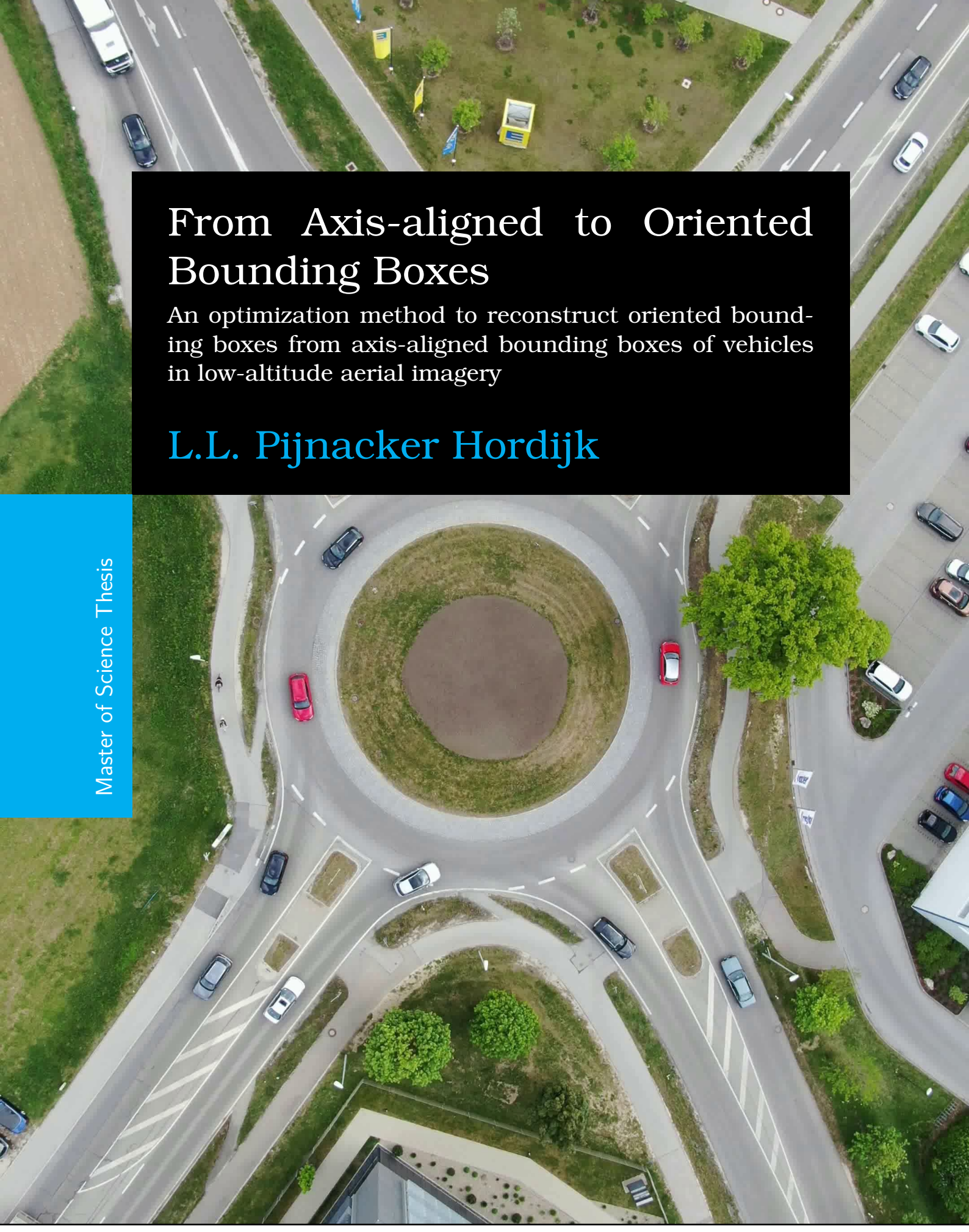


From Axis-aligned to Oriented Bounding Boxes

An optimization method to reconstruct oriented bounding boxes from axis-aligned bounding boxes of vehicles in low-altitude aerial imagery

L.L. Pijnacker Hordijk

Master of Science Thesis

TU Delft Delft University of Technology

Delft Center for Systems and Control

From Axis-aligned to Oriented Bounding Boxes

An optimization method to reconstruct oriented bounding boxes from axis-aligned bounding boxes of vehicles in low-altitude aerial imagery

Master of Science Thesis

For the degree of Master of Science in Mechanical Engineering at Delft University of Technology

L.L. Pijnacker Hordijk

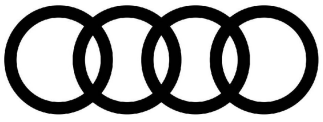
December 4, 2020

Faculty of Mechanical, Maritime and Materials Engineering (3mE) · Delft University of Technology

Confidentiality Note

The thesis as presented is based on internal, confidential data and information of the company AUDI AG.

The thesis may only be made available to the first and second examiner as well as to authorised members of the examination board. Publication and duplication of the thesis - including excerpts thereof - shall not be permitted. The explicit permission of the author and the company shall be required before the thesis may be inspected by unauthorised parties.



Abstract

An increase in the intelligence of autonomous driving functionalities demands detailed analysis of the behaviour of traffic participants. This level of analysis requires datasets that accurately describe the movement of all objects in a specific scene. Recent developments in small Unmanned Aerial Vehicles (sUAVs) and drones introduce an interesting approach to extract this type of data: applying object detection and tracking methods to imagery captured by sUAVs/drones to extract kinematic parameters of traffic participants.

Analysing the behaviour of vehicles in urban traffic has been the topic of research for many years. However, vehicle detection and tracking pipelines used in the proposed approaches lack the required accuracy in estimating a vehicle's orientation and contours. Vehicle detection methods used in recent approaches detect vehicles as Axis-aligned Bounding Boxes (ABBs) that contain no information about a vehicle's orientation and contours. Alternatively, vehicles could be detected as Oriented Bounding Boxes (OBBs) that do contain this information. For low-altitude aerial imagery, however, the amount of publicly available data to train methods using OBB detections is scarce.

To fill the gap in data availability, this thesis introduces the novel 'Axis-aligned to Oriented Bounding Boxes' methodology (A2OBB). Based on a non-linear least squares approximation of the geometrical relationship between an ABB and an OBB, A2OBB finds a set of OBB parameters that fit optimally in all provided ABBs. As such, A2OBB can be applied (1) as an enhancement tool to convert existing ABB annotations to OBBs or (2) as an extension to existing detection networks that perform detections in ABB format.

This thesis contributes by formulating the novel A2OBB methodology and by performing a detailed analysis on its OBB reconstruction performance. In four experiments, an analysis is performed investigating the impact of two assumptions made in the proposed methodology on the reconstruction performance.

The results of the experiments illustrate a sensitivity to: (1) a change in perspective between the observed vehicle and the camera, (2) a mismatch between the perceived shape of vehicles and the assumed rectangular shape and (3) the error introduced by the detection network (in the application of extending ABB detection networks). To reduce the impact of the shape

mismatch, a correction factor is introduced that significantly improves the reconstruction performance.

As such, A2OBB can be used to reconstruct OBBs from a set of manually annotated ABBs with a length and width approximation accurate within 5% and 15% of their respective ground-truths and an optimal orientation with an average error below 2.5°. Overall, manually annotated ABBs are converted to OBBs with an average Intersection over Union (IoU) of 85%. Detections resulting from an in-house out-of-the-box implementation of a YOLOv3 network are reconstructed with a length and width bias of 4.5% and 10% of their respective ground-truth and with an orientation error below 5° but above 2.5°. Overall, the reconstructed OBBs for these detected ABBs have an IoU of 77%.

Note that the orientations resulting from A2OBB's reconstructions comprise one of four possible solutions. A subsequent step is required to find the correct version.

Table of Contents

Acknowledgements

First of all, I would like to express my greatest appreciation to supervisors Moritz and Julian for the encouraging guidance throughout the creation of my thesis. Especially Moritz, your ease in communication and willingness to help at any time inspired me greatly and helped me pull through to the very end. Thank you very much!

I would also like to thank the people at AUDI for providing me with the necessary supplies and a very comfortable place to work. I really enjoyed my time in the AI Lab and I hope to return to you somewhere in the future.

Lastly, I would like to share my gratitude to my friends, roommates and family at home, both in Ingolstadt and in Delft. Without your tremendous encouraging support and inspiring conversations, this thesis would not have been possible.

Chapter 1

Introduction

With an increasing number of autonomous driving functionalities introduced in traffic, low-level analysis and prediction of traffic participants' behaviour have both attracted a lot of attention. Self-driving vehicles must not only be able to plan their own trajectories, but should be able to anticipate the movement of objects in their environment as well. Therefore, accurately predicting low-level traffic behaviour is of vital importance for traffic safety in the future.

Both the analysis and prediction of trajectories of traffic participants require datasets that describe the dynamic state of vehicles, cyclists, pedestrians and other objects on the road. Ideally, these dataset should provide an accurate description of the position, orientation and contours of objects surrounding one specific traffic participant. A conventional method to obtain this type of data is via the use of an vehicle equipped with LiDAR, RADARs and camera systems [1, 2, 3].

However, this method has some downsides and limitations. Especially around roundabouts and intersections, accurately tracking objects proves to be difficult. Both the low manoeuvring flexibility of vehicles and relatively small range that can be captured limits the effectiveness and efficiency of this method [4].

Recent developments in small Unmanned Aerial Vehicles (sUAVs) and drone imagery introduce an interesting alternative: using object tracking methods on imagery captured by sUAVs/drones to extract kinematic data of traffic participants. An illustration of such a low-altitude aerial image is depicted in figure 1-1. Where early research using sUAV/drone imagery focused more on robustly detecting traffic participants, both the resolution and stability of drones are nowadays sufficient to extract an accurate estimate of their position and velocity as well [5]. In addition, sUAVs/drones are easily manoeuvrable and can capture a wide area in a naturalistic way for relatively low costs [4]. Therefore, applying drones to extract kinematic parameters appears to be more beneficial compared to the conventional method.

Several studies are already devoted to extracting kinematic properties of vehicles in traffic from low-altitude aerial imagery [6, 7] and one approach is commercially available under the

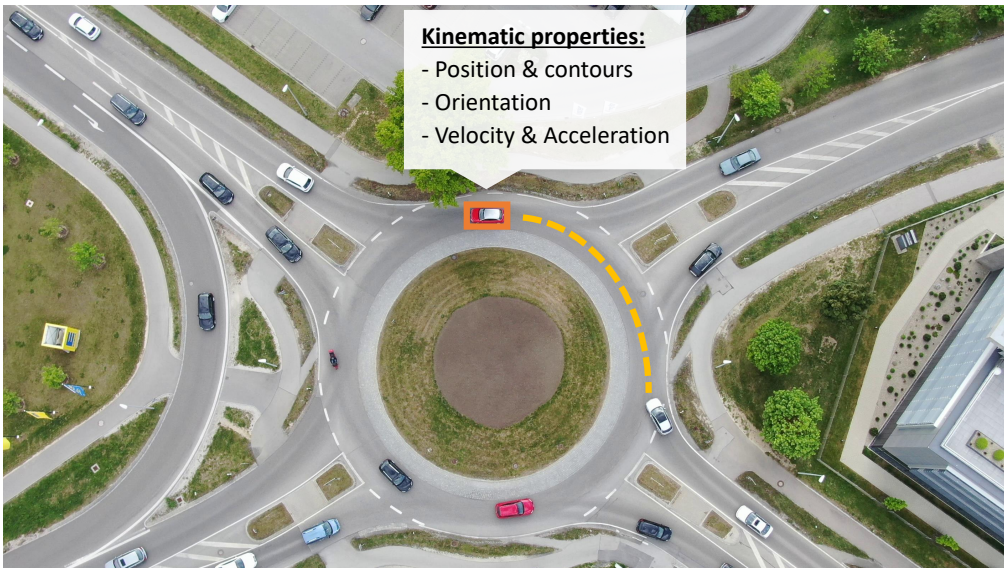


Figure 1-1: An illustration of a roundabout captured in-house by a drone from low-altitude aerial perspective.

name Flow [8]. While position estimation and robustness of detections are already sufficient [5], little attention is dedicated to extracting accurate object contours and orientations. As a consequence, these approaches do not have the accuracy in contour and orientation estimation of traffic participants that is required for low-level trajectory analysis and prediction.

To improve these existing vehicle tracking pipelines, object detection methods can be used that are seen in other applications. In most recent detection methods, objects are detected either in the form of a 2D Axis-aligned Bounding Box (ABB) (e.g. YOLO [9]) or a 2D Oriented Bounding Box (OBB) (e.g. R^3-Net [10]). Both types of bounding boxes are illustrated in figure 1-2. Detecting objects in ABB format allows for extracting the position of vehicles, containing no information about its orientation or contours, whereas detections in OBB format do include an estimation of an object's orientation. In the application of vehicle detection in low-altitude aerial imagery, these methods also allow for an accurate estimate of the vehicle's contours due to their rectangular shape.

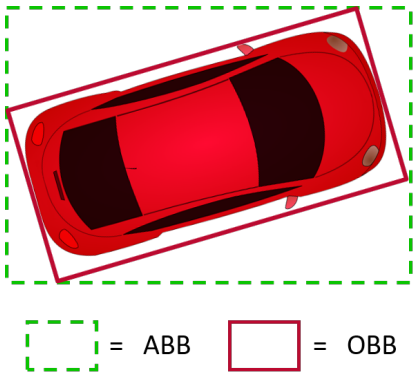


Figure 1-2: An illustration of the difference between an Axis-aligned Bounding Box (ABB), depicted as the green-dotted box, and an Oriented Bounding Box (OBB), depicted in solid red.

A major issue in applying OBB detection methods, however, is the lack of available data to train them. For low-altitude imagery, the limited number of public datasets that exist are, to the authors knowledge, all annotated in ABB format (e.g. the Visdrone dataset [11]). If these existing datasets could be extended to contain data about a vehicle's orientation, essentially converting them to OBBs, it would open a lot of doors for further research in this field, without having to (re)annotate data manually again.

Interestingly, OBB annotations of higher altitude imagery are in some applications also used to evaluate ABB detections by converting the OBB ground-truths to ABBs based on geometry [12]. In this thesis, it is hypothesized that a reverse of this operation can be used to reconstruct OBBs from existing ABBs, given a set of ABBs describing one vehicle in multiple orientations.

Benefits of this method would be twofold: (1) Existing dataset containing ABB annotations can be converted to OBB format. (2) The method can be used as a post-processing extension to existing detection networks, converting results from ABB to OBB format.

For the purpose of investigating this hypothesis, a novel method is introduced called Axis-aligned to Oriented Bounding Boxes (A2OBB). A2OBB exploits the difference between ABB dimensions describing one vehicle in multiple orientations, as depicted in figure 1-3. This figure illustrates that as the orientation of a specific vehicle with respect to recording device changes, so do the dimensions of the ABB describing this vehicle. Based on the difference between ABB dimensions of subsequent observations, a reconstruction can be made of an OBB for every observation.

Note that the orientation of each reconstruction is one of four possible solutions. Therefore, a subsequent step is required to find the correct angle. Examples to perform this step are briefly mentioned, but are not the main focus of this thesis.

Figure 1-3: An illustration of a change in the size and shape of an ABB (orange box) between multiple observations of the same vehicle in different orientations.

The core principle of A2OBB is the assumption that the difference in ABB dimensions is mainly caused by rotation of the vehicle with respect to the image frame. However, in practice, other effects will play a role as well reducing the accuracy of the reconstructions.

To find out how accurately OBBs can be reconstructed using A2OBB in practice, the focus of this thesis is to answer the following research question:

"Can the geometrical relationship between OBBs and ABBs be used to reconstruct OBBs from a set of ABBs describing a vehicle in multiple orientations in one scene captured from a low-altitude aerial perspective?"

An answer to this question is formulated based on a four-step evaluation of the novel A2OBB method, gradually moving from an ideal, theoretic scenario to the method's intended scenes of application. To provide a clear structure, each step is focussed on answering a separate research question with increasing complexity. These questions are as follows:

1. "Can OBBs be reconstructed from ABBs that describe multiple observations of a rectangular vehicle with constant dimensions?"
2. "To what extent can OBBs with varying dimensions be reconstructed from a set of ABBs generated around OBBs using geometry?"
3. "To what extent can OBBs be reconstructed from manually annotated ABBs?"
4. "To what extent can OBBs be reconstructed from ABBs created by a detection network?"

The structure of this report is as follows. Chapter 2 provides information about studies related to this topic, sketching the general pipeline in which A2OBB is intended to operate and stating the contributions of this thesis. Subsequently, chapter 3 states the novel proposed A2OBB methodology, explaining its mathematical preliminaries, zooming in on the assumptions made and describing the method's implementation details. After that, chapter 4 describes the experiments performed in this thesis to answer the research questions and discusses their results. Finally, the thesis is concluded in chapter 5.

Chapter 2

Related Work

The method introduced in this thesis is aimed at enhancing vehicle detection tools applied to low-altitude aerial imagery for the purpose of behavioural predictions of traffic agents. These tools are designed to operate in a pipeline that extracts accurate kinematic parameters of all vehicles in a specific scene. This chapter provides an overview of the relevant literature and the datasets associated with this topic.

Firstly, this chapter describes the steps that are required to extract kinematic parameters from aerial imagery. It starts by briefly discussing the general pipeline and then zooms in on the vehicle detection step. This part is responsible for determining the location and contours of vehicles and is therefore vital in extracting the vehicle's location, orientation and contours accurately. Subsequently, this chapter discusses the datasets that are publicly available for training vehicle detection networks. After that, the chapter states the main contributions of this thesis.

2-1 Vehicle tracking pipelines

Extracting traffic information from imagery taken at a top-down perspective has been done for various purposes over the past decades. Some research, for example, is focussed on examining high level traffic flow on a city scale [13], whereas other types of research focus on examining individual traffic agents for the purpose of analysing lane following, lane changing and lane merging behaviour on the highway [14, 15].

Generally, in vehicle detection and tracking pipelines, three core steps can be identified. Although the terminology and order of these steps might vary per approach, all essentially perform a form of (1) image processing, (2) vehicle detection and (3) vehicle tracking. A schematic of this pipeline is illustrated in figure 2-1.

Figure 2-1: A visualization of a standard vehicle detection and tracking pipeline applied to low-altitude aerial imagery.

The image processing step comprises operations performed on the imagery that compensate distortions that occur while capturing the imagery, relate pixels to metric coordinates through Geo-referencing and transfer all imagery to a fixed frame using stabilization filters. The accuracy of the Geo-referencing step and the ability to compensate imagery for both short-term and long-term movement of the recording device are vital in acquiring accurate results. An analysis of these topics is provided in [4] and [5].

The vehicle detection step, as the name suggests, localizes and classifies vehicles that appear in the imagery. This step is discussed more elaborately in the next section.

Thirdly, the vehicle tracking step traces all detected vehicles throughout the image sequence. The approach for tracking objects depends on the format in which that object is detected and the desired level of accuracy in the results. Pipelines seen in literature use four different approaches: (1) Nearest-Neighbour tracking [16, 17], (2) feature-based tracking [18, 19], (3) Kalman filters [20] or (4) particle filters [6].

2-2 Vehicle detection methods

Detecting vehicles in imagery is a very widely covered subject applied to many different areas. The literature study of [21] divides the approaches seen in literature elegantly in two categories: (1) traditional methods and (2) Deep Learning methods.

Traditional methods comprise techniques that have been developed in the earlier stages in the history of object detection. Due to a lack of effective image representations at that time, these approaches were built around sophisticated hand-crafted features [21]. Several of these hand-crafted features are also used as an approach in existing vehicle tracking pipelines, such as the Viola Jones detector [6] and HOG detectors [18, 22, 23], as well as detectors based on SIFT features [24, 25].

Recent approaches more often opt for the newer Deep Learning based models. Deep Learning methods comprise approaches that utilize some form of Convolutional Neural Network (CNN). This type of network is able to learn robust and high-level feature representations of an image and therefore is very suitable for object detection. [21]

In Deep Learning, two genres can be identified: the two-stage detectors and the one-stage detectors, where the former approaches the detection as a coarse-to-fine process while the later aims perform the detection in a single step [21]. Examples of one-stage detectors that are used in vehicle detection in aerial imagery are You Only Look Once (YOLO) [9], its most recent versions, YOLOv3 [26] and YOLOv4 [27] and the Single Shot MultiBox Detector (SSD)[28]. Examples of Two-stage detectors are the Regional CNN (RCNN) [29], its more advanced version Faster-RCNN, and the Spatial Pyramid Pooling method (SPP) [30].

The previously mentioned Deep Learning methods generally produce object detections in the form of an 2D Axis-aligned Bounding Box (ABB). This type of bounding boxes provides a description of an object's location, but the contour description is inaccurate and the boxes contain no information about the object's orientation. As such, this type of bounding box and these types of networks are not suitable in an application where the orientation and contour estimation of objects is vital.

Although the majority of state-of-the-art object detection networks are not specifically designed for oriented objects [12], some adaptions have been researched to perform detections in the form of an 2D Oriented Bounding Box (OBB). The approach described in [31], for example, applies a modified version of YOLOv3 by incorporating an additional angular term during the logistic regression step, the paper of [12] introduces a modified version of the Faster R-CNN model and the authors of [10] introduce a rotatable region proposal network (R-RPN) combined with a rotatable detection network (RDN) that makes use of batch averaging rotatable (BAR) anchor boxes.

Note, however, that each of these networks is applied to high-altitude aerial imagery (e.g. Satellite imagery) and requires a large amount of training data. Unfortunately, the altitude combined with the movement speed of the recording device introduces a large change in perspective between subsequent imagery and limits the number of observations that can be taken from one specific traffic scene. An illustration of the difference in perspective of two images in the same sequence is depicted in figure 2-2. As a consequence, high-altitude imagery is less suitable for evaluating traffic for behaviour prediction purposes.

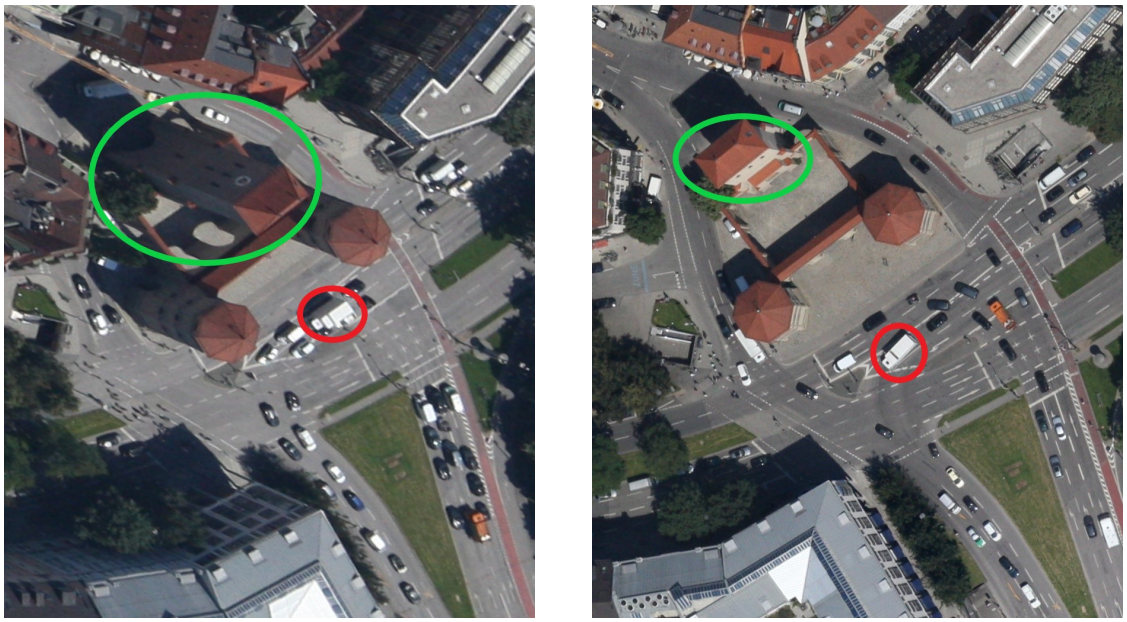


Figure 2-2: Two examples of the KIT AIS dataset [32] illustrating the change in perspective between two images of the same sequence in high-altitude imagery. Not only does the shape of the scenery change (e.g. the tower of the building circled in green), but shape of vehicles changes as well (illustrated by the vehicle circled in red).

In addition, vehicles in high-altitude aerial imagery are perceived as relatively small objects compared to vehicles in low-altitude aerial imagery. Detecting smaller objects requires a different type of network architecture. As such, detection methods that detect vehicles in OBB format in high-altitude aerial imagery can not be applied directly to low-altitude imagery [12].

Networks that are designed to detect oriented objects in low-altitude aerial imagery are scarce, if not non-existent.

2-3 Available datasets

Deep Learning approaches generally require large amounts of training samples. This section provides a brief description of some datasets that are publicly available. The datasets described in this section are divided between high-altitude imagery (satellite/high-altitude UAV imagery) and low-altitude imagery recorded by sUAVs or drones.

High-altitude imagery has been used for traffic analysis for many years. Therefore, the amount of data that is available in this field is extensive and supports the development of detection networks in both ABB and OBB format. Examples of such datasets are the DOTA dataset [12], the VEDAI dataset [33] and the KIT AIS dataset [32].

