**COMPUTER VISION**

**Computer vision background**

Computer vision began in earnest during the 1960s at universities that viewed the project as a stepping stone to artificial intelligence. Early researchers were extremely optimistic about the future of these related fields and promoted artificial intelligence as a technology that could transform the world.

**Functions**

Computer vision is a field of artificial intelligence (AI) that enables computers and systems to derive meaningful information from digital images, videos and other visual inputs, and take actions or make recommendations based on that information. If AI enables computers to think, computer vision enables them to see, observe and understand.

Computer vision works much the same as human vision, except humans have a head start. Human sight has the advantage of lifetimes of context to train how to tell objects apart, how far away they are, whether they are moving and whether there is something wrong in an image.

Computer vision trains machines to perform these functions, but it has to do it in much less time with cameras, data and algorithms rather than retinas, optic nerves and a visual cortex. Because a system trained to inspect products or watch a production asset can analyze thousands of products or processes a minute, noticing imperceptible defects or issues, it can quickly surpass human capabilities.

**Application of computer vision**

* Computer vision for intruder detection- through hyperspectral cameras it is possible to differentiate between a fruit and a stone, which allows, especially in food, safer products for the consumer. Hyperspectral cameras are able to differentiate the type of material through the measurement they make of the wavelength. In this way, it is possible to differentiate a stone from a fruit, a plastic from a metal or other combination while the material is different.
* Computer vision for defect detection – With computer vision we can detect defects such as cracks in metals, paint defects, bad prints etc, in sizes smaller than 0.05mm. Much better than the human eye, these vision cameras need an algorithm that is the intelligent brain which is able to differentiate what is a defect and what is not. This algorithm is designed and trained specifically for each particular application through images with defect and without defects.
* Computer vision for metrology-It is one of the most important applications of the computer vision. What has been done so far with complex laser metrology equipment or probes, now it can be measured using computer vision. The key on this topic, is to make a good adjustment of the reference to be able to measure with the necessary precision, and above all, use the appropriate lighting for each type of material and work environment.
* Computer vision for Assembly verification- Every day more and more complex assemblies are made, with more parts or connections. Computer vison allows us to verify, step by step, that each piece is in its place, or at the end of the process, that the final assembly is correct. This application is very useful for the assembly of machinery, equipment, electronic boards or pre-assemblies with a lot of complexity. These systems significantly reduce cycle times of very complex operations and reoperation times.
* Computer vision for screen reader- Sometimes it is not possible to extract data from a display screen either because it is a closed supplier system or because that system is incompatible with the one installed. A solution to this problem is to install a computer vision camera to read the screen and extract data that appears on it for example temperatures, codes, tension and any useful information that appears on the screen and you need it. To accomplish this, we look for the interest regions in which the information is located, we use a character recognition algorithm (OCR) to extract it.
* Computer vision plus robotics for bin picking- This is also one of the applications which involves combining collaborative robotics with computer vision to be able to perform bin picking of pieces placed disorderly. The pieces are messy and therefore we need to optimize the trajectories and detect the grip coordinates. The robot needs help from the computer vision, telling it what a piece is and where it is, so it can decide what is the best way to pick up the piece.

**How we will use computer vision in our project**

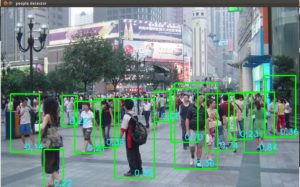
Computer vision for defect detection- this is one of the applications of the computer vision and this can help us detect holes in the farm area or rather ditches which human eye cannot detect. The holes are in this case defects which we are supposed to get rid of, and the computer vision will help us identify it and we will be able to get rid of it.

Computer vision for intruder detection- in our case the computer vision will help us detect foreign materials in the farm. Since our mission is navigation and tillage, we will have to identify foreign materials that we will most likely to encounter. For example during tillage, a tree is an obstacle which the computer vision will help us detect in the farm area when performing the task.

**Computer vision with the Raspberry Pi**

Computer Vision (CV) allows your Raspberry Pi to see things. In practical terms, this means that your Raspberry Pi can analyze an image, looking for items of interest and even recognizing faces and text.

If you link this with a camera to supply the images, then all sorts of possibilities open up.





## **Challenges and solutions in computer vision for autonomous vehicles**

In autonomous vehicles the quality and reliability of computer vision solutions can be a matter of life or death, either of the driver or of others on the road.

The agricultural sector has witnessed a lot of contributions when it comes to artificial intelligence (AI) and computer vision in areas like plant health detection and monitoring, planting, weeding, harvesting, and advanced analysis of weather conditions.

