System Architecture for an Intersection Assistant Fusing Image, Map, and GPS information

Stefan K. Gehrig
HPC T 728
DaimlerChrysler AG
70546 Stuttgart, Germany
stefan.gehrig@dcx.com

Sebastian Wagner
Stammheimer Str.10
Innovative Navigation GmbH
70806 Kornwestheim, Germany
sebastian.wagner@innovative-navigation.de

Uwe Franke
HPC T 728
DaimlerChrysler AG
70546 Stuttgart, Germany
uwe.franke@dcx.com

Abstract

Street intersections account for a large portion of road accidents everywhere around the world. Therefore, a driver assistance system supporting the driver to negotiate intersections is highly desirable and has a tremendous potential of reducing the number of collisions at intersections.

What makes intersection assistance more challenging than other traffic scenarios such as highways and rural roads? On one hand, the environment all around the vehicle must be monitored. A very large field of view for the local sensors is required. We decided to use both omnidirectional cameras for a panoramic view and active cameras to inspect smaller areas around us at a high resolution. On the other hand, intersections come in all shapes, sizes, and signal regulations. Hence, some prior knowledge of the signal mode and the intersection geometry is very helpful. This can be achieved via GPS in conjunction with an attributed digital map.

We propose a system architecture for intersection assistance using image, map, and GPS information to issue the typical intersection warnings such as cross traffic warning or traffic light warning.

1 Introduction

Intersections are dangerous. A look at the accident statistics in Germany shows, that more than one quarter of all accidents occur at street intersections ([1], see Figure 1). Statistics from other countries exhibit a similar peak for collisions at intersections.

In the past, a lot of research has been devoted to driver assistance for highway driving (see e.g. [2]). Several car manufacturers already offer adaptive cruise control (ACC) functions as optional features in their models. In recent years, recognition tasks for inner city tasks have been addressed in research (see e.g. [3]) to enable an urban city assistant function. This so called Urban City Assistant uses cameras as its main sensors.

Driver assisting systems in general can choose from a variety of sensors. Onboard automotive sensors such as ultra-sonic, LIDAR, RADAR, or image sensors have already been deployed heavily (see e.g. [4]). Ultra-sonic sensors are used for parking aid, LIDAR and RADAR are available for adaptive cruise control systems, and vision is used for night view systems or for lane departure warning systems (see e.g. [5]).

Besides these local sensors, navigation systems deploy the satellite infrastructure to localize the vehicle via GPS (global positioning system). Combined with a digital map, routing information can be obtained.

Since an essential part of intersection assistance is traffic light monitoring to keep drivers from running a red light, image sensors must be deployed unless we would...
assumed additional infrastructure (e.g., transponders). Image sensors introduced in the references mentioned above exhibit a limited field of view which is insufficient for intersection monitoring of cross traffic. We extend the limited field of view of conventional cameras in two ways:

- **Active Cameras:** Similar to the human driver using his eyes with a limited field of view, we use a conventional camera and mount it on a pan-tilt unit that allows us to inspect areas of our surroundings in (almost) any direction. When we have to yield to vehicles coming from the right, we would turn our head in that direction. The same can be done with an active camera.

- **Panoramic Cameras:** Already popular in the field of robotics, we use a camera and a parabolic mirror to obtain a 360°-view of the scene. Since the projection properties are known, a rectified image can be obtained in any direction. However, the resolution of the scene is much smaller.

How is this paper organized? In Section 2 some previous work in intersection assistance is listed. Section 3 describes our approach to intersection assistance and derives our system architecture. Some fusion strategies of image processing information and map information are detailed in Section 4. Section 5 describes how we obtain such attributed maps. Preliminary results obtained with our demonstrator are shown in Section 6. Conclusions and future work comprise the final Section.

2 Related Work

In this brief literature survey, we limit ourselves to papers who address intersection assistance directly.

In [6], model-based intersection recognition is performed to extract intersection topology and lane structure. Here, a camera mounted on a house was used. We extract this type of information from a digital map. Tracking objects using a stationary camera is described in [7]. Again, a stationary camera was used. Jocoy [8] uses GPS and RADAR to detect potential collisions ahead of an intersection. In [9], driver behavior at unsignalized intersections is presented. From there, one can derive the strategies how a human driver negotiates this type of intersection. An early paper tries to assess the danger of an intersection depending on the position of vehicles in different parts of the intersection and with different motion vectors (see [10]). Once again, this work was performed using stationary sensors - not onboard sensors.

Summarizing, it is striking that, to our knowledge, intersection assistance (especially traffic light recognition) has not been approached with vision in the literature.

3 System Architecture

3.1 Hardware

Our demonstrator for an intersection assistant system is equipped with a digital stereo camera system, that is mounted on pan-tilt units (PTU). The pan-tilt units are attached to the dashboard. The camera system and the pan-tilt units are connected to the image processing computer, which is an Intel-based system, running under SuSE-Linux 8.0. The digital video information is transmitted over network cables and processed by a digital framegrabber card. The pan-tilt units are basically servo/stepping motors that can turn the mounted cameras to (almost) any direction. This helps us to track any desired object, even traffic lights that are mounted unfavorably high above or besides the street.

