We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS BOOK CITATION INDEX INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

22

4D Ground Plane Estimation Algorithm for Advanced Driver Assistance Systems

Faisal Mufti[1], Robert Mahony[1] and Jochen Heinzmann[2]
[1]*Australian National University*
[2]*Seeing Machines Ltd.*
Australia

1. Introduction

Over the last two decades there has been a significant improvement in automotive design, technology and comfort standards along with safety regulations and requirements. At the same time, growth in population and a steady increase in the number of road users has resulted in a rise in the number of accidents involving both automotive users as well as pedestrians. According to World Health Organization, road traffic accidents, including auto accidents and personal injury collisions account for the deaths of an estimated 1.2 million people worldwide each year, with 50 million or more suffering injuries (Organization, 2009). These figures are expected to grow by 20% within the next 20 years (Peden et al., 2004). In the European Union alone the imperative need for Advanced Driver Assistance Systems (ADAS) sensors can be gauged from the fact that every day the total number of people killed on Europe's roads are almost the same as the number of people killed in a single medium-haul plane crash (Commission, 2001) with 3^{rd} party road users (pedestrian, cyclist, etc) comprising the bulk of these fatalities (see Figure 1 for proportion of road injuries) (Sethi, 2008). This transforms into a direct and indirect cost on society, including physical and psychological damage to families and victims, with an economic cost of 160 billion euros annually (Commission, 2008). These statistics provide a strong motivation to improve the ADAS ability of automobiles for the safety of both passengers and pedestrians.

The techniques to develop vision based ADAS depend heavily on the imaging device technology that provides continuous updates of the surroundings of the vehicle and aid

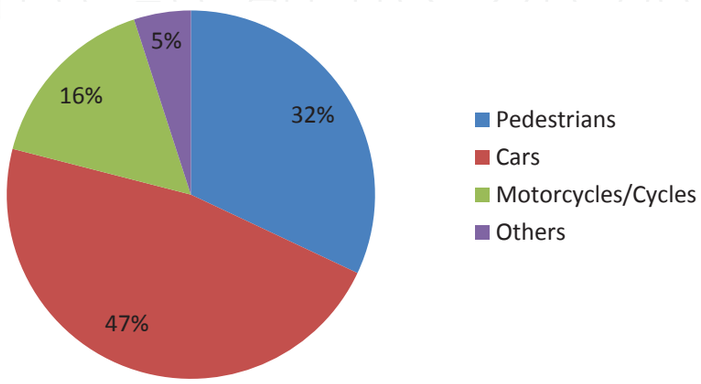


Fig. 1. Proportion of road traffic injury deaths in Europe (2002-2004).

drivers in safe driving. In general these sensors are either spatial devices like monocular CCD cameras, stereo cameras or other sensor devices such as infrared, laser and time-of-flight sensors. The fusion of multiple sensor modalities has also been actively pursued in the automotive domain (Gern et al., 2000). A recent autonomous vehicle navigation competition DARPA (US Defense Advanced Research Projects Agency) URBAN Challenge (Baker & Dolan, 2008) has demonstrated a significant surge in efforts by major automotive companies and research centres in their ability to produce ADAS that are capable of driving autonomously in an urban terrain.

Range image devices based on the principle of time-of-flight (TOF) (Xu et al., 1998) are robust against shadow, brightness and poor visibility making them ideal for use in automotive applications. Unlike laser scanners (such as LIDAR or LADAR) that traditionally require multiple scans, 3D TOF cameras are suitable for video data gathering and processing systems especially in automotive that often require 3D data at video frame rate. 3D TOF cameras are becoming popular for automotive applications such as parking assistance (Scheunert et al., 2007), collision avoidance (Vacek et al., 2007), obstacle detection (Bostelman et al., 2005) as well as the key task of ground plane estimation for on-road obstacle and obstruction avoidance algorithms (Meier & Ade, 1998; Fardi et al., 2006).

The task of obstacle avoidance has normally been approached as by either (a) directly detecting obstacles (or vehicles) and pedestrian or (b) estimating ground plane and locating obstacles from the road geometry. Ground plane estimation has been tackled using methods such as least squares (Meier & Ade, 1998), partial weighted eigen methods (Wang et al., 2001), Hough Transforms (Kim & Medioni, 2007), and Expectation Maximization (Liu et al., 2001), amongst others. Computationally expensive semantic or scene constraint approaches (Cantzler et al., 2002; Nüchter et al., 2003) have also been used for segmenting planar features. However, these methods work well for dense 3D point clouds and are appropriate for laser range data. A statistical framework of RANdom SAmple Concensus (RANSAC) for segmentation and robust model fitting using range data is also discussed in literature (Bolles & Fischler, 1981). Existing work in applying RANSAC to 3D data for plane fitting uses single frame of data (Bartoli, 2001; Hongsheng & Negahdaripour, 2004) or tracking of data points (Yang et al., 2006), and does not exploit the temporal aspect of 3D video data.

In this work, we have formulated a spatio-temporal RANSAC algorithm for ground plane estimation using 3D video data. The TOF camera/sensor provides 3D spatial data at video frame rate and is recorded as a video stream. We model a planar 3D feature comprising two spatial directions and one temporal direction in 4D. We consider a linear motion model for the camera. In order that the resulting feature is planar in the full spatio-temporal representation, we require that the camera rotation lies in the normal to the ground plane, an assumption that is naturally satisfied for the automotive application considered. A minimal set of data consisting of four points is chosen randomly amongst the spatio-temporal data points. From these points, three independent vector directions, lying in the spatio-temporal planar feature are computed. A model for the 3D planar feature is obtained by computing the 4D cross product of the vector directions. The resulting model is scored in the standard manner of RANSAC algorithm and the best model is used to identify inlier and outlier points. The final planar model is obtained as a Maximum likelihood (ML) estimation derived from inlier data where the noise is assumed to be Gaussian. By utilizing data from a sequence of temporally separated image frames, the algorithm robustly identifies the ground plane even when the ground plane is mostly obscured by passing pedestrians or cars and in the presence of walls (hazardous planar surfaces) and other obstructions. The fast segmentation of the obstacles

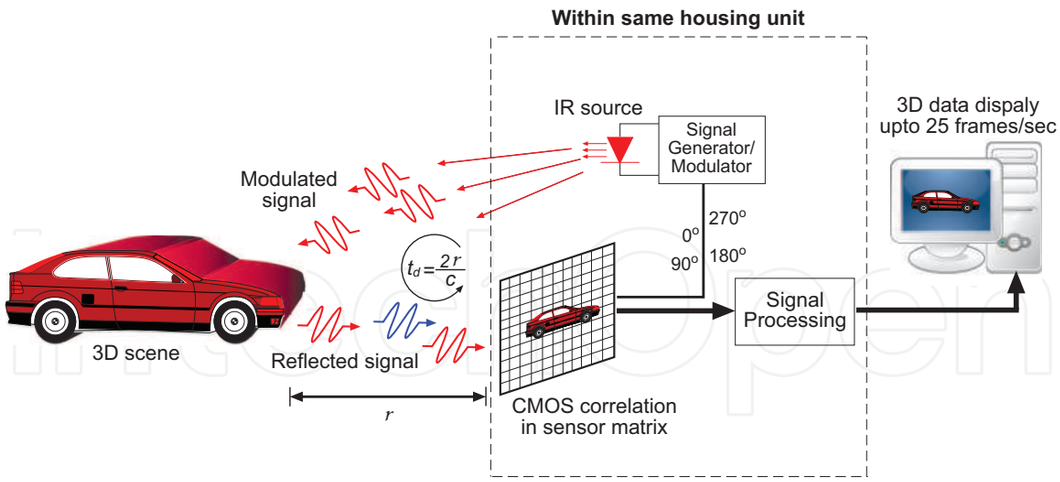


Fig. 2. Basic principle of TOF 3D imaging system.

