e-book

LeddarTech

LeddarVision
Sensor Fusion and Perception Technology Overview

LeddarVision
Sensor Fusion and Perception Technology Overview // e-book

Table of Contents

1. Introduction

In an autonomous vehicle, reliable and accurate perception of the environment is critical to enable safe driving decisions. The output from the various sensors needs to be fused with no loss of information to produce an accurate model of the environment that captures every surrounding object.

The standard approach used by perception platforms currently available in the ADAS and AD market is object fusion, where information about object detection performed by each type of sensor is brought together to support the driving decision-making.

The main limitation of this approach, however, is that no single sensor on its own possesses sufficient information to support all possible driving scenarios. For example, HD cameras do not see depth, while distance sensors such as LiDARs and radars may lack resolution.

The LeddarVision™ sensor fusion and perception solution provides a different, innovative approach to understanding the vehicle's changing environment with raw-data fusion. Through a leveraged algorithmic approach, the software solution encompasses 3D reconstruction, artificial intelligence (AI) and computer vision to turn sparse data into a most precise dense 3D environmental model, contributing to improving the perception system's performance, an essential component of autonomous vehicles.

This e-book will explain and illustrate the main features and benefits related to this advanced, cost-effective sensor fusion and perception solution, which requires low computational power and makes optimal use of available sensor suites.

2. Features and Challenges

2.1. System Overview

LeddarVision integrates raw data from sensors to generate a 3D RGBD[1] model, as shown in Figure 1 below. It also incorporates 3D reconstruction to enrich and add robustness to the model. Sensor data is fused intelligently, adding temporal information (i.e., information from multiple frames) and more accurate representations of a single measurement (i.e., multiple measurements of a single object enable reducing the measurement error by √n).

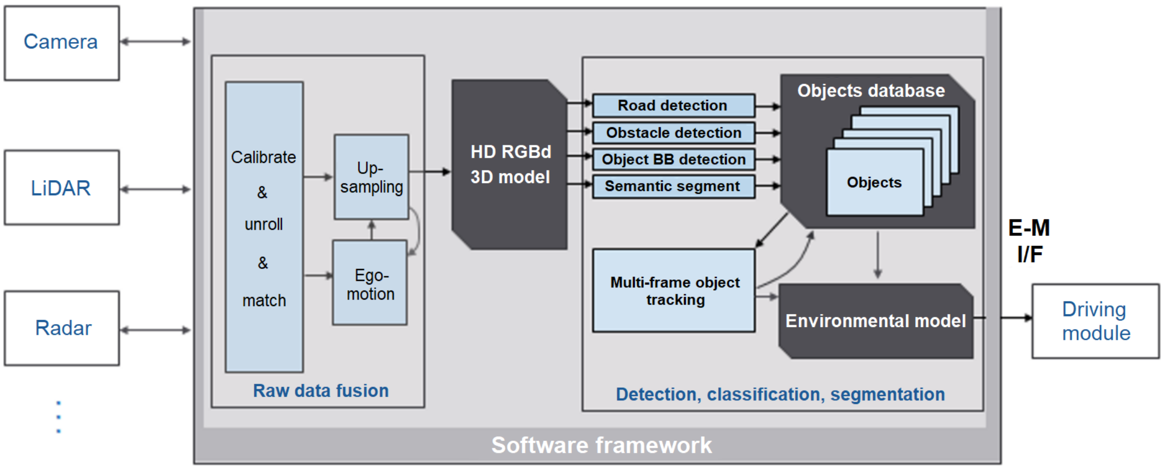


Figure 1 – System overview

The 3D RGBD model also provides an abstraction layer between the sensors and the perception algorithms. This model allows the use of different sensors' brand, resolution and positioning without modifying the algorithms. This provides a significant benefit as it reduces the testing, verification and validation required for the system when using a different set of sensors.

2.2. Object-Level Fusion vs. Raw-Data Fusion

While sensor fusion (fusion of data from different sensors) and perception (an online collection of information about the surrounding environment) is already in use in current ADAS and AD applications, the technology still has one major drawback: each detection is based on sub-optimal information (sensing data from the camera, radar, LiDAR, etc.), resulting in partial, or even contradictory information, which can lead the system to make a wrong decision.

In a traditional object-level fusion approach, perception is done separately on each sensor (Figure 2). This is not optimal because when sensor data is not fused before the system makes a decision, it may need to do so based on contradicting inputs. For example, if an obstacle is detected by the camera but was not detected by the LiDAR or the radar, the system hesitates as to whether the vehicle should stop.

[1] Red, green, blue + depth.

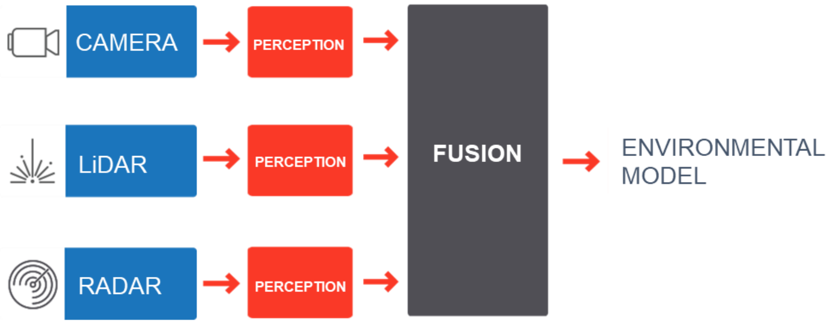


Figure 2 – Object-level fusion

In a raw-data fusion approach, objects detected by the different sensors are first fused into a dense and precise 3D environmental RGBD model, then decisions are made based on a single model built from all the available information (Figure 3). Fusing raw data from multiple frames and multiple measurements of a single object improves the signal-to-noise ratio (SNR), enables the system to overcome single sensor faults and allows the use of lower-cost sensors. This solution provides better detections and less false alarms, especially for small obstacles and unclassified objects.

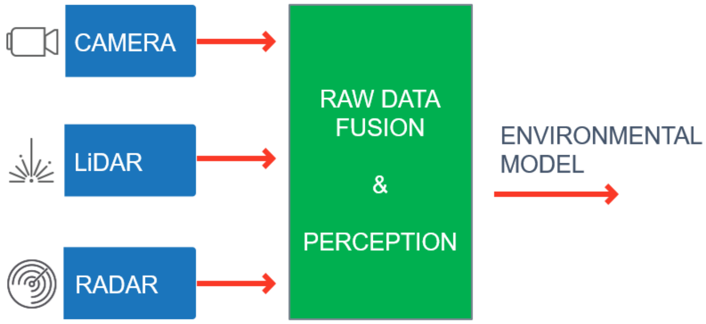


Figure 3 – Raw-data fusion

2.3. 3D Reconstruction and Upsampling

The core advantage that LeddarVision provides is a 3D reconstruction, which generates a high-density 3D image of the vehicle's surroundings from the camera, LiDAR points and/or radar measurements. By using the HD image from the vision sensor (camera), the algorithm divides the surroundings between static and dynamic objects. The LiDAR measurements on the static objects are accumulated over time, which allows the allocating of a larger portion of the distance measurements to moving targets. The acquired LiDAR measurements are further interpolated based on similarity cues from the HD image.

