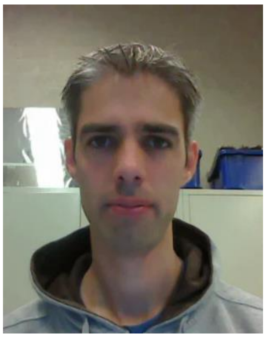
VISTA: VISION SUPPORTED TRUCK DOCKING ASSISTANT

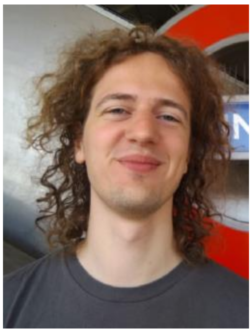
K. Kural, MSc, PhD., Senior Researcher at HAN University of Applied Sciences.



J. van Kolfschoten, MSc. Senior Researcher at HAN University of Applied Sciences.

J. Benders, MSc., Program Manager Control Systems, HAN University of Applied Sciences

K. Essig, Prof., MSc, PhD, professor at Rhine-Waal University of Applied Sciences



Jeroen Delcour, MSc., AI and Computer Vision Engineer. Track 32.

P. Ribeiro, MSc., Research Engineer, Rhine-Waal University of Applied Sciences

D. Devasia, MSc., Research Engineer, HAN University of Applied Sciences

P. Hasse, M.A., Creative Director at CODUCT GmBH

Abstract

The logistics sector is constantly looking for innovations that can help improve the service level and profitability of the freight transport. These innovations have led to reduced fuel consumption and emissions in recent decades, and the automatization inside the warehouses and distribution centers, among other things. Yet, the automatization outside the distribution centers, at the parking areas, does not emerged so far. The docking of the vehicle combination towards the loading dock is still done manually by the drivers alike decades ago, even though the risk the driver will not handle the maneuver properly and in safe way still exists. To improve the performance of the driver a novel support system is proposed in this paper. The framework consists of a computer vision-based system which localizes the vehicle combination with respect to the loading dock, an optimal path planner, and the human machine interface which feeds the driver with the outputs of the tracking controller that ensures the reference path is followed. The paper furthermore reports on the first steps taken to the demonstrator development and integration of subsystems.

Keywords: Vehicle Safety, Intelligent Vehicle System, Driver Support, Articulated Vehicles, Low-speed Maneuvering

1. Introduction

The profitability and safety are essential in the logistics sector. Given the trend of the road freight transport demand in Europe, which can be seen in Figure 1., it may be expected that more vehicles on the roads and in the distribution centers will be needed to satisfy the transport demand in the future. To increase the efficiency, experiments are being carried out nowadays on a large scale with automatization of trucks on various levels.

Many of these tested innovations have in common that additional systems in the vehicle take over the role of the driver at the cost of complex technology which can only be used in new vehicles. Hence it appears, that for the deployment and implementation of these systems on a global scale many more steps need to be taken, and human drivers will still be highly needed.

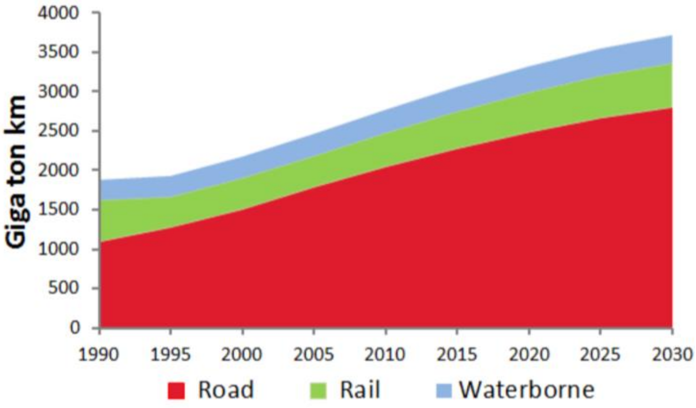


Figure 1. Transport demand in EU.[1]

As confirmed by the measurements with human drivers, documented in [2], during bi-directional low-speed maneuvering with articulated vehicle combinations the driver primarily suffers from a lack of view from the cabin, which is limited to the frontal outlook and the rear mirrors. Moreover, he/she is challenged to control the naturally unstable vehicle combination during reversing at the area which is very limited in space. The situation gets further complicated by the fact that typically no sensors on the trailing vehicle units are available, which otherwise would ease the operation for the driver to localize the position and orientation of the vehicle combination with respect to the distribution center and to other operating vehicle combinations. Potentially placed conventional GPS attached to the vehicle bodies does not provide sufficient accuracy to determine the vehicle combination pose (i.e. position and orientation).