is achieved using the statistical distribution of the feature and then employing a statistical threshold. The proposed algorithm is simple as no spatio-temporal tracking of data points is required. It is computationally inexpensive without the need of image/feature selection, calibration or scene constraint and is easy to implement in fewest possible steps.
This chapter is organized as follows: Section 2 describes the time-of-flight camera/sensor technology, Section 3 presents the structure and motion model constraints for planar feature, Section 4 describes formulation of spatio-temporal RANSAC algorithm, Section 5 describes application of the framework and Section 6 presents experimental results and discussion, followed by conclusion in Section 7.

2. Time-of-flight camera

Time-of-Flight (TOF) sensors estimate distance to a target using the time of flight of a modulated infrared (IR) wave between the sender and the receiver (see Fig. 2). The sensor illuminates the scene with a modulated infrared waveform that is reflected back by the objects and a CMOS (Complementary metal-oxide- semiconductor) based lock in CCD (charge-coupled device) sensor samples four times per period. With the precise knowledge of speed of light c, each of these (64×48) smart pixels, known as Photonic Mixer Devices (PMD) (Xu et al., 1998), measure four samples a_0, a_1, a_2, a_3 at quarter wavelength intervals. The phase φ of the reflected wave is computed by (Spirig et al., 1995)

$$\varphi = \arctan \frac{a_0 - a_2}{a_1 - a_3}.$$

The amplitude A (of reflected IR light) and the intensity B representing the gray scale image returned by the sensor are respectively given by

$$A = \frac{\sqrt{(a_0 - a_2)^2 + (a_1 - a_3)^2}}{2}, \quad B = \frac{a_0 + a_1 + a_2 + a_3}{4}.$$

With measured phase φ, known modulation frequency f_{mod} and precise knowledge of speed of light c it is possible to measure the un-ambiguous distance r from the camera,

$$r = \frac{c.\varphi}{4\pi f_{\text{mod}}}. \tag{1}$$

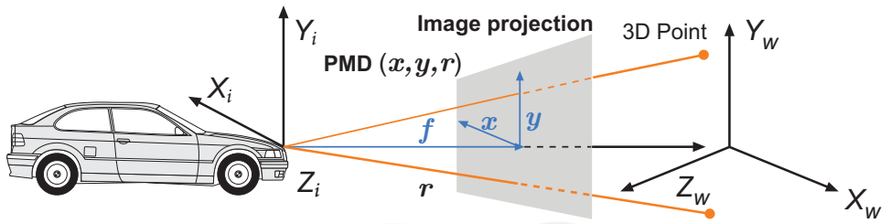


Fig. 3. Time-of-Flight sensor geometry

With a modulation wavelength of λ_{mod}, this leads to a maximum possible unambiguous range of $(\lambda_{\text{mod}}/2)$. For a typical camera such as PMD 3k-S (PMD, 2002), f_{mod}=20Mhz and with a speed of light c given by 3×10^8 m/s, the non-ambiguous range r_{max} of the TOF camera is given as

$$r_{\text{max}} = \frac{c}{2f_{\text{mod}}} = \frac{3 \times 10^8}{2 \cdot 20 \times 10^6} = 7.5\text{meters}.$$

The sensor returns a range r value for each pixel as a function of pixel coordinates (x,y) as shown in Fig. 3.
The range values are used to compute 3D position $\mathbf{X}(X,Y,Z)$ of the point

$$Z = r(x,y).\frac{f}{\sqrt{f^2+x^2+y^2}};\;\; X = Z\frac{x}{f};\;\; Y = Z\frac{y}{f}, \tag{2}$$

where f is the focal length of the camera.

3. Structure and motion constraints

In the following section we will discuss the motion model and the planar feature parameters essential to derive the spatio-temporal RANSAC formulation for a planar feature.

3.1 Motion model

Consider a TOF camera moving in space. Let $\{i\}$ denote the frame of reference at time stamp $i, 1 \leq i \leq n$, attached to the camera. Let $\{W\}$ denote the fixed world reference frame. The rigid body transformation

$${}^W_i\mathcal{M} : \mathbb{R}^3 \rightarrow \mathbb{R}^3; \qquad \mathbf{X}_i \mapsto \mathbf{X}_W := {}^W_iR\mathbf{X}_i + {}^WT_i \tag{3}$$

is defined as the coordinate mapping from frame $\{i\}$ to world frame $\{W\}$ with rotation $({}^W_iR)$ and translation $({}^WT_i)$ respectively. Let $\bar{\mathbf{X}} \in \mathbb{R}^4$ denote the homogenous coordinates of $\mathbf{X} \in \mathbb{R}^3$, then the transformation (3) in matrix form is given by

$${}^W_i\bar{\mathcal{M}} : \mathbb{R}^4 \rightarrow \mathbb{R}^4; \tag{4}$$

$$\bar{\mathbf{X}}_W = \begin{bmatrix} \mathbf{X}_W \\ 1 \end{bmatrix} = \begin{bmatrix} {}^W_iR & {}^WT_i \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{X}_i \\ 1 \end{bmatrix} = {}^W_i\bar{\mathcal{M}}\bar{\mathbf{X}}_i. \tag{5}$$

Let ${}^i_j\bar{\mathcal{M}}$ be the rigid body mapping from frame $\{j\}$ to frame $\{i\}$ then,

$${}^i_j\bar{\mathcal{M}} \;=\; {}^i_W\bar{\mathcal{M}}\;{}^W_j\bar{\mathcal{M}} \;=\; ({}^W_i\bar{\mathcal{M}})^{-1}{}^W_j\bar{\mathcal{M}}.$$

Hence

$${}^i_j\bar{\mathcal{M}} = \begin{bmatrix} ({}^W_iR^\top)({}^W_jR) & ({}^W_iR^\top)({}^WT_j - {}^WT_i) \\ 0 & 1 \end{bmatrix}. \tag{6}$$

3.2 Equation of planar feature with linear motion

Let P be a 2D planar feature that is stationary during the video sequence considered. Let $\eta_i \in \{i\}$ be the normal vector to P in frame $\{i\}$, then η_i is a direction that transforms between frames of reference as

$$\eta_i = {}^j_i R^\top \eta_j = \quad {}^i_j R \eta_j. \tag{7}$$

The homogenous coordinates of a direction (free vector) such as η_i are given by

$$\bar{\eta}_i = \begin{bmatrix} \eta_i \\ 0 \end{bmatrix} \in \mathbb{R}^4, \ {}^j_i\bar{\mathcal{M}}\bar{\eta}_i = {}^j_i\bar{\mathcal{M}} \begin{bmatrix} \eta_i \\ 0 \end{bmatrix} = \begin{bmatrix} {}^j_i R \eta_i \\ 0 \end{bmatrix} = \bar{\eta}_j. \tag{8}$$