* **Autonomous car sensors**

Sensors are basically the senses of an autonomous vehicle and the foundation of its further actions. Currently, there are four leading sensor types used to check on surroundings and provide the controller neural network with the information required to make decisions:

1. Camera
2. Lidar
3. Radar
4. Ultrasonic sensors

**Problems:**

There is a wide array of hardware and software configurations, resulting in millions of possible pitfalls. Even a slight change in one or another sensor configuration, for example the sensitivity or the position regarding an axis of a vehicle, can alter the readings. Because of this, the neural network can have a problem with making proper interpretations.

Also, as it is in the case of humans, only a combination of sensors provides the driver with information accurate enough to ensure the safe movement of speeding, heavy machinery. When it comes to a human driver, there is a dependence on sight, hearing, and motion sense. An autonomous vehicle needs much more than that, using only the sensors mentioned above. The challenge is in building a comprehensive image of the outer world and processing it effectively.

* **Datasets**

We can say that an AI solution can only be as good as the data it was trained or validated on. In this case, gathering and processing datasets for the training of autonomous vehicles is a challenge in itself.

**Problems:**

When it comes to building datasets for autonomous drivers, there are at least two challenges to consider:

1. Gathering data- the best idea is to wander around in a car and gather basically everything.
2. Labeling data- all the data gathered needs to be properly labeled, usually requiring heavy, manual human labor.

The first challenge can be approached either by gathering data through existing semi-autonomous cars and driving support systems or to gather data in an artificial, simulated reality based on computer game engines. The most popular environment of this latter variety is called Carla.

For labeling the data there is no simple answer with the Captcha mechanism being one of the most successful. Within the slightly annoying mechanism of proving one is not a robot, a user is provided with two words or an image and needs to spot the correct elements. When it comes to words, one is verified to be correct, while the other is labeled by the user.

The same goes for the image. With thousands of Internet users tirelessly labeling images in search of traffic lights, road signs, or pedestrians, it has become possible to gather a decent number of labeled images that are further used in training image recognition-based solutions.

On the other hand, though, a company needs to be extremely careful when it comes to gathering training data, so sometimes we must evaluate the labels and ensure 100% accuracy of the dataset. When this consists of thousands of images, it is a challenge in itself.

* **Object recognition**

The prime and the most basic task of computer vision algorithms is to recognize an object in a picture. While computers outperform humans in multiple image recognition tasks, there are several which are particularly interesting in the context of autonomous vehicles.

**Problems:**

1. Object recognition needs to be done in a real-time environment. When it comes to input from a camera, it is sometimes based on a set of lines that are constantly flowing from the sensor and are used to refresh an ever changing image on a screen, rather than on a series of complete and whole images. Thus, there is a need to recognize objects without actually seeing them.
2. There are multiple elements in an environment that can be confusing for an autonomous system – a truck trailer in front of a car can be a good example.

This first challenge can be solved by training a model on the data delivered by the sensor as an output, practically switching the model toward signal analysis rather than image recognition.

The second challenge is an example of a typical problem of AI being unable to generalize and having no prior knowledge on a subject. An interesting solution comes from enriching image recognition with partial evidence which is a technique that enables the neural network to use a piece of additional information (for example context) to exclude the least probable outcome.

So if there is a car hovering higher than five meters off the ground, it is probably a billboard and there is no need to decelerate.

**Traffic Signs**

Traffic sign recognition is an iconic task for an autonomous vehicle’s neural network. The challenge in traffic sign recognition is in doing it quickly and in a highly volatile environment. A sign can be dirty, covered by leaves, twisted into an odd angle, or modified in this or another way.

Also, it is common to rearrange signs on the road or put some up temporarily, for example informing about a detour or road construction. So the net needs to be swift in processing and vigilant in spotting signs.

**Pedestrians**

An even more significant challenge comes with pedestrians. The machine needs not only to recognize one without a moment’s doubt but also needs to be able to estimate one’s pose. If the pedestrian’s motion indicates that he or she is going to cross the road, the vehicle needs to spot that and react quickly.

* **Semantic Segmentation And Semantic Instance Segmentation**

Semantic segmentation and semantic instance segmentation are intuitively similar problems with different sets of challenges.

Semantic segmentation is about detecting multiple entities in one image and providing each one with a separate label. Obviously, there can be a car, a road sign, a biker, and a truck on the road at the same time – and that’s what semantic segmentation is all about.