While LeddarVision's raw data fusion uses low-level data to construct an accurate RGBD 3D point cloud, upsampling algorithms enable the software to increase the sensors' effective resolution. This means that lower-cost sensors can be enhanced with LeddarVision and provide a high-resolution understanding of the environment. The sample results of the 3D reconstruction algorithm for a camera and LiDAR are shown in Figure 4. The top picture represents the camera-only frame, the lower-left the original LiDAR point cloud and the lower-right the LeddarVision 3D reconstruction with upsampling.

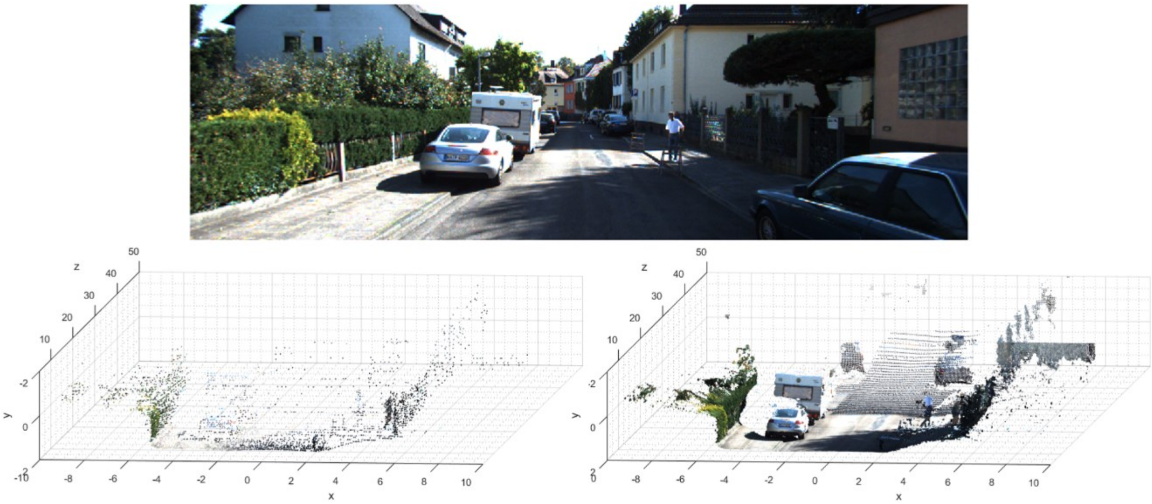


Figure 4 – 3D reconstruction example

For example, this method can be applied for low-density scanning LiDARs, as is demonstrated in Figure 4, thus enabling the use of a lower-cost scanning LiDAR, e.g., with 32 beams, instead of an expensive high-density 64-beam LiDAR, to achieve the required performance.

The RGBD model created through raw sensor fusion is sent to a state-of-the-art “RGBD object detection” module that detects 3D objects in a 4D domain. Meanwhile, the free space and road lanes are identified and accurately modeled in three dimensions, leading to an accurate geometric occupancy grid. This bird’s-eye-view grid of the world surrounding the AV is more accurate than using a camera-alone estimator. The RGBD model allows very accurate keypoint matching in 3D space, thus enabling very accurate ego-motion estimation.

With its novel approach based on proprietary IP, LeddarVision also adds another layer of raw (plus intermediate) data fusion and improves all aspects of the perception outcomes, as detailed in the following sections: detections, 3D information, occupancy grid and ego-motion.

2.4. Object Detection and Classification

For safe driving, all objects on the road and in the vicinity must be detected, identified and tracked for proper path planning. An example is shown below. The system marks an object by placing a tight 3D bounding box around it and defining its position (x,y,z), dimensions (width, height, depth) and orientation (αx,αy,αz) with respect to the 3D scene and the car location. Furthermore, the bounding box is projected onto the image, and a 2D or 3D bounding box is marked on the image. Finally, the system classifies the object in one of the following categories: vehicle, pedestrian, cyclist or unknown. The “unknown” category is used to classify obstacles on the road that do not belong to any other category (e.g., a cat crossing the street, or a hole).

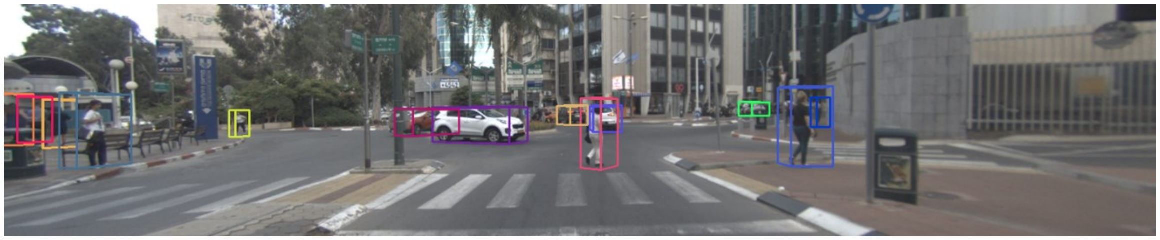


Figure 5 – 3D detection and enhancement (camera view)

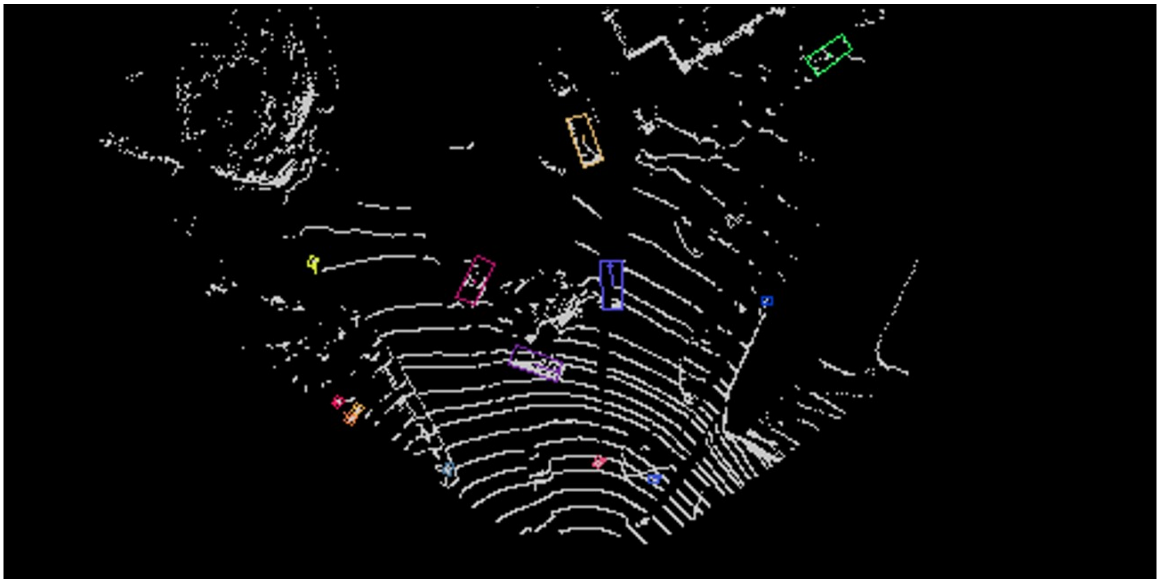


Figure 6 – 3D detection and enhancement (bird's-eye view based on LiDAR data)