Let $\boldsymbol{X}_i, \boldsymbol{X}_j \in P$ be different elements of the planar feature P observed in different frames $\{i\}$ and $\{j\}$. Note that $\boldsymbol{X}_i \neq {}^i_j\mathcal{M}\boldsymbol{X}_j$ in general as the points do not correspond to the same physical point in the plane, however, $(\boldsymbol{X}_i, {}^i_j\mathcal{M}\boldsymbol{X}_j)$ must both lie in P in $\{i\}$. Since η_i is a normal to P in $\{i\}$, one has

$$\langle (\bar{\boldsymbol{X}}_i - {}^i_j\mathcal{M}\bar{\boldsymbol{X}}_j), \bar{\eta}_i \rangle = 0. \tag{9}$$

Thus

$$\begin{aligned}
&\left\langle \begin{bmatrix} \boldsymbol{X}_i \\ 1 \end{bmatrix} - \begin{bmatrix} ({}^W_i R^\top)({}^W_j R)\boldsymbol{X}_j + ({}^W_i R^\top)({}^W T_j - {}^W T_i) \\ 1 \end{bmatrix}, \begin{bmatrix} \eta_i \\ 0 \end{bmatrix} \right\rangle = 0 \\
&\left\langle \boldsymbol{X}_i - ({}^W_i R^\top)({}^W_j R)\boldsymbol{X}_j - ({}^W_i R^\top)({}^W T_j - {}^W T_i), \eta_i \right\rangle = 0 \\
&\left\langle \boldsymbol{X}_i - ({}^W_i R^\top)({}^W_j R)\boldsymbol{X}_j, \eta_i \right\rangle - \left\langle ({}^W_i R^\top)({}^W T_j - {}^W T_i), \eta_i \right\rangle = 0 \\
&\left\langle \boldsymbol{X}_i - \boldsymbol{X}_j, ({}^W_j R)({}^W_i R^\top)\eta_i \right\rangle - \left\langle ({}^W T_j - {}^W T_i), ({}^W_i R)\eta_i \right\rangle = 0.
\end{aligned} \tag{10}$$

Let $V \in \{W\}$ denote the linear velocity then the rigid body dynamics for a moving body (an automotive) is modelled by

$$\begin{aligned}
\dot{T} &= V; \quad T(0) = T_1 \\
\dot{R} &= \widehat{\omega} R; \ R(0) = R_1,
\end{aligned} \tag{11}$$

where $\omega \in \{W\}$ is the angular velocity and $\widehat{\omega} \in \mathbb{R}^{3\times 3}$ denote the skew symmetric matrix that corresponds to vector cross product operation in 3D.

Assumption: We assume that the angular velocity ω of the camera is parallel to $\eta \in \{W\}$, the normal to the ground plane at all times and the translation velocity V in the direction normal to the ground plane is constant such that

$$\eta \times \omega = 0 \ \text{ and } \ \langle V, \eta \rangle = \text{constant}, \tag{12}$$

where $\times$ represents a cross product between two vectors. For normal motion of a vehicle, roll and pitch rotations are negligible compared to yaw motion associated with angular velocity of the turning vehicle Gracia et al. (2006) and corresponds to common ground-plane constraint (GPC) Sullivan (1994) (see Figure 4).

In real environments for motion captured at nearly video frame rate, the piecewise linear velocity along the normal direction can be assumed constant as evident from the experiments

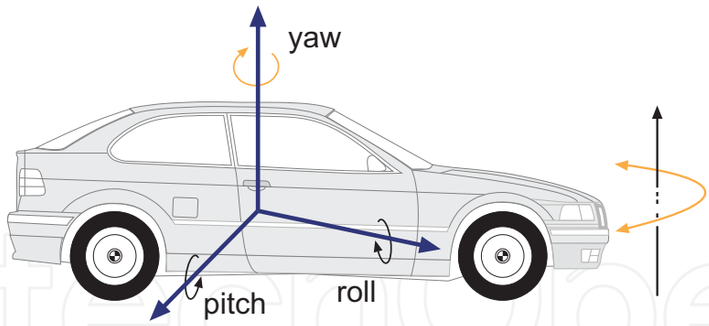


Fig. 4. Vehicle with roll, pitch and dominant yaw motion

in Section 4. This is to be expected in the case where the camera is attached to a vehicle that moves on a plane P, precisely the case for the automotive example considered. In practice, this degree of motion is important to model situations where the car suspension is active and is also used to identify non-ground features that the vehicle may be approaching with constant velocity.
As a consequence of (12)

$$\omega = s(t)\eta \in \{W\};\ s : \mathbb{R} \to \mathbb{R} \ \text{ in time } t. \tag{13}$$

Following (13) one can re-write (11) as

$$\dot{R} = s(t)\widehat{\eta}R; \quad R(0) = R_1.$$

Therefore the continuous rotation motion $R(t) : \mathbb{R} \to SO(3)$ for the automobile trajectory is expressed as

$$R(t) = \exp(\theta(t_i)\widehat{\eta})R_1;\ \ \theta(t_i) = \int_0^{t_i} s(\tau)d\tau \tag{14}$$

where t_i time is at frame $\{i\}$ and

$$^{W}_{i}R = R(t_i).$$

By definition $^{W}_{i}R\eta_i = \eta$ and hence,

$$\begin{aligned} \eta_i &= {}^{W}_{i}R^{\top}\eta \\ &= R_1^{\top}\exp(\theta(t_i)\widehat{\eta})^{\top}\eta \\ &= R_1^{\top}\eta = \eta_1 \end{aligned} \tag{15}$$

Using (15), we can re-write (10) as

$$\left\langle \mathbf{X}_i - \mathbf{X}_j, \eta_1 \right\rangle - \left\langle {}^{W}T_j - {}^{W}T_i, \eta_1 \right\rangle = 0. \tag{16}$$

We assume the frames are taken at constant time interval δt and hence $t_i = \delta t(i-1) + t_1$. Since $\langle V, \eta \rangle$ is constant and $t_1 = 0$, the linear translation motion $^{W}T_i$ satisfies

$$\left\langle {}^{W}T_i, \eta_i \right\rangle = \langle V, \eta \rangle\, \delta t(i-1) + \langle T_1, \eta_1 \rangle\,. \tag{17}$$

Using assumption (12), define $\alpha \in \mathbb{R}$ to be

$$\alpha = \langle V, \eta \rangle\, \delta t = \text{constant}. \tag{18}$$

Thus, from (16) and (17), the structure and motion constraint that $\mathbf{X}_i, \mathbf{X}_j$ lie in the plane P can be expressed as

$$\left\langle \mathbf{X}_i - \mathbf{X}_j, \eta_1 \right\rangle - \alpha(j-i) = 0. \tag{19}$$