For low-altitude aerial imagery, the amount of available training data is much more scarce and the publicly available datasets are all annotated in ABB format. Examples of such datasets are the Visdrone dataset [11], the UAV-DT dataset [34] and the AU-AIR dataset [35]. To the authors knowledge, public datasets that contain annotations in OBB format do not exist.

2-4 Axis-aligned to Oriented Bounding Boxes

The fact that low-altitude imagery has not been covered extensively for oriented object detection could be caused by a lack of publicly available training data annotated in OBB format. As Deep Learning methods generally require large amounts of training samples, a lack of available data would demotivate researchers to investigate this field. To fill the gap in data availability, this thesis proposes the novel method called Axis-aligned to Oriented Bounding Boxes (A2OBB). This method is inspired by how ABB detections are evaluated using the DOTA dataset [12].

Interestingly, the DOTA dataset only contains OBB annotations, but does support the evaluation of ABB detection networks. To evaluate such networks, the annotated OBBs are converted to ABBs based on the geometrical relationship between an OBB and its surrounding ABB. The novel A2OBB method aims to revert this operation to create OBBs from ABB. Note, however, that creating OBBs from ABBs is more complex and requires additional information and assumptions to the perceived shape of the object described by these bounding boxes.

The novel proposed A2OBB method obtains this additional information by observing multiple observations of a single vehicle in different orientations. By assuming the perceived size and shape of the vehicle to be constant in all observations and by assuming a perfect fit of the ABB annotations, an estimation can be made of an OBB that fits in all annotated ABBs based on the same geometrical relationship used in the DOTA evaluation.

2-5 Contributions

The contributions of this thesis are threefold. Firstly, this thesis contributes by formulating the A2OBB methodology and by providing proof of the method's working principle under the condition that the assumptions made in the methodology are valid. As such, A2OBB can be applied as enhancement tool to convert existing ABB annotations to OBBs or as an extension to existing detection networks that perform detections in ABB format.

Secondly, a well-structured analysis is performed on the validity of the assumptions that are made in the proposed method in practice. Based on a four-step evaluation, gradually moving from an ideal scenario to the method's intended applications, this analysis provides a good understanding of how each assumption affects A2OBB's reconstruction performance and how their impact can be minimized.

Lastly, this thesis proposes a correction factor that accounts for one of the main potential sources of error and improves the OBB reconstruction performance. In practice, vehicles observed from aerial perspective are perceived as a rounded rectangle instead of the assumed regular rectangle. As a consequence, the reconstructed OBB is smaller than the actual shape of the vehicle. To compensate, a correction factor is formulated based on the theoretical difference between regular rectangles and rectangles with ellipse-shaped corners.

2-6 Conclusion

To summarize, this chapter provides a brief description of vehicle detection and tracking pipelines applied to aerial imagery for the purpose of extracting kinematic parameters of vehicles in traffic. It zooms in on the detection methods, describing the different approaches and types of datasets seen in literature. Most recent approaches opt for the use of Deep Learning methods, that generally detect objects in ABB format. Some research is devoted to detecting objects in OBB format, but these are mainly designed for high-altitude imagery.

An absence of approaches devoted to detecting oriented objects in low-altitude aerial imagery could be caused by a lack of publicly available datasets annotated in OBB format. If existing ABB datasets could be extended to contain data about a vehicle's orientation, essentially converting them to OBBs, this data gap would be filled. It would open a lot of doors for further research in this field without having to (re)annotate data manually.

That is where the novel A2OBB methodology comes in. Based on the geometrical relationship between ABBs and OBBs, A2OBB reconstructs OBBs from a set of ABBs describing a single vehicle in multiple orientations. As such, A2OBB can be applied as enhancement tool to convert manually annotated or automatically detected ABBs describing vehicles in low-altitude aerial imagery to annotations in OBB format. This thesis contributes by formulating the methodology, analysing its performance in practice and by proposing a correction factor for a main potential source of error.

Chapter 3

Method

This chapter introduces the novel Axis-aligned to Oriented Bounding Boxes (A2OBB) methodology. This method is proposed as a post-processing enhancement tool for annotations that used for detecting vehicles in low-altitude aerial imagery. The goal of the method is to convert 2D Axis-aligned Bounding Box (ABB) annotations to 2D Oriented Bounding Boxes (OBBs). As such, this method can be used in two different applications: (1) to convert existing ABB annotations to OBBs and (2) as an extension to existing detection networks that perform detections in ABB format.

To perform this ABB-to-OBB conversion, the method requires a set of ABB annotations of that describe multiple observations of one vehicle. These ABBs must describe this vehicle perceived from aerial perspective and in different orientation. By assuming a constant, rectangular shape for the vehicle in all observations, a set of equations can be constructed that describe the geometrical relationship between that constant shape, described by the length and width of an OBB, and all the provided ABBs. Finding a set of OBB dimensions that fits in each of these equations will allow a reconstruction of the OBB for every individual ABB annotation.

In practice, vehicles will not be perceived with a constant, rectangular shape in all observations. As a consequence, one unique OBB that fits in all geometrical equations will not exist. Therefore, the goal of A2OBB is to find one OBB that fits best in each of these equations. To do so, the method formulates a non-linear least squares problem based on the aforementioned set of geometrical equations. Minimizing the error of this least-squares problem will yield the optimal set of OBB parameters that fit best in the provided ABB annotations.

This chapter starts off by describing the preliminaries of bounding boxes, introducing the necessary parameters and the geometrical relations between ABBs and OBBs that are required to formulate A2OBB's least-squares problem. This section also illustrates how an OBB that fits in each of the provided ABBs can be found under ideal circumstances.

The second section of this chapter zooms in on the assumptions made when constructing the set of equations, describing how they hold in practice. This section will provide a brief

theoretical analysis on how violations to both assumptions affect the optimality of the OBB reconstructions.

After that, the non-linear least squares problem, central to the proposed A2OBB method, is introduced. In this section, a description is provided of A2OBB's implementation, discussing the problem formulation and the optimizer used for the non-linear least-squares approximation. This section also introduces a correction factor to the least-squares problem that compensates for one of the main possible sources of error and mentions the method's limitation.

3-1 Preliminaries of bounding boxes

This section introduces the preliminaries of both ABBs and OBB that are required to construct the least-squares problem central to the A2OBB methodology. To start off, this section introduces the main parameters and notations that are used throughout this chapter. After that, the geometrical relationship between ABBs and OBBs is described and it is demonstrated how a set of OBBs can be reconstructed given a set of ABBs that describe vehicle under ideal conditions.

3-1-1 Bounding box parameters and notations

The A2OBB methodology exploits the differences between the individual dimensions of ABBs that describe one vehicle in multiple orientations. Both types of bounding boxes are described by a set of parameters indicating the location and size of the object in question. The general concept of the method revolves around the geometrical relationship between the two sets of parameters for rectangular objects, illustrated in figure 3-1.

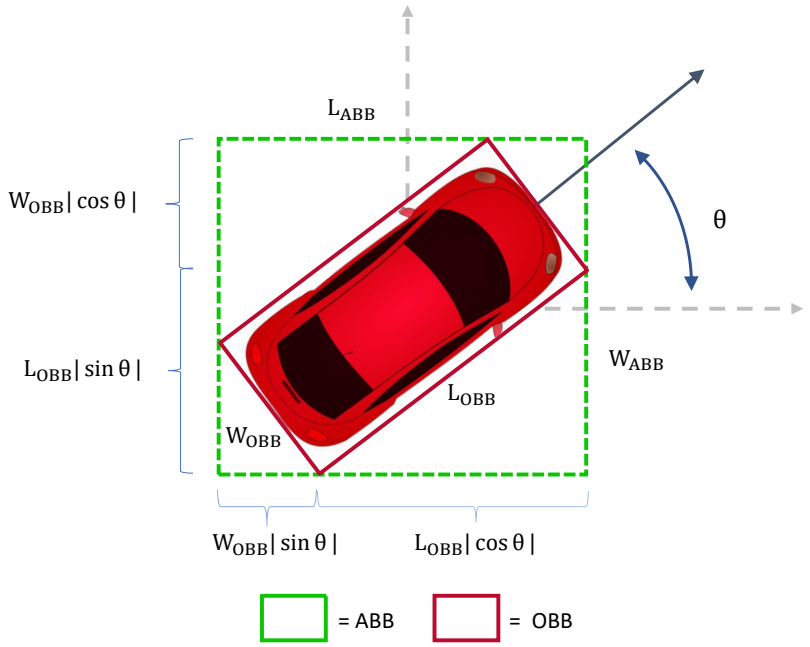


Figure 3-1: An illustration of the geometrical relationship between an ABB and an OBB that describe the same rectangular object. L_{ABB} and L_{OBB} describe the length of the ABB and OBB respectively, whereas W_{ABB} and W_{OBB} describe their width. θ represents the vehicle's orientation w.r.t. the image axes.

The location of the object described by the bounding box is generally indicated by an x-, or y-coordinate in pixel coordinates. For vehicles, the location of both the ABB and OBB are the same. Converting one to the other will not affect these parameters. As such, the location of the bounding box is not mentioned in further formulations.

The size of an object is roughly defined by the length and the width of the bounding box, also quantified in pixel coordinates. In further formulations, the length and width of ABBs are

denoted as L_{ABB} and W_{ABB}. The length and width of an OBB are illustrated in figure 3-1 as L_{OBB} and W_{OBB}. However, to allow for a convenient formulation of optimization problem later in this chapter, the length of the bounding box will be expressed in terms of the OBB's aspect ratio, R, and W_{OBB}:

$$L_{OBB} = R \cdot W_{OBB} \tag{3-1}$$

An OBB also contains information of the orientation of the object it describes. The orientation of the object with respect to the image axes is denoted as θ, quantified in degrees.

3-1-2 Geometrical relationship between ABBs and OBBs

Figure 3-1 suggests that the dimensions of the ABBs, L_{ABB} and W_{ABB}, can be calculated as a function of the dimensions of the OBBs, L_{OBB} and W_{OBB}, and the vehicle's orientation with respect to the image axes, θ. Expressing L_{OBB} using equation 3-1, the geometrical relationship between OBBs and ABBs for a rectangular object can be formulated as:

$$\begin{aligned} L_{ABB} &= W_{OBB}(R|\cos\ \theta| + |\sin\ \theta|) \\ W_{ABB} &= W_{OBB}(|\cos\ \theta| + R|\sin\ \theta|) \end{aligned} \tag{3-2}$$

Under ideal conditions, equations 3-2 can be solved for W_{OBB}, R and θ when the dimensions of two or more ABBs are available that describe the same object in multiple orientations. These ideal conditions state that two assumptions should hold for the set of ABBs and the objects they contain. Firstly, all ABBs describe the same object in multiple orientations and the perceived size of that object does not change between subsequent observations. Thus, (1) the dimensions of the OBB in all ABBs, R and W_{OBB} are assumed to be constant. Secondly, the equations 3-2 describe the ABB-OBB relationship for rectangular objects. A2OBB aims to extract OBBs from vehicle annotations. Therefore, (2) vehicles observed from aerial perspective are assumed to be rectangular.

To illustrate how these equations can be solved for W_{OBB}, R and θ, the scenario of a single ABB is initially considered to establish some constraints and alternate formulations of the ABB-OBB relationship. These constrains and formulations are subsequently used to find the OBB parameters for in a situation with two or more ABBs. Note that in each of the derivations in this section, both the ABBs and OBBs are assumed to adhere to the aforementioned ideal conditions.

3-1-3 A single ABB

When considering a single ABB, equations 3-2 can be constructed to describe the ABB-OBB relationship. The number of unknown values (3: W_{OBB}, R and θ) in this set of equations exceeds the number of available equations (2). Therefore, the amount of available information is not yet sufficient to solve the equations and find a unique solution. Nonetheless, these equations, along with the visualisation of the geometrical relationship in figure 3-1 allow for the definition of three constraints to the unknown values.

The first two constraints are relatively straightforward. Firstly, by definition, the length of a vehicle always exceeds its width. Therefore, R will always be larger than 1. Secondly, W_{OBB} describes the width of the vehicle. This parameter will always be larger than zero, but equal to or smaller than the smallest value of the ABB dimensions.

A third constraint that can be formulated is that equations 3-2 must contain a solution having an orientation on the interval $[0, \pi/2]$. To illustrate why this constraint must hold, notice that equations 3-2 contain the moduli of the $\cos\theta$ and $\sin\theta$. Both terms are used to calculate the length of line segments and should therefore always be positive. As a result, equations 3-2 essentially comprise four different sets of equations, each describing one quadrant of the possible solutions for θ. If a solution to equations 3-2 exists in one of the quadrants, a similar solution also exists in the other three quadrants. The solutions in the different quadrants are mutually related as illustrated in figure 3-2.

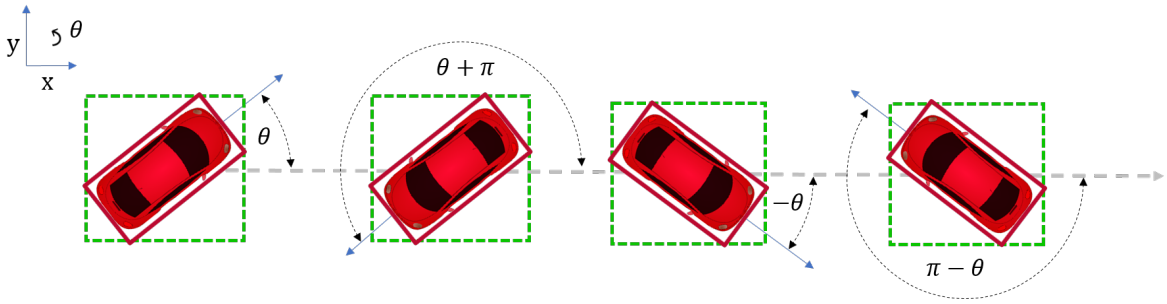


Figure 3-2: An illustration of the four possible OBBs that result in the same ABB. For every OBB with an orientation θ^*, three alternative solutions exist having orientations of $\theta = \theta^* + \pi$, $\theta = -\theta^*$ or $\theta = \pi - \theta^*$.

On the one hand, this means that if a certain set of OBB dimensions fits to equations 3-2, four different orientations will be possible. Finding the true orientation out of these four angles requires an additional step.

On the other hand, if a solution for θ exists, one of these solutions will lie in the first quadrant of θ and this solution will be unique. This allows the formulation of the constraint that θ must lie in the first quadrant on the interval $[0, \pi/2]$. Note that restricting the orientation to lie in the other quadrants will work as well, but the equations that follow will be slightly different.

Using these constraints, equations 3-2 can be reformulated to:

$$\begin{aligned} L_{ABB} &= W_{OBB}(R\cos\ \theta + \sin\ \theta) \\ W_{ABB} &= W_{OBB}(\cos\ \theta + R\sin\ \theta) \end{aligned} \qquad \text{where: } \begin{cases} R > 1, \\ 0 \leq W_{OBB} \leq \min(L_{ABB}, W_{ABB}), \\ 0 \leq \theta \leq \pi/2 \end{cases} \tag{3-3}$$

Furthermore, equations 3-3 can be rewritten to express W_{OBB} and R in terms of the dimensions of the ABB, L_{ABB} and W_{ABB}, and the orientation θ:

$$R = \frac{L_{ABB}\cos\theta - W_{ABB}\sin\theta}{W_{ABB}\cos\theta - L_{ABB}\sin\theta}, \qquad W_{OBB} = \frac{W_{ABB}}{\frac{L_{ABB}\cos\theta - W_{ABB}\sin\theta}{W_{ABB}\cos\theta - L_{ABB}\sin\theta}\cos\ \theta + \sin\ \theta};$$

$$\text{where: } \begin{cases} R > 1, \\ 0 \leq W_{OBB} \leq \min(L_{ABB}, W_{ABB}), \\ 0 \leq \theta \leq \pi/2 \end{cases} \tag{3-4}$$

Given a specific set of ABB dimensions, these equations can be used to calculate all possible pairs of W_{OBB} and R for orientations on the interval $[0, \pi/2]$. Each pair of values represents a point in the R-W_{OBB} domain. As such, the set of all possible pairs of these dimensions will be represented as a line illustrated in figure 3-3. This line will represent all possible OBB dimensions that the given set of ABB dimensions can contain.

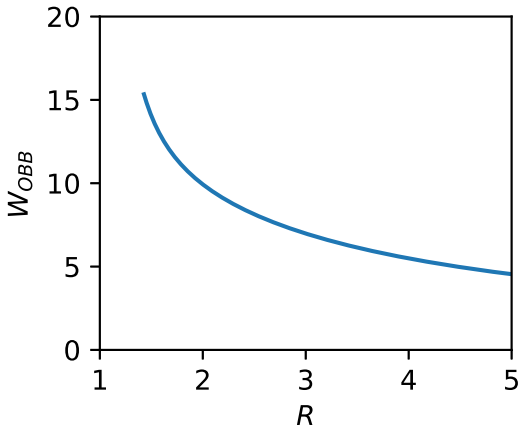


Figure 3-3: The line illustrated in this figure represents all possible pairs of OBB dimensions (W_{OBB} and R) that can be contained by a given pair of ABB dimensions (L_{ABB} and W_{ABB}). These points are calculated as a function of θ using equations 3-4.

3-1-4 Multiple ABBs

For a single ABB, the amount of available information is not sufficient to extract the desired OBB dimensions and orientation. However, in case of two or more ABBs, the amount of information is sufficient and a unique set of OBB dimensions and orientatations can be reconstructed.

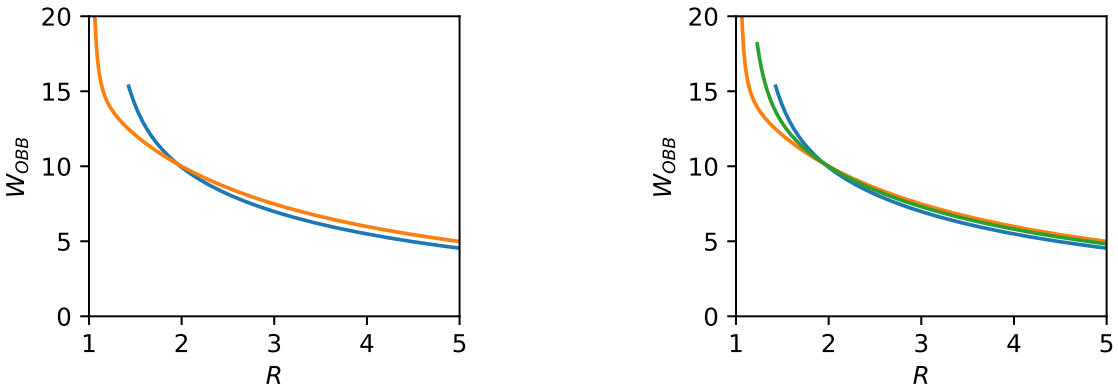
Two ABBs

To illustrate how this works, a situation of two ABBs that belong to one OBB in multiple orientations is considered. For both ABBs, equations 3-3 can be formulated adhering to the constraints mentioned in the previous section. As such, the following set of equations can be formulated:

$$\begin{aligned} L_{ABB_1} &= W_{OBB}(R\cos\ \theta_1 + \sin\ \theta_1) \\ W_{ABB_1} &= W_{OBB}(\cos\ \theta_1 + R\sin\ \theta_1) \\ L_{ABB_2} &= W_{OBB}(R\cos\ \theta_2 + \sin\ \theta_2) \\ W_{ABB_2} &= W_{OBB}(\cos\ \theta_2 + R\sin\ \theta_2) \end{aligned} \quad \text{where:} \quad \begin{cases} R > 1, \\ 0 \leq W_{OBB} \leq \min(L_{ABB_i}, W_{ABB_i}), \\ 0 \leq \theta_i \leq \pi/2 \end{cases} \tag{3-5}$$

with a constant set of unknown OBB dimensions R, W_{OBB} and two different unkown orientations, θ_1 and θ_2.

The system of equations comprises 4 equations containing a total of 4 unknown values. Thus, the amount of information in this set of equations should be sufficient to find these unknown values. In fact, these values can be determined numerically by observing all possible pairs of W_{OBB} and R, given both sets of ABB dimensions. To recall from the previous section, all possible pairs of OBB dimensions that fit in both sets of ABB dimensions can be represented as a line in the R-W_{OBB} domain. Two lines for two different sets of ABB dimensions are depicted in figure 3-4a. Provided that the constraints defined in the previous section are met, these lines will intersect at only a single point. The values of W_{OBB} and R at that point describe the dimensions of the single OBB that fits in both provided ABBs. These dimensions can subsequently be used to determine orientations θ_1 and θ_2.



(a) Two ABBs **(b)** Three ABBs

Figure 3-4: An illustration of all possible pairs of W_{OBB} and R in the R-W_{OBB} domain for multiple sets of ABB dimensions. The left figure depicts the possible pairs for two sets of ABBs, where the right shows this for three sets of ABBs.

Essentially, for a set of two ABBs that describe the same OBB in two orientations, the dimensions of that OBB and its orientation in both orientations ABBs can be determined.

Three or more ABBs

The scenario for two ABBs can be extended to three or more ABBs. To generalise, when considering N ABBs, a set of 2xN-equations can be constructed to describe the geometrical relationship between all ABBs and all possible OBBs they can contain:

$$\begin{aligned} L_{ABB_1} &= W_{OBB}(R\cos\ \theta_1 + \sin\ \theta_1) \\ W_{ABB_1} &= W_{OBB}(\cos\ \theta_1 + R\sin\ \theta_1) \\ &\vdots \\ L_{ABB_N} &= W_{OBB}(R\cos\ \theta_N + \sin\ \theta_N) \\ W_{ABB_N} &= W_{OBB}(\cos\ \theta_N + R\sin\ \theta_N) \end{aligned} \quad \text{where:} \begin{cases} R > 1, \\ 0 \leq W_{OBB} \leq \min(L_{ABB_i}, W_{ABB_i}), \\ 0 \leq \theta_i \leq \pi/2 \end{cases} \tag{3-6}$$

Under ideal conditions, this set of equations contains a constant set of OBB dimensions (R, W_{OBB}) in N-different orientations, (θ_1, ..., θ_N). Similar to the situation with two ABBs, it can be shown that this set of equations has a unique solution for each of these parameters.

Again, each set of equations can be represented as a line in the R-W_{OBB} domain and these lines intersect at the point where the ABBs contain an OBB with the same dimensions. A situation for three different ABBs is depicted in figure 3-4a. Under the ideal conditions state in the previous section, these lines will all intersect at a single point. Again, the values of W_{OBB} and R at that point describe the dimensions of the single OBB that fits in all provided ABBs and these dimensions can be used to determine orientations θ_1, ..., θ_N using equations 3-2.

3-1-5 Conclusion

To summarize, for a given set of ABBs that describe several observations of a rectangular object, a set of equations can be constructed that describe the geometrical relationship between all ABBs and the OBBs they can contain. Provided that the perceived size of this object remains constant over all observations, this set of equations will have a unique solution for the length and width of the OBB in all ABBs that can be determined numerically.

These dimensions can subsequently be used to determine the orientation of the object in every observation. Note, however, that four orientations will be possible for every observation. A subsequent step is required to find the correct version.

Vehicles observed from low-altitude aerial perspective will not be perfectly rectangular and their perceived size will vary between consecutive observations. Therefore, the next section will focus on how the ideal conditions used to formulate this concept will hold in practice.

3-2 Assumptions and possible error sources

The concept behind the proposed methodology considers a set of ABBs that describe one vehicle in multiple orientations. Under ideal conditions, defined in section 3-1-2, a unique solution can be found that allows for the reconstruction of the set of OBBs this set of ABBs can contain. Essentially, a set of ABBs acquired in ideal conditions should adhere to two assumptions. This section analyses the extent to which these assumptions hold for vehicles annotated in low-altitude aerial imagery.

Essentially, this method comprises a non-linear least squares approximation of a set of equations describing the geometrical relationship between ABBs and OBBs. A unique solution to these equations can be found, given validity of two assumptions for the set of ABBs. These assumptions are:

- Assumption 1: "The dimensions of the OBBs that describe the object in the provided set of ABBs do not change over the set of observations."
- Assumption 2: "The shape of vehicles described by the ABBs is rectangular."

The first assumption states that the perceived size of objects should remain constant over the set of observations. For vehicles, however, the perceived dimensions will slightly differ due to variations introduced while acquiring the annotations and the imagery. These variations are mainly caused by:

- lens distortions;
- a change in perspective between the drone and the vehicle;
- annotation errors.

The second assumption states that the objects described by the ABBs should be rectangular. To be more specific, the set of ABBs should be constructable from a set of OBBs using equations 3-2. In practice, however, the ABB dimensions can vary due to:

- a mismatch between the rectangular shape estimation and the actual vehicle shape;
- annotation errors.

As a result, the ABB-OBB relationship will be slightly different to equations 3-2.

For each of the above-mentioned aspects, a short description is provided on how they affect the perceived vehicle size and the geometrical relationship and how their impact on the validity of the assumptions can be minimized.

3-2-1 Lens distortions

A first source that induces variations in the perceived vehicle size occurs while acquiring the imagery. Due to the fact that a camera observes a scene through a specific lens, the observed scene is deformed based on the shape of the lens. Depending on the shape of the lens, this deformation takes the form of barrel distortion or pincushion distortion (both illustrated in figure 3-5). Generally, in aerial imagery the lens induces barrel distortion.

