Statistical Modelling of Object Detection in Stereo Vision Based Driver Assistance

Jan Erik Stellet[1], Jan Schumacher[1], Oliver Lange[2], Wolfgang Branz[1], Frank Niewels[1], and J. Marius Zöllner[3]

[1] Robert Bosch GmbH, Corporate Research, Vehicle Safety and Assistance Systems, Schwieberdingen, Germany
[2] Robert Bosch GmbH, Corporate Research, Automotive Video Systems, Hildesheim, Germany
[3] FZI Forschungszentrum Informatik, Karlsruhe, Germany

Abstract. In this work, a statistical analysis of object detection for stereo vision based driver assistance systems is presented. Analytic modelling has not been attempted previously due to the complexity of dense disparity maps and state of the art algorithms. To approach this problem, a simplified algorithm for object detection in stereo images which allows studying error propagation is considered. In order to model the input densities, vehicle contours are approximated by Gaussian Mixture Models and distance dependent measurement noise is taken into account. Theoretical results are verified with Monte Carlo methods and real world image sequences. Using the proposed model, a prediction on the uncertainty in object location and optimal threshold selection can be obtained.

Keywords: driver assistance, stereo vision, object detection, statistical modelling, error propagation

1 Introduction

To an increasing degree, surround environment perception is used in recent developments in the driver assistance domain. Currently, employing the stereo vision principle to infer relevant information from the surroundings is an active research topic [1, 2, 10].

This contribution considers statistical modelling of stereo vision based driver assistance. The main motivation is to enhance system understanding for further improvement and prediction of system performance. Experimental approaches with ground truth information retrieved using additional sensors [11] or robotic vehicles [12] are costly. Comprehensive simulation-based evaluation on the other hand still poses unresolved challenges. Generating realistic synthetic stereo images is difficult due to the high complexity of relevant scenes and is approached by augmentation of real images [9]. Furthermore, because of the complex nature of dense disparity maps, state of the art algorithms in driver assistance applications feature heuristic methods and dependencies. Hence, rigorous analytic treatment often becomes infeasible in closed form. Previously, this has been considered only for a very simple scheme in the robotics domain [8].

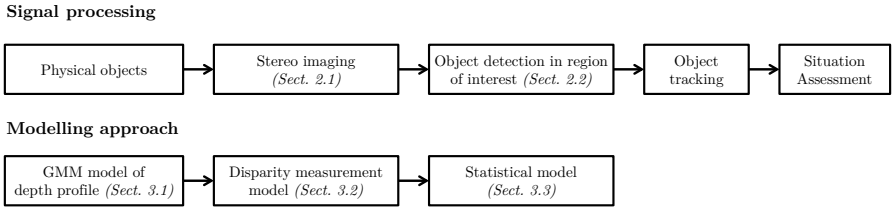


Fig. 1. Signal processing chain (top) and overview of modelling approach (bottom).

In this work, a simplified object detection algorithm for collision warning in car following scenarios is assumed. The intention is to provide means for statistical modelling of actual state of the art algorithms which can be derived from the abstract approach. The proposed scheme is based on stereo imaging and evaluates disparity measurements in central columns of the image. Similar principles are used in common full-scale methods where column-wise aggregation is performed not only in the central region but on the whole image [2, 10].

For this abstract detection algorithm, analytic error propagation to first order is studied. The statistical modelling is based on a probabilistic description of disparity measurement inputs. Compared to the challenge of modelling whole images, the limited scope of a column-wise aggregation gives an easier path towards finding input densities. These are composed of the depth profile of an object and measurement noise. A Gaussian Mixture Model is proposed to model an object's contour. The number of measurements on a single object is modelled depending on object height and distance to the camera. It is assumed that this distribution is corrupted by additive correlated Gaussian noise [1] which gives the final disparity measurement distribution.

Given this probabilistic model of disparity measurements, propagation to the detection algorithm output is analysed. The predictions given by the model are compared to empirical results from a set of real world image sequences.

First, a description of the exemplary system design is provided in Sect. 2. Section 3 presents error propagation from disparity map to detection output. Expected theoretical values are compared to empirical results in Sect. 4.

2 Background

Approaches to scene understanding from high resolution stereo images are typically split up into separate steps as depicted in Fig. 1 (top). In this work, the sensor signal processing part is considered. It comprises image generation, global image analysis and object detection. Subsequent steps include object tracking over time, estimation of motion parameters and situation assessment.

