Hindawi Publishing Corporation
EURASIP Journal on Embedded Systems
Volume 2007, Article ID 70256, 7 pages
doi:10.1155/2007/70256

Research Article

StereoBox: A Robust and Efficient Solution for Automotive Short-Range Obstacle Detection

Alberto Broggi, Paolo Medici, and Pier Paolo Porta

VisLab, Dipartimento Ingegreria Informazione, Università di Parma, 43100 Parma, Italy

Received 30 October 2006; Accepted 15 April 2007

Recommended by Gunasekaran S. Seetharaman

This paper presents a robust method for close-range obstacle detection with arbitrarily aligned stereo cameras. System calibration is performed by means of a dense grid to remove perspective and lens distortion after a direct mapping between image pixels and world points. Obstacle detection is based on the differences between left and right images after transformation phase and with a polar histogram, it is possible to detect vertical structures and to reject noise and small objects. Found objects' world coordinates are transmitted via CAN bus; the driver can also be warned through an audio interface. The proposed algorithm can be useful in different automotive applications, requiring real-time segmentation without any assumption on background. Experimental results proved the system to be robust in several envitonmental conditions. In particular, the system has been tested to investigate presence of obstacles in blind spot areas around heavy goods vehicles (HGVs) and has been mounted on three different prototypes at different heights.

1. INTRODUCTION

Problems concerning traffic mobility, safety, and energy consumption have become more serious in most developed countries in recent years. The endeavors to solve these problems have triggered the interest towards new fields of research and applications, such as automatic vehicle driving. New techniques are investigated for the entire or partial automation of driving tasks. A recently defined comprehensive and integrated system approach, referred to as intelligent transportation systems (ITSs), links the vehicle, the infrastructure, and the driver to make it possible to achieve more mobile and safer traffic conditions by using state-of-the-art electronic communication and computer-controlled technology.

In fact, ITS technologies may provide vehicles with different types and levels of "intelligence" to complement the driver. Information systems expand the driver's knowledge of routes and locations. Warning systems, such as collision-avoidance technologies, enhance the driver's ability to sense the surrounding environment. Driver assistance and automation technologies simulate the driver's sensor-motor system to operate a vehicle temporarily during emergencies or for prolonged periods.

Human-centered intelligent vehicles hold a major potential for industry. Since 1980, major car manufacturers and other firms have been developing computer-based in-vehicle navigation systems. Today, most developed/developing systems around the world have included more complex functions to help people to drive their vehicles safely and efficiently. New information and control technologies that make vehicles smarter are now arriving on the market either as optional equipment or as specialty after-market components. These technologies are being developed and marketed to increase driver safety, performance, and convenience. However, these disparate individual components have yet to be integrated to create a coherent intelligent vehicle that complements the human driver, fully considering his requirements, capabilities, and limitations.

In particular, concerning heavy goods vehicles (HGVs), many accidents involving trucks are related to the limited field of view of the driver: there are large blind spots all around the vehicle (see Figure 1). Some of these blind areas can be at least partly covered by additional mirrors. However, this is not always an optimal solution considering the aerodynamic effects and also the resulting complex driver interface.

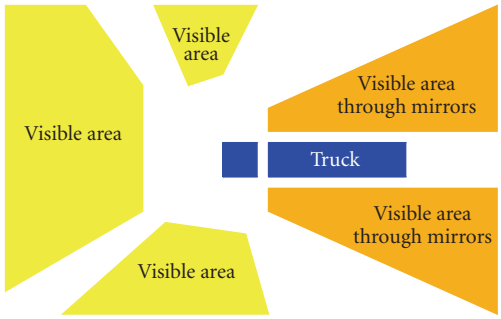


Figure 1: Field of view of a truck driver.

Figure 2: Typical dangerous situation.

Examples of traffic situations where the limited field of view can result in conflicts are

(i) starting from stationary at crosswalks or other places where a person or an object can be close in front of the vehicle,
(ii) lane change and turn situations to the passenger side,
(iii) situations with cross-road traffic sideways,
(iv) backup situations especially when ranging up to a loading dock.

This type of accidents accounts for approximately 10% of all accidents between trucks and unprotected road users and about 20% of all fatal accidents between trucks and unprotected road users.

The most effective single measure would be to improve the forward vision from HGV cabs so that an average size pedestrian could be seen even when standing right up against the front of the vehicle, see Figure 2. This would have been likely to save the lives of 12% of the pedestrians killed by HGVs. Changing the design of the front of a truck in this way is not an easy task. Similar benefits can be achieved by using sensors to detect the presence of a pedestrian or an obstacle and to warn the driver and also to prevent the vehicle from taking off when something is present in the forward blind spot: this is called start-inhibit, see Figure 3.