Hence, this paper is introducing an intermediate form of driver support system for bi-directional low-speed maneuvering of articulated vehicles in distribution centers that brings advantages of automation into the existing fleet, without excessive technological costs which are restricting the implementation. With the use of VIsion Supported Truck docking Assistant (VISTA) technology, the technological changes to the truck are minimal and the driver continues to control the vehicle combination himself while being supported to use optimal paths and steering input to enhance the safety and productivity. A potential side-effect is that less experienced younger drivers can be employed more easily, as there is now a large shortage of drivers with enough experience.

2. Research Method

The general functionality of VISTA framework can be explained through Figure 2. Assuming the driver enters the distribution center, at first, he/she needs to report at a check-in desk where the loading dock number is being assigned to her/him and where the vehicle combination should be subsequently parked. Here the driver receives a smart device (currently assumed to be similar like a tablet, or wearable computer) which provides further navigation to the dock. The device is permanently connected to warehouse control system and is temporarily coupled by the driver to a specific vehicle combination which is to be parked.

At next, the vehicle combination pose is being accurately determined by the real-time camera-based localization system (RTLS) that uses the plane fitting approach with machine learning based algorithms for robust localization of all vehicle units. Given the actual pose availability, which can be considered as a start of the maneuver, the optimal path can be determined by the path planner towards the final destination. The resulting reference path ensures both the kinematic viability, and the length that is not excessively long. Moreover, the optimal path is pleasant to negotiate by the driver in both forward and reverse direction (i.e. without path segments having excessive curvature).

Figure 2. VISTA framework general principle

Based on the error between the actual pose of the vehicle combination and the proposed reference path the steering instructions are being translated to the driver to minimize the tracking error. The steering instructions on Human Machine Interface (HMI) follow the output of the path tracking controller which is applicable for both single or multiple articulated vehicles using the controller structure explained in [3]. In case the tracking error during the maneuver exceeds the maximal limit, the reference path needs to be re-planned such the vehicle can be still parked to the loading dock.

As for the form of HMI it is accounted to combine acoustic, haptic and visual channels to augment the reality and intuitively navigate the driver without excess of data which would increase the driver workload. The basis for the HMI definition will be interviews with the drivers, but also driver’s eye-tracking data recorded while performing the docking maneuvers at the distribution centers.

The paper is organized as follows; at first the work done on developing the real-time localization system is described, followed by the description of the VISTA framework which employs the virtual reality simulator for the development. The simulator integrates the motion planner, path following controller and human machine interface, description of which is provided hereafter. The paper is concluded with the description of next steps and expected results.

3. Real Time Localization System (RTLS)

The main role of RTLS is to deliver accurate information on the location and orientation of the truck and trailer bodies relative to the loading dock of the distribution centre. This is achieved using a camera system combined with and deep learning techniques. Cameras are placed above the docks such that the truck is always in view during the docking maneuver. To train the system the approach is firstly developed on scaled basis, and later will be deployed and tested full scale.

Truck pose estimation is achieved in number of steps show in Figure 3. First, the 2D position of a set of keypoints on the truck are detected in the image using deep learning. Deep learning for human pose estimation using keypoints is a proven technology, so a model originally developed for human keypoint detection was adapted to detect truck cabin and trailer keypoints. From a 256x256 resolution input image, the deep learning model produces a heatmap for each keypoint of the truck, as shown in the Figure 4. The maximum of each heatmap is then taken as the 2D position of the keypoint, using curve fitting to achieve subpixel accuracy.