2.1 Stereo Image Measurements

Two cameras which form a rectified stereo system [2] with base-width b_w and focal length in pixel c are assumed for disparity generation. A 3-D point $\boldsymbol{p}^w = (x^w, y^w, z^w)^\top$ is given in Cartesian world coordinates where the x^w-axis is pointing in the driving direction. A Cartesian sensor coordinate system is aligned to the image plane of the left camera with the optical axis as x^c-axis. Transformation using the extrinsic camera parameters $\boldsymbol{R}, \boldsymbol{t}$ is:

$$\boldsymbol{p}^c = \boldsymbol{R}\boldsymbol{p}^w - \boldsymbol{t} \; . \tag{1}$$

For simplified expressions it is assumed that camera mounting angles are usually small and $\boldsymbol{R} \approx \boldsymbol{I}$. Without loss of generality a setup with $t_x = t_y = 0$ is considered. The origin of the Cartesian sensor coordinate system lies at pixel location (u_0, v_0) in the image. A point $\boldsymbol{p}^c$ is mapped to the image with a distance dependent disparity value d yielding uvd-coordinates:

$$\begin{pmatrix} u \\ v \\ d \end{pmatrix} = \frac{c}{x^c} \begin{pmatrix} y^c \\ -z^c \\ b_w \end{pmatrix} + \begin{pmatrix} u_0 \\ v_0 \\ 0 \end{pmatrix} \; . \tag{2}$$

2.2 Object Detection Algorithm

In the domain of driver assistance, one crucial task is to detect vehicles or other earthbound objects. Motivated by symmetry considerations, disparity images are commonly aggregated column-wise, either globally in a *vDisparity* representation [7] or locally in *stixels* [2, 10]. Here, simplification is achieved by only considering one central region of the image. It is assumed that within this narrow image region, relevant objects appear as similar to a vertical plane. Thus, a correspondence measure between measurements and an idealised vertical plane is calculated and used to indicate the presence of an object.

First, disparity measurements $\tilde{d}\,(u, v)$ are taken from Δu central columns $u \in [u_0 - 1/2\Delta u, u_0 + 1/2\Delta u]$. There are $n_{\Delta v}$ rows $v \in [\underline{v}, \underline{v} + \Delta v]$. Under the assumption of spatial homogeneity of an object contour, disparity values are condensed to their row-wise mean value to form *mDisparity* values $d\,(v)$:

$$d\,(v) = \frac{1}{n_{\Delta u}} \sum_{u=u_0-\frac{1}{2}\Delta u}^{u_0+\frac{1}{2}\Delta u} \tilde{d}\,(u, v) \; . \tag{3}$$

For an exemplary scene, this is visualised in Fig. 2(a)-2(b).

Secondly, detection of objects from measurements $\boldsymbol{d} = \{d_i\}_{i=1}^{n_{\Delta v}}$ is performed using a template matching approach in 3-D space [3]. A template is understood as a parametrised representation of all possible measurement realisations. It is assumed that relevant objects are ideally described as vertical planes frontoparallel to the camera with $x_0^w = \text{const}$. This approximately maps to disparity values $d_0 = \text{const}$. [10] and the template is thus defined by a single parameter d_0.

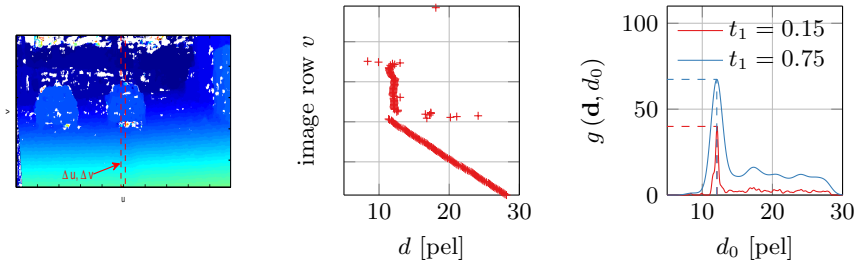
In order to decide whether a measurement comprises a relevant object and to determine its position, a similarity measure to the template is calculated. Here, this is evaluated using a Gaussian window on the distance $(d_i - d_0)$. Scaling the distance with t_1 (a system parameter) governs the tolerance to small deviations. Assuming independent measurements, summation over all image rows generates a correspondence measure $g(\boldsymbol{d}, d_0) \in [0, n_{\Delta v}]$ as shown in Fig. 2(c):

$$g(\boldsymbol{d}, d_0) = \sum_{i=1}^{n_{\Delta v}} \exp\left(-\frac{1}{2}\frac{(d_i - d_0)^2}{t_1^2}\right) . \tag{4}$$

Maximising (4) over the template parameter d_0 gives the realisation with $\hat{d}_{\text{obj}}$ that corresponds best to the measurements:

$$\hat{d}_{\text{obj}} = \arg\max_{d_0} g(\boldsymbol{d}, d_0) . \tag{5}$$

Distance estimates $\hat{x}^w_{\text{obj}}$ are calculated by transformation of $\hat{d}_{\text{obj}}$ to Cartesian coordinates according to (1)-(2). Relevant objects are differentiated from false measurements or clutter by a detection threshold on $g_{\max} := g\left(\boldsymbol{d}, \hat{d}_{\text{obj}}\right)$.



