Version
This is the Accepted Manuscript version. This version is defined in the NISO recommended practice RP-8-2008 http://www.niso.org/publications/rp/

Suggested Reference

Copyright
Items in ResearchSpace are protected by copyright, with all rights reserved, unless otherwise indicated. Previously published items are made available in accordance with the copyright policy of the publisher.

© 2013 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.


https://researchspace.auckland.ac.nz/docs/uoa-docs/rights.htm
Stereo Accuracy for Collision Avoidance for Non-linear Collision Trajectories

Waqar Khan and Reinhard Klette

Abstract—In this study, we have generalized our previous tool for assisting a safety engineer in assessing collision trajectories by extending from colliding objects with constant velocity to more general variable velocity ones. We have also highlighted that a linear system cannot be relied upon for handling a colliding object with variable velocity. To deal with such trajectories, past observations are weighted depending on velocities at those locations; priority is given to locations with reduced velocity. Based on this hypothesis, we have shown that the weighted system outperforms a linear one. The benefit is that it always issues a timely warning, even if the trajectory of the colliding object keeps on changing over time.

I. INTRODUCTION

Stereo vision faces similar limitations as various other computer science applications - an improvement has a certain cost. Accuracy of stereo improves by increasing image resolution but at the cost of an increased disparity range. This range is proportional to the memory required for a real-time (yet accurate) hardware implementation of a stereo matching algorithm like Dynamic Programming stereo (DP), Semi-Global Matching stereo (SGM), Belief Propagation stereo (BP) [10], or Graph Cut stereo (GC); see, for example, [4], [7], [13], [17]. For example, memory needed for BP is \(O(MNd_{\text{max}})\) [15], for images of size \(M \times N\) and \(d_{\text{max}}\) disparity levels.

If the image resolution is increased, the camera sensor captures more detail in the scene but the sensor still contains discrete pixels, hence the disparity values are still integral. Accordingly, the depth is measured at discrete steps [12]. Stereo matching algorithms often generate sub-pixel disparities, but still the disparities remain a discrete representation of a continuous world. In other words, for every measured depth, there is an inherent inaccuracy involved. In a controlled environment with known location of the object of interest, this stereo limitation is significantly reduced by an appropriate off-line stereo configuration. However, when the locations of objects of interest are unknown as in the case of dynamic stereo imaging, then determining the accuracy of estimated depth (or trajectory) becomes an interesting problem.

In stereo photogrammetry, the depth resolution is better at close distance than for objects farther away. So, with an object approaching an ego-vehicle (i.e., the vehicle the stereo system is operating in), the accuracy of the estimated trajectory would appear to improve over time, allowing for better estimates of the possibility of a collision scenario [11].

W. Khan and R. Klette are with the Computer Science Department, Tamaki Innovation Campus, The University of Auckland, New Zealand. Contact email: wkha011@aucklanduni.ac.nz

[6]. Due to the integral measurable depths, the stereo system installed on an ego-vehicle for driver assistance, will have some uncertainties for estimating an object trajectory and issuing a warning [8]. Timing of the warning and its accuracy are crucial for setting the confidence of a driver over the system. An earlier accurate warning improves the driver’s confidence over the system, whereas false or late warnings degrade this confidence [1], [2]. For a collision scenario, an accurate and timely warning issued by the safety system is a challenge.

As highlighted previously [9], for a collision scenario with a colliding object moving with constant velocity, a stereo-based safety system can issue timely warnings. However, due to the uncertainties in estimated trajectories, warnings can also be false. Here, we generalize the previously used model further by considering an object moving with constant speed but variable direction.

II. ASSUMPTIONS

A colliding object is rigid and is of size \(L \times W \times H\), and is travelling with constant speed \(V\) on a flat surface with \(Y = 0\), however its direction of motion may vary along \(XZ\)-components over time. The object’s \(L\) dimension is always in the direction of the object’s motion. The colliding object first appears at \(O^\top(0) = [O_x(0), O_y(0), O_z(0)]\). As the direction of motion changes, the pose of the object will change accordingly.