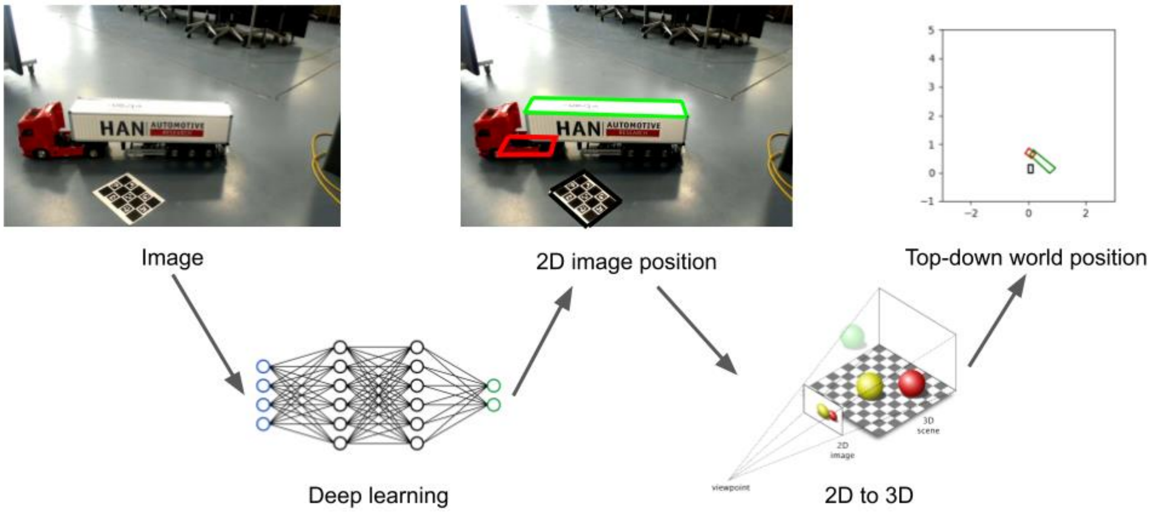


Figure 3. Overview of steps in vehicle pose estimation.

The next step is to transform these 2D points to 3D positions in the world. Using the camera's known focal length and 3D position and orientation in the world, 2D points in the image can be projected onto the ground plane to obtain their real-world position. Note that the height of the key point off the ground plane must be known in advance so it can be corrected for during projection to obtain the full 3D position of the key point in the world. While this approach requires a few more known parameters (notably camera pose and key point height relative to the ground plane) when compared to depth stereo camera systems, it has several advantages which make it more suitable for truck pose estimation. It relies on only a single monocular camera, reducing installation costs and increasing flexibility. Most notably, it achieves an accuracy of a few centimeters at several meters, beyond the operating range of typical stereo camera systems.

The deep learning model operates in real-time, achieving a minimum of 5 frames per second on an NVIDIA GeForce 950m graphics card. Higher speeds could easily be achieved on better hardware. However, with the addition of a kinematic predictive model to account for camera lag, 5 fps is typically sufficient for low-speed maneuvering. The scaled system has been verified to work in low light conditions using a night vision infrared camera.

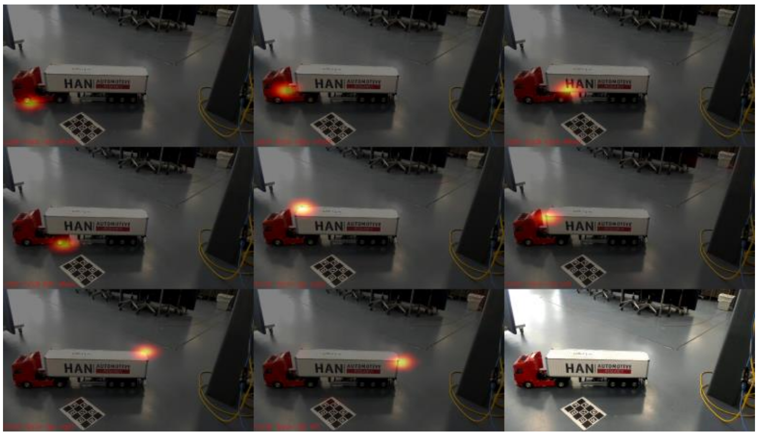


Figure 4. Heatmaps produced by the deep learning model to detect 2D position of truck key points.