An alternative setup would be one pan-tilt unit with two cameras mounted on top. However, this setup is not acceptable when it comes to packaging issues. Single PTU-cameras can be packaged in small units (see Figure 2, you see a conventional fire-lighter on the right side), whereas a Stereo-PTU automatically needs at least the size of the stereo baseline. In addition, two PTU-cameras can also be controlled independently to perform multiple tasks (e.g. tracking a stoplight and a leading vehicle) at the same time. Our current camera setup is displayed in Figure 3.

![Figure 2: Example of an active camera. The model here is smaller than a fire-lighter. In our demonstrator, however, we use a larger version due to accuracy reasons (see next figure).](image-url)

Another stereo camera set, with 2 panoramic camera-mirror systems, is planned to be mounted to the left and right of the rearview mirror. These cameras can inspect the full area around the vehicle at low resolution (see Figure 4 for a typical view).

Furthermore, two additional PCs collect information...
Figure 3: Right camera with pan-/tilt units installed in our demonstrator.

Figure 4: Image of our panoramic camera. For intersection assistance, especially the area facing forward is relevant.

Figure 5: System architecture of our demonstrator.

3.2 Software
The software for this project is written in C++. We decided to use Parallel Virtual Machine (PVM) for host communication ([11]) developed for parallel computing. PVM is a TCP socket-based program. It regulates the access to shared data and spawns processes. The image processing PC is the PVM master. It can spawn processes and transfer data to or from the GPS PC and the vehicle PC.

3.3 GPS signal accuracy
In a nutshell, the GPS uses satellite infrastructure to localize itself via triangulation by receiving at least signals from 4 different satellites simultaneously.

A few years ago, the GPS information was degraded by an artificial noise, resulting in a poor position quality (selective availability). But around the turn of the century this noise was removed and the GPS signal became interesting for commercial applications with adequate accuracy. Still, the accuracy of GPS in unobstructed areas is about 5m, improving to about 2m if there is a local differential adjustment signal being received (differential GPS, dGPS). Future satellite systems like the European positioning system EGNOS will be about the same quality. To further improve this accuracy, and to deal with GPS dropouts, we fuse yaw rate and velocity of our vehicle to improve the localization accuracy and from other sensors: A 19" Motorola-based VME Computer, running on LynxOS, gathers car-based information from the onboard sensors via CAN-Bus (vehicle PC). This includes e.g. velocity, yaw rate, direction signal status ... The GPS signal is received by a laptop that runs under SuSE-Linux 7.2 (GPS PC). All PC's are interconnected via an Ethernet hub on one broadcast domain.

An overview of our system architecture is summarized in Figure 5.
obtain an accuracy of about 1 meter. In urban areas with obstructed views and multi-scattering problems, we yield a signal accuracy of about 10 m at all times with off-the-shelf GPS and dGPS receivers, with a potential of obtaining an accuracy of 2 m by improving the localization software.

4 Fusion of Image Information and Map Information

The fusion of image processing and GPS based map applications will increase the flexibility and performance of our application. Using an attributed digital map, it is possible to incorporate not only the human recognition but also human knowledge about a certain scene. The map gives us valuable information. It informs us about the upcoming situation even though it might be out of the visual range, almost independent of the weather conditions and it could provide the position of significant objects in front such as traffic lights or stop signs.

A fusion scheme of image processing and map information has already been implemented for lane recognition in [12]. Here the limited lookahead of the image processing algorithm was augmented by the large lookahead derived from the digital map.

For the intersection scenario, especially the following tasks can be efficiently improved by map information:

- Traffic light recognition: The map can trigger the traffic light recognition task in a certain area based on map data. Hence, the detection task is largely facilitated. Besides the location, also the size of the traffic light is available. Obviously, the recognition of the traffic light status cannot be retrieved from the map data. This task is left to an image processing algorithm. See Figure 6 for an illustration.

- Traffic sign recognition: Since the traffic sign information is available in the map database, the recognition task is facilitated to a mere verification task (is it still there?).

- Cross traffic recognition: Since the crossing geometry is known from the map, the image processing algorithms can specifically investigate the intersection area. In addition, the precedence regulations are also known from the map (signalized intersection, traffic sign intersection, unsignalized intersection). In order to detect cross traffic, image processing algorithms are used. A fusion of optical flow information and stereo information will be used to detect cross traffic very fast (see [13]).

- Traffic light assignment: Information of the lane geometry around intersections is available, including information which traffic light belongs to which lane and what maneuvers are possible starting from that lane. In addition the vehicle signal status is evaluated to incorporate the driver commands. This helps to assign the correct traffic light to the current position (see Figure 7).

Figure 6: Traffic light position and traffic light direction information is stored as map attributes in addition to the lane information.