Embedded systems have to be compact and well designed for integration, but at the same time easy to use and to configure. In particular for a market ready product, there are some production aspects that get a central importance, for example, calibration procedure.

In all vision systems, calibration is one of the main topics, because it deeply affects algorithms performance; with our method the system is hardware-independent. In fact in case of accident or generic camera misalignment, the system can be restored after a recalibration (that could be done with automatic procedure with the vehicle parked in front of a grid). Even in case of cameras substitution for damage or commercial, reason system restoring would be done in the same way.

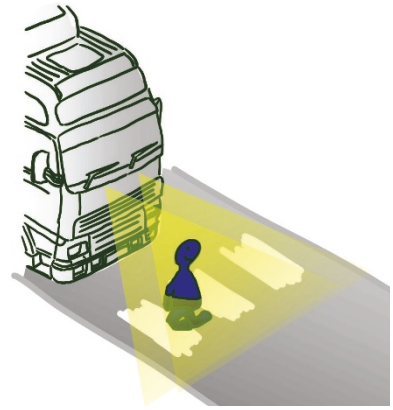


Figure 3: Start-inhibit protection system.

This is a strong point of StereoBox because it allows an easy installation and maintenance.

The system is composed of two cameras with sferic lenses to get a wide field of view, but introducing a strong distortion on images. They are placed in front of the truck and are arbitrary aligned, as will be discussed in Section 5. In particular, only the frontal driver blind spot area is framed by the cameras.

Two well-known approaches for stereo obstacle detection have been considered:

(i) the computation of the disparity of each pixel [1],
(ii) the use of stereo inverse perspective mapping [2].

An obstacle detection algorithm for offroad autonomous driving is presented in [1]. The dominant surface (e.g., the ground) is found through a v-disparity image [3] computation, while the obstacles come from a disparity space image (DSI) analysis. In this case, the cameras axes of the stereo system are almost parallel to the ground. Unfortunately, this approach is not suitable for start-inhibit, because one of the most important design issues is not to force a specific cameras alignment. In fact, the approach described in [1] requires a perfect camera alignment and precise constraints on cameras orientation.

Therefore a stereo inverse perspective mapping-based approach has been considered. The whole processing is performed by means of two main steps:

(i) lens distortion and perspective removal from both stereo images,
(ii) obstacle detection.

Concerning the first step, the problems of distortion removal and inverse perspective mapping without the knowledge of the intrinsic and extrinsic parameters of cameras

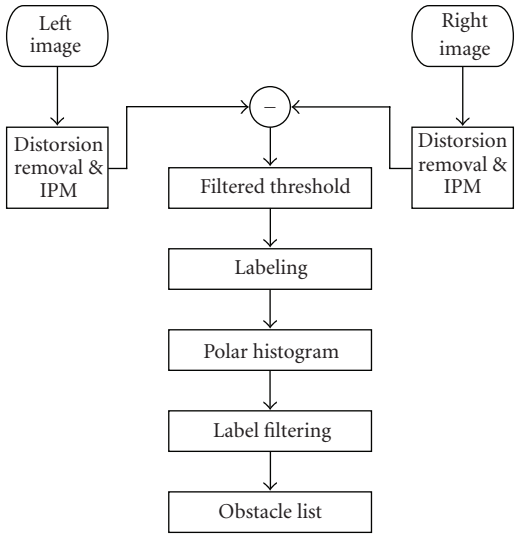


Figure 4: Algorithm's block diagram.

have to be solved. Lens distortion is usually modeled as polynomial radial distortion [4, 5] and it is removed by estimating the coefficients of this polynomial. After the distortion removal phase, extrinsic parameters are obtained [6], nevertheless, the highly complex mathematical model of the sferic lens may affect the computational time.

Therefore, a graphic interface to remove lens distortion has been designed to manually associate the grid points of the source image to their homologous points on a square grid on the *IPM* image [2] as explained in Section 2. This preprocessing is performed offline and the result are stored in a lookup table for a quicker online use.

In order to detect obstacles, two different approaches have been tested.