4. VISTA Driver support framework development

The functionality of the VISTA driver support framework with the human in the loop is developed and tested in the Virtual Reality (VR) based simulator. The approach can be described as follows, a user will wear VR glasses which will display the environmental model of the driver's outlook from the vehicle cabin whilst operating in the distribution center, and HMI should support the user to successfully park the vehicle combination at the loading dock. The environmental model consists of a distribution center, its surroundings, and the vehicle combination that is developed in Unity software. The environmental model is powered by the kinematic model running in MATLAB/Simulink and the HMI is powered by controller which runs in MATLAB - Simulink as well. The input for both the controller and the kinematic model is provided by the user through the physical steering wheel in terms of steer angle and pedal positions as a response to the visual inputs, as can be seen in Figure 5. Hereafter we describe the components of the VISTA VR-Simulator.

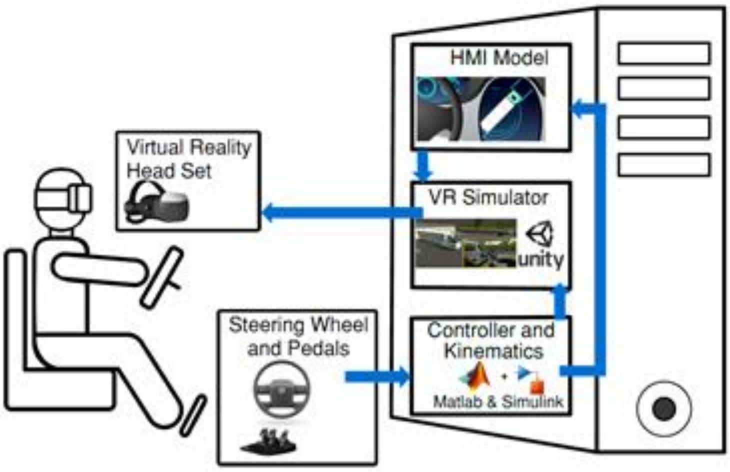


Figure 5. VISTA VR-Simulator architecture.

A. Simulator Hardware

To achieve high level of realism the simulator is built inside a real Volvo FH16 truck cabin, see Figure 6a). Herewith we used the Head-Mounted Display (HMD) HTC VIVE Pro Eye [4] which provides stereoscopic vision with a resolution of 1440 x 1600 pixels per eye (2880 x 1600 pixels combined), refresh rate of 90 Hz and a field of view of 110 degrees. This provides a high immersive VR experience while the user can freely observe the surrounding virtual world in any direction just by turning the head. The HMD is equipped with infrared sensors to track the user's position in real time by detecting infrared pulses coming from two emitters. The HMD has an integrated binocular eye-tracker with a sampling rate of 120 Hz, an estimated measured gaze accuracy of 0.5° as well as a trackable field of view of 110°. The gaze data recorded in the simulator is not only used to determine in real time at which objects the driver is directly focusing at but also which ones he/she is looking at in virtual the mirror. From the recorded gaze data we can not only identify the objects the driver is looking at but also in which temporal order the driver picks up internal and external information for the truck docking process. Additionally, eye-tracking is also used to improve the simulator performance by using foveated rendering technique, allowing the dynamic toggling of the mirrors and the camera used by the HMI. For instance, depending on the current gaze position, mirrors in/outside the visual periphery gets disabled in order to reduce the processing effort and to assure an optimal frame rate.

The hand tracking is realized with the help of optical based Leap motion controller [5] that is integrated in HMD (see Figure 6c). The user's hand motions and gestures can be thus captured and simultaneously mapped onto the virtual hands of the driver. In this way, the user controls the virtual hands inside the simulator and interacts with different objects, like the steering wheel or elements of the HMI, such as buttons. The properties of the virtual hands can be adjusted according to the user's needs (e.g. skin colour, arm length).

