ÓBUDAI EGYETEM
ÓBUDA UNIVERSITY

PhD Thesis

Towards a Transfer Concept for Camera-Based Object Detection

From Driver Assistance to the Assistance of Visually Impaired Pedestrians

Judith Jakob

Supervisor:
Dr. habil. József Tick

Doctoral School of Applied Informatics and Applied Mathematics

June 8, 2020

Comprehensive Exam Committee:

Prof. Dr. habil. László Horváth, professor emeritus, Óbuda University, Budapest

Prof. Dr. Róbert Fullér, DSc, Széchenyi István University, Győr

Dr. Zoltán Vámossy, Director of Institute, Óbuda University, Budapest

Public Defense Committee:

Chair of the Committee:

Prof. Dr. Péter Nagy, DSc, professor emeritus, Óbuda University, Budapest

Secretary of the Committee:

Dr. Adrienn Dineva, Óbuda University, Budapest

Opponents:

Dr. Henriette Komoróczki-Steiner, Óbuda University, Budapest

Prof. Dr. Elmar Cochlovius, Furtwangen University of Applied Sciences, Furtwangen

Members of the Committee:

Prof. Dr. András Benczúr, professor emeritus, Eötvös Loránd University, Budapest

Prof. Dr. István Fekete, Eötvös Loránd University, Budapest

Prof. Dr. habil. László Horváth, professor emeritus, Óbuda University, Budapest

Acknowledgement

First and foremost, I express my gratitude to my supervisor Dr. habil. József Tick for his patient guidance, constant support, and excellent advice.

Furthermore, I thank the leaders of the *Doctoral School of Applied Informatics and Applied Mathematics* of *Óbuda University*, Prof. Dr. Aurél Galántai, Prof. Dr. László Horváth, and Prof. Dr. József Tar, for their valuable feedback in the course of my research and Dr. Zoltán Vámossy for his educational courses about image processing and computer vision.

I am grateful for the support of my superiors and colleagues at *Hochschule Furtwangen University*: Prof. Dr. Peter Fleischer for his uncomplicated guidance, Dr. Pirmin Held for the continuous exchange of experiences, and Kordula Kugele for her sound advice regarding qualitative research and her always open office door.

This project was supported by a scholarship of *Cusanuswerk*, a young talent program funded by the *German Federal Ministry of Education and Research*. In addition to financial assistance, I greatly benefited from the exchange with other scholarship holders and the informative seminars whose topics went beyond my fields of expertise.

Finally, I am deeply grateful to my family and friends for their unconditional love and support.

Statutory Declaration

I declare that I have authored this thesis independently, that I have not used others than the declared sources and resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

Budapest, June 8, 2020

Judith Jakob

Contents

List of Figures

List of Tables

Acronyms

CoPeD Comparable Pedestrian Driver iv, vi, 2, 14–16, 40–42, 44, 46, 47, 54, 66, 70, 72–74, 76, 78

AAL Ambient Assisted Living 3, 37

ADAS Advanced Driver Assistance Systems iv, vi, 1, 2, 9–12, 14–16, 18, 31, 33, 38–40, 46–48, 50, 53, 59, 62, 67, 69–78

ASVI Assistive Systems for the Visually Impaired iv–vi, 1–6, 9, 10, 12, 14–16, 18, 31, 33, 38–40, 46, 47, 49–52, 59, 61, 62, 67, 69–78

CNN Convolutional Neural Network 10, 34, 48

EDF Edge Distribution Function 62–64, 69, 73

EI Expert Interviewee(s) 22–32, 37, 38

ERANSEC Extended Random Sample Consensus 34, 57

GPS Global Positioning System 3, 8, 17, 24–27, 77

HD High Definition 15, 41

HOG Histogram of Oriented Gradients 34

ML Machine Learning v–vii, 12, 32, 48, 49, 52, 54, 56, 65–68, 73, 76

MSER Maximally Stable Extremal Region 34, 57

MTG Member(s) of the Target Group vii, 22–24, 27, 29–31, 35, 36, 38

NN Neural Network vii, 10, 11, 32, 48, 49, 66–69, 76

OCR Optical Character Recognition 27, 30, 32, 35, 36, 38, 39, 71

RBS Road Background Segmentation v–vii, 16, 47–50, 53, 54, 59, 62, 65–70, 72–74, 76, 78

ROI Region Of Interest 16, 59, 62, 67, 69, 70, 72, 73, 78

SSD Sensory Substitution Device(s) 7, 8, 12

SVM Support Vector Machine vii, 10, 32, 34, 48, 57, 58, 66–69, 76

Chapter 1

Introduction

According to [1], worldwide 36 million people were estimated to be blind and 216.6 million people were estimated to have a moderate to severe visual impairment in 2015. The according numbers have increased in recent years: in comparison to 1990, there was a rise of 17.5 % in blind people and 35.5 % in people with moderate to severe visual impairment. Due to demographic change a higher occurrence of diseases causing age-related vision loss is to be expected in the near future, leading to a decrease in the autonomous personal mobility of the affected [2]. In their study about how the visually impaired perceive research about visual impairment, Duckett and Pratt underline the importance of mobility for this group: "The lack of adequate transport was described as resulting in many visually impaired people living in isolation. Transport was felt to be the key to visually impaired people fulfilling their potential and playing an active role in society," [3].

In order to increase their independent mobility, I research possibilities of assisting visually impaired people in traffic situations by camera-based detection of relevant objects in their immediate surroundings. Although some according research concerning Assistive Systems for the Visually Impaired (ASVI) exists, there is a much higher amount of research in the field of Advanced Driver Assistance Systems (ADAS). Hence, it is suitable to profit from the progress in driver assistance and make it applicable for visually impaired pedestrians. As it is in general not possible to use driver assistance algorithms without any adaptations, e. g. due to known and stable camera positions, a generalized concept for the transfer is needed. Therefore, the presented research leads the way towards a universal transfer concept from ADAS to ASVI for camera-based detection algorithms.

1.1 Topic and Structure

This thesis deals with the question of whether and how camera-based object detection algorithms from ADAS can be adapted so that they can be used to support visually impaired people in traffic situations. The goal is to formulate a transfer concept that contains instructions on how to adapt algorithms from ADAS to ASVI.

Figure 1.1 illustrates the strategy for the development of the transfer concept. First, a literature review of camera-based ADAS and camera-based ASVI reveals the need for and novelty of a transfer concept. Then, the traffic scenarios where the visually impaired need support are gathered by conducting and evaluating a qualitative interview study. This results in a complete collection of relevant vision use cases. Afterwards, the overlap with use cases addressed in ADAS is built by analysing according ADAS literature. The overlapping use cases then have to be examined concerning the possibilities of adaptation from ADAS to ASVI. For the evaluation of the according algorithms, comparable video data from pedestrian and driver perspective are needed and therefore the Comparable Pedestrian Driver (*CoPeD*) data set is created. Finally, the adaptation procedures for all considered use cases have to be summarized in a concept for transferring ADAS algorithms to ASVI.

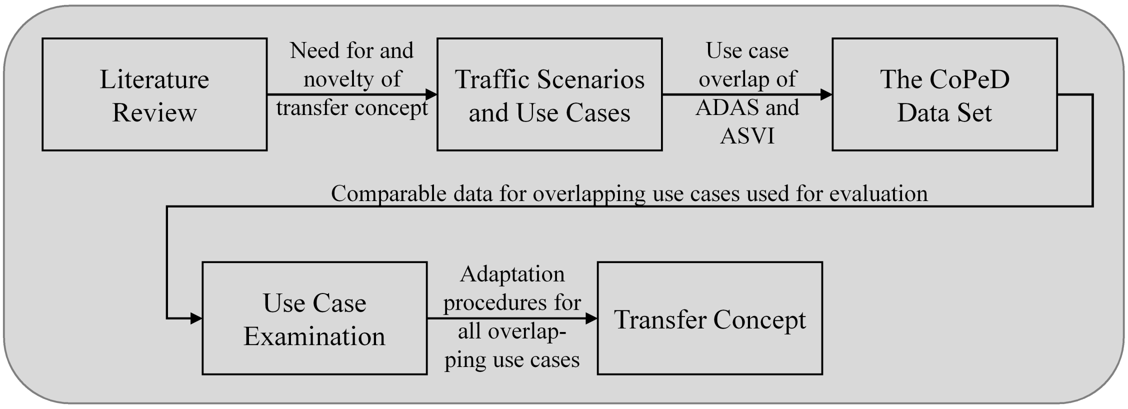


Figure 1.1: Research Strategy towards a Transfer Concept from ADAS to ASVI

Based on this strategy, the thesis is structured as follows: The next chapter 2 presents a literature review of both relevant fields, camera-based ADAS and camera-based ASVI, and from that the motivation for and novelty of the presented research is derived. Afterwards, chapter 3 introduces research categories including an overview of used methods and presents the objectives that have to be discussed and achieved in the course of the thesis. The subsequent chapters 4, 5, and 6 are each dedicated to one research category. Chapter 4 describes design and evaluation of the qualitative study that is used to determine the vision use cases that can support the visually impaired in traffic scenarios. In chapter 5, the *CoPeD* data set for traffic scenarios is introduced. The according data are used in chapter 6 in order to evaluate the algorithms adapted from ADAS to ASVI that are examined in this chapter. This thesis presents adaptations for three use cases, namely Road Background Segmentation (RBS), lane detection, and crosswalk detection. Chapter 7 outlines the confirmed theses based on the objectives formulated in chapter 3 and summarizes the scientific contributions made in this thesis. The final chapter 8 first points out perspectives for future research derived from the presented research and afterwards concludes the thesis.

1.2 Application Scenario

A transfer concept from ADAS to ASVI for camera-based object detection algorithms makes it possible to integrate the resulting ASVI algorithms into a camera-based mobile assistive system.

Adding detection algorithms for the remaining use cases that are of relevance for the visually impaired in traffic scenarios (see chapter 4) results in an assistive system that offers comprehensive support.

Although no such system is built in the course of the thesis' research, a sketch of a camera-based ASVI using a commercially available smartphone is shown in Figure 1.2 to demonstrate a possible application scenario of the presented content. A camera as well as earphones or a hearing aid to provide Text To Speech (TTS) output are connected to the smartphone. Computationally expensive image processing calculations are exported to a cloud service and relevant external information that can support image detection (e. g. Global Positioning System (GPS) coordinates of crosswalks or construction sites) is extracted and provided through a cloud module. Due to the diversity of visually impaired people, it is important to take profiling and personalization into account when building an assistive system. This is achieved by using the Ambient Assisted Living (AAL) platform described in [4].

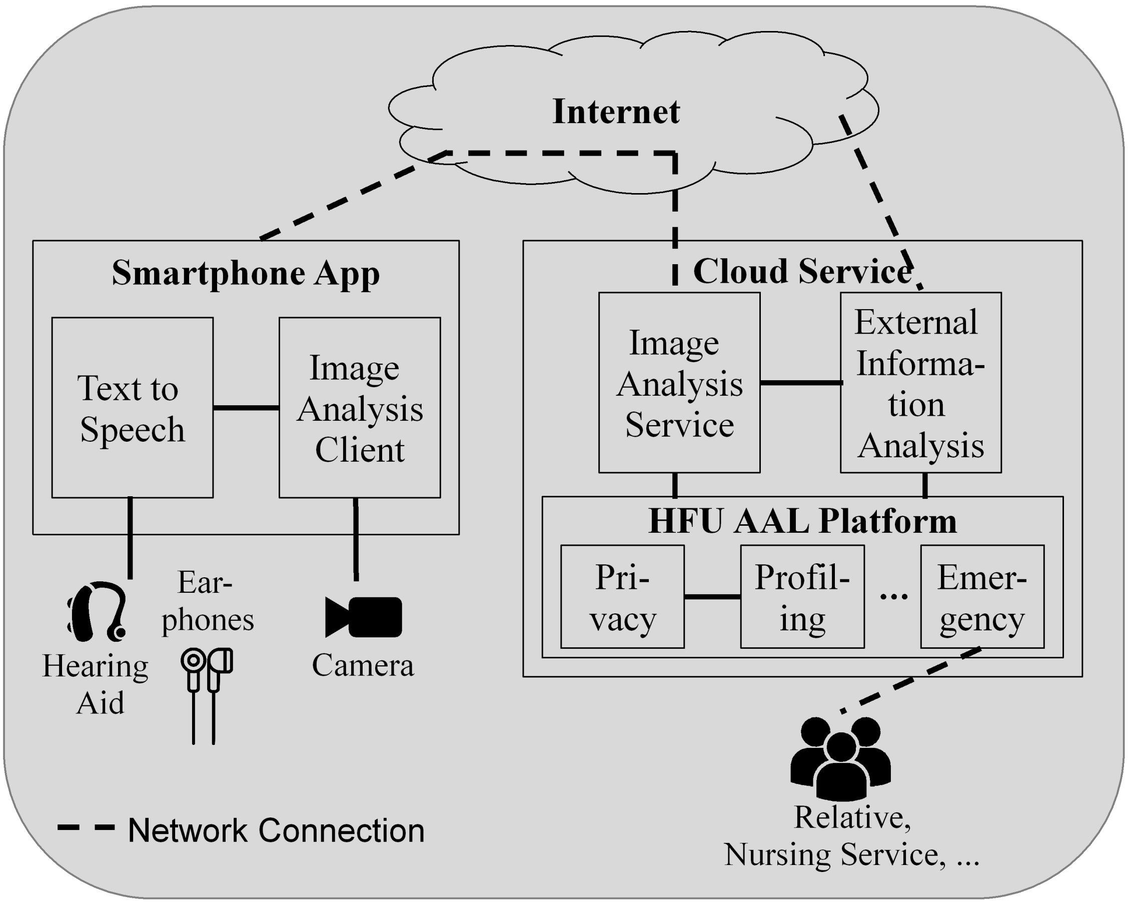


Figure 1.2: Sketch for a Camera-Based Mobile Assistive System

Chapter 2

Literature Review and Novelty

This chapter first presents a literature review of the two fields that are of relevance for the formulation of the transfer concept: Camera-based assistance of the visually impaired as well as camera-based driver assistance. From that, the need for and novelty of a transfer concept is derived.

2.1 Camera-Based ASVI

The use of cameras in conjunction with intelligent image processing and recognition algorithms is a widespread approach to assisting visually impaired people in everyday life. The spectrum ranges from a suitable representation of images (e. g. Oxsight[1]) on the interpretation of image content in certain situations (e. g. crosswalk detection [5]) to sensory substitution devices that transfer the visual perception to a different sense (e. g. *Sound of Vision* [6]). Thereby, different technologies are used: from smartphone applications to mobile assistive systems developed for a specific application. I present an overview of existing solutions divided according to different application areas in 2.1.2. Before that, I describe the general composition of ASVI that I developed from the systems described in the literature in 2.1.1. This section's statements are based on a translatation of my publication [7].

2.1.1 System Design of ASVI

Figure 2.1 shows the system design of ASVI I created based on a literature review. Abstract function modules that describe a definable partial functionality of the entire system - without making any assumptions about the implementation in hardware or software - are used. In general, ASVI consist of (a) an *Image Capture*, (b) an *Image Processing*, and (c) an *Output Unit*.

(a) The *Image Capture Unit* contains a video camera or image sensor that captures the present scene. The *Capture & Preprocessing* module decomposes the data stream into individual frames and adjusts for example brightness, contrast, colour space, and other parameters. If the image acquisition and image processing units are spatially separated from one another, the adapted image signals must be suitably transmitted to the image processing unit. Due to the high

[1] `https://www.oxsight.co.uk/`, accessed on June 6, 2020

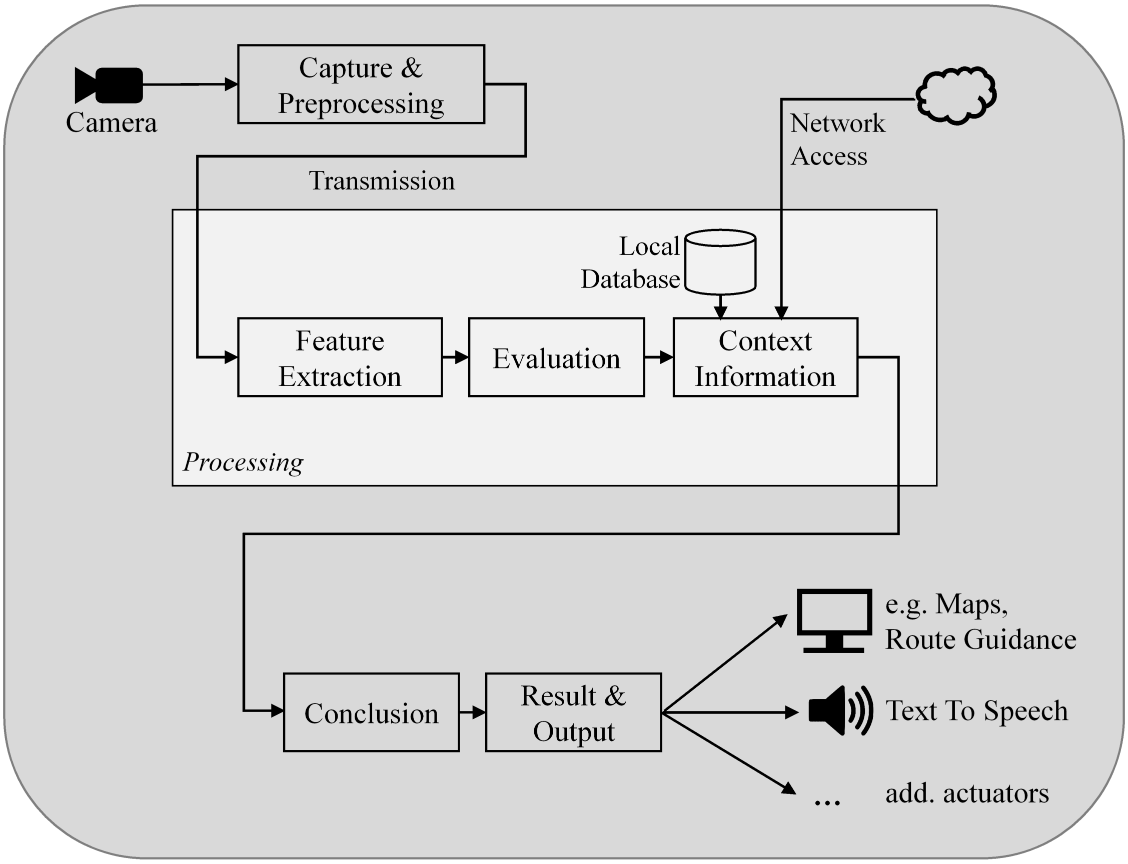


Figure 2.1: System Design of ASVI

bandwidth requirements of raw video data, this first requires video encoding for compression.

(b) In the *Image Processing Unit*, if necessary, the received video data are decoded and the individual original frames are restored. The function module *Feature Extraction* analyses the image content of a single frame. This can involve the application of basic transformations to identify edges, coherent objects, or areas of uniform colour. The result is passed on to the next stage of the image processing pipeline. In the function module *Evaluation*, a first interpretation of the frame content is made on the basis of its extracted features. By comparison with predecessor frames, motion vectors of objects can be identified, both absolutely in the world coordinate system as well as relative to one another, and their respective speeds can be determined.

Depending on the application, in the next step, relevant *Context Information* retrieved via a local database or network access is used to provide further information or to either substantiate or discard previously established hypotheses.

(c) In the *Output Unit*, a user-relevant result is generated from the previous calculations and resulting hypotheses by drawing an appropriate and possibly context-dependent *Conclusion*. For example, in a navigation application, a route guidance is calculated to guide the user to a detected crosswalk. Communicating with the user in the last step requires the generation of a multimodal output which can be visually displayed in a way suitable for the visually impaired, audible via the audio channel, or tactile via alternative feedback devices (Force Feedback).

2.1.2 Application Areas of ASVI

I clustered the literature about camera-based ASVI into the four application areas: Reading out text, recognizing faces and objects, perceiving the environment, and navigation and collision avoidance. Further applications are summarized in the final paragraph.

Reading Out Text

There are numerous smartphone apps that assist the visually impaired with the perception of texts by analysing a recorded image and creating an according TTS output (e. g. KFNB Reader[2], ABBYY Textgrabber[3]). In addition, there are mobile assistive systems that record text from a running camera and give a real time TTS feedback to the user. Examples are *OrCam's* commercially available system *MyEye*[4] and *FingerReader* [8]. The system *MyEye* consists of a camera mounted on the user's glasses that is connected to a small computation unit. The user points in the direction of the text which is then submitted via headphones. *FingerReader* consists of a small camera which is attached to the index finger like a ring and connected to a computer. The user moves their index finger along a line of text that is rendered to synthesized speech. With the help of motors, the user receives tactile feedback about in which direction the finger must be moved so that the line can be correctly detected by the camera and further processed by the system.

Recognizing Faces and Objects

Numerous applications and assistive systems allow the user to recognize previously stored faces and objects. The above described *MyEye* for example informs the user if there is a person known to them in the vision field of the camera. With the object recognition it is possible, amongst other things, to find the desired product in a supermarket. Other applications, however, are designed to analyse the facial expression and movements of a conversation partner and communicate it to the user.

The system *Expression* [9] is based on Google Glass, while *iFEPS* (improved Facial Expression Perception through Sound) [10] and *iMAPS* (interactive Mobile Affect Perception System) [11] are Android applications with a server connection to a computer. *Expression* submits the counterpart's head movements (looks to the right/left/up/down), facial expressions (e. g. laughter), and behaviours (e. g. yawning) to the user via TTS.

The system *iMAPS*, on the other hand, analyses the conversation partner according to the dimensional approach [12] which defines the range of human sensations during a conversation through the three values valence, arousal, and dominance. The type of feedback to the user has yet to be determined.

[2] `https://knfbreader.com/`, accessed on June 6, 2020
[3] `http://www.textgrabber.pro/en/`, accessed on June 6, 2020
[4] `https://www.orcam.com/en/`, accessed on June 6, 2020

The system *iFEPS* is based on the Facial Acting Coding System [13] which defines facial expressions as a combination of motion units. The movements of eyebrows, eyes, and lips are encoded to sound and transmitted to the user.

The purpose of the *Listen2dRoom* project [14] is to help visually impaired users to understand the layout of an unfamiliar space by detecting certain objects in the room, such as tables, chairs, and trash cans, with the use of a smartphone or portable camera and by communicating the detected objects via TTS. The order of the output allows to draw conclusions about the objects' arrangement in the room.

Furthermore, a variety of smartphone applications provide assistance in recognizing objects (e. g. *LookTel Recognizer*[5]) or detecting special objects such as banknotes (e. g. *LookTel Money Reader*[6]). The app *TapTapSee*[7] describes images with arbitrary content but, as can be read in the privacy policy, *TapTapSee* uses both, an algorithmic analysis of the image as well as crowdsourcing.

There is some research in the detection of traffic-related objects. For the crossing of roads, [5, 15, 16] describe the detection of crosswalks. In doing so, it is important to distinguish between crosswalks and other objects consisting of parallel lines, such as staircases; [15] and [16] address and handle this problem. Information about the according algorithms is given in section 6.3.1. Another way of safely crossing the road are traffic lights. In [17] similar shapes are used in order to detect traffic lights for the visually impaired.

Perceiving the Environment

While many assistive applications for the visually impaired focus on specific topics, there is a variety of systems that follow a more comprehensive approach and try to make the environment as a whole perceivable for the user. In most cases, so-called Sensory Substitution Device(s) (SSD) are developed, which convert perceptions so that they can be detected by a sense other than the impaired one.

In 1992, Meijer created the basis for *vOICe*, a SSD that converts camera images into audio signals [18]. The software is now available free of charge for several platforms and can be used on standard hardware equipped with a camera[8]. The project *Sound of Vision* [6] is an example for a modern SSD that provides tactile feedback in addition to auditory feedback.

Since the sense of hearing is of great importance to perceive the environment, especially for visually impaired people, some researchers concentrate on relieving the sense of hearing by transferring perception to the sense of touch. Due to its sensitivity that allows a differentiated perception, the tongue can be used for SSD. For this purpose, images are recorded with a camera, further processed, and finally converted into electronic impulses which are sent to an electrode matrix located on the tongue. The foundations for this procedure were first presented in [19]; recent research based

[5] `http://www.looktel.com/recognizer`, accessed on June 6, 2020
[6] `http://www.looktel.com/moneyreader`, accessed on June 6, 2020
[7] `https://taptapseeapp.com/`, accessed on June 6, 2020
[8] `https://www.seeingwithsound.com/`, accessed on June 6, 2020

on that can for example be found in [20] and [21]. Other research translates the processed image material onto belts equipped with several small motors (e. g. [22]), tactile displays (e. g. [23]), or transmits the signal to the fingertips (e. g. [24]). Instead of making the environment perceivable through a sense other than the sense of sight, it is possible to optimize the representation of objects in the user's vicinity by means of image enhancement to adapt them to the residual vision of non-blind visually impaired people (e. g. Oxsight[9]).

Navigating and Avoiding Collision

The primary technology used for navigation is GPS (e. g. the smartphone apps *LookTel Breadcrumbs*[10] and *Blindsquare*[11]), but GPS has limitations in close range. In addition to the traditional assistive systems, white cane and guide dog, ultrasound is often used to detect obstacles and prevent collisions [25]. Besides, SSD contribute to navigation and collision avoidance.

With the continuous development of image recognition and computer vision, more camera systems are being used to enhance GPS-based navigation as well as obstacle detection respectively collision avoidance. The project *IMAGO* [26] expands and refines GPS-based pedestrian navigation using image recognition techniques. Current images of the user's smartphone camera are compared with previously recorded images of the same route. In contrast, the project *Crosswatch* [27] uses satellite images from Google Maps and compares them with user-made 360-degree panoramic images to pinpoint the user's location.

In the area of obstacle detection and collision avoidance, some research aims to improve the already existing white cane by equipping it with additional sensors such as infrared and ultrasound but also cameras (e. g. [28]). Further research concentrates on replacing the guide dog with robots that have the appropriate skills. In [29], a commercially available smartphone camera is used to detect and process obstacles and traffic information.

Systems that are independant of the traditional assistive systems can be sorted into three categories. In the first case, motion estimation with optical flow is applied to data from a 2D camera, often simple smartphone cameras (e. g. [30, 31, 32]). The second case processes data from RGB-D cameras that provide depth information in addition to RGB colour data (e. g. [15, 33]), and in the third case, stereovision is used (e. g. [34, 35]). Especially important for the visually impaired is the detection of elevated objects and of stairs leading downwards, as they are generally difficult or impossible to detect with the white cane. The works [15] and [34] address these topics.

Further Areas of Application

In the following, I present further applications and assistive systems that cannot be assigned to one of the before described application areas.

[9] `https://www.oxsight.co.uk/`, accessed on June 6, 2020
[10] `http://www.looktel.com/breadcrumbs`, accessed on June 6, 2020
[11] `http://www.blindsquare.com/`, accessed on June 6, 2020

An according scanner app (e. g. [36]) can read barcodes and thus identify the corresponding products, while other smartphone apps concentrate on the identification of colours and lighting conditions (e. g. *ColorSay*[12]). The app *BeMyEyes*[13] follows a different approach by not using computer vision but connecting visually impaired people via video chat to a volunteer sighted person who can then help out with whatever problem there might be.

An assistive system consisting of a camera, a computer, a microphone, and earphones which informs users of colours and patterns of clothing is described in [37]. There are eleven different colours as well as the patterns plaid, striped, patternless, and irregular. When recording videos with portable cameras which are often used in ASVI, blurred frames occur frequently. This considerably impedes the extraction of features. Therefore, an algorithm that distinguishes which frames of a video stream are out of focus and which are not is presented in [38]. Only the frames with good quality are processed further; if several blurred frames follow each other, a de-blurring algorithm is used.

For a comparison of the presented apps and assistive systems concerning composition, prize and availability, usability as well as functionality and performance, I refer to my publication [7].