(a) Disparity map (b) Row-wise mean (c) Correspondence measure

Fig. 2. Scene with a vehicle at $x^w_{\text{obj}} = 12\,\text{m} \leftrightarrow d_{\text{obj}} \approx 12\,\text{pel}$. It can be recognised how the vehicle's rear is mapped to similar disparity values over multiple image rows v. The value of $g(\boldsymbol{d}, d_0)$ is calculated for two different values of the system parameter t_1. Increasing t_1 yields higher maximum values and less sensitivity to an object's contour but reduced distinction between object and background.

3 Statistical Modelling

Based on the system description given in Sect. 2, a statistical analysis that is structured according to Fig. 1 (bottom) is performed. First, an approach to model relevant aspects of object appearance is proposed. The obtained input

densities are then propagated to disparity space. In a third step, propagation to expected value and standard deviation of the detection algorithm's output is calculated.

3.1 Modelling of Objects

The appearance of objects in the narrow region of interest that is considered here depends on object dimensions and the depth profile along the vertical axis.

Coverage of Image Rows. The number of image rows $n_{\Delta v,\mathrm{obj}}$ that an object is mapped to depends on object height and distance to the camera. In the following, a deterministic expression is derived from geometrical considerations. Depending on environmental conditions and object's texture, invalid disparity measurement may occur in certain image regions. This can be modelled as a stochastic influence which overlays the deterministic one that is considered here.

An object with height h_{obj} at distance x^w_{obj} is vertically confined by its ground contact position at $\boldsymbol{p}^{w-}_{\mathrm{obj}}$ and upper edge at $\boldsymbol{p}^{w+}_{\mathrm{obj}}$ in Cartesian world coordinates. Evaluating the transformation to image coordinates (1)-(2) then gives:

$$v^+ = v_0 - \frac{cz^{c-}_{\mathrm{obj}}}{x^{c-}_{\mathrm{obj}}} \quad v^- = v_0 - \frac{cz^{c+}_{\mathrm{obj}}}{x^{c+}_{\mathrm{obj}}} \,. \tag{6}$$

These theoretical values might exceed the image's dimensions $(\underline{v}, \underline{v} + \Delta v)$. This has to be considered when calculating the difference $n_{\Delta v,\mathrm{obj}}$ between upper and lower image row. Neglecting this effect gives a simplified formula:

$$n_{\Delta v,\mathrm{obj}} = \min\left(\underline{v} + \Delta v, v^+\right) - \max\left(\underline{v}, v^-\right) \approx \frac{c_k h_{\mathrm{obj}}}{x^w_{\mathrm{obj}}} \,. \tag{7}$$

The exact and approximate solution in (7) are compared in Fig. 3 for three typical values of h_{obj}. Moreover, empirical values originating from a vehicle with $h_{\mathrm{obj}} = 1.4\,\mathrm{m}$ are displayed. For the simplified approximation, deviations can be observed at small distances where upper and lower edges lie outside of the field of view of the camera. Good correspondence is achieved for $x^w_{\mathrm{obj}} \geq 6\,\mathrm{m}$.

Modelling of Depth Profile. In order to model the distribution of relative depth values Δx which describe a contour, a finite mixture model is employed [4]. Here, a weighted sum of K Gaussians is chosen to allow closed-form expressions. The probability density function is defined by the parameters $\boldsymbol{\mu}_{\Delta x}, \boldsymbol{\sigma}_{\Delta x}, \boldsymbol{\rho}_{\Delta x}$:

$$p\left(\Delta x\right) = \sum_{k=1}^{K} \rho_{\Delta x,k} \mathcal{N}\left(\mu_{\Delta x,k}, \sigma^2_{\Delta x,k}\right) \,. \tag{8}$$

To describe the profile of a real vehicle, the parameters of a Gaussian mixture model (GMM) with $K = 3$ components are estimated from a sequence of measurements x^w taken in a stationary scene with $x^w_{\mathrm{obj}} = 6\,\mathrm{m}$. The centred depth measurements in Cartesian coordinates $\Delta x = x^w - x^w_{\mathrm{obj}}$ are used to estimate the model parameters with the expectation maximisation principle [4] (Fig. 4 left).