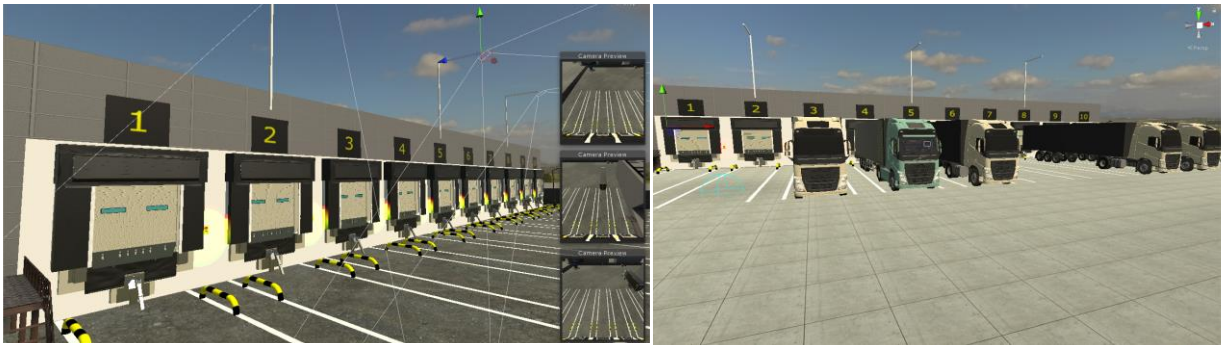
The driver interacts with the VR world by means of original steering wheel, shifting lever, and pedals, see Figure 6b, which are directly connected to the computer on which the simulation is running.

a) b) c)

Figure 6. VISTA VR-Simulator: a) cabin exterior, b)cabin interior, c) Leap motion controller mounted on the HMD

B. VR Simulation Modelling

The VR simulation software was developed using the Unity 3D Platform [6], which is a powerful platform for the development of a VR environments, and which is offering a simple integration with other components of the simulator. The VR simulation software allows the 3D representation of the Volvo FH16 tractor unit and a trailer (see Figure 7b) as well as the complete environment of the distribution center shown in Figure 7a. The distribution center environment was modeled based on dimensions measured in the real world and consists of ten numbered loading docks and three parameterizable docking assist cameras. The floor of the distribution center area has floor marking guidelines aligned with the loading dock. Additionally, the texture of the floor can be easily changed whenever the simulator requires different conditions. The surrounding environment is realized with a realistic skydome. Each loading dock also integrates a red and green light to inform the driver about the remaining distance of the truck to the docking door in order to support the docking process. Additionally, there is also the possibility to determine which is the targeted loading dock and to park trucks in any lane. The 3D models used for the construction of the distribution center were purchased from the Unity Asset Store or from a 3D model supplier.



a) b)

Figure 7. a) Distribution center loading docks and the three docking assist cameras, b) Distribution center with a parametrized number of parked trucks.

The Volvo FH16 tractor unit and the trailer are realistic 3D models that were acquired from a 3D model supplier. Before importing the 3D models in Unity, we used Blender [7] to reduce the mesh density and to detach parts of the truck as fully independent objects e.g. mirror, steering wheel and so on. Previous field studies with drivers wearing binocular mobile eye-tracking glasses while doing the truck docking process in the real world revealed that they mainly switch their gaze movements from the inside instruments to the mirrors. These results illustrate how important it is to simulate the mirrors in a realistic way. For this, the curvature of the mirrors is of particular importance. In that sense, a reflection probe is currently used as a solution (Figure 8a). Although this technique offers a realistic experience, it also comes with some limitations. A reflection probe is similar to a camera that captures a spherical view of the 3D scene in all directions. This implies that the rendering of objects located outside the field of view increase the rendering effort per frame considerably. Since the truck has six mirrors it can add up to significant effects on the frame rate due to computation demand.

a) b)

Figure 8. a) Mirrors with real-time reflections using reflection probes, b) HMI installed in a central position and interacted with the driver.

In order to overcome this problem, we implemented a solution that dynamically deactivates or decreases the resolution and refresh rate of the reflection probes based on the current camera direction as well as drivers' field of view determined from the recorded eye-tracking data. Note, that the driver can also rotate each of the six mirrors in the simulator in order to meet his/her preferences. Finally, we also integrated an additional assistive device in the interior of the Volvo FH16 truck cabin. To be more specific, we installed a tablet device for virtually running the HMI and to present graphical instructions/driving feedback to the user. The tablet can be placed at on arbitrary and actuated by means of leap motion sensor which tracks the hands as in Figure 8b.