This is an equation for a plane P parameterized by $\eta_1 \in S^2$ $(\|\eta_1\| = 1)$ and motion parameter $\alpha \in \mathbb{R}$. An additional parameter, the distance $h \in \mathbb{R}$ of the plane P from the origin in frame $\{1\}$ in the direction η_1, completes the structure and motion constraints of planar feature. Note that α is the component of translational camera velocity in the direction normal to the planar feature P. The component α will be the defining parameter for the temporal component of the 3D planar feature that is identified in the RANSAC algorithm (see Section 4).
Let $\bar{\bar{\mathbf{X}}}_i$ be a 4D spatio-temporal coordinate that incorporates both spatial coordinates $\mathbf{X}_i$ and a reference to the frame index or time coordinates i

$$\bar{\bar{\mathbf{X}}}_i = \begin{bmatrix} \mathbf{X}_i \\ i \end{bmatrix}. \tag{20}$$

Associated with this we define a normal vector that incorporates the spatial normal direction η_1 and the motion parameter α

$$\bar{\bar{\eta}} = \begin{bmatrix} \eta_1 \\ \alpha \end{bmatrix}. \tag{21}$$

Using these definitions (19) may be re-written as

$$\left\langle \bar{\bar{\mathbf{X}}}_i - \bar{\bar{\mathbf{X}}}_j, \bar{\bar{\eta}} \right\rangle = 0. \tag{22}$$

4. Spatio-temporal RANSAC algorithm

In this section we present the spatio-temporal RANSAC algorithm and compute a 3D spatio-temporal planar hypothesis based on the structure and motion model derived in Section 3.2 and a minimal data set.

4.1 Computing a spatio-temporal planar hypothesis

Equation (19) provides a constraint that $(\bar{\bar{\mathbf{X}}}_i - \bar{\bar{\mathbf{X}}}_j) \in \mathbb{R}^4$ lies in the 3D spatio-temporal planar feature P in $\mathbb{R}^4$ with parameters $\eta_1 \in S^2$, $\alpha \in \mathbb{R}$ and $h \in \mathbb{R}$. Given a sample of four points $\{\bar{\bar{\mathbf{X}}}_{i_1}, \bar{\bar{\mathbf{X}}}_{i_2}, \bar{\bar{\mathbf{X}}}_{i_3}, \bar{\bar{\mathbf{X}}}_{i_4}\}$, one can construct a normal vector $\bar{\bar{\eta}}$ to P by taking the 4D cross product (see Appendix A)

$$\bar{\bar{\eta}}_o = \mathbf{cross}_4(\bar{\bar{\mathbf{X}}}_{i_1} - \bar{\bar{\mathbf{X}}}_{i_2}, \bar{\bar{\mathbf{X}}}_{i_1} - \bar{\bar{\mathbf{X}}}_{i_3}, \bar{\bar{\mathbf{X}}}_{i_1} - \bar{\bar{\mathbf{X}}}_{i_4}) \in \mathbb{R}^4, \tag{23}$$

where $\bar{\bar{\mathbf{X}}}_i \in \{\{1\}, \ldots, \{n\}\}$. To apply the constraint $\eta_1 \in S^2$ we normalize $\bar{\bar{\eta}}_o = (\bar{\bar{\eta}}_o^x, \bar{\bar{\eta}}_o^y, \bar{\bar{\eta}}_o^z, \bar{\bar{\eta}}_o^t)$ by

$$\bar{\bar{\eta}} = \frac{1}{\beta}\bar{\bar{\eta}}_o; \qquad \beta = \sqrt{(\bar{\bar{\eta}}_o^x)^2 + (\bar{\bar{\eta}}_o^y)^2 + (\bar{\bar{\eta}}_o^z)^2}. \tag{24}$$

The resulting estimate $\bar{\bar{\eta}} = (\eta_1, \alpha)$ is an estimate of the normal $\eta_1 \in S^2$ and α, the normal vector component of translation velocity (18).
Note that the depth parameter h can be determined by

$$h_1 = \langle \mathbf{X}_i, \eta_1 \rangle - \alpha(i-1). \tag{25}$$

However, the parameter h is not required for the robust estimation phase of the RANSAC algorithm and is evaluated in the second phase where a refined model is estimated.

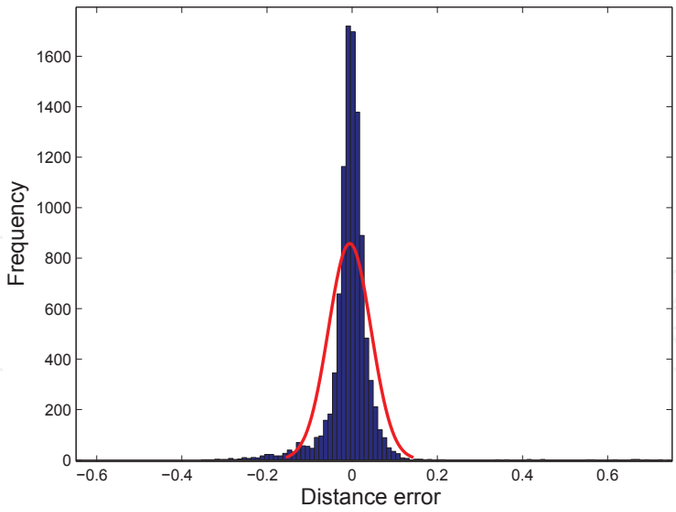


Fig. 5. Statistical distribution of planar feature data points derived from experimental data documented in Section 6.

4.2 Statistical distribution of 4D data points

The spatio-temporal data points that have a probability p of lying in the planar feature are defined as inliers. Due to Gaussian noise in range measurements of TOF camera, the distance of these inliers from the model (planar feature) have a Gaussian distribution with $\mathcal{N}(0,\sigma)$ as shown in Fig. 5.
As a consequence, the point square distance $a_{\perp}^2$,

$$a_{\perp}^2 = (\langle(\bar{\bar{\boldsymbol{X}}} - \bar{\bar{\boldsymbol{X}}}_{i_1}), \bar{\bar{\eta}}\rangle)^2; \;\; \bar{\bar{\boldsymbol{X}}} \in \text{all spatio-temporal data points},$$

of the inliers (Hartley & Zisserman, 2003) from the planar feature associated with the data point $\bar{\bar{\boldsymbol{X}}}_i$, have a chi-squared distribution χ^2. Since we consider a spatio-temporal planar feature, there are three degrees of freedom in the chi-squared distribution. Let $F_{\chi_3^2}$ denote the cumulative frequency of three degree of freedom of chi-squared distribution χ_3^2 then one can define the threshold coefficient q^2 by

$$q^2 = F_{\chi_3^2}^{-1}(p)\sigma^2. \tag{26}$$

Thus, the statistical test for inliers is defined by

$$\begin{cases} \text{inliers} & a_{\perp}^2 < q^2 \\ \text{outliers} & a_{\perp}^2 \geq q^2. \end{cases} \tag{27}$$

In the experiments documented in Section 6, we use a value of $p = 0.95$. In this case the threshold is $q^2 = 7.81\sigma^2$ where σ is determined empirically. Spatial ground plane estimation algorithms using single 3D images (Cantzler et al., 2002; Bartoli, 2001; Hongsheng & Negahdaripour, 2004) are associated with two degree of chi-squared distribution since they lack temporal dimension. As a result the same analysis leads to a threshold of $q^2 = 5.99\sigma^2$ (for $p = 0.95$). The additional threshold margin for the proposed spatio-temporal algorithm quantifies the added robustness that comes from incorporating the temporal dimension along with the data available by incorporating multiple images from the

video stream. This leads to significant improvement in robustness and performance of the proposed algorithm over single image techniques. The resulting spatio-temporal RANSAC algorithm is outlined in Algorithm 1.