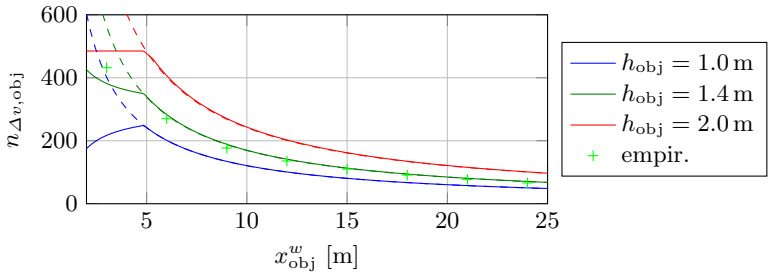


Fig. 3. Number of image rows an object with heigth h_{obj} at distance x^w_{obj} is mapped to. Exact solution (solid) and approximation (dashed) according to (7) compared to empirical values obtained with $h_{\mathrm{obj}} = 1.4\,\mathrm{m}$.

3.2 Disparity Measurement Model

In order to model statistical properties on *mDisparity* level, the contour description (8) is propagated through a measurement model including noise.

Measurement Errors. Several models have been presented for measurement errors in stereo vision. Error propagation through triangulation and quantisation on individual image pixels is considered in [14]. Algorithms that achieve sub-pixel accuracy lead to correlated errors over adjacent pixels. The overall disparity errors are commonly approximated as zero mean Gaussians with variance σ^2_{disp} [1, 8] and correlation coefficient ρ_{disp} between n_{corr} neighbouring pixels.

Taking the row-wise mean over $n_{\Delta u}$ disparity values according to (3) gives for the standard deviation σ_{noise} of measurement noise in *mDisparity* values:

$$\sigma_{\mathrm{noise}} = \frac{\sigma_{\mathrm{disp}}}{n_{\Delta u}} \sqrt{n_{\Delta u} + \rho_{\mathrm{disp}} L \left(2 n_{\Delta u} - L - 1\right)} \tag{9}$$

where $L = \min\left(n_{\Delta u}, n_{\mathrm{corr}}\right)$. In the following, correlation is only considered in the estimation of σ_{noise} and neglected between mean values from different rows.

Propagation to mDisparity Space. It is assumed that the observed vehicle's depth profile Δx is independent of distance. Minor perspective effects are hereby neglected. Then, the GMM $p\left(\Delta x\right)$ of an object at position x^w_{obj} is transformed to a GMM $p\left(d\right)$ in disparity domain. Expected values are transformed by the nonlinear mappping (2) between 3-D positions and disparity:

$$\mu_d = \frac{c b_w}{x^w_{\mathrm{obj}} + \mu_{\Delta x}} \,. \tag{10}$$

Secondly, the variance is calculated by Gaussian error propagation. Measurement noise as described by (9) is assumed independent of the object's contour:

$$\sigma_d^2 = \left(\frac{c b_w}{{x^w_{\mathrm{obj}}}^2}\right)^2 \sigma^2_{\Delta x} + \sigma^2_{\mathrm{noise}} \,. \tag{11}$$

The weights $\boldsymbol{\rho}_d$ are left identical to $\boldsymbol{\rho}_{\Delta x}$ [13].

A comparison of empirical measurements to an analytical GMM distribution $p(d)$ that is propagated according to (10)-(11) is shown in Fig. 4. Good correspondence is achieved for situations with $x^w_{\mathrm{obj}} > 6\,\mathrm{m}$ where the vehicle is fully visible in the image. In the first situation with $x^w_{\mathrm{obj}} = 3\,\mathrm{m}$ the vehicle is only partly visible from the car boot upwards. That is why measurements of the road in front of the vehicle (higher disparity values) are not present in the histogram.

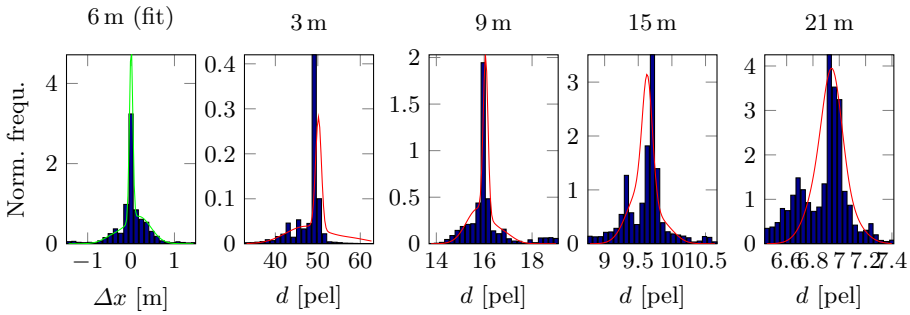


Fig. 4. Approximation of depth profile Δx from measurements at $x^w_{\mathrm{obj}} = 6\,\mathrm{m}$ (green, left) and propagation to disparity space according to (10)-(11) (red). Empirical measurements taken at the respective distance are shown for comparison (histograms).