The ego-vehicle starts at \(O^\top(0) = [O_x^0(0), O_y^0(0), O_z^0(0)]\) and travels at a constant velocity \(V^2\) in the direction of the \(Z\)-axis so that \(O_x^0 = V^2_t = O_y^0 = V^0_y = 0\). We define an exclusion zone around the ego-vehicle of radius \(r_{\text{exc}}\).

At frame 0, \(n\) equidistant and binocularly visible feature points are chosen over the extent of an object. The closest of them is the reference point for the object.

A. Single vs. More Feature Points

In object tracking, an object is often referred to as by a single reference point. However, when the pose of an object is changing, as is here, then the single reference point may observe the change in object position later than closer feature points. This is particularly important when the extent of object covers more than one disparity values.

Figure 1A shows an example scenario. In this scenario, a toy-car of dimensions \([L \times W \times H] = [0.1, 0.1, 0.3]\) \(m\) is rotating around its \(XZ\)-centre. We use a stereo system in canonical configuration, defined by parameters \(f = 16mm\), \(b = 60mm\), \(\delta s = 0.03s\) and \(\tau = 4.65\mu m\). The stereo
system uses a DP matcher \cite{14} to generate dense disparity maps at each pixel (no sub-pixel accuracy). After background subtraction, we colour-code the disparity levels over the extent of the toy-car.

As the pose of the toy-car changes over time, the observed disparities over the toy-car’s surface also change accordingly. Due to non-linear uncertainties, only few pixels over the surface are observed at a different disparity in $\delta s$. Therefore, instead of using a single feature point, multiple feature points can assist in improving the system’s decision.

As common for a safety system, the nearest object position is considered to safely avoid the worst case scenario. For the sequence used in Figure II-A with some post-processing, the nearest feature point can be chosen as a top-corner point, central point, or a bottom-corner point. Post-processing involves excluding smaller real-world contours on the toy-car, i.e. smaller than $[L \times W \times H] = [0,0.10,0.05] \text{m}$.

The output is the measured distance of each chosen feature point with respect to the central point, or a bottom-corner point. Post-processing involves excluding smaller real-world contours on the toy-car, i.e. smaller than $[L \times W \times H] = [0,0.10,0.05] \text{m}$.

The text for further explanations.

\begin{equation}
\bar{V}^r(k) = \begin{bmatrix}
V^r \cos \zeta^r_k \\
0 \\
V^r \sin \zeta^r_k
\end{bmatrix}
\end{equation}

with object’s $L$ dimension in the direction of $\bar{V}(k)$. After each system observation $k-1$, the object moves to a new position $O^r(k) = O^r(k-1) + (\bar{V}^r(k-1) \cdot \delta s)$.

The model assumes that object changes its trajectory after every observation too. This effectively would also change object pose, derived by repositioning $n$ feature points (see Section IV). The object trajectory changes from $\zeta^r_{k-1}$ to $\zeta^r_k \pm \delta c$, where $\delta c$ is the rate of change of angular velocity and is input to the model (see Figure III).

Notation $\pm$ can either be $+ \text{ or } -$. Figure 3 illustrates the use of the $\pm$ notation. The same procedure is also later explained in Algorithm 1. At $k = 0$, labelled as position A in Figure 3, object trajectory $\zeta^r_0 = \zeta^r_L$. For the following observations, $\pm = +$ is used by the model, to change object trajectories from A to C. Later, if $\pm = +$ is used any further, then $(\zeta^r_k + \delta c) \geq \zeta^r_R$, where $\zeta^r_R = \tan^{-1}\left(\frac{-O^r_y(k)}{-O^r_z(k)}\right) + \sin^{-1}\left(\frac{r_{exc}}{D^r(k)}\right)$ is the right tangent to the ego-vehicle exclusion zone from the object at observation $k$. At this trajectory, the object would pass behind the ego-vehicle. So, for a collision scenario, $\pm = -$ is used for trajectories between C and E.
IV. OBJECT POSE

The pose of an object is determined based on its real-world velocity derived from $\overrightarrow{V}(k)$ as in Equation 1, defined by

$$\overrightarrow{V}(k) = \begin{bmatrix} V_x^r(k) \\ V_y^r(k) + V_z^r \end{bmatrix}$$  \hspace{1cm} (2)

The angle of the actual object trajectory, relative to the X-axis, is $\eta_k = tan^{-1}\left(\frac{V_y^r(k)}{V_x^r(k)}\right)$, and it is used by the model to determine the object pose before observation $k$ (and after system observation $k-1$).