C. Driver Support Human Machine Interface (HMI)

The HMI of the VISTA support system was designed to be intuitive and as flexible as possible. Therefore, the interface was also built in Unity 3D, so it can be easily integrated in the VR simulator environment as well as in the real world on tablet hardware of any manufacturer and can be adapted to various VR and AR headsets. The version of HMI used currently in the simulator is focused on showing the actual driving advice of the system. It is also capable of onboarding the driver and giving him instructions on where to drop off the hardware as well as generate user feedback.

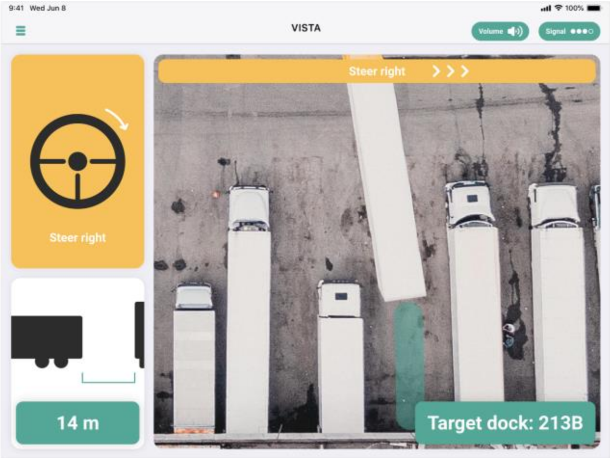


Figure 9. Prototype of HMI

Driving advice is communicated via several channels being; a visual representation of the steering wheel and it’s positioning. a simplified steering indicator, a target dock name plate, a visual representation of the distance to the docking port that is calculated from the steering system, an embedded video feed of the camera system (tracking system) that is overwatching the whole docking area of the distribution centre. First version of HMI prototype can be seen in Figure 9. Additionally, the HMI supports audio cues that represent steering angles and distance to the dock in order to give a passive indicator that does not need a visual affordance of the driver. All those indicators can be quickly adapted to integrate into the FOV (field of view) of the user's VR glasses as well so the direct impact can be measured as well.

D. Vehicle Path Planner and Path Following Controller

The main role of the path planner is to establish an optimal bi-directional reference path for the vehicle combination which connects the initial pose and the terminal pose represented by the loading dock. On the other side the role of the path tracking controller is to derive steering angle based on the error between the actual pose of the vehicle and previously established reference path, which can be subsequently provided as in a form of an advice to the driver. Given the low operational speed (below 2m/s), both subsystems are based on the assumption the vehicle combination driving behavior is perfectly kinematic, and thus no tyre slip is assumed. Since the path following controller was already extensively described in [3], in this paper we will concentrate on the path planner developed for VISTA, note that full description is provided in [8].

The path planner is using a so-called lattice-based approach in combination with motion primitives, where the environment is divided into a set of discrete states which can be connected and which can create the complete solution for the path. Hence at first, the operational environment is described in terms of free and restricted space using the polygons, which specify the location, shape and size of the obstacles in the sensed environment. The blue polygon in Figure 10a). depicts the available space whereas the red polygons represent the obstacles, whose position is also consistent with the environmental model of the VR. The state of the vehicle combination is defined by the position of the center of the semi-trailer axle in the global coordinate system x_2 and y_2, the yaw angle of the semitrailer θ_1, and the articulation angle γ.

Thus, the path planning problem is reduced to finding the path between two discrete states in the path planning environment where the path segments called motion primitives are generated, see Figure 10b). This is achieved by solving an optimal control problem from one of the discrete states to the other while using the kinematic equations of the articulated vehicle and differential constraints which limits the path curvature and steering angle rate and acceleration. It guarantees that created path segments are kinematically viable for the vehicle and still negotiable by the human driver given his/her physiological limitations in actuation of the steering wheel. The formulation of the optimal control problem with a desired cost function to optimize some parameters, the kinematic constraints and its solutions can be found in [8] and [9].