3.3 Detection Algorithm

In the previous section, the input distribution of disparity on objects has been modelled. Now, propagation to the detection algorithm (5) will be analysed.

Expected Values. Given $n_{\Delta v,\mathrm{obj}}$ values with a GMM density $p(d)$, the expected value of the correspondence measure $g_{\max}$ is to be estimated. As correlation between *mDisparity* values of different rows is neglected, these are treated separately with $\mathbb{E}[g(\boldsymbol{d}, d_0)] = n_{\Delta v,\mathrm{obj}}\mathbb{E}[g(d, d_0)]$. One obtains:

$$\begin{aligned}\mathbb{E}[g(\boldsymbol{d}, d_0)] &= n_{\Delta v,\mathrm{obj}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\frac{(d-d_0)^2}{t_1^2}\right) p(d)\,\mathrm{d}d \\ &= n_{\Delta v,\mathrm{obj}} \sum_{k=1}^{K} \frac{\rho_{d,k} t_1}{\sqrt{t_1^2+\sigma_{d,k}^2}} \cdot \exp\left(-\frac{1}{2}\frac{(\mu_{d,k}-d_0)^2}{t_1^2+\sigma_{d,k}^2}\right) . \end{aligned} \tag{12}$$

In order to calculate the expected value of $g_{\max}$, the order of expectation and maximisation operators is changed:

$$\mathbb{E}[g_{\max}] \approx \max_{d_0}\left(\mathbb{E}[g(\boldsymbol{d}, d_0)]\right) . \tag{13}$$

Because an analytic solution to (13) does not exist in all cases for $K \geq 2$, the exponential in (12) is approximated by a first order series expansion:

$$\mathbb{E}\left[g\left(\boldsymbol{d}, d_0\right)\right] \approx n_{\Delta v,\mathrm{obj}} \sum_{k=1}^{K} \frac{\rho_{d,k} t_1}{\sqrt{t_1^2 + \sigma_{d,k}^2}} \cdot \left(1 - \frac{1}{2} \frac{\left(\mu_{d,k} - d_0\right)^2}{t_1^2 + \sigma_{d,k}^2}\right) . \tag{14}$$

Taking the derivative yields a unique maximum at:

$$\mathbb{E}\left[\hat{d}_{\mathrm{obj}}\right] \approx \left[\sum_{k=1}^{K} \frac{\rho_{d,k} \mu_{d,k}}{\left(t_1^2 + \sigma_{d,k}^2\right)^{\frac{3}{2}}}\right] \cdot \left[\sum_{k=1}^{K} \frac{\rho_{d,k}}{\left(t_1^2 + \sigma_{d,k}^2\right)^{\frac{3}{2}}}\right]^{-1} . \tag{15}$$

Finally, the expected value of the correspondence measure $g_{\max}$ is calculated from (12) evaluated at $\mathbb{E}\left[\hat{d}_{\mathrm{obj}}\right]$ from (15).

Variance in Object Distance. One is naturally interested in the uncertainty of the distance estimates $\hat{x}_{\mathrm{obj}}^{w}$ that are calculated from the disparity $\hat{d}_{\mathrm{obj}}$. For the considered detection algorithm this value is an implicit function $\phi\left(\boldsymbol{d}\right) := \arg\max g\left(\boldsymbol{d}, d_0\right)$ of the measurements $\boldsymbol{d}$. A linear approximation of the variance is obtained for $i = 1 \ldots n_{\Delta v}$ measurements distributed independently with $d_i \sim \mathcal{N}\left(\mu_{d,i}, \sigma_{d,i}^2\right)$. Then, Gaussian error propagation gives:

$$\mathrm{Var}\left(\hat{d}_{\mathrm{obj}}\right) \approx \sum_{i=1}^{n_{\Delta v}} \sigma_{d,i}^2 \left(\left.\frac{\partial \phi\left(\boldsymbol{d}\right)}{\partial d_i}\right|_{\mu_{d,i}}\right)^2 . \tag{16}$$

As there is no explicit expression for $\phi\left(\boldsymbol{d}\right)$, the implicit function theorem is used to calculate the partial derivatives [5]:

$$\frac{\partial \phi\left(\boldsymbol{d}\right)}{\partial d_i} = \frac{\partial \hat{d}_{\mathrm{obj}}}{\partial d_i} = -\left(\frac{\partial^2 g\left(\boldsymbol{d}, d_0\right)}{\partial^2 d_0}\right)^{-1} \left(\frac{\partial^2 g\left(\boldsymbol{d}, d_0\right)}{\partial d_i \partial d_0}\right) . \tag{17}$$