2.5. Small-Object Detection

LeddarVision's raw data sensor fusion and 3D reconstruction provide more information that enables the detection algorithms to detect small objects farther than otherwise possible and smaller obstacles that would otherwise escape detection. It is this capability that provides the ability to safely drive faster and in more demanding conditions. LeddarVision's novel RGBD-based detection utilizes the depth features of obstacles, allowing the recognition of never-before-seen obstacles. This technique helps bridge the gap between the outstanding performance of image-based deep-learning approaches for object detection and the geometry-based logic of an obstacle. The solution also reduces the number of false alarms, such as those that can occur due to discoloration on the road, reflections or LiDAR errors. The example in Figure 7 highlights LeddarVision's ability to detect such small obstacles even at large distances.

Figure 7 – Small-object detection, example 1

Figure 8 below illustrates a small object on the road. Although it is still far away, the object is detected by the system, identified by a rectangle and classified as "unknown." The obstacle is actually a rock on the road, as can be seen in Figure 9.

Figure 8 – Small-object detection, example 2 (object is detected at a distance)

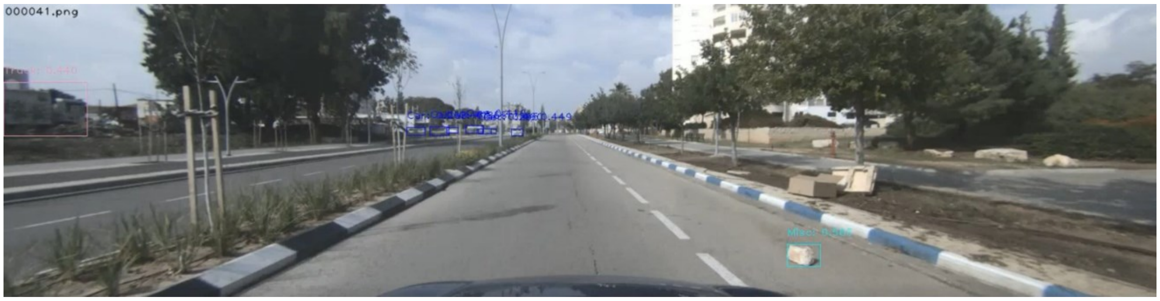


Figure 9 – Small-object detection, example 2 (object is identified)

2.6. Object Trajectories

Once detected, each object must be assigned with velocity estimation. The algorithms track and follow each detected object's motion path in 3D space by tracking the angular velocity through image frames and the radial velocity through depth-image frames. When Doppler radar data is available, this too is used. This allows the system to distinguish between dynamic and static objects. The method generates the 3D trajectory motion for dynamic objects, which will be used later for path planning and crash prevention. In the figure below, each 3D bounding box color represents a unique ID, and each vector in the bird's-eye-view display represents the velocity of the object.

Figure 10 – Sample object trajectories (3D + bird's-eye view)

2.7. Ego-Motion

LeddarVision can also provide advanced algorithms for visual odometry self-localization, based on estimating the motion between every two consecutive frames of a camera image and LiDAR point cloud to create the motion vector and trajectory of the host car. This solution is based on LiDAR, camera, inertial measurement unit (IMU) and CAN-bus data, enabling operation in global navigation satellite system (GNSS) denied regions such as in a tunnel. If integrated with GNSS, the GNSS accuracy is then increased by an order of magnitude; for example, the module includes drift compensation based on GNSS readings and correction of GNSS reflections. Figure 11 shows an example of the ego-motion function (blue is ego-motion, pink is GNSS as ground truth).

The image was generated using only a single camera and a single LiDAR, with the vehicle driving approximately 500 meters in the streets of Or-Yehuda, Israel.

Figure 11 – Basic localization from a bird's-eye view (blue = ego-motion, pink = GNSS as ground truth)

Keypoints with a corresponding depth can be used for optimal ego-motion results, as shown in Figure 12 and Figure 13. Keypoints are points in the frame that are static. In addition to their estimated depth, the ego-motion calculates the delta from the previous frame. The ones with the lowest estimated error are considered.

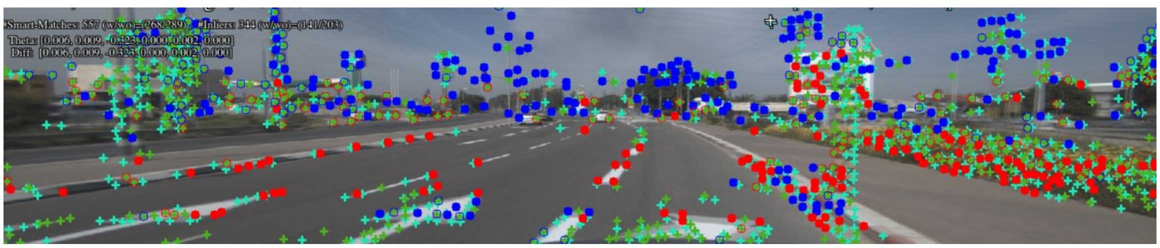


Figure 12 – Ego-motion keypoints (red = with depth, blue = without)

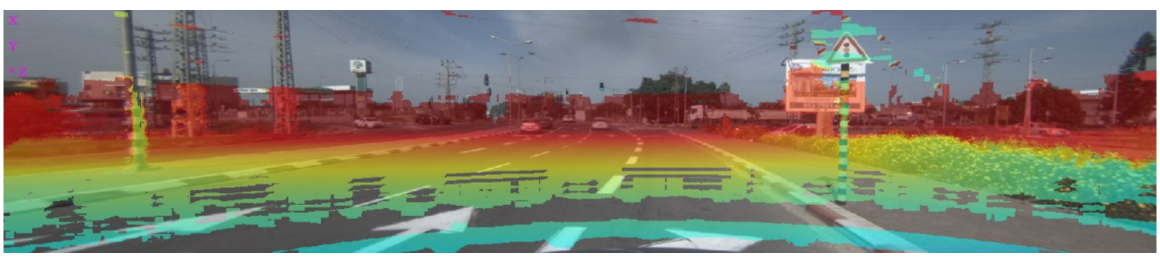


Figure 13 – Depth map (close = blue to distant = red)

2.8. Free-Space Detection

One of the significant outputs for ADAS and AD applications provided by LeddarVision is an accurate estimation of the free space available for the car to drive in, which is achieved by classifying each pixel in a "road / non-road" category. Proprietary algorithms take into account both HD-image and HD-depth map as created by the 3D-reconstruction block.