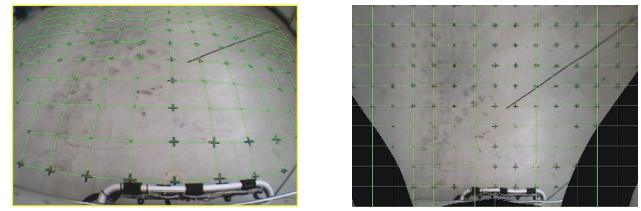
(1) The first searches for connected blocks on the thresholded image generated from the difference between left and right images after distortion removal and inverse perspective mapping (see [7]).
(2) The other one is based on the use of a polar histogram (see [8, 9]).

These two approaches have been fused into one algorithm to get the best from both. The whole algorithm flowchart is described in Figure 4 and is discussed in the following.

2. CALIBRATION

Camera calibration is one of the most important topics for vision systems especially when fielding systems that must be installed on real vehicles which have to operate in real scenarios.

In our case, highly distorting cameras are used without any knowledge about the intrinsic and extrinsic camera parameters. An analytic approach to calibration would be computationally prohibitive: the equations that are normally used to model sferic lenses become too complex when wide-angle lenses are used.



(a) (b)

Figure 5: Original and undistorted images of the grid.

Therefore, an empiric strategy has been used: during an offline preprocessing, a lookup table that allows a fast pixel remapping is generated; namely each pixel of the distorted image is associated to its corresponding pixel on the undistorted image. Images of a grid, painted on a stretch of flat road in front of the truck, are used to compute the lookup table (see Figure 5). A manual system to pinpoint all the crossing points on the source image is used.

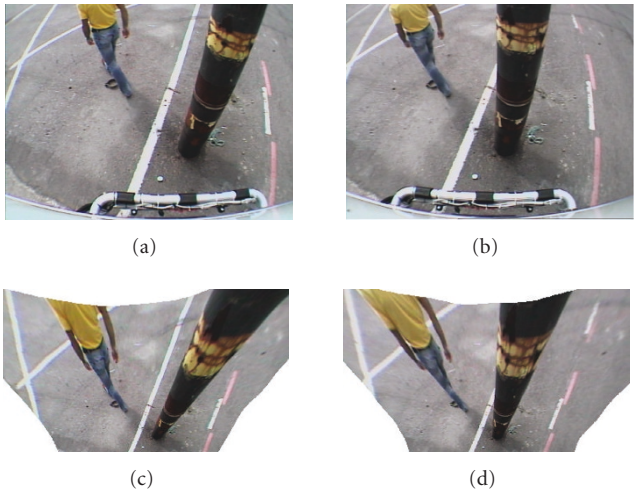
Thanks to the knowledge of the relative position of the truck with respect to the grid itself and to the assumption that the road can be considered nearly flat in the proximity of the vehicle, it is possible to compute a new image (the *IPM* image) removing both the perspective effect and camera distortion at once. A nonlinear interpolation function is used to remap the pixels of the source image that are not cross-points.

The process to determine coordinates (x, y) of the source image from the (i, j) pixels of the IPM image is divided into two steps.

Let us assume to have a grid with N vertical lines and M horizontal lines. For each vertical line of the grid, a function f_n is defined, where $n \in [1, N]$ is the line number. The spline creation is constrained by the correspondences between the crossing points of each line in the source image and in the IPM image; see (2) as an example, assuming x_1, y_1, x_2, y_2, and so forth, as the coordinates of the cross-points on the source image:

$$f_n(j) : \mathbb{R} \longrightarrow \mathbb{R}^2, \qquad f_n(j) = \begin{cases} f_n^x(j) \longrightarrow x, \\ f_n^y(j) \longrightarrow y, \end{cases} \tag{1}$$

$$\begin{aligned} f_1^x(0) &= x_1, \\ f_1^y(0) &= y_1, \\ f_1^x(1) &= x_2, \\ f_1^y(1) &= y_2, \\ &\vdots \\ f_1^x(N) &= x_N, \\ f_1^y(N) &= y_N. \end{aligned} \tag{2}$$



(a) (b) (c) (d)

Figure 6: Perspective and distortion removal: (a) left source image; (b) right source image; (c) left IPM image; (d) right IPM image.

Using functions $f_1(j), f_2(j), \ldots, f_N(j)$, another class of functions can be built, called $g_j(i)$ and defined as described in (3) with (4) as constraint:

$$g_j(i) : \mathbb{R} \longrightarrow \mathbb{R}^2, \qquad g_j(i) = \begin{cases} g_j^x(i) \longrightarrow x, \\ g_j^y(i) \longrightarrow y, \end{cases} \tag{3}$$

$$\begin{aligned} g_j(1) &= f_1(j), \\ g_j(2) &= f_2(j), \\ &\vdots \\ g_j(N) &= f_N(j). \end{aligned} \tag{4}$$

In this way, all the pixels of the IPM image have a correspondence to a pixel of the source image and the cubic spline interpolation method allows to get the best match between the two sets of pixels. An example of the resulting images obtained using these equations is shown in Figure 6.