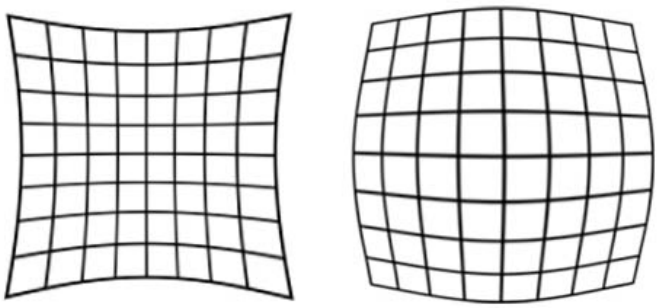


Figure 3-5: An illustration of pincushion distortion (left) and barrel distortion (right).

This type of distortion is widely covered in literature and several methods are suggested to compensate this effect. A commonly used method to do so is Zhang's Calibration method [36]. In current state-of-the art detection methods, compensation for lens distortion is mostly applied prior to extracting annotations or applying a detection network (e.g. in [6] and [16]). Therefore, the impact of lens distortions on the perceived vehicle size is assumed to be fully compensated and will not have an affect on A2OBB's performance.

3-2-2 Change in perspective

A second source of introducing variations in the perceived vehicle size is a change in perspective between the vehicle and the recording device. Both the vehicle and the recording device move as time progresses. As a result, the perspective between both objects changes, introducing a variation in the perceived vehicle size.

The movement of the recording devices and camera are usually very limited. Most state of the art drones are equipped with stabilization mechanisms that compensate for short-term movement and long-term drift of the drone can be corrected via video stabilization filters [16]. One of such methods is implemented in this study as well. More details on this filter are discussed in appendix A. As a result of applying this filter, the camera position is assumed to be fixed.

Nonetheless, the vehicle will still move with respect to the camera. For vehicles moving away from the image centre, parts of the vehicles side increasingly appear in the image that would be occluded when observing the vehicle from top-down. These 'top-down occluded sides' deform and increase the perceived vehicle size. An illustration of this effect in 2D is illustrated in figure 3-6. The figure illustrates that the perceived vehicle length, denoted as L^*, depends on the x-coordinate of the vehicle.

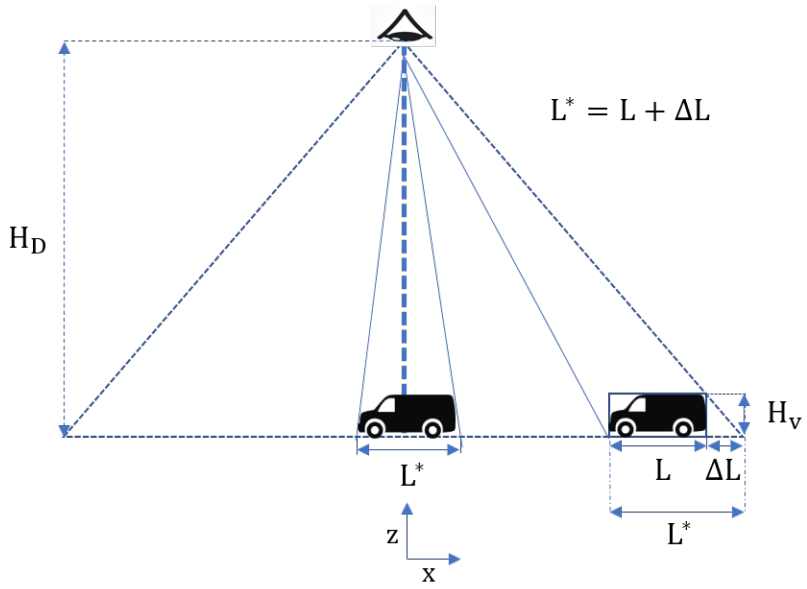


Figure 3-6: An illustration of the effect of a difference in perspective on the perceived length of a vehicle in 2D, denoted as L^*.

The position and height of the drone (H_D in the figure 3-6) remain fixed in this scenario. Therefore, the deformation of the vehicle will depend on two factors: (1) the position (x) of the vehicle with respect to the image centre and (2) the shape of the 'top-down occluded side' that starts to appear. For the simplified version depicted in figure 3-6, the shape of the 'top-down occluded side' mainly depends on the height of the vehicle (denoted as H_v).

For a box-shaped vehicle, where the included backside of the vehicle is perpendicular to the roof of the vehicle, the change in observed length (ΔL) can be calculated as a function of x, H_D and H_v using similar triangles. This relationship is described by the following equation:

$$\Delta L = \frac{H_v}{H_D} x \tag{3-7}$$

In 3D, the magnitude of the deformation is split between the perceived length and width of the vehicle, depending on the orientation of the vehicle with respect to the vehicle's displacement from the image centre. A very basic example of how this distribution depends on this orientation is depicted in figure 3-7a.

In essence, when assuming that the vehicle's orientation with respect to the image centre constant, the perspective transformation can be reduced to a 2D problem, illustrated in figure 3-6 and described by equation 3-7. This equation suggests that for box-shaped vehicles, ΔL linearly increases with the vehicle's distance to the image centre as long as θ_d remains constant. For non-box-shaped vehicles (e.g. a regular passenger vehicle), this relationship becomes slightly more complex depending on the inclination angle of the sides pointed to and from the image centre. As illustrated in figure 3-7b, H_v will be smaller for these vehicles and the magnitude of the perspective transformation will be smaller as well. In addition, the location of the 'top-down occluded side' for non-box-shaped vehicles depends on the inclination angle of the vehicle and the vehicle's distance to the image centre. However, for

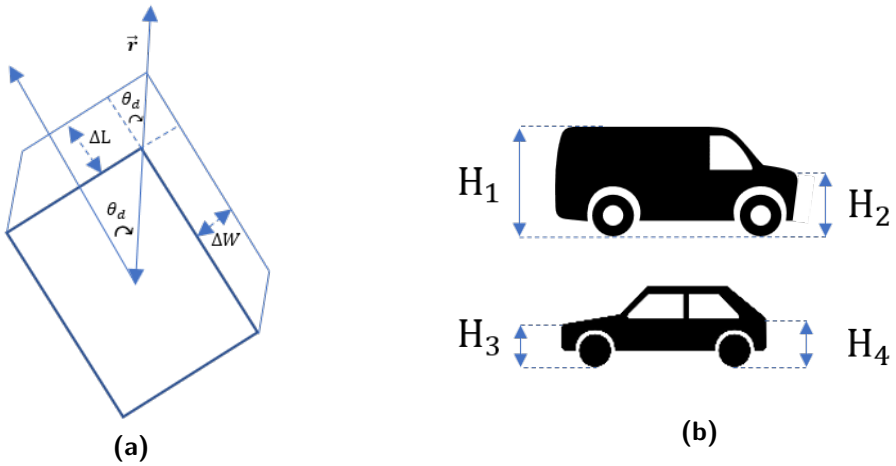


Figure 3-7: Illustrations of the perspective transformation. Figure 3-7a illustrates the distribution of increase in size over the length and width of the vehicle depending on angle θ_d, the angle between the vehicle orientation and the vehicle's displacement vector with the origin ($\vec{r}$.). Figure 3-7b shows a length intersection for a van (top) and a regular passenger vehicle (bottom), illustrating the height of the sides that affect the perspective transformation.

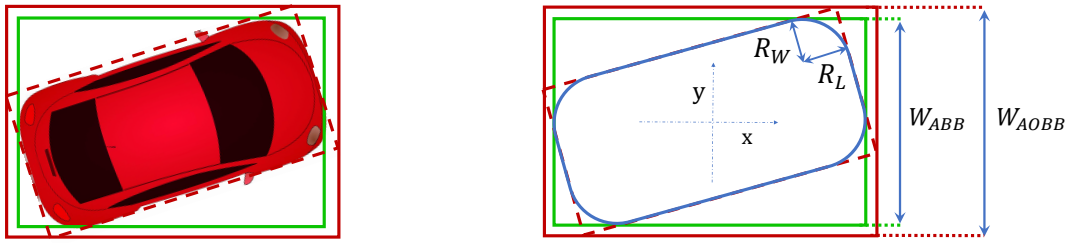
the scope of this thesis, both of these additional complexities are neglected and all vehicles are assumed to be box-shaped.

Neglecting these extra complexities, one could state that for vehicles moving in a direction aligned with their orientation with respect to the image centre remains constant, an increase in their perceived size should be linearly dependent on the vehicle's distance to the image centre. In addition, knowing that the perspective transformation depends on the distance of a vehicle with respect to the image centre, only considering observations that are equidistant from the image centre could reduce the variance in OBB size between different observations.

3-2-3 Mismatch between the assumed and actual vehicle shape

A third source of deviation to the assumptions is caused by a mismatch between the assumed, rectangular shape of vehicles and their actual, perceived shape. Instead of seen as a regular rectangle, vehicles observed from an aerial perspective are rather perceived as rounded-rectangles. Consequently, annotated ABBs will be smaller than the ABB constructed from the theoretical OBBs using equations 3-2 as illustrated in figure 3-8. For the remainder of the thesis, this effect is referred to as the 'rounded-corner effect'. For further reference, an ABB constructed from an OBB using equations 3-2 is referred to as an AOBB.

The magnitude of the difference between both bounding boxes depends on the orientation of the vehicle and the shape of the vehicle's corners. By estimating the shape of the vehicle's corners, a prediction can be made of the difference annotated ABB parameters and AOBBs. For example, if the corners are assumed to be ellipse-shaped, the difference in the length and width between both bounding boxes, respectively denoted as ΔL_A and ΔW_A, can be expressed in terms of the radii that define the elliptic shape of the corners. The radii of the ellipse-shaped corner in the direction of the vehicle length and the width are denoted as R_L



(a) A visual example. **(b)** A schematic view.

Figure 3-8: A visual and schematic illustration of the 'rounded-corner effect' for a vehicle with ellipse-shaped corners in an orientation θ. Both figures depict the discrepancy between annotated ABBs (green box) and the AOBBs (solid red box), generated around the OBB (red-dotted box) using equations 3-2. The right figure illustrates the difference between the widths of ABBs and AOBBs (W_{ABB} and W_{AOBB}) in a rough sketch, also showing the radii that define the elliptic shape of the corners. R_L is the radius along the length of the vehicle, whereas R_W represents the radius along the vehicle's width.

and R_W and the orientation of the vehicle with respect to the image axes as θ. A derivation of this expression is included in appendix B.

To summarize the derivation, the 'rounded- corner effect' for vehicles with ellipse-shaped corners can be expressed as:

$$\begin{bmatrix} \Delta L_A \\ \Delta W_A \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \cos\theta & \sin\theta \end{bmatrix} \begin{bmatrix} \Delta L_{A_{x'}} \\ \Delta L_{A_{y'}} \end{bmatrix} \tag{3-8}$$

where:

$$\Delta L_{A_{x'}} = 2\left|\sqrt{\frac{R_L^2 R_W^2}{R_W^2 + R_L^2\tan^2\theta}}\right|, \qquad \Delta L_{A_{y'}} = 2\left|\sqrt{\frac{R_L^2 R_W^2\tan^2(\theta)}{R_W^2 + R_L^2\tan^2\theta}}\right| \tag{3-9}$$

Note that these formulas describe the effect for a vehicle where all corners are shaped equally and the vehicle is orientated on the interval $[0, \frac{\pi}{2}]$. In this case, an elliptic shape is chosen over a circular shape, as it describes the curvature of a vehicle's corners more accurately.

Equations 3-8 and 3-9 allow for a decent description of the 'rounded-corner effect' for objects with ellipse shaped corners. As such, this information can be used to compensate errors that are caused by this effect. To do so, section 3-3-2 introduces a correction factor to the least-squares problem central to the A2OBB methodology.

3-2-4 Annotation errors

Lastly, vehicle annotations will differ from the ideal conditions due to annotation errors. Annotations in general are an approximation of the location and shape of the actual vehicle.

The fact that annotations are an approximation means that every type of annotation always contains some degree of error. The magnitude of the error does depend on the type of annotation. In this application, two types of annotations can be identified: (1) Annotations performed manually by a human and (2) annotations resulting from an automatic detection network.

Manual annotations provide the closest approximation of the actual location and contours of vehicles in imagery. This type of annotation still contain inaccuracies due to human error, but at this point in time no programmed annotations will yield better results. For that reason and with a lack of a better alternative, manual annotations are considered as the ground-truth. Unfortunately, this leaves little to say about the accuracy of this type of annotation. For now, human annotation errors introduced in manually annotated bounding boxes are considered to be small.

Detected annotations are annotations that result from an automatic detection network, designed to locate a desired object in an image. Essentially, these networks attempt to detect objects based on manually annotated examples. Therefore, detected annotations will never be consistently more accurate than manual annotations. On top of that, detection networks contain detection errors. The magnitude and shape of the error introduced by the network depend on the training procedure, the training data and the network architecture.

3-2-5 Conclusion

To summarize, ABBs that describe a vehicle in low-altitude aerial imagery will deviate from the ideal conditions stated in section 3-1. The perceived size of vehicles will vary between subsequent observations due to lens-distortions, a change in perspective and annotation errors. In addition, the ABB-OBB relationship will be slightly different from equations 3-2 as vehicles are perceived as rounded-rectangles instead of regular rectangles.

3-3 A2OBB: From Axis-aligned to Oriented Bounding Boxes

Section 3-1 poses the idea that under ideal conditions a set of OBB annotations can be reconstructed based on a set of ABB describing an object in multiple orientations. A2OBB applies this idea in practice, considering a set of ABB annotations describing a vehicle observed from aerial perspective.

ABB annotations of vehicles will deviate from the ideal conditions as described in section 3-2. As a consequence, slight perturbations are introduced in the set of geometrical equations describing the ABB-OBB relationship and the resulting equations will not contain a consistent set of OBB dimensions that will fit to all observations.

Nonetheless, a set of OBB dimensions can be found that closely approximates the dimensions of the actual OBB in every observation. A2OBB aims to find these optimal OBB dimensions and corresponding orientations by means of formulating and solving a non-linear least-squares problem. This procedure requires an error function and a set of constraints that define the problem and a non-linear least-squares solver that can find a solution. This section discusses each of these aspects.

One of the main possible sources of error comprises the 'rounded-corner effect', introduced by a mismatch between the perceived shape of vehicles and the assumed rectangular shape, as described in sectin 3-2-3. To compensate for this effect, this section also introduces a correction factor based on the approximation of this effect for objects with ellipse-shaped corners.

Lastly, this section mentions the method's limitations, along with possible follow-up steps to overcome them.

3-3-1 Error functions and constraints

A2OBB's goal is to find a set of OBB parameters that closely resembles the true OBB in every ABB. Essentially, a set of OBB dimensions and a set of orientations must be found for each observation to which the error in equations 3-6 is smallest. These errors can be formulated for every i-th ABB as:

$$\begin{aligned} E_{L_i} &= L_{ABB_i} - W_{OBB}(R\cos\theta_i + \sin\theta_i) \\ E_{W_i} &= W_{ABB_i} - W_{OBB}(\cos\theta_i + R\sin\theta_i) \end{aligned} \tag{3-10}$$

where E_{L_i} and E_{W_i} represent the error in the ABB length and width calculations respectively. In addition, based on what is previously discussed in section 3-1-3, a set of constraints can be defined for the least-squares problem. To recall, the OBB parameters for every ABB comply to the following constraints:

1. $R > 1$;
2. $0 \leq W_{OBB} \leq \min(L_{ABB_i}, W_{ABB_i})$;
3. $0 \leq \theta_i \leq \pi/2$.

Combining error functions 3-10 for all N-bounding boxes, the following set of equations can be defined:

$$\begin{bmatrix} E_{L_1} \\ E_{W_1} \\ E_{L_2} \\ E_{W_2} \\ \vdots \\ E_{L_N} \\ E_{W_N} \end{bmatrix} = \begin{bmatrix} L_{ABB_1} \\ W_{ABB_1} \\ L_{ABB_2} \\ W_{ABB_2} \\ \vdots \\ L_{ABB_N} \\ W_{ABB_N} \end{bmatrix} - W_{OBB} \begin{bmatrix} R & 1 & & & & & \\ 1 & R & & \mathbf{0} & \cdots & & \mathbf{0} \\ & & R & 1 & & & \\ & \mathbf{0} & 1 & R & \cdots & & \mathbf{0} \\ \vdots & & \vdots & & \ddots & \vdots & \\ & & & & & R & 1 \\ & \mathbf{0} & & \mathbf{0} & \cdots & 1 & R \end{bmatrix} \begin{bmatrix} \cos\theta_1 \\ \sin\theta_1 \\ \cos\theta_2 \\ \sin\theta_2 \\ \vdots \\ \cos\theta_N \\ \sin\theta_N \end{bmatrix} \tag{3-11}$$

To find an optimal OBB that fits best in every ABB, these errors must be minimized for all observations. These errors are used to define the non-linear least-squares problem. Note that as the number of bounding boxes in the set of ABBs increases, equations 3-11 become increasingly sparse.

3-3-2 Correction factor to the 'rounded-corner effect'

As mentioned in section 3-2-3, one of the main possible sources of error comprises the 'rounded-corner effect' introduced by a mismatch between the perceived shape of vehicles and the assumed rectangular shape. To allow for a method to reduce the impact of this effect on the reconstruction performance, this section proposes a correction factor. For every i-th bounding box observation, two correction factors, denoted as C_L and C_W, are formulated as:

$$\begin{aligned} C_{L_i} &= 2\left(\cos\theta_i \left|\sqrt{\frac{R_L^2 R_W^2}{R_W^2 + R_L^2\tan^2\theta_i}}\right| - \sin\theta_i \left|\sqrt{\frac{R_L^2 R_W^2 \tan^2\theta_i}{R_W^2 + R_L^2\tan^2\theta_i}}\right|\right) + B_L \\ C_{W_i} &= 2\left(\cos\theta_i \left|\sqrt{\frac{R_L^2 R_W^2}{R_W^2 + R_L^2\tan^2\theta_i}}\right| + \sin\theta_i \left|\sqrt{\frac{R_L^2 R_W^2 \tan^2(\theta_i)}{R_W^2 + R_L^2\tan^2\theta_i}}\right|\right) + B_W \end{aligned} \tag{3-12}$$

The first two terms of these equations represent the 'rounded-corner effect' described by equations 3-8 and 3-9. In addition, to accommodate compensation for a possible bias in the bounding box annotations, two constants, B_L and B_W, are introduced. The values for each of these parameters can be approximated based on the size and shape of vehicles that appear in real-world examples.

Combining the suggested correction factor with error functions 3-10, the error functions that takes the 'rounded-corner effect' into account is formulated as:

$$\begin{aligned} E_{L_i} &= L_{ABB_i} - W_{OBB}(R\cos\theta_i + \sin\theta_i) + C_{L_i} \\ E_{W_i} &= W_{ABB_i} - W_{OBB}(\cos\theta_i + R\sin\theta_i) + C_{W_i} \end{aligned} \tag{3-13}$$

3-3-3 Non-linear least squares solver

The next step is to find a suitable solver that can find a solution to the non-linear least squares problem described by equations 3-11. Note that finding the best solver for this problem is not the main focus of this thesis. As such, the rate of convergence and the computational effort required to find a solution are considered to be of lesser importance. Nonetheless, the solver used in A2OBB must at least be able to handle a sparse, bounded non-linear least squares problem.

Based on these demands, an implementation of the Trust Region Reflective method [37] provided by the SciPy library [38] is selected. The Trust Region Reflective method iteratively solves trust-region sub-problems augmented by a special diagonal quadratic term and with trust-region shapes determined by the distance from the bounds and the direction of the gradient. This method works robustly with a sparse, bound-constrained minimization problem.

The method minimizes the sum of squared residuals calculated by equation 3-11. In total, the loss function minimized by the Trust Region Reflective method is:

$$L = \sum_{i=1}^{N} (E_{L_i}^2 + E_{W_i}^2) \tag{3-14}$$

To reduce the amount of computational effort that is required to find an optimal solution, the least-squares solver is provided with an initial estimate for the width of the OBB defined as W_{OBB_0}. This value is based on lowest value of the ABB parameters:

$$W_{OBB_0} = min(L_{ABB_i}, W_{ABB_i}) \tag{3-15}$$

3-3-4 Limitation

Based on the non-linear least squares problem defined in the previous section, a set of optimal OBBs can be reconstructed that should closely resemble the true OBBs in every observation. Note, however, that the method has one main limitation: the orientation of all bounding boxes is constrained to lie in the first quadrant of θ on the interval $[0, \frac{\pi}{2}]$. Additionally, as briefly mentioned in section 3-1, this solution will always have a similar solutions in the other three quadrants as illustrated in figure 3-4. Finding the correct version from these four possible solution is impossible based on this least-squares solution. Additional post-processing steps are required.

A2OBB is intended to operate in a vehicle tracking pipeline. Therefore, determining the correct solution could be performed during the tracking procedure. Nonetheless, it is worth mentioning two methods that are fairly straightforward to implement due to the fact that the set of ABBs describe a set of consecutive observations in the same scene. Firstly, the direction of the bounding box displacement can be used as an initial estimate. Given a high enough frame rate in the imagery, the displacement of the bounding box should be oriented relatively aligned with the vehicle's orientation. Picking the solution that is closest to the bounding box displacement should result in the correct solution. Secondly, the true solution can be found based on the vehicle's location in the image. Due to the fact that the scenery

remains the same for all images, a mask can be created to define what the orientation of every vehicle roughly should be based on the vehicle's position. The downside to this method is that it is not applicable to locations where vehicles can move in multiple directions.

3-3-5 Conclusion

To summarize, A2OBB applies the idea that OBBs can be extracted from ABBs describing a single rectangular object in multiple orientations in practice. Due to annotation errors and variations in the perceived size of vehicles between consecutive observations, a perfect OBB that fits in all ABBs can not be found. Therefore, the goal of A2OBB is to reconstruct an optimal set of OBB dimensions and corresponding orientations that fits closest to every observation.

A2OBB finds this optimal set of dimensions by solving a non-linear least-squares problem based on the ABB-OBB relationship for rectangular objects. To solve this, an out-of-the-box implementation of the Trust-Region Reflective algorithm is implemented.

A2OBB reconstructs all OBBs with an orientation in the first quadrant of θ on the interval $[0, \frac{\pi}{2}]$. Note that similar solutions also exist in the remaining three quadrants. A2OBB alone does not provide enough information to find the correct version. However, this can be done in post-processing steps.

Chapter 4

Experiments

The newly proposed A2OBB is designed to reconstruct a set of OBBs from a set of ABBs describing a vehicle in multiple orientations. To do so, a least-squares optimization is formulated making two assumptions. The more a given dataset deviates from these assumptions, the less optimal the reconstructed OBBs will be. In section 3-2, it is argued that the intended input data will indeed deviate from these assumptions. The experiments descibed in this chapter investigate the impact of a deviation from these assumptions on A2OBB's reconstruction performance.

A2OBB is evaluated in a four-step evaluation, gradually moving from an ideal, theoretic scenario to the method's intended scenes of application. The first two steps focus on artificially created data to provide a basis for analysing the impact of violations to each assumption on the reconstruction performance. The latter two experiments focus on assessing the method's reconstruction performance in its intended applications. Each step consists of an experiment evaluating the reconstruction performance on a dataset recreating the proper scenario.

The chapter starts by describing the datasets and evaluation metrics used in the experiments. Subsequently, the setup and results of each evaluation step are explained, formulating an answer to the research questions stated in chapter 1.

4-1 Data acquisition and Metrics

This section describes the methods of acquisition and the characteristics of the collected and created datasets used during the experiments. The first experiment makes use of synthetically generated data, whereas the other experiments are based on in-house collected data. The parameters used for generating the synthetic dataset are based on what is seen in the in-house collected data. To understand the magnitude of the parameters used for the synthetic dataset, the in-house collected data acquisition and the properties of vehicles seen in the dataset are discussed first.

The second and third sub question work with manually annotated and synthetically generated bounding boxes, whereas the fourth question involves the evaluation of detected ABBs. To obtain the detected ABBs, an out-of-the-box implementation of Yolov3 [26] is used. The training procedure for this network also briefly described in this section.

Lastly, this section introduces the four metrics used for evaluating the research questions.

4-1-1 In-house collected datasets

For the purpose of evaluating A2OBB's performance on real-world examples, an in-house collected dataset is created. This dataset consists of imagery recorded by a Mavic 2 Zoom drone in 4K resolution at an altitude of 100 meters above a roundabout. In total, two sequences are collected, each counting 400 images at 5 fps. Both sequences cover the same scene under different weather conditions. The imagery is subsequently processed in a fashion similar to the general vehicle tracking pipeline described in section 2-1.

Image preprocessing

The imagery is rectified using Zhang's calibration method [36] to compensate for lens distortions and the image sequences are compensated using a stabilization filter described in appendix A. This filter is a modified version of the pipeline described in [39]. As a result, the imagery is considered to be taken from an fixed position. Geo-referencing is not relevant in this thesis and is therefore omitted. 'Vehicle detections' are in this case created both through manual annotation and as a result of the in-house trained YOLOv3 network.

Vehicle detection: Manual annotations

The imagery is annotated manually in two formats: ABBs and OBBs. An example of an annotated image is depicted in figure 4-1. It illustrates an image containing vehicles manually annotated in ABB (green boxes) and OBB (red boxes) format. The blue circles indicate the maximum matching radius used in the tracking procedure explained later in this section.

The A2OBB methodology assumes the vehicle to be rectangular in shape. Therefore, the manually annotated datasets only contain rigid-body vehicles. For vehicles with trailers, only the leading vehicle is annotated.

All 800 images are annotated in ABB format using the OpenLabeling tool [40]. The purpose of this type of annotation is to serve as input for the method's evaluation, to serve as source

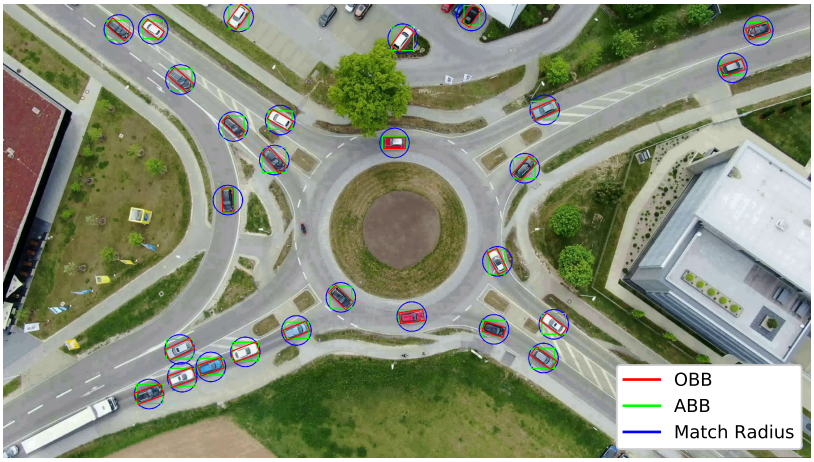


Figure 4-1: An example of annotations in OBB (red boxes) and ABB format (green boxes). The image also depicts a matching radius (blue circle) of 70 pixels around each vehicle for tracking purposes.

for investigating the validity of the aforementioned assumptions and to serve as training data for the YOLOv3 detection network. In total, the set of ABB annotations consists of 14.860 vehicle observations, describing the movement of 126 vehicle instances.

In addition, every 4th image of the first image sequence is annotated in OBB format using the labelImg_OBB tool [41]. These OBBs are used for investigating the assumptions as well and it serve as ground-truth during the method's evaluation in the latter two experiments. The set of OBB annotations describes a total of 2296 vehicle observations in 100 images covering 63 vehicle instances.

Vehicle detection: AOBBs

After obtaining the OBBs, a third dataset is created synthetically by generating axis-aligned bounding boxes around the OBB annotations using equations 3-2. These bounding boxes are again referred to as AOBBs.

Vehicle detection: YOLO Detections

A set of detections is created to evaluate A2OBB's performance as an extension to existing detection networks. Detections are created by an in-house trained out-of-the-box implementation of the YOLOv3 network [26], performing detections in ABB format.

The detection network is trained in the Darknet neural network framework [42]. For the training procedure, 350 images, 175 in each sequence, are used to train and evaluate the network in 10.000 epochs, using a batch size of 64 and a learning rate of 0.001. The optimizer used in this procedure is the standardly implemented Stochastic Gradient Decent (SGD).

Vehicle tracking

A2OBB works with a sequence of ABBs describing multiple observations. Therefore, to track annotations through the image sequence, a Nearest-Neighbour tracking algorithm is applied. The maximum displacement of a vehicle between consecutive image is determined by the frame rate of the image sequence. The higher the frame rate of the video, the smaller the distance between consecutive observations of a one vehicle will be.

Given a high enough frame rate, the displacement will be small enough to re-identify a specific vehicle instance in the next image. For a bounding box belonging to a certain vehicle instance, the closest bounding box in the next image will belong to the same instance. This procedure only works when all vehicle observations are annotated. For manually annotated datasets, this is usually the case. However, for datasets resulting from a detection network, some observations may be missed or falsely annotated. As a consequence, some vehicle observations will be assigned to incorrect vehicle instances and some vehicle instances will be divided in multiple smaller instances.

To minimize the influence of faulty detections, a maximum matching radius is introduced based on the maximum displacement of a vehicle between consecutive images. Only vehicles within that radius are considered in the tracking procedure. For example, given a vehicle travelling at 60 km/h (17 m/s) in imagery with a frame rate of 5 fps, the vehicle displacement is equal to 3.4m. In the recorded dataset, this translates to a displacement of 68 pixels. Assuming all observed vehicles to travel at velocities under 60 km/h, which is reasonable around a roundabout, the maximum displacement of bounding boxes of the same vehicle instance in consecutive frames must be lower than 68 pixels. Figure 4-1 illustrates a matching radius of 70 pixels (depicted by the blue circles around each vehicle observation).

Missing observations in vehicle instances will cut that instance in multiple smaller instances with a fewer observations. As long as the number of observations and the variance in orientation in the resulting set are large enough, this will not introduce any problems.

The tracking procedure is used to obtain vehicle instances for the manually annotated ABBs and the YOLO detections. For the annotated OBBs, the displacement of vehicles between subsequent images is too large for the Nearest-Neighbour approach. Therefore, the OBBs are matched per frame to the ABB vehicle instances.

Dataset characteristics

The dimensions of the observed vehicles, described by the dimensions of the annotated OBBs, are used to generate a synthetic dataset. The shapes of all vehicle observations are illustrated in figure 4-2. As previously mentioned, OBB dimensions are usually expressed in width and height. However, to avoid confusion with vehicle shapes, the OBB dimensions are expressed in length and width.

Figure 4-2 shows that vehicle instances observed in the OBB dataset are similar in width, but vary in length. The majority of the pixel values for the width lie between 40 and 60, whereas the length varies between 65 and 155. The diversity in vehicle shapes can also be identified when observing the vehicle aspect ratios, ranging from 1.5 to 3.1.

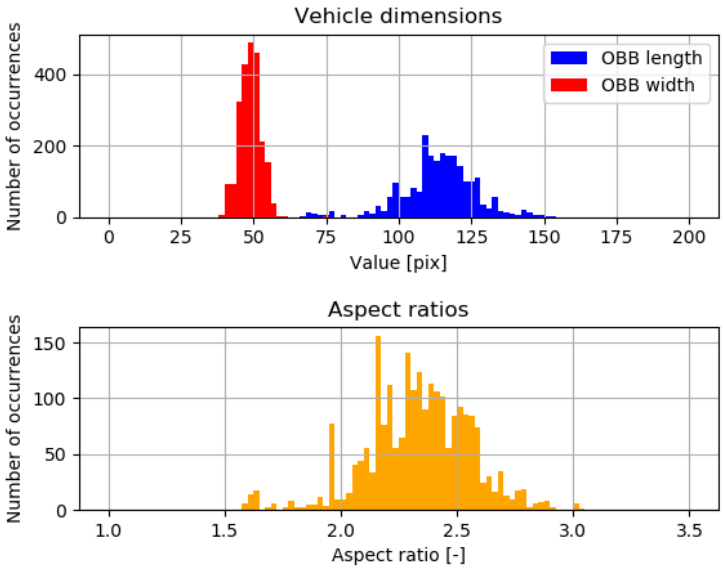


Figure 4-2: An overview of vehicle dimensions in pixel values (top) and vehicle aspect ratios (bottom). Blue and red describe the OBB length and width respectively and orange depicts the aspect ratio.

A2OBB requires the vehicle to change orientation in the set of ABBs. To ensure this in the dataset, only vehicle sequences with an angular variance above 5 ° are considered. This leaves a total of 47 vehicle sequences in the OBB annotation set for further evaluation.

4-1-2 Synthetic data generation

To validate the A2OBB's working principle in a scenario where all of its required assumptions hold, a set of synthetically generated ABBs is created. These assumptions state that a set of OBBs describing an ideal vehicle should consist of multiple rectangles, each of the same size but in different orientations. Based on these rectangles, a set of ABBs can be constructed using equations 3-2.

Recreating this scenario synthetically is done by generating a specific number of rectangles having a predefined length and width. These rectangles are rotated by an angle taken from a specific distribution, creating a set of OBBs. Subsequently, for each OBB, its surrounding ABB is calculated using equations 3-2. Note that the resulting set of bounding boxes is different to the previously mentioned set of AOBBs.

The dimensions of the rectangles are based on vehicle dimensions seen in the in-house collected dataset (depicted in figure 4-2). These vehicles vary much in length ($65 < L < 160$), but less in width ($W \simeq 50$). For that reason, three different sets of dimensions are used, covering a varying length ($L = 75, L = 100, L = 150$) but a constant width ($W = 50$). Additionally, to mimic performance at lower and higher resolutions, these dimensions are also halved and doubled.

Altogether, the following shapes are considered:

- 75×50
- 100×50
- 150×50
- 37.5×25
- 50×25
- 75×25
- 150×100
- 200×100
- 300×100

The number of rectangles in each set is based on the number of observations per vehicle instance seen in the in-house collected dataset. In the set of manually annotated OBBs, the number of observations per vehicle vary between 5 and 100. Therefore, to simulate this synthetically, sets of 5, 25, 50 and 100 rectangles are generated.

The angles that describe the rectangle's orientation are taken from a distribution that describes the manoeuvre that is simulated. At the roundabout depicted in figure 4-1, three types of manoeuvres can be distinguished, assuming a vehicle moving through the scene: a vehicle making one, two or three turns, changing its orientation for the duration of the turn, but maintaining its orientation outside.

The hardest of these manoeuvres to reconstruct is a single-turn manoeuvre as it contains the smallest variation of the vehicle's orientation. In addition, a two- and three-turn manoeuvre essentially comprise multiple single-turn manoeuvres. For that reason, the analysis is restricted to a single-turn manoeuvre is simulated in this dataset. Also note that the synthetic dataset does not cover stationary vehicles or vehicles moving in a straight line as their orientation will not change over time.

To simulate a single-turn manoeuvre, an angular distribution is generated comprising an entering section, a section describing the turn itself, and a departing section. The starting and departing sections each consist of a fixed value randomly generated over the interval $[0, 2\pi]$, representing a vehicle in all possible orientations. The turn itself describes an orientation linearly increasing or decreasing from the entering to the departing orientation.

In total, the synthetic dataset covers 36 different scenarios. Each of these scenarios is ran 1000 times.

4-1-3 Metrics

This section describes the four metrics used to evaluate A2OBB's reconstruction performance. These metrics differ from performance metrics conventionally used for object detection networks. Object detection networks are commonly evaluated using methods that assess correctness of detection rather than observing their accuracy. However, in the scope of A2OBB, these detections are already provided, rendering the testing of the correctness of detections irrelevant. Instead, this research reverts to more elementary metrics that describe the accuracy of the reconstructed OBBs in more detail.

The first three metrics are based on three parameters that define an OBB in 2D. In general, an OBB in 2D is defined by five parameters: the bounding box location, defined by its centre coordinates (x, y), its length and width (L, W) and its orientation (θ). However, as described in section 3-1-1, the location of a bounding box remains unchanged by A2OBB. Thus, the accuracy of the location of the reconstructed OBBs depends fully on the accuracy of the

location of the ABBs. For that reason, only the latter three parameters (L, W and θ) are considered. For all three parameters, the error of the predicted dimensions with respect to the ground-truth is considered. To allow for a comparison of vehicles of multiple sizes, the length and width are evaluated relative to the value of the ground truth.

Due to a deviation from conventionally used performance metrics, benchmarking the proposed methodology proves to be difficult. To still allow for a basic comparison, the Intersection over Union (IoU) is also evaluated. This metric describes the ratio of the intersection between the reconstructed OBB (OBB_R) with its ground-truth (OBB_{GT}) over their union, calculated using the the following equation:

$$IoU = \frac{OBB_R \cap OBB_{GT}}{OBB_R \cup OBB_{GT}} \tag{4-1}$$

Note that section 3-3-4 mentions the limitation of A2OBB to find one of 4 possible angles. Selecting the correct angle from these options is not part of this approach. Therefore, for the evaluation of the IoU and the angle, the solution resulting in the best IoU and the smallest angular error is taken from the four possible options.

Lastly,

For every evaluation metric, the value is taken as an average over all observations belonging to one vehicle instance.

Summarized, the metrics used for the evaluation are:

- E_L, E_W: the average relative error in the OBB length and width to evaluate the vehicle dimension reconstruction;
- E_θ: the average error of the vehicle orientation to evaluate the orientation reconstruction;
- IoU: the average Intersection over Union of the vehicle with its ground-truth to evaluate the contour reconstruction.

4-2 Experiment 1: Performance in the ideal case

This section describes the setup and results of the first experiment. In this experiment, A2OBB's performance is evaluated on a dataset recreating a scenario where the assumptions made in the method hold. To create this scenario, synthetic ABBs are generated using the procedure described in 4-1-2.

The main goal of this question is to answer the following research question:

"Can OBBs be reconstructed from ABBs that describe multiple observations of a rectangular vehicle with constant dimensions?"

As suggested in section 3-1, for a set of ABBs that describe a rectangular vehicle with constant dimensions, a set of equations can be constructed that describe the geometrical relationship. This set of equations will have a unique solution given that the orientation of all OBBs lies on the interval $[0, \frac{\pi}{2}]$ and an aspect ratio for which $R > 1$. Approximating this reconstruction using a non-linear least squares optimization should be able to find the correct solution for every set of bounding boxes containing two or more observations. For that reason, the errors for are expected to be negligible for each scenario covered in the synthetic dataset and all IoU values are expected to be equal to 1.

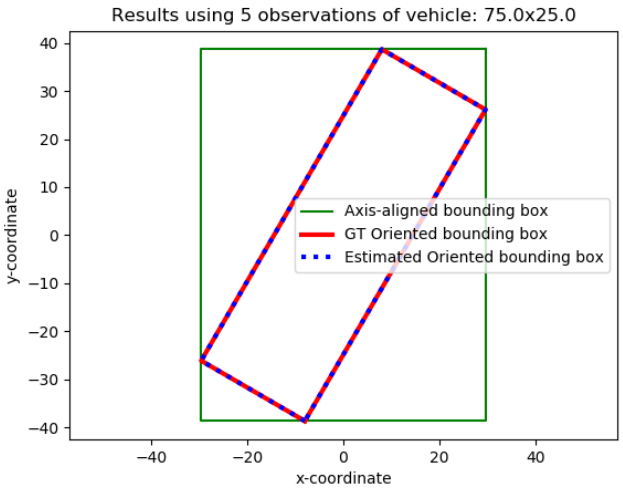


Figure 4-3: An example of a reconstructed bounding box based on synthetically generated data. This dataset mimics a vehicle of size 75x25, observed five times. The red line shows the original OBB, green depicts the constructed ABB and the blue-dotted line illustrates the reconstructed OBB.

	Vehicle size: 75x25			
Observations	Average E_L [pix]	Average E_W [pix]	Average E_θ[rad]	Average IoU
5	$3.79\text{x}10^{-16}$	$7.67\text{x}10^{-16}$	$-8.90\text{x}10^{-17}$	1
25	$1.32\text{x}10^{-16}$	$2.13\text{x}10^{-15}$	$-7.47\text{x}10^{-17}$	1
50	$7.58\text{x}10^{-17}$	$3.15\text{x}10^{-15}$	$-6.91\text{x}10^{-17}$	1
100	$5.68\text{x}10^{-17}$	$3.82\text{x}10^{-15}$	$-6.96\text{x}10^{-17}$	1

Table 4-1: Results of the reconstruction for a vehicle with dimensions 75x25 performing a single turn. The error is described as the average error over all 1000 vehicle sequences per scenario. Note that the errors in this experiment are in the order of magnitude of numeric precision errors.

4-2-1 A2OBB's performance

A2OBB's reconstruction performance is evaluated using the synthetically generated dataset. In total, 36 different scenarios have been evaluated, covering nine vehicle sizes and four different numbers of observations.

Figure 4-3 illustrates an example of a reconstruction based on a vehicle observed five times making a single turn. The vehicle's length and width are equal to 75 and 25 respectively. The figure illustrates that the reconstruction seems to overlap perfectly with its ground-truth.

To provide a more accurate analysis and a numerical performance indication, the method's evaluation metrics are illustrated table 4-1. The table depicts the results for a vehicle of size 75x25 pixels the different numbers of observations. Note that for this experiment the errors are described in pixel values, since all vehicles have the same size. For all four scenarios covered in the table, the average reconstruction errors in length, width and orientation are in the order of magnitude of numeric precision errors and the IoU is equal to 1. Based on these values, the conclusion can be drawn that for the simulated scenarios covering a vehicle of size 75x25 pixels, the reconstruction is nearly perfect. This adheres to what was expected.

Performing the reconstruction on datasets covering the remaining 32 scenarios yield similar results. The error values for every parameter in each scenario are in the order of magnitude of numeric precision errors.

4-2-2 Conclusion

To summarize, this experiment analyses A2OBB's performance on a dataset where the assumptions made in A2OBB's approach hold to provide proof for the method's working principle and to validate the implementation. A synthetic dataset is constructed based on in-house collected examples using the procedure described in section 4-1-2.

Based on the theoretical analysis, it is expected that a unique solution can be found for every set of bounding boxes containing two or more vehicle observations For every vehicle instance simulated in the synthetic dataset, the set geometrical of equations has an optimal solution that is closely approximated by the least-squares implementation. These results would suggest that the optimizer is implemented correctly. In addition, the results suggest that for an ideal scenario, where the required assumptions hold, OBBs can indeed be reconstructed using a set of ABBs using A2OBB.

4-3 Experiment 2: Reconstruction for AOBBs with changing vehicle dimensions

This section describes the setup and results of the second experiment. A2OBB's approach makes the assumption that the perceived size of vehicles in low-altitude aerial imagery remains constant as a vehicle moves through the captured scene. However, as argued in sections 3-2-4 and 3-2-2, the perceived size of vehicles will vary due to lens distortions, due to a change of perspective between the camera and the vehicle and due to inaccuracies in individual annotations.

While collecting the imagery, lens distortions are have already been compensated. However, the change in perspective and annotations errors will still affect how a vehicle is perceived. To find out how both factors affect the perceived vehicle size and, ultimately, how they affect A2OBB's OBB reconstruction performance, the goal of this experiment is to answer the following research question:

"To what extent can OBBs with varying dimensions be reconstructed from a set of ABBs generated around OBBs using geometry?"

An answer to this question is formulated in three steps. The first step assesses how the perception of a vehicle size changes between subsequent observations and argues how this effect can be minimized. The second step then evaluates the impact of a change in OBB dimensions on the reconstruction performance by applying A2OBB to the set of AOBBs. Initially, the reconstruction performance without any corrections to the 'rounded-corner effect' is evaluated. After that, the correction factor introduced in section 3-3-2 is tested to see how the quality of the reconstructions improves.

4-3-1 OBB dimension analysis

This section describes the analysis performed to identify the magnitude and shape of a change in the perceived size of vehicles in low-altitude UAV imagery. This analysis is performed based on the set of manually annotated OBBs. The dimensions of OBB annotations form a fair representation of the shape of a vehicle observed from top-down. Therefore, an analysis of OBB dimensions should in turn be a fair approximation of how the observed vehicle dimensions change.

Overall dimensional change

As described in section 4-1-1, the set of OBBs is grouped into vehicle instances. Figure 4-4 illustrates the change in OBB length and width for every vehicle instance. It depicts the dimensions of all observations for each vehicle instance expressed as the difference between the dimension of that observation and the average dimensions of that instance. The figure illustrates that when observing a vehicle over the whole scene, its dimensions vary significantly.

For most of the observed vehicle instances, the width shows a highly volatile behaviour around the mean value. This volatility is likely caused by a high sensitivity to inaccurate annotations

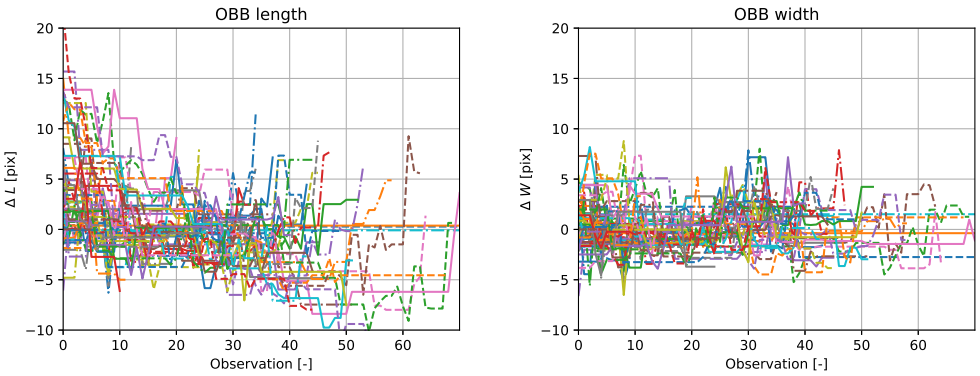


Figure 4-4: An illustration of the change in OBB dimensions for all vehicle instances over the course of their observations.

due to the relatively low magnitude of this dimension. To demonstrate, a (not unlikely) annotation error of 3 pixels already has a significant impact on the observed width ($\simeq 50$).

The change in observed length exhibits a similar volatile behaviour. However, the figure also illustrates a more gradual change. The length of the bounding boxes appears to decrease gradually over the first observations of each vehicle and seems to increase in the last observations.

The gradual change in length is likely to be caused by a change in perspective between the vehicle and the camera. As described in section 3-2-2, as a vehicle approaches the image centre, the perceived size of the vehicle decreases. In 3D, this effect depends on three parameters: (1) the vehicle's distance to the image centre, (2) its orientation with respect to the image centre and (3) the shape of the vehicle. The influence of the vehicle's distance to the image centre and the shape of the occluded and included sides on the vehicle's perceived size can both be identified in the recorded data.

Note that the influence of the perspective transformation also depends on the altitude of the drone. The higher the altitude of the drone, the smaller the influence of the perspective transformation will be. However, since all annotations are obtained from the same dataset, recorded at the same altitude, the influence of this parameter can not be analysed.

Vehicle's distance to the image centre

A dependency of the perspective transformation on the vehicle's distance to the image centre can best be illustrated by observing the OBB dimensions in parts of the image where the vehicle's orientation with respect to the image centre remains constant. To recall from section 3-2-2, vehicle's orientation with respect to the image centre affects how the change in the observed vehicle size is distributed over the length and width of the vehicle. Keeping this orientation relatively constant will also keep this distribution constant and allows clear identification of the dependency of the perspective transformation on the vehicle's distance to the image centre.

By keeping the vehicle's orientation with respect to the image centre constant, the influence of the perspective tranformation can be approximated as a 2D problem, as described in section

3-2-2 and depicted in figure 3-6. Section 3-2-2 suggests that in this case, the influence of the perspective transformation should linearly depend on a vehicle's distance to the image centre.

Figure 4-5: An illustration of bottom-right and bottom-left exit in the in-house captured imagery.

To verify this statement, the OBB dimensions of vehicles appearing in the bottom-right and bottom-left exit of the roundabout (figure 4-5) are examined. Most vehicles appearing in these regions move in a relatively straight line, aligned with their displacement with respect to the centre of the image, and are observed a sufficient number of times to reduce the influence noise introduced by annotation errors. Since the vehicles are aligned with their displacement to the image centre, the change in perspective will mainly affect the perceived length.

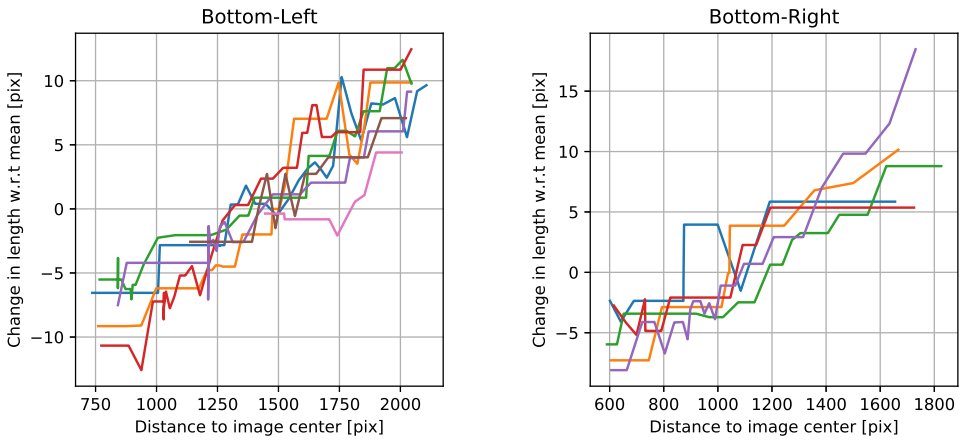


Figure 4-6: The dependency of the change in observed vehicle length (ΔL) on the distance vehicle's distance to the image center. The left and right image illustrate vehicles observed in the bottom-left and bottom-right of the image respectively

The ΔL of OBBs for instances appearing more than five times is depicted in figure 4-6. The figure shows the difference in length between individual OBBs of an instance with respect to the instance's average length as a function of the OBBs distance to the image centre.

Though some instances contain noise due to annotation errors, all vehicles do show a positive correlation between ΔL and the distance to the image centre. And, as previously suggested, for most vehicle instances, the relationship even appears to be linear.

Shape of the vehicle

The dependency of the perspective transformation on the vehicle shape can be illustrated by comparing vehicles based on their shape and size. To do so, the sets of OBB annotation are divided in two categories: 'Vans' and 'Passenger vehicles'. The category 'Vans' also includes buses and trucks.

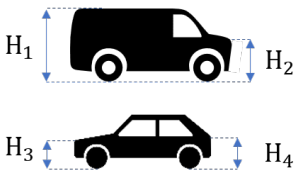


Figure 4-7: Length intersection for a van (top) and a regular passenger vehicle (bottom), illustrating the height of the sides that affect the perspective transformation.

The shape of vehicles affects the perceived dimensions depending on the height of sides that start to appear or disappear when the perspective changes. The height of these sides are illustrated for basic examples of both categories in 4-7. The height of the sides for the more box-shaped Van is relatively large compared to the more complex-shaped Passenger vehicle. Therefore, the influence of the perspective transformation for Vans should be larger than for the passenger vehicles.

To illustrate the influence of the shape in the data, the change in the annotated OBB length is depicted for five Vans and for five Passenger vehicles in figure 4-8

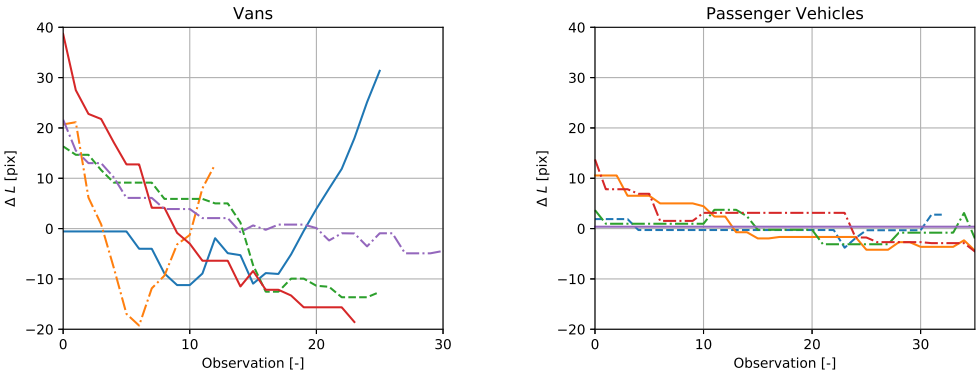


Figure 4-8: Change in the observed OBB length per vehicle instance compared to the average length of that vehicle instance. The left column depicts the change for five Vans and the right column for five Passenger Vehicles.

Figure 4-8 clearly illustrates a large difference in the change in observed length between both categories. For the majority of the observed vans and trucks, the length changes up to 20 pixels from the average value, whereas for most passenger vehicles, the difference is no more than 10 pixels. The difference in magnitude for the change in the perceived size between

both categories indicates that the shape of the vehicle indeed plays an important role in the perspective transformation. For vans, the impact of the perspective transformation is fairly large, while for passenger vehicles it is relatively low. Therefore, A2OBB's overall reconstruction performance is expected to be much better for passenger vehicles compared to vans.

Interestingly, figure 4-8 also illustrates the negative correlation between the perceived vehicle length and the vehicle's distance to the image centre for Vans. For example, the yellow line describes a truck approaching and receding from the image centre, whereas the green, purple and red line describe a van only approaching the image centre. These relationships can be easily identified, since vans and trucks are box-shaped. The relationship is less strong for passenger vehicles.

Conclusion

To summarize, the perceived size of OBB annotations describing a vehicle in low-altitude aerial imagery changes as the vehicle moves through the scene. In both dimensions, the influence of incorrect annotations can be identified in the form of noise. For the length, a more gradual change can also be seen. This change is likely to be caused by a perspective tranformation between the vehicle and the camera.

The influence of the perspective transformation can be identified by viewing the dimensions of vehicles over straight road sections aligned with its displacement from the image centre. Over these sections, the increase in length appears to be linear with the distance of vehicles to the image centre. In addition, the magnitude of the impact made by the perspective transformation depends on the shape of vehicles. Vans and trucks are box-shaped and will therefore be affected more significantly than more complex-shaped passenger vehicles, where the height of the 'top-down occluded' side is lower.

Using this information, the impact of the perspective transformation can be minimized by considering only observations that are relatively equidistant from the origin. Also, based on the linear relationship, observations can be inter- and extrapolated over road sections that are aligned with the vehicle's displacement from the image centre.

4-3-2 Reconstruction performance with changing vehicle dimensions

This section describes the analysis performed to quantify the impact of a change in observed vehicle size on A2OBB's reconstruction performance. To allow for a well-defined analysis, the method is applied to the sythetically constructed AOBBs, creating a scenario where only the second assumption is enforced to hold.

Comparing A2OBBs reconstructions for AOBBs to the results from experiment 1 (section 4-2) will indicate the impact of a change in OBB dimensions on the reconstruction performance. The first experiment (section 4-2-1) illustrates that for an ideal scenario the reconstructions are accurate. Errors that occur are in the order of magnitude of numeric precision errors. Therefore, any error found A2OBB's reconstructions based on AOBBs should be the result of changing OBB dimensions.

An example of the OBBs reconstructed from AOBBs is depicted in figure 4-9. This figure depicts the ground-truth OBBs (red), the generated AOBBs (green) and the reconstructed OBBs (blue). To allow for an interpretable image, only the centre of the image is depicted.

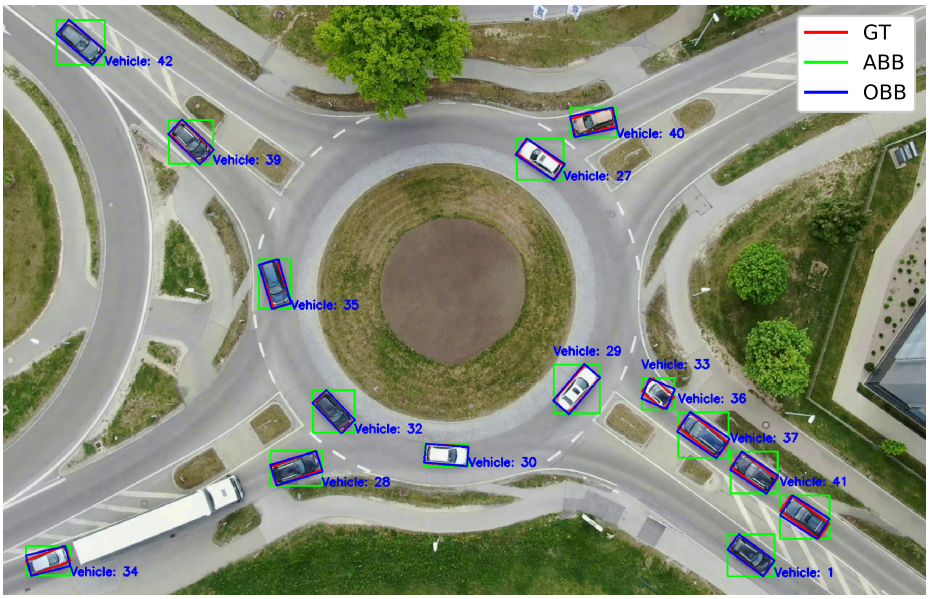


Figure 4-9: An example image of the reconstruction performance based on ABBs fitted around manually annotated OBBs. Red boxes depict the ground-truth OBBs, green illustrate the constructed ABBs and blue shows the reconstructed OBBs. The depicted image is part of the in-house collected imagery.

For the majority of the reconstructed vehicles, the OBBs overlap quite well with the ground-truth. It is interesting to note the reconstructions for vehicles 34 (bottom-left), 33, 36, 37 and 41 (bottom-right). These vehicles are approximated to be thicker but shorter compared to their ground-truth contours. Each of these vehicles spends time waiting to enter the roundabout. As a consequence, a vehicle is perceived multiple times at the same location. The perceived size at that location occurs multiple times in the vehicle sequence possibly introducing a bias. In addition, in the annotation process vehicle observations that do not move are copied to the next frame. If the initial annotation at that location contains annotation errors, these errors are also copied to the other observations at that location.

Another result to note is the reconstruction for vehicle 33. This vehicle relatively has small vehicle parameters. Consequently, annotation errors will have a larger impact on the observed vehicle size and thus on the reconstruction performance. In this case, the annotated OBB is slightly shorter than the actual vehicle. As a result, the reconstructed orientation is slightly off.

A more accurate analysis can be made by observing the evaluation metrics for the AOBB reconstructions. Figure 4-10 depicts the average relative errors in length, width, the average error in orientation and the average IoU for each vehicle instance. First off, the figure shows that a change in the observed vehicle size does have an impact on the reconstruction

performance. It illustrates that the overall reconstructions contain a very small bias towards shorter but thicker vehicles. This is likely caused by the effect that occurs for vehicles 33, 34, 36, 37 and 41, described earlier.

In addition, the impact of a change in OBB dimensions is relatively larger for the width compared to the length. The magnitude of the width is smaller, therefore the relative impact of both the change in perspective and annotation errors is larger.

Consequently, the vehicles are reconstructed having an average orientation that slightly deviates from the actual orientation. Nonetheless, the majority of the reconstructions are accurate within 2.5°. As a result, the reconstructed IoU is above 85% for the majority of reconstructions.

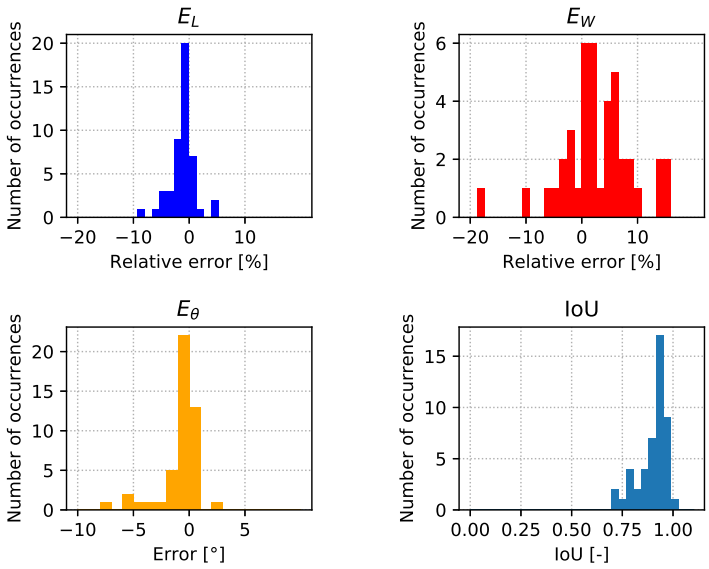


Figure 4-10: Performance metrics for AOBBs. The figure displays histograms of the average error in length, width and orientation and the average IoU per vehicle instance.

4-3-3 Reducing impact of the perspective transformation

In essence, A2OBB's reconstruction performance drops significantly when considering vehicles with a changing object size. However, the analysis performed in section 4-3-1 suggests that a change in observed vehicle size can be reduced significantly by considering only a selection of the observations. By filtering out observations that are equidistant to the image centre, the overall influence of the perspective transformation can be minimized, improving the reconstruction performance.

In addition, the reconstructions in the previous section also suggest that a bias might be introduced by a vehicle that is observed at the same location multiple times. The specific size of vehicles at that location along with possible annotation errors introduce a bias in the optimization. By filtering out these 'duplicates', this bias can also be minimized.

Figure 4-11: The part of the image describing the actual roundabout.

To test these suggestions, A2OBB is applied to sets of AOBBs only describing observations that appear on the actual roundabout. Vehicles observed in this region of the image will not only be relatively equidistant from the image centre, but will mostly be non-stationary. An illustration of the part of the image that captures the actual roundabout is depicted in figure 4-11.

The performance metrics for A2OBB's reconstructions only using observations on the actual roundabout are depicted in figure 4-12. Comparing these results to the reconstructions using all observations described in figure 4-10, illustrates a strong improvement. The error in the length is almost completely compensated and the error in the width is significantly reduced.

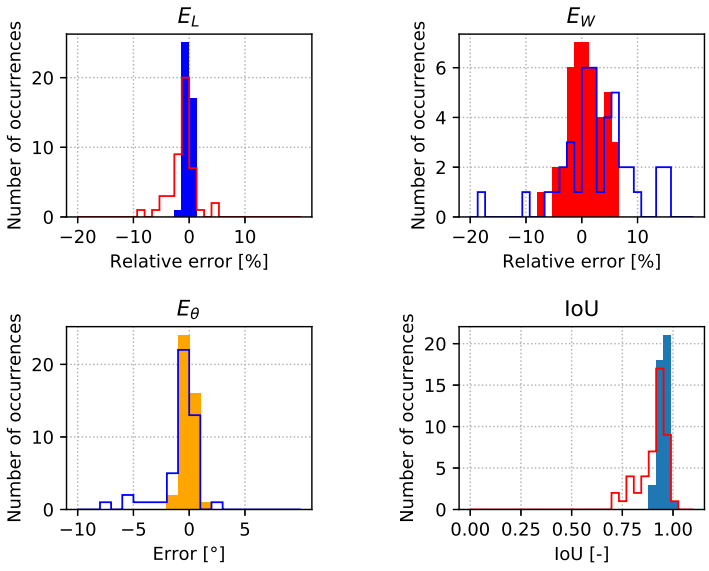


Figure 4-12: Performance metrics for AOBBs observed on the actual roundabout. The figure displays histograms of the average error in length, width and orientation and the average IoU per vehicle instance.

Vehicle observations omitted in the selected sequences are mostly affected in the observed length. Therefore, the length approximation improves more significantly than the estimation of the width. As a consequence of both improvements, the orientation approximation is also more accurate. All orientations are estimated within a range of 2°. In addition, the average IoU for all reconstructions is 95%. In literature, this is considered to be near-perfect [21]. Therefore, it is fair to say that using only observations in the middle of the image significantly reduces OBBs can be accurately reconstructed from AOBBs.

4-3-4 Conclusion

To summarize, this experiment was focussed on answering the following research question: *"To what extent can OBBs with varying dimensions be reconstructed from a set of ABBs generated around OBBs using geometry?"*

An answer to this question is formulated by analysing the size of OBBs describing observations of vehicles in low-altitude aerial imagery. These OBBs indicate that the perceived size of vehicles change as they move through the scene due to errors in annotations and a perspective transformation between the camera and the vehicle.

The perspective transformation depends on three parameters, two of which can be identified in the data: (1) the vehicle's distance to the image centre and (2) the shape and height of the vehicle. In-house annotated data show a linear increase of the perceived size of vehicles moving in a direction relatively aligned with the vehicle's displacement from the image centre. In addition, a comparison between observations of Vans and regular passenger vehicles indicate a dependency of the magnitude of the perspective transformation on the shape of the vehicle.

The change in dimensions induced by the perspective transformation can be minimized by considering observations relatively equidistant to the image centre and by filtering out multiple observations of a single vehicle at the same location. Using a selection of the OBBs, all observed at the actual roundabout, the length can be reconstructed within 2%, the width within 7% and orientation within an error of 2 °, resulting in an average IoU of 95%.

Due to the identified linearity, it should be possible to inter- and extrapolate vehicle sizes over road sections that are relatively aligned with their displacement from the image centre. Additionally, including the shape of the model and the orientation with respect to the vehicle's displacement in approximating the effect of the perspective transformations might further improve the results.

4-4 Experiment 3: Reconstruction for manually annotated ABBs

This section describes the setup and results of the third experiment. This experiment is the first of two experiments investigating the impact of a difference between ABBs and AOBBs on the reconstruction performance. The focus of this experiment lies on evaluating manually annotated ABBs.

A2OBB's approach states that the given set of ABBs should be constructable from a target set of OBBs using geometry. Essentially, this holds if the objects described by the bounding boxes are perfectly rectangular and the ABBs coincide AOBBs. However, this is generally not the case for vehicles perceived from aerial perspective. As a consequence of the previously described perspective transformation, the 'rounded-corner effect'(section 3-2-3 and errors that occur in the annotation process, a misfit between manually annotated ABBs and AOBBs is introduced. The purpose of this experiment is to assess how this misfit affects the reconstruction performance and to find out what choices can be made to reduce its impact on the quality of the reconstructions.

The influence of the perspective transformation is addressed in the previous experiment (section 4-3). Therefore, the focus in this experiment lies on assessing the influences of rounded corners and annotation errors. In addition, results of this experiment will allow a fair judgement of the method's applicability as a means to extend a set of existing manually annotated ABBs to OBBs. Combining both purposes, the research question of this experiment is:

"To what extent can OBBs be reconstructed from manually annotated ABBs?"

An answer to this question is formulated in three steps, similarly to the approach used in section 4-3. First, an analysis is performed on the shape of the misfit between manually annotated ABBs and AOBBs. Secondly, the impact of the misfit on A2OBB's reconstruction performance is evaluated by comparing the results using manually annotated ABBs to results of the previous experiment. Subsequently, the correction factor introduced in section 3-3-2 is applied and evaluated to see if the impact of the 'rounded-corner effect' can be mitigated.

4-4-1 ABB-AOBB difference for manual annotations

This section describes the analysis performed to identify the shape and magnitude of the mismatch between manually annotated ABBs and AOBBs. To reduce the impact of the perspective transformation, only bounding boxes describing vehicles on the actual roundabout, depicted in figure 4-11, are considered.

A histogram of the difference between the length and width of ABBs with respect to corresponding AOBBs is depicted in figure 4-13. Note that the errors in the length and width of the ABBs are illustrated respectively as E_x and E_y to differentiate these error from the errors used for the OBB predictions.

This figure clearly illustrates that there is indeed a significant mismatch between both bounding boxes. The average manually annotated ABBs is shorter and/or narrower compared to AOBBs and this is in line from what can be expected based on the 'rounded-corner effect' described in section 3-2-3 and illustrated in figure 3-8a. Basically, due to the 'rounded-corner effect' ABBs will always be equal or smaller compared to AOBBs depending on the orientation of the vehicle.

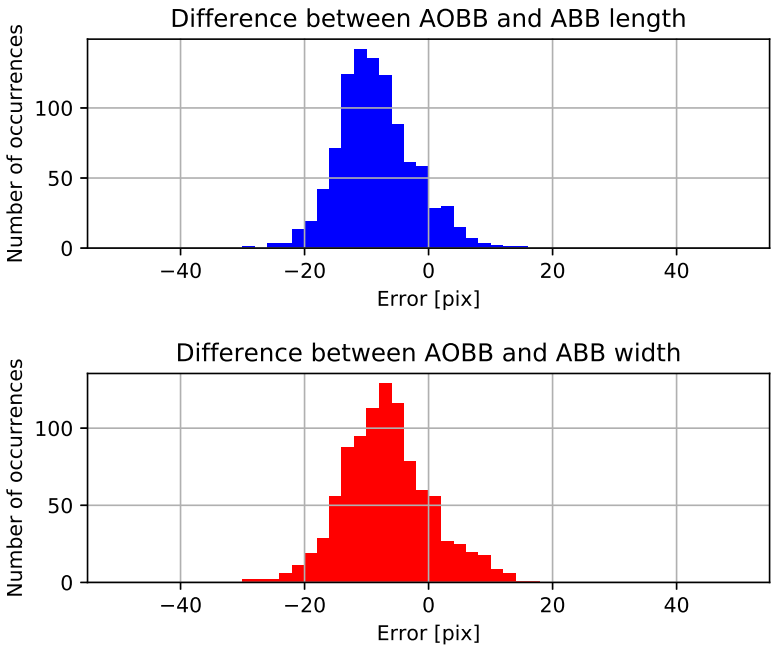


Figure 4-13: Errors in the dimensions of annotated ABBs and AOBBs

In addition to the negative bias for both dimensions, figure 4-13 also exhibits a significant amount of variance for both errors. For some vehicles, the manually annotated bounding boxes are even longer and/or wider compared to the AOBBs in contrast to what can be expected from the 'rounded-corner effect'.

To provide more insight into why this variance and these larger manually annotated ABBs occur, the error between both bounding boxes is plotted as a function of the vehicle's orientation in figure 4-14. In addition, to illustrate that the variance of the error corresponds to the variance between the errors prescribed by the 'rounded-corner effect' for corners with different radii, the figure also depicts the errors predicted for three vehicles differently rounded vehicles having ellipse-shaped corners with different radii. To recall from section 3-2-3, the ABB-AOBB difference predicted by the 'rounded-corner effect' can be approximated for vehicles oriented over an interval of $[0, \pi/2]$ using equations 3-8 and 3-9. In figure 4-14, the radii used for the corners of the vehicles are: $(R_L, R_W = 2$ pix), $(R_L, R_W = 5$ pix) and $(R_L, R_W = 10$ pix).

Figure 4-14 clearly illustrates a dependency of the difference between both bounding boxes on the orientation of the vehicle. The closer an orientation approaches an angle of $\pm 45°$, the smaller the annotated ABB becomes compared to the AOBB. Similarly, as the orientation approaches an angle of $0°$ or $\pm 90°$, ABB becomes relatively larger.

The shape of the ABB-AOBB difference appears to be similar to the shape of the errors predicted by the 'rounded-corner effect'. The figure exhibits a fairly large variance for the difference in dimension for vehicles in the same orientation, which appears to be similar to the difference between the three predicted vehicle shapes. Therefore, the variance observed in the data could be explained by the fact that the dataset contains vehicles with differently

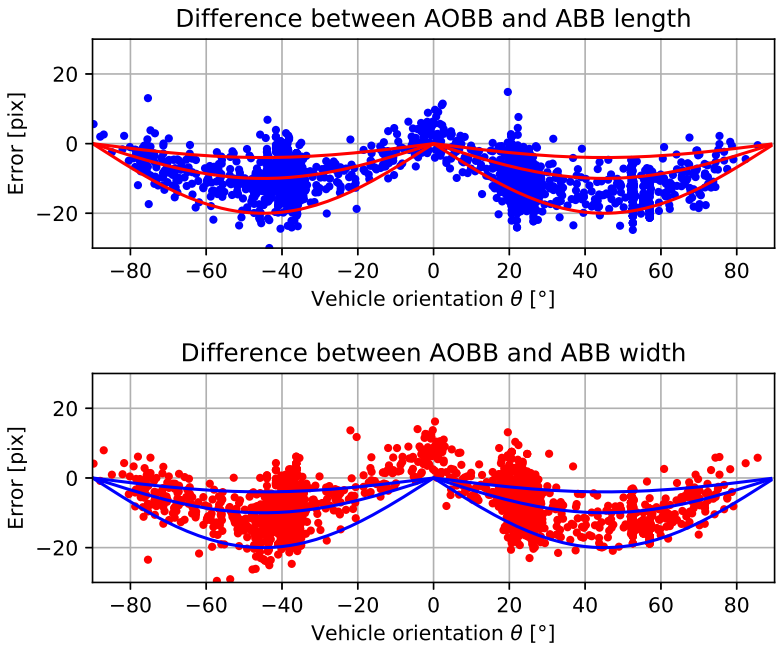


Figure 4-14: Error between ABBs and AOBBs as a function of the ground-truth orientation. The lines in both figures depicts the 'rounded-corner effect' for three differently shaped vehicles, calculated using equations 3-8 and 3-9. The radii of the ellipse-shaped corners are: $(R_L, R_W = 2)$, $(R_L, R_W = 5)$ and $(R_L, R_W = 10)$

shaped corners.

However, what is interesting about figure 4-14 is that for almost all observations having an orientation close to 0°, the annotated ABB is larger than the corresponding AOBB. These results contradict what is expected from the 'rounded-corner effect'.

To illustrate what causes this problem, three examples of vehicles oriented between $-5°$ and $5°$ are depicted in figure 4-15. The figures show that some green bounding boxes, illustrating the manually annotated ABBs, are indeed larger compared to the blue bounding boxes, depicting the AOBBs.

Figure 4-15: Three examples depicting vehicles oriented between $-5°$ and $5°$. The green boxes depict manually annotated ABBs. The blue boxes depict AOBBs generated around manually annotated OBBs.

The fact that the ABBs are larger appears to mainly be caused by faulty annotations. It is important to note that different annotation tools were used for obtaining both types of

bounding boxes and the annotations were performed at different occasions. Whether the discrepancy between both bounding boxes is caused by the fact that different tools were used or an error introduced by the annotator remains unclear.

Nonetheless, ABBs describing objects oriented around the image axes appear to be annotated with a positive bias in both dimensions. This error could just apply to those annotations, but could also be applicable to all observations. In the latter case, the error introduced would be systematic and would lead to a positive bias.

To summarize, the ABB-AOBB difference consists of two factors. The first factor comprises the 'rounded-corner effect' due to vehicles being observed as rounded rectangles. A second factor consists of error introduced by the annotator, introducing a positive bias for observations aligned with the image axes. The latter factor could point to a tendency for humans to annotated vehicles larger than their actual observed shape.

4-4-2 Reconstruction performance on manually annotated ABBs

This section describes the analysis performed to quantify A2OBB's reconstruction performance for manually annotated ABBs. This analysis is performed by applying A2OBB to the set of manually annotated ABBs. The result of this analysis will provide a measure of how accurately OBBs can be reconstructed for A2OBB's first intended application and will illustrate the impact of the ABB-AOBB difference on the quality of the reconstructions.

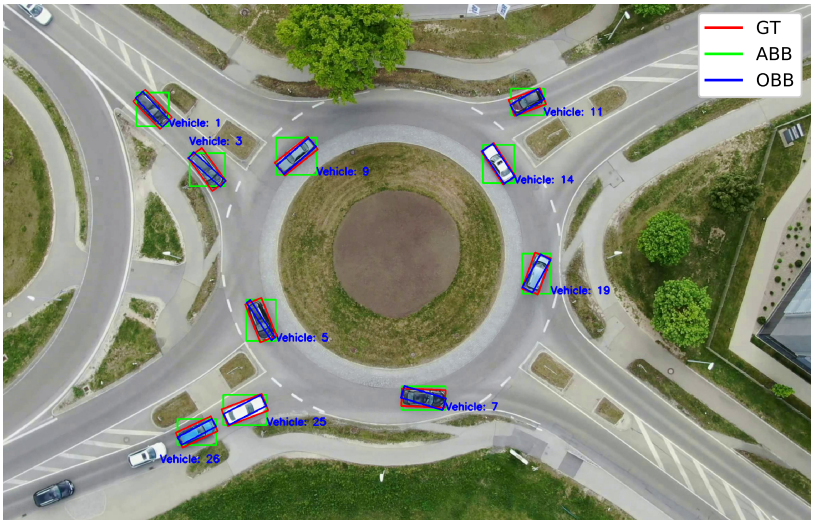


Figure 4-16: Reconstructions for manually annotated ABBs, depicted by the blue bounding boxes. The red bounding boxes illustrate the ground-truth OBBs and the green show the manually annotated ABBs.

Once more, to reduce the impact of the perspective transformation on the reconstruction performance, only observations in the middle of the image are considered. A visual example of the reconstructions for manually annotated ABBs is depicted in figure 4-16.

Note that the original image differs from the image depicted in figure 4-9. This particular image is selected to highlight the errors using manually annotated ABBs with and without correction.

The results depicted in this figure clearly illustrate the impact of the ABB-AOBB difference on the reconstruction quality. All vehicles depicted in this image are reconstructed much smaller, especially thinner, compared to their corresponding ground-truth. Note that the reconstructed length and orientation are still fairly accurate. This would suggest that mainly the width reconstruction is affected by the ABB-AOBB difference. This makes sense. The variation in both length and orientation is restricted. The magnitude and orientation of the vehicle's length axis can only vary between of the diagonal and the longest dimension of the ABB.

To obtain a more wholistic view on A2OBB's performance for manually annotated ABBs, the performance metrics for the reconstructions are depicted in figure 4-17. Note that the domain of the figure describing the errors for the width has increased to [-100%,100%] to include all width reconstructions. Similar to the visual reconstructions, these results illustrate a significant negative bias for the width. In addition, the figure also exhibits a large amount of variance.

The length on the other hand is reconstructed fairly accurately, mostly with an average error within 8% of the ground-truth. As a result, the orientation of most vehicles can be reconstructed with a maximum error of 6° and a slight positive bias. Observing the IoU illustrates a significant drop in performance of the overall contour reconstruction. The majority of the

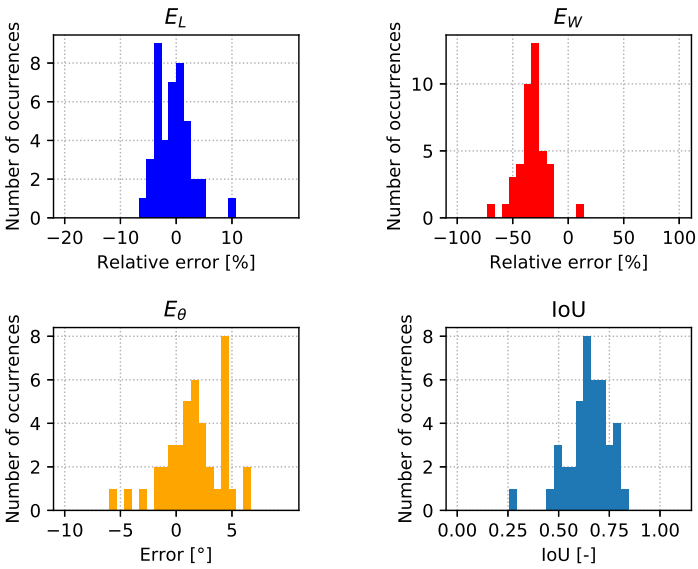


Figure 4-17: Performance metrics for manually annotated ABBs. The figure displays histograms of the average error in length, width and orientation and the average IoU per vehicle instance.

reconstructions has an average IoU below 75%, with outliers below 60%. This drop in performance is mostly induced by faulty width reconstructions. The reconstruction of the other two parameters are still fairly accurate. Judging the reconstruction performance based on the IoU alone might therefore paint the wrong picture. In the application of behaviour prediction for traffic agents for example, the orientation reconstruction is much more relevant.

To conclude, the ABB-AOBB difference for manually annotated ABBs introduces inconsistencies in the set of equations to such an extent, that the optimal solution to the set of equations results in an OBB that is much thinner compared to the actual observed shape. As a result, the quality of the average contour reconstruction drops significantly. On the other hand, the length and orientation of vehicles can be reconstructed with decent accuracy.

4-4-3 Compensating the ABB-AOBB difference

The previous section illustrates a significant drop in A2OBB's reconstruction performance as a consequence of the ABB-AOBB difference. Section 4-4-1 illustrates that this difference can be approximated based on the 'rounded-corner effect' for ellipse-shaped corners. This raises questions about whether modelling this 'rounded-corner effect' in the set of equations used for A2OBB's optimization can improve the reconstruction performance. To touch upon whether this is indeed the case, the two correction factors, C_W and C_L, introduced in section 3-3-2 and described by equations 3-12, are introduced and evaluated.

Each of the parameters used to define the correction factors is determined by fitting equations 3-12 to the results described in 4-14. To do so, the same non-linear least squares optimizer is used as the one used in A2OBB (described in 3-3-3).

The cost functions for this fit are:

$$L_{correction} = \sqrt{(C_L - E_L)^2} + \sqrt{(C_W - E_W)^2} \tag{4-2}$$

and are taken as the difference between the correction factors (C_L and C_W) and errors between the annotated ABBs and corresponding AOBBs (E_L and E_W).

To illustrate the value of the added constants, B_L and B_W, two types of correction factors are illustrated in figure 4-18. The first set of factors contains and the second set lacks the added constants (B_L and B_W), respectively depicted by a continuous and dotted line.

Figure 4-18 illustrates a better fit for the correction factor with the added constant. For the length, the difference is small, but for the width, the added constant significantly improves the fit. Based on this, one could argue that the error seen in the data can be approximated by the 'rounded-corner effect' plus an added bias.

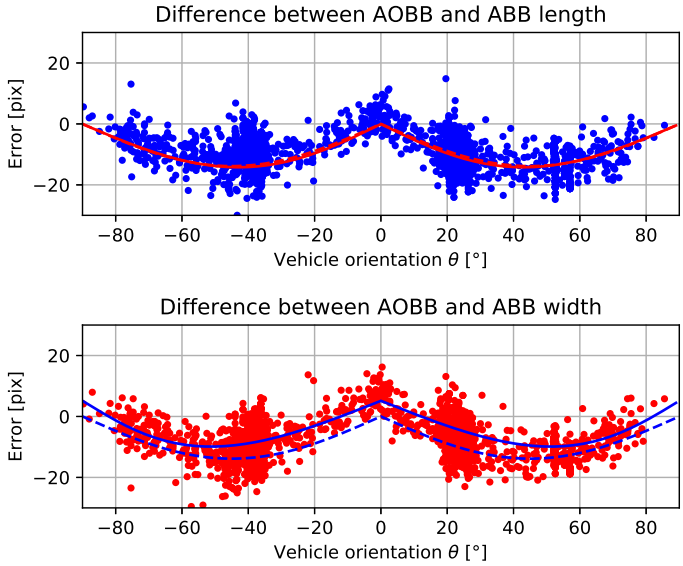


Figure 4-18: A visualisation of the correction factor for manually annotated ABBs. The dotted line depicts correction without and the continuous line with the added constants B_L and B_W

The effectiveness of the added bias could indicate that the annotation error introduces a positive bias, confirming the suggestion made in the previous section that the annotator has a tendency to annotate vehicles slightly larger than their perceived size.

To prevent from over-fitting the correction factor on the data, the final parameters used in the optimization are determined based on the ABB-AOBB difference of 50% of all observations. The resulting values are depicted in table 4-2. These parameters are also applied to detected ABBs, as described in section 4-5-4.

Parameter	Value
R_L	7.4
R_W	8.3
B_L	0.6
B_W	5.9

Table 4-2: Parameters used to calculated the correction factor for manually annotated ABBs.

Impact on the reconstruction performance

The impact of these correction factors is tested by applying A2OBB to the set of manually annotated ABBs once more. The resulting reconstructions are depicted in figure 4-19. The figure illustrates a significant improvement in A2OBB's performance. The width is reconstructed much more accurately and, as a result, the length and orientation improve as well. Note, however, that this improvement is not as strong for every vehicle. For example vehicles 3, 5 and 7 are still reconstructed thinner but longer compared to their actual size, implying that the ABB-AOBB difference is not fully compensated. This makes sense since the introduced correction factors are based on a general prediction of the 'rounded-corner effect'. In reality, each vehicle has its own shape, leading to stronger or weaker ABB-AOBB differences. Therefore, the degree to which the reconstruction improves is different per vehicle shape. Customizing the radii used in equations 3-12 per individual vehicle would improve the reconstruction performance further.

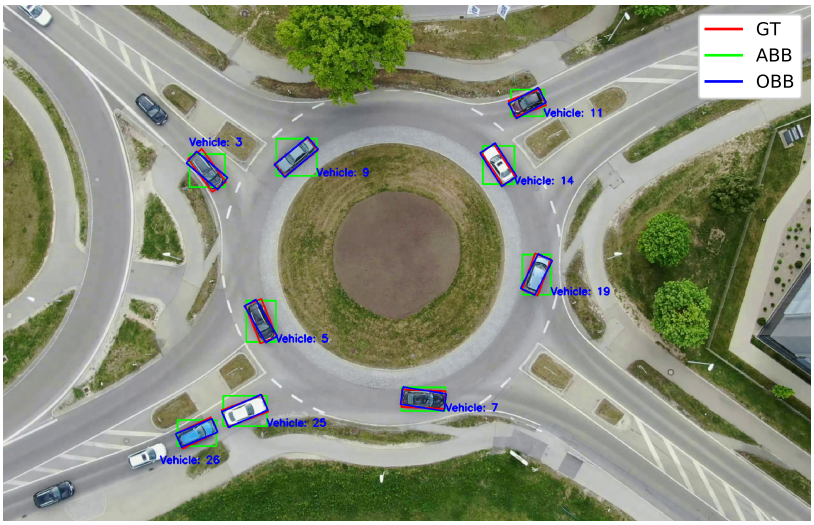


Figure 4-19: An example image of the reconstruction performance based on manually annotated ABBs, corrected for the 'rounded-corner effect'. Red boxes depict the ground-truth OBBs, green illustrate the detected ABBs and blue shows the reconstructed OBBs.

To illustrate how these improvements are reflected in the performance, the performance metrics of these reconstructions are displayed in figure 4-20. Comparing these metrics to the reconstructions for manually annotated ABBs without the added correction factors, described in figure 4-17, shows significant improvements as well. Firstly, the negative width bias is much less, resulting in reconstructions with an average relative width error below 15% of the ground-truth for the majority of the vehicles. In addition, the variance in the average error of the length is reduced to and the orientation estimation has improved. Most OBBs are reconstructed with a length having a maximum relative error of 5% and an orientation error below 2.5°. As a result, the overall contour estimation improves remarkably to an average of 85%.

Altogether, introducing basic correction factors based on the 'rounded-corner effect' improves the reconstruction performance significantly. Using these correction factors, OBBs can be reconstructed from manually annotated ABBs with an average accuracy of length and width within 5 and 15% of the ground-truth respectively, an orientation with an overall average error of below 2.5° and an average IoU of 85%. Note that the correction factors only describe the 'rounded-corner effect' generalized over a portion of the vehicles within 5 and 15% respectively vehicles. Formulating a correction factor per vehicle specifically would improve the results even more.

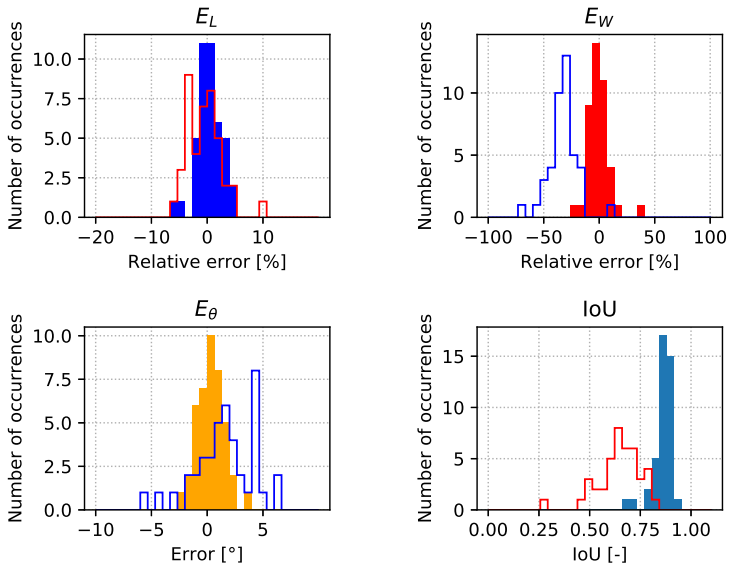


Figure 4-20: Performance metrics for manually annotated ABBs, corrected for the 'rounded-corner effect'. The figure displays histograms of the average error in length, width and orientation and the average IoU per vehicle instance. The shelled histograms depict the metrics without correction factors.

4-4-4 Conclusion

To summarize, the main focus of this experiment was to answer the research question:

"To what extent can OBBs be reconstructed from manually annotated ABBs?"

To find out how the manually annotated ABBs differ from the previously evaluated AOBBs, an analysis is performed on the difference between the two types of bounding boxes for vehicles observed from aerial perspective. A comparison between the two bounding boxes illustrates an ABB-AOBB difference consisting of two factors. The first factor comprises the 'rounded-corner effect' introduced due by the fact that vehicles observed from top-down are shaped as rounded rectangles. The rounded corners of these vehicles cause manually annotated ABBs to be smaller than corresponding AOBBs. The second factor can not be identified based on this analysis, although the shape of the errors points a tendency for humans to annotate vehicles larger than their actual perceived size.

Subsequently, A2OBBs performance to reconstruct OBBs from manually annotated ABBs is determined by applying the proposed method to the set of manually annotated ABBs. The results show that the ABB-AOBB difference introduces inconsistencies in the set of equations using in the optimization to such an extend that the optimal solution found by A2OBB is much thinner compared to the ground-truth OBBs. As such, the average contour reconstruction, evaluated by the IoU metric, drops significantly. Nonetheless, reconstruction values for the length and orientation of the vehicle are decent. Therefore, one could conclude that for ABB-OBB conversion, the IoU metric does not present a fair evaluation in the context of extracting kinematic parameters of vehicles in aerial imagery.

To improve the reconstruction performance and to attempt to compensate for the ABB-AOBB

difference, a correction factor is added based on the 'rounded-corner effect' formulated for ellipse-shaped corners and an additional constant. The parameters used to calculate these correction factors are determined based on the ABB-AOBB difference observed in the dataset. As a result of adding these correction factors, the reconstruction performance improves significantly.

Altogether, using the correction factor calculated based on the observed ABB-AOBB difference for 50% of all vehicles, the majority of OBBs can be reconstructed from manually annotated ABBs with an average accuracy of length and width within 5 and 15% of the original ground-truth respectively, an orientation with an overall average error of below 2.5° and an average IoU of 85%. In addition, specifying a correction factor per individual vehicle would improve the results even more.

4-5 Experiment 4: Reconstruction using ABBs from a detection network

This section describes the setup and results for the last experiment, focussing on assessing A2OBB's reconstruction performance as an extension to existing ABB detection networks. This experiment also comprises the second experiment focussing on analysing the ABB-AOBB difference, in this case for ABB detections. This experiment focusses on answering the question:

"To what extent can OBBs be reconstructed from ABBs created by a detection network?"

ABB detections essentially are reconstructions of manually annotated ABBs, constructed by an object detection network. Therefore, detected ABBs will contain the same differences with respect to AOBBs as manually annotated ABBs. On top of those difference, the object detection network introduces detection errors. Since the ABB-AOBB differences for manually annotated ABBs have been analysed in the previous experiment, this section will focus mainly on the error introduced by the detection network.

An answer to this question is formulated in four steps. The first step analyses the detection error for the in-house trained YOLOv3 network. Subsequently, a second step analyses how this detection error is reflected in the total ABB-AOBB difference for detected ABBs. Thirdly, the A2OBB reconstruction performance on detected ABBs is quantified, again focussing on the impact of the detection error. Finally, a fourth step assesses how this performance can improve using the correction factor introduced in section 3-3-2.

4-5-1 Detection errors

The shape and magnitude of detection errors for neural networks depend very much on the type of detection network that is used. Therefore, this analysis will not be applicable to all detection networks. Nonetheless, performing this analysis will provide a basic indication of how A2OBB performs as an extension to a current state-of-the-art network.

For the purpose of this analysis, a version of the YOLOv3 network [26] is trained. The procedure and data used are described in section 4-1-1. As explained in section 4-1-3, the metrics used for this evaluation differ from the conventional methods. Detections are already provided. Therefore, instead of focussing on metrics such as Average Precision (AP) or Average Recall (AR) evaluating the overall detection accuracy, the metrics assess the accuracy of the dimensions of detected bounding box: length and width.

The detection error is found by comparing the detected bounding boxes to their corresponding ground-truth, in the form of the manually annotated ABBs. The errors in both dimensions are displayed in figure 4-21.

Figure 4-21 illustrates that the detected bounding boxes differ from manually annotated boxes in two ways. Firstly, the length of the bounding boxes is consistently larger, with an average of 6 pixels, whereas the width is detected with a negative bias of 2.5 pixels. Thus, vehicles are detected to be thinner, but longer compared to the annotated bounding boxes. Secondly, the detector introduces a significant amount of variance in both dimensions.

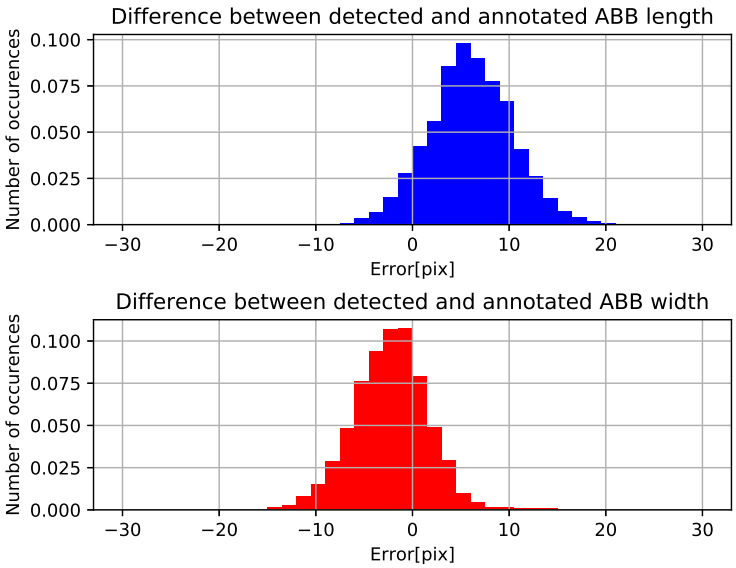


Figure 4-21: A figure illustrating the detection error of an in-house trained YOLOv3 network.

As illustrated in section 4-4-1, the ABB-AOBB difference observed for manually annotated bounding boxes depends on the orientation of the vehicle. To see whether a similar dependency exists for the detection error, the difference in dimensions is also plotted as a function of the orientation of the OBB ground-truth in figure 4-22. Observing this figure illustrates that no clear dependency on the orientation can be identified. The variance appears to be roughly the same for all orientations in both the length and the width.

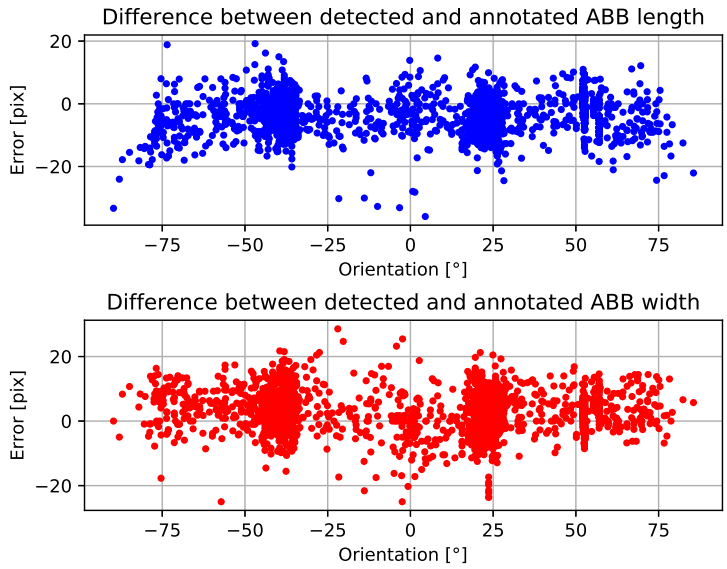


Figure 4-22: A figure illustrating the detection error of an in-house trained YOLOv3 network as a function of the orientation of the underlying vehicle.

4-5-2 ABB-AOBB difference for detections

The previous section illustrates that detection errors for the in-house trained YOLOv3 network introduce a bias and additional variance in both dimensions compared to manually annotated bounding boxes. To see how these errors are reflected in the total ABB-AOBB difference for detected ABBs, the error between the dimensions of the ABB detections and corresponding AOBBs is depicted in figure 4-23 as a function of the orientation.

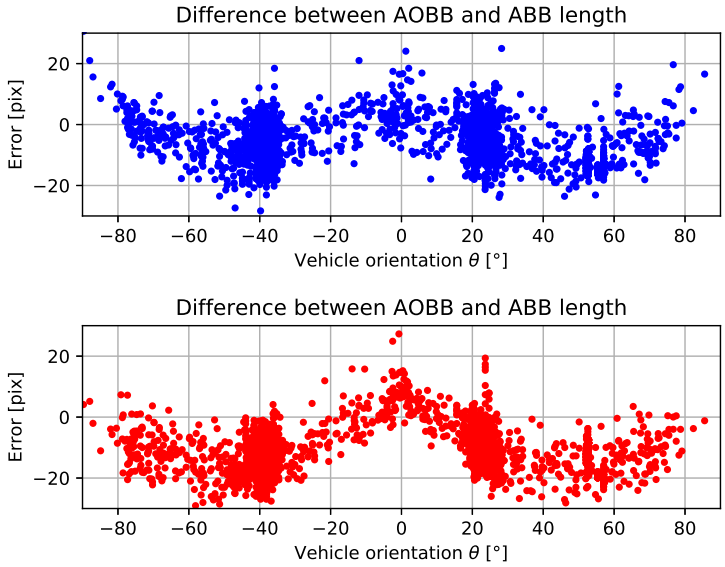


Figure 4-23: A figure illustrating the total ABB-AOBB difference for bounding boxes detected by the YOLOv3 network.

Comparing these errors to the results depicted in figure 4-14 illustrates a similar shape. Again, the detected boxes become increasingly smaller compared to the AOBBs as the orientation approaches $\pm 45°$ and increases for orientations aligned with the image axes. Note however, that the influence of the bias and variance introduced by the detection error can also be identified. The error in length is slightly more positive, whereas the difference width is slightly more negative. In addition, the variance in the error for vehicles at similar orientations is higher.

4-5-3 Reconstruction performance for detected ABBs

The detection error introduces a positive bias for the length and a negative bias for the width of ABBs compared to AOBBs and introduces additional variance. To see how both factors affect the reconstruction performance, A2OBB is applied to the set of detections resulting from the trained YOLOv3 network. The detections in individual frames are matched based on the displacement of the bounding boxes, described in section 4-1-1.

The reconstructions based on the YOLOv3 detections are visualized in figure 4-24. To illustrate similarities and differences with the reconstructions based on manually annotated ABBs, the same frame is depicted as in figure 4-16.

Similarly to the results based on the manually annotated ABBs, most of the reconstructions based on detected ABBs the value for the width is inaccurate. Again, this points to the fact that the difference between the detected ABBs and generated AOBBs introduces inconsistencies to such an extent that the optimal vehicle shape is much thinner compared to the ground-truth. Comparing the reconstructions to the results depicted in figure 4-16 illustrates a similar performance, although most detections are reconstructed to be thinner compared to manually annotated ABBs. Note that the accuracy of the bounding box fit differs per individual image. Vehicle reconstructions depicted in this image might fit better or worse compared to the results in other images.

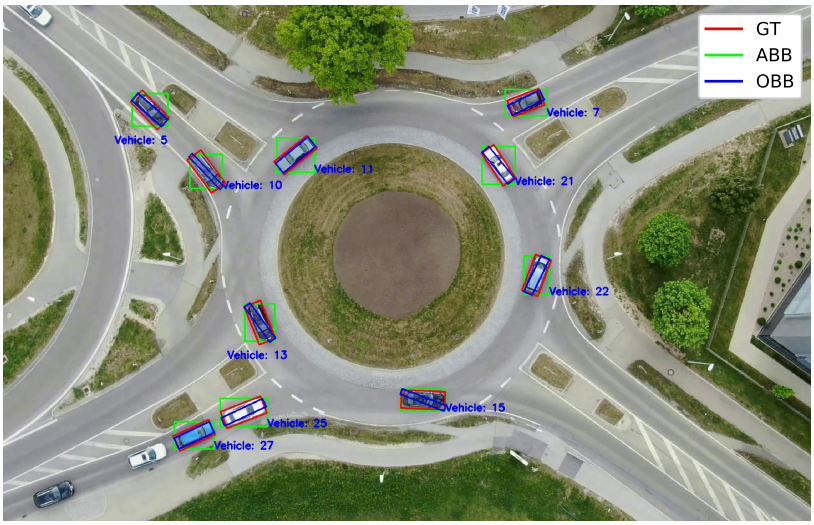


Figure 4-24: An example image of the reconstruction performance based on ABBs obtained from a YOLOv3 detection network. Red boxes depict the ground-truth OBBs, green illustrate the detected ABBs and blue shows the reconstructed OBBs.

To see how this is reflected in the reconstruction metrics, the results are quantified in figure 4-25 and compared to the reconstruction metrics based on manually annotated ABBs, depicted in figure 4-17. A comparison illustrates an increase in variance for the errors in all dimensions. The length is reconstructed with a larger error and the width is reconstructed less accurate. As a result, the orientation estimation is worse and IoUs also drop.

Numerically, the error for the length remains below 10% of the ground-truth value for most vehicle sequences. However, almost all vehicles are reconstructed with an average width error of at least 10% of the ground-truth width, averaging at an error of 40% . Nonetheless, the orientation is reconstructed fairly decently with an average error of most sequences below 5°. On the other hand, the IoU appears to drop significantly to an average of 50%.

To conclude, the ABB-AOBB difference for detected ABBs introduces inconsistencies in the set of equations to such an extent, that the optimal solution to the set of equations results in an OBB that is much thinner compared to the actual observed shape. Similar to the manually annotated ABBs, the overall quality of the reconstructions drops significantly compared to the AOBB reconstructions.

Altogether, without any correction factors, A2OBB's reconstruction performance is too low to be applied as an extension to existing detection networks.

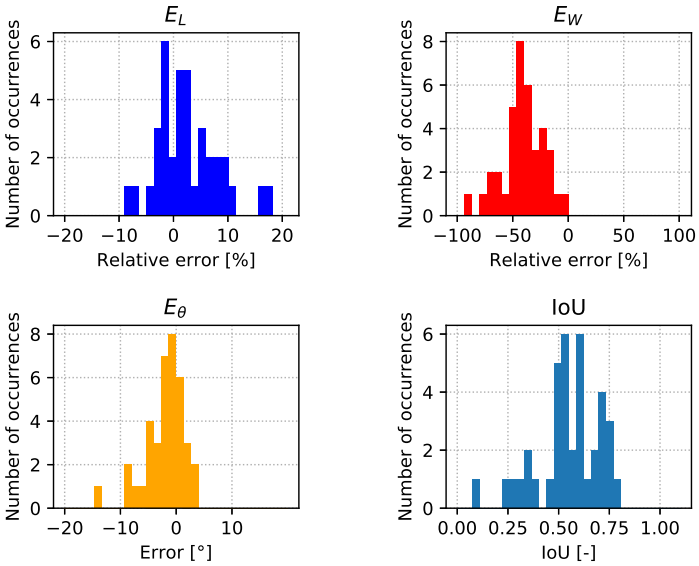


Figure 4-25: Performance metrics for ABBs detected by a YOLOv3 network. The figure displays histograms of the average error in length, width and orientation and the average IoU per vehicle instance.

4-5-4 Compensating the ABB-AOBB difference

Observing the total ABB-AOBB difference for detected ABBs, depicted in figure 4-5-2, a shape can be identified similar to the difference for manually annotated ABBs. In the previous section, this shape pointed to an influence of the 'rounded-corner effect' for ellipse shaped corners, which could be compensated by using the correction factors calculated using equations 3-12. The similarity between both shapes suggests that the reconstructions for detected ABBs can also improve when introducing similar correction factors. To test whether this is true, the same correction factors are applied, calculated using the parameters depicted in table 4-2.

The resulting reconstructions are visualized in figure 4-26. Comparing these reconstructions to the results without correction depicted in figure 4-24 again illustrates significant improvements. The majority of the vehicles is reconstructed with a width comparable to the ground-truth. As a result, the reconstruction for the length and the orientation also improve. Interestingly, a portion of the vehicles (e.g. vehicles 10, 13, 15 and 25) is reconstructed longer and thinner compared to the ground-truth. For these vehicles, the ABB-AOBB difference is not fully compensated due to the fact that the correction factors are generalized over multiple vehicles and do not account for the detection errors. On the other hand, for some vehicles (e.g. vehicle 7, 11, 22 and 27), the correction factors appear to work decently, resulting in fairly accurate reconstructions.

Reconstructions being thinner but longer compared to the ground-truth can also be identified

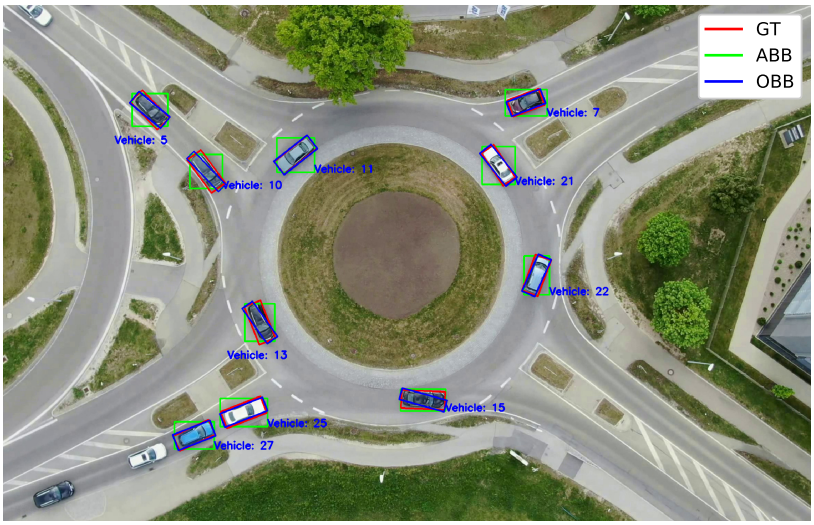


Figure 4-26: An example image of the reconstruction performance based on ABBs obtained from a YOLOv3 detection network. Red boxes depict the ground-truth OBBs, green illustrate the detected ABBs and blue shows the reconstructed OBBs.

when observing the performance metrics for these reconstructions, depicted in figure 4-27. To compare, the figure also depicts the reconstructions without corrections in shelled histograms. The figure illustrates a slight drop in the reconstruction performance of the length. Most of the reconstructed lengths are larger compared to the ground-truth and the magnitude of the errors has slightly increased, introducing a average bias of +4.5% of the ground-truth value. Nonetheless, comparing the results of the remaining performance metrics, illustrates several improvements. The width is reconstructed more accurately, with most vehicles reconstructed within 20% of the ground truth value and the negative bias is significantly reduced to 10%. As a consequence, the orientation estimation also improves slightly with the majority of the errors below 7°. However, the orientation estimation does exhibit a slight negative bias, averaging at -2.5°. Overall, the contour reconstruction performance improves significantly to an average IoU of 77% over all vehicles.

Altogether, the introduction of the correction factors in the optimization indeed improves the reconstruction performance significantly. However, the impact of the improvement is not as much as for manually annotated ABBs due to the variance and bias introduced by the detection error.

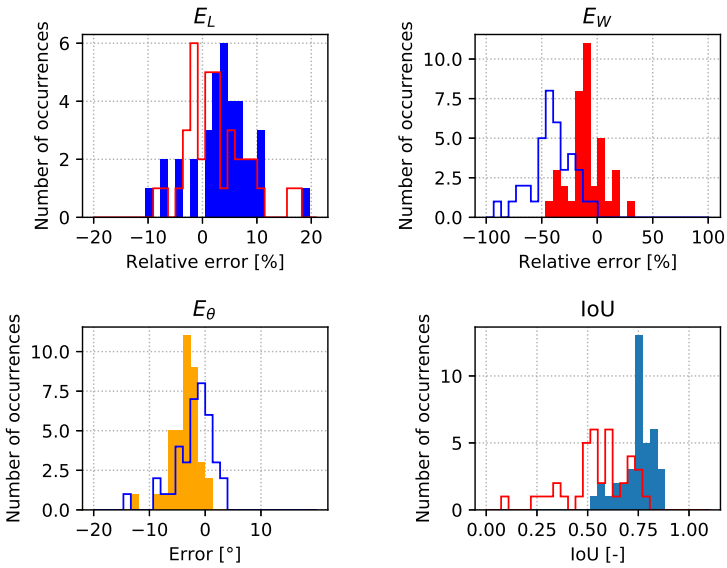


Figure 4-27: Performance metrics for ABBs detected by a YOLOv3 network, corrected for the 'rounded-corner effect'. The figure displays histograms of the average error in length, width and orientation and the average IoU per vehicle instance. The shelled histograms depict the metrics without correction factors.

4-5-5 Conclusion

To summarize, the main focus of this experiment was to answer the research question:

"To what extent can OBBs be reconstructed from ABBs created by a detection network?"

Detected ABBs are reconstructions of manually annotated ABBs and will therefore contain the same ABB-AOBB differences. On top of that, the detection network introduces detection errors. For the in-house trained YOLOv3 network, the detection errors introduce a positive bias for the length and a negative bias for the width, along with additional variance in both dimensions. These biases and the added variance can both be identified in the total ABB-AOBB difference for detected ABBs. Nonetheless, the shape of the ABB-AOBB difference is similar to the difference observed for manually annotated ABBs.

Subsequently, A2OBBs performance to reconstruct OBBs from detected annotated ABBs is determined by applying the proposed method to the set of detected ABBs. The results again show that the ABB-AOBB difference introduces inconsistencies in the set of equations using in the optimization to such an extent that the optimal solution found by A2OBB is much thinner compared to the ground-truth OBBs. As a result, the quality of the reconstructions drops significantly. Essentially, without additional improvements, OBB reconstructions using based on detected ABBs resulting from this trained version of YOLOv3 are too poor to be useful further on.

As the ABB-AOBB difference for detected ABBs appears similar to the difference for manually annotated ABBs, the two correction factors compensating for the 'rounded-corner effect' introduced in the previous experiment are also evaluated to improve the methods performance,

again calculated from the ABB-AOBB difference as a function of the orientation. The error in length has a positive bias of +4.5%, whereas the width has a mean average error of 10%. The majority of the vehicles is reconstructed with an orientation error below 5°. Note however, that all reconstructions have an average orientation error 2.5°. Overall, the contour reconstruction performance improves significantly to an average IoU of 77% for all vehicles.

Chapter 5

Conclusions and Recommendations

This thesis proposed the novel A2OBB method, designed to reconstruct a set of Oriented Bounding Boxes (OBBs) from a set of Axis-aligned Bounding Boxes (ABBs) that describe multiple observations of a vehicle in low-altitude aerial imagery. This method is intended as an enhancement tool for datasets and vehicle detection methods used in a vehicle tracking pipeline.

Based on a non-linear least squares approximation of the geometrical relationship between all ABBs and OBBs defined for rectangular objects, A2OBB finds an optimal set of OBB dimensions that fits best in all provided ABB.

A2OBB's approach assumes that vehicles observed from low-altitude aerial perspective are perceived as rectangles and that their size remains constant for all observations. In practice, both assumptions will not hold, affecting the optimality of the reconstructions. To find out how accurately OBBs can be reconstructed using A2OBB, the objective of this thesis was to answer the following research question:

"Can the geometrical relationship between OBBs and ABBs be used to reconstruct OBBs from a set of ABBs describing a vehicle in multiple orientations in one scene captured from a low-altitude aerial perspective?"

An answer to the main research question is formulated by answering four sub questions, each covered in a separate experiment. This chapter briefly summarizes the conclusions of each of the experiments reflects upon their results and formulates an answer to their sub questions. Subsequently, the discussion section reflects on the scope of the thesis and the results in general and touches upon the method's limitations. Lastly, the thesis is concluded by making recommendations for further research topics.

5-1 Findings

This section briefly repeats the research question, summarizes the findings for each experiment and provides the answer to the question in bold.

- Experiment 1: "Can OBBs be reconstructed from ABBs that describe multiple observations of a rectangular vehicle with constant dimensions?"
  - The non-linear least squares problem that describes the geometrical relationship between ABBs and OBBs is solvable for ABBs under ideal conditions;
  - **OBBs can be reconstructed perfectly from ABBs using A2OBB for rectangular objects with a constant size.**
- Experiment 2: "To what extent can OBBs with varying dimensions be reconstructed from a set of ABBs generated around OBBs using geometry?"
  - The perceived size of vehicles in low-altitude aerial imagery changes as the vehicle moves through the scene. The magnitude of this change depends on the shape of the vehicle and the vehicle's change in displacement from the image centre;
  - The impact of a changing size on the A2OBB's reconstruction performance can be minimized by considering observations that are equidistant to the image centre;
  - **OBBs describing an object with varying dimensions can be reconstructed from ABBs with an average IoU of 95%. The length and width of OBB are reconstructed with maximum error of 1% and 7% of the ground-truth respectively, the orientation can be reconstructed with a maximum error of within 2°.**
- Experiment 3: "To what extent can OBBs be reconstructed from manually annotated ABBs?"
  - Vehicles observed from low-altitude aerial perspective are perceived as rounded rectangles instead of regular rectangles. As a consequence, ABBs around vehicles are smaller than ABBs that generated around the desired OBBs (AOBBs). This effect is referred to as the 'rounded-corner effect';
  - Manually annotated ABBs differ from AOBBs as a consequence of this 'rounded-corner effect' and due to annotation errors. This ABB-AOBB difference reduces the optimality of the OBB reconstructions;
  - To compensate this ABB-AOBB difference, a correction factor is formulated based on the 'rounded-corner effect' for elliptically shaped corners. Applying this compensation factor significantly improves A2OBB's reconstruction performance;
  - **After applying a correction for the 'rounded-corner effect', OBBs can be reconstructed from manually annotated ABBs with an average IoU of 85%. The length and width are accurate within 5% and 15% of the ground-truth and the orientation is reconstructed with an overall average error of below 2.5°.**
  - IoU metric does not present a fair evaluation in the context of extracting kinematic parameters of vehicles in aerial imagery.

- Experiment 4: "To what extent can OBBs be reconstructed from ABBs created by a detection network?"
  - On top of the 'rounded-corner effect, ABBs created by a detection network suffer from the detection error introduced by the network. The magnitude and shape of this detection error depend on the detection network;
  - To obtain an indication of A2OBBs performance on a state-of-the-art detection network, an out-of-the-box YOLOv3 network is trained on in-house collected data. The detection error of the YOLOv3 network leads to a positive bias in the length of ABB detections and a slight negative bias for its width;
  - The correction factor to compensate the 'rounded-corner effect' also improves the reconstruction results for detected ABBs;
  - **After applying a correction for the 'rounded-corner effect', OBBs can be reconstructed from detected ABBs with an average IoU of 77%. The length and width are reconstructed with a bias of 4.5% and 10% of the ground-truth and the orientation is reconstructed with an overall average error of below 5° but above 2.5°.**

5-2 Discussion

This section reflects on the scope of the thesis, A2OBB's evaluation and performance in general and on its application.

5-2-1 Scope of the thesis

The scope of this thesis is set to a point where a set of detected or manually annotated ABBs is available. This set of ABBs describe several observations of a vehicle in imagery that is assumed to be pre-processed, compensating lens distortions and stabilizing the frames. In practice, not all imagery might be processed in a similar fashion. Without these preprocessing steps, shapes in the imagery will be distorted reducing the optimality of A2OBB's reconstructions.

The goal of this thesis was investigate the hypothesis that the geometrical relationship between ABBs and OBBs for rectangular objects can be used to reconstruct OBBs from ABBs. The focus in selecting a least squares solver lay on proving the hypothesis and providing a method that is easy to implement, rather than on selecting the best suited solver. The results of the experiment illustrate that the chosen solver works. However, other solvers might be more suitable.

In addition, the scope of thesis is set to cover the reconstruction of the OBB dimensions and, to a certain extent, their orientation. As explained in section 3-3-4, A2OBB is restricted to perform an orientation estimation over the interval $[0, \pi/2]$. The resulting orientation is one of four possible solutions. During the evaluation, the orientation that results in the highest IoU with the ground-truth is selected, depicting the most ideal result. Finding the correct orientation from the four possibilities is not included in the scope of this thesis.

5-2-2 A2OBB's evaluation and performance

The experiments performed in this thesis provide a well-structured analysis of A2OBB's reconstruction performance on annotated and detected ABBs. The analysis starts from an ideal scenario, working its way to real-world datasets by releasing one assumption at a time. This provides a clear visualisation of the impact of deviations to each assumption.

The datasets used for evaluation A2OBB's reconstruction performance are fairly representative to the methods intended applications. However, experiments 2 (section 4-3) and 3 (section 4-4) do illustrate a significant impact of errors made while annotating the imagery. These annotation errors are caused by human error, but might also be caused by the use of a different annotation tool for the different bounding box formats. These annotation errors affect not only the ABBs used as an input to the experiments, but also the OBBs used as ground-truth used during the evaluation. As such, the results presented in this thesis paint a slightly distorted picture.

Nonetheless, the annotation errors do not distort the results too much and still allow us to draw relevant conclusions. Furthermore, as argued in section 3-2-4, at this point in time no automated annotation tool will yield a better evaluation.

Interestingly, experiment 3 illustrates a large impact of the 'rounded-corner effect', inducing an ABB-AOBB difference for vehicles observed from aerial perspective. Manually annotated ABBs differ from ABBs generated around OBBs by the fact that vehicles are observed as rounded- instead of regular rectangles. The implications of this effect are two-fold. The first implication is addressed in this thesis (section 3-2-3. The equations used in the least squares approximation are not completely accurate. As a result, the reconstructed ABBs will also be inaccurate. The 'rounded-corner effect' can be compensated to a decent extend using the proposed correction factor.

More importantly, the presence of the 'rounded-corner effect' means that for non-rectangular objects, ABBs generated around annotated OBBs are not suitable for evaluating ABB detections. The evaluation method used current publicly available datasets (e.g. DOTA [12]) to evaluate ABB detections based on an OBB dataset does not provide an accurate measure of performance.

In general, A2OBB's reconstruction performance for manually annotated ABBs appears to be accurate. For detected ABBs, the performance is worse, although the reconstructed orientation is fairly accurate. Note that A2OBB's performance on detected ABBs heavily depends on the detection error introduced by the network. If the accuracy of the detection network improves, A2OBB's reconstructions will improve as well.

Unfortunately, comparing the results of this thesis to the state-of-the-art proves to be difficult. Object detection networks are commonly evaluated using methods that assess correctness of detection rather than observing their accuracy. In addition, converting annotations from ABB to OBB is, to the authors knowledge, unprecedented and therefore no comparison is possible. To still allow for an indication of performance, the IoU metric is used. However, as concluded the results in experiment 3 4-4, this metric does not present a fair evaluation in the context of extracting kinematic parameters of vehicles in aerial imagery.

5-2-3 A2OBB's applicability

A2OBB is intended as a data and detection enhancement tool in vehicle tracking pipelines. A2OBB is only applicable to video sequences captured from an aerial perspective in which the size of objects remain relatively constant. This scenario can be created using low-altitude aerial imagery, captured by sUAVs/drones. sUAVs/drones can remain at a relatively fixed position, keeping its perspective with respect to the scene fairly constant.

The in-house collected dataset used to evaluate A2OBB's performance covers one specific scene. Although only one scene is covered, it is believed that the results are also representative of other scenarios that cover a similar scenario. As long as the camera remains relatively fixed and lens distortions are compensated, the perspective and shape of vehicles should be similar, and thus A2OBBs performance should be similar as well.

Note that A2OBB is not applicable to low-altitude aerial imagery where the camera's perspective moves significantly between subsequent frames. For high-altitude imagery (e.g. satellite imagery), the perspective between camera and scene changes significantly between consecutive frames (illustrated in figure 2-2). As a result the perceived shape of objects also changes significantly. Thus, A2OBB is also not suitable for high-altitude aerial imagery.

5-3 Further research

To conclude the thesis, this section provides a short description of the topics that are interesting for further research. In general, the following topics could be of interest:

- **Choosing the correct orientation.** First and foremost, the main limitation to A2OBB is the fact that the orientations of the resulting OBB all comprise one of four possible solutions. Selecting the correct version is not included in this thesis, but some recommendations are made in section 3-3-4.
- **Using alternate shapes.** The proposed methodology employs equations that describe the ABB-OBB relationship for rectangular objects. It might be interesting to investigate the use of different shapes or even non-parametric shapes as a basis for the methodology. Using different shapes might allow this methodology to be applicable to alternately shaped objects as well;
- **Including the change in perspective.** As illustrated in the second experiment of this thesis, the change in perspective as a vehicle moves through the scene causes a change in the perceived size of objects. Modelling this change could be done based on the shape of the vehicle and the vehicle's distance and orientation with respect to the image centre. Including this model in the method's calculations could improve the reconstruction results.
- **Tailoring the correction factor for the 'rounded-corner effect' to specific vehicle types.** Experiment 3 illustrates a significant impact of the 'rounded-corner effect' to which a compensation is introduced based on a model this effect for ellipse-shaped corners. The parameters of this correction factor are generalised for all types of vehicles. However, the 'rounded-corner effect' is stronger for some type of vehicles than

others. As such, customize the correction factor per vehicle type is expected to improve the reconstruction performance.

- **Determining optimal number of observations.** In this thesis, the amount of computational effort required to find a solution was deemed to be of lesser importance and no elaborate analysis is made on the number of observations required. However, as the number of observations considered for the reconstruction increases, so does the accuracy of the reconstruction, but also the computational effort required to find a solution. Therefore, a trade-off exists in the desired accuracy and the required computational effort with an optimum in the middle.

Appendix A

Implementation of the video stabilization filter

This appendix provides a short description of the video stabilization filter used in this thesis to compensate drone imagery for long term drift. The filter comprises a feature-based motion estimation between consecutive imagery based on the procedure described in [39]. The method proposed in this paper strikes a good balance on computational efficiency and robustness [39]. Its pipeline is depicted in figure A-1.

The procedure in [39] attempts to capture the global motion of the drone based on the movement of feature points detected in the imagery. First, ORB-features [43] are extracted for every individual image and matched between subsequent frames. The matches between frames are used to calculate inter-image geometric transformations and these are combined to describe the global motion. Lastly, a Gaussian smoothing filter is applied to the global motion vectors between all frames and the transformations are applied to the imagery.

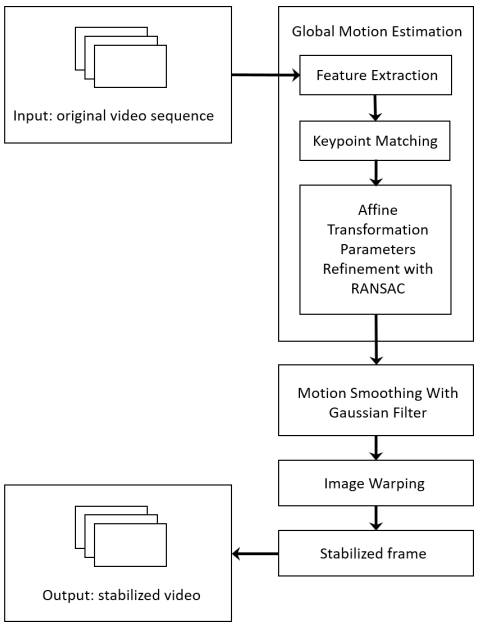


Figure A-1: The pipeline of the video stabilization filter [39].

Note however, that aerial imagery recorded for traffic analysis contains a large number of traffic participants. As depicted in figure A-2, a portion of the detected features will therefore lie on these participants. The position of these traffic participants varies slightly from image to image. As such, matches that are found using these feature points will introduce noise in the inter-image movement calculations.

Figure A-2: An example of ORB-feature points (depicted as purple circles) detected in one of the image in the recorded dataset. Note that some of the features lie on stationary points whereas others lie on moving vehicles.

To filter out these noisy matches and improve the final video stability, the procedure suggested in [39] is slightly altered. This procedure is depicted in figure A-3. Instead of calculating the inter-image movement using matches between subsequent imagery, all images are matched to one specific 'mean image'. This 'mean image' is taken as the mean image of the imagery in the stabilized video sequence resulting from [39]. By taking the average of all frames in the video, an image is created that contains no traffic participants, as depicted in figure A-4.

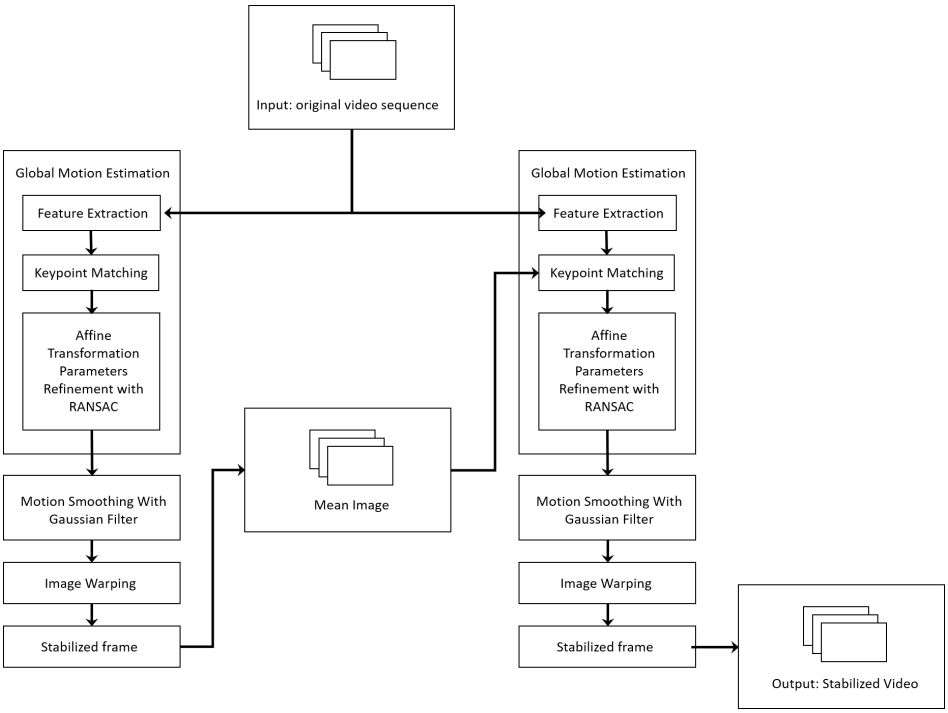


Figure A-3: The pipeline of the video stabilization filter used in this thesis.

After calculating the 'mean image', its ORB-features are extracted and matched with the feature points in each of the images in the sequence. The resulting matches are used to calculate the motion of each individual image with respect to the mean image. Subsequently, the motion vectors are again passed through a gaussian filter and applied to the imagery. The resulting set of images is stabilized and compensated for long-term drift.

Figure A-4: A visualisation of the calculated mean image.

Appendix B

Derivation of the 'rounded-corner effect' for elliptically shaped corners

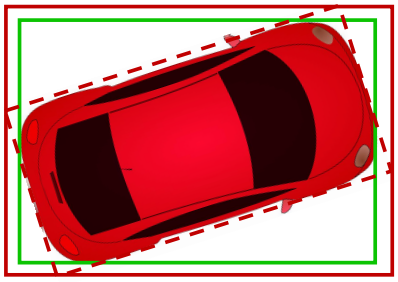
This appendix describes the mathematical derivation of the 'rounded-corner effect' a vehicle with elliptically shaped corners. As mentioned in section 3-2-3, the 'rounded-corner effect' refers to a difference between a manually annotated ABB and an ABB generated around an OBB, referred to as AOBB, illustrated in figure B-1a. The derivation in this section will show how the ABB-AOBB difference depends on the orientation of the vehicle with respect to the image axes.

In this derivation, it is assumed that the vehicle is oriented at an angle θ with respect to the image axes lying on the interval $[0, \frac{\pi}{2}]$. A sketch of the situation is depicted in figure B-1b. The green boxes represents the annotated ABB, the red-dotted boxes illustrate the OBB and the solid-red boxes show the AOBB, generated around the OBB. The blue rounded rectangle represents the contours of the vehicle.

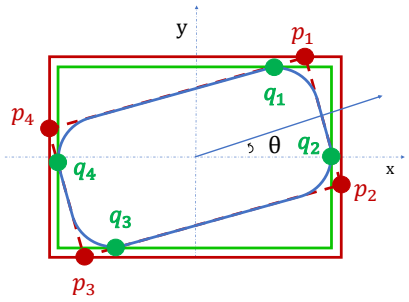
Figure B-1b suggests that the ABB-AOBB difference is determined by two sets of points. The first set of points consists of the corner points of the OBB, denoted as p_i. The second set comprises the points on the ellipse-shaped corners of the vehicle where the slope of the tangent line is parallel to the main axis of the image. These points are denoted as q_i. The dimensions of the ABB and AOBB terms of the axis of the image (x, y) will be denoted as L_A and W_A. Essentially, the difference between the dimensions of the ABB and AOBB can be calculated as:

$$\Delta L_{A_x} = |p_{1_x} - q_{1_x}| + |p_{3_x} - q_{3_x}|, \qquad \Delta W_{A_y} = |p_{2_y} - q_{2_y}| + |p_{4_y} - q_{4_y}| \tag{B-1}$$

In this derivation, it is assumed that the corners of the back of the vehicle are shaped the same as the corners of the front of the vehicle. As a consequence, the differences between sets p_3-q_3 and p_4-q_4 will be the same as the differences between points p_1-q_1 and p_2-q_2. Therefore, equations B-1 can be rewritten to:



(a) A visual example.



(b) A schematic view.

Figure B-1: A visual and schematic illustration of the 'rounded-corner effect' for a vehicle in an orientation θ, depicting the discrepancy between annotated ABBs (green box) and the AOBBs (solid-red box), generated around the OBB (red-dotted box). In the right figure, points p_i represent the corners of the OBB, points q_i illustrate the point on the ellipse-shaped corners where the tangent lines are parallel to the image axis and the blue rounded-rectangle depicts the contours of the vehicle.

$$\Delta L_{A_x} = 2|p_{1_x} - q_{1_x}|, \qquad \Delta W_{A_y} = 2|p_{2_y} - q_{2_y}| \tag{B-2}$$

Thus, the goal of this derivation is to express the displacement between points p_1-q_1 and p_2-q_2 in a coordinate system aligned with the axes of the image (x, y). To find these displacements between p_1-q_1 and p_2-q_2, the problem is first viewed in the coordinate system aligned with the axes of the vehicle, denoted as (x', y'). Subsequently, the results are converted to the coordinate system of the image by means of a rotation matrix.

ABB-AOBB displacement with respect to the vehicle axis

The first step is to determine the displacement between both points in terms of x' and y':

$$\begin{aligned} \Delta L_{A_{x'}} &= 2|p_{1_{x'}} - q_{1_{x'}}|, \quad \Delta L_{A_{y'}} = 2|p_{1_{y'}} - q_{1_{y'}}|, \\ \Delta W_{A_{x'}} &= 2|p_{2_{x'}} - q_{2_{x'}}|, \quad \Delta W_{A_{y'}} = 2|p_{2_{y'}} - q_{2_{y'}}| \end{aligned} \tag{B-3}$$

To obtain these displacement, it is beneficial to determine the location of all four points with respect to the main axes of the vehicle, denoted as (x', y). For this purpose, a more detailed view of the difference between the ABBs and the AOBBs is illustrated in figures B-2. Both figures depict a sketch of a vehicle in a coordinate system aligned with the main axes of the vehicle, . Figure B-2a is zoomed in on the front of the vehicle, whereas figure B-2b shows a more detailed view of the top right corner of the vehicle. Again, the green lines depict the sides of the ABB, the blue lines illustrate the vehicle contours and the red-dotted lines show

the sides of the OBB. Additionally, the figures include the radii that describe the elliptical shape of the corners, denoted as R_L and R_W. The slope of the green tangent lines with respect to the vehicle's main axes, that define the location of q_1 and q_2, are denoted as θ_1 and θ_2. Both angles can be determined using the vehicle's orientation θ, where:

$$\theta_1 = -\theta, \qquad \theta_2 = \pi - \theta \tag{B-4}$$

The location of all points can be expressed with respect to the centre of curvature of each corresponding corner. To illustrate, for points p_1 and q_1 this point is denoted as O in figure B-2b. As a result, the location of point p_1 can be expressed as:

$$p_{1_{x'}} = R_L, \qquad p_{1_{y'}} = R_W \tag{B-5}$$

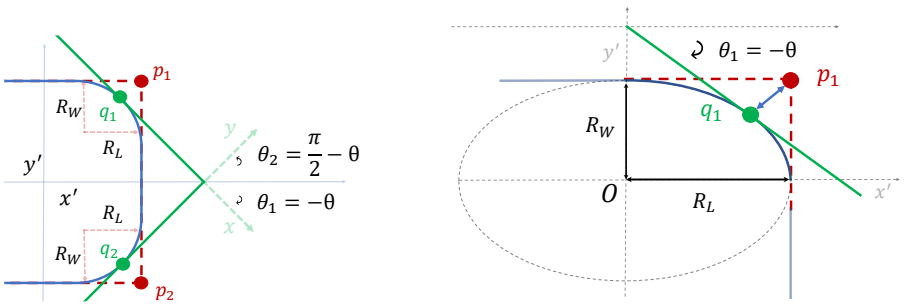
Similarly, the location of point p_2 can be expressed with respect to the centre of curvature for the lower right corner as:

$$p_{2_{x'}} = R_L, \qquad p_{2_{y'}} = -R_W \tag{B-6}$$

Points q_1 and q_2 must to lie on an ellipse defined by radii R_L and R_W. For an ellipse with radii R_L and R_W, all points on this ellipse must lie on a curve described by:

$$\frac{x'^2}{R_L^2} + \frac{y'^2}{R_W^2} = 1 \tag{B-7}$$

Based on this formula, the x'-coordinate of every point on the ellipse can be expressed in terms of the corresponding y'-coordinate and vice-versa:



(a) A detailed view of the front of the vehicle. **(b)** A detailed view of one corner of the vehicle.

Figure B-2: The 'rounded-corner effect' depicted in the coordinate system of the vehicle (x', y'). The corner points of the AOBB are depicted as p_1 and p_2, whereas the manually annotated corner points are denoted by q_1 and q_2. R_L and R_W represent the radii that define the elliptical shape of the corners. θ represents the orientation of the OBB.

$$y' = R_W\sqrt{1-\frac{x'^2}{R_L^2}}, \qquad x' = R_L\sqrt{1-\frac{y'^2}{R_W^2}} \tag{B-8}$$

Additionally, the derivative of function B-7 with respect to x', defined as:

$$\frac{dy'}{dx'} = -\frac{R_W^2 x'}{R_L^2 y'} \tag{B-9}$$

will describe the slope of the tangent to every point on this ellipse. Using equations B-8, this derivative can be expressed as a function of x' and as a function of y':

$$\frac{dy'}{dx'} = -\frac{-R_W x'}{R_L^2\sqrt{1-\frac{x'^2}{R_L^2}}}, \qquad \frac{dy'}{dx'} = -\frac{R_W^2\sqrt{1-\frac{y'^2}{R_W^2}}}{R_L y'} \tag{B-10}$$

Based on figures B-2, it can be shown that the slope of the ellipse for q_1 must be:

$$q_1 : \frac{dy'}{dx'} = \tan(-\theta) \tag{B-11}$$

Using equations B-10, this slope can be expressed in terms of $q_{1'_x}$ and $q_{1'_y}$:

$$\tan(-\theta) = -\frac{-R_W q_{1_{x'}}}{R_L^2\sqrt{1-\frac{q_{1_{x'}}^2}{R_L^2}}}, \qquad \tan(-\theta) = -\frac{R_W^2\sqrt{1-\frac{q_{1_{y'}}^2}{R_W^2}}}{R_L q_{1_{y'}}} \tag{B-12}$$

These equations can subsequently be rewritten to express the location of q_1 as a function of the vehicle's orientation θ:

$$q_{1_{x'}} = \pm\sqrt{\frac{R_L^4\tan^2\theta}{R_W^2+R_L^2\tan^2\theta}}, \qquad q_{1_{y'}} = \pm\sqrt{\frac{R_W^4}{R_W^2+R_L^2\tan^2\theta}} \tag{B-13}$$

Following a similar procedure, for point q_2, the slope of the tangent to the ellipse can be defined as:

$$q_2 : \frac{dy'}{dx'} = \tan(\frac{\pi}{2}-\theta) = \cot(-\theta) \tag{B-14}$$

and location of q_2 can be expressed as:

$$q_{2_{x'}} = \pm\sqrt{\frac{R_L^4}{R_L^2+R_W^2\tan^2\theta}}, \qquad q_{2_{y'}} = \pm\sqrt{\frac{R_W^4\tan^2\theta}{R_L^2+R_W^2\tan^2\theta}} \tag{B-15}$$

Based on figure B-2b, points $q_{1'_x}$, $q_{1'_y}$ and $q_{2'_x}$ must be larger than zero, but $q_{2'_y}$ should be below zero. This information can subsequently be filled into equations B-3 combined with the locations of p_1 and p_2 described by equations B-5 and B-6:

$$\begin{aligned} \Delta L_{A_{x'}} &= 2\left|R_L - \sqrt{\frac{R_L^4 \tan^2\theta}{R_W^2 + R_L^2 \tan^2\theta}}\right|, \qquad \Delta L_{A_{y'}} = 2\left|R_W - \sqrt{\frac{R_W^4}{R_W^2 + R_L^2 \tan^2\theta}}\right|, \\ \Delta W_{A_{x'}} &= 2\left|R_L - \sqrt{\frac{R_L^4}{R_L^2 + R_W^2 \tan^2\theta}}\right|, \qquad \Delta W_{A_{y'}} = 2\left|-R_W + \sqrt{\frac{R_W^4 \tan^2\theta}{R_L^2 + R_W^2 \tan^2\theta}}\right| \end{aligned} \tag{B-16}$$

Subsequently, these equations can be reduced further to the form:

$$\begin{aligned} \Delta L_{A_{x'}} &= 2\left|\sqrt{\frac{R_L^2 R_W^2}{R_W^2 + R_L^2 \tan^2\theta}}\right|, \qquad \Delta L_{A_{y'}} = 2\left|\sqrt{\frac{R_L^2 R_W^2 \tan^2\theta}{R_W^2 + R_L^2 \tan^2\theta}}\right| \\ \Delta W_{A_{x'}} &= 2\left|\sqrt{\frac{R_L^2 R_W^2 \tan^2\theta}{R_W^2 + R_L^2 \tan^2\theta}}\right|, \qquad \Delta W_{A_{y'}} = 2\left|\sqrt{\frac{R_L^2 R_W^2}{R_W^2 + R_L^2 \tan^2\theta}}\right| \end{aligned} \tag{B-17}$$

Converting the ABB-AOBB displacement to the image axes

Equations B-17 provide a description of the ABB-AOBB difference expressed in a coordinate system aligned with the main vehicle axes. The second step of this derivation is to convert these equations to the image axes (x, y). This can be done by applying a rotation matrix:

$$\begin{bmatrix} x \\ y \end{bmatrix} = R \begin{bmatrix} x' \\ y' \end{bmatrix}, \qquad R = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \tag{B-18}$$

where θ represents the angle between the two coordinate systems. Note that this angle is equal to the orientation of the vehicle. In this case, the goal is to calculate the ΔL_{A_x} and ΔW_{A_y}. ΔL_{A_x} can be calculated using the first row of equation B-18, whereas ΔW_{A_y} is calculated with the second row:

$$\begin{aligned} \Delta L_{A_x} &= \cos\theta \Delta L_{A_{x'}} - \sin\theta \Delta L_{A_{y'}} \\ \Delta W_{A_y} &= \sin\theta \Delta W_{A_{x'}} + \cos\theta \Delta W_{A_{y'}} \end{aligned} \tag{B-19}$$

Note that:

$$\Delta L_{A_{x'}} = \Delta W_{A_{y'}}, \qquad \Delta L_{A_{y'}} = \Delta W_{A_{x'}} \tag{B-20}$$

Therefore, equations B-19 can be reduced further to:

$$\begin{bmatrix} \Delta L_{A_x} \\ \Delta W_{A_y} \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \cos\theta & \sin\theta \end{bmatrix} \begin{bmatrix} \Delta L_{A_{x'}} \\ \Delta L_{A_{y'}} \end{bmatrix}, \tag{B-21}$$

where:

$$\Delta L_{A_{x'}} = 2\left|\sqrt{\frac{R_L^2 R_W^2}{R_W^2 + R_L^2 \tan^2\theta}}\right|, \qquad \Delta L_{A_{y'}} = 2\left|\sqrt{\frac{R_L^2 R_W^2 \tan^2(\theta)}{R_W^2 + R_L^2 \tan^2\theta}}\right| \tag{B-22}$$

Summary

To summarize, the 'rounded-corner effect' for vehicles with elliptically shaped corners can be described as a function of the orientation of the vehicle by the following two functions:

$$\begin{bmatrix} \Delta L_{A_x} \\ \Delta W_{A_y} \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \cos\theta & \sin\theta \end{bmatrix} \begin{bmatrix} \Delta L_{A_{x'}} \\ \Delta L_{A_{y'}} \end{bmatrix}, \tag{B-23}$$

where:

$$\Delta L_{A_{x'}} = 2\left|\sqrt{\frac{R_L^2 R_W^2}{R_W^2 + R_L^2 \tan^2\theta}}\right|, \qquad \Delta L_{A_{y'}} = 2\left|\sqrt{\frac{R_L^2 R_W^2 \tan^2\theta}{R_W^2 + R_L^2 \tan^2\theta}}\right| \tag{B-24}$$

In these equations, θ describes the vehicle's orientation and R_L and R_W represent the parameters that define the shape of the corners. Note that these equations describe the effect for vehicles where all corners are equally shaped and the vehicle orientation remains on an interval $[0, \frac{\pi}{2}]$.

Bibliography

[1] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3354–3361, June 2012.

[2] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The Cityscapes Dataset for Semantic Urban Scene Understanding," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3213–3223, 2016.

[3] F. Yu, H. Chen, X. Wang, W. Xian, Y. Chen, F. Liu, V. Madhavan, and T. Darrell, "BDD100K: A Diverse Driving Dataset for Heterogeneous Multitask Learning," *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2633–2642, 2020.

[4] E. N. Barmpounakis, E. I. Vlahogianni, J. C. Golias, and A. Babinec, "How accurate are small drones for measuring microscopic traffic parameters?," *The International Journal of Transportation Research (Transportation Letters)*, pp. 1–9, July 2017.

[5] G. Guido, V. Gallelli, D. Rogano, and A. Vitale, "Evaluating the accuracy of vehicle tracking data obtained from Unmanned Aerial Vehicles," *International Journal of Transportation Science and Technology (IJTST)*, vol. 5, pp. 136–151, Oct. 2016.

[6] J. Apeltauer, A. Babinec, D. Herman, and T. Apeltauer, "Automatic vehicle trajectory extraction for traffic analysis from aerial video data," *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (ISPRS Archives)*, vol. 40, pp. 9–15, 2015.

[7] G. Salvo, L. Caruso, A. Scordo, G. Guido, and A. Vitale, "Traffic data acquirement by unmanned aerial vehicle," *European Journal of Remote Sensing (EuJRS)*, vol. 50, pp. 343–351, Jan. 2017.

[8] DataFromSky, "Flow." https://datafromsky.com, 2020.

[9] J. Redmon, S. Divvala, R. B. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 779–788, 2016.

[10] Q. Li, L. Mou, Q. Xu, Y. Zhang, and X. Zhu, "R3-Net: A Deep Network for Multioriented Vehicle Detection in Aerial Images and Videos," *IEEE Transactions on Geoscience and Remote Sensing (TGRS)*, vol. 57, pp. 5028–5042, 2019.

[11] D. et al., "VisDrone-DET2019: The Vision Meets Drone Object Detection in Image Challenge Results," *IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pp. 213–226, 2019.

[12] G.-S. Xia, X. Bai, J. Ding, Z. Zhu, S. J. Belongie, J. Luo, M. Datcu, M. Pelillo, and L. Zhang, "DOTA: A Large-Scale Dataset for Object Detection in Aerial Images," *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3974–3983, 2018.

[13] J. Sewall, D. Wilkie, P. Merrell, and M. C. Lin, "Continuum Traffic Simulation," *Computer Graphics Forum (CGF)*, vol. 29, pp. 439–448, May 2010.

[14] H. Bi, T. Mao, Z. Wang, and Z. Deng, "A Data-driven Model for Lane-changing in Traffic Simulation," *Symposium on Computer Animation (SPA)*, pp. 149–158, 2016.

[15] S. H. Park, B. Kim, C. M. Kang, C. C. Chung, and J. W. Choi, "Sequence-to-Sequence Prediction of Vehicle Trajectory via LSTM Encoder-Decoder Architecture," in *IEEE Intelligent Vehicles Symposium (IV)*, pp. 1672–1678, 2018.

[16] R. Krajewski, J. Bock, L. Kloeker, and L. Eckstein, "The highD Dataset: A Drone Dataset of Naturalistic Vehicle Trajectories on German Highways for Validation of Highly Automated Driving Systems," in *IEEE International Conference on Intelligent Transportation Systems (ITSC)*, pp. 2118–2125, Nov. 2018.

[17] S. A. Ahmadi and A. Mohammadzadeh, "A simple method for detecting and tracking vehicles and vessels from high resolution spaceborne videos," in *IEEE Joint Urban Remote Sensing Event (JURSE)*, pp. 1–4, Mar. 2017.

[18] Loughborough University Department of Computer Science, Loughborough, UK and Q. Meng, "Self-Learning Vehicle Detection and Tracking from UAVs," *International Journal of Multidisciplinary Educational Research (IJMER)*, 2016.

[19] L. Wang, F. Chen, and H. Yin, "Detecting and tracking vehicles in traffic by unmanned aerial vehicles," *Automation in Construction*, vol. 72, pp. 294–308, Dec. 2016.

[20] E.-J. Kim, H.-C. Park, S.-W. Ham, S.-Y. Kho, and D.-K. Kim, "Extracting Vehicle Trajectories Using Unmanned Aerial Vehicles in Congested Traffic Conditions," *Journal of Advanced Transportation (JAT)*, pp. 1–16, Apr. 2019.

[21] Z. Zou, Z. Shi, Y. Guo, and J. Ye, "Object Detection in 20 Years: A Survey," *arXiv:1905.05055 [cs]*, May 2019.

[22] X. Chen and Q. Meng, "Robust vehicle tracking and detection from UAVs," in *International Conference of Soft Computing and Pattern Recognition (SoCPaR)*, pp. 241–246, Nov. 2015.

[23] A. Ajay, V. Sowmya, and K. P. Soman, "Vehicle Detection in Aerial Imagery using Eigen Features," in *IEEE International Conference on Communication and Signal Processing (ICCSP)*, pp. 1620–1624, Apr. 2017.

[24] T. Moranduzzo and F. Melgani, "A SIFT-SVM method for detecting cars in UAV images," in *IEEE International Geoscience and Remote Sensing Symposium(IGARSS)*, pp. 6868–6871, July 2012.

[25] X. Chen and Q. Meng, "Vehicle Detection from UAVs by Using SIFT with Implicit Shape Model," in *IEEE International Conference on Systems, Man, and Cybernetics(SMC)*, pp. 3139–3144, Oct. 2013.

[26] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," *arXiv:1804.02767 [cs]*, Apr. 2018.

[27] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," *arXiv preprint arXiv:2004.10934 [cs]*, 2020.

[28] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *European Conference on Computer Vision (ECCV)*, pp. 21–37, Springer, 2016.

[29] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 580–587, 2014.

[30] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 37, no. 9, pp. 1904–1916, 2015.

[31] J. Lei, C. Gao, J. Hu, C. Gao, and N. Sang, "Orientation Adaptive YOLOv3 for Object Detection in Remote Sensing Images," in *Chinese Conference on Pattern Recognition and Computer Vision (PRCV)*, pp. 586–597, 2019.

[32] F. Schmidt, "KIT AIS Data Set." http://www.ipf.kit.edu/downloads_data_set_AIS_vehicle_tracking.php. Accessed: 2020-11-16.

[33] S. Razakarivony and F. Jurie, "Vehicle Detection in Aerial Imagery: A small target detection benchmark," *Journal of Visual Communication and Image Representation*, vol. 34, pp. 187–203, Jan. 2016.

[34] H. Yu, G. Li, W. Zhang, Q. Huang, D. Du, Q. Tian, and N. Sebe, "The Unmanned Aerial Vehicle Benchmark: Object Detection, Tracking and Baseline," *International Journal of Computer Vision (IJCV)*, vol. 128, no. 5, pp. 1141–1159, 2020.

[35] I. Bozcan and E. Kayacan, "AU-AIR: A Multi-modal Unmanned Aerial Vehicle Dataset for Low Altitude Traffic Surveillance," *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 8504–8510, 2020.

[36] Z. Zhang, "A Flexible New Technique for Camera Calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 22, pp. 1330–1334, Nov. 2000.

[37] M. A. Branch, T. F. Coleman, and Y. Li, "A Subspace, Interior, and Conjugate Gradient Method for Large-Scale Bound-Constrained Minimization Problems," *SIAM Journal on Scientific Computing (SISC)*, vol. 21, pp. 1–23, Jan. 1999.

[38] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, *et al.*, "Scipy 1.0: fundamental algorithms for scientific computing in python," *Nature methods*, vol. 17, no. 3, pp. 261–272, 2020.

[39] J. Xu, H.-w. Chang, S. Yang, and M. Wang, "Fast Feature-Based Video Stabilization without Accumulative Global Motion Estimation," *IEEE Transactions on Consumer Electronics*, vol. 58, pp. 993–999, Aug. 2012.

[40] J. Cartucho, "Labelimg." https://github.com/Cartucho/OpenLabeling, 2018.

[41] Tzutalin, "Openlabeling." https://github.com/tzutalin/labelImg, 2015.

[42] J. Redmon, "Darknet: Open source neural networks in c." http://pjreddie.com/darknet/, 2013–2016.

[43] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *IEEE International Conference on Computer Vision (ICCV)*, pp. 2564–2571, Nov. 2011.

Glossary

List of Acronyms

A2OBB	Axis-aligned to Oriented Bounding Boxes
ABB	2D Axis-aligned Bounding Box
OBB	2D Oriented Bounding Box
AOBB	Axis-aligned Bounding Boxes generated around Oriented Bounding Boxes
sUAVs	small Unmanned Aerial Vehicles
IoU	Intersection over Union