Figure 14 – Example of free-space detection (light green)

LeddarVision's free-space algorithm uses the front-view image to predict the per-pixel probability of a surface being drivable or not, using a deep convolutional neural network trained on extensive data collected over diverse and challenging scenarios. This prediction is projected onto a bird's-eye-view perspective using LeddarVision's non-linear bird's-eye-view manifold estimation, which does not assume a single road plane. This leads to a more robust model of the road surface compared to a first-degree estimation of the road surface commonly used.

Figure 15 – Free space (in green), front and bird's-eye views

2.9. Tracking

In order to utilize temporal information to further improve detection results and to add a velocity vector to each object, LeddarVision utilizes an advanced tracking algorithm. This has a significant benefit in filtering out false alarms and improving all fields' accuracy associated with an object's detection, including velocity. As part of this process, each object is given an ID that can be tracked throughout a sequence which infuses measures of 2D detections with temporal measures, including ID switches, which are shown in Figure 16 and Figure 17 below.

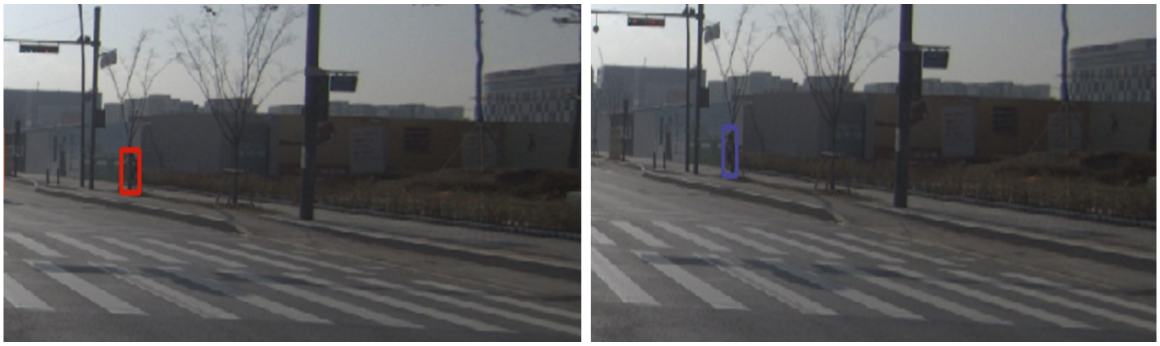


Figure 16 – ID switch on an occluded pedestrian (left: ID No. 1 = red; right: ID No. 2 = purple)

Figure 17 – ID switch on one distant and one occluded vehicle (left: distant = purple, occluded = blue. Right: distant = green, occluded = pink)

2.10. Road and Lane Detection

LeddarVision can identify lane lines and divide the road into lanes, which is achieved through a mix of image processing algorithms to detect markers and supervised deep neural networks (DNN) to identify and classify the markers. The algorithm then generates a segmentation of lanes and lane borders, as shown in Figure 18 below.

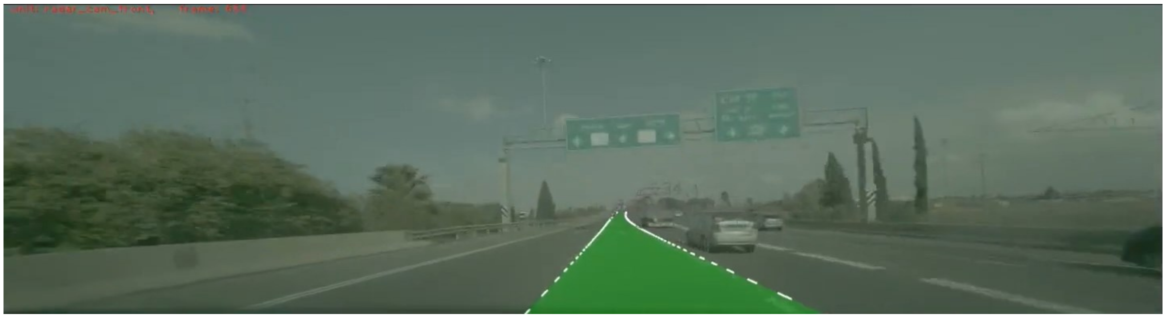


Figure 18 – Example of road and lane detection

Once the surrounding lane markers are identified, LeddarVision analyzes and returns important information per lane marker, such as its type (dashed or solid), single or double and its color (white or yellow). Additionally, LeddarVision models each lane using a polynomial fit with this model. Information on each lane's width, known length and center is provided. Figure 19 and Figure 20 below show examples of this method.

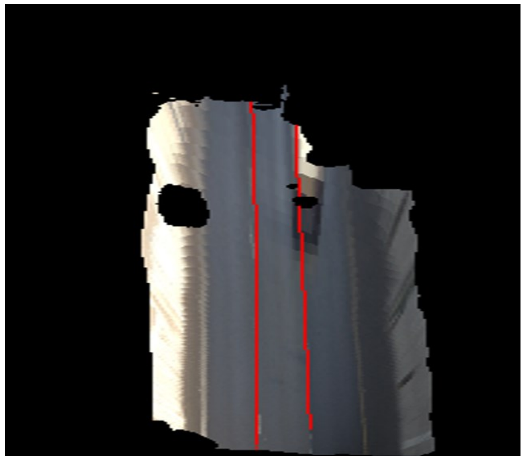


Figure 19 – Bird's-eye-view projection with a polynomial fit of the driving lane

Figure 20 – Driving lane and its middle visualization

2.11. Occupancy Grid

Occupancy grid[2] is a bird's-eye view of the vehicle's surroundings, including the free space location and all surrounding objects, which enables path planning. Figure 21 below shows a typical example of an occupancy grid, where green areas represent free space, red represent objects and arrows represent velocity.

---

2 "The basic idea of the occupancy grid is to represent a map of the environment as an evenly spaced field of binary random variables, each representing the presence of an obstacle at that location in the environment" (Wikipedia).

