self-similarity. Proc. SPIE Vol. Vol., 9445 (2015)

14. Prun, V., Polevoy, D., Postnikov, V.: Forward rectification—spatial image normalization for a video from a forward facing vehicle camera. In: Ninth International Conference on Machine Vision, International Society for Optics and Photonics, pp. 103410W–103410W (2017)

15. Prun, V. et al.: Geometric filtration of classification-based object detectors in realtime road scene recognition systems. In: Eighth International Conference on Machine Vision (ICMV 2015), 9875p (2015)