Being the system based on stereo vision, two tables, one for each camera and both fixed under the same reference frame, are computed with this procedure. The lookup table generation is a time-consuming step, but it is computed only once when the cameras are installed or when their position is changed.

3. ALGORITHM

Starting from the *IPM* images, a difference image D is generated by comparing every pixel i of left image (L) to its homologous pixel of the right one (R) and computing their absolute distance:

$$D_i = |L_i - R_i|. \tag{5}$$

In particular, working on RGB color images, the distance used is the average of absolute differences of each color channel.

(a) (b)

Figure 7: Difference image between Figures 6(c) and 6(d) and result of labeling.

Then a particular threshold filter is applied on the resulting image D. In particular, for each pixel we define a square area A centered on it; the average value m of all the pixels in that area is computed and a threshold γ is applied on m. The resulting value is assigned to the pixel T_i as shown in the following equation:

$$\forall i \in D, \quad m = \frac{\sum_{\forall j \in A} D_j}{N_A}, \quad T_i = \begin{cases} 0 & \text{if } m < \gamma, \\ 1 & \text{if } m > \gamma, \end{cases} \tag{6}$$

where N_A represents the number of pixels in A.

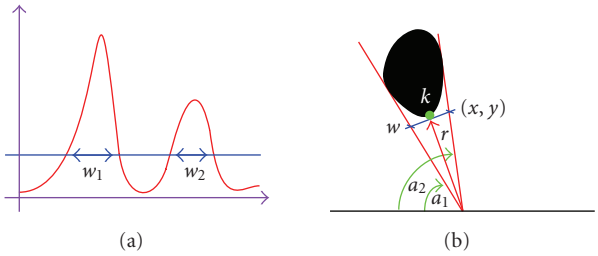
This is a kind of lowpass filtering and is useful to find the most significant differences in these images. Compared to similar methods like a thresholding followed by a morphological opening, it is faster because it is easy to be optimized and, nevertheless, works on the whole range of values of grey images.

Connected areas appearing in the resulting image are localized for and labeled: a progressive number is assigned to each label for further identification (as shown in Figure 7). A polar histogram is computed for each region. The focus used to compute the polar histogram is the projection of the mid point between the two cameras onto the road plane. These regions produce peaks on the polar histogram. Thus, the presence of strong peaks can be used to detect obstacles.

Some specific configurations of this histogram have to be considered, due to regions that are weakly connected or too thin to be a real obstacle. Therefore, it is necessary to further filter the polar histograms to remove regions that cannot be considered as obstacles.

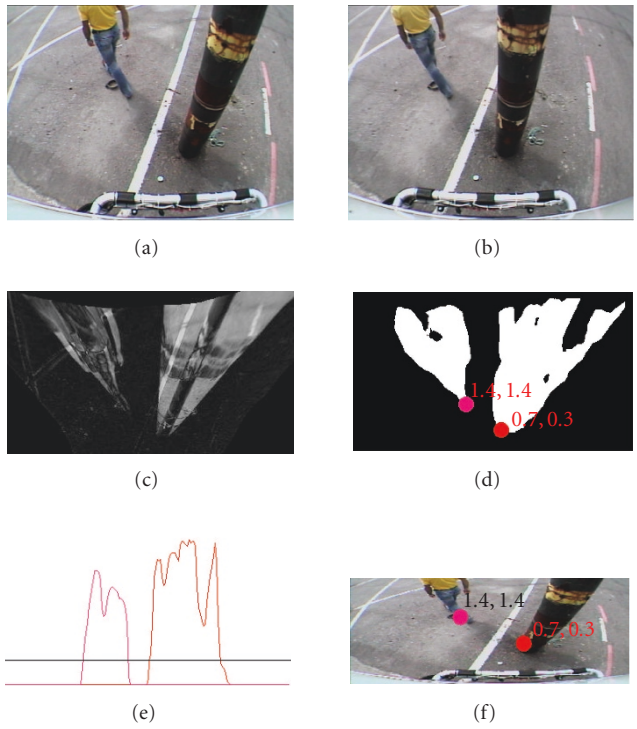
This filtering is performed considering the width of the histogram for the region of interest. The width of the histogram is computed in correspondence to a given threshold. When a polar histogram features several peaks, different values of width (w_1, w_2, etc.) are generated (see Figure 8(a)). If $\max\{w_1, w_2, \ldots, w_n\} > w_{\min}$ (where $w_{\min}$ is a width threshold), then the region previously labeled is maintained, otherwise it is discarded.

