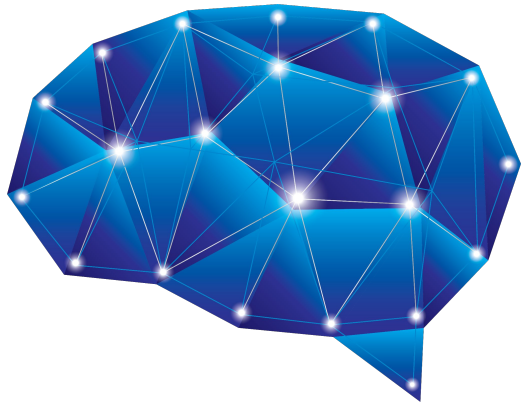
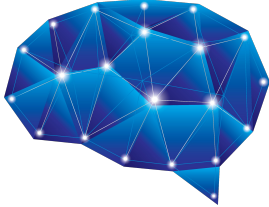
Artificial Intelligence

Replacing LiDAR with Computer Vision for Depth and Distance Perception

Presented to :- Dr. Deepti Malhotra
Presented By :-
Krish Gupta (23BECSE28)
Gagandeep Singh (23BECSE19)
Sameer Ansari (23BECSE47)
Keshav Bhatt (23BECSE26)



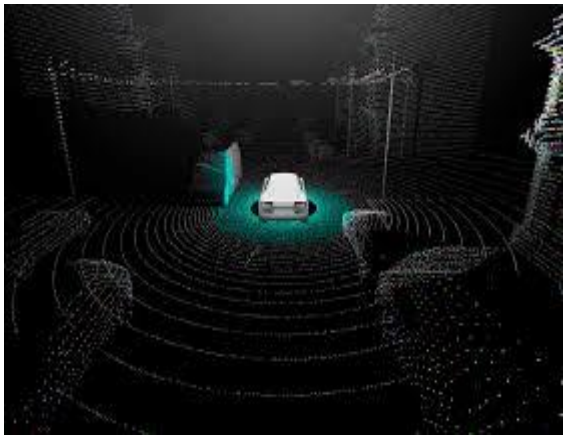


??..**Autonomy :-**

- In the sense of robotics and machines, autonomy means to the ability of machines (e.g. drones, Autonomous vehicles etc.) to perform tasks without human control either fully or to some degree.
- The degree of autonomy is classified into 6 levels. The 4th and 5th level of autonomy refers to the degree of automation in which a self driving car can handle all driving functions in all the scenarios and environments but still driver should be present for avoiding accidental situations.
- In a self driving car a core functionality of automation system is measuring the distance between objects (e.g. other cars or pedestrians) or more technically Depth and Distance Perception accurately.
- Most of the self-driving car companies uses LiDAR technology but Tesla uses the Computer vision for this.
- As the driving system and environment is operated through human visions so computer vision can definitely be the best option.

??..LiDAR and how it works:-

LiDAR sends out light pulses—which originates from a pulsed laser—to measure various distances by collecting the reflected light pulses. The internal systems calculate how long it takes for the light to return to the sensor and then calculates the distance through time of flight measurements.



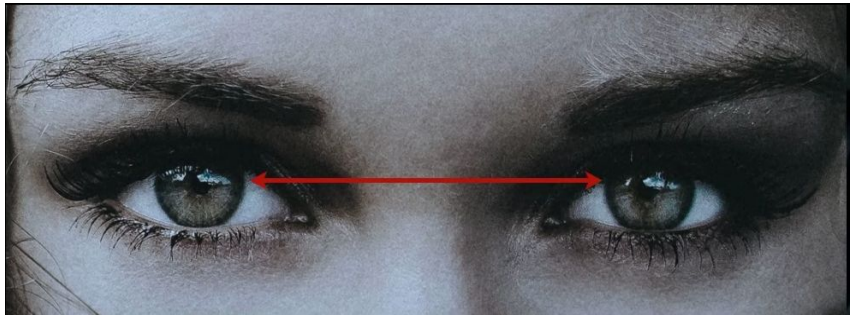
LiDAR Sensors are mounted on the vehicle

Sensors sends and receives light rays and calculate distance

Based on distances a 3D depth mapping of env. Is created

??..Human Stereo visions :-

- Humans do not emit and receives lasers for depth perception. Actually Human eyes and Neural network derives all the measurement in the 3D understanding of world, just from vision.
- Humans have two eyes aligned horizontally which is called stereo vision, at every single step we get two independent measurement and our brain stitches these measurement to arrive at some depth estimation and creates a 3D mapping of env.



Human stereo vision provides depth perception from images also as in the image, it seems that upper blue bar is longer than lower one, even both of them are of same length

??.. Computer Vision and Neural Network :-

- For the stereo vision like human, multiple stereo cameras are placed at different parts of car, real time data collected by cameras is processed through neural networks and a 3D depth map is created.

Click on the videos below

Stereo Camera vision

3D reconstructed mapping of surrounding

??.. How to use computer vision for solving the problem :-

- *SfM (Structure from Motion)* is one of the technique that can be used to solve it.
- SfM is based on the idea of capturing a series of images of an object or a scene from different viewpoints, and then reconstructing the 3D scene and the camera positions.
- The key concept is triangulation, where the same point in the scene is observed from multiple viewpoints, and the depth (distance to the camera) can be estimated by determining the position of the point in 3D space.
- The basic workflow of sfm involves 5 steps:-

1. **Feature Detection**: Identify key points or features (such as corners, edges, or textures) in the images.
2. **Feature Matching**: Match these key points across multiple images.
3. **Camera Pose Estimation**: Estimate the position and orientation (pose) of the camera for each image.
4. **Triangulation**: Use the matched feature points and camera poses to triangulate 3D points in the scene.
5. **3D Reconstruction**: Reconstruct the full 3D structure of the scene or objects by combining all the triangulated points and camera poses.

??.. How much we can do :-

- According to the knowledge we have about AI, we are able to perform few steps for building this model.

  1. Data Collection
  2 .Feature detection and matching. (Semantic segmentation)
  3. Monocular Depth estimation. (Using Deep Learning)

- **Data Collection** : We need a set of images (from a camera) that capture the scene or object from different angles. These could be:

  Monocular images (single camera).
  Stereo images (two cameras).
  Video frames (for continuous motion).

- **Feature detection and matching** :- To reconstruct the 3D structure, we first need to detect and track features in the images.

1. We will use **Mask-RCNN** to segment the images into different parts so that our model can detect features (eg. car, person) from real time images. It will help in feature matching and triangulation (computing the 3D coordinates of a point in the scene.)
2. Once feature will be detected we can use feature matching algorithms like **FLANN** (Fast Library for Approximate Nearest Neighbors).

3. In the feature matching step algorithm match features across consecutive images. The goal is to identify which features correspond to the same 3D point in the scene, despite the camera's movement.

- **Monocular Depth Estimation**:- After obtaining the 3D points (In 3D Reconstruction step), we can generate depth maps. Depth maps represent the distance from the camera to each point in the scene.

  For monocular (single camera) setups, we can convert the triangulated 3D points into a depth map using the camera's internal parameters (like focal length and principal point).

  Then we should use CNN-based models, like Monodepth2, which estimate depth from a single camera image, and combine this with SfM for real-time processing.

But still there are few important steps are remaining for building the model:-

1. **Camera Pose estimation** (estimate the camera's position and orientation (pose) for each image.)
2. **Triangulation** (computing the 3D coordinates of a point in the scene.)
3. **3D Reconstruction** (3D model of the scene by combining the 3D points into a point cloud.)

Thank You