Calculating these derivatives for $g\left(\boldsymbol{d}, d_0\right)$ from (4) one obtains for (16):

$$\mathrm{Var}\left(\hat{d}_{\mathrm{obj}}\right) = \frac{\sum_{i=1}^{n_{\Delta v}} \sigma_{d,i}^2 \left[\exp\left(-\frac{1}{2} \frac{\left(\mu_{d,i} - \hat{d}_{\mathrm{obj}}\right)^2}{t_1^2}\right) \left(1 - \frac{\left(\mu_{d,i} - \hat{d}_{\mathrm{obj}}\right)^2}{t_1^2}\right)\right]^2}{\left[\sum_{i=1}^{n_{\Delta v}} \exp\left(-\frac{1}{2} \frac{\left(\mu_{d,i} - \hat{d}_{\mathrm{obj}}\right)^2}{t_1^2}\right) \left(1 - \frac{\left(\mu_{d,i} - \hat{d}_{\mathrm{obj}}\right)^2}{t_1^2}\right)\right]^2} . \tag{18}$$

One difficulty is that evaluating this expression requires the distribution parameters $\mu_{d,i}, \sigma_{d,i}$ for all image rows. A lower bound is obtained using the Cauchy-Schwarz inequality and assuming that the variance on objects is $\sigma_{d,i} = \sigma_d$:

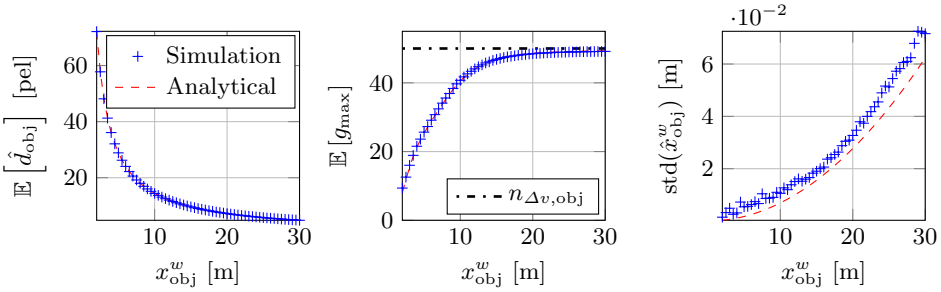
$$\mathrm{Var}\left(\hat{d}_{\mathrm{obj}}\right) \geq \frac{\sigma_d^2}{n_{\Delta v,\mathrm{obj}}} . \tag{19}$$

This concurs with the intuition that aggregating over image rows reduces the variance similar to averaging over $n_{\Delta v,\mathrm{obj}}$ measurements. Finally, the variance in distance $\hat{x}_{\mathrm{obj}}^{w}$ is obtained from the variance in $\hat{d}_{\mathrm{obj}}$ by linearisation of (2).

Monte Carlo Experiment. In a Monte Carlo simulation with $N = 1.000$ iterations, vehicle contour samples of $n_{\Delta v,\mathrm{obj}} = 50$ values are drawn from a GMM distribution $p(\Delta x)$. Each sample is mapped to disparity space for varying distances x^w_{obj} according to (1)-(2) and correlated Gaussian measurement noise is added. In every simulation run, the detection algorithm output (5) is calculated.

First, mean values of $\hat{d}_{\mathrm{obj}}$ and $g_{\max}$ are compared to the expected values (15) and (13) as visualised in Fig. 5(a). The closed-form expressions give a good estimate of the simulation results. As has been shown in Fig. 4, the disparity distribution narrows for higher distances. That is why $g_{\max}$ approaches $n_{\Delta v,\mathrm{obj}}$ as it is expected for the ideal case $d_0 = \mathrm{const}$.

Secondly, the standard deviation in $\hat{x}^w_{\mathrm{obj}}$ is calculated and shown in Fig. 5(b). In this second Monte Carlo simulation, only the realisations of the measurement noise are sampled whereas the vehicle contour distribution is drawn once. It is argued that while the contour can be conveniently modelled as a random variable Δx, it will typically not vary over time for the same object. Therefore, the analytic lower bound (19) is calculated with $\sigma_d = \sigma_{\mathrm{noise}}$ from (9). Due to the inversely proportional relationship (2) the uncertainty increases with distance.



(a) Expected value (b) Standard deviation

Fig. 5. Monte Carlo simulation results and analytic expressions.