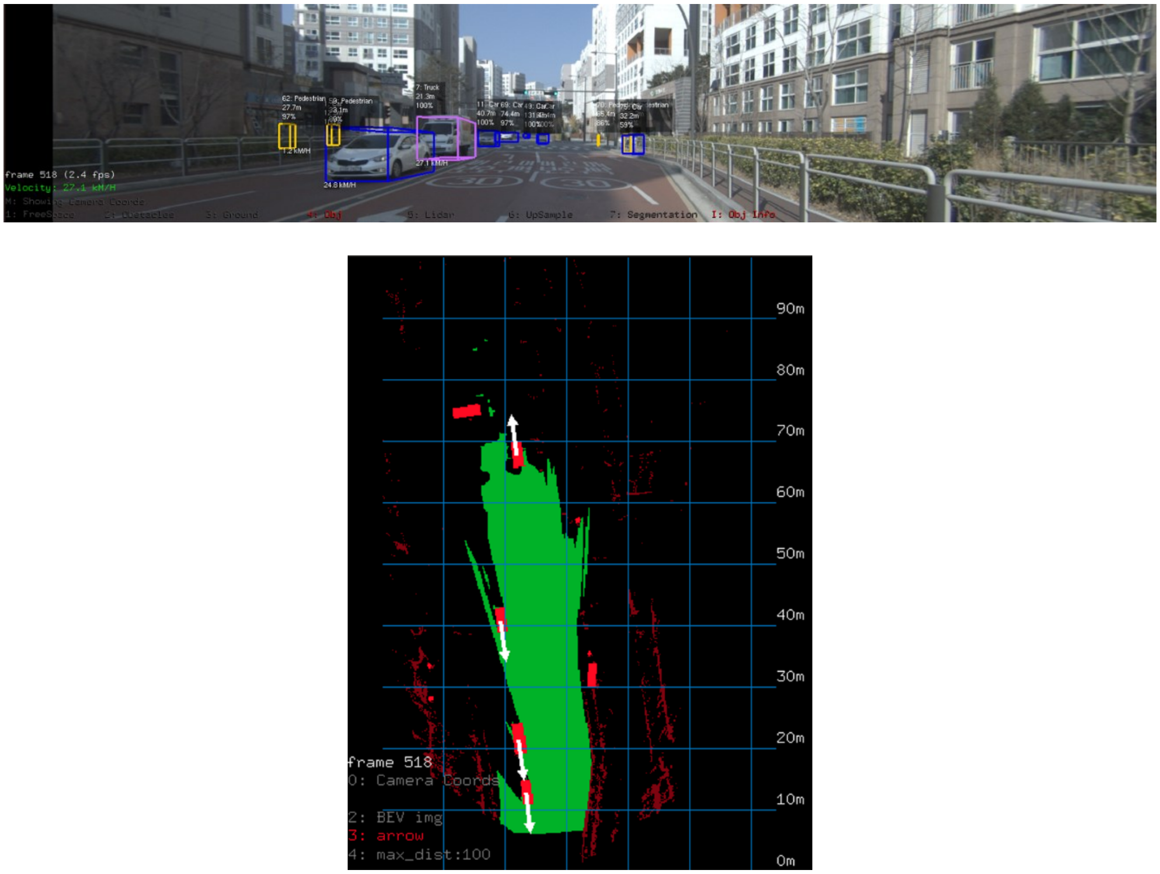


Figure 21 – Front view with detections (top) and corresponding occupancy grid (bottom)

2.12. 2D Detection and Classification

LeddarVision's RGBD-based detection utilizes the scene's image-based and depth-based features in parallel to accurately find all objects in the vehicle's surroundings. This technique has several advantages:

- Increased detection rate of objects that are hard to recognize in the image but contain good depth features (e.g., dark/saturated areas of the image).
- Reduced false alarms on posters, billboards, etc. due to lack of depth consistency.
- Generally improved results: real-world objects tend to have consistency between their image and depth-based features.

The detector recognizes the following classes: vehicles (cars, trucks, buses, trains, etc.), and humans (a pedestrian, person sitting, cyclist, etc.).

The following example compares camera-only detection working just on the RGB scheme and just on the camera, and LeddarVision's fusion algorithms working on the RGBD. In the top picture, a poster installed on a fence features flat images of cars. Whereas the camera-only system detects cars on the poster (blue rectangles) that do not actually exist, the RGBD detection model can detect real cars that are located much farther away (top left in the bottom picture) and discriminate them from the 2D representations on the poster.

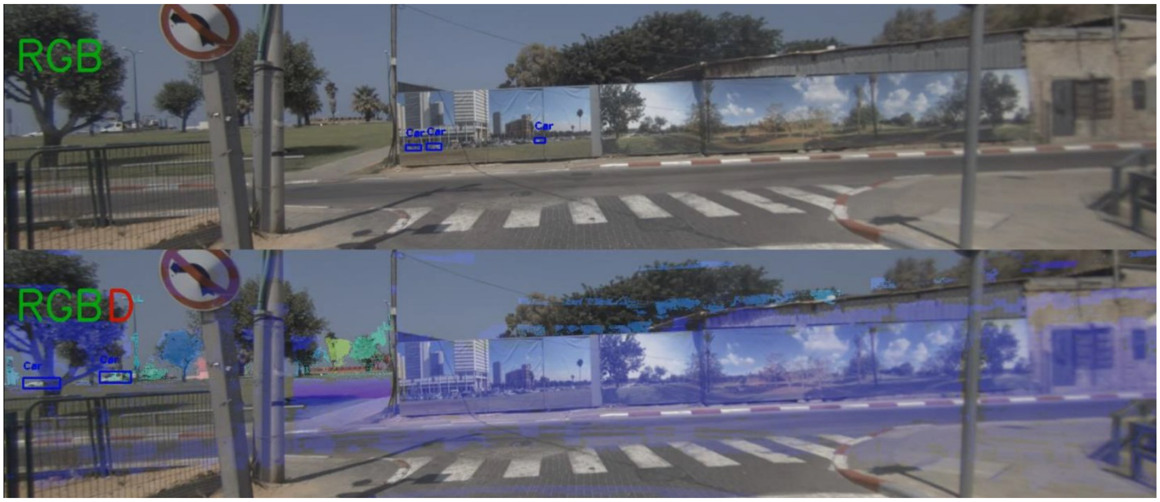


Figure 22 – Top: cars with no depth are detected (RGB detection).
Bottom: only "real" cars are detected (RGBD detection).

Another significant capability is that the system correctly detects partially occluded pedestrians, as illustrated below, which demonstrates the system's ability to perceive a pedestrian who is *about* to cross the road.