For each resulting region, the point k closest to the origin of the polar reference system and the angles of view (a_1, a_2) under which the region is seen are computed (see Figure 8(b)).



(a) (b)

FIGURE 8: (a) Polar histogram thresholding and filtering and (b) information extracted from the detected obstacle.



(a) (b) (c) (d) (e) (f)

FIGURE 9: (a) left source image, (b) right source image, (c) difference image, (d) connected components labeling, (e) polar histograms, (f) resulting image.

A rough width (w) of the detected object is computed as well, applying the following equation and considering r as the distance of k from the focus:

$$w = 2r \cdot \tan\left(\frac{a2 - a1}{2}\right). \tag{7}$$

Working on the *IPM* image, the location of point k in world coordinates can be estimated through the same lookup table previously used.

Figure 9 shows the complete set of intermediate results starting from the left and right original images; the difference and labeled images; the polar histogram whose filtering allows detecting one obstacle and discarding the small road curb; and finally the left original image with a red marker indicating the obstacle.

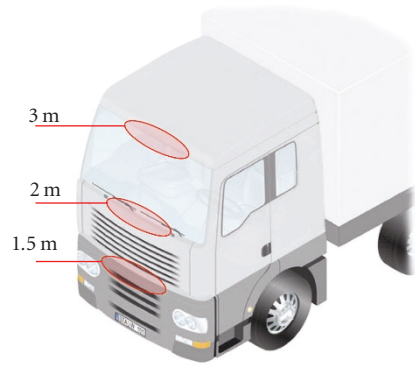


FIGURE 10: Possible position of the stereo pair.

(a) (b)

FIGURE 11: Two examples of StereoBox hardware.

4. COMPUTATIONAL REQUIREMENTS

The system presented in this paper was tested in several situations and with different architectures.

The algorithm can be applied to both progressive and interlaced images, widen the range of possible applications and hardware. Applied to a pair of 768×576 pixels interlaced color images, it takes approximately 30 milliseconds to be executed on an off-the-shelf Pentium4 running at 3.2 GHz. On the same architecture, working on stereo 640×480 progressive image retrieved from Bayer Pattern CCD sensor, the algorithm takes only 20 milliseconds to be executed on each frame.

Due to the small amount of resources required, the system was ported also on cheaper architectures. On Via EPIA EN15000 running at 1.5 GHz, analyzing stereo 640×480 progressive images, algorithm takes about 80 miiliseconds and it is thus capable to run up to 10 Hz.

5. SYSTEM SETUP

The stereo pair is placed right above the region of interest: in particular in all the different set-ups tested so far the cameras have been fixed in the front side of the vehicle.

The system was tested with cameras installed at several different heights: 3 m, 2 m, and 1.5 m, as shown in Figure 10. Stereo baseline and camera lenses must be changed accordingly. Values for baseline and focal length shown in Table 1 were chosen in order to view a given area.

Another important degree of freedom is cameras convergence: especially in case of large baselines or low heights, it is hard to view the whole region of interest with both cameras when their optical axes are parallel. Since images are

Figure 12: Result images showing typical algorithm output. A red dot shows the closest point of contact of each obstacle with the ground.

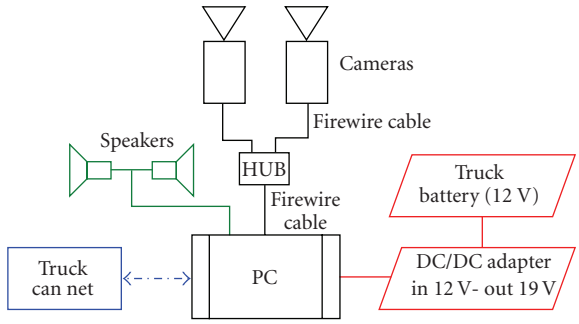


Figure 13: Block schema of the system.

preprocessed with a lookup table (as explained in Section 2) every effect introduced by freely placing of cameras is removed together with distortion and perspective.

In Figure 11 are shown different systems developed for two different projects.

The system is able to provide several types of output on several peripherals (typical application layout is shown in Figure 13).