4 Application and Experimental Results

Two possible applications of the theoretical modelling will be explored in the following: First, expected correspondence measure and standard deviation in position estimates of a vehicle are derived for different sensor noise levels. These are compared to empirical results from a set of real world image sequences. Secondly, the correspondence measure is predicted for different classes of objects. This allows to choose appropriate detection threshold values in order to separate unwanted detections from true objects.

4.1 Vehicle Detection

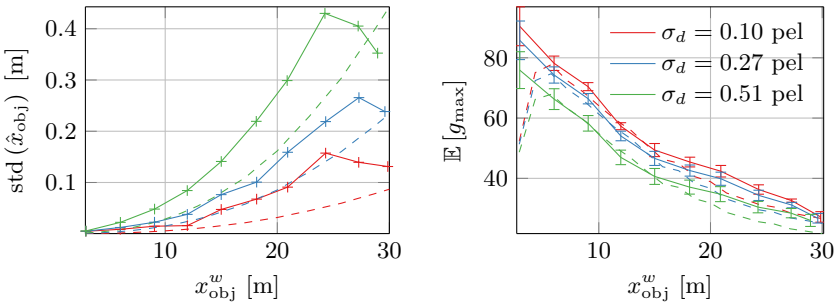
A stereo video sensor records sequences of a standing vehicle at positions $x^w_{\mathrm{obj}} \in \{3\,\mathrm{m}, 6\,\mathrm{m}, \ldots, 30\,\mathrm{m}\}$ as illustrated in Fig. 2(a). In order to emulate different sensor configurations or environmental conditions, additional correlated Gaussian noise with $\sigma_{d,\mathrm{add}} \in \{0.25\,\mathrm{pel}, 0.50\,\mathrm{pel}\}$ is added to the disparity images. Each sequence comprises approximately 230 frames. The detection algorithm from Sect. 2.2 with $t_1 = 0.5$ is applied to these measurements.

Analytic predictions on the detection algorithm results are determined as follows: Given the probabilistic description of the vehicle's contour from Sect. 3.1, the expectation of the correspondence measure $g_{\max}$ is calculated from (12)-(15). The uncertainty in distance estimates is evaluated from the lower bound (19).

Fig. 6(a) shows that standard deviation in $\hat{x}^w_{\mathrm{obj}}$ increases with distance and stereo image noise. In Fig. 6(b) the expected value of $g_{\max}$ is compared to the measurements. Good correspondence to the empirical results can be observed.

Overall, the number of image rows covered by an object's appearance decreases with distance which concurs with lower values of the correspondence measure (4). It should be noted however, that the reduction in $n_{\Delta v,\mathrm{obj}}$ as seen in Fig. 3 is much steeper than that of the corresponding values of $g_{\max}$, especially for distances $x^w_{\mathrm{obj}} \leq 12\,\mathrm{m}$. This effect has already been noted in the simulation results in Fig. 5(a) and is explained by the object's contour which causes a higher variation in disparity at small distances.

One limitation is that the vehicle's contour depth density has been derived for a vehicle which is fully visible in the image and thus does not perfectly represent the case of a partly visible vehicle at $x^w_{\mathrm{obj}} = 3\,\mathrm{m}$. As has been seen in Fig. 4 the depth distribution is then assumed too wide and therefore the predicted value of $g_{\max}$ becomes too small.



(a) Standard deviation in object position.

(b) Correspondence measure $g_{\max}$.

Fig. 6. Comparison of measurements from real stereo images (solid) and theoretical predictions (dashed). Variation of $g_{\max}$ within a sequence of measurements is illustrated by the $\pm 1\sigma$ error bars.

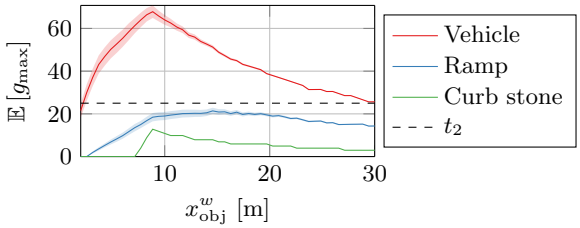


Fig. 7. Prediction of correspondence measure and $\pm 1\sigma$-bounds for different objects.

4.2 Threshold Selection

For the application of collision avoidance in vehicle following scenarios, collision relevant obstacles have to be separated from unwanted detections. One question is thus how the threshold on the continuous similarity measure has to be chosen.