2.2 Camera-Based ADAS

Analogous to the examination of ASVI, I give details about the composition of camera-based ADAS before analysing the according application areas. The following paragraphs are based on my publication [39].

2.2.1 System Design of ADAS

In [40], Ranft and Stiller give an overview of computer vision techniques used for intelligent vehicles, including ADAS and autonomous driving. The three main, subsequent steps are *Early Vision*, *Maps and Localization*, and *Recognition and Classification*. The following explanations summarize the information provided in [40].

Early Vision refers to methods that are usually used in the first step of the computer vision pipeline leading to intermediate results [41]. Ranft and Stiller differentiate between methods used for single, dual, or multi frame systems and they assume that lens distortions are already corrected.

Early Vision methods for single frames include, but are not limited to, edge detection based on gradients (e. g. [42]), the estimation of vanishing points through intersecting lines (e. g. [43]), the detection of shapes and patterns (e. g. circular and triangular shapes indicate traffic signs [44]), and the reduction of noise by Gaussian filters or non-linear transforms (e. g. the bilateral filter [45]). For dual frames, stereo vision and

[12] `https://www.whitemarten.com/en/colorsay`, accessed on June 6, 2020
[13] `https://www.bemyeyes.com/`, accessed on June 6, 2020

optical flow are the major *Early Vision* methods. They both aim at finding corresponding features in image pairs. Stereo vision uses the fact that the distance between the two cameras is known and that the scene is captured at the exact same time in order to compute the scene's depth. However, optical flow provides information about the velocity and motion of objects by finding corresponding features in a time-shifted image pair [40].

For multi frames, the evaluation of image quadruples consisting of two pairs of stereo image pairs, the computation of scene flow, and 6-D vision (e. g. [46]) are considered. Whereas scene flow results in a motion field showing the 3-D velocity, 6-D vision additionally compensates ego-motion in order to specify results with respect to world instead of camera coordinates. These techniques can be extended to longer sequences of frames in order to resolve ambiguity problems and to remove outliers occurring in the computation of optical flow and disparity [40].

The second step in Ranft and Stiller's model, following *Early Vision*, is *Maps and Localization*. Visual localization is performed by comparing existing aerial images with images taken by on-board cameras (e. g. [47]). With the goal of further improving the accuracy, 3-D maps consisting of three layers are created (e. g. [48]). The localization layer contains a 6-D pose (3-D position, 3-D orientation), the tactical layer includes lane geometry as well as traffic rules and signs, and the dynamic layer holds information on static and moving objects.

The third step is dedicated to *Recognition and Classification*. For the recognition of specified objects in larger distances, appearance cues such as symmetry, shadows, local texture, and gradients are used (e. g. the vehicle detection described in [49]). Closer objects can be identified by examining the 3-D scene geometry (e. g. the lane and obstacle detection presented in [50]), especially the disparity, and by grouping image regions with similar optical flow motion vectors (e. g. the obstacle detection proposed in [51]). With the use of segmentation and classification methods, it is possible to assign labels such as road, person, or vehicle to detected objects. This procedure is known as semantic labelling. It is usually done either with classical methods like Support Vector Machine (SVM) [52] and conditional random fields [53] or with modern Convolutional Neural Network (CNN) [54].

Loce et al.'s book about computer vision and imaging in intelligent transportation systems [55] presents in its introduction a general computer vision pipeline for transportation applications which include ADAS (see Figure 2.2). For video capture in ADAS, there are cameras positioned outside the vehicle to capture the environment and cameras positioned inside the vehicle to monitor the driver. Furthermore, transmission of the video data to the processing units has to be considered in this context. After *Preprocessing*, analogous to preprocessing in ASVI respectively *Early Vision* in the previous paragraphs, a *Feature Extraction* suitable for the respective problem is performed. Based on the features and on further context information, such as models obtained from training data or information from other sensors, a *Decision/Inference* is made and communicated appropriately. In the case of Neural Networks (NN), the blocks *Feature Extraction* and *Decision/Inference* are one logical unit.

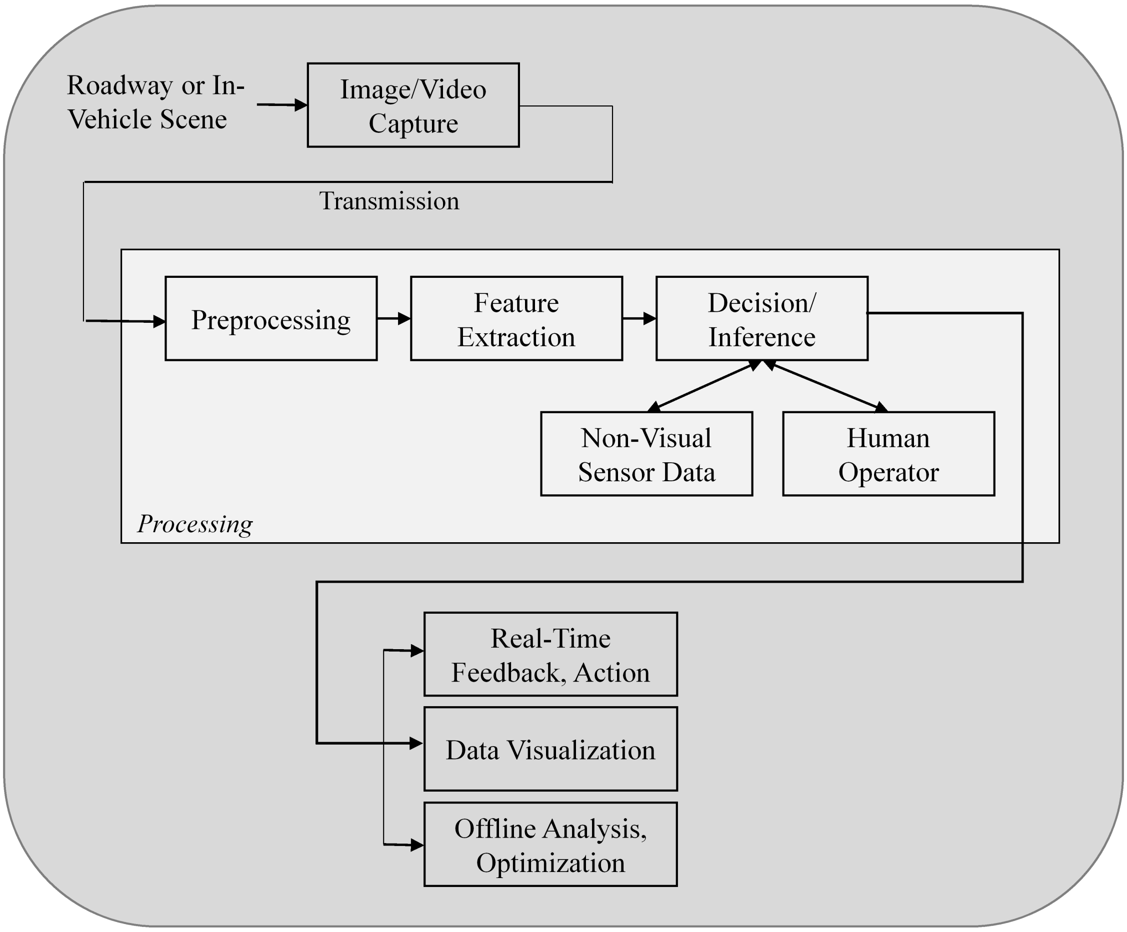


Figure 2.2: Computer Vision Pipeline for Transportation Applications. Adapted from [55].

2.2.2 Application areas of ADAS

In the following, I present an overview of application areas in ADAS. I will describe ADAS solutions for different uses cases in more detail in sections 4.2.3, 6.2.1, 6.3.1, and 6.4.1.

In the area of ADAS, there is already a large number of camera-based assistance functions that are used in latest car models. They include the detection of traffic signs [56] and road lanes [57] as well as collision avoidance [58]. In addition to traditional algorithms, modern NN are used in driver assistance algorithms [59].

Further research in the detection of relevant traffic objects, such as crosswalks, traffic lights, and pedestrians, is being conducted. In [60], Chiu et al. use the Hough transform and a comparison with test images for a real-time traffic light detection. However, Choi et al. present in [61] a combined detection of traffic lights and crosswalks based on the HSV colour space. For the detection of pedestrians, Benenson et al. give an overview of and compare the best known existing algorithms [62].

Besides, the monitoring of the driver and its condition has become a large field of research for ADAS. Yan et al. for example describe a real-time drowsiness detection system based on the percentage of eyelid closure over the pupil over time which is computed by means of image processing [63].

2.3 Novelty

Based on the presented literature review, this section shows that there is no assistive system offering all support visually impaired pedestrians need in traffic situations. Furthermore, the section explains why using and adapting ADAS algorithms to ASVI is a promising idea based on what was learned from the literature. That transferring algorithms from ADAS to ASVI is indeed possible will be addressed and shown in chapter 6.

In the course of this chapter, I presented the computer vision pipeline of ADAS according to [55], the steps of computer vision in ADAS according to [40], and the general system design of ASVI that I developed on the basis of a literature review of camera-based ASVI. Even though different terminologies are used, three phases are in the center of data processing in all described cases: Preprocessing/Early vision, feature extraction/recognition, and evaluation/classification/decision. Maps and localization is only mentioned in [40], but a similar procedure can be observed in the ASVI project *Crosswatch* [27]. Additionally, context information is used in ADAS as well as ASVI in different forms, data transmission to the processing unit has to be addressed, and appropriate ways of communicating the results have to be developed. In summary it can be stated that the overall system design, especially the computer vision steps, of camera-based ADAS and ASVI are similar.

Furthermore, the identified application areas of ADAS and ASVI reveal that there is in general an overlap in addressed and needed use cases, e. g. collision avoidance and support at crossings. At the same time, the amount of research in the field of ADAS is much higher than in ASVI, with not only research institutions but also major players such as Google, Bosch, or Tesla involved. Furthermore, companies like Intel, Nvidia, or Mobileye are well-known for their Machine Learning (ML) solutions in the field of ADAS. In addition to the amount of research, the quality of detection algorithms used in latest car models has to be very high to ensure the safety of the driver and other road users. It is therefore desirable to make latest and future advancements in ADAS applicable for visually impaired people.

Hence, it follows that using ADAS algorithms in an ASVI is possible because of the similar compositions of such systems and because of the overlap of relevant use cases. The transfer concept is necessary because ADAS algorithms can in general not be used for pedestrians without any adaptations. This is mainly due to a priori assumptions, such as known and stable camera position, that are only valid when used in cars.

Most of the ASVI described above perform a specific task or several tasks that can be assigned to one topic. These systems have the potential to take on a variety of functions based on their composition. However, this is usually not exploited since often only prototypes for a specific research purpose are built. An exception is the commercial system *MyEye* which in addition to its main task of recognizing and reading text can recognize objects and faces; further applications are under development. Naturally, the SSD have no other functions because of their approach of transferring the visual perception to other senses. Some of the introduced systems cover single traffic topics such as crossings (e. g. [27]), but there is no system or research project that covers the needs of visually impaired people in all significant traffic scenarios by camera-based detection.

Therefore, the research presented in this thesis leads the way to provide comprehensive assistance in all traffic scenarios that are of relevance for the visually impaired based on the evaluation of camera footage.

Chapter 3

Research Categories and Objectives

This chapter gives an overview of the three research categories treated in the thesis. An exploration of traffic scenarios with relevance to the visually impaired, see section 3.1, results in a complete collection of vision use cases that could support visually impaired pedestrians. The overlap of these vision use cases with the ones addressed in ADAS needs to be considered in the following. First, the video data set *CoPeD* containing comparable video sequences from pedestrian and driver perspective for the identified overlapping use cases is created, see section 3.2. These data are used in order to evaluate the algorithms that are adapted from ADAS to ASVI, see section 3.3.

In the following, I describe the objectives of each research category that will be discussed in the course of this thesis. I furthermore name used methods and tools.

3.1 Definition of Traffic Scenarios and Vision Use Cases for the Visually Impaired

Objectives

This category's purpose is the understanding of needs visually impaired people have as pedestrians in traffic situations. From the gathered insights, a list with relevant vision use cases has to be acquired and the overlap with vision use cases addressed in ADAS has to be built.

In detail, the following objectives have to be achieved throughout the research in this category. The according results will be summarized in Thesis 1 in section 4.4.

(O1.1) All traffic scenarios that are of interest for visually impaired pedestrians have to be defined.

(O1.2) All vision use cases that can support the visually impaired in traffic situations have to determined.

(O1.3) The overlap of vision use cases addressed in ADAS and needed in ASVI has to be determined.

(O1.4) The idea of using (adapted) software engineering methods to cluster and present qualitative data has to be introduced.

Besides, the research in this category is expected to answer the following questions:

(Q.a) Are there differences in gender and/or age of visually impaired people when it comes to dealing with traffic scenarios?

(Q.b) Is the use of technology common among visually impaired people?

(Q.c) How do visually impaired people prepare for a trip to an unknown address?

(Q.d) Are visually impaired people comfortable with having to ask for support or directions?

(Q.e) Which identified vision use cases are the most important?

Methods and Tools

To achieve these goals, I create, conduct, and evaluate a qualitative interview study consisting of expert interviews and interviews with MTG, namely visually impaired pedestrians. I use Witzel's problem-centered method [64] and Meuser and Nagel's notes on expert interviews [65]. Transcription and analysis are performed with the software MAXQDA Version 12 [66]. I code the interviews with inductively developed codes as proposed by Mayring [67].

By clustering the data, I determine different traffic scenarios and according vision use cases that can support the visually impaired in the respective scenario. I summarize the evaluation of the interview study in scenario tables adapted from software engineering [68]. By comparing the collection of ASVI use cases with an ADAS literature review, I determine the desired overlap.

3.2 The *CoPeD* Data Set for Traffic Scenarios

Objectives

This category addresses the acquisition of video data that are needed to compare ADAS algorithms with their ASVI adaptations that I will develop in the next research category. For the evaluation of these algorithms, comparable video data from both perspectives, driver and pedestrian, have to be gathered. It is important that the video data cover all identified overlapping vision use cases from ADAS and ASVI. Although there are numerous data sets covering traffic scenarios, they are mostly from driver perspective and no according comparable data from driver and pedestrians perspective exists. Therefore, I create the *CoPeD* data set for traffic scenarios. The data set is made publicly available and others are permitted to use, distribute, and modify the data.

In detail, the following objective has to be achieved throughout the research in this category. The according results will be summarized in Thesis 2 in section 5.4.

(O2) The data set *CoPeD* containing comparable video data from driver and pedestrian perspective and covering the overlapping use cases from ADAS and ASVI has to be created.

Methods and Tools

Review and analysis of according scientific literature reveal that no comparable data exist. Therefore, I create and publish the *CoPeD* data set for traffic scenarios. For the planning of the data set, I use activity diagrams from software engineering [68]. The sequences are filmed in High Definition (HD) with a *Kodak PIXPRO SP360 4K* camera.

3.3 Use Case Examination

Objectives

The overlapping use cases have to be examined concerning their possibilities of adaptation from ADAS to ASVI. For each use case, appropriate algorithms from ADAS have to be chosen and adapted algorithms have to be developed. It is important to show that the adapted algorithms perform at least as good as the underlying ADAS algorithms so that they are applicable in an assistive system.

In detail, the following objectives have to be achieved throughout the research in this category. The according results will be summarized in Thesis 3 in section 6.7.

(O3.1) It has to be shown that determining the Region Of Interest (ROI) for ASVI detection algorithms can in general not be taken from ADAS and that adapting a RBS from ADAS to ASVI solves this problem.

(O3.2) Adaptations of algorithms from ADAS to ASVI have to developed and implemented. The adapted algorithms have to achieve similar hit rates as the underlying ADAS algorithms.

Methods and Tools

I develop adaptations to use ADAS algorithms in ASVI and implement the adapted algorithms in Matlab Version R2017b [69]. Afterwards, they are evaluated on several sequences from the *CoPeD* data set.

Chapter 4

Definition of Traffic Scenarios and Vision Use Cases for the Visually Impaired

In order to get insights into the difficulties visually impaired pedestrians face and to reach the objectives introduced in section 3.1, I design, conduct, and evaluate a qualitative interview study with experts and MTG. First, design and conduction of the study including information on related work, methods, data collection and analysis, and participants are described. In the following, the evaluation containing social insights, traffic and non-traffic scenarios as well as use case importance is presented. After a conclusion of the qualitative interview study, the before formulated objectives for this research category are discussed. Aside from the last section, this chapter is based on my publication [70].

4.1 Qualitative Interview Study: Design and Conduction

4.1.1 Related Work and Methods

The project *CrossingGuard* [71] examines information requirements of the visually impaired but is limited to crossings and does not take the use of camera information into account. From the results of a formative study with four visually impaired individuals and two specialists for orientation and mobility, Guy and Truong develop the system *CrossingGuard* that delivers "sidewalk to sidewalk" [71] directions at crossings. Afterwards, they test their system by means of a user test with ten visually impaired participants.

Quiñones et al. [72] present a study concerning the needs in navigation of visually impaired people focusing on localization via GPS. In 20 semi-structured interviews with participants from the U.S. and South Korea, they discuss routine, infrequent, and unfamiliar routes. They go deeper into wayfinding techniques and the differences between known and unknown routes than the presented study, but they do not discuss which camera-based object detection algorithms could provide support during navigation. As in contrast to my research, Quiñones et al. do not intend to use a camera in their GPS system, both studies, theirs and mine, are tailored to the specific problem the research groups address and therefore complement each other.

The goals of my study are to create an overview of relevant traffic scenarios for the visually impaired and to collect the corresponding vision use cases that each represent a specific camera-based object detection. Furthermore, I state which of the identified use cases are also of relevance in ADAS. The overlap from both fields then has to be examined for the development of the transfer concept.

Before interviewing MTG, I conducted expert interviews [65], whose results are presented in [73] and [74]. As the results needed to be extended by the new findings from the interviews with target group members, I present a common evaluation of both interview types here and in [70].

Creswell and Creswell [75] define four different worldviews (Postpositivism, Constructivism, Transformative, Pragmatism), each as a "general philosophical orientation about the world and the nature of research that a researcher brings to a study" [75, p. 5]. I see myself as a representative of the pragmatic worldview. Pragmatism includes that "researchers are free to choose the methods, techniques, and procedures of research that best meet their needs and purposes" [75, p. 10]. For this study, I chose to conduct qualitative interviews because their narrative nature allows me insights into the daily life of the interviewees themselves and, in the case of experts, the people they are in contact with, while simultaneously leading to the collection of traffic scenarios and vision use cases I need for my research. Additionally, I decided to summarize the results by means of software engineering in order to make them easily accessible for developers from the field of computer science.

Creswell and Creswell [75] state further that pragmatists, and in a similar way mixed method researchers, "look to many approaches for collecting and analysing data rather than subscribing to only one way," [75, p. 10]. Although I used exclusively qualitative methods in this study, I approached the problem from different perspectives by interviewing experts as well as MTG. At the same time, my study can be seen as the first phase of an "exploratory sequential" [75, p. 218] mixed method design. Although, I will not pursue this in the near future because it is not needed for the success of my research towards a transfer concept from ADAS to ASVI, the results of this study can be used as the foundation for quantitative studies. Based on the traffic scenarios and vision use cases I collected and described, it is e. g. possible to create a quantitative study that explores the correlation between the nature and degree of a visual impairment and needed vision use cases in traffic scenarios.

The word "problem-centered" [75, p. 6] is used as one of four dimensions to outline the pragmatic worldview. As a researcher with strong connections to applied computer science, my research usually evolves around specific problems, in this case around the question of how a camera-based assistive system could support visually impaired people in traffic situations. This is in accordance with Patton [76] who states that looking for solutions to problems is central to the pragmatic approach. Not only the research question itself is problem-centered; I also used a problem-centered method [64] to design the interviews I conducted.

Witzel's problem-centered method [64] is a semi-structured interview form in which the guideline is handled flexibly so that the interview can turn into a conversation between interviewer and interviewee. In addition to the guideline as a structuring framework for the interview, Witzel names three more required instruments: A short questionnaire to gather basic information about the interviewee, a recording for later transcription, and a postscript to write down non-verbal aspects of the conversation and spontaneous ideas for the evaluation.

I apply this method to two different sets of interviews: The expert interviews concentrate on accessing the "contextual knowledge" [65] the interviewees have gained concerning the visually impaired, whereas the interviews with MTG focus on the interviewee's personal experiences. I define experts as persons who are regularly, through voluntary or professional work, in contact with a diverse group of visually impaired people concerning age, gender, and impairment. An own visual impairment is possible but not a requirement. A person is considered a MTG if they have a visual impairment and are frequent road users.

4.1.2 Interview Guidelines

Both interview types essentially followed the same guideline. After going through the privacy policy, the interviewer gathered basic information about the interviewee in the short questionnaire. Then, traffic situations were discussed and after finishing the interview, the interviewer wrote a postscript.

In the privacy policy part, the interviewees agreed to recording and transcription of the interview and were ensured that their identity will not be revealed throughout the research. Furthermore, they were informed about the possibility to end the interview at any time and withdraw their consent to recording and transcription.

The short questionnaire as well as the discussion of traffic scenarios differed for the two interview types. In the short questionnaire, expert interviewees were asked about their age, gender, own impairment, profession, and if their expert work with visually impaired people is voluntary or professional. In addition, the interviewer asked about age and gender distribution, kinds of visual impairments, and affinity to technology of the visually impaired persons the experts are regularly in contact with. For the interviews with MTG, the focus was on personal characteristics: age, gender, kind of impairment as well as use of smartphones and computers.

To discuss traffic scenarios, the interviewer asked the experts about the three biggest challenges visually impaired pedestrians face and if there are differences in the problems for people of different age, gender, and degree of impairment. Starting to talk about the three biggest challenges usually resulted in the discussion of further problems. For the case that the conversation stopped, a prepared list with traffic situations helped to give impulses to the interviewee. Towards the end of the interview, the interviewer asked about the differences between the visually impaired and the sighted when it comes to the preparation for a trip to an unknown address.

For the interviews with MTG, I chose a different approach: The interviewees were given a concrete situation involving traffic challenges and had to talk the interviewer through the process of solving this situation while keeping in mind to mention when and how a camera-based assistive system could provide support. They were told to imagine a system that has no limitations in the identification and communication of objects captured by the camera. The four discussed situations were:

(a) Familiar route, familiar surroundings:
You want to go to a concert that takes place in your home city/village. You know the concert hall and have already been to concerts there.

(b) Unfamiliar route, familiar surroundings:
You have a doctor's appointment in your home city/village. You do not know the doctor and have never been to their office.

(c) Familiar route, unfamiliar surroundings:
You want to visit a friend who lives in another city/village. You have known the friend for a long time and have visited them frequently.

(d) Unfamiliar route, unfamiliar surroundings:
You want to travel to a city in which you have not yet been because there is an event that interests you.

After the first interviews, I observed that the discussion of four problems took too long and produced a lot of repetitions. The interviews were then reduced to the discussion of two topics, one in familiar and one in unfamiliar surroundings. It depended on what was learned to this point about the interviewee which ones in particular were chosen. For example, if the person already said that they do not travel unfamiliar routes, the familiar ones were discussed.

4.1.3 Data Collection and Analysis

The interviews were conducted in German language via phone. I translated the quotes presented in this thesis from German to English. Originally, the interviews were planned to take place in person, but while planning the interview dates and locations with the interviewees, it became clear that travelling to a meeting location, even close to their home, meant a great effort for some interviewees. Therefore, two interviews were scheduled via phone to see if the course of the interviews met the expectations without meeting in person. As this was the case, I decided to conduct all interviews via phone and thus reduced the effort for both, interviewer and interviewee. All interviewees live in Germany meaning that their statements refer to German traffic situations and not all results may be transferable to other countries. The experts come from different parts of the country, whereas the MTG all live in the south west of Germany in rural as well as urban areas.

For transcription as well as data analysis and evaluation, I used the software MAXQDA Version 12 [66]. I defined codes in order to categorize the answers of the interviewees according to the six identified scenarios that I present in Figure 4.1. The codes were not predefined but developed inductively as proposed for example by Mayring [67]. Every time an interviewee talked about a new situation, another code was added. I then merged some of the codes, e. g. *Crosswalk*, *Traffic Light*, and others to *Crossing A Road*, in order to have a manageable number of codes. The three categories *Orientation*, *Pedestrian*, and *Public Transport* were added at the very end after reviewing the coding to provide a better overview.

As the subjects addressed in the interviews are social situations that can be clearly separated, I decided to have the coding done by one single person.

In the course of the expert interviews, data saturation was reached after three interviews; in the course of the interviews with MTG after six. To ensure data saturation, I conducted one more expert and four more interviews with MTG. Data saturation, in this case, is understood as the moment where further interviews did not lead to new insights concerning traffic scenarios and according vision use cases.

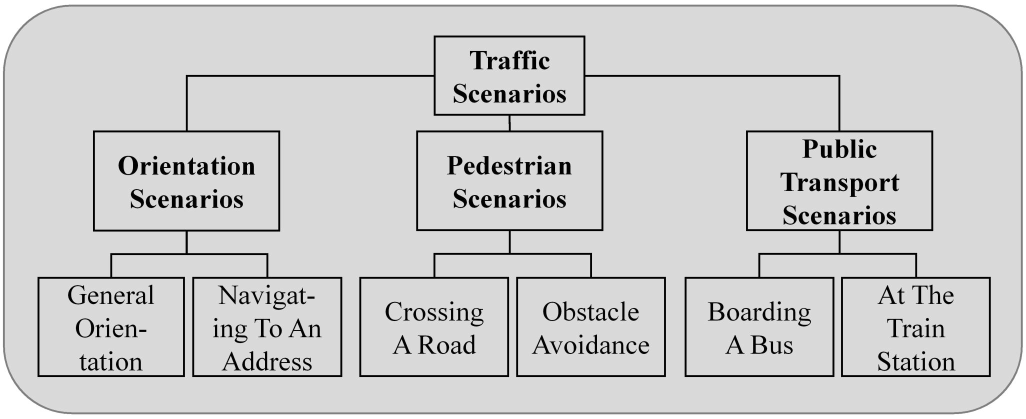


Figure 4.1: Hierarchy of the Identified Traffic Scenarios

4.1.4 Participants

Table 4.1 presents information about the interviewees of both interview types. Four experts, all being male and covering an age range from over 40 to over 70, were interviewed. Three of them are blind, whereas the fourth person has no visual impairment. Two are active members of interest groups who work as volunteers and the others are employees of educational institutes. The visually impaired people the experts work with cover a wide range of age, gender, and degree of impairment. Concerning age, the students of one of the educational institutes are generally not older than 19, but the interest groups consist mostly of older members, due to demographic reasons and the fact that visual impairments are often age-related. Furthermore, ten MTG were interviewed, three female and seven male, covering an age range from 43 to 76. Three are blind, whereas the others have residual vision. Two of the interviewees are gainfully employed, the others are in retirement. They all use at least smartphone or computer but mostly both. The predominant used operating systems are Apple's iOS for smartphones and Windows for computers.

4.2 Qualitative Interview Study: Evaluation

After discussing social insights I gained from both interview types, I present details about the traffic scenarios extracted from the interview data and analyse them concerning use cases that can be solved by computer vision and would facilitate the according traffic scenario.