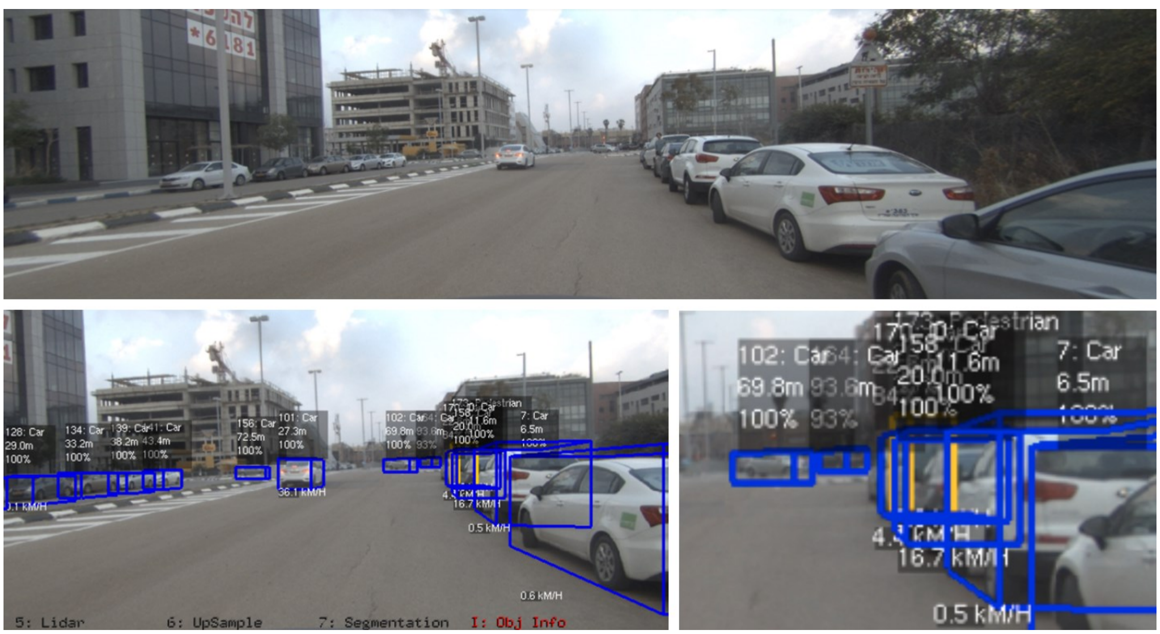


Figure 23 – Top: original image, showing partially occluded pedestrian detection.
Bottom left: pedestrian (in yellow) is detected between two cars;
bottom right: zoom-in of the scene.

2.13. 3D Detection and Classification

Each object detected by LeddarVision has a predicted position, size and orientation. The results of these predictions depend on the amount and accuracy of the LiDAR points returned from each object. LeddarVision's algorithm relies on a neural network's accurate predictions. Also, a non-machine learning (ML) fallback option is used in cases of high uncertainty.

The following four figures show the advantage of detecting 3D objects assisted by the 3D reconstruction algorithm in low-visibility cases.

Figure 24 – Cars (on the left) detected with blinding sun above

Figure 25 – Cars (on the left) and pedestrians (on the right) detected with blinding sun above

Figure 26 – Good detections in different lighting conditions

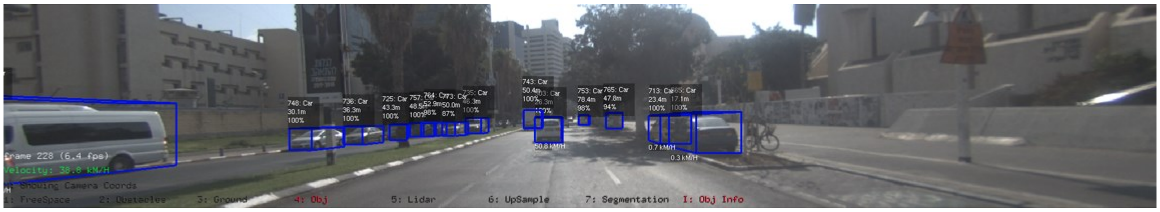


Figure 27 – Good detections in area of light transition (bright to shade on the right)

3. System Requirements and Architecture

3.1. Hardware Requirements

Processing requirements for the "sensor fusion" system can be divided into three sub-groups:

- CPU: basic multi-processor, multi-thread system. The CPU subsystem is currently moderately loaded.
- AI acceleration: a dedicated DNN acceleration could be used. Currently, DNN acceleration is done on the graphics processing unit (GPU).
- Image/data processing: general-purpose calculation could be executed on any available floating-point digital signal processor (DSP) machine.

3.2. Synchronization Requirements

Fusing data from different sensors, both at the object level and raw data level, requires synchronization. For the latter, the accuracy of the synchronization is even more critical. The system uses the camera frames as its synchronization points and aligns the LiDAR and radar inputs accordingly. Inter-sector synchronization is done in the object level tracking.

The system's synchronization scheme does not require any hardware-based time synchronization between sensors. Soft synchronization is used between different sensor data by using a timestamp-based mechanism. The time source for all sensors is the host computer's internal clock.

Since time measurements on a non-real-time operating system may fluctuate, proprietary methods have been invented and implemented to adjust timestamps according to the sensor's estimated origin hardware clock.

Highly accurate calibration between the LiDARs, cameras and radars is critical for raw data fusion and 3D reconstruction. This includes the camera's intrinsic calibration and extrinsic calibration for each sensor with respect to the vehicle. A translation error of <5 cm and a rotational error of <1 mrad is necessary for a proper RGBD model to be created, assuming the intrinsic calibration is perfect.

Three types of calibrations are performed with LeddarVision: factory calibration (using calibration targets), offline calibration ("in the wild") and online calibration (during driving). Figure 28 denotes an example of calibration between a LiDAR point cloud to the image plane of the camera. The distance of the LiDAR point is color-coded (red is near, purple is far). As can be seen, calibration here is very accurate.

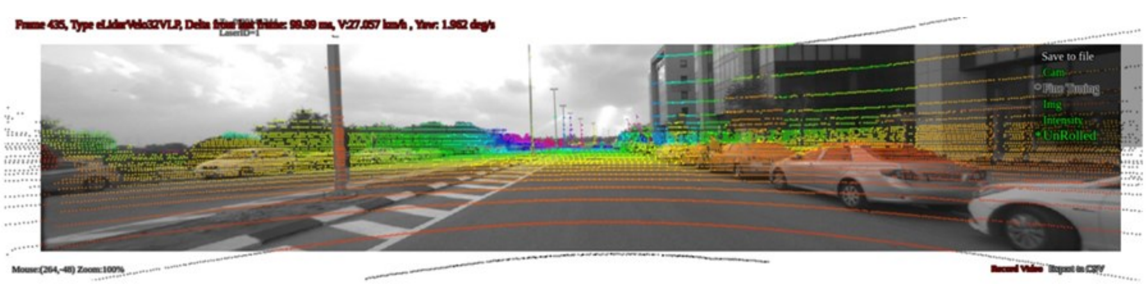


Figure 28 – Example of calibration between LiDAR point cloud and camera image

As opposed to the above, Figure 29 clearly shows miscalibrated LiDAR and camera data (left).