Here, the proposed model is used to predict the correspondence measure for three exemplary objects: A vehicle, a ramp with slope of 8% as found in parking decks and a curb stone with height 20 cm. The depth profiles are modelled in form of a GMM for the vehicle's rear, a uniform distribution for the ramp and constant distance for the curb stone. Note that many depth profiles can be approximated by a GMM which allows to apply the derived analytical expressions. Results shown in Fig. 7 indicate that a detection threshold $t_2 = 25$ on $g_{\max}$ yields successful separation of a relevant vehicle from the two non-relevant cases.

5 Conclusion

This contribution has addressed the increasing need for model based performance evaluation in the domain of driver assistance systems. An object detection algorithm for dense stereo images as starting point for study of analytic error propagation has been proposed. Closed-form propagation of Gaussian input densities has been performed with first-order approximations. Standard deviation of position estimates has been obtained with Monte Carlo methods and an analytic lower bound has been derived. Future works will focus on more efficient calculation of propagation, e.g. the unscented transformation.

Having analytic expressions for the algorithm's performance also facilitates to draw conclusions in the inverse direction. Given a specification of detection threshold, sensor and system parameters one is able to formally characterise the detected physical objects in terms of height, distance and depth distribution.

One insight is that inspection of subsequent processing steps can be used to reduce the dimensionality of modelled input densities. In the exemplary system design, disparity measurements are aggregated row-independently in central image columns. Therefore, instead of describing high-dimensional stereo images, only a one-dimensional density is needed. A Gaussian Mixture Model is proposed to model the depth profiles of relevant objects. Exemplary results demonstrate

the approximation of a vehicle's rear and how the distribution is propagated to disparity space.

Propagation of uncertainty in object measurements to object tracking and situation assessment is a further relevant topic. The effect of sensor errors on collision warning algorithms has been analysed in [6,15] for idealised noise processes and can be extended to the measurement error model from this work.

References

1. Badino, H.: Binocular Ego-Motion Estimation for Automotive Applications. Ph.D. thesis, Goethe Universitaet Frankfurt am Main (2008)
2. Barth, A.: Vehicle Tracking and Motion Estimation Based on Stereo Vision Sequences. Ph.D. thesis, Friedrich-Wilhelms-Universitaet zu Bonn (2010)
3. Chang, P., Hirvonen, D., Camus, T., Southall, B.: Stereo-based object detection, classification, and quantitative evaluation with automotive applications. In: Computer Vision and Pattern Recognition - Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on. pp. 62–62 (2005)
4. DasGupta, A.: Asymptotic theory of statistics and probability. Springer (2008)
5. Fessler, J.: Mean and variance of implicitly defined biased estimators (such as penalized maximum likelihood): applications to tomography. Image Processing, IEEE Transactions on 5(3), 493–506 (1996)
6. Hillenbrand, J., Kroschel, K.: A study on the performance of uncooperative collision mitigation systems at intersection-like traffic situations. In: Cybernetics and Intelligent Systems, 2006 IEEE Conference on. pp. 1–6 (2006)
7. Labayrade, R., Aubert, D., Tarel, J.P.: Real time obstacle detection in stereovision on non flat road geometry through "v-disparity" representation. In: Intelligent Vehicle Symposium, 2002. IEEE. vol. 2, pp. 646–651 vol.2 (2002)
8. Matthies, L., Grandjean, P.: Stochastic performance, modeling and evaluation of obstacle detectability with imaging range sensors. Robotics and Automation, IEEE Transactions on 10(6), 783–792 (1994)
9. Nilsson, J.: Handbook of Augmented Reality, chap. Using Augmentation Techniques for Performance Evaluation in Automotive Safety, pp. 631–649. Springer (2011)
10. Pfeiffer, D.: The Stixel World. Ph.D. thesis, Humboldt-Universitaet zu Berlin (2012)
11. Pfeiffer, D., Morales, S., Barth, A., Franke, U.: Ground truth evaluation of the stixel representation using laser scanners. In: Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on. pp. 1091–1097 (2010)
12. Schneider, N., Gehrig, S., Pfeiffer, D., Banitsas, K.: An evaluation framework for stereo-based driver assistance. In: Real-World Scene Analysis 2011, LNCS. p. 2751 (2012)
13. Terejanu, G., Singla, P., Singh, T., Scott, P.D.: Uncertainty propagation for nonlinear dynamical systems using Gaussian mixture models. Journal of Guidance, Control, and Dynamics 31(6), 1622–1633 (2008)
14. Zhang, T., Boult, T.: Realistic stereo error models and finite optimal stereo baselines. In: Applications of Computer Vision (WACV), 2011 IEEE Workshop on. pp. 426–433 (2011)
15. Zheng, P., McDonald, M.: The effect of sensor errors on the performance of collision warning systems. In: Proceedings of the Intelligent Transportation Systems. pp. 469–474 (2003)