5. Application

The planar feature estimation algorithm in 4D is an approach that can be utilized in multiple scenarios with reference to automotive domain. Since the dominating planar feature for an automotive is a road, we have presented an application of the proposed algorithm for robust ground plane estimation and detection.
A constant normal velocity component α (18) helps to detect ground plane due to the fact that piecewise linear velocity in the normal direction of the automotive motion is small and constant over the number of frames recorded at frame rate. Detection of ground plane in spatio-temporal domain provides an added advantage for cases where there is occlusion and single frame detection is not possible. Section 6 presents number of examples for ground plane.

Algorithm 1: Pseudo code Spatio-temporal RANSAC algorithm

Initialization: Choose a probability p of inliers. Initialize a sample count $m = 0$ and the trial process $N = \infty$.

repeat

a. Select at random, 4 spatio-temporal points $(\bar{\bar{\boldsymbol{X}}}_{i_1}, \bar{\bar{\boldsymbol{X}}}_{i_2}, \bar{\bar{\boldsymbol{X}}}_{i_3}, \bar{\bar{\boldsymbol{X}}}_{i_4})$.

b. Compute the temporal normal vector $\bar{\bar{\eta}}$ according to (23) and (24).

c. Evaluate the spatio-temporal constraint (22) to develop a consensus set C_m consisting of all data points classified as inliers according to (27).

d. Update N to estimate the number of trials required to have a probability p so that the selected random sample of 4 points is free from outliers as (Fischler & Bolles, 1981),

$$N = \log(1-p)/\log\left(1 - \frac{\text{number of inliers}}{\text{number of points}}\right)^4.$$

until *at least N trials are complete*
Select the consensus set C_m^* that has the most inliers.
Optimize the solution by re-estimating from all spatio-temporal data points in C_m^* by maximizing the likelihood of the function ϕ

$$\phi(\bar{\bar{\eta}}, h) = \sum_{\bar{\bar{\boldsymbol{X}}} \in C_m^*} (\langle \bar{\bar{\eta}}, \bar{\bar{\boldsymbol{X}}} \rangle - h)^2 \tag{28}$$

$$\mathcal{L}(\phi) = \prod_{\bar{\bar{\boldsymbol{X}}} \in C_m^*} \phi(\bar{\bar{\boldsymbol{X}}} | \bar{\bar{\eta}}, h); \qquad (\hat{\bar{\bar{\eta}}}, \hat{h}) = \arg\max_{\bar{\bar{\eta}}, h}(\mathcal{L}),$$

where we assume a normal distribution in observed depth.

An obstacle detection algorithm can be applied once a robust estimation of planar ground surface is available. In the proposed framework, the algorithm evaluates each spatio-temporal data point and categorizes traversable and non-traversable objects or obstacles. Traversable objects are the points that can be comfortably driven over in a vehicle. We are inspired by a similar method proposed in (Fornland, 1995). The estimated Euclidean distance $\hat{d}$ to the plane for an arbitrary data point $\bar{\bar{\boldsymbol{X}}}$ is defined as

$$\hat{d} = \langle \bar{\bar{\boldsymbol{X}}}, \hat{\bar{\bar{\eta}}} \rangle - \hat{h}. \tag{29}$$

Objects (in each frame) are segmented from the ground plane by a threshold τ as

$$\bar{\bar{\boldsymbol{X}}} = \begin{cases} \text{Obstacle} & |\hat{d}| \geq \tau_o \\ \text{Traversable object} & |\hat{d}| < \tau_o, \end{cases} \tag{30}$$

where τ_o is set by the user for the application under consideration. This threshold segmentation helps in reliable segregation of potential obstacles. The allowance of larger threshold in inliers for plane estimation makes obstacle detection phase robust for various applications especially for on road obstacle detection.

6. Experimental results and discussions

Experiments were performed using real video data recorded from PMD 3k-S TOF camera mounted on a vehicle with an angle varying between 2° to 20° to the ground. The camera records at approx 20 fps and provides both gray scale and range images in real time. The sensor has a field of view of $33.4° \times 43.6°$. The video sequences depict scenarios in an under cover car park. In particular, we consider cases with pedestrians, close by vehicles, obstacles, curbs/foothpaths and walls etc. Five experimental scenarios have been presented to evaluate the robustness of the algorithm against real objects and also compared with standard 3D RANSAC algorithm. The gray scale images shown represent the first and the last frame of video data. It is not possible to have a 4D visualization environment, therefore a 3D multi-frame representation (each data frame represented in different color) provides a spatio-temporal video range data. The estimated spatio-temporal planar feature is represented in frame $\{1\}$. The final solution is rotated for better visualisation.

In the first set of experiments shown in Fig. 6 and Table. 1(sequence 1-4), four different scenarios are presented. The first scenario shows multiple walls at varying level of depth and a ground plane. The algorithm correctly picks the ground plane rejecting other planar features. In the next scenario, a truck in close vicinity is obstructing the clear view but the ground plane has been identified by exploiting the full video sequence of the data. A number of obstacles including cars, wall and a person are visible while the car is manoeuvring a turn in the third scenario. The algorithm clearly estimates actual ground plane. In the fourth scenario the result is not perturbed by passing pedestrians and the algorithm robustly identifies the ground plane. In a typical sequence a 8-10 frame data is enough to resolve a ground plane even in the presence of some kind of occlusion.

In another experiment shown in Fig. 7a (sequence 5 with single frame data), the standard RANSAC algorithm is applied using a single frame data for comparison.

The obvious failure of a standard RANSAC algorithm is due to the bias of planar data points towards the wall. On the other hand, the proposed algorithm has correctly identified the ground surface in Fig. 7b by simply incorporating more frames (10 frames and $|\alpha| = 0.0018$) due to the availability of temporal data without imposing any scene constraint.