Table 1: System specifications for different cameras heights.

Height (m)	Baseline (m)	Focal length (mm)
3	0.8	2.3
2	0.5	2.3
1.5	0.5	2.2

(i) The system can provide a visual output (e.g., on a display). This output consists in dedistorted image with mark on detected obstacles. A blinking red frame notify to driver danger condition.

(ii) An audio output: an intermittent sound is modulated according to distance and position of obstacles.

(iii) Through CAN bus, detected object's world coordinates are sent and a system can use this information to perform a high-level fusion with others' sensors.

(iv) Using CAN (or serial/ethernet) interface, the system can drive directly others' warning device (e.g., load torque on throttle command).

6. CONCLUSION AND FUTURE WORK

This paper presents an easy, fast, and reliable stereo obstacle detection technique for a start-inhibit system. Cameras mounted on a vehicle are arbitrary aligned, meaning that no special alignment is required by specialists or IT professionals. The choice of using a stereo vision system instead of radar or ultrasonic devices stems from the fact that the driver can see directly the image and can understand what caused the alarm.

Tests were made in several environmental conditions considering different kinds of road and obstacles, even with different illumination conditions. Low illumination conditions do not affect the system behavior because headlamps light up only the interesting part of vertical obstacles, easing the detection. To avoid light reflection, polarizing filters could be mounted in front of cameras.

Figure 12 shows some examples of the algorithm output remapped onto the original image. Red circles are used to mark obstacles positions. On the long tests performed, no false negatives were found: every single pedestrian and every tall enough obstacle were detected. Some false positives were generated by reflective road surfaces (water, e.g.).

Taking advantage of the stereo approach, the road texture, road markings, and shadows are successfully filtered out. Moreover, the algorithm easily detects large obstacles, rejecting most of smaller ones, like sidewalk borders. In general, due to the particular configuration of the system, vertical objects are correctly detected, thus the use of image tracking or temporal comparisons seems not mandatory.

Future developments will be centered on providing an automated algorithm to calibrate the system. A standard grid with easily recognizable markers will be placed in front of vehicle and an automated calibration procedure will be engaged by an operator. This procedure will become necessary only after major vehicle changes and/or maintenance.

ACKNOWLEDGMENT

The work described in this paper has been developed in the framework of the Integrated Project APALACI-PReVENT, a research activity funded by the European Commission to contribute to road safety by developing and demonstrating preventive safety technologies and applications.

REFERENCES

[1] A. Broggi, C. Caraffi, R. I. Fedriga, and P. Grisleri, "Obstacle detection with stereo vision for off-road vehicle navigation," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, p. 65, San Diego, Calif, USA, June 2005.

[2] M. Bertozzi, A. Broggi, and A. Fascioli, "Stereo inverse perspective mapping: theory and applications," *Image and Vision Computing*, vol. 16, no. 8, pp. 585–590, 1998.

[3] R. Labayrade, D. Aubert, and J.-P. Tarel, "Real time obstacle detection on non flat road geometry through "v-disparity" representation," in *Proceedings of IEEE Intelligent Vehicles Symposium*, vol. 2, pp. 646–651, Versailles, France, June 2002.

[4] D. Claus and A. W. Fitzgibbon, "A rational function lens distortion model for general cameras," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, vol. 1, pp. 213–219, San Diego, Calif, USA, June 2005.

[5] F. Devernay and O. Faugeras, "Straight lines have to be straight," *Machine Vision and Applications*, vol. 13, no. 1, pp. 14–24, 2001.

[6] R. Tsai, "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses," *IEEE Journal of Robotics and Automation*, vol. 3, no. 4, pp. 323–344, 1987.

[7] M. Bertozzi, A. Broggi, P. Medici, P. P. Porta, and A. Sjögren, "Stereo vision-based start-inhibit for heavy goods vehicles," in *Proceedings of IEEE Intelligent Vehicles Symposium (IVS '06)*, pp. 350–355, Tokyo, Japan, June 2006.

[8] M. Bertozzi and A. Broggi, "GOLD: a parallel real-time stereo vision system for generic obstacle and lane detection," *IEEE Transactions on Image Processing*, vol. 7, no. 1, pp. 62–81, 1998.

[9] K. Lee and J. Lee, "Generic obstacle detection on roads by dynamic programming for remapped stereo images to an overhead view," in *Proceedings of IEEE International Conference on Networking, Sensing and Control (ICNSC '04)*, vol. 2, pp. 897–902, Taipei, Taiwan, March 2004.