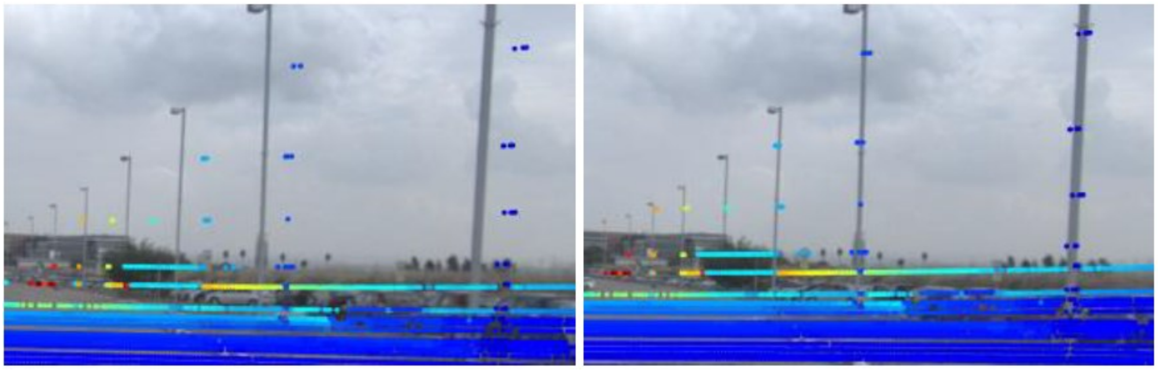


Figure 29 – Camera + LiDAR before calibration (left) and after calibration (right)

3.3. Modular Architecture

The LeddarVision solution is fully modular, in that it can support multiple features and combinations of features, starting from basic L1 (e.g., 1 camera + 1 radar) to advanced L2 (e.g., 2 cameras, 2 radars) or even full L3 implementation (e.g., highway pilot function, 6 cameras, 5 radars, long-range front LiDAR, etc.). All of the information will be routed into raw data fusion, thus enabling a precise 3D model, block-based architecture tailored to the client requirements as depicted in Figure 30.

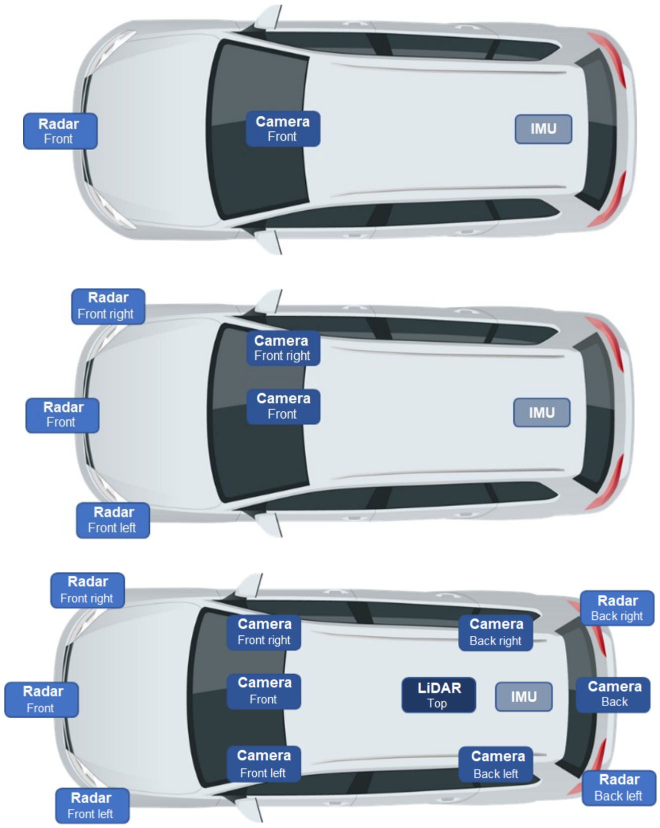


Figure 30 – Examples of sensor architecture, from basic to advanced

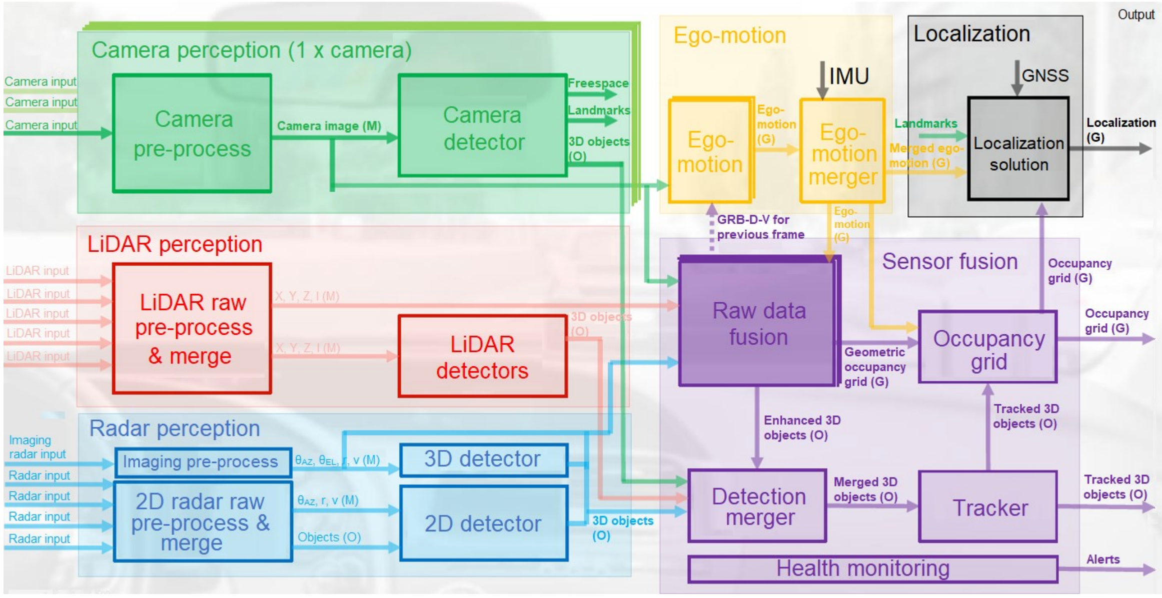


Figure 31 – System scalability and functional safety

4. Conclusion

As exemplified above, LeddarVision is a novel, comprehensive and scalable sensor fusion and perception solution that provides unrivaled performance even in harsh environmental conditions. This open software platform implements both a range of powerful and context-aware computer vision algorithms and deep neural networks that deliver highly accurate 3D environmental models for ADAS implementation and autonomous vehicles. While the system leverages data from various possible combinations of sensor modalities (including LiDAR, radar and camera), upsampling enables the software to increase the sensors' effective resolution. The system detects the various objects in the scene, including vehicles, pedestrians, bicycles, drivable road, obstacles, signs, lanes, lane lines, etc. LeddarVision also detects very small obstacles on the road with better detection rates and less false alarms than legacy "object fusion" solutions. Unclassified obstacles are also detected, providing an additional layer of safety to the vehicle.

With its modular architecture, robust design and multiple benefits over existing technologies, the LeddarVision sensor fusion and perception solution is strategically positioned to become an essential building block in future ADAS/AD implementations for all levels of autonomy.