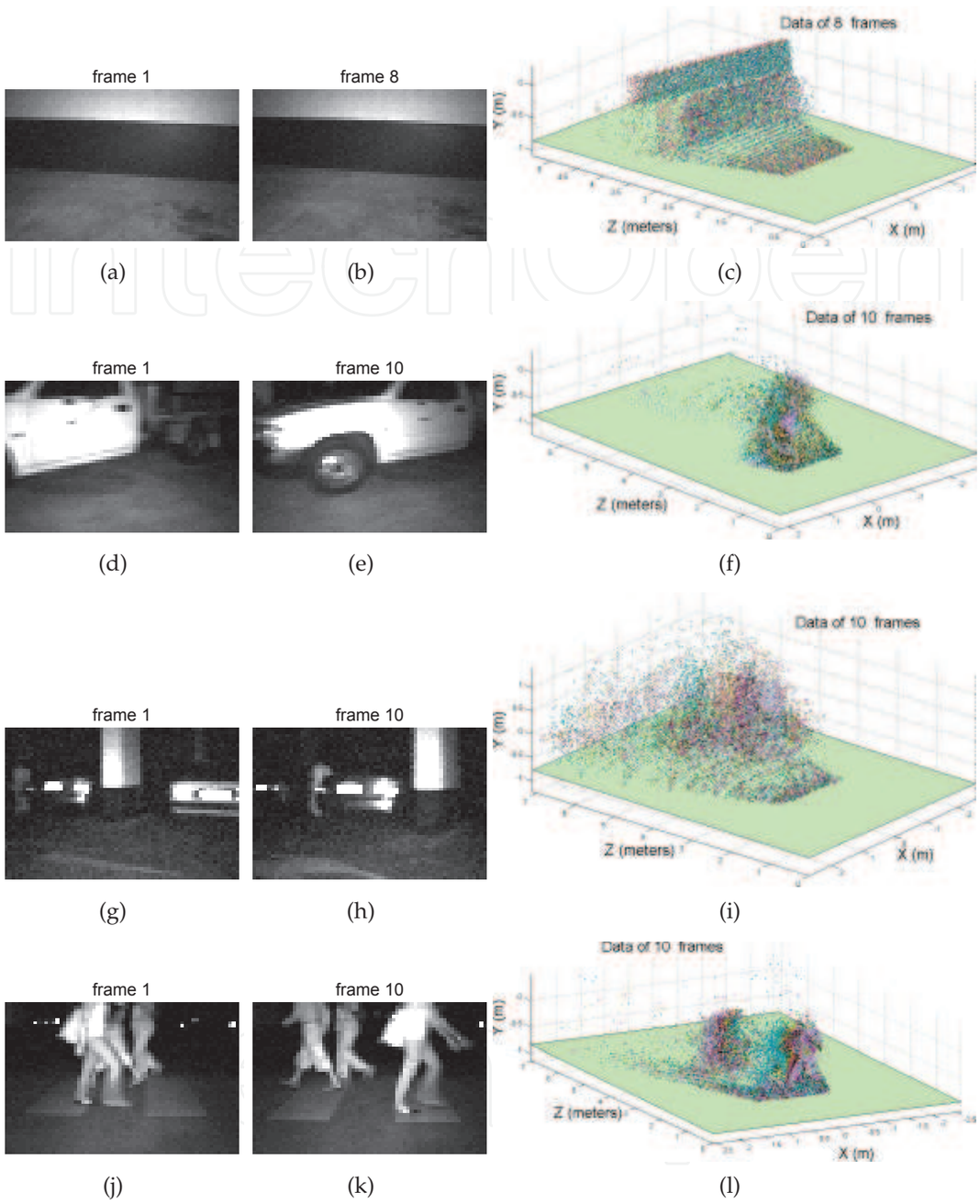
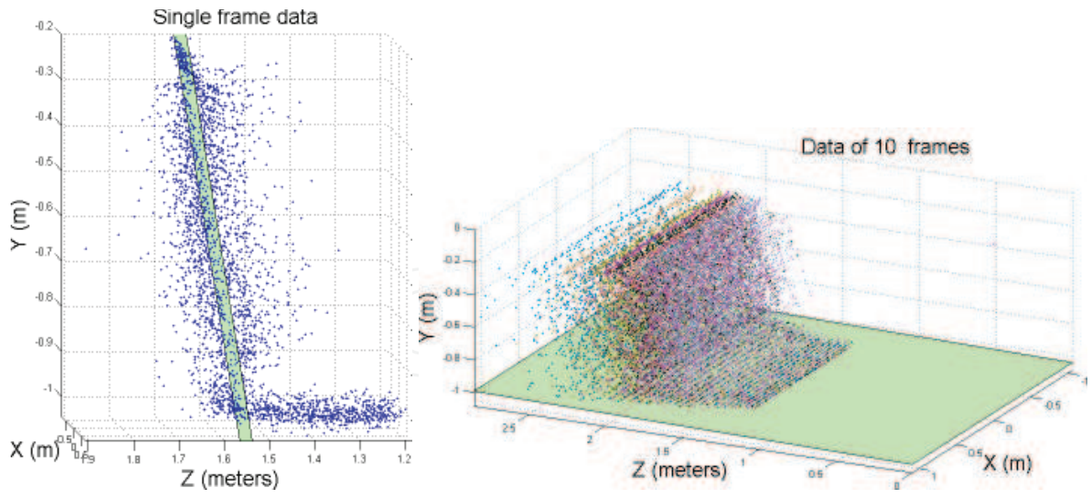


Fig. 6. Experimental data shown in a three column format. First two columns show gray scale image of first and last video frame and third column shows spatio-temporal fit on 4D data. Each frame of 3D data is represented by a different color. (a-b) Gray scale images of a double wall and ground plane at turning (c) Spatio-temporal ground plane fitting of 8 frames at t=1. (d-e) A truck in close vicinity (f) Corresponding spatio-temporal ground plane fit of 10 frames. (g-h) Cars, wall and a person as obstacles at turning. (i) Corresponding spatio-temporal ground plane fit. (j-k) Pedestrians. (l) Ground plane fit.



(a) (b)

Fig. 7. Using data from sequence 5, (a) Standard RANSAC plane fitting algorithm picks the wall with a single frame data. (b) Spatio-temporal RANSAC algorithm picks the correct ground plane (10 frames).

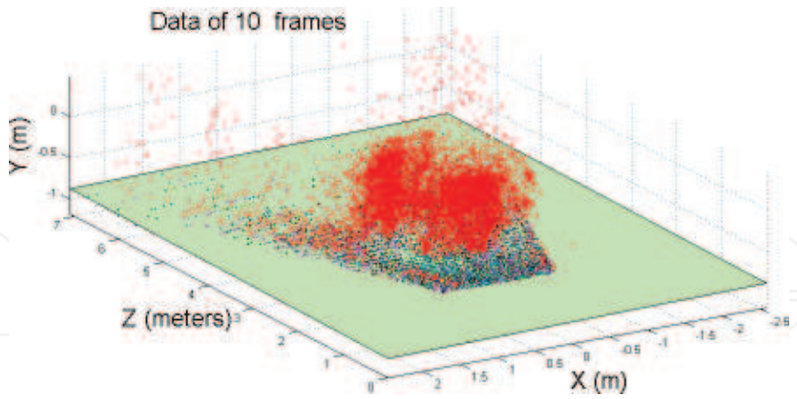
Obstacle detection algorithm is effectively applied after robust estimation of ground plane. In the experiment shown in Fig. 8, pedestrians are segmented with $\tau_o = 0.1$ by the obstacle detection algorithm after correct identification of ground plane. This threshold implies that objects with a height greater than 10 cm (shown in red color) are considered as obstacle where data points close to ground plane are ignored (traversable objects) with this threshold.
The experimental results are straightforward and show excellent performance. The proposed 4D spatio-temporal RANSAC algorithm's computation cost is associated with picking the normal vector to the 3D planar feature by random sampling (please note that this is the only computation cost associated with 4D spatio-temporal RANSAC algorithm). This eliminates any computation cost associated with pre-processing images unlike conventional algorithms. The experiments were performed on a PC machine with Intel Core 2 Duo 3GHz processor and 2 GB RAM. The algorithm is implemented in MATLAB. The computation cost varies with the number of inliers and the planar surface occlusion in the range data as shown in Fig. 9.