Table 4.1: Characteristics of the Interviewees (EI: Expert Interviewee, MTG: Member of the Target Group)

ID	Age	Gender	Impairment
EI1	40-50	Male	Blind
EI2	40-50	Male	Sighted
EI3	70-80	Male	Blind
EI4	50-60	Male	Blind
MTG1	50-60	Female	Residual vision
MTG2	60-70	Female	Blind
MTG3	50-60	Female	Blind
MTG4	70-80	Male	Residual vision
MTG5	50-60	Male	Residual vision
MTG6	60-70	Male	Residual vision
MTG7	50-60	Male	Residual vision
MTG8	40-50	Male	Residual vision
MTG9	70-80	Male	Blind
MTG10	40-50	Male	Residual vision

Each scenario is then summarized in the form of tables inspired by software engineering. Sommerville [68] suggests to record scenarios in tables with the keywords: *Initial Assumption*, *Normal*, *What Can Go Wrong*, *Other Activities*, and *State Of Completion*. As the last two keywords are of no relevance in the presented research, I modify the table by deleting them. In the added first line, *Quotes*, the scenario is introduced by citing one of the interviewees to underline the importance of the respective scenario. The line *Normal* is used to explain the current procedure to solve the scenario and the line *What Can Go Wrong* to determine problems that can occur. In the added last line, *Vision Use Cases*, I record vision use cases which can be solved by means of computer vision derived from the line *What Can Go Wrong*. This form of scenario record, inspired by [68], results in a clustered overview of the needs of visually impaired people in traffic situations.

I present first versions of the six descriptive tables, one for each scenario, in [73] and [74]. Here, the final tables updated with the data from the interviews with MTG (see Table 4.2 to Table 4.7) are shown. The interviews with MTG brought new insights to the tables about *General Orientation*, *Crossing a Road*, and *Obstacle Avoidance*. The others remain unchanged in comparison with the ones presented in [73] and [74].

Afterwards, I mention some non-traffic scenarios that came up in the course of the interviews and present insights on the importance of the use cases collected through the interviews with the MTG.

4.2.1 Social Insights

The expert interviewees were asked about differences in gender and age when it comes to dealing with traffic situations. Only one EI, EI1, named a particularity concerning gender. In his experience, girls and young women are more likely to attend voluntary mobility workshops than boys and young men. The workshops' purpose is to provide additional advice and to pass on further knowledge beyond the mandatory mobility training. According to the EI, age is less important to solve problems in traffic situations than life experience with visual impairment. However, it has to be noted that an increased age often causes further limitations, e. g. in hearing and motor skills. The data from the target group interviews are insufficient in order to make a statement about differences in gender and age.

The experts attest to the community of the visually impaired a certain openness regarding the handling of technology ("When you have a limitation, you depend on technology and of course you use it," (EI4)). This is underlined by the fact that all interviewed MTG use at least PC or smartphone.

When visually impaired people prepare for a trip to an unknown address, they essentially cover the same topics as the sighted, but the amounts of needed information differ. One of the blind EI summarized it in the sentence "I just simply need more precise information" (EI4), but "if one does so [collect detailed information before going on a trip], surely also depends on the personality" (EI1), no matter if a person is sighted or visually impaired. According to the impressions of the EI, a minority of the visually impaired attempt to travel to an unknown address on their own. The results from the interviews with MTG confirm this. Many interviewees state that they do not travel alone before knowing a route. Mostly, it is difficult for these interviewees to imagine how and if an assistive system could change that. From my research, I can therefore not make a statement about if the use of an assistive system would encourage more visually impaired people to travel unknown routes on their own.

A topic often discussed in the course of the target group interviews is that visually impaired people frequently have to ask for support, e. g. ask the bus driver about bus number and direction. Whereas some interviewees say that they do not mind asking and like to be in contact with people ("Even if I know the way, I always let myself be helped. You then start a conversation and communicate with the people and I find that very important," (MTG2)), others report bad experiences such as unfriendly and false answers ("(...) because it has happened that I asked passers-by and they told me the wrong [bus] line," (MTG6)). For the latter group of people, an assistive system offering support and reducing the dependency on asking would significantly improve their daily life.

One EI pointed out that when discussing differences between the visually impaired and sighted, we have to keep in mind that "the blind and visually impaired are as different individuals as you and your colleagues," (EI4). Duckett and Pratt underline the importance of the acknowledgement of diversity when doing research for visually impaired people: "Participants were opposed to being clumped together in large groups of visually impaired people with whom all they shared was owning the same diagnostic label," [3, p. 827].

4.2.2 Traffic Scenarios

As shown in Figure 4.1, I extracted a total of six traffic scenarios from the expert interviews that can be clustered into the three categories: *Orientation*, *Pedestrian* and *Public Transport* scenarios. Each category contains two scenarios: *General Orientation* and *Navigating to an Address* are *Orientation* scenarios, whereas *Crossing a Road* and *Obstacle Avoidance* form the *Pedestrian* scenarios. The *Public Transport* scenarios consist of *Boarding a Bus* and *At the Train Station*. Scenarios for subways and trams were not defined because in terms of scenario description and according use cases they can be seen as a mixture of *Boarding a Bus* and *At the Train Station*.

General Orientation

It can be difficult for a visually impaired person to know their exact location and to be aware of their direction and surroundings. The two expert interviewees EI1 and EI3 explained this as follows:

> The main problem for a blind person is to have an inner idea of an unfamiliar way, (...), the psychologists call it a mental map. (...) That is of course relatively difficult for a blind person and maybe one can program some sort of exploration mode where the camera says "Ok, I'll just try to detect objects and describe the ones close to you." (EI1)
>
> (...) the orientation, when I'm in an unknown environment, I first have to know, have to ask the question, where am I? How can I cope there? (...) [It is] problematic in general to *keep an eye* on your destination. (...) You can simply get lost easily. (...) And this is a really burdening point for us. (EI3)

Special navigational apps such as *Blindsquare*[1] can help to keep track of direction and also surroundings, as they announce points of interest such as shops and restaurants. However, the navigational app can only announce places in the database; if the data are not maintained or if the person is in a remote area, there might not be much information: "*Blindsquare* announces for example partially which shops there are but not every shop. There, I would wish (...) that the camera identified the names [of the shops]," (MTG3). Furthermore, not every important detail is announced by the app, e. g. position of the curb ("The curb is very important. If you don't have that you can't orient yourself while walking down the street. The curbside is very important," (MTG2).), sign posts ("I could make my way from (...) to (...), I'd know the way, but I'd not be able to decipher signposts," (MTG7).), street name plates ("What I often don't recognize is the street name itself," (MTG5).), or traffic signs ("Or are there traffic signs that I need to be made aware of," (MTG9)?). The inaccuracy of GPS is another problem and will be discussed in the next scenario.

On a smaller scale, Tactile Ground Guidance System (TGGS) help with orientation, but in an unknown area a visually impaired person might not know if and where to find them and as they are not directed, it is possible to walk them in the wrong direction. Interviewee EI1 states:

> What can generally be a problem with TGGS is (...) that you know that there is something, but you need to find the guidance system first. In a sense, you would need a guidance system to the guidance system. (...) Let's say I as a blind hit the TGGS with the cane and know it guides me somewhere, but I don't know exactly where. (...) They are not oriented. They lead from A to B and from B to A. (...) There [at a station] is a hustle and you drift away from the TGGS, you get back there,

[1] `http://www.blindsquare.com/`, accessed on June 6, 2020

> and now the question is: Did I stupidly turn 180 degree while I sidestepped other people and am walking back now? That actually happened. (EI1)

If a person gets lost, they can ask passers-by or use the app *BeMyEyes*[2] which connects them via video chat to a seeing person who is willing to help. As stated in the section about social insights, asking passers-by can be unpleasant for some people. More details about *BeMyEyes* are given in the next scenario.

Based on this information, the following use cases would be helpful in this scenario: Describing the surroundings and lane detection (identifies the course of the road) to help form a "mental map" (EI1), analogous and digital displays to detect and read different signs (shop signs, signposts, street plates), curb(-sides), traffic signs with importance to pedestrians, and TGGS.

Table 4.2 summarizes the findings of the *General Orientation* scenario.

Navigating to an Address

If a visually impaired person wants to walk to a certain address that is not part of their routine ways, they use a GPS-based app in addition to aids used anyway such as cane or guide dog. However, due to accuracy, the GPS app can generally not lead them to the exact place of the entrance door. To find the right door and possibly the doorbell sign, they have to ask people who pass by or call someone:

> The blind-specific [apps]. (...) There is a well-known one called *Blindsquare* and it always warns you when the accuracy falls under a certain critical value. (...) And then, this has relatively little use for you, I'd say. Ok, you know, you are close, but without experimenting or asking someone, you really wouldn't find the door. (...) Doorbell signs are a typical camera task. There is already the app *BeMyEyes*, you might have heard of it before. It is a social network. I call some nice people and they get my camera picture and say "Yes, it's the third [doorbell] from the top". (EI1)

> We have a certain amount of uncertainty, but you get roughly informed that you are in the vicinity of this property and that's already half the rent. And then, of course, you move on with the white cane and (...) you have to make sure to (...) find an entrance somewhere. Then, you have to ask [someone], is this the right place or not? (EI3)

From this, it is extracted that it would be useful to automatically detect house numbers, doors, and text on doorbell signs. With the help of this information, it would be possible to compensate the inaccuracy of GPS-based systems and help navigate the user to the right door(-bell) without them having to ask for direction.

Table 4.3 summarizes the presented evaluation of the *Navigating to an Address* scenario.

[2] `https://www.bemyeyes.com/`, accessed on June 6, 2020

Table 4.2: General Orientation

Quote	"[It is] problematic in general to *keep an eye* on your destination. (...) You can simply get lost easily. (...) And this is a really burdening point for us," (EI3).
Initial Assumption	A visually impaired person wants to know where they are and be aware of their direction and surroundings in order not to get lost.
Normal	Navigational smartphone apps, e. g. *Blindsquare*, help the person to keep track of their direction and also their surroundings as the app can announce crossings, shops, restaurants, and such. On a smaller scale, TGGS help the person to find their way. If they get lost anyway, they can ask a passer-by or use the app *BeMyEyes* which connects them via video chat to a seeing person that is willing to help.
What Can Go Wrong	The navigational app can only announce places in the database. If the data are not maintained or if the person is in a remote area, there might not be enough information for the person to create a "mental map" (EI1) of their surroundings. Not every important detail of the surroundings can be announced by the app (e. g. size and height of the curb, course of the road, street names). Depending on the location of the person, GPS can be inaccurate. As TGGS are not directed, it is possible that they do not know, if they are walking it in the right direction. The person cannot find the TGGS even though it is there.
Vision Use Cases	TGGS Detection, Description of the surroundings, Traffic Sign Detection (e. g. to find pedestrian zones), Curb Information, Display Detection (to find shop signs or street name plates), Lane Detection

Crossing a Road

When a visually impaired person wants to cross a road, they first have to find a crosswalk or traffic light which can be hard in unfamiliar areas. Then, they have to make sure that it is safe to cross the road. Intermediate platforms can make them unsure how to proceed.

> Well, then I would say, it's exactly these three problems. Finding possibilities to cross the road, meaning finding traffic lights and crosswalks. The second would be to safely cross the street (...) which starts with knowing when the light is green and in which direction I have to go exactly and there's the question: If there is a traffic island, do I have to stop there or can I cross over? Is it still green on the other side? (EI1)

Table 4.3: Navigating to an Address

Quote	”You have to make sure to (...) find an entrance somewhere. Then, one has to ask [someone], is this the right place or not?” (EI3).
Initial Assumption	A visually impaired person wants to walk to a certain address.
Normal	The person enters the address into a GPS-based navigational smartphone app. The app leads them to the specified address.
What Can Go Wrong	Due to GPS accuracy, the navigational app cannot lead the person directly to the entrance of the building. If the building is unknown to them, they have to ask in order to get to the entrance and possibly find the right door bell.
Vision Use Cases	House Number Detection, Door Detection, Optical Character Recognition (OCR) (for doorbell signs)

TGGS and acoustic signals offer support for finding crosswalks respectively traffic lights but are not always available. Even when available, it can be hard to find the TGGS (as discussed before) or the traffic light and its pole to activate the acoustic signal: ”There, I often look for the traffic light to cross the street. I often don’t find it. Then, people tell me ”There it is”. And I always have to ask where ”there” is,” (MTG2); ”Then, it would be important that it [the camera] finds me traffic light poles. As a blind person (...), I always have to search a bit until I find the pole and can press the traffic light,” (MTG3).

If there is no support in the form of crosswalks or traffic lights, crossing a road involves risks: ”Crossing a road with high traffic frequency without any safeguarding is always a very big danger for a blind person,” (EI3). The visually impaired person has to rely on their hearing to identify a safe moment to cross the road which can be difficult, e. g. because ”Electric cars are a problem. (...) On the other side, it’s nice for the sighted, when the traffic flows quieter. For the sighted, this is quite good, but for us blind people, it’s really a problem,” (MTG3).

No matter if there are traffic lights, crosswalks, or no support, curbs can be a problem for a visually impaired person who is crossing a road: ”One time, I did not have my white cane with me and the tram (...) had its exit on a higher curb. And I crossed the road and stumbled on the curb,” (MTG7).

To help find a crosswalk, a camera could detect the TGGS leading there, the crosswalk itself and/or the according traffic sign. For traffic lights, it is important to detect the traffic light itself, its pole and the state (red, green). Other relevant vision use cases are the detection of driving vehicles to know when it is safe to cross the road, lane detection to extract details about the road’s size and course, and curb detection to obtain information about its position and height.

Table 4.4 summarizes the presented evaluation of the *Crossing a Road* scenario.

Table 4.4: Crossing a Road

Quote	”Crossing a road with high traffic frequency without any safeguarding is always a very big danger for a blind person,” (EI3).
Initial Assumption	A visually impaired person needs to cross a road.
Normal	With the help of TGGS or acoustic signals, the person finds a crosswalk or traffic light and safely crosses the road.
What Can Go Wrong	There is no TGGS in front of the crosswalk or the traffic light does not offer acoustic signals. The person cannot find the TGGS even though it is there. The person does not find the traffic light pole in order to activate the acoustic signal. An intermediate platform may make them unsure on how to proceed. Not knowing the height of the curb can cause stumbling and falling. It can make them unsure, if they do not know the size of the road.
Vision Use Cases	Crosswalk Detection, Traffic Sign Detection (to detect the crosswalk sign), Traffic Light Detection, Traffic Light State Detection (red, green), Traffic Light Pole Detection, TGGS Detection, Lane Detection (to extract information about the road’s size and course), Driving Vehicle Detection (to know if the road can be crossed), Curb Information

Obstacle Avoidance

When moving in traffic situations, a visually impaired person has to check constantly for obstacles in order not to collide with someone or something and thereby they have to be careful not to lose direction and orientation: ”They [obstacles] impede the walking flow, they interrupt you, you lose direction,” (EI4).

Guide dog and white cane are the usual aids in this situation. Thereby, a distinction is made between grounded and elevated obstacles. Whereas guide dogs are usually trained to detect both types, it is not possible to detect elevated obstacles with the white cane:

> Concerning the obstacles, a trash can or garbage bin or whatever someone has put there (...), these are the so-called grounded obstacles. You’ll catch them with the cane in any case and also a guide dog would go around it. In obstacles, the so-called elevated obstacles are problematic. Drooping branches or cargo areas of lorries, perhaps very low hanging signs or anything that can hang into the pavement from above. A good guide dog is trained on them, but with the cane you will under-swing them. (EI1)

Other obstacles that have to be considered specifically are construction sites ("The biggest problems in traffic are actually construction sites and of course unknown situations. (...) Construction sites can be very confusing," (EI2).), stairs ("When steps are ahead, for example. (...) You have to be pretty careful. I can feel it with the cane, but (...) I have to slow down," (MTG3).), and bike-riders ("It's hard to recognize them [bike-riders]. Especially if they drive very undisciplined. (...) And this undisciplined behaviour of bike-riders is extremely fatal for the visually impaired. They drive everywhere, they drive on the pavement and everything," (MTG4).).

Therefore, besides a general obstacle detection, the following vision use cases are of importance: The detection of construction sites and the according traffic sign, TGGS detection in case one loses it while moving around obstacles, stairs, and bicycle detection.

Table 4.5 summarizes the presented evaluation of the *Obstacle Avoidance* scenario.

Table 4.5: Obstacle Avoidance

Quote	"They [obstacles] impede the walking flow, they interrupt you, you lose direction," (EI4).
Initial Assumption	A visually impaired person is on the move as a pedestrian in traffic situations and has to take care not to collide with obstacles.
Normal	With the help of the white cane or a guide dog, obstacles are detected and avoided.
What Can Go Wrong	Whereas guide dogs are usually trained to detect ground as well as elevated obstacles, it is not possible to detect elevated obstacles with the white cane. The detection and avoidance of a construction site can be difficult. While moving around an obstacle the person can lose orientation and/or drift away from the TGGS (see Table 4.2). Although stairs are detected by white cane and guide dog, the person might still be unsure. Bike-riders can be difficult to perceive because of their speed in combination with their often unpredictable behaviour.
Vision Use Cases	Obstacle Detection, Construction Site Detection, Traffic Sign Detection (to detect the construction site sign), TGGS Detection, Stairs Detection, Bicycle Detection

Boarding a Bus

Taking the bus is challenging for the visually impaired: "The bus is also the most difficult means of transport because it is so flexible," (EI3). After having found the bus stop, they usually wait on the entry field marked with TGGS to enter the bus that stops directly in front of the field. Problems occur at larger stops where several buses stop at once. In this case, it is hard to find the right bus and the door.

> You have a so called entry field at the beginning of the bus stop. You stand there and when the bus arrives, it stops exactly with the door at the entry field. You enter and that's it. But if you have two or three buses behind each other, you don't get it. (...) You have a lot of difficulties finding the entrance door. (...) You then have to feel along the vehicle to find the doors. (EI3)

A further limitation at bus stops is information that is presented in written form such as displays ("At bus stops, there are displays announcing the next bus or tram. I can't read that," (MTG4).), timetables ("The system could help me read the timetable at the bus stop," (MTG7).), or number and direction of the bus ("At bus stops where only one bus departs, it's no problem at all, but if there are several buses (...), I have to try to see the bus number so that I don't get on the wrong bus," (MTG4).).

The vision use cases that could support the visually impaired in this traffic scenario are traffic sign detection to find the bus stop, TGGS detection to find the entry field, display and text detection to find and read information, and door detection to find the entrance to the bus.

Table 4.6 the presented evaluation of the *Boarding a Bus* scenario.

Table 4.6: Boarding a Bus

Quote	"The bus is also the most difficult means of transport because it is so flexible." (EI3)
Initial Assumption	A visually impaired person wants to board a bus.
Normal	The person waits at the bus stop on the entry field marked with TGGS. When the bus arrives, they enter.
What Can Go Wrong	The person cannot find the bus stop. They cannot find the entry field even though it is there. There is no entry field and the person has to rely on hearing to find the door. They do not know, if the arriving bus has the right number and direction and they might not want to ask every time. At a larger stop where several buses stop at once, it is difficult to find the right bus.
Vision Use Cases	Traffic Sign Detection (to detect bus stop signs), TGGS Detection, Display Detection (to detect displays with important information), OCR (to read the text on the detected displays), Door Detection

At the Train Station

Although train stations are "easier to overlook and (...) at least at most train stations, there is some logic that you can understand" (EI1), there are some problems that can occur.

The bigger train stations in Germany offer some guidance for the visually impaired such as TGGS leading from the entrance hall to the information counter and platforms, oral announcements, and Braille indicators under the handrails. Additionally, the German railway company offers a mobility service meaning that employees guide visually impaired travellers at the train station. Problems occur for example at smaller stations with less infrastructure and when someone cannot find the supporting infrastructure or does not know that it exists.

The following quotes from interviewees underline this problematic:

> There is a guidance system with which I can reach the information [counter]. From there, I use the mobility service. It's often difficult when the platform where the train departs changes. The people from the [mobility] service know that faster. Until I figure it out, the train is already gone. (MTG2)

> At new or newly renovated stations, you usually have TGGS that lead from the hall to the platform. (...) It's more of a problem at smaller stations where you have little to no announcements. (EI2)

> If the station is unknown to you, you usually have to ask because not every handrail has Braille indicators or (...) it may exist, but you have to know that in the first place and you have to get there. (EI1)

Again, all kinds of signs such as departure boards ("(...) the departure of some trains can change. It would be quite good if the system could read that," (MTG3)), coach numbers ("There is a certain problem that the coach number on the ICE [Inter City Express] is hard to recognize," (MTG4).), seat numbers ("Even if I have reserved a seat: How do I find it," (MTG3)?), or platform sections ("Imagine the platform, 400 meters long, and you are looking for section C where your coach with the number 12 stops," (EI1).) are hard to impossible to receive for the visually impaired.

At train stations, the following vision use cases could therefore support visually impaired travellers: TGGS detection to find the guidance system leading to points of interest, traffic sign detection to detect different signs such as platform or section indicators, display and text detection to find and read important information, and door detection to find the door of the right coach.

Table 4.7 summarizes the presented evaluation of the *At the Train Station* scenario.

4.2.3 Related ADAS Work for Relevant Vision Use Cases

In the further course of the thesis, I focus on the derived use cases, especially the ones that are also of relevance in the field of ADAS. Therefore, Figure 4.2 gives an overview of the extracted use cases sorted by the hierarchy categories presented in Figure 4.1, and Figure 4.3 lists the overlapping use cases from both fields, ADAS and ASVI. In the following, I summarize ADAS solutions for the overlapping use cases.

In their chapter about related work, Yuan et al. [77] give an overview of existing methods for traffic sign detection and recognition. Traffic sign recognition usually consists

Table 4.7: At the Train Station

Quote	”It [the train station] is easier to overlook and (...) at least at most train stations, there is some logic that you can understand,” (EI1).
Initial Assumption	A visually impaired person wants to travel by train.
Normal	A TGGS leads the person to the platforms. They find the right platform with the help of Braille indicators under the handrails or they know the design of the station. Additionally, they can use the mobility service of the German railway company.
What Can Go Wrong	There is no TGGS leading to the platforms or the person cannot find the TGGS even though it is there. There are no Braille indicators or they do not find the handrails. The mobility service is not available. Small train stations do not always offer announcements. They do not find the right track section which matches their seat reservation.
Vision Use Cases	TGGS Detection, Traffic Sign Detection (to detect platform numbers and platform section signs), Display Detection (to detect displays with important information), OCR (to read the text on the detected displays), Door Detection

of two blocks: detection and classification. In addition, tracking is used to increase the recognition rate. For detection, there are on the one hand methods based on colour and shape (e. g. [78]) and on the other hand ML approaches, e. g. based on SVM [79]. Tracking detected signs with Kalman filters is applied with several intentions. For example, tracking can be used to include detection results from different frames [80] or to eliminate the results which cannot be found in successive frames [81].

As traffic sign classification is a typical object classification problem, the according algorithms are applied. Following feature extraction, SVM [79], NN [82], and other techniques are used. As traffic signs who set out lane arrangements are often neglected by traffic sign recognition systems, [83] specifically addresses this problem.

A vehicle detection for urban areas under consideration of different angles in which the vehicle is recorded is presented in [84]. The paper also contains an analysis of the state of the art for vision-based vehicle detection sorted by certain characteristics such as on-road environment (highway, urban, night time), vehicle views (front, rear, side), or if it is covers partial occlusions and part-based detection. Rubio et al. [85] for example present a vehicle detection for night time highways under consideration of front and rear views and without addressing partial occlusions or part based detection.

Vehicle detection can be seen as part of obstacle detection, as many obstacle detection developments in ADAS focus on specific objects. Besides vehicles, pedestrians

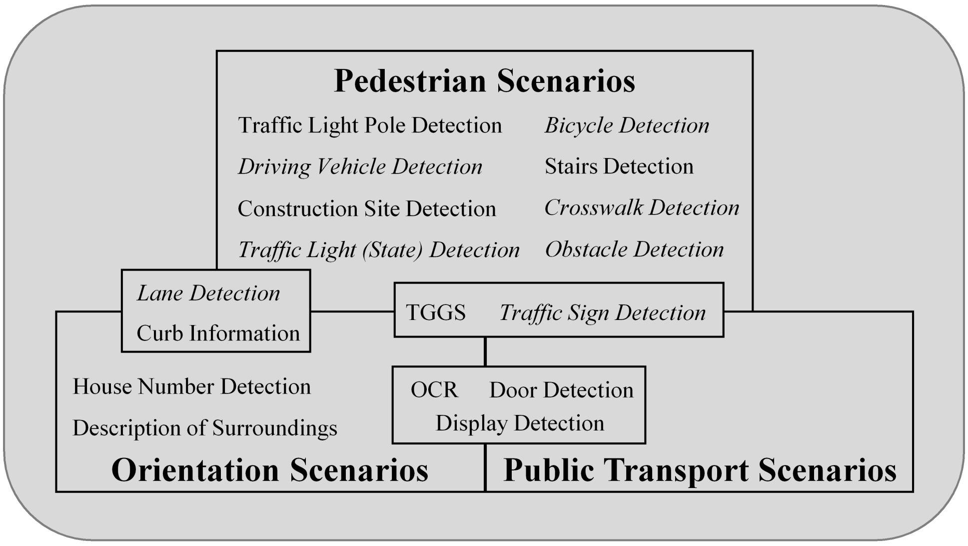


Figure 4.2: Traffic Scenarios and their Overlapping Use Cases. Italic Use Cases are of Relevance in ADAS.

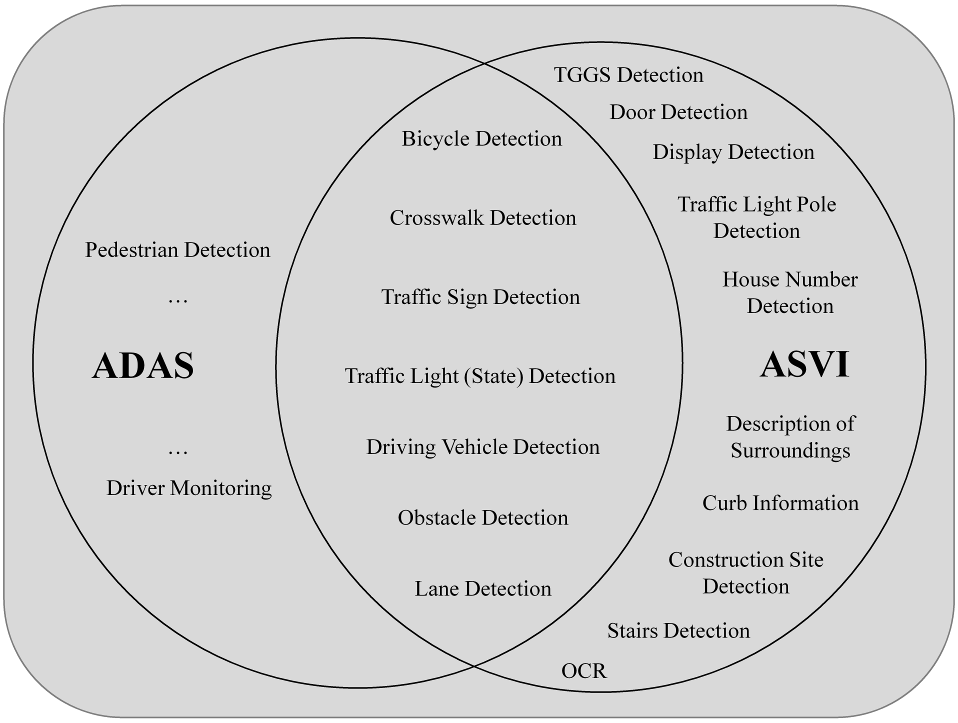


Figure 4.3: Overlapping Vision Use Cases from ADAS and ASVI

and bicycles are detected (e. g. [86]). In these cases, a priori knowledge about the obstacles' texture, colour, and shape is used to train models. On the contrary, Yang and Zheng [87] present a system that responds to every approaching object by exploiting motion patterns of a single dashboard camera. Other work that responds to all kinds of approaching objects can for example be found in [88, 89]. These general methods are the ones that have to be considered for visually impaired pedestrians because the kinds of obstacles they face are extremely distinctive.

Jung et al. [90] present a lane detection based on spatio-temporal images and summarize the general procedure of existing lane detection algorithms. They consist of two phases: detecting the lane and fitting it to a parametric curve. In addition to edge detection followed by Hough transform, the main parts of the detection phase, some preprocessing steps, such as gray-level conversion, contrast or gradient enhancement, and noise reduction are executed. For the fitting phase, various interpolation methods such as B-spline, cubic, or other polynomial curves are applied. For example, [91] introduces a lane detection method that first enhances the gradients and thus the lane marks. The lane is then detected by Canny edge detection and Hough transform, before a fitting based on a quadratic curve model is performed. To reduce occurring problems caused by changing illumination or missing parts of lane markings, motion vectors or optical flow are used (e. g. [92]).