LeddarVision™

- Best-in-class perception platform
- Custom embedded solutions
- Raw data fusion from multiple sensors
- LiDAR full-waveform analysis
- Context-aware algorithms

Detection of small objects absent in training sets

Depth data assigned to every pixel in camera picture

Accurate shape definition of vehicles, humans, and any other object

Superior Raw Data Fusion and Perception for Reliable and Safe ADAS & AD

LeddarVision is a sensor fusion and perception solution that delivers highly accurate 3D environmental models for ADAS and autonomous driving applications. The full software stack supports all SAE autonomy levels by fusing raw data from radar and camera for L2-L2+ applications and camera, radar and LiDAR for L3-L5 applications.

Bringing together industry-leading sensing technologies, LeddarVision applies AI and computer vision algorithms and deep neural networks with computational efficiency to scale up the performance of AV sensors and hardware essential for planning the driving path.

LeddarVision combines the strengths of each sensor type. Its state-of-the-art sensor fusion uses raw sensor data with upsampling to detect the various objects in the scene, including vehicles, pedestrians, bicycles, drivable road, obstacles, signs, lanes, lane lines and more. LeddarVision also detects very small obstacles on the road with better detection rates and less false alarms than legacy “object fusion” solutions. Unclassified obstacles are also detected, providing an additional layer of safety to the vehicle.

From factories to mining sites, an increasing number of industrial vehicles are being equipped with environmental perception solutions aimed at providing advanced driver assistance capabilities, increasing productivity or fully automating the vehicles’ operations. Delivering 360° perception in real-time and state-of-the-art unidentified obstacle detection based on raw data fusion from LiDARs, radars, cameras and other sensors, **LeddarVision** offers customizable, high-performance environmental perception solutions for all levels of industrial vehicle autonomy.

www.leddartech.com/leddarvision

A Leader in Environmental Sensing Solutions

About LeddarTech

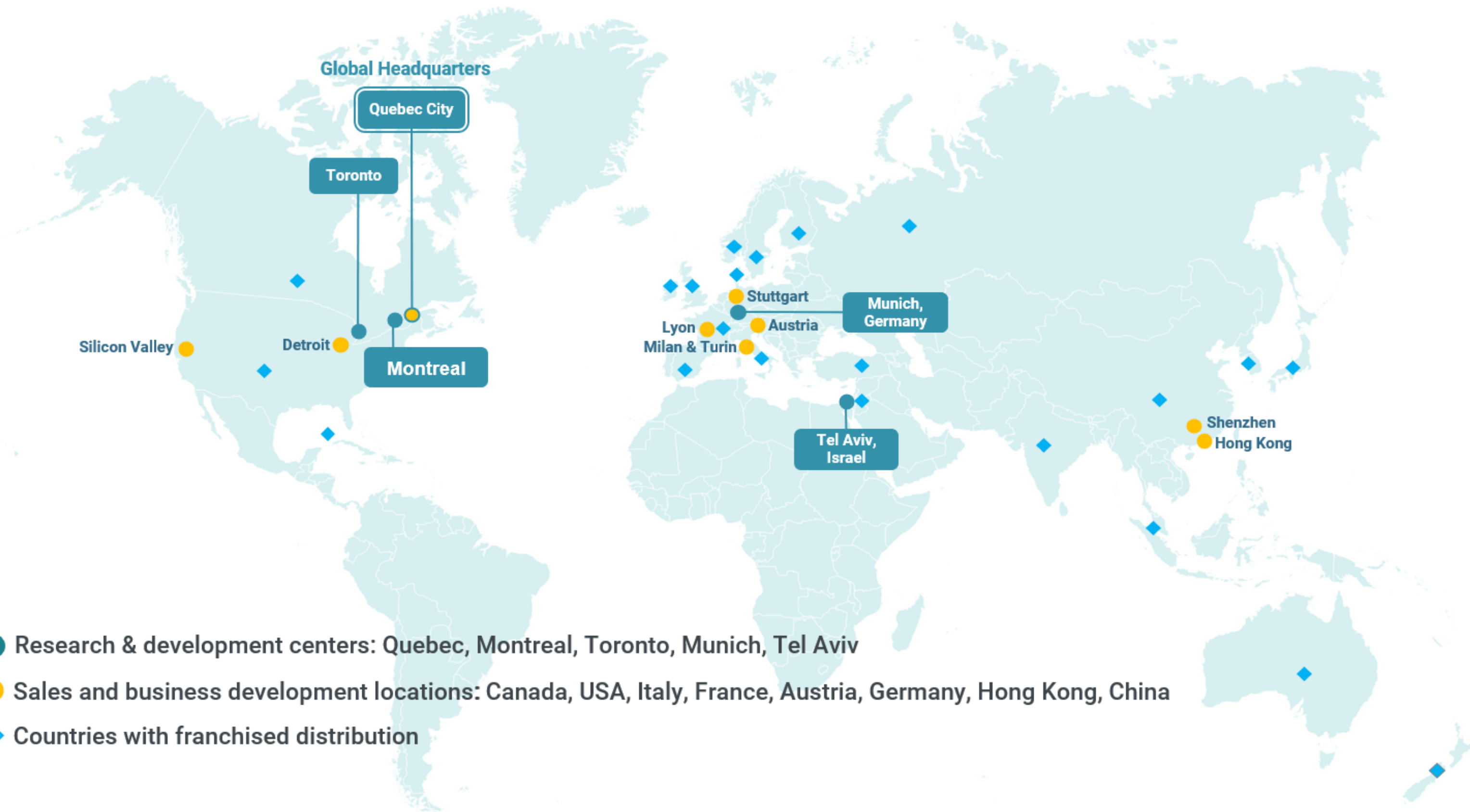
LeddarTech, a global software company founded in 2007, develops and provides comprehensive perception solutions that enable the deployment of ADAS and autonomous driving applications.

LeddarTech's automotive-grade software applies AI and computer vision algorithms to generate highly accurate 3D models of the environment, allowing for better decision making and safer navigation. This high-performance, scalable, cost-effective technology is leveraged by OEMs and Tier 1-2 suppliers to efficiently implement automotive and off-road vehicle solutions.

LeddarTech is responsible for several remote-sensing innovations, with over 150 patents granted or applied for that enhance ADAS and AD capabilities.

Reliable perception is critical in making global mobility safer, more efficient, sustainable and affordable: this is what drives LeddarTech to become the most widely adopted sensor fusion and perception software solution.

Global Expertise,
Global Presence.



- Research & development centers: Quebec, Montreal, Toronto, Munich, Tel Aviv
- Sales and business development locations: Canada, USA, Italy, France, Austria, Germany, Hong Kong, China
- Countries with franchised distribution

LeddarTech®

CANADA – USA – AUSTRIA – FRANCE – GERMANY – ITALY – ISRAEL – HONG KONG – CHINA

Head Office
4535, boulevard Wilfrid-Hamel, Suite 240
Québec (Québec) G1P 2J7, Canada
leddartech.com

Phone: + 1-418-653-9000
Toll-free: 1-855-865-9900

V2.2_EN / 20230224 / TF ID 028586