7. Conclusion

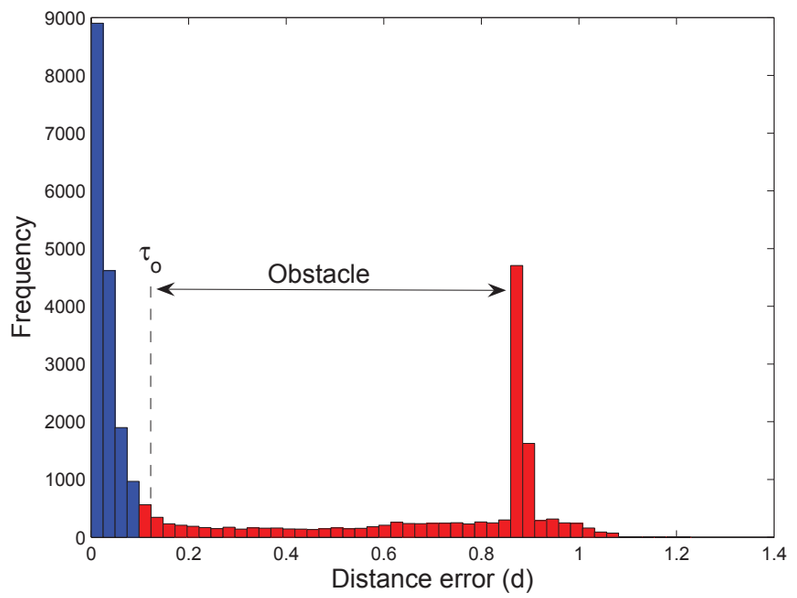
Many vision based applications use some kind of segmentation and planar surface detection as a preliminary step. In this paper we have presented a robust spatio-temporal RANSAC framework for ground plane detection for use in ADAS of automotive industry. Experimental

Sequence no	Sequence	No of frames used	$\lvert\alpha\rvert$ (m/frame)
1	Double Wall	8	0.0016
2	Moving truck	10	0.0017
3	Multiple objects	10	0.0021
4	Pedestrian	10	0.0020
5	Front wall	10	0.0018

Table 1. Experimental data for ground plane estimation



(a)



(b)

Fig. 8. (a) Potential obstacles and Pedestrians are shown in red color. (b) Histogram of ground plane and obstacles.

results validate the structure and motion model of a 3D spatio-temporal planar feature in 4D. Since the algorithm does not involve any tracking or feature selection, it is highly robust, simple and practical to implement. The algorithm is suitable not only for automotive industry but also in general computer vision applications that satisfy the particular motion constraint ($\eta \times \omega = 0$). This constraint ensures that a spatial planar feature generates a planar feature in spatio-temporal domain. The spatio-temporal constraints increases reliability in planar surface estimation that is otherwise susceptible to noisy data in any algorithm developing a single frame data. Further improvement in computation cost can be achieved through dedicated hardware implementation.

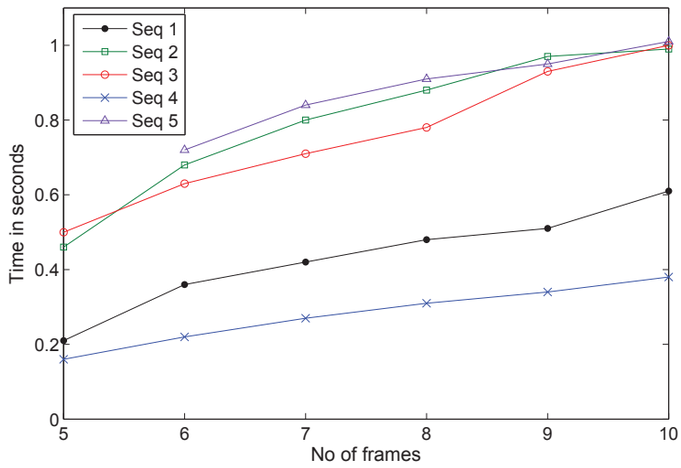


Fig. 9. Performance plots for Spatio-temporal RANSAC for all the sequences.

8. Appendix A

Given an orthonormal basis $\{\mathbf{e}_1,\ldots,\mathbf{e}_n\} \in \mathbb{R}^n$ the 'Levi-Civita' (ε) antisymmetric tensor is defined as (Shaw, 1987)

$$\varepsilon_{i,j,\ldots,n} = \begin{cases} +1 & \text{if } (i,j,\ldots,n) \text{ an even permutation of } (1,2,\ldots,n) \\ -1 & \text{if } (i,j,\ldots,n) \text{ an odd permutation of } (1,2,\ldots,n) \\ 0 & \text{if } (i,j,\ldots,n) \text{ not a permutation of } (1,2,\ldots,n) \end{cases}$$

The cross product of three vectors $\mathbf{a},\mathbf{b},\mathbf{c} \in \mathbb{R}^4$ is defined as

$$\mathbf{cross}_4(\mathbf{a},\mathbf{b},\mathbf{c}) = (\mathbf{a}\times\mathbf{b}\times\mathbf{c}) = \sum_{i,j,k,l=1}^{n=4} \varepsilon_{ijkl} a_j b_k c_l \mathbf{e}_i \tag{31}$$

The vector cross product of the three vectors in $\mathbb{R}^4$ has the following properties (amongst others).

1. *Trilinearity:* For $\alpha,\beta,\gamma \in \mathbb{R}$, $\alpha\mathbf{a}\times\beta\mathbf{b}\times\gamma\mathbf{c} = \alpha\beta\gamma(\mathbf{a}\times\mathbf{b}\times\mathbf{c})$.
2. *Linear dependence:* $\mathbf{cross}_4(\mathbf{a},\mathbf{b},\mathbf{c}) = 0$ iff $\langle\mathbf{a},\mathbf{b},\mathbf{c}\rangle$ are linearly dependent.
3. *Orthogonality:* Let $\mathbf{d} = \mathbf{a}\times\mathbf{b}\times\mathbf{c} \Rightarrow \langle\mathbf{d},\mathbf{a}\rangle = \langle\mathbf{d},\mathbf{b}\rangle = \langle\mathbf{d},\mathbf{c}\rangle = 0$

9. References

Baker, C. & Dolan, J. (2008). Traffic interaction in the urban challenge: Putting boss on its best behavior, *Proc. International Conference on Intelligent Robots and Systems (IROS 2008)*, pp. 1752–1758.

Bartoli, A. (2001). Piecewise planar segmentation for automatic scene modeling, *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition (CVPR '01)*.