Semantic instance segmentation is about spotting the difference between each object in the scene. For a car’s system it is not enough that there are simply three cars on the road – it needs to be able to differ between them easily in order to track their individual behavior. While semantic segmentation frames each car, each tree and each pedestrian, semantic instance segmentation labels each as car1, car2, tree1, tree2 etc.

This challenge is to deliver a bit of additional information about the number of objects on the road and their positions regarding each other without naming them.

**Problems:**

1. Performance – as mentioned above, it is challenging to deliver truly real-time object recognition due to the limitations of the sensor itself. The same feature is limiting both for semantic segmentation and semantic instance segmentation in the same way it is for object recognition.
2. Confusion – while machines are increasingly efficient in their tasks, one needs to remember that there is always a present factor of unpredictability in the artificial neural network. Thus the network can be confused by several factors like unusual lighting conditions or weather.

Tackling these problems is tied to acquiring and leveraging bigger datasets that provide the neural network with more examples to generalize on. Providing the network with artificial data either generated manually or via Generative Adversarial Networks is one of the simplest ways to tackle this challenge.

* **Stereovision And Multi-Camera Vision**

Depth estimation is one of the key factors in ensuring the safety of a vehicle and its passengers. While there are multiple tools available, including camera radar and lidar, it is common to support them with multiple camera vision.

Knowing the distance between camera lenses and the exact location of an object on images taken by them is the first step toward building a stereo vision system. In theory, the rest is simple – the system uses triangulation to make depth estimations.

**Problems:**

1. Camera configuration – the distance between lenses and the sensitivity of the sensor can differ, delivering additional challenges for a depth estimation system.
2. Non-parallel representation – the cameras used in autonomous cars can deliver slightly different images, without pixel-to-pixel world representation. Thus, if there is a hardware shift in a pixel representation of an image, the system can find it more challenging to calculate the distance.
3. Perspective distortion – the bigger the distance between the camera lenses, the better the depth estimation. Yet this comes with another challenge – perspective distortion, which needs to be accounted for in the depth estimation.

The Tooploox engineers have overcome this challenge by feature engineering and combining the stereoscopic data with information from lidar and radar devices. Also, it is a matter of feature engineering – with enough artificial data generated to tweak the algorithm and real data to validate the effects, the team was able to deliver top-class results.

A good example of multi-objective camera usage in the automotive industry is Light, one of Tooploox’s clients.

* **Object Tracking**

Object tracking aims to provide the autonomous vehicle control system with information about the current motion and motion prediction of the object. While object recognition informs the system that a particular object is a car, a truck, or a tram, this feature delivers information on whether or not that object is accelerating, decelerating, or maneuvering.

**Problems:**

1. Risk estimation – the network needs to not only predict movement but also to anticipate behavior. In the same way, a human driver is often more careful when driving near cyclists than when driving along with other cars. Any incident would be more dangerous for a person on a bicycle.
2. Volatile background – when tracking an object, the network needs to deal with changes in the background. Other vehicles are approaching, the road changes color or there are trees instead of fields behind it. This is not a problem for a human driver, but it can be utterly confusing for a neural network.
3. Confusing objects – objects on a road are fairly repetitive, with dozens of red cars passing by each day. The tracking software can possibly mismatch one red car with another and thus provide the controller network with inaccurate information.

Providing the controller neural network with multimodal data gathered by Lidar and Radar is a good answer for this challenge. While Lidar can struggle with the identification of the type of a particular object, it delivers the exact position of it with pinpoint-accuracy. Radar provides less accurate data, yet independent of the scene and rarely affected by other factors like weather conditions.

* **3D Scene Analysis**

Combining all information gathered from the techniques described above, the controlling system should construct a 3D representation of the surrounding world. This can be compared to one’s imagination, where the computations are done, the effects estimated and the outcome produced.

**Problems:**

1. Accuracy – in 3D scene analysis even the slightest inaccuracies tend to stack into a larger mistake and result in a drift. What appears to be harmless at low speeds can become visible when velocity rises.
2. Multiple unpredictable objects – what is fairly straightforward on a highway gets complicated in urban traffic, where the street network grows more complex as well as the intentions of the other actors on the road.

Without the supporting role of Lidar and Radar, an effective and fast 3D scene analysis would be extremely challenging. Tooploox engineers and researchers have approached this challenge by working with point cloud data on object identification.

By this method, the system controlling the autonomous vehicle receives accurate information about objects in the scene from two independent sources – a camera with image recognition capabilities and a Lidar system with point cloud data analysis and identification.