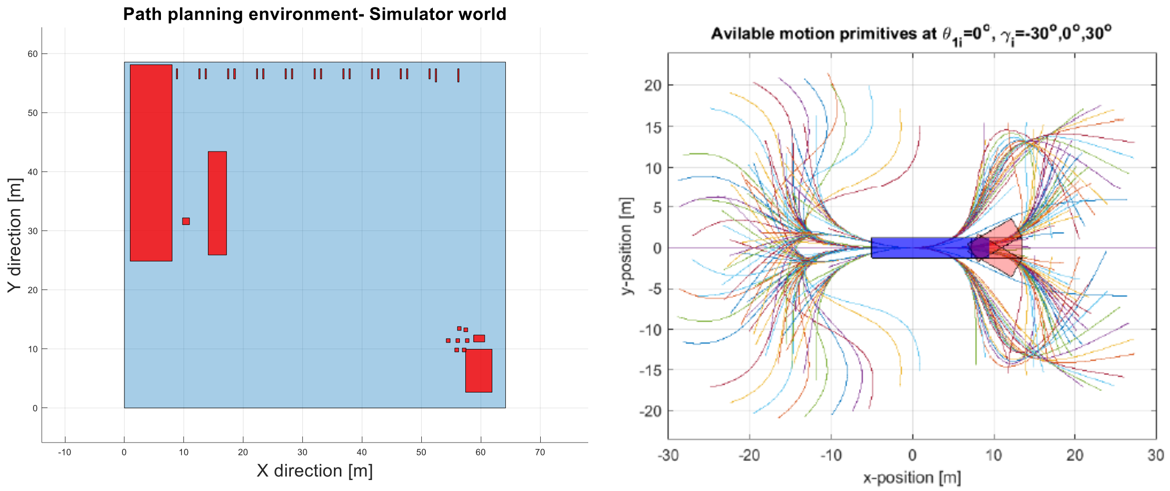


Figure 10. a) Path planning environment with polygons, (b) Motion Primitives.

Given the library of motion primitives defined, an algorithm is required to find optimal combination of motion primitives to traverse from one point to other while using the free space within the operational environment yet avoiding the obstacles. For this purpose, the graph search algorithm A* algorithm is used with a collision detection module which checks for collisions while planning. Customized A* algorithm is also extensively explained in [8].

5. Research Outlook and Next Steps

Given the simulator framework established, the first generation of HMI will be extensively tested in the simulator environment with human drivers in the loop. It is foreseen that obtained results will lead and consecutive re-iteration of HMI, Path Planner, and Path Following controller where fundamentally new approach based on Model Predictive Control will be tested. As for the RTLS, next steps include support for multiple trucks in the same image, which requires detecting multiple of the same keypoint on different trucks and grouping keypoints belonging to the same truck. Herewith, a deep learning architecture will be tested, which jointly detect individual truck instances and their associated keypoint positions, keeping the processing time per image constant. Another functionality that requires solution before the system can be full scale implemented is tracking vehicle units instances through multiple video frames which presents additional challenges across long distances, where multiple cameras are required to cover the full maneuvering area and where the information from multiple cameras need to be robustly merged.

6. References

[1]. European Commission, Impact Assessment - accompanying document to the White Paper, pp. 137, Brussels 2011
[2]. Kural, K., Besselink, I.J.M., Pauwelussen, J.P., Nijmeijer, H. & Patel, K., Analysis of driver behavior during reverse driving of double articulated vehicles, Proceedings HVTT13, San Luis, Argentina, 2014
[3]. K. Kural, Analysis of high-capacity vehicles for Europe: application of performance

based standards and improving maneuverability, Doctoral dissertation, Technische Universiteit Eindhoven, 2019
[4] “VIVE Pro Eye.” https://www.vive.com/eu/product/vive-pro-eye/overview/ (accessed Jun. 09, 2021).
[5] “Leap Motion Controller.” https://www.ultraleap.com/ (accessed Jun. 09, 2021).
[6] U. Technologies, “Unity 3D Platform.” https://unity.com/ (accessed Jun. 09, 2021).
[7] B. Foundation, “Blender,” blender.org. https://www.blender.org/ (accessed Jun. 10, 2021).
[8] D. Devaisa, "Motion planning with obstacle avoidance of single articulated vehicle," Master Thesis, HAN University of Applied Sciences, Arnhem, 2019.
[9] M. Kannan, "Automated docking maneuvering of an articulated vehicle in the presence of obstacles," Master Thesis, HAN University of Applied Sciences, Arnhem, 2019.