In the model, we consider $n$ equidistant feature points on the surface of the colliding object. The nearest observed point in real-world coordinates is the reference point $\overrightarrow{O_j}(k)$ (at $j = 1$), and it is at position $\overrightarrow{O}(k)$.

As the object trajectory changes with respect to $\overrightarrow{V}(k)$, $\eta_k$ also changes. This leads to a change in positions of each feature point on the object’s surface with respect to the object reference point $\overrightarrow{O_j}(k)$.

The feature point $j$ at distance $c$ from $\overrightarrow{O_j}(k)$ along the $W$ dimension moves to

$$\overrightarrow{O_j(k)} = \begin{bmatrix} O_{1x}(k) + c \cos(\eta_k - \frac{\pi}{2}) \\ 0 \\ O_{1z}(k) + c \sin(\eta_k - \frac{\pi}{2}) \end{bmatrix}$$  \hspace{1cm} (3)

Similarly, the feature point $j$ at distance $c$ from $\overrightarrow{O_j}(k)$ along the $L$ dimension moves to

$$\overrightarrow{O_j(k)} = \begin{bmatrix} O_{1x}(k) + c \cos(\eta_k - \pi) \\ 0 \\ O_{1z}(k) + c \sin(\eta_k - \pi) \end{bmatrix}$$  \hspace{1cm} (4)

It is assumed that the system makes observation $k$, after, the object pose has been pre-adjusted by the model. Algorithm 1

**GeneralCollisionTrajectories**($\delta s, V_r, \delta c, r_{exc}, \overrightarrow{V}(0), \overrightarrow{O}(0)$)

returns system state $S$ at each observation $k$;
Initialize state, $S \leftarrow S_0$ and observations $k \equiv -1$;
Object initial relative position $\overrightarrow{O}(0) = \overrightarrow{O}(0)$;

for each observation $k$ do

Relative object distance $D^r(k) = \|\overrightarrow{O^r(k)}\|$;
Relative collision trajectory to the ego-vehicle is $\zeta_k = tan^{-1}\left(\frac{-\overrightarrow{O^r(k)}}{\overrightarrow{O}(k)}\right)$;
Left tangent to exclusion zone in polar coordinates is $\zeta^L = \zeta_k \cdot sin^{-1}\left(\frac{r_{exc}}{\overrightarrow{V}(k)}\right)$;
Right tangent to exclusion zone in polar coordinates is $\zeta^R = \zeta_k + sin^{-1}\left(\frac{r_{exc}}{\overrightarrow{V}(k)}\right)$;

if First observation at $k = 0$ then

Initial collision trajectory is the left tangent, $\zeta^L = \zeta^R$;

For $\pm$ notation SET $\text{MinusFlag} \leftarrow \text{TRUE}$;

else

if $\zeta^L$ is supposed to increase ($\text{MinusFlag} == \text{FALSE}$) then

if ($\zeta^L_{k-1} + \delta c > \zeta_k$) then

$\zeta^L_k$ has to decrease, SET $\text{MinusFlag} \leftarrow \text{TRUE}$;

else

$\zeta^L_k$ has to increase, SET $\text{MinusFlag} \leftarrow \text{FALSE}$;

end if

else

if ($\zeta^L_{k-1} + \delta c < \zeta_k$) then

$\zeta^L_k$ has to increase, SET $\text{MinusFlag} \leftarrow \text{FALSE}$;

else

$\zeta^L_k$ has to decrease, SET $\text{MinusFlag} \leftarrow \text{TRUE}$;

end if

end if

if $\text{MinusFlag} == \text{TRUE}$ then

Decrease $\zeta^L_k = \zeta^L_{k-1} - \delta c$;

else

Increase $\zeta^L_k = \zeta^L_{k-1} + \delta c$;

end if

Use Equation 7 to derive the relative object collision velocity $\overrightarrow{V^r(k)}$ at observation $k$;
Use Section IV to determine the pose of the object;
Use either the variable velocities in Section V-A to determine the system state $S$;
Before the next observation, the object moves to new position $\overrightarrow{O^r(k)} = \overrightarrow{O^r(k)} + \delta s \overrightarrow{V^r(k)}$;

end for

Fig. 5. Algorithm 1, modelling a collision scenario with variable velocity and decision making.
(see Fig. 5) shows the steps followed by the model to change object velocity and its pose before each system observation.

V. STEREO-BASED DRIVER ASSISTANCE SYSTEM

This model has the same design constraints as discussed in Section II. D. of [9]. Therefore, due to the discrete nature of the stereo measurements, the stereo measurements have uncertainty in them. This leads to a possible range of trajectories derived from the range \((min_j V^T \text{ to } max_j V^T)\) of velocities (see Equations 10 and 11 of [9]). To handle the changing trajectories of an object travelling with variable velocity, we compute the weighted average of velocity ranges. This average is computed by first affiliating a weight \(w^k = [w^k_u, 0, w^k_d]\) to each feature point at observation \(k\).

A. Weighted System

For simplicity, we assume that the following procedure would be applicable to each feature point. We also assume that the maximum legally permitted speed in a traffic scenario, the system assumes that the object speed would not exceed \(V_{\text{max}} = sV_{\text{limit}}\), where \(s\) is the speeding factor. Parameter \(s\) is input to the model.

At observation \(k\), weight \(w^k\) is directly related to the time \(\Delta t\) for which a feature point is observed at the same position \((u, d)\) with some uncertainty \((\pm \nu_u, \pm \nu_d)\). \(w^k\) is also indirectly related to the uncertainty at the measured position \((\Delta X(u, d), \Delta Z(d))\).

\[
-w^k = \begin{bmatrix}
\Delta t_y \\
\left| \hat{X}_{1,k} - \hat{X}_{0,k} \right| \\
0 \\
\Delta t_x \hat{Z}_{1,k} \\
\left| \hat{Z}_{1,k} - \hat{Z}_{0,k} \right|
\end{bmatrix}
\]  

(5)

where \(\hat{Z}_{0,k}\) is the observed \(Z\)-position at \(\hat{Z} = (u, d)\) at observation \(k\). Similarly, \(\hat{X}_{0,k}\), \(\hat{X}_{1,k}\), \(\hat{Z}_{0,k}\), and \(\hat{Z}_{1,k}\) are the observed positions at the boundary of the measured pixel \(\hat{X} = (u + \nu_u, d - \nu_d)\), \(\hat{X} = (u - \nu_u, d + \nu_d)\), \(\hat{Z} = (u, d - \nu_d)\), and \(\hat{Z} = (u, d + \nu_d)\). \(\hat{Z}_{0,k}\) is used in Equation 5 as a scaling factor for the \(Z\)-term of \(w^k\).

Due to the uncertainty of stereo measurement, the measured velocity is a range as already discussed in Equations 10 and 11 of [9]. To compute their weighted outcome at observation \(k\), we generalize the measured velocity term as \(mV^k\). With \(wmV^k\) as their corresponding weighted velocity, where \(\frac{1}{w^k}\) keeps the feedback from previous observations and is computed at previous observation \(k - 1\),

\[
\frac{1}{wV^k} = \begin{bmatrix}
wV^k - \sum_{i=0}^{k-1} w_i \sum_{j=0}^{k} V_j + \sum_{i=0}^{k} w_i \\
\sum_{i=0}^{k-1} \sum_{j=0}^{k} V_{i,j} + \sum_{i=0}^{k} w_i \\
0 \\
\sum_{i=0}^{k-1} \sum_{j=0}^{k} V_{i,j} + \sum_{i=0}^{k} w_i \\
\sum_{i=0}^{k-1} \sum_{j=0}^{k} V_{i,j} + \sum_{i=0}^{k} w_i
\end{bmatrix}
\]  

(7)

At \(k = 0\), we have that \(wV^k = wV^k = 0\) and \(wmV^k = 0\).

Velocity Extrema: For each feature point \(j\), the velocity extrema (before weighted average) range between \(\min_j V^T\) (see Equation 10 of [9]) and \(\max_j V^T\) (see Equation 11 of [9]). Whereas, after weighted average using Equation 6 the weighted velocity extrema range between \(\min_j wV^T\) and \(\max_j wV^T\).

Previously, the object was assumed to be moving with constant velocity and the system computed the velocity extrema to be the maximum values consistent with all previous observations. Here, however, since the object is moving with variable velocity, therefore the system chooses the best outcome based on the weight derived in Equation 7.

The object is still assumed to be rigid. Therefore, the system computes the largest minimum and the smallest maximum as a range of velocities being consistent with all feature point observations. Thus, for the whole object we have that

\[
\frac{\max V^T}{\min V^T} = \begin{bmatrix}
\min \left( \max_j V^T \right) \\
\min \left( \max_j V^T \right) \\
0 \\
\max \left( \min_j V^T \right) \\
\max \left( \min_j V^T \right)
\end{bmatrix}
\]  

(8)

\[
\frac{\min V^T}{\max V^T} = \begin{bmatrix}
\min \left( \max_j V^T \right) \\
\min \left( \max_j V^T \right) \\
0 \\
\max \left( \min_j V^T \right) \\
\max \left( \min_j V^T \right)
\end{bmatrix}
\]  

(9)

The range of trajectory angles, \((\rho_L, \rho_R)\), is then computed from the truncated \(\min V^T\) and \(\max V^T\).

In order to judge a collision, the system follows the same procedures as in [9], Sections II.E.2, II.E.3, and II.E.4. Algorithm 1 of [9] becomes slightly different for modelling the motion of feature point(s) position based on Equations 5 and Equation 4 instead of \(O_{k,j} = O_{0,j} + V^T t\). Over the course of time the system stays at one of the following states. \(S_0\) for first observation, \(S_1\) if object object is safe, \(S_2\) if system uncertain of object course and can safely wait for another observation, \(S_3\) if system uncertain of object course and not safe to wait for additional observations, or \(S_4\) if object is on a collision course. For cases \(S_3\) and \(S_4\) system issues a braking warning.

VI. RESULTS AND DISCUSSION

The model has a wide range of parameters. We use the standard parameters from Table 1. We assume that the disparity
map is with sub-pixel accuracy of up to two sub-pixels on a pixel.

In our previous study of linear systems, we highlighted that a stereo-based safety system is prone to issuing too many precautionary warnings. We also used two hypotheses to avoid generating the precautionary warnings when possible. Firstly, estimated high speeds were truncated to be not faster than $V_{max}$.

A. Experiment 1a: Linear vs. Weighted

Using the typical parameters from Table I with $n = 1$, Figure 6(a) and Figure 6(b) shows the maximum tolerable speeds of a linear and a weighted system. The tolerable speeds are only for a collision scenario for objects first appearing at any of these locations within stereo CFoV. In Figure 6(a) a weighted system outperforms a linear one. Both systems have similar behaviour at closer distances as both issue timely precautionary warnings. However at farther distances, while the object takes more time in colliding with the ego-vehicle, the system can safely delay the precautionary warning. More time should permit systems to improve their estimates. But, as the trajectory keeps changing, the linear system estimates that the object is safely avoiding the ego-vehicle. So, it fails to recognize a collision scenario. Whereas, the weighted system, gives weight to each measurement based on the time it is observed at the same location, hence avoids such false estimates.

B. Experiment 1b: Anomalies of a Linear System

For the typical system with Table I configuration parameters, a linear system shows certain anomalies at farther distances (see Figure 6(a)). To analyse these anomalies we assume two similar sized objects first appearing at two different locations $A = [4, 112]^{T} \text{ m}$ and $B = [4, 114]^{T} \text{ m}$ from the ego-vehicle. A being closer should have lower tolerable speed than B, however, the maximum tolerable speed of A (25.4 ms$^{-1}$) is unexpectedly higher than that of B (1.1 ms$^{-1}$). To analyse this anomaly, we consider the collision speed of $V = 1.2$ ms$^{-1}$ tolerable by B but intolerable by A. A being closer is first observed at $d = 13.5$, while B being farther away is first observed at $d = 13.0$ (sub-pixel level). After the 2nd observation, the system can determine the range of trajectories for the first time. At this stage, the system estimates that the objects could possibly collide, but as the objects are quite far from the ego-vehicle, so the system can safely wait for additional observations. After 61 observations for A with the object at (2.6, 76.6)m, the system can no longer wait for additional observations and hence issues a timely precautionary warning. Whereas, B is farther away at (2.7, 78.6)m from ego-vehicle, thus the system chooses to consider additional observations. After observation 64 for B, the system has the estimated range of trajectories $(\rho_L, \rho_R) = (262.9, 266.8)^\circ$. As, this range is not crossing the ego-vehicle’s exclusion zone tangents $(\zeta_L, \zeta_R) = (266.9, 269.9)^\circ$, so the system falsely estimates that B is safe.

### Table I

**System parameters used in the model**

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>Typical value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f$</td>
<td>Focal length</td>
<td>9 mm</td>
</tr>
<tr>
<td>$\tau$</td>
<td>Pixel size</td>
<td>4.7 $\mu$m</td>
</tr>
<tr>
<td>$b$</td>
<td>Baseline length</td>
<td>750 mm</td>
</tr>
<tr>
<td>$d_{max}$</td>
<td>Maximum disparity</td>
<td>127</td>
</tr>
<tr>
<td>$\phi$</td>
<td>Vergence angle</td>
<td>0$^\circ$</td>
</tr>
<tr>
<td>$\delta s$</td>
<td>Sampling interval</td>
<td>0.03 s</td>
</tr>
<tr>
<td>$r_{exc}$</td>
<td>Radius of vehicle’s exclusion zone</td>
<td>2 m</td>
</tr>
<tr>
<td>$V^t$</td>
<td>Vehicle speed</td>
<td>17 ms$^{-1}$ (60 kmh)</td>
</tr>
<tr>
<td>$V_{crit}$</td>
<td>Maximum collision speed</td>
<td>2.77 ms$^{-1}$ (10 kmh) [18]</td>
</tr>
<tr>
<td>$V_{limit}$</td>
<td>Maximum speed limit</td>
<td>17 ms$^{-1}$ (60 kmh)</td>
</tr>
<tr>
<td>$s$</td>
<td>Speeding factor</td>
<td>1.5</td>
</tr>
<tr>
<td>$t_d$</td>
<td>Driver response time</td>
<td>0.5 s [2], [3]</td>
</tr>
<tr>
<td>$\mu$</td>
<td>Coefficient of friction</td>
<td>0.4 [16], [5]</td>
</tr>
<tr>
<td>$t_p$</td>
<td>Object detection and classification time</td>
<td>1.5 ms</td>
</tr>
<tr>
<td>$(L \times H \times W)$</td>
<td>Object size</td>
<td></td>
</tr>
<tr>
<td>$\delta_e$</td>
<td>Rate of collision angle</td>
<td>0.1$^\circ$</td>
</tr>
<tr>
<td>$n$</td>
<td>Maximum number of feature points</td>
<td>9</td>
</tr>
</tbody>
</table>

**Derived Values**

| $V_{max}$   | Object maximum speed                            | 25.5 ms$^{-1}$ (90 kmh) |
| $Z_{min}(d_{max})$ | Minimum depth in CFoV                              | 11.3 m     |
| $\theta$    | Half angle of stereo field-of-view               | 14.9$^\circ$   |
| $t_b$       | Vehicle braking time                             | 1.8 s [16])    |
| $D_b$       | Maximum safe braking distance                    | 44.4 m        |

**Glossary**

| $\rho_L, \rho_R$ | Range of trajectory angles for an object       |
| $\zeta_L, \zeta_R$ | Trajectory angles with (or tangents to) vehicle’s exclusion zone |
VII. CONCLUSIONS

The designed tool can assist a safety engineer in assessing the effect of various stereo configurations, and of vehicle and object parameters. Previously we designed a similar tool for constant object trajectories only. Now, modelling variable collision trajectories, we have shown that a linear system can falsely estimate a safe travel.

To mitigate this problem, we designed a weighted system which always issues a timely warning for a collision scenario - provided the object is binocularly visible CFoV. Although these warnings are timely, mostly they are precautionary as well, as the system is mostly unsure about the exact object trajectory to some extent, thus issuing a warning based on a possible worst case trajectory.

REFERENCES