Figure 7: The signal status of our demonstrator is evaluated to find out the corresponding traffic light.

5 Map Generation

5.1 Introduction

Currently, a lot of efforts are made to enhance the available digital street maps with additional attributes. Several map manufacturers have announced releases of maps including lane information. Since our needs
for map attributes are not fully covered with currently available maps, we decided to create our own maps using the ECDIS map format, a standard format used in ship navigation.

A recorded stereo image sequence of the road, that is to be surveyed and mapped, is used to generate the digital map database. Using stereo images allows us to extract 3D information easily (see e.g. [3]). The recorded sequence includes the GPS information of the vehicle at the time of recording. The GPS position is stored for each image as a meta text attachment. This insures synchronization between the image and the GPS signal. The sequences are analyzed offline which allows an interactive selection of the traffic lights in the scene.

5.2 Coordinate system transformation
One step of mapping of a landmark results in a position, that is given relative to the vehicle’s motion direction. In order to associate several images, an arbitrary reference system is chosen. For every recorded image pair, a relative position is computed and transformed into the reference coordinate system. Figure 8 shows a bird’s eye view of the scenario.

![Figure 8: Coordinate transformation from the vehicle's coordinate system \((a_v)\) to the reference system \((a_{ref})\).](image)

The \(x\)-axis points to East and the \(z\)-axis to the North. The height of the landmark is not changed and taken from the 3D-surveying of the image.

The new position within the reference system is computed by a rotation and translation as shown in Equation 1. \(\xi\) and \(r_v\) are heading direction and vehicle radius, respectively. See Figure 8 for an explanation of the used quantities. The vehicle radius is computed using standard vehicle kinematics (see e.g. [14]).

\[
\begin{pmatrix}
\cos(\xi) & 0 & -\sin(\xi) \\
0 & 1 & 0 \\
\sin(\xi) & 0 & \cos(\xi)
\end{pmatrix}
\cdot
\begin{pmatrix}
a_{x_{ref}} \\
a_{y_{ref}} \\
a_{z_{ref}}
\end{pmatrix}
=
\begin{pmatrix}
\frac{r_v^x}{r_v} \\
0 \\
\frac{r_v^z}{r_v}
\end{pmatrix}
\tag{1}
\]

5.3 Averaging
Every image that has been taken into account results in a position within the reference system. The average position can be determined taking the measurement errors into account. Equation 2 shows the formula that was used for averaging. The disparity (displacement of a structured point between left and right camera) is inverse proportional to the distance of that structured point. The distance uncertainty of such a measured point is proportional to the square of the disparity, and the weighted regression of multiple measurements yields a dependence on the square of the distance uncertainty. As can be seen from this equation, by far the most weight is put on the measurement, where the traffic light is closest to the camera system.

\[
m = \frac{\sum_i x_i \cdot \text{disp}_{i}^4}{\sum_i \text{disp}_{i}^4}.
\tag{2}
\]

GPS position uncertainty \((\sigma_{\text{gps}})\) is not accounted for yet but will be included in the future. The averaging formula extends in such a way that,

\[
\text{disp}_{i}^4 \rightarrow (\text{const} \cdot \text{disp}_{i}^4 + \sigma_{\text{gps}}^2).
\tag{3}
\]

6 Preliminary Results

6.1 Map Generation Results
We compared our auto-generated map data with map data available from official sources. The accuracy of our maps is well within 5m, even in the presence of comparatively tall buildings that generate signal dropouts and multi-scattering. Figure 6.1 shows a preliminary comparison, the black crosses denote the position of a traffic light. The black GPS trace is the localization of the vehicle driving around the Esslingen innercity ring.

6.2 Traffic Light Recognition Results
Currently, only the active cameras are deployed to perform traffic light recognition. Tracking of the traffic light with the pan-tilt units allows us to keep the traffic light in the field of view until we have passed it. Since our traffic light recognition operates with
Figure 9: GPS-track and traffic light positions (crosses) overlayed with a street map.

greyscale cameras, only robust recognition in daylight is possible. Color information is necessary for night time driving.

7 Conclusions and Future Work

In this paper an intersection assistant was presented using vision and map information as its primary sources. Since cross traffic is potentially dangerous, early detection of dangerous cross traffic is vital for an intersection assistant. The required large field of view in combination with the need to detect the traffic light status led us to use cameras with a large field of view: Active cameras and panoramic cameras. We believe that such a system can help reducing collisions at intersections significantly.

Future work will include the following steps:

- Stereo evaluation of the two cameras mounted on two pan-/tilt units: This requires synchronous control of the PTUs and rectification information in all possible directions.

- Integration of two panoramic cameras in the demonstrator. We also plan a stereo evaluation of the vehicle’s environment with these cameras. A back-of-the-envelope calculation indicates that depth information up to 15m from the sensors can be obtained.

- Early detection of cross traffic with active cameras using stereo and flow methods.

- Incorporation of lane information to make traffic light association more reliable (see [12]).

References