In [93], a traffic light (state) detection based on HSV colour space, Maximally Stable Extremal Region (MSER), Histogram of Oriented Gradients (HOG) features, SVM, and CNN is presented. Additionally, [93] describes the general structure of algorithms for traffic light (state) detection. First, candidates for traffic lights are identified by their colour for which different colour spaces can be used. The number of false candidates is then reduced by shape features, e. g. not round candidates are eliminated with the Hough transform [94]. Furthermore, structure information of the traffic light is extracted using global and local image feature descriptors like HOG [95] or Haar-like features [96]. The performance and robustness of traffic light detections is improved by using a combination of colour, shape, and structure features. For the recognition of the traffic light's state different classifiers such as SVM [95] or CNN [97] can be used.

For crosswalk detection, [98] proposes to extract crosswalk regions under different illuminations by means of MSER and to eliminate false candidates using Extended Random Sample Consensus (ERANSEC). The fact that crosswalks have a horizontal structure from driver perspective is used by [61] and [99]. Choi et al. [61] use a 1-D mean filter and examine the difference image between original and filtered image, whereas Kummert and Hasselhoff [99] take advantage of the bipolarity and the straight lines of crosswalks by applying Fourier and Hough transform.

4.2.4 Non-Traffic Scenarios

When asked where a camera-based assistive system could offer support for the handling of the problems described in the interview guideline, the MTG often mentioned not only traffic situations but also indoor scenes. Because of the problems they were confronted with, they named navigational help inside of buildings in general and different use cases inside the doctor's office as well as the concert hall where camera-based detection could support them. Although, no scenario containing a supermarket was presented, two interviewees pointed out that navigational support in supermarkets, e. g. finding a certain product category, would improve their daily life. I will not pursue these scenarios any further in the course of the thesis because the focus is on traffic situations, but it is important to point out that detecting or recognizing any kinds of signs or displays and reading the text, if there is any, are very important use cases in almost every situation of daily life, indoor and outdoor.

4.2.5 Use Case Importance

Table 4.8 gives an overview on the importance of the discussed use cases sorted by the MTG that were interviewed. As the interview's focus was to gather as many use cases as possible, not every use case was discussed with every interviewee. This is why in some cases, it cannot be known from the answers of the interviewees if a certain use case is needed or not. Furthermore, the following effect has to be considered: MTG1 does not use public transport on her own and therefore states that she does not need the according use cases, but it was not discussed in the course of the interview if she would consider travelling by public transport on her own, if she had the support of an assistive system.

The use case overview in Table 4.8 confirms what I stated in the section about social insights: visually impaired road users are a very distinctive group with different needs. While for the blind interviewees (MTG2, MTG3, and MTG9) almost all discussed use cases are marked as needed, the interviewees with residual vision give very differing answers. This has in general two reasons: First, it depends on the concrete impairment and second, the person's personality and autonomy in road traffic plays an important role on how much support is required by an assistive system. For the conceptual design of such a system, it is therefore important to take possibilities of personalization into account.

Considering the lines about traffic sign detection, OCR, display detection, and house number detection, it can be seen that the ability of perceiving different kinds of displays and signs, indoor as well as outdoor, is essential for many visually impaired persons. Additionally, obstacle detection in general and bicycle detection in particular as well as traffic light (state) detection are often needed use cases.

4.3 Conclusion

In this study, the problems visually impaired pedestrians face in traffic situations are collected by conducting four expert interviews and ten interviews with MTG using Witzel's problem-centered method [64] in both cases. From the data a set of six different scenarios are extracted (*General Orientation*, *Navigating to an Address*, *Crossing a Road*, *Obstacle Avoidance*, *Boarding a Bus*, *At the Train Station*) clustered into the three categories *Orientation*, *Pedestrian*, and *Public Transport Scenarios*. For each of the six scenarios, a descriptive table is created summarizing the usual procedure of the respective scenario, the problems that can occur and which vision-based use cases could help to overcome the problems. As the objective of this thesis is to work towards the formulation of a transfer concept for vision-based use cases and according algorithms from driver assistance to the assistance of visually impaired people, the following research is based on the overlapping use cases pointed out in Figure 4.3. With the data from the interviews, some further findings that go beyond the collection of traffic scenarios and are of importance for researchers who want to build a camera-based assistive system can be formulated.

Table 4.8: Traffic Use Cases with Importance to MTG

Use Case/ID	MTG1	MTG2	MTG3	MTG4	MTG5	MTG6	MTG7	MTG8	MTG9	MTG10
Traffic Light Pole Detection	-	✓✓	✓✓	✓✓	-	✓✓	×	-	-	-
Bicycle Detection	✓	✓✓	✓✓	✓✓	×	✓✓	-	×	✓✓	✓✓
Stairs Detection	✓✓	-	✓✓	-	-	✓	×	-	-	-
Construction Site Detection	-	✓✓	-	×	-	✓	×	-	-	✓✓
Driving Vehicle Detection	-	×	✓✓	✓✓	×	✓✓	×	-	-	✓✓
Crosswalk Detection	✓	✓✓	-	×	-	✓✓	×	-	-	✓✓
Traffic Light (State) Detection	✓	✓✓	-	✓✓	-	✓✓	×	✓✓	-	✓✓
Obstacle Detection	✓✓	✓✓	✓✓	✓	×	✓✓	×	✓✓	✓✓	✓✓
Lane Detection	✓✓	-	✓	-	-	✓✓	-	-	✓✓	-
Curb Information	✓✓	✓✓	✓✓	-	×	✓	✓	✓✓	✓✓	-
TGGS Detection	-	-	-	-	-	-	-	-	-	-
Traffic Sign Detection	-	✓✓	✓✓	-	✓✓	✓✓	-	×	✓✓	✓✓
House Number Detection	✓✓	✓✓	✓✓	✓✓	✓✓	✓✓	✓✓	✓✓	✓✓	✓✓
Description Of Surroundings	-	-	✓✓	-	-	-	-	-	-	-
OCR	✓✓	✓✓	✓✓	✓✓	✓✓	✓✓	✓✓	✓✓	✓✓	✓✓
Door Detection	×	✓✓	✓✓	×	-	-	×	-	-	✓✓
Display Detection	✓✓	✓✓	✓✓	✓✓	✓✓	✓✓	✓✓	✓✓	✓✓	✓✓

✓✓: Needed, ✓: Useful Addition, -: Not Discussed, ×: Not Needed

Although there are some use cases that seem to be of general importance (see Table 4.8), the diversity of visually impaired people concerning kind and degree of impairment(s) as well as personality leads to very differing needs. Therefore, it is necessary to consider possibilities of profiling and personalization when building an assistive system, e. g. by using an AAL platform as presented in [4] and as suggested in the sketch for an assistive system in section 1.2. Among the generally important use cases, the perception of any kinds of signs and displays has to be emphasized because of its importance in numerous situations in daily life. Shen and Coughlan [100] for example address this problem.

With the help of the interview data, I can answer the questions formulated in section 3.1:

(Q.a) Are there differences in gender and/or age of visually impaired people when it comes to dealing with traffic scenarios?
Concerning gender, one EI stated that girls and young women are more likely to attend additional voluntary mobility workshops than boys and young men. According to the EI, life experience with visual impairment is more important than age.

(Q.b) Is the use of technology common among visually impaired people?
One EI states: "When you have a limitation, you depend on technology and of course you use it,". This is underlined by the fact that all interviewed MTG use at least PC or smartphone.

(Q.c) How do visually impaired people prepare for a trip to an unknown address?
Visually impaired people essentially cover the same topics as the sighted, but in general they need more information and it takes more time to gather the information. It also depends on a person's personality, no matter if sighted or visually impaired, how much information is gathered before a trip.

(Q.d) Are visually impaired people comfortable with having to ask for support or directions?
There are opposing reports from the interviewees. It depends on a person's personality how comfortable they are with asking people for support or directions.

(Q.e) Which identified vision use cases are the most important?
From Table 4.8, it can be seen that almost all discussed use cases are of importance for the blind interviewees. Contrarily, interviewees with residual vision give very differing answers. It depends on a person's concrete impairment but again also on their personality, which and how much support is required. Nevertheless, it can be stated that the perception of all kinds of signs, indoor and outdoor, as well as obstacle detection are very important use cases for the visually impaired.

When discussing differences between the sighted and visually impaired and talking about needs of visually impaired people, it is important to keep in mind that "the blind and visually impaired are as different individuals as you and your colleagues," (EI4).

That the idea of a camera-based assistive system in traffic situations is met with approval is underlined by the following two interviewee quotes:

> One would be much more independent. It could help in all areas of life. (MTG1)
>
> We currently have the rapid development of smartphones, and with that we are also experiencing more and more comfort. And in this context, such a development and research as yours is of utmost importance, so that one can achieve more safety in road traffic. (EI3)

4.4 Thesis 1

In the following, the four parts of the objective formulated in 3.1 are discussed. Objectives (O1.1), (O1.2), and (O1.3) are achieved because of the data saturation of the qualitative interview study meaning that the enumerations of traffic scenarios and use cases are exhaustive.

(O1.1): All traffic scenarios that are of interest for visually impaired pedestrians have to be defined.
Inductively coding the interview data leads to six traffic scenarios that are of interest for the visually impaired and that can be clustered into three categories (see also Figure 4.1): *Orientation Scenarios* (*General Orientation*, *Navigating to an Address*), *Pedestrian Scenarios* (*Crossing a Road*, *Obstacle avoidance*), and *Public Transport Scenarios* (*Boarding a Bus*, *At the Train Station*).

(O1.2): All vision use cases that can support the visually impaired in traffic situations have to determined.
From the tables created for each traffic scenario (Table 4.2 to Table 4.7), I extract and gather all 17 relevant vision use cases (see also Figure 4.2): (1) Traffic light pole detection, (2) traffic light (state) detection, (3) bicycle detection, (4) (driving) vehicle detection, (5) stairs detection, (6) construction site detection, (7) crosswalk detection, (8) obstacle detection, (9) lane detection, (10) curb information, (11) TGGS detection, (12) traffic sign detection, (13) house number detection, (14) description of surroundings, (15) OCR, (16) door detection, and (17) display detection.

(O1.3): The overlap of vision use cases addressed in ADAS and needed in ASVI has to be determined.
I determine the overlap in use cases of relevance for the visually impaired and the ones addressed in ADAS by comparing the above created ASVI list of 17 use cases with ADAS literature. This results in seven overlapping use cases (see also Figure 4.3): (1) Lane detection, (2) crosswalk detection, (3) traffic sign detection, (4) traffic light (state) detection, (5) (driving) vehicle detection, (6) obstacle detection, and (7) bicycle detection.

(O1.4): The idea of using (adapted) software engineering methods to cluster and present qualitative data has to be introduced.
In my publication [74], I discuss objective (O1.4). The literature provides several procedures for the analysis of data acquired in qualitative research, including codes which I used for the evaluation of the presented study. But as suitable structuring and representation of the gained knowledge are highly depending on the concrete problem, there is no "simple step that can be carried out by following a detailed, "mechanical," approach. Instead, it requires the ability to generalize, and think innovatively, and

so on from the researcher," [101, p. 63]. I master this challenge by adapting methods from software engineering. Using qualitative methods to improve the software development process has been applied in the past (see e. g. [102, 103]), but the reverse – using software engineering methods to improve the representation of qualitative data – is a new approach. The many structuring elements, such as different tables and diagrams, found in software engineering are powerful tools that can help to generalize qualitative data and other procedures beyond software.

Hence, all four objectives are achieved and can be summarized in Thesis 1.

Thesis 1: Traffic Scenarios and Use Cases

I defined the significant traffic scenarios for visually impaired pedestrians and determined all vision use cases of relevance in these scenarios. From that, I determined the overlap of vision use cases between ADAS and ASVI. Besides, I introduced the idea of using software engineering methods for the presentation of qualitative data.

(T1.1) I showed that the traffic scenarios of interest for visually impaired pedestrians are: *Orientation Scenarios* (*General Orientation*, *Navigating to an Address*), *Pedestrian Scenarios* (*Crossing a Road*, *Obstacle avoidance*), and *Public Transport Scenarios* (*Boarding a Bus*, *At the Train Station*).

(T1.2) I determined all vision use cases that can support the visually impaired in traffic situations: (1) Traffic light pole detection, (2) traffic light (state) detection, (3) bicycle detection, (4) (driving) vehicle detection, (5) stairs detection, (6) construction site detection, (7) crosswalk detection, (8) obstacle detection, (9) lane detection, (10) curb information, (11) TGGS detection, (12) traffic sign detection, (13) house number detection, (14) description of surroundings, (15) OCR, (16) door detection, and (17) display detection.

(T1.3) I determined the overlap of vision use cases addressed in ADAS and needed in ASVI: (1) Lane detection, (2) crosswalk detection, (3) traffic sign detection, (4) traffic light (state) detection, (5) (driving) vehicle detection, (6) obstacle detection, and (7) bicycle detection.

(T1.4) I introduced the idea of using (adapted) software engineering methods to cluster and present qualitative data.

Own publications supporting Thesis 1 are: [70, 73, 74, 104].

Chapter 5

The *CoPeD* Data Set for Traffic Scenarios

For the overlapping use cases from ADAS and ASVI, comparable video data from driver and pedestrian perspective are needed for the evaluation of ADAS algorithms and their ASVI adaptations. I first cite related work in section 5.1 and thereby show that existing data sets only cover driver but not pedestrian perspective. Consequently, I develop the *CoPeD* data set for traffic scenarios. Conditions and content of *CoPeD* are described in section 5.2. After a conclusion of this chapter, objective (O2) is discussed. The presented work is based on my publication [105].

5.1 Related Work

There are several databases that are used for developing and testing algorithms from driver and pedestrian perspective. In the following, according data sets covering the seven considered use cases are outlined by example.

In [106] and [107], publicly available data sets for lane detection are described. Whereas the first data set consists of four clips with a total of 1224 frames, the *KITTI-ROAD* data set from [107] contains approximately 600 images with held-out annotations for evaluation via a website. Le et al. [108] developed an algorithm for pedestrian lane detection at traffic junctions. They therefore created a not publicly available data set consisting of about 1000 images.

Although there are several researchers who work on crosswalk detection from driver or pedestrian perspective (e. g. [98, 109]), no publicly available data set exists to my knowledge. Concerning traffic light (state) detection from driver perspective, the *LISA Traffic Light Dataset* ([110] and [111]) provides comprehensive material.

There are several driver data sets available for the detection of obstacles in general (e. g. [88] uses the already mentioned *KITTI-ROAD* data set) or specific obstacles such as vehicles ([112], Caltech[1]) or pedestrians ([113]). In addition to data acquired from inside the car, there are applications where data from surveillance cameras are used to match vehicles (e. g. [114]). To my knowledge, no publicly available database for bicycles in traffic situations exists.

[1] `http://www.vision.caltech.edu/archive.html`, accessed on June 6, 2020

Examples for traffic sign detection and recognition data sets are the *LISA Traffic Sign Dataset* [115] for the USA and the German Traffic Sign Detection Benchmark [116]. They both cover traffic signs that are of importance for drivers but in general not for pedestrians.

In my experience, there are many researchers who test their algorithms and systems on own data sets which they do not make publicly available. Whereas some data sets for driver use cases exist, there are very few data sets from pedestrian perspective. As comparable sequences from both perspectives and for all use cases that have to be considered are needed, it was inevitable that I create my own database.

5.2 Conditions and Content

All video sequences are recorded under good weather conditions meaning that there is no precipitation. In the car, the camera is mounted on the rear view mirror, whereas for the pedestrian scenes, the camera is carried in the hand. Wherever possible, the driver and pedestrians scenes are shot in the same place right after another to ensure similar conditions. The HD video sequences are recorded with the camera model *Kodak PIXPRO SP360 4K*, the traffic sign images are taken with different smartphone cameras.

All people that can be seen in the sequences signed a consent to the publication of the material. The *CoPeD* data set is licensed under the Creative Commons Attribution 4.0 International License[2] and hosted publicly[3]. The license allows everyone, even in commercial contexts, to use, change, and redistribute the material. In return, one must give appropriate credit by citing my publication [105] and indicate if any changes to the material were made. Additionally, it is not allowed to apply legal terms or technological measures that legally restrict others from doing anything the license permits.

The folder available for download is structured as follows: Besides the read-me and license files, the main folder contains four sub folders, one for each category. The folders *Lanes*, *Crossings*, and *Obstacles* each contain the sub folders *Driver* and *Pedestrian* which then contain the comparable video sequences. Some of these folders contain another sub folder called *Others* consisting of further video sequences that belong to the according topic but do not have a comparable scene from driver respectively pedestrian perspective. The folder *Traffic Sign* encloses four sub folders that contain several images with the according traffic sign. The sub folders are: *Bus Stop*, *Crosswalk*, *Pedestrian and Bicycle Path*, and *Others*.

The folder *Lanes* covers the use case lane detection, the folder *Crossings* the use cases crosswalk, traffic light (state), and (driving) vehicle detection. The folder *Obstacles* provides material for bicycle detection as well as obstacle detection in general. Figure 5.1 illustrates in which *CoPeD* sub folder the data covering the respective use case can be found.

[2] `https://creativecommons.org/licenses/by/4.0/`, accessed on June 6, 2020
[3] `http://dataset.informatik.hs-furtwangen.de/`, accessed on June 6, 2020

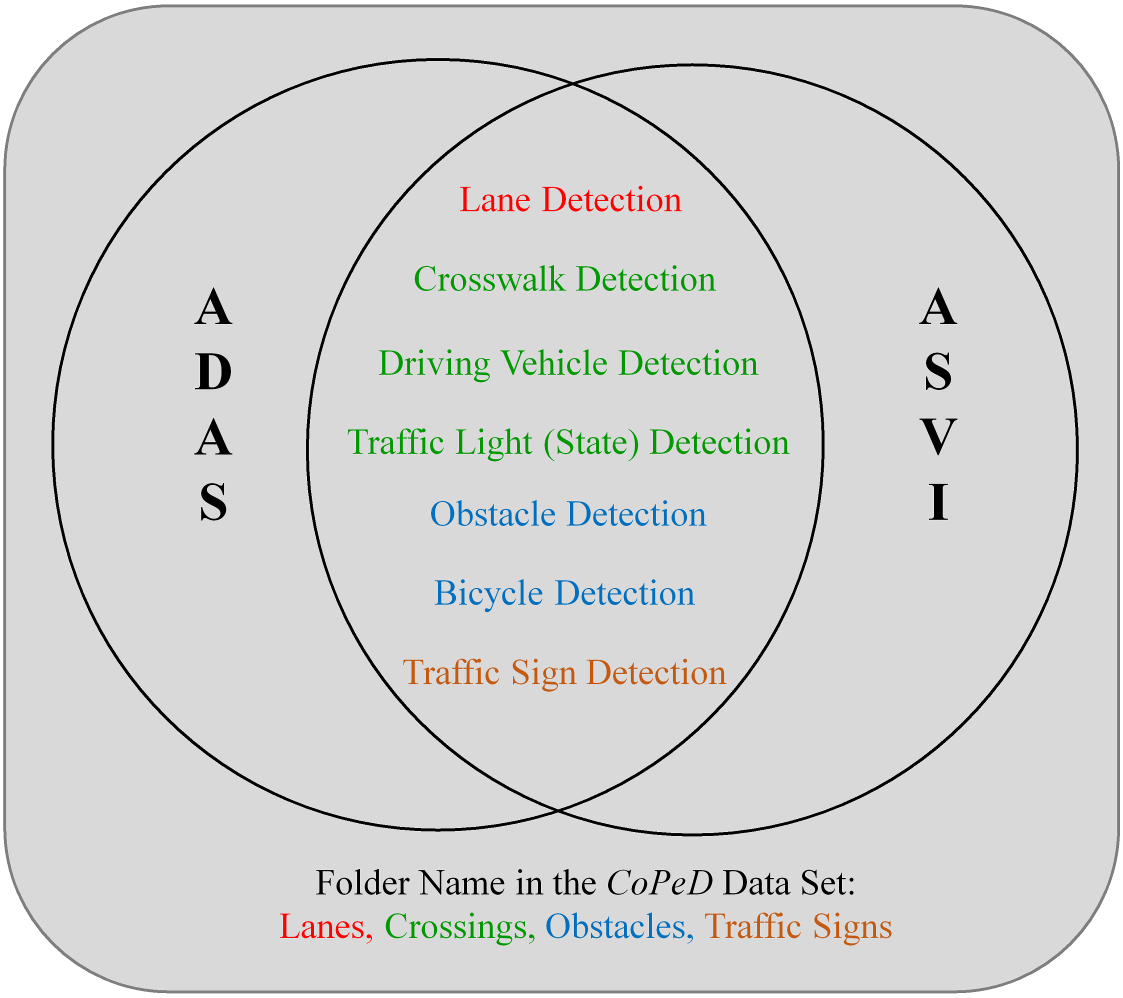


Figure 5.1: Coverage of Use Cases of the *CoPeD* Data Set

Figure 5.2 shows the combined activity diagram for lane detection and crossings from driver and pedestrian perspective. As there are three paths in the lane detection diagram, three sequences from each perspective, driver and pedestrian, following a straight road as well as a right and a left turn were filmed.

Crossings are divided depending on if there is a crosswalk, a traffic light, or no according support. In each path, there are two possibilities: one can cross the road respectively drive on or one has to wait. Therefore, there are in total six sequences for pedestrians and drivers each. Sequences where a pedestrian approaches a crossing with crosswalk, traffic light, or without support were filmed. In the first case, when there is no traffic, the traffic stopped, or the pedestrian light shows green, the pedestrian immediately crosses the road. Otherwise, they wait until it is safe to cross the road. The filming from driver perspective took place at the same locations. The driver either has to wait or can pass the crossing immediately.

The obstacles drivers and pedestrians face in traffic situations are very different. Hence, *CoPeD* contains sequences with varying obstacles. From both perspectives, bicycles and construction sites are addressed.

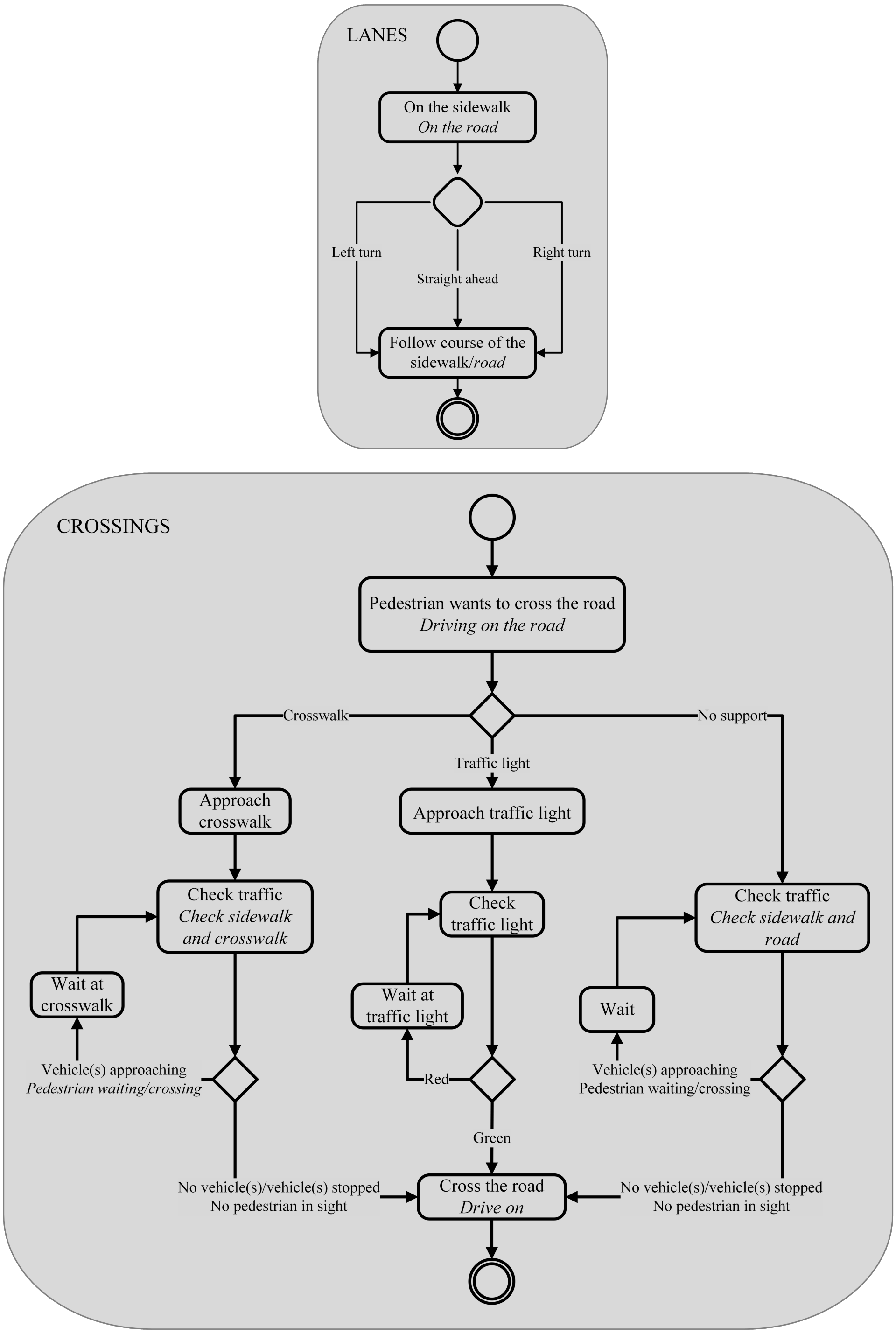


Figure 5.2: Activity Diagrams for Lanes (Above) and Crossings (Below). Italic Statements refer to Drivers.

From driver perspective, there are additional sequences with vehicles and waste on the roadway; from pedestrian perspective sequences with other pedestrians and trash cans on the pavement were added.

The usual data sets for traffic sign recognition refer to signs that are of importance for drivers but not for pedestrians. Therefore, the *CoPeD* data set contains a collection of images for traffic signs that are of importance for pedestrians. Figure 5.3 presents some partial images of the folder containing German bus stop signs.

Figure 5.3: Collection of Partial Images Containing the Traffic Sign *Bus Stop*

Figure 5.4 shows example images from driver and pedestrian perspective for lanes, crossings, and obstacles. A straight road is shown in the case of lanes. For crossings, three comparable frames from driver and pedestrian perspective are displayed. The driver approaches a crosswalk where a pedestrian is about to cross the road, a red traffic light, and wants to turn into a road where a pedestrian is crossing the road. The pedestrian waits at the crosswalk for the car to stop, at the traffic light for it to turn green, and crosses a road into which a car is about to turn. Concerning obstacles, the figure shows comparable frames for bikes and construction sites: A driver overtaking a bike-rider, a bike-rider overtaking a pedestrian, and a car as well as a pedestrian passing a construction site. In all example frames, the sky is cropped for better visibility of important content.

5.3 Conclusion

The literature review in section 5.1 shows that existing data sets for traffic scenarios do not cover all vision use cases needed in my research. Besides, the existing data are mostly from driver perspective and no comparable data from driver and pedestrian perspective exist. Thus, I presented the publicly available *CoPeD* data set for traffic scenarios. Divided into the four categories lanes, crossings, obstacles, and traffic signs, I created comparable video sequences from driver and pedestrian perspective.

In the case of traffic signs, single images containing traffic signs with importance to pedestrians are collected.

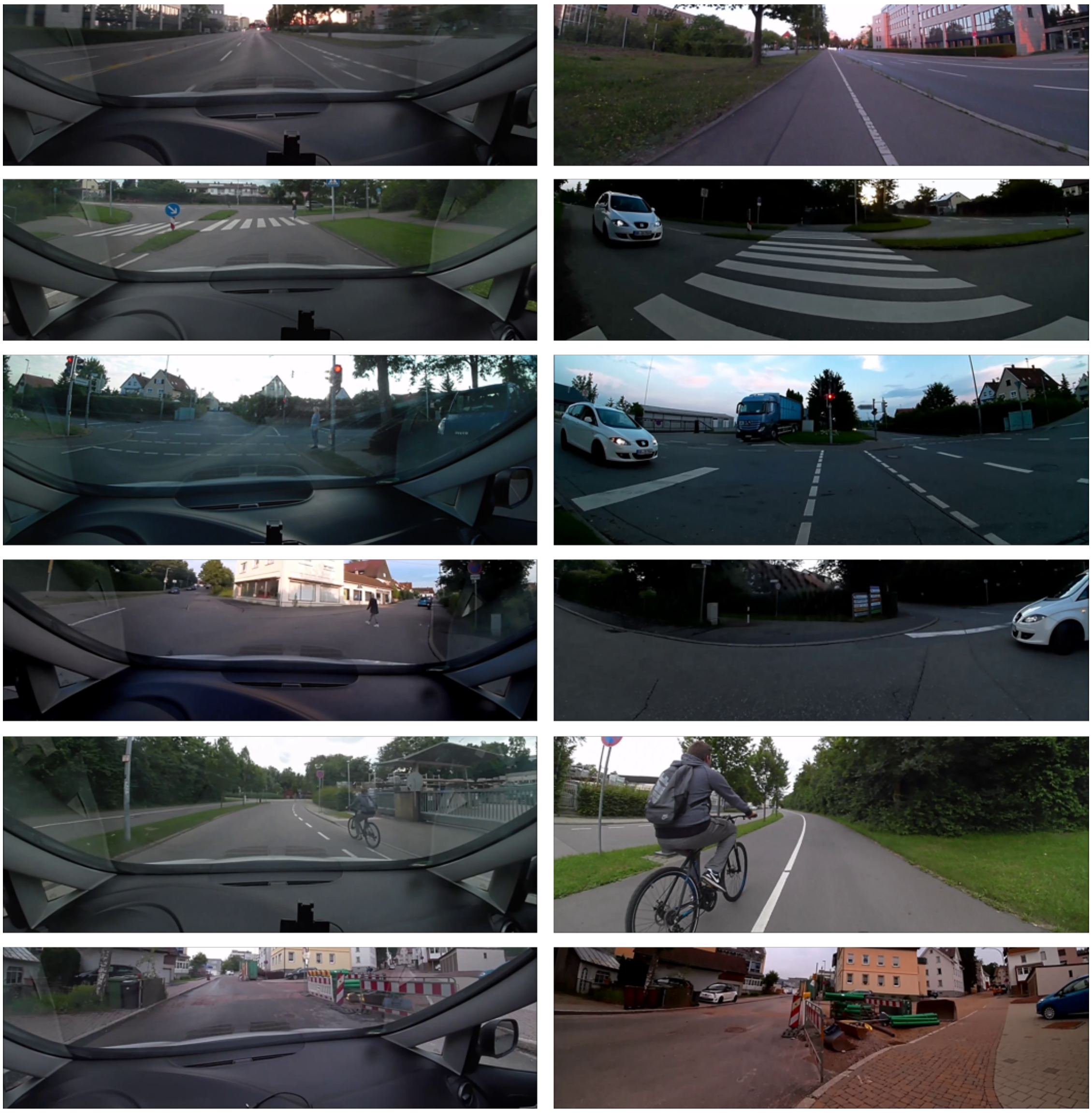


Figure 5.4: Example Images (Cropped) from Driver and Pedestrian Perspective for Lanes (Straight), Crossings (Crosswalk, Traffic Light, No Support), and Obstacles (Bicycle, Construction Site).

The *CoPeD* data set makes it possible to compare the performances of ADAS algorithms and their adapted ASVI versions that will be developed in the following chapter. It is licensed under the Creative Commons Attribution 4.0 International License and hosted publicly which makes it available for all researchers.

5.4 Thesis 2

In the following, the objective formulated in 3.2 is discussed.

(O2): The data set *CoPeD* containing comparable video data from driver and pedestrian perspective and covering the overlapping use cases from ADAS and ASVI has to be created.
Existing traffic data sets are usually recorded from driver and not from pedestrian perspective. The newly created *CoPeD* data set on the other hand contains traffic scenarios filmed from both perspectives. Recording the according scenarios in the same place and directly one after the other results in comparable sequences that are suitable to compare the evaluation of algorithms from driver and pedestrian perspective. The fact that all overlapping use cases are covered is shown in Figure 5.1.

Hence, the objective is achieved and can be summarized in Thesis 2.

Thesis 2: Video Data Acquisition

Thesis (T2): I created the data set *CoPeD* containing comparable video data from driver and pedestrian perspective and covering the overlapping use cases from ADAS and ASVI.

Own publications supporting Thesis 2 are: [104, 105].

Chapter 6

Use Case Examination

The seven use cases identified before and covered in the *CoPeD* data set have to be examined concerning possibilities of adaptation from ADAS to ASVI. I evaluate every adapted use case on *CoPeD* frames and compare the performances with the ones of the underlying ADAS algorithms.

In addition to these use cases, I introduce RBS as first step of the computer vision pipeline in section 6.1. The three following section 6.2, 6.3, and 6.4 are each dedicated to one of the so far addressed use cases, namely RBS, crosswalk detection, and lane detection. In each according section, I first cite related work and explain which ADAS procedures were chosen for adaptation to ASVI. Afterwards, each section contains a detailed description of proposed ASVI solutions based on chosen ADAS algorithms. I present a common evaluation of all three adapted use cases in section 6.5. After this chapter's conclusion in section 6.6, objective (O3) is discussed in section 6.7. Aside from the last section, this chapter is based on my publications [117] and [118].

6.1 Use Case Partition based on RBS

In addition to the seven overlapping use cases, I introduce RBS as first step of the computer vision pipeline. In ADAS, position and angle of the cameras are usually known and this knowledge can be exploited for object detection. Choi et al. [61] for example developed their crosswalk detection for a tilted camera that shows only the road but no background. The lane detection in [119] contains a vanishing point computation based on a camera aligned straight ahead in front of the car. However, in ASVI no such assumptions can be made. Therefore, RBS is considered as an additional use case. RBS allows some detection algorithms to be executed on a subset of the original image (see Figure 6.1). Traffic sign and traffic light detection can be executed on the background image, whereas crosswalk and lane detection are carried out on the road part of the image. Obstacles, including vehicles and bikes, are detected on the complete image. In the course of my research, I examined the use cases of the middle path of Figure 6.1: RBS as well as the "on-road" use cases crosswalks and lanes.

Table 6.1 shows the input and output variables for each procedure. It can be seen that RBS is the foundation for the other two algorithms because the output from RBS is used as input for crosswalk and lane detection. In the complete chapter, (x, y) refers to a pixel position in a given image.

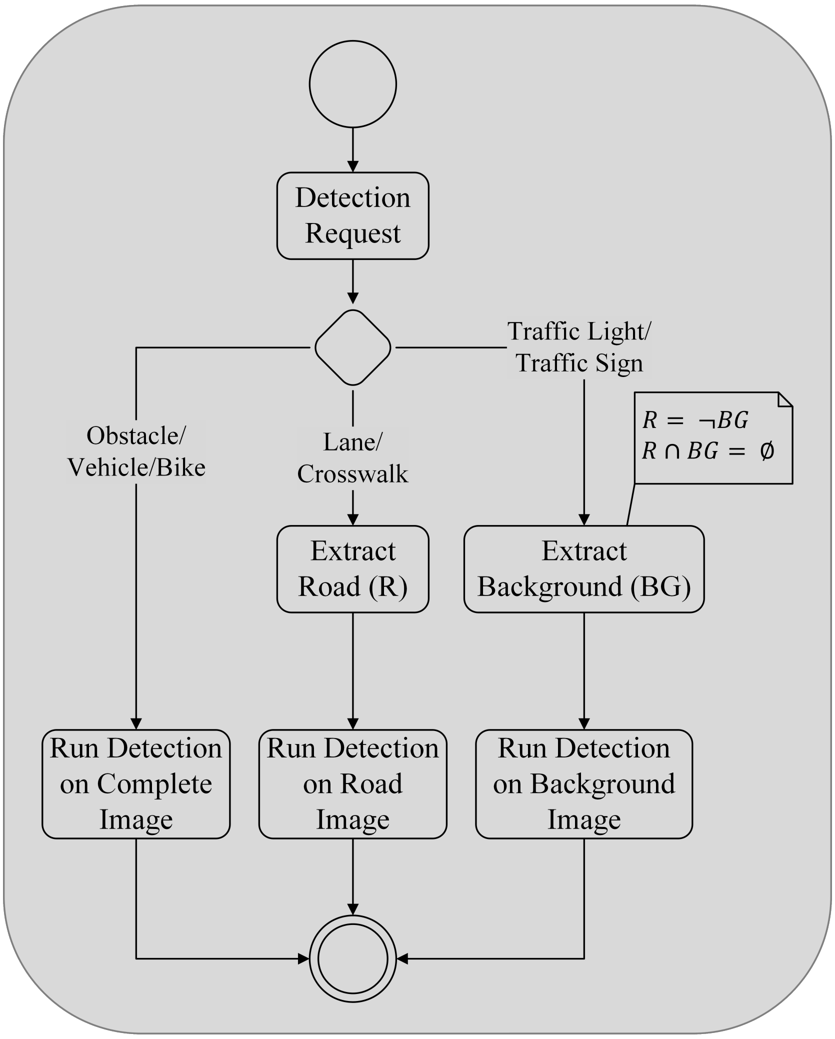


Figure 6.1: Partition of Detection with RBS

6.2 Road Background Segmentation

6.2.1 Related Work

In the 1990s, road segmentation for driver assistance using morphological watersheds was introduced [120]. The algorithm uses markers situated in front of the car; hence, it cannot be used without adaptations on data from pedestrian perspective.

From about 2008 on, Alvarez et al. proposed road segmentation based on illumination-invariant models (see e. g. [54, 121, 122]). Their procedure poses two problems when used from pedestrian perspective: They use seeds in front of the car and a camera calibration is needed. The latter makes it difficult to use on any smartphone which is the platform an assistive system for the visually impaired is most likely to use.

Many researchers in the field of ADAS use different ML techniques for road segmentation. NN are for example used in [123], CNN in [124], SVM in [125], and deep learning in [126]. In this context, Kavzoglu states that "[the] accuracy achieved by a supervised classification is largely dependent upon the training data provided by the analyst.

Table 6.1: Input and Output Variables for RBS, Crosswalk Detection, and Lane Detection

	RBS	**Crosswalk Detection**	**Lane Detection**
Input	I: RGB Input Image	I, R	I, R
Output	R: Binary image; road pixels are white. BG: Binary image; background pixels are white ($BG = \neg R$).	CW: Binary image; crosswalk pixels are white.	$Text$: Indicates if the road is straight or takes a right/left turn, provided any lanes are detected.

The use of representative training data sets is of significant importance for the performance of all classification methods," [127]. However, there are few video and image data sets consisting of traffic situations from pedestrian perspective and according publicly available annotated data sets which makes it difficult to test ML approaches in the context of ASVI. Foedisch and Takeuchi [123] overcome the problem of needing annotated data by presenting an adaptive road detection that uses NN. They exploit the known areas in the video data from driver perspective as can be seen in Figure 6.2. The green rectangles usually belong to the road and the blue ones to the background composed of sky and adjacent areas.

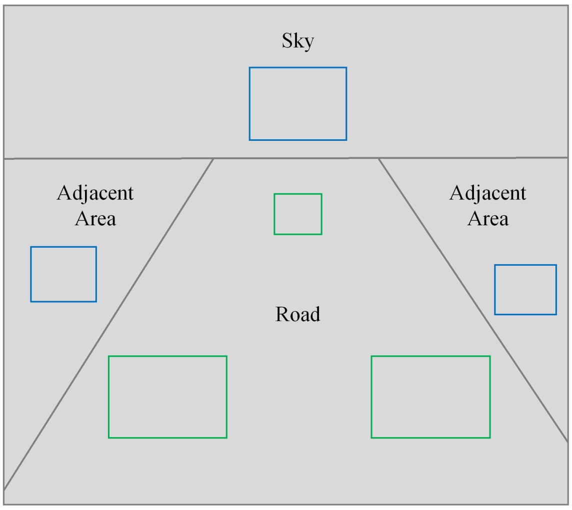


Figure 6.2: Areas in Road Images from Driver Perspective (adapted from [123]).

In the following sections, 6.2.2 and 6.2.3, two adaptations for RBS from pedestrian perspective are presented. The first adaptation is based on Beucher et al.'s work using watersheds [120], the second is based on the work of Foedisch and Takeuchi [123] using ML.

6.2.2 Proposed ASVI Adaptation based on Watersheds

The presented algorithm consists of five steps: After preprocessing and watershed computation (1), properties of the according catchment basins are determined (2). With the help of this information, the mosaic image is built by assigning to each pixel the mean gray value of the catchment basin it belongs to (3). Afterwards, adjacent catchment basins are merged if their gradient is below a threshold (4). Finally, a decision is made which uniform region from the merged mosaic corresponds to the road part of the image and a morphological postprocessing of this region is carried out (5).

Figure 6.3 shows the procedure of the adapted algorithm and the differences to the underlying ADAS algorithm.

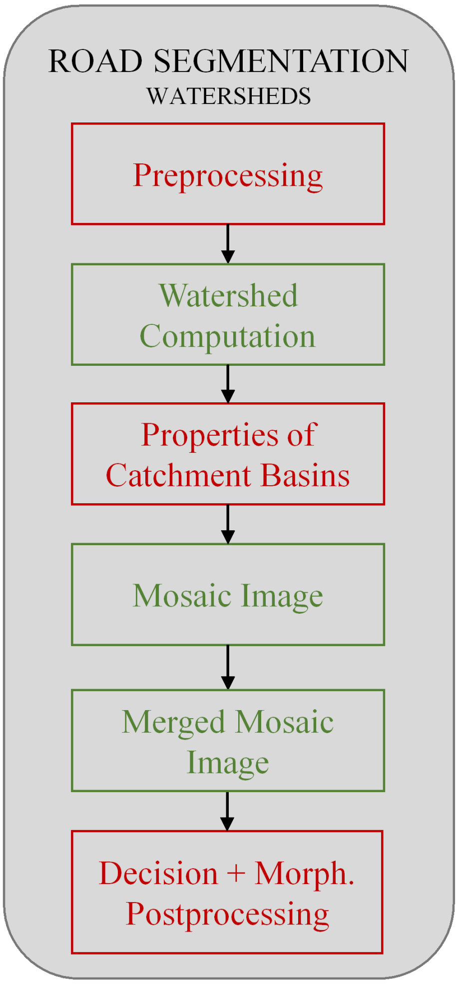


Figure 6.3: Procedure of Presented RBS based on Watersheds. Green boxes are taken from [120] with small changes. Red boxes required major changes or are newly introduced.

Contrarily to [120], I do not use a morphological gradient image as input of the watershed computation in order to reduce the number of catchment basins. To achieve this goal, some catchment basins are set to zero, depending on their mean value and mean saturation; thus, the boxes *Preprocessing* and *Properties of Catchment Basins* are marked red in Figure 6.3. Additionally, Beucher et al. use seeds in front of the car to expand the road from which is not possible in ASVI. To replace this procedure, step (5) *Decision + Morph. Postprocessing* is developed.

(1) Preprocessing and Watershed Computation
Watersheds are based on a topological interpretation of gray value images where the pixels, as spatial coordinates, are plotted against their intensity values. Local maxima are defined as watershed lines that separate the image into different regions. The inner parts of the regions are called catchment basins.

The following notations are considered throughout this section:

I RGB input image, smoothed with a Gaussian filter.

$gray$ Gray value version of I.

s s-channel (saturation) from I converted into HSV colour space.

L Watershed image of $gray$. Every pixel has an assigned number k according to the catchment basin it belongs to. L is set to zero for watershed lines.

N Number of catchment basins.

(2) Properties of Catchment Basins
For every catchment basin $k, 1 \leq k \leq N$, the mean intensity and mean saturation value are computed:

$$\overline{gray}(k) = \sum_{\{(x,y)|L(x,y)=k\}} \frac{gray(x,y)}{\#\{(x,y)|L(x,y)=k\}}$$
$$\overline{s}(k) = \sum_{\{(x,y)|L(x,y)=k\}} \frac{s(x,y)}{\#\{(x,y)|L(x,y)=k\}}$$

Thereby, $\#$ stands for the number of elements in a set.

(3) Mosaic Image
For $L(x,y) = k$, the mosaic image is then defined as

$$mosaic(x,y) = \begin{cases} \overline{gray}(k), & \text{if } k \neq 0 \wedge \overline{s}(k) < th_s \wedge th_1 < \overline{gray}(k) < th_2 \\ 0, & \text{else.} \end{cases}$$

Depending on the thresholds th_s, th_1, and th_2, the value of the mosaic image is either set to the mean gray value or to zero.

If $mosaic(x,y)$ is set to zero, $L(x,y)$ is simultaneously set to zero. This means that the value of each pixel is set to the mean gray value of the corresponding catchment basin, provided mean gray value and mean saturation of the basin are in ranges that make them possible parts of the road.

(4) Merged Mosaic Image
In this step, adjacent catchment basins are merged, if the gradient between them

falls below a threshold th_{gray}. Let k and l be the identification numbers of adjacent catchment basins. Their gradient is then defined as

$$\nabla(k,l) = |\overline{gray}(k) - \overline{gray}(l)|.$$

If it holds that $\nabla(k,l) < th_{gray}$, $\overline{gray}(k)$ and $\overline{gray}(l)$ are updated to their mean value

$$\overline{gray}(k) = \overline{gray}(l) = \frac{\overline{gray}(k) + \overline{gray}(l)}{2}$$

and the mosaic image is updated to the merged mosaic image

$$mosaic(x,y) = \begin{cases} \overline{gray}(k), & \text{if } (L(x,y) = k \vee L(x,y) = l) \\ & \wedge \nabla(k,l) < th_{gray} \\ mosaic(x,y), & \text{else.} \end{cases}$$

Furthermore, $L(x,y)$ is set to $L(x,y) = k$ for all $\{(x,y)|L(x,y) = l\}$. This process is repeated for all adjacent catchment basins.

(5) Decision and Morphological Postprocessing

Let max_i be the gray value that occurs most often in the merged mosaic image. The binary road image R is then defined as

$$R(x,y) = \begin{cases} 1, & \text{if } mosaic(x,y) = max_i \\ 0, & \text{else.} \end{cases}$$

A morphological postprocessing by applying filling, opening, largest component, and closing on R is performed.

The background is then determined as the complement of R

$$BG = \neg R.$$

For stability reasons, the union of the last n frames is used as a tracking step. Hence, the road of the n-th frame in a sequence is defined as

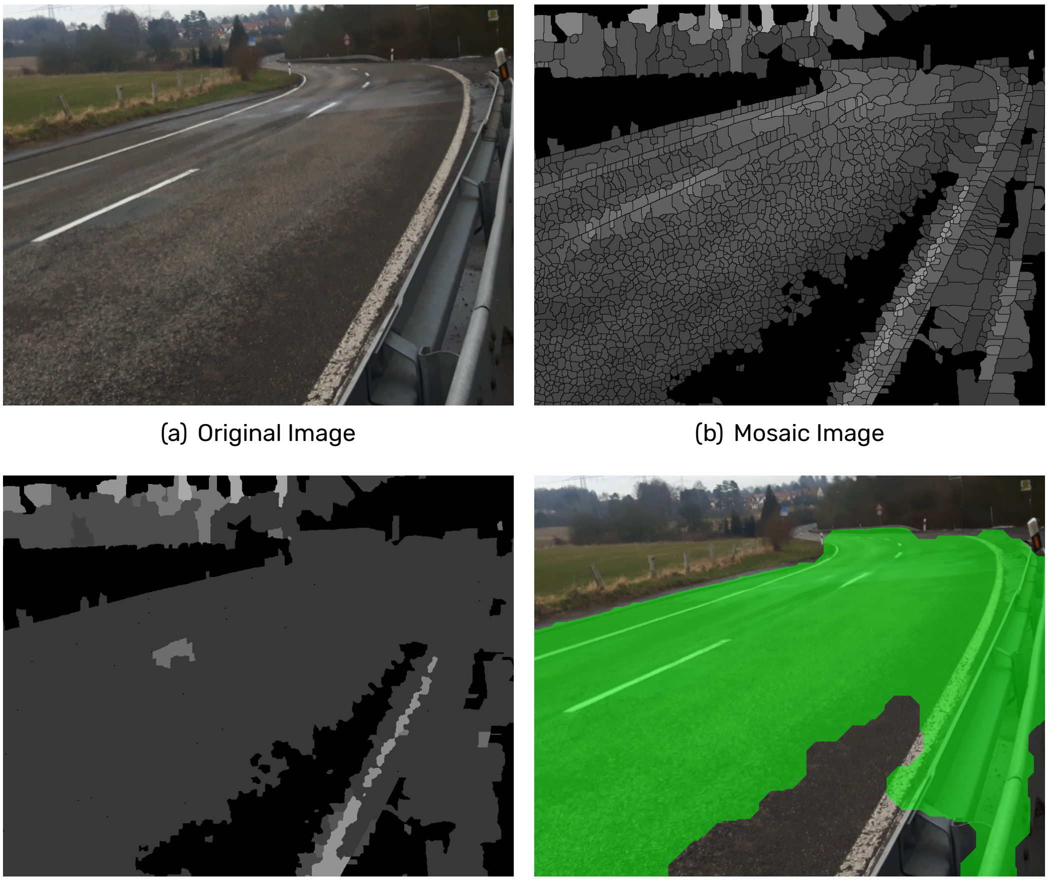
$$R_n = \bigcup_{i=0}^{n-1} R_{n-i}.$$

In tests, $n = 10$ led to good results.

Figure 6.4 shows intermediate steps of this algorithm for an example image. Even though not the complete road is extracted, the segmentation counts as successful because there are enough road details to determine the course of the road. Furthermore, it is important that the bus stop traffic sign on the right side was not extracted as road.

6.2.3 Proposed ASVI Adaptation based on ML

The proposed algorithm is based on [123] and has three stages which are shown in Figure 6.5.



(a) Original Image
(b) Mosaic Image
(c) Merged Mosaic Image
(d) Result: Marked Road

Figure 6.4: Intermediate Steps of RBS based on Watersheds (Cropped Images).

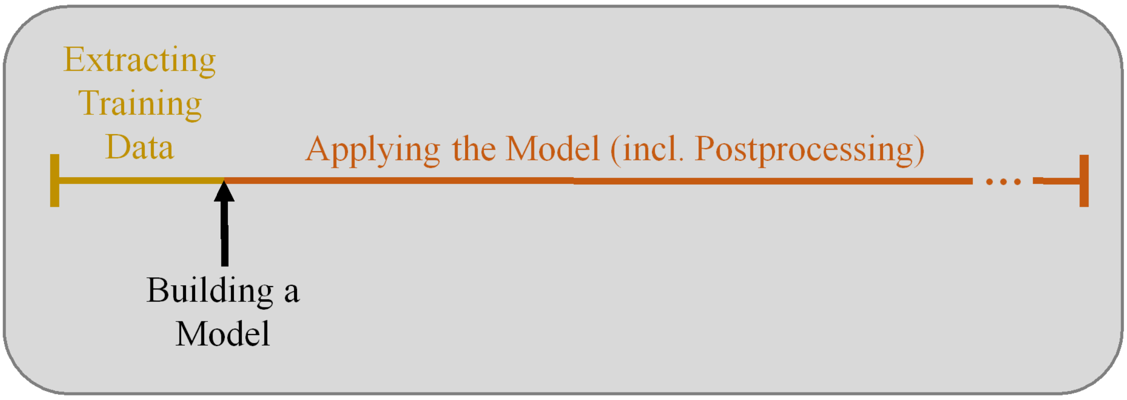


Figure 6.5: Timeline of a Video Sequence

The first frames of a video sequence are used to automatically extract training data. Afterwards, these data are used to build a model which can then be applied to the remaining frames of the scene. Application of the model includes a postprocessing step.

Figure 6.6 shows the procedure of the adapted algorithm and the differences to the underlying ADAS algorithm.

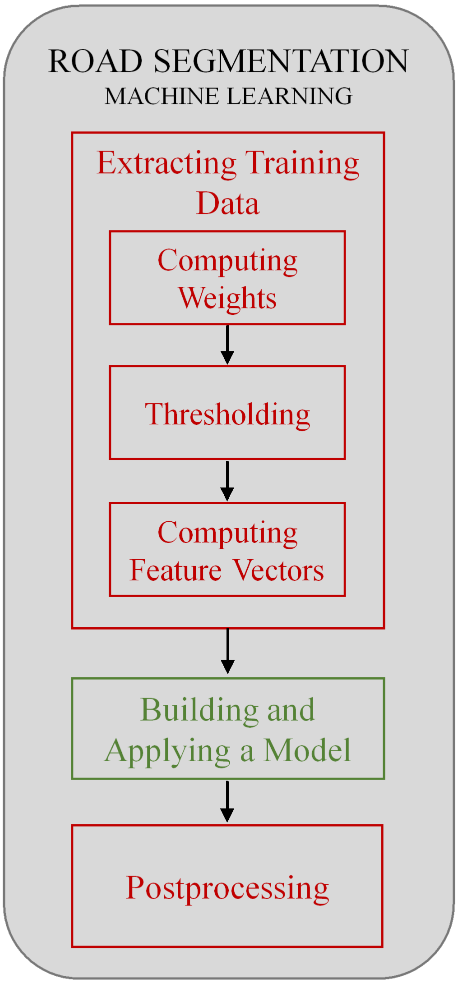


Figure 6.6: Procedure of Presented RBS based on ML. Green boxes are taken from [123] with small changes. Red boxes required major changes or are newly introduced.

Contrarily to [123], a stable template for road and background areas as in Figure 6.2 cannot be used to extract training data. An according example for frames from the *CoPeD* data set is shown in Figure 6.7. With the same template, training data are chosen correctly from driver perspective (on the left) but only partially correct from pedestrian perspective (on the right). Therefore, I developed a new system to extract training data (1) by computing weights of image blocks (1.1) and deciding via treshold-ing (1.2) which blocks are most likely to be part of the road respectively background. Afterwards, feature vectors of these blocks are computed (1.3). The feature vectors in [123] consist of 26 entries: For each RGB colour channel, independent eight bin colour histograms are computed. Additionally, x and y coordinates are stored. The presented algorithm uses only four entries. The last step, postprocessing (3), consists of morphological operations and simple tracking.

(1) Extracting Training Data
The first frames of a sequence are used to extract training data for a ML method. The extraction consists of three steps: The frames are divided into non-overlapping blocks. In the first step, weights for each block are computed indicating their likeliness of belonging to the road or background. The second step is a thresholding process that

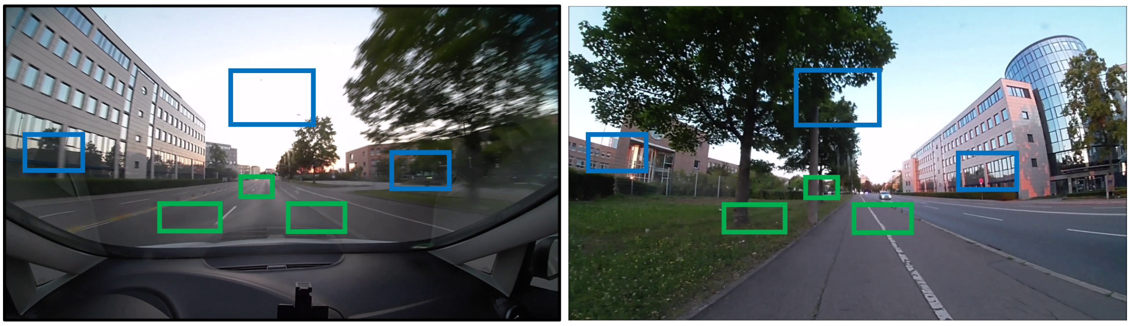


Figure 6.7: Left: Frame from Driver Perspective with a Training Template according to [123]. Right: Same Template applied to a Frame from Pedestrian Perspective. Green: Road. Blue: Background.

labels certain blocks as road respectively background based on the weights whereas other blocks are not considered for training. Finally, feature vectors of the labelled blocks are computed for training purposes.

(1.1) Computing Weights
For each block, two weights are computed. The first weight, w_R, indicates if the block is likely to be part of the road, whereas the second weight, w_{BG}, implies the likelihood of the block belonging to the background. Both weights are weighted means, which are made up of three additional weights depending on gray value (w_g), saturation (w_s), and y coordinate (w_y).

Let $g(b)$ be the mean gray value of a block b, $s(b)$ the mean saturation, and $y(b)$ the y coordinate of the block's middle pixel. The three values are normalized to $[0, 1]$. The weights are then defined as

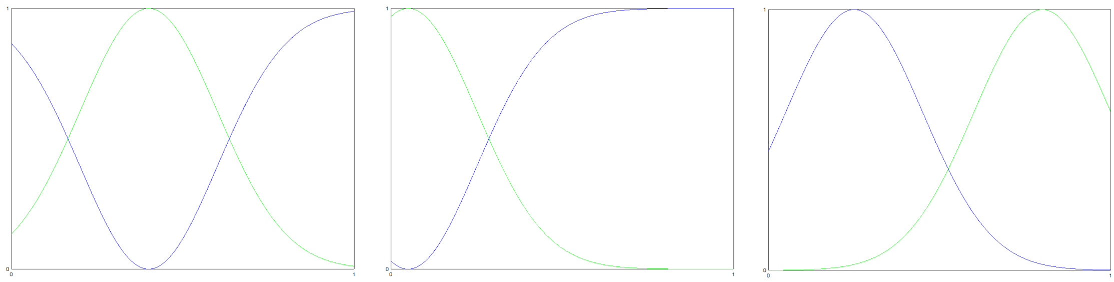
$$\begin{aligned}
w_{g_R}(b) &= G_\sigma(g(b), m_g) \\
w_{g_{BG}}(b) &= 1 - w_{g_R}(b) \\
w_{s_R}(b) &= G_\sigma(s(b), m_s) \\
w_{s_{BG}}(b) &= 1 - w_{s_R}(b) \\
w_{y_R}(b) &= G_\sigma(y(b), m_{y_R}) \\
w_{y_{BG}}(b) &= G_\sigma(y(b), m_{y_{BG}}) \\
w_R(b) &= \frac{2 \cdot w_{g_R}(b) + 2 \cdot w_{s_R}(b) + w_{y_R}(b)}{5} \\
w_{BG}(b) &= \frac{2 \cdot w_{g_{BG}}(b) + 2 \cdot w_{s_{BG}}(b) + w_{y_{BG}}(b)}{5}
\end{aligned}$$

with

$$G_\sigma(x, m) = \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right).$$

The idea behind these weights is that road regions fall in a specific gray value range with low saturation. Furthermore, the road is more likely to appear in the lower regions than in the upper regions of an image, and vice versa for the background. Gray value and saturation are given higher weights than the y coordinate because they can be predicted more easily. The values $m_g, m_s, m_{y_R}, m_{y_{BG}}$ are variables that have to be set for each sequence.

Figure 6.8 shows the weights for road (green) and background (blue) with $m_g = 0.4, m_s = 0.05, m_{y_R} = 0.8, m_{y_{BG}} = 0.25$, and $\sigma = 0.2$.



(a) Gray Value: w_{g_R} (green) and $w_{g_{BG}}$ (blue) with $m_g = 0.4$ and $\sigma = 0.2$

(b) Saturation: w_{s_R} (green) and $w_{s_{BG}}$ (blue) with $m_s = 0.05$ and $\sigma = 0.2$

(c) y Coordinate: w_{y_R} (green) and $w_{y_{BG}}$ (blue) with $m_{y_R} = 0.8$, $m_{y_{BG}} = 0.25$, and $\sigma = 0.2$

Figure 6.8: Weights for (a) Gray Value, (b) Saturation, and (c) y Coordinate. Green: Road. Blue: Background.

(1.2) Thresholding
After computing the weights $w_R(b)$ and $w_{BG}(b)$ for each block, thresholds th_R and th_{BG} decide which blocks are labelled as road (1), background (0), or not considered for training:

$$\text{label}(b) = \begin{cases} 1, & \text{if } w_R(b) > th_R \\ 0, & \text{if } w_{BG}(b) > th_{BG} \\ \text{None}, & \text{else} \end{cases}$$

(1.3) Computing Feature Vectors
The feature vectors have four entries: (1) Mean gray value, (2) mean saturation, (3) y coordinate of a block's middle pixel, and (4) x coordinate of a block's middle pixel. For each labelled block, several feature vectors are computed by creating smaller half-overlapping blocks within the blocks. Every feature vector of a smaller block inherits the label of the bigger block it belongs to. The feature vectors and their labels form the automatically extracted training data that can now be passed to a ML method.

(2) Building and Applying a Model
The training data extracted from the first frames of a sequence are passed to a ML method. The corresponding model is used on the remaining frames of a sequence to separate road and background pixels.

For every frame, a matrix Z of the same size filled with zeros is initialized. For overlapping blocks, it is predicted if the block belongs to the road or the background depending on the weights of this block (w_R, w_{BG}), the thresholds (th_R, th_{BG}), and the trained model (Mdl).

For every block b, the four dimensional feature vector $v(b)$ and the weights $w_R(b)$ and $w_{BG}(b)$ are computed.

The matrix Z is then altered as follows:

$$Z(b) = \begin{cases} Z(b) + 1, & \text{if } w_R(b) > th_R \\ Z(b), & \text{if } w_{BG}(b) > th_{BG} \\ Z(b) + \text{predict}(Mdl, v(b)), & \text{else} \end{cases}$$

The trained model is only used if the weights are below their respective threshold which reduces the computation time.

Because of the overlapping blocks, every pixel is considered up to four times. After adding up the labels in the matrix Z, each pixel p with a value of four is declared as road R whereas the others are considered background:

$$R(p) = \begin{cases} 1, & \text{if } Z(p) = 4 \\ 0, & \text{else} \end{cases}$$

(3) Postprocessing
The postprocessing consists of morphological operations applied to the road mask R and a simple tracking step. Based on the assumption that the road sections of an image are connected, it is part of the morphological postprocessing to select the largest connected component. As a simple tracking step, the pixel-wise union of the last n frames is computed. Hence, the road of the n-th frame in a sequence is determined as:

$$R_n(p) = \bigcup_{i=0}^{n-1} R_{n-i}(p)$$

In our tests, we set $n = 10$. Finally, the background BG is defined as the complement of the road R:

$$BG(p) = \neg R(p)$$

Figure 6.9 exemplary shows two images for each considered video sequence. On the left side, one of the training images can be seen with green and blue boxes indicating if a block's content is used as training data for road respectively background. The right column shows a later frame from the according sequence classified with a SVM. Pixels detected as road are marked green.

6.3 Crosswalk Detection

6.3.1 Related Work

The adapted crosswalk detection algorithm presented in this chapter is based on the work of Choi et al. [61]. They introduce a combined detection of crosswalks and traffic lights. Their crosswalk detection is based on a 1-D mean filter in horizontal direction. Another example of an algorithm using the fact that crosswalks are horizontal structures from driver's perspective is described in [99]. Haseloff and Kummert apply Fourier and Hough Transform and thus make use of the bipolarity and straight lines which characterize crosswalks. Zhai et al. propose a crosswalk detection based on MSER and ERANSEC [98].

Figure 6.9: Results for Crosswalk w/ Traffic, Crosswalk w/o Traffic, Straight, Left, Right (from top to bottom). Left: Training Image (Green: Road, Blue: Background). Right: Later Frame Classified with a SVM (Road Marked Green). Cropped Images.

Furthermore, there is some research concerning crosswalk detection in the area of ASVI. The algorithms described in [16] and [15] are based on parallel lines that are extracted by Hough Transform. Cheng et al. [5] extract the bright crosswalk stripes by adaptive thresholding. They address challenging scenarios, such as partial occlusion, low contrast and distant crosswalks, and different illuminations. In addition, they offer an extensive literature review on crosswalk detection algorithms.

6.3.2 Proposed ASVI Adaptation

The presented algorithm is based on Choi et al.'s work [61]. It consists of four steps: After preprocessing and computing the ROI (1), horizontal and vertical 1-D mean filters are applied (2). The differences between original and filtered images are the foundation for binarization and morphological postprocessing (3). Finally, the two resulting masks are combined via a bitwise or-operation and a decision is made if the remaining pixels form a crosswalk or not (4). The original image as well as the extracted road from the road segmentation algorithm are the input variables for this algorithm.

Figure 6.10 shows the procedure of the adapted algorithm and the differences to the underlying ADAS algorithm. Choi et al. [61] developed their algorithm for an ADAS containing a tilted front camera so that the resulting images show only the road but no background. Therefore, no ROI was needed in [61] and I developed the before presented RBS as ROI. Furthermore, a vertical filter was added in order to be able to detect crosswalks from every angle by combining it with the horizontal filter. As two filters are considered, the handling of the masks differs from the one in [61]. Finally, Choi et al. [61] gave no detailed description of their postprocessing and decision step so that I had to develop my own version.

(1) Preprocessing and ROI
As preprocessing, I apply a Gaussian filter for noise reduction. Alternatively, I use the already smoothed RGB image I from the road segmentation algorithm as input. Let $gray$ be the gray value version of I. The ROI is defined by the before computed road R. Therefore, the background pixels in $gray$ are set to zero and the values belonging to the road are kept:

$$gray(x,y) = \begin{cases} gray(x,y), & \text{if } R(x,y) = 1 \\ 0, & \text{else.} \end{cases}$$

(2) 1-D Mean Filters
Choi et al. [61] propose a horizontal 1-D mean filter to detect crosswalks because from driver's perspective crosswalks are horizontal structures. For pedestrians, however, crosswalks appear in every possible angle. Therefore, an additional 1-D mean filter in vertical direction is used. By combining the two filters, it is possible to detect crosswalks occurring in any angle. In order to respect the ROI, pixels having the value zero are excluded from the computation. With that, the results for filtering in horizontal respectively vertical direction are G_x and G_y:

$$G_x(x,y) = \sum_{k=-s}^{s} \frac{gray(x+k,y)}{\#\{(x+k,y) | -s \leq k \leq s \wedge gray(x+k,y) \neq 0\}}$$

$$G_y(x,y) = \sum_{l=-s}^{s} \frac{gray(x,y+l)}{\#\{(x,y+l) | -s \leq l \leq s \wedge gray(x,y+l) \neq 0\}}.$$

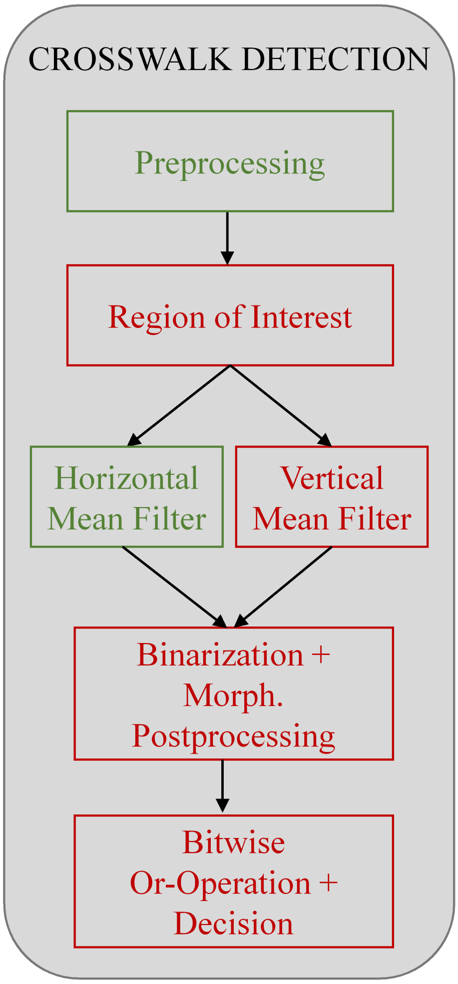


Figure 6.10: Procedure of Presented Crosswalk Detection. Green boxes are taken from [61] with small changes. Red boxes required major changes or are newly introduced.

The size of the filter is in both cases $(2 \cdot s + 1)$.

(3) Binarization and Morphological Postprocessing
The differences between the original image $gray$ and the filtered results G_x and G_y are computed as D_x and D_y:

$$D_x(x, y) = |gray(x, y) - G_x(x, y)|$$
$$D_y(x, y) = |gray(x, y) - G_y(x, y)|.$$

Applying a threshold th results in the masks M_x and M_y:

$$M_x(x, y) = \begin{cases} 1, & \text{if } D_x(x, y) > th \\ 0, & \text{else.} \end{cases}$$
$$M_y(x, y) = \begin{cases} 1, & \text{if } D_y(x, y) > th \\ 0, & \text{else.} \end{cases}.$$

Afterwards, closing followed by opening with a larger structuring element is performed on both masks independently.

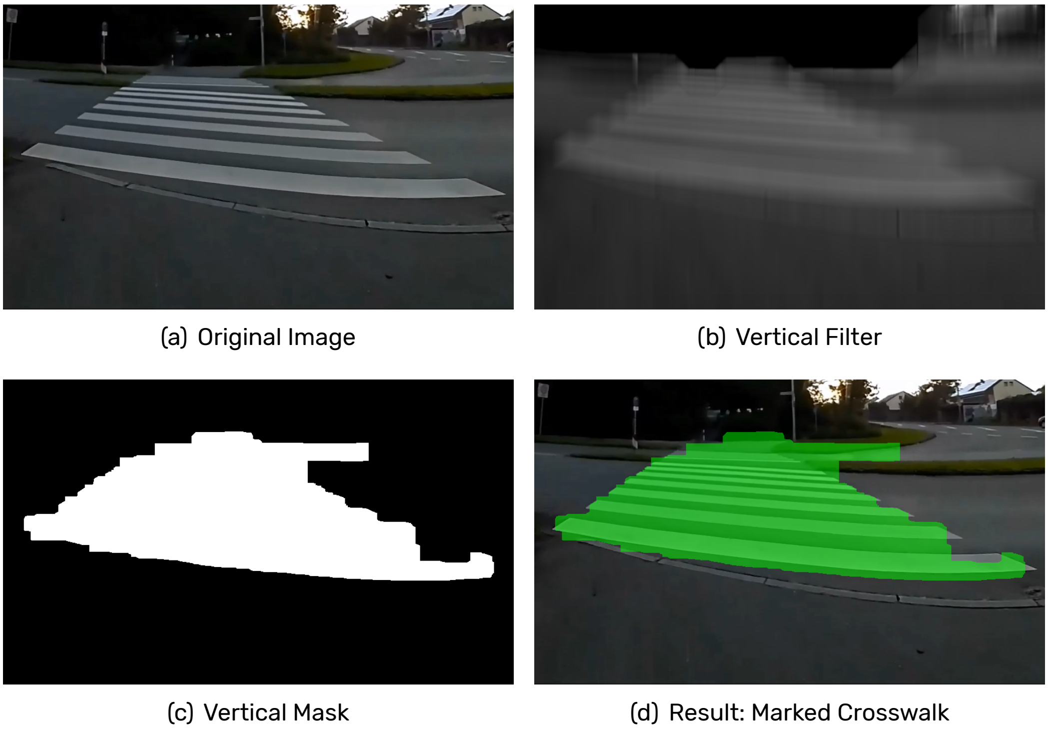


Figure 6.11: Intermediate Steps of the Presented Crosswalk Detection (Cropped Images)

(4) Bitwise Or Operation and Decision
Before combining the two masks, components that cover less than 1.5% of the image are deleted. Then, the two masks are combined to one mask M by applying a pixelwise or operation

$$M(x, y) = M_x(x, y) \vee M_y(x, y).$$

The largest component CW of M is a crosswalk candidate, provided it also exceeds the 1.5% mark. The decision is made based on two values: The extent of CW indicating the relative number of pixels that are set inside a box surrounding CW and the ratio between minor and major axis length of CW.

If both, ratio and extent, are higher than thresholds th_r and th_e, CW is considered a crosswalk. Otherwise, no crosswalk is present and CW is set to zero.

Figure 6.11 shows intermediate results of the crosswalk detection algorithm for an example image. The figure only shows intermediate steps for vertical direction because the crosswalk in this example is a vertical structure. Therefore, the horizontal mask is zero and the vertical mask is identical with the result.

6.4 Lane Detection

6.4.1 Related Work

In ASVI, the goal of lane detection is to help with general orientation by giving information about the course of the road, meaning indicating if the road goes on straight ahead or takes a left respectively right turn. On the other hand, in ADAS, lane detection is used to warn the driver if they risk to depart the lane. Lane departure warning systems consist of three steps, namely lane detection, lane tracking, and communication with the driver [55].

For ASVI, I therefore concentrate on the first step. According to [55], single frame lane detection is generally composed of four steps: (1) Image acquisition and preprocessing, (2) edge detection, (3) stripe identification by Hough Transform or Edge Distribution Function (EDF), (4) line fitting.

As Hough Transform is not suitable to detect even slight curves, the presented research concentrates on using the EDF which is the histogram of the gradient magnitude with respect to the corresponding edge angle. The work described in this thesis is based on the lane detection using EDF presented by Lee [119]. After computing the ROI with the help of the vanishing point, edge extraction as well as EDF construction are carried out. Based on an EDF analysis, Lee identifies if the car is safely within the lane or in risk of departing the lane boundaries.

6.4.2 Proposed ASVI Adaptation

The presented algorithm is based on Lee's work [119]. Six subsequent steps are carried out: After preprocessing and computing the ROI based on the RBS (1), the ROI is divided into a total of eight subimages (2). For every subimage, computation (3) and analysis (4) of the EDF are carried out, resulting in an angle for every subimage. Afterwards, the angles are interpolated (5) and the course of the road ahead is decided (6) based on the concavity of the interpolated function.

Figure 6.12 shows the procedure of the adapted algorithm and the differences to the underlying ADAS algorithm. Lee's ROI computation [119] is based on the vanishing point. Because this procedure cannot be used in ASVI, the before extracted road forms the ROI. In contrast, EDF computation and analysis are mostly taken from [119]. The partition into subimages and the decision steps, interpolation and curve analysis, are newly developed. As the purposes of lane detection differ for ADAS and ASVI, there are no according steps in [119]. Interpolation by linear parabolic fitting is carried out similarly to [55] but is adapted to the presented procedure.

(1) Preprocessing and ROI
This step is analogous to the before presented crosswalk detection algorithm. Additionally, the first row of the road R containing a non-zero entry is computed as min_y.

(2) Partition into Subimages
The image is divided from the bottom of the image to min_y into eight parts. From bottom to top, the first two subimages consist of one fourth of the available rows, the next two of one eighth, and the last four of one sixteenth.

The following two steps are carried out for each of the eight subimages. This will be indicated with the index j, where $j = 1$ refers to the subimage at the bottom which is the one closest to the user.

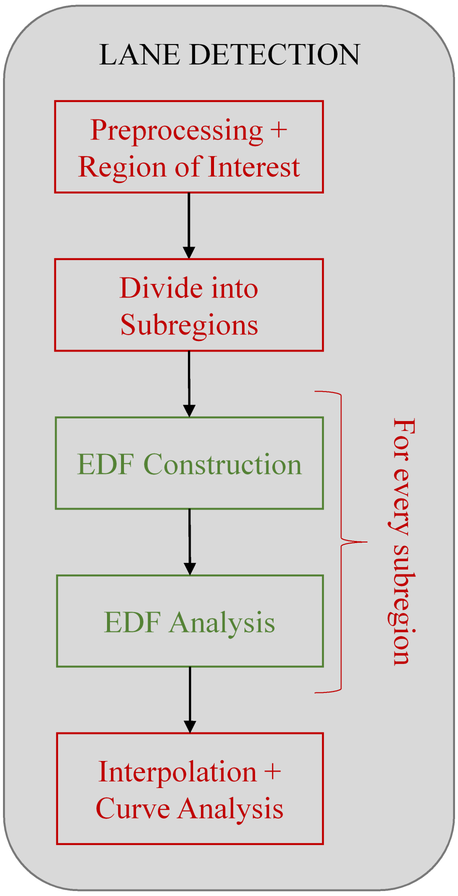


Figure 6.12: Procedure of Presented Lane Detection. Green boxes are taken from [119] with small changes. Red boxes required major changes or are newly introduced.

(3) EDF Computation
The gradient $G_j(x, y) = (G_{j_x}(x, y), G_{j_y}(x, y))^T, j = 1, ..., 8$ is computed by convolving the image with the Sobel masks

$$H_x = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix}$$

and

$$H_y = H_x^T.$$

The magnitude $M_j(x, y)$ is then defined as

$$M_j(x, y) = \sqrt{G_{j_x}^2(x, y) + G_{j_y}^2(x, y)}, j = 1, ..., 8$$

and the angle $\alpha(x, y)$ is computed as

$$\alpha(x, y) = \tan^{-1}\left(\frac{G_{j_y}(x, y)}{G_{j_x}(x, y)}\right), j = 1, ..., 8,$$

so that the result is in the range $[1, 180]$. The degree values are rounded to integers.

Before determining the EDF, the 97%-quantile of M_j is computed and the values below it are set to zero. The EDF is then defined as

$$EDF_j(i) = \sum_{i=\alpha(x,y)} M_j(x,y), j = 1, ..., 8.$$

(4) EDF Analysis
First, $EDF_j, j = 1, ..., 8$ is smoothed with a 1-D filter of size 15. Afterwards, the relative value of the highest peak $p_j \in [0,1], j = 1, .., 8$ and its position $\theta_j \in [1, 180], j = 1, , 8$ are computed. In case of no occurring peak, p_j is set to zero and θ_j is set to 180.

(5) Interpolation: Linear Parabolic Fitting
For interpolation, the lane in the first two subimages from the bottom, the ones closest to the user, is considered as linear and the remaining lane as parabolic. In order to be able to interpolate independently from image size, the values are normalized to $[0,1]$. Furthermore, the axes are changed, meaning that the x-values are defined according to the size computed in step (2). With that, the interpolation function $f(x)$ is defined as

$$f(x) = \begin{cases} a \cdot (x - 0.5) + b, & \text{if } x \leq 0.5 \\ a \cdot (x - 0.5) + b + c \cdot (x - 0.5)^2, & \text{else.} \end{cases}.$$

and the x-values are $x = (0, 0.25, 0.5, 0.625, 0.75, 0.8125, 0.875, 0.9375, 1)^T$.

The according y-values are computed by determining a line with angle θ_j:

$$\begin{aligned} y_1 &= 0 \\ y_j &= y_{j-1} + (x_j - x_{j-1}) \cdot \tan(\theta_{j-1} - 90), j = 2, ..., 9. \end{aligned}$$

It is necessary to subtract 90 from the angle because the axes are rotated. After computing the y-values, they are normalized to $[0,1]$.

To get the interpolation function f, the mean square error solution of the following over-determined system of equations is computed:

$$\begin{pmatrix} x_1 - 0.5 & 1 & 0 \\ x_2 - 0.5 & 1 & 0 \\ x_3 - 0.5 & 1 & 0 \\ x_4 - 0.5 & 1 & (x_4 - 0.5)^2 \\ x_5 - 0.5 & 1 & (x_5 - 0.5)^2 \\ x_6 - 0.5 & 1 & (x_6 - 0.5)^2 \\ x_7 - 0.5 & 1 & (x_7 - 0.5)^2 \\ x_8 - 0.5 & 1 & (x_8 - 0.5)^2 \\ x_9 - 0.5 & 1 & (x_9 - 0.5)^2 \end{pmatrix} \cdot \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \end{pmatrix}$$

(6) Decision
There are four possible outcomes for the algorithm: The input image does not contain lanes, the road goes straight ahead, or it takes a right or left turn.

First, it is determined if the image contains lanes by checking if the mean value of the highest peaks $\overline{p} = \sum_{j=1}^{8} p_j$ is within a certain range $[th_{p_{low}}, th_{p_{high}}]$. This is based on the idea that there is a characteristic amount of pixels belonging to the lane boundaries that indicate the direction. In tests, the according pixels made up between 1.4% and 3% of the subimage.

If $\overline{p}$ lies outside the specified range, the output text is set to "No lane detected". Otherwise, the course of the road is checked by examining the parameter c from the interpolation function and the variance of the θ-values $\overline{\overline{\theta}} = \frac{1}{7} \cdot \sum_{j=1}^{8}(\theta_j - \overline{\theta})$, where $\overline{\theta}$ is the mean θ value. In general, a negative c value means that the parabola is concave down and the road takes a right turn; accordingly, a positive value of c means that the parabola is concave up and the road takes a left turn. The higher the absolute value of c, the tighter the curve of the road; the lower the absolute value of c, the straighter the road. At the same time, the variance of the angles within the image is higher when the curve is tighter. With this knowledge, a four-step process to determine the course of the road depending on several thresholds is set up:

1. If $|c| < th_{c_{low}}$, the output text is "Straight ahead".
2. If $th_{c_{low}} \leq |c| < th_{c_{mid}}$, the output is determined as

$$\text{text} = \begin{cases} \text{"Straight ahead"} & \text{if } \overline{\overline{\theta}} < th_{var_{high}} \\ \text{"Left turn"}, & \text{if } \overline{\overline{\theta}} \geq th_{var_{high}} \wedge c > 0 \\ \text{"Right turn"}, & \text{else.} \end{cases}$$

3. If $th_{c_{mid}} \leq |c| < th_{c_{high}}$, the output is determined as

$$\text{text} = \begin{cases} \text{"Straight ahead"} & \text{if } \overline{\overline{\theta}} < th_{var_{low}} \\ \text{"Left turn"}, & \text{if } \overline{\overline{\theta}} \geq th_{var_{low}} \wedge c > 0 \\ \text{"Right turn"}, & \text{else.} \end{cases}$$

4. If $|c| \geq th_{c_{high}}$, the output is determined as

$$\text{text} = \begin{cases} \text{"Left turn"}, & \text{if } c > 0 \\ \text{"Right turn"}, & \text{else.} \end{cases}$$

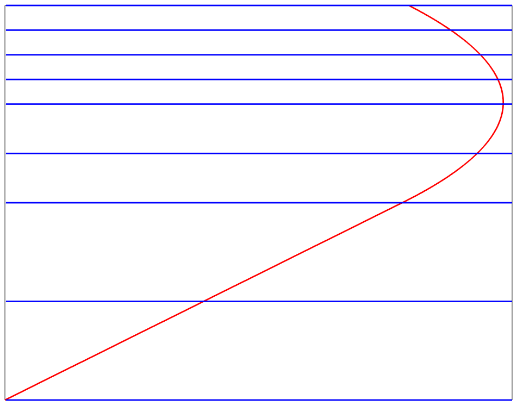
The mean values of $\overline{p}$ and $\overline{\overline{\theta}}$ for the last 15 frames are computed and the output text is set to the one occurring the most in the last 15 frames.

Figure 6.13 shows the interpolated function for an example image. In general, the interpolated function does not match the course of the road exactly but is a good approximation and has identical concavity.

6.5 Evaluation

The four before described algorithms are implemented in Matlab Version R2017b [69] using the *Image Processing and Computer Vision Toolbox*. For the ML based RBS, I additionally used the *Deep Learning* as well as *Statistics* and *Machine Learning Toolboxes*.

(a) Original Image (Cropped)



(b) Interpolated Function (Blue lines mark the subimages)

Figure 6.13: Lane Detection Result

The algorithms are tested on a subset of the *CoPeD* data set [105]. In the following, the relative path to the used sequences inside the folder that can be downloaded from the *CoPeD* website[1] is given:

- Crosswalk with traffic (CW w/ traffic), first 243 frames:
  CoPeD\2 Crossings\Pedestrian\Pedestrian_Crosswalk_Traffic.mp4
- Crosswalk without traffic (CW w/o traffic), first 195 frames:
  CoPeD\2 Crossings\Pedestrian\Pedestrian_Crosswalk_NoTraffic.mp4
- Straight, 74 frames:
  CoPeD\1 Lane Detection\Pedestrian\Others\Pedestrian_Straight_2.mp4
- Left, 489 frames:
  CoPeD\1 Lane Detection\Pedestrian\Others\Pedestrian_Left_2.mp4
- Right, 585 frames:
  CoPeD\1 Lane Detection\Pedestrian\Others\Pedestrian_Right_2.mp4

Before evaluating the algorithms, an evaluation of the training data that is passed to the ML method in the case of the ML-based RBS is presented. Table 6.2 shows how many of the blocks used for training are correctly respectively wrongly detected. Blocks which contain both, road and background, are treated as correct detection. For all sequences, blocks of size 60×60 were used in the first five frames of the sequence. Training data was collected in overlapping blocks of size 40×40 inside the detected 60×60 blocks. The numbers in Table 6.2 refer to the number of detected 60×60 blocks. With a hit rate of 99.84 % very reliable data are passed to the ML methods.

Tables 6.3, 6.4, 6.5, 6.6, and 6.7 show the evaluation for RBS based on watersheds, RBS based on ML using SVM (Matlab command `fitcsvm`), RBS based on ML using NN (Matlab command `patternnet`), crosswalk detection, and lane detection. In the first four cases, nine frames less than stated above are considered, in the last case 14 frames less. This is because the union of the last ten frames is computed for RBS and the mean of different values of the last 15 frames for lane detection is formed. The output

[1] `http://dataset.informatik.hs-furtwangen.de/`, accessed on June 6, 2020

Table 6.2: Evaluation of Training Data Blocks (CD: Correct Detection, WD: Wrong Detection, CW: Crosswalk, BG: Background)

	CW w/ traffic	CW w/o traffic	Straight	Left	Right	Total
CD Road	237	124	362	858	442	2023
CD BG	1620	1680	801	1136	806	6043
Sum CD	**1857**	**1804**	**1163**	**1994**	**1248**	**8066**
WD Road	0	0	0	6	0	6
WD BG	0	0	0	0	7	7
Sum WD	**0**	**0**	**0**	**6**	**7**	**13**
Hit Rate	**100%**	**100%**	**100%**	**99.7%**	**99.44%**	**99.84%**

of the watershed-based RBS is used as input for the ROI computation of crosswalk and lane detection.

For RBS, an over-detection means that in addition to the road, important data belonging to the background according to Figure 6.1 (traffic lights and signs) close to the user, at least 20×20 pixels in size, are extracted. An under-detection means that not enough of the road is extracted in order to detect the road's markings.

For crosswalk detection, the goal is to detect the crosswalk in all frames of the first two sequences and no crosswalk in the other sequences. In the case of lane detection, it is the other way around: For the first two sequences, the result should be "No lane detected", whereas for the latter three the respective course of the road should be determined.

For RBS, the hit rates are with 99.87 % (watersheds), 99.41 % (ML-based with SVM), and 99.87 % (ML-based with NN) in comparable ranges. The proposed ML-based algorithm offers advantages in speed and topicality of the used methods compared to the watershed-based algorithm. The hit rate of crosswalk detection is 98.64 % whereas the proposed lane detection reaches a hit rate of 97.89 %.

It would be useful to compare the evaluations with results from the underlying ADAS algorithms, but this is difficult to achieve because they were tested on other, not publicly available data. Choi et al. [61] report 96.2 % correctly detected, present crosswalks and 0.66 % false positives for a data set of 21864 frames from which 1053 contain a crosswalk. In my case, the according numbers are with 96.9 % and 0.76 % in a similar range. For lane detection, Lee [119] tested on 1200 frames, if the car departed from the road or not. The reached hit rate is 96.42 %. Beucher et al. [120] present their results on a small number of single frames. Foedisch and Takeuchi [123] report less than 5 % misclassifications. Even though it is not possible in all cases to compare my hit rates with the ones of the underlying ADAS algorithms, it can be stated that the adapted algorithms perform well and reach satisfying hit rates that make them applicable for ASVI.

Table 6.3: Evaluation RBS based on Watersheds (NF: Number of Frames, CD: Correct Detection, OD: Over-Detection, UD: Under-Detection)

	CW w/ traffic	CW w/o traffic	Straight	Left	Right	Total
NF	234	186	65	480	576	1541
CD	234	186	65	478	576	1539
OD	0	0	0	2	0	2
UD	0	0	0	0	0	0
Hit Rate	**100 %**	**100 %**	**100 %**	**99.59 %**	**100 %**	**99.87 %**

Table 6.4: Evaluation RBS based on ML with SVM (NF: Number of Frames, CD: Correct Detection, OD: Over-Detection, UD: Under-Detection)

	CW w/ traffic	CW w/o traffic	Straight	Left	Right	Total
NF	234	186	65	480	576	1541
CD	234	186	65	471	576	1532
OD	0	0	0	4	0	4
UD	0	0	0	5	0	5
Hit Rate	**100 %**	**100 %**	**100 %**	**98.13 %**	**100 %**	**99.41 %**

Table 6.5: Evaluation RBS based on ML with NN (NF: Number of Frames, CD: Correct Detection, OD: Over-Detection, UD: Under-Detection)

	CW w/ traffic	CW w/o traffic	Straight	Left	Right	Total
NF	234	186	65	480	576	1541
CD	234	186	65	478	576	1539
OD	0	0	0	2	0	2
UD	0	0	0	0	0	0
Hit Rate	**100 %**	**100 %**	**100 %**	**99.59 %**	**100 %**	**99.87 %**

Table 6.6: Evaluation Crosswalk Detection (NF: Number of Frames, CD: Correct Detection, ND: Not Detected if crosswalk is present, FP: False Positive)

	CW w/ traffic	CW w/o traffic	Straight	Left	Right	Total
NF	234	186	65	480	576	1541
CD	226	181	65	472	576	1520
ND	8	5	0	0	0	13
FP	0	0	0	8	0	8
Hit Rate	**96.71 %**	**97.37 %**	**100 %**	**98.36 %**	**100 %**	**98.64 %**

Table 6.7: Evaluation Lane Detection (NF: Number of Frames, CD: Correct Detection, WD: Wrong Detection)

	CW w/ traffic	CW w/o traffic	Straight	Left	Right	Total
NF	229	181	60	475	571	1516
CD	229	181	60	473	541	1484
WD	0	0	0	2	30	32
Hit Rate	**100 %**	**100 %**	**100 %**	**99.58 %**	**94.75 %**	**97.89 %**

6.6 Conclusion

In the course of this thesis, I identified a total of seven use cases that have to be considered for the development of a transfer concept. This chapter focused on the detection of markings on the road, namely crosswalk and lane detection.

In order to solve the ROI problem and to reduce the computational cost, I first introduced a RBS. Thereby, crosswalk and lane detection can be carried out on the road only. Algorithms for traffic sign and traffic light detection to be developed in the future will be applied to the background part of the image, whereas obstacles will be detected on the whole image.

Two adaptation from ADAS to ASVI for RBS are presented. The first algorithm is based on Beucher et al.'s [120] use of morphological watersheds and in addition uses thresholds for the mean gray and saturation values of each catchment basin. The adaptation reached a hit rate of 99.87 %. The second algorithm is based on Foedisch and Takeuchi's [123] idea of using NN and introduces weights to determine if a block of an image is likely to be part of the road respectively the background. The according hit rates are 99.41 % when using a SVM and 99.87 % when using a NN.

The basis of the developed crosswalk detection is Choi et al.'s idea of using a horizontal mean filter [61]. I combined the horizontal with a vertical filter and defined a decision process. The recognition rate of the adapted algorithm was 98.64 %.

The suggested lane detection uses the EDF. Instead of applying the EDF to the whole image as in Lee's ADAS work [119], I first divided the image into subimages of decreasing size from bottom to top. Afterwards, interpolation of the angles and analysis of the according function returned the course of the road. Correct detection occurred in 97.89 % of the examined frames.

With that, it can be stated that the presented ASVI algorithms, adapted from according ADAS methods, achieved overall good hit rates and thus are applicable for ASVI.

6.7 Thesis 3

In the following the two parts of the objective formulated in 3.3 are discussed.

(O3.1): It has to be shown that determining the ROI for ASVI detection algorithms can in general not be taken from ADAS and that adapting a RBS from ADAS to ASVI solves this problem.
As explained at the beginning of this chapter, knowing position and angle of the camera is exploited in ADAS. However, in ASVI no such assumptions can be made. Therefore, a RBS as an additional use case is proposed. Adapting a RBS from ADAS to ASVI makes it possible to carry out some detection algorithms on a subset of the original image (see Figure 6.1). Altogether, this confirms that objective (O3.1) is achieved.

(O3.2): Adaptations of algorithms from ADAS to ASVI have to developed and implemented. The adapted algorithms have to achieve similar hit rates as the underlying ADAS algorithms.
As seen in section 6.5, a comparison between the performances of the underlying ADAS algorithms and their ASVI adaptations is difficult to achieve because so far they have not been tested on the same data. In the future, the ADAS algorithms can be implemented and tested on the according comparable video sequences from the *CoPeD* data set. Nevertheless, (O3.2) is achieved for RBS, crosswalk detection, and lane detection because the presented hit rates are similar to the ones reported in the literature (provided hit rates are reported) and high enough to make the algorithms applicable in ASVI.

Hence, objective (O3.1) is achieved and (O3.2) is achieved for RBS, lane detection, and crosswalk detection. These findings are summarized in Thesis 3.

Thesis 3: Adaptation Possibilities

I showed that adapting a RBS from ADAS to ASVI solves the ROI problem and that ASVI adaptations for RBS, lane detection, and crosswalk detection achieve similar hit rates as the underlying ADAS algorithms.

(T3.1) I showed that determining the ROI for ASVI detection algorithms can in general not be taken from ADAS and that adapting a RBS from ADAS to ASVI solves this problem.

(T3.2) I developed and implemented adaptations of algorithms from ADAS to ASVI for RBS, lane detection, and crosswalk detection. I proved that the adapted algorithms in these three cases achieve satisfying and similar hit rates as the underlying ADAS algorithms.

Own publications supporting Thesis 3 are: [39, 74, 104, 105, 117, 118, 128].

Chapter 7

Theses and Contributions

This chapter summarizes the theses confirmed throughout the project and lists the contributions made in the course of this thesis.

7.1 Theses

Three groups of objectives were addressed in the course of the research resulting in three confirmed theses. In the following, I repeat the theses and list my own publications with relation to the respective thesis.

Thesis 1: Traffic Scenarios and Use Cases

I defined the significant traffic scenarios for visually impaired pedestrians and determined all vision use cases of relevance in these scenarios. From that, I determined the overlap of vision use cases between ADAS and ASVI. Besides, I introduced the idea of using software engineering methods for the presentation of qualitative data.

(T1.1) I showed that the traffic scenarios of interest for visually impaired pedestrians are: *Orientation Scenarios* (*General Orientation*, *Navigating to an Address*), *Pedestrian Scenarios* (*Crossing a Road*, *Obstacle avoidance*), and *Public Transport Scenarios* (*Boarding a Bus*, *At the Train Station*).

(T1.2) I determined all vision use cases that can support the visually impaired in traffic situations: (1) Traffic light pole detection, (2) traffic light (state) detection, (3) bicycle detection, (4) (driving) vehicle detection, (5) stairs detection, (6) construction site detection, (7) crosswalk detection, (8) obstacle detection, (9) lane detection, (10) curb information, (11) TGGS detection, (12) traffic sign detection, (13) house number detection, (14) description of surroundings, (15) OCR, (16) door detection, and (17) display detection.

(T1.3) I determined the overlap of vision use cases addressed in ADAS and needed in ASVI: (1) Lane detection, (2) crosswalk detection, (3) traffic sign detection, (4) traffic light (state) detection, (5) (driving) vehicle detection, (6) obstacle detection, and (7) bicycle detection.

(T1.4) I introduced the idea of using (adapted) software engineering methods to cluster and present qualitative data.

Own publications related to Thesis 1: [70, 73, 74, 104].

Thesis 2: Video Data Acquisition

(T2): I created the data set *CoPeD* containing comparable video data from driver and pedestrian perspective and covering the overlapping use cases from ADAS and ASVI.

Own publications related to Thesis 2: [104, 105].

Thesis 3: Adaptation Possibilities

I showed that adapting a RBS from ADAS to ASVI solves the ROI problem and that ASVI adaptations for RBS, lane detection, and crosswalk detection achieve similar hit rates as the underlying ADAS algorithms.

(T3.1) I showed that determining the ROI for ASVI detection algorithms can in general not be taken from ADAS and that adapting a RBS from ADAS to ASVI solves this problem.

(T3.2) I developed and implemented adaptations of algorithms from ADAS to ASVI for RBS, lane detection, and crosswalk detection. I proved that the adapted algorithms in these three cases achieve satisfying and similar hit rates as the underlying ADAS algorithms.

Own publications related to Thesis 3: [39, 74, 104, 105, 117, 118, 128].

7.2 Contributions

Below, I first summarize the contributions made in the respective chapters of this thesis and then list my own publications connected to this thesis.

Related Work and Novelty

I reviewed the literature in the fields of camera-based ASVI and camera-based ADAS. I derived a general composition of ASVI from the literature; for ADAS the according composition was already discussed in the literature. Furthermore, I clustered camera-based ASVI into four application areas: Reading out text, recognizing faces and objects, perceiving the environment as well as navigation and collision avoidance. From the similarity of the compositions, the overlap in use cases, and the fact that there is no comprehensive assistive system for the visually impaired in traffic situations, I concluded the need for and novelty of a transfer concept for camera-based algorithms from ADAS to ASVI.

Own publications related to this topic: [7, 39, 104].

Traffic Scenarios and Vision Use Cases

I reviewed the literature containing studies about the requirements of visually impaired people in traffic situations. From that, I concluded the necessity of conducting an own study. In the following, I designed, conducted, and evaluated a qualitative interview study with four experts and ten MTG. With the help of the acquired data, I defined six traffic scenarios and 17 vision use cases with importance to visually impaired pedestrians. Forming the overlap with use cases addressed in ADAS revealed seven use cases. I presented a literature review of ADAS solutions for these use cases. Furthermore, I answered questions concerning age, gender, use of technology, trips to unknown addresses, asking for support, and use case importance, and I introduced the idea of using (adapted) software engineering methods for clustering and presentation of qualitative data.

The *CoPeD* Data Set

I reviewed literature about video and image data sets for traffic scenarios from driver and pedestrian perspective. As the existing data sets did not cover all needed use cases from both perspectives, I designed and developed the publicly hosted *CoPeD* data set containing comparable sequences from driver and pedestrian perspective for all seven overlapping use cases. It is licensed under the Creative Commons Attribution 4.0 International License which allows everyone, even in commercial contexts, to use, modify, and redistribute the data as long as appropriate credit is given.

Use Case Examination

I introduced RBS as a further use case to be considered in order to solve the ROI problem. I performed a literature review for RBS, crosswalk detection, and lane detection. I then developed adaptations to ASVI for Beucher et al.'s watershed-based RBS [120], Foedisch and Takeuchi's ML-based RBS [123], Choi et al.'s crosswalk detection based on a 1-D mean filter [61], and Lee's EDF-based lane detection [119]. The adaptations were implemented in Matlab [69] and evaluated on sequences from *CoPeD*.

List of Own Publications

The following list contains my publications with relation to this thesis. The list is arranged by decreasing date so that the most recent publication is on the top. Parts of these works were presented and cited in this thesis.

- J. Jakob and J. Tick: "Extracting training data for machine learning road segmentation from pedestrian perspective," in IEEE 24th International Conference on Intelligent Engineering Systems, Virtual Event, July 2020, Accepted for Publication, [118]:
  - Literature review for RBS.
  - Description of proposed ASVI adaptation for RBS based on ML.
  - Evaluation of the proposed algorithm on *CoPeD* sequences.

- J. Jakob and J. Tick: "Camera-based on-road detections for the visually impaired," in Acta Polytechnica Hungarica, vol. 17, no. 3, pp. 125 - 146, 2020, [117]:
  - Literature review for RBS, crosswalk detection, and lane detection.
  - Description of proposed ASVI adaptations for RBS based on watersheds, crosswalk detection, and lane detection.
  - Evaluation of the proposed adaptations on *CoPeD* sequences.
- J. Jakob and J. Tick: "Towards a transfer concept from camera-based driver assistance to the assistance of visually impaired pedestrians," in IEEE 17th International Symposium on Intelligent Systems and Informatics, pp. 53 - 60, Subotica/Serbia, September 2019, [104]:
  - Preliminary versions of the objectives treated in this thesis.
  - Summary of camera-based ASVI and camera-based ADAS.
  - Summary of the qualitative interview study.
  - Summary of *CoPeD*.
  - Summary of adapted algorithms for RBS, crosswalk detection, and lane detection.
- J. Jakob and J. Tick: "CoPeD: Comparable pedestrian driver data set for traffic scenarios," in IEEE 18th International Symposium on Computational Intelligence and Informatics, pp. 87 - 92, Budapest/Hungary, November 2018, [105]:
  - Literature review of existing data sets from driver and pedestrian perspective.
  - Conditions and content of the *CoPeD* data set.
  - Preliminary version of the crosswalk and lane detection presented in this thesis.
- J. Jakob, K. Kugele, and J. Tick: "Defining traffic scenarios for the visually impaired," in The Qualitative Report, 2018, Under Review (Accepted into Manuscript Development Program), [70]:
  - Literature review containing studies about the requirements of visually impaired people in traffic situations.
  - Design of the qualitative interview study (expert interviews and interviews with MTG).
  - Evaluation of the qualitative interview study (expert interviews and interviews with MTG).
- J. Jakob and J. Tick: "Traffic scenarios and vision use cases for the visually impaired," in Acta Electrotechnica et Informatica, vol. 18, no. 3, pp. 27 - 34, 2018, [74]:
  - Design and evaluation of the expert interviews which are one part of the qualitative study.
  - Literature review for the overlapping use cases.
  - Preliminary version of the lane detection presented in this thesis.

- J. Jakob, K. Kugele, and J. Tick: "Defining camera-based traffic scenarios and use cases for the visually impaired by means of expert interviews," in IEEE 14th International Scientific Conference on Informatics, pp. 128 - 133, Poprad/Slovakia, November 2017, [73]:
  - Design and evaluation of the experts interviews (which are one part of the qualitative study).
  - Literature review for the overlapping use cases.
- J. Jakob and J. Tick: "Concept for transfer of driver assistance algorithms for blind and visually impaired people," in IEEE 15th International Symposium on Applied Machine Intelligence and Informatics, pp. 241 - 246, Herl'any/Slovakia, January 2017, [39]:
  - Literature review for camera-based ASVI and camera-based ADAS.
  - Summary of preliminary version of the crosswalk detection presented in this thesis.
  - Sketch of the future work towards a transfer concept from ADAS to ASVI.
- J. Jakob and E. Cochlovius: "OpenCV-basierte Zebrastreifenerkennung für Blinde und Sehbehinderte," in Software-Technologien und -Prozesse: Open-Source Software in der Industrie, KMUs und im Hochschulumfeld - 5. Konferenz STeP, pp. 21 - 34, Furtwangen/Germany, May 2016, [128]:
  - Detailed description of a preliminary version of the crosswalk detection presented in this thesis.
- J. Jakob, E. Cochlovius, and C. Reich: "Kamerabasierte Assistenz für Blinde und Sehbehinderte - State of the Art," in informatikJournal, vol. 2016/17, pp. 3 - 10, 2016, [7]:
  - Comprehensive literature review of camera-based ASVI.

Chapter 8

Perspectives and Conclusion

Before concluding the thesis, I describe directions for possible future work based on the presented research.

8.1 Perspectives

The results of the qualitative interview study presented in chapter 4 can be used as the first phase of an exploratory sequential mixed method according to Creswell and Creswell [75]. Based on the traffic scenarios and according use cases defined by the qualitative study, quantitative studies can be conducted, for example in order to examine correlations between type and degree of visual impairment and needed support in traffic scenarios.

The *CoPeD* data set presented in chapter 5 contains video sequences under good weather and lightning conditions. It can be expanded by the addition of scenes under different conditions such as snow, rain, and twilight. Further improvement can be reached by labelling traffic sign images with the coordinates of the sign(s) within the image and other annotations. Labels and annotations provide training data for ML techniques like SVM and NN and make it possible to check the according results automatically. Currently, there are no publicly available annotated data sets from pedestrian perspective. Therefore, adding annotations to *CoPeD* will be an important contribution to the use of ML methods in ASVI.

Chapter 6 concentrates on the adaptation of the "on-road" use cases crosswalk and lane detection as well as RBS. In order to formulate a generalized transfer concept from ADAS to ASVI, the remaining overlapping use cases - namely obstacle, bicycle, vehicle, traffic sign, and traffic light (state) detection - have to be examined and objecitve (O3.2) has to be achieved for these use cases as well. Afterwards, the adaptation procedures of all overlapping use cases have to be inspected and clustered into a concept. For notation of the concept, methods from software engineering [68] and project management [129] can be used and adjusted.

To improve the evaluation of objective (O3.2), the ADAS algorithms on which the ASVI adaptations are based on can be implemented and the performances can be compared by using the *CoPeD* data set. As the ADAS algorithms are generally not described in detail in the according literature, their implementation is a challenging task.

The work presented in this thesis concentrates on a subset of the use cases that are of importance in ADAS as well as ASVI. Besides examining the remaining overlapping use cases in order to formulate the transfer concept, it is important to consider all use cases identified through the evaluation of the qualitative interview study described in chapter 4 when developing a camera-based ASVI that offers support in all relevant traffic situations. The study showed that the perception of all kinds of signs, in traffic as well as other situations, is a substantial need for many visually impaired people. Furthermore, possibilities of personalization should be taken into account so that the system can reflect on the differing needs of the visually impaired.

To improve the hit rates of detection algorithms, external information can be taken into account. The sketch of a camera-based ASVI presented in Figure 1.2 therefore provides the module *External Information Analysis* as part of the cloud service. The idea is to extract information from the internet, e. g. GPS locations of crosswalks, so that there is a priori information about the image content that makes it possible to specify the algorithm accordingly. A similar approach is used for the interpretation of satellite images using data from Geographic Information Systems (GIS). In [130] Liu et al. show that the detection of roads in satellite images can be significantly improved by the addition of GIS data, while Yu et al. [131] use different maps in addition to GIS data to differentiate urban from non-urban areas in satellite images.

In the course of this research, the developed algorithms were implemented in Matlab [69] and run on a PC. In order to make them applicable for the visually impaired, it is essential to implement them on a mobile assistive system, e. g. the one proposed in section 1.2 where the image processing and computer vision algorithms are mainly executed in the cloud.

8.2 Conclusion

In this thesis, I presented the completed research towards a transfer concept for camera-based object detection from ADAS to ASVI.

Based on a literature review of camera-based ASVI, I described the composition of such systems and summarized the according application areas. The composition of ADAS is already described in the literature, see [40, 55]. Furthermore, the literature review yielded the application areas of ADAS. In both cases, the computer vision steps are preprocessing, feature extraction, and decision, although different terminologies are used. Additionally, context information, data transmission, and ways of communicating the results are topics addressed in ADAS as well as in ASVI. The literature review also revealed an overlap between addressed use cases in both fields, e. g. collision avoidance, and there is no comprehensive camera-based assistive system for the visually impaired in traffic situations. Therefore, developing a transfer concept from ADAS to ASVI is a new and meaningful approach.

In order to determine relevant traffic scenarios for the visually impaired and according vision use cases, a qualitative study consisting of interviews with experts and visually impaired people was designed, conducted, and evaluated. The study reached data saturation, meaning that further interviews did not bring any new insights concerning traffic scenarios and use cases. By coding the interview data, six traffic scenarios in three categories were extracted: *Orientation Scenarios* (*General Orientation*, *Navigating to an Address*), *Pedestrian Scenarios* (*Crossing a Road*, *Obstacle avoidance*), and *Public Transport Scenarios* (*Boarding a Bus*, *At the Train Station*). Evaluating the study revealed all vision use cases that could support the visually impaired in traffic situations. Afterwards, this collection was compared with the vision use cases addressed in ADAS literature. The overlap is built by the seven use cases (1) lane detection, (2) crosswalk detection, (3) traffic sign detection, (4) traffic light (state) detection, (5) (driving) vehicle detection, (6) obstacle detection, and (7) bicycle detection. The qualitative data was clustered and presented in adapted scenario tables inspired by software engineering [68].

Furthermore, I created the video data set *CoPeD* containing comparable video sequences from driver and pedestrian perspective in order to be able to evaluate the algorithms developed in the following. The creation of an own data set was necessary because no comparable data from both perspectives existed. *CoPeD* is hosted publicly[1] and licensed under the Creative Commons Attribution 4.0 International License[2].

In the course of the thesis, adaptations for two of the identified seven overlapping use cases, namely crosswalk and lane detection, were discussed. Additionally, RBS was introduced as a further use case and two adaptations were presented. RBS solves the ROI problem and makes it possible to run certain detection algorithms on the road part of the image respectively the background part. For all three considered use cases, I developed adaptations of ADAS algorithms to ASVI and implemented the algorithms in Matlab [69]. I evaluated the newly developed algorithms and compared the results with the ones of the underlying algorithms. Implementation of the underlying ADAS algorithms and evaluation on *CoPeD* will provide more insights in the future. Furthermore, the remaining overlapping use cases have to be examined and it has to be shown that objective (O3.2) can be achieved for them as well. After examining all use cases, the adaptation steps can be summarized in a transfer concept from ADAS to ASVI.

The presented research leads the way towards a generalized transfer concept of camera-based algorithms from ADAS to ASVI that will make latest and future advancements in ADAS applicable for visually impaired pedestrians. Thus, the content of this thesis makes an important contribution to the autonomous mobility of visually impaired people.

[1] `http://dataset.informatik.hs-furtwangen.de/`, accessed on June 6, 2020
[2] `https://creativecommons.org/licenses/by/4.0/`, accessed on June 6, 2020

Bibliography

[1] R. Bourne, S. Flaxman, T. Braithwaite, M. Cicinelli, A. Das, J. Jonas, J. Keeffe, J. Kempen, J. Leasher, H. Limburg, *et al.*, "Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: a systematic review and meta-analysis," *The Lancet Global Health*, vol. 5, no. 9, pp. 888–897, 2017.

[2] H.-W. Wahl and V. Heyl, "Die psychosoziale dimension von sehverlust im alter," *Forum für Psychotherapie, Psychiatrie, Psychosomatik und Beratung*, vol. 1, no. 45, pp. 21–44, 2015.

[3] P. Duckett and R. Pratt, "The researched opinions on research: visually impaired people and visual impairment research," *Disability & Society*, vol. 16, no. 6, pp. 815–835, 2001.

[4] H. Kuijs, C. Rosencrantz, and C. Reich, "A context-aware, intelligent and flexible ambient assisted living platform architecture," in *The Sixth International Conference on Cloud Computing, GRIDs and Virtualization*, (Nice, France), pp. 70–76, March 2015.

[5] R. Cheng, K. Wang, K. Yang, N. Long, W. Hu, H. Chen, J. Bai, and D. Liu, "Crosswalk navigation for people with visual impairments on a wearable device," *Journal of Electronic Imaging*, vol. 26, no. 5, p. 053025, 2017.

[6] S. Caraiman, A. Morara, M. Owczarek, A. Burlacu, D. Rzeszotarski, N. Botezatu, P. Herghelegiu, F. Moldoveanu, P. Strumillo, and A. Moldoveanu, "Computer vision for the visually impaired: the sound of vision system," in *IEEE International Conference on Computer Vision*, (Venice, Italy), pp. 1480–1489, IEEE, October 2017.

[7] J. Jakob, E. Cochlovius, and C. Reich, "Kamerabasierte Assistenz für Blinde und Sehbehinderte," in *informatikJournal*, vol. 2016/17, pp. 3–10, 2016.

[8] R. Shilkrot, J. Huber, C. Liu, P. Maes, and S. Nanayakkara, "FingerReader: a wearable device to support text reading on the go," in *CHI'14 Extended Abstracts on Human Factors in Computing Systems*, (Toronto, ON, Canada), pp. 2359–2364, ACM, April 2014.

[9] A. Anam, S. Alam, and M. Yeasin, "Expression: a google glass based assistive solution for social signal processing," in *16th international ACM SIGACCESS conference on Computers & accessibility*, (Rochester, NY, USA), pp. 295–296, ACM, October 2014.

[10] M. Tanveer, A. Anam, M. Yeasin, and M. Khan, "Do you see what i see?: designing a sensory substitution device to access non-verbal modes of communication," in *15th International ACM SIGACCESS Conference on Computers and Accessibility*, (Bellevue, WA, USA), ACM, October 2013.

[11] A. Rahman, M. Tanveer, A. Anam, and M. Yeasin, "IMAPS: A smart phone based real-time framework for prediction of affect in natural dyadic conversation," in *Visual Communications and Image Processing*, (San Diego, CA, USA), pp. 1–6, IEEE, November 2012.

[12] J. Russell, "Evidence of convergent validity on the dimensions of affect," *Journal of personality and social psychology*, vol. 36, no. 10, p. 1152, 1978.

[13] P. Ekman and W. Friesen, *Manual for the facial action coding system*. Consulting Psychologists Press, 1978.

[14] M. Jeon, N. Nazneen, O. Akanser, A. Ayala-Acevedo, and B. Walker, "Listen2dRoom: Helping blind individuals understand room layouts," in *CHI'12 Extended Abstracts on Human Factors in Computing Systems*, (Austin, TX, USA), pp. 1577–1582, ACM, May 2012.

[15] S. Wang and Y. Tian, "Detecting stairs and pedestrian crosswalks for the blind by rgbd camera," in *IEEE International Conference on Bioinformatics and Biomedicine Workshops*, (Philadelphia, PA, USA), pp. 732–739, IEEE, October 2012.

[16] S. Se, "Zebra-crossing detection for the partially sighted," in *IEEE Conference on Computer Vision and Pattern Recognition*, (Hilton Head Island, SC, USA), pp. 211–217, IEEE, June 2000.

[17] J. Ying, J. Tian, and L. Lei, "Traffic light detection based on similar shapes searching for visually impaired person," in *Sixth International Conference on Intelligent Control and Information Processing*, (Wuhan, China), pp. 376–380, IEEE, November 2015.

[18] P. Meijer, "An experimental system for auditory image representations," *IEEE Transactions on Biomedical Engineering*, vol. 39, no. 2, pp. 112–121, 1992.

[19] P. B. y Rita and K. Kaczmarek, "Tongue placed tactile output device," August 2002. US Patent 6,430,450.

[20] T. Nguyen, T. Nguyen, T. Le, T. Tran, N. Vuillerme, and T. Vuong, "A wearable assistive device for the blind using tongue-placed electrotactile display: Design and verification," in *International Conference on Control, Automation and Information Sciences*, (Nha Trang, Vietnam), pp. 42–47, IEEE, November 2013.

[21] A. Anbuvenkatesh, S. Pownkumar, R. Abinaya, M. Rajaram, and S. Sasipriya, "An efficient retinal art for blind people," in *IEEE International Conference on Computational Intelligence and Computing Research*, (Tamilnadu, India), pp. 1–4, IEEE, December 2014.

[22] K. Moller, F. Toth, L. Wang, J. Moller, K. Arras, M. Bach, S. Schumann, and J. Guttmann, "Enhanced perception for visually impaired people," in *3rd International Conference on Bioinformatics and Biomedical Engineering*, (Beijing, China), pp. 1–4, IEEE, June 2009.

[23] R. Velazquez, E. Fontaine, and E. Pissaloux, "Coding the environment in tactile maps for real-time guidance of the visually impaired," in *International Symposium on Micro-Nano Mechatronics and Human Science*, (Nagoya, Japan), pp. 1–6, IEEE, November 2006.

[24] S. Horvath, J. Galeotti, B. Wu, R. Klatzky, M. Siegel, and G. Stetten, "FingerSight: Fingertip haptic sensing of the visual environment," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 2, pp. 1–9, 2014.

[25] D. Aguerrevere, M. Choudhury, and A. Barreto, "Portable 3d sound/sonar navigation system for blind individuals," in *Second Latin American and Caribbean Conference for Engineering and Technology*, (Miami, Florida, USA), June 2004.

[26] S. Jonas, E. Sirazitdinova, J. Lensen, D. Kochanov, H. Mayzek, R. de Heus, R. Houben, H. Slijp, and T. Deserno, "IMAGO: Image-guided navigation for visually impaired people," *Journal of Ambient Intelligence and Smart Environments*, vol. 7, no. 5, pp. 679–692, 2015.

[27] V. Murali and J. Coughlan, "Smartphone-based crosswalk detection and localization for visually impaired pedestrians," in *IEEE International Conference on Multimedia and Expo Workshops*, (San Jose, CA, USA), pp. 1–7, IEEE, July 2013.

[28] S. Bouhamed, J. Eleuch, I. Kallel, and D. Masmoudi, "New electronic cane for visually impaired people for obstacle detection and recognition," in *IEEE International Conference on Vehicular Electronics and Safety*, (Istanbul, Turkey), pp. 416–420, IEEE, July 2012.

[29] Y.Wei and M. Lee, "A guide-dog robot system research for the visually impaired," in *IEEE International Conference on Industrial Technology*, (Busan, South Korea), pp. 800–805, IEEE, February 2014.

[30] R. Tapu, B. Mocanu, and T. Zaharia, "Real time static/dynamic obstacle detection for visually impaired persons," in *IEEE International Conference on Consumer Electronics*, (Las Vegas, NV, USA), pp. 394–395, IEEE, January 2014.

[31] D. Liyanage and M. Perera, "Optical flow based obstacle avoidance for the visually impaired," in *IEEE Business Engineering and Industrial Applications Colloquium*, (Kuala Lumpur, Malaysia), pp. 284–289, IEEE, April 2012.

[32] S. Pundlik, M. Tomasi, and G. Luo, "Collision detection for visually impaired from a body-mounted camera," in *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, (Portland, OR, USA), pp. 41–47, IEEE, June 2013.

[33] Y. Lee and G. Medioni, "Rgb-d camera based navigation for the visually impaired," in *RSS 2011 RGB-D: Advanced Reasoning with Depth Camera Workshop*, (Los Angeles, CA, USA), pp. 1–6, June 2011.

[34] J. Martinez and F. Ruiz, "Stereo-based aerial obstacle detection for the visually impaired," in *Workshop on Computer Vision Applications for the Visually Impaired*, (Marseille, France), pp. 1–14, October 2008.

[35] H. Fernandes, P. Costa, V. Filipe, L. Hadjileontiadis, and J. Barroso, "Stereo vision in blind navigation assistance," in *World Automation Congress*, (Kobe, Japan), pp. 1–6, IEEE, September 2010.

[36] E. Tekin, D. Vásquez, and J. Coughlan, “Sk smartphone barcode reader for the blind,” in *Journal on technology and persons with disabilities: Annual International Technology and Persons with Disabilities Conference*, vol. 28, pp. 230–239, NIH Public Access, 2013.

[37] X. Yang, S. Yuan, and Y. Tian, “Assistive clothing pattern recognition for visually impaired people,” *IEEE Transactions on Human-Machine Systems*, vol. 44, no. 2, pp. 234–243, 2014.

[38] L. Tian, Y. Tian, and C. Yi, “Detecting good quality frames in videos captured by a wearable camera for blind navigation,” in *IEEE International Conference on Bioinformatics and Biomedicine*, (Shanghai, China), pp. 334–337, IEEE, December 2013.

[39] J. Jakob and J. Tick, “Concept for transfer of driver assistance algorithms for blind and visually impaired people,” in *IEEE 15th International Symposium on Applied Machine Intelligence and Informatics*, (Herl’any, Slovakia), pp. 241–246, January 2017.

[40] B. Ranft and C. Stiller, “The role of machine vision for intelligent vehicles,” *IEEE Transactions on Intelligent Vehicles*, vol. 1, no. 1, pp. 8–19, 2016.

[41] C. Tomasi, “Early vision,” *Encyclopedia of Cognitive Science*, 2006.

[42] J. Canny, “A computational approach to edge detection,” in *Readings in Computer Vision*, pp. 184–203, Elsevier, 1987.

[43] S. Ramalingam and M. Brand, “Lifting 3d manhattan lines from a single image,” in *IEEE International Conference on Computer Vision*, (Sydney, NSW, Australia), pp. 497–504, IEEE, December 2013.

[44] M. García-Garrido, M. Sotelo, and E. Martín-Gorostiza, “Fast road sign detection using hough transform for assisted driving of road vehicles,” in *International Conference on Computer Aided Systems Theory*, (Las Palmas de Gran Canaria, Spain), pp. 543–548, Springer, February 2005.

[45] C. Tomasi and R. Manduchi, “Bilateral filtering for gray and color images,” in *Sixth International Conference on Computer Vision*, (Bombai, India), pp. 839–846, IEEE, January 1998.

[46] C. Rabe, T. Müller, A. Wedel, and U. Franke, “Dense, robust, and accurate motion field estimation from stereo image sequences in real-time,” in *11th European conference on computer vision*, (Heraklion, Greece), pp. 582–595, Springer, September 2010.

[47] O. Pink and C. Stiller, “Automated map generation from aerial images for precise vehicle localization,” in *13th International IEEE Conference on Intelligent Transportation Systems*, (Funchal, Portugal), pp. 1517–1522, IEEE, September 2010.

[48] J. Ziegler, P. Bender, M. Schreiber, H. Lategahn, T. Strausss, C. Stiller, T. Dang, U. Franke, N. Appenrodt, C. Keller, *et al.*, “Making bertha drive — an autonomous journey on a historic route,” *IEEE Intelligent Transportation Systems Magazine*, vol. 6, no. 2, pp. 8–20, 2014.

[49] M. Bertozzi, A. Broggi, and S. Castelluccio, "A real-time oriented system for vehicle detection," *Journal of Systems Architecture*, vol. 43, no. 1-5, pp. 317–325, 1997.

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

[51] W. Kruger, W. Enkelmann, and S. Rossle, "Real-time estimation and tracking of optical flow vectors for obstacle detection," in *Intelligent Vehicles Symposium*, (Detroit, MI, USA), pp. 304–309, IEEE, September 1995.

[52] T. Scharwächter, M. Enzweiler, U. Franke, and S. Roth, "Efficient multi-cue scene segmentation," in *German Conference on Pattern Recognition*, (Saarbrücken, Germany), pp. 435–445, Springer, September 2013.

[53] B. Triggs and J. Verbeek, "Scene segmentation with crfs learned from partially labeled images," in *Advances in neural information processing systems*, (Vancouver, BC, Canada), pp. 1553–1560, December 2008.

[54] J. Alvarez, T. Gevers, Y. LeCun, and A. Lopez, "Road scene segmentation from a single image," in *European Conference on Computer Vision*, (Florence, Italy), pp. 376–389, Springer, October 2012.

[55] R. Loce, R. Bala, and M. Trivedi, *Computer Vision and Imaging in Intelligent Transportation Systems*. John Wiley & Sons, 2017.

[56] B. Hoferlin and K. Zimmermann, "Towards reliable traffic sign recognition," in *IEEE Intelligent Vehicles Symposium*, (Xi'an, China), pp. 324–329, IEEE, June 2009.

[57] X. Miao, S. Li, and H. Shen, "On-board lane detection system for intelligent vehicle based on monocular vision," *International Journal on Smart Sensing & Intelligent Systems*, vol. 5, no. 4, pp. 957–972, 2012.

[58] M. Lu, K. Wevers, and R. V. D. Heijden, "Technical feasibility of advanced driver assistance systems (adas) for road traffic safety," *Transportation Planning and Technology*, vol. 28, no. 3, pp. 167–187, 2005.

[59] J. Li, X. Mei, D. Prokhorov, and D. Tao, "Deep neural network for structural prediction and lane detection in traffic scene," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 3, pp. 690–703, 2017.

[60] Y.-T. Chiu, D.-Y. Chen, and J.-W. Hsieh, "Real-time traffic light detection on resource-limited mobile platform," in *IEEE International Conference on Consumer Electronics - Taiwan*, (Taipei, Taiwan), pp. 211–212, IEEE, May 2014.

[61] J. Choi, B. Ahn, and I. Kweon, "Crosswalk and traffic light detection via integral framework," in *19th Korea-Japan Joint Workshop on Frontiers of Computer Vision*, (Incheon, South Korea), pp. 309–312, IEEE, January 2013.

[62] R. Benenson, M. Omran, J. Hosang, and B. Schiele, "Ten years of pedestrian detection, what have we learned?," in *European Conference on Computer Vision*, (Zurich, Switzerland), pp. 613–627, Springer, September 2014.

[63] J.-J. Yan, H.-H. Kuo, Y.-F. Lin, and T.-L. Liao, "Real-time driver drowsiness detection system based on perclos and grayscale image processing," in *International Symposium on Computer, Consumer and Control*, (Xi'an, China), pp. 243–246, IEEE, July 2016.

[64] A. Witzel, "The problem-centered interview," *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, vol. 1, no. 1, 2000.

[65] M. Meuser and U. Nagel, "Das Experteninterview — konzeptionelle Grundlagen und methodische Anlage," in *Methoden der vergleichenden Politik-und Sozialwissenschaft*, pp. 465–479, Springer, 2009.

[66] VERBI GmbH, "MAXQDA Version 12." https://www.maxqda.de/, Download: 2017-08-02.

[67] P. Mayring, "Qualitative content analysis," *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, vol. 1, no. 2, 2000.

[68] I. Sommerville, *Software Engineering*. Addison-Wesley, 9th edition ed., 2011.

[69] MathWorks, "MATLAB R2017b." https://de.mathworks.com/, Download: 2017-09-21.

[70] J. Jakob, K. Kugele, and J. Tick, "Defining traffic scenarios for the visually impaired," *The Qualitative Report*, 2018. Under Review (Accepted into Manuscript Development Program).

[71] R. Guy and K. Truong, "CrossingGuard: exploring information content in navigation aids for visually impaired pedestrians," in *SIGCHI Conference on Human Factors in Computing Systems*, (Austin, TX, USA), pp. 405–414, ACM, May 2012.

[72] P.-A. Quinones, T. Greene, R. Yang, and M. Newman, "Supporting visually impaired navigation: a needs-finding study," in *CHI'11 Extended Abstracts on Human Factors in Computing Systems*, (Vancouver, BC, Canada), pp. 1645–1650, ACM, May 2011.

[73] J. Jakob, K. Kugele, and J. Tick, "Defining camera-based traffic scenarios and use cases for the visually impaired by means of expert interviews," in *IEEE 14th International Scientific Conference on Informatics*, (Poprad, Slovakia), pp. 128–133, IEEE, November 2017.

[74] J. Jakob and J. Tick, "Traffic scenarios and vision use cases for the visually impaired," *Acta Electrotechnica et Informatica*, vol. 18, no. 3, pp. 27–34, 2018.

[75] J. Creswell and J. Creswell, *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications, 2017.

[76] M. Q. Patton, *Qualitative evaluation and research methods*. SAGE Publications, 1990.

[77] Y. Yuan, Z. Xiong, and Q. Wang, "An incremental framework for video-based traffic sign detection, tracking, and recognition," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, pp. 1918–1929, July 2017.

[78] H. Fleyeh and M. Dougherty, "Road and traffic sign detection and recognition," in *16th Mini-EURO Conference and 10th Meeting of EWGT*, (Poznań, Poland), pp. 644–653, September 2005.

[79] F. Zaklouta and B. Stanciulescu, "Real-time traffic sign recognition in three stages," *Robotics and autonomous systems*, vol. 62, no. 1, pp. 16–24, 2014.

[80] C. Bahlmann, Y. Zhu, V. Ramesh, M. Pellkofer, and T. Koehler, "A system for traffic sign detection, tracking, and recognition using color, shape, and motion information," in *IEEE Intelligent Vehicles Symposium*, (Las Vegas, NV, USA), pp. 255–260, IEEE, June 2005.

[81] A. Ruta, Y. Li, and X. Liu, "Real-time traffic sign recognition from video by class-specific discriminative features," *Pattern Recognition*, vol. 43, no. 1, pp. 416–430, 2010.

[82] D. Cireşan, U. Meier, J. Masci, and J. Schmidhuber, "A committee of neural networks for traffic sign classification," in *International Joint Conference on Neural Networks*, (San Jose, CA, USA), pp. 1918–1921, July 2011.

[83] Z. Fazekas and P. Gáspár, "Computerized recognition of traffic signs setting out lane arrangements," *Acta Polytechnica Hungarica*, vol. 12, no. 5, pp. 35–50, 2015.

[84] S. Sivaraman and M. M. Trivedi, "Vehicle detection by independent parts for urban driver assistance," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, pp. 1597–1608, December 2013.

[85] J. C. Rubio, J. Serrat, A. M. Lopez, and D. Ponsa, "Multiple-target tracking for intelligent headlights control," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, pp. 594–605, June 2012.

[86] D. Greene, J. Liu, J. Reich, Y. Hirokawa, A. Shinagawa, H. Ito, and T. Mikami, "An efficient computational architecture for a collision early-warning system for vehicles, pedestrians, and bicyclists," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, pp. 942–953, Dec 2011.

[87] M. T. Yang and J. Y. Zheng, "On-road collision warning based on multiple foe segmentation using a dashboard camera," *IEEE Transactions on Vehicular Technology*, vol. 64, pp. 4974–4984, November 2015.

[88] W. Song, M. Fu, Y. Yang, M. Wang, X. Wang, and A. Kornhauser, "Real-time lane detection and forward collision warning system based on stereo vision," in *IEEE Intelligent Vehicles Symposium*, (Los Angeles, CA, USA), pp. 493–498, IEEE, June 2017.

[89] S. Santhanam, V. Balisavira, and V. K. Pandey, "Real-time obstacle detection by road plane segmentation," in *IEEE 9th International Colloquium on Signal Processing and its Applications*, (Kuala Lumpur, Malaysia), pp. 151–54, March 2013.

[90] S. Jung, J. Youn, and S. Sull, "Efficient lane detection based on spatiotemporal images," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, pp. 289–295, Jan 2016.

[91] H. Yoo, U. Yang, and K. Sohn, "Gradient-enhancing conversion for illumination-robust lane detection," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, pp. 1083–1094, Sept 2013.

[92] R. Sharma, G. Taubel, and J. S. Yang, "An optical flow and hough transform based approach to a lane departure warning system," in *11th IEEE International Conference on Control Automation*, (Taichung, Taiwan), pp. 688–693, June 2014.

[93] S. Saini, S. Nikhil, K. R. Konda, H. S. Bharadwaj, and N. Ganeshan, "An efficient vision-based traffic light detection and state recognition for autonomous vehicles," in *IEEE Intelligent Vehicles Symposium*, (Redondo Beach, CA, USA), pp. 606–611, June 2017.

[94] M. Omachi and S. Omachi, "Traffic light detection with color and edge information," in *2nd IEEE International Conference on Computer Science and Information Technology*, (Beijing, China), pp. 284–287, August 2009.

[95] Q. Chen, Z. Shi, and Z. Zou, "Robust and real-time traffic light recognition based on hierarchical vision architecture," in *7th International Congress on Image and Signal Processing*, (Dailan, China), pp. 114–119, IEEE, October 2014.

[96] J. Gong, Y. Jiang, G. Xiong, C. Guan, G. Tao, and H. Chen, "The recognition and tracking of traffic lights based on color segmentation and camshift for intelligent vehicles," in *IEEE Intelligent Vehicles Symposium*, (San Diego, CA, USA), pp. 431–435, June 2010.

[97] V. John, K. Yoneda, B. Qi, Z. Liu, and S. Mita, "Traffic light recognition in varying illumination using deep learning and saliency map," in *17th International IEEE Conference on Intelligent Transportation Systems*, (Qingdao, China), pp. 2286–2291, IEEE, October 2014.

[98] Y. Zhai, G. Cui, Q. Gu, and L. Kong, "Crosswalk detection based on mser and eransac," in *IEEE 18th International Conference on Intelligent Transportation Systems*, (Las Palmas, Spain), pp. 2770–2775, IEEE, September 2015.

[99] A. Haselhoff and A. Kummert, "On visual crosswalk detection for driver assistance systems," in *IEEE Intelligent Vehicles Symposium*, (San Diego, CA, USA), pp. 883–888, June 2010.

[100] H. Shen and J. Coughlan, "Towards a real-time system for finding and reading signs for visually impaired users," in *International Conference on Computers for Handicapped Persons*, (Linz, Austria), pp. 41–47, Springer, July 2012.

[101] P. Runeson, M. Host, A. Rainer, and B. Regnell, *Case study research in software engineering: Guidelines and examples*. John Wiley & Sons, 2012.

[102] M. John, F. Maurer, and B. Tessem, "Human and social factors of software engineering: workshop summary," *ACM SIGSOFT Software Engineering Notes*, vol. 30, no. 4, pp. 1–6, 2005.

[103] C. Seaman, "Qualitative methods in empirical studies of software engineering," *IEEE Transactions on Software Engineering*, vol. 25, no. 4, pp. 557–572, 1999.

[104] J. Jakob and J. Tick, "Towards a transfer concept from camera-based driver assistance to the assistance of visually impaired pedestrians," in *IEEE 17th International Symposium on Intelligent Systems and Informatics*, (Subotica, Serbia), pp. 53–60, IEEE, September 2019.

[105] J. Jakob and J. Tick, "CoPeD: Comparable pedestrian driver data set for traffic scenarios," in *IEEE 18th International Symposium on Computational Intelligence and Informatics*, (Budapest, Hungary), pp. 87–92, IEEE, November 2018.

[106] M. Aly, "Real time detection of lane markers in urban streets," in *IEEE Intelligent Vehicles Symposium*, (Eindhoven, Netherlands), pp. 7–12, IEEE, June 2008.

[107] J. Fritsch, T. Kuhnl, and A. Geiger, "A new performance measure and evaluation benchmark for road detection algorithms," in *16th International IEEE Conference on Intelligent Transportation Systems*, (The Hague, Netherlands), pp. 1693–1700, IEEE, October 2013.

[108] M. Le, S. Phung, and A. Bouzerdoum, "Pedestrian lane detection for assistive navigation of blind people," in *21st International Conference on Pattern Recognition*, (Tsukuba Science City, Japan), pp. 2594–2597, IEEE, November 2012.

[109] V. Ivanchenko, J. Coughlan, and H. Shen, "Crosswatch: A camera phone system for orienting visually impaired pedestrians at traffic intersections," in *International Conference on Computers for Handicapped Persons*, (Linz, Austria), pp. 1122–1128, Springer, July 2008.

[110] M. P. Philipsen, M. Jensen, A. Møgelmose, T. Moeslund, and M. Trivedi, "Traffic light detection: A learning algorithm and evaluations on challenging dataset," in *IEEE 18th international conference on intelligent transportation systems*, (Las Palmas, Spain), pp. 2341–2345, IEEE, September 2015.

[111] M. Jensen, M. Philipsen, A. Møgelmose, T. Moeslund, and M. Trivedi, "Vision for looking at traffic lights: Issues, survey, and perspectives," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 7, pp. 1800–1815, 2016.

[112] S. Sivaraman and M. Trivedi, "A general active-learning framework for on-road vehicle recognition and tracking," *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, no. 2, pp. 267–276, 2010.

[113] P. Dollar, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: An evaluation of the state of the art," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 4, pp. 743–761, 2012.

[114] G. Kertész, S. Szénási, and Z. Vámossy, "Multi-directional image projections with fixed resolution for object matching," *Acta Polytechnica Hungarica*, vol. 15, no. 2, pp. 211–229, 2018.

[115] A. Møgelmose, M. Trivedi, and T. Moeslund, "Vision-based traffic sign detection and analysis for intelligent driver assistance systems: Perspectives and survey.," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 4, pp. 1484–1497, 2012.

[116] S. Houben, J. Stallkamp, J. Salmen, M. Schlipsing, and C. Igel, "Detection of traffic signs in real-world images: The German Traffic Sign Detection Benchmark," in *International Joint Conference on Neural Networks*, (Dallas, TX, USA), pp. 1–8, IEEE, August 2013.

[117] J. Jakob and J. Tick, "Camera-based on-road detections for the visually impaired," *Acta Polytechnica Hungarica*, vol. 17, no. 3, pp. 125–146, 2020.

[118] J. Jakob and J. Tick, "Extracting training data for machine learning road segmentation from pedestrian perspective," in *IEEE 24th International Conference on Intelligent Engineering Systems*, (Virtual Event), IEEE, July 2020. Accepted for Publication.

[119] J. W. Lee, "A machine vision system for lane-departure detection," *Computer vision and image understanding*, vol. 86, no. 1, pp. 52–78, 2002.

[120] S. Beucher, M. Bilodeau, and X. Yu, "Road segmentation by watershed algorithms," in *Pro-art vision group PROMETHEUS workshop*, (Sophia-Antipolis, France), April 1990.

[121] J. Alvarez, A. Lopez, and R. Baldrich, "Illuminant-invariant model-based road segmentation," in *IEEE Intelligent Vehicles Symposium*, (Eindhoven, Netherlands), pp. 1175–1180, IEEE, June 2008.

[122] J. Alvarez and A. Lopez, "Road detection based on illuminant invariance," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 1, pp. 184–193, 2011.

[123] M. Foedisch and A. Takeuchi, "Adaptive real-time road detection using neural networks," in *7th International IEEE Conference on Intelligent Transportation Systems*, (Washington, WA, USA), pp. 167–172, IEEE, October 2004.

[124] F. Zohourian, B. Antic, J. Siegemund, M. Meuter, and J. Paul, "Superpixel-based road segmentation for real-time systems using cnn," in *13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, (Funchal, Portugal), pp. 257–265, January 2018.

[125] M. Song and D. Civco, "Road extraction using svm and image segmentation," *Photogrammetric Engineering & Remote Sensing*, vol. 70, no. 12, pp. 1365–1371, 2004.

[126] G. Oliveira, W. Burgard, and T. Brox, "Efficient deep models for monocular road segmentation," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, (Daejeon, South Korea), pp. 4885–4891, IEEE, October 2016.

[127] T. Kavzoglu, "Increasing the accuracy of neural network classification using refined training data," *Environmental Modelling & Software*, vol. 24, no. 7, pp. 850–858, 2009.

[128] J. Jakob and E. Cochlovius, "OpenCV-basierte Zebrastreifenerkennung für Blinde und Sehbehinderte," in *Software-Technologien und -Prozesse: Open-Source Software in der Industrie, KMUs und im Hochschulumfeld - 5. Konferenz STeP*, (Furtwangen, Germany), pp. 21 – 34, May 2016.

[129] P. Rosenberger and J. Tick, "Suitability of PMBOK 6th edition for agile developed it projects," in *IEEE18th International Symposium on Computational Intelligence and Informatics*, (Budapest, Hungary), pp. 241–246, November 2018.

[130] M. Liu, L. Guo, X. Lu, Z. Wang, and C. Li, "Target extraction from sar image under the guide of gis," in *12th International Conference on Signal Processing*, (Hangzhou, China), pp. 905–908, IEEE, October 2014.

[131] S. Yu, M. Berthod, and G. Giraudon, "Toward robust analysis of satellite images using map information-application to urban area detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 4, pp. 1925–1939, 1999.