Bolles, R. C. & Fischler, M. A. (1981). A RANSAC-based approach to model fitting and its application to finding cylinders in range data, *Proc. Seventh Int. Joint Conf. Artificial Intelligence*, pp. 637–643.

Bostelman, R., Hong, T. & Madhavan, R. (2005). Towards AGV safety and navigation advancement obstacle detection using a TOF range camera, *Proc. 12th Int. Conf. Advanced Robotics (ICAR '05)*.

Cantzler, H., Fisher, R. B. & Devy, M. (2002). Improving architectural 3D reconstruction by plane and edge constraining, *Proc. British Machine Vision Conf. (BMVC '02)*, pp. 43–52.

Commission, E. (2001). European transport policy for 2010: time to decide (white paper), *Technical Report COM(2001) 370 final*, Commission of the European Communities, Brussels.

Commission, E. (2008). CARS 21 mid-term review high level conference conclusions and report, *Technical report*, European Commission, Enterprise and Industry.

Fardi, B., Dousa, J., Wanielik, G., Elias, B. & Barke, A. (2006). Obstacle detection and pedestrian recognition using a 3D PMD camera, *Proc. IEEE Intell. Vehicles Symp.*, pp. 225–230.

Fischler, M. A. & Bolles, R. C. (1981). Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography, *Communications of the ACM* 24, Issue 6 June 1981: 381–395.

Fornland, P. (1995). Direct obstacle detection and motion from spatio-temporal derivatives, *Proc. 6th Intl. Conf. Comp. Anal. of Images and Patterns*, pp. 874–879.

Gern, A., Franke, U. & Levi, P. (2000). Advanced lane recognition-fusing vision and radar, *Proc. IEEE Intelligent Vehicles Symposium IV 2000*, pp. 45–51.

Gracia, G. A., Jimenez, F., Paez, J. & Narvaez, A. (2006). Theoretical and experimental analysis to determine the influence of the ageing process of the shock-absorber on safety, *Int. J. Vehicle Design* 40(1/2/3): 15–35.

Hartley, R. & Zisserman, A. (2003). *Multiple View Geometry in Computer Vision*, Cambridge Universty Press, chapter Estimatio-2D Projective Transformation. pp. 118.

Hongsheng, Z. & Negahdaripour, S. (2004). Improved temporal correspondences in stereo-vision by RANSAC, *Proc. 17th Int. Conf. Pattern Recognition (ICPR '04).*, Vol. 4, pp. 52– 55.

Kim, E. & Medioni, G. Lee, S. (2007). Planar patch based 3D environment modeling with stereo camera, *Proc. 16th IEEE Int. Symp. Robot and Human Interactive Communication*, Jeju island, Korea.

Liu, Y., R., E., Charabarti, D., Burgard, W. & Thrun, S. (2001). Using EM to learn 3D models of indoor environments with mobile robots, *Proc. 18th Int. Conf. Machine Learning*, pp. 329–336.

Meier, E. B. & Ade, F. (1998). Object detection and tracking in range image sequences by separation of image features, *Proc. IEEE Int. Conf. Intelligent Vehicles*.

Nüchter, A., Surmann, H. & Hertzberg, J. (2003). Automatic model refinement for 3D reconstruction with mobile robots, *Proc. Fourth Int. Conf. 3-D Digital Imaging and Modeling (3DIM '03)*, pp. 394– 401.

Organization, W. H. (2009). Global status report on road safety: Time for action, *Technical report*, World Health Organization.

Peden, M., Scurfield, R., Sleet, D., Mohan, D., Hyder, A. A., Jarawan, E. & Mathers, C. (2004). *World Report on Road Traffic Injury Prevention*, World Health Organization(WHO).

PMD (2002). PMD tech., http://www.pmdtec.com.
URL: *http://www.pmdtec.com*

Scheunert, U., Fardi, B., Mattern, N., Wanielik, G. & Keppeler, N. (2007). Free space determination for parking slots using a 3D PMD sensor, *Proc. IEEE Intelligent Vehicles Symposium*, pp. 154–159.

Sethi, D. (2008). Road traffic injuries among vulnerable road users.

Shaw, R. (1987). Vector cross products in n dimensions, *Int. J. Math. Educ. Sci. Technol.* 18(6): 803–816.

Spirig, T., Seitz, P., Vietze, O. & Heitger, F. (1995). The lock-in CCD-two dimensional synchronous detection of light, *IEEE J. Quantum Electron.* 31: 1705–1708.

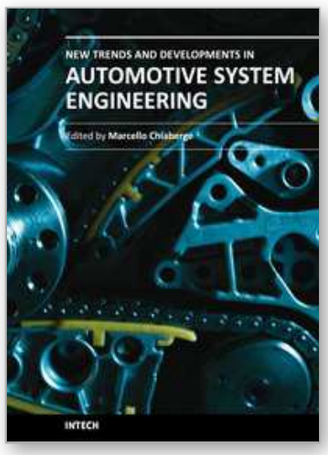
Sullivan, G. D. (1994). *Real-time Computer Vision*, Cambridge University Press, Cambridge, chapter Model-based vision for traffic scenes using the ground plane constraint, pp. 93–115.

Vacek, S., Schamm, T., Schroder, J. & Dillmann, R. (2007). Collision avoidance for cognitive automobiles using a 3D PMD camera, *Proc. 6th IFAC Symp. on Intell. Autonomous Vehicles Symp.*

Wang, C., Tanahashi, H., Hirayu, H., Niwa, Y. & Yamamoto, K. (2001). Comparison of local plane fitting methods for range data, *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR '01)*.

Xu, Z., Schwarte, R., Heinol, H., Buxbaum, B. & Ringbeck, T. (1998). Smart pixel - photonic mixer device (PMD), new system concept of a 3D-imaging camera-on-a-chip, *5th Int. Conf. Mechatronics and Machine Vision in Practice*, Nanjing, pp. 259–264.

Yang, A., Rao, S. & Ma, Y. (2006). Robust statistical estimation and segmentation of multiple subspaces, *Proc. Conference on Computer Vision and Pattern Recognition Workshop CVPRW '06*, pp. 99–99.



New Trends and Developments in Automotive System Engineering
Edited by Prof. Marcello Chiaberge

ISBN 978-953-307-517-4
Hard cover, 664 pages
Publisher InTech
Published online 08, January, 2011
Published in print edition January, 2011

In the last few years the automobile design process is required to become more responsible and responsibly related to environmental needs. Basing the automotive design not only on the appearance, the visual appearance of the vehicle needs to be thought together and deeply integrated with the â€œpowerâ€ developed by the engine. The purpose of this book is to try to present the new technologies development scenario, and not to give any indication about the direction that should be given to the research in this complex and multi-disciplinary challenging field.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following:

Faisal Mufti, Robert Mahony and Jochen Heinzmann (2011). 4D Ground Plane Estimation Algorithm for Advanced Driver Assistance Systems, New Trends and Developments in Automotive System Engineering, Prof. Marcello Chiaberge (Ed.), ISBN: 978-953-307-517-4, InTech, Available from: http://www.intechopen.com/books/new-trends-and-developments-in-automotive-system-engineering/4d-ground-plane-estimation-algorithm-for-advanced-driver-assistance-systems

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen