Go, change the world®

ARTIFICIAL INTELLIGENCE IN AUTONOMOUS VEHICLES

VII semester
Department of AIML
RV College of Engineering

Course Incharge: Dr. Viswavardhan Reddy Karna

2024-2025

Course Contents

Go, change the world®

Semester: VII						
ARTIFICIAL INTELLIGENCE IN AUTONOMOUS VEHICLES **Category: Professional Core Elective** **(Theory)**						
Course Code	:	21AI73G1		CIE	:	100Marks
Credits: L: T: P	:	3:0:0		SEE	:	100 Marks
Total Hours	:	40T		SEE Duration	:	3.00 Hours

Unit-I	8 Hrs.
Introduction to Autonomous Driving: Autonomous Driving Technologies Overview, Autonomous Driving Algorithms, Autonomous Driving Client System, Autonomous Driving Cloud Platform Autonomous Vehicle Localization: Localization with GNSS, Localization with LiDAR and High-Definition Maps, Visual Odometry, Dead Reckoning and Wheel Odometry, Sensor Fusion	
Unit – II	8 Hrs.
Perception in Autonomous Driving: Introduction, Datasets, Detection, Segmentation, Stereo, Optical Flow, and Scene Flow, Tracking Deep Learning in Autonomous Driving Perception: Convolutional Neural Networks., Detection, Semantic Segmentation, Stereo and Optical Flow	
Unit –III	8 Hrs.
Prediction and Routing: Planning and Control Overview, Traffic Prediction, Lane Level Routing Decision, Planning, and Control: Behavioral Decisions, Motion Planning, Feedback Control	

Course Contents – Reference Books

Go, change the world®

Reference Books	
1.	Creating Autonomous Vehicle Systems, Second Edition Shaoshan Liu, Liyun Li, Jie Tang, Shuang Wu, and Jean-Luc Gaudiot ,2nd Edition, September 2020, ISBN: ISBN: 9781681739366
2.	George Dimitrakopoulos, Aggelos Tsakanikas, Elias Panagiotopoulos, Autonomous Vehicles Technologies, Regulations, and Societal Impacts, 1st Edition,Elsevier Publications, 2021 , ISBN-10 1681730073
3.	Hanky Sjafrie, "Introduction to Self-Driving Vehicle Technology", 1st Edition, Published December 11, 2019 by Chapman and Hall/CRC, ISBN: 978-0-323-90137-6
4	Creating Autonomous Vehicle Systems Shaoshan Liu, Liyun Li, Jie Tang, Shuang Wu, and Jean-Luc Gaudiot October 2017, ISBN-10 1681730073

Course Contents – Unit - II

Go, change the world®

Perception inAutonomous Driving:

- Introduction
- Datasets, detection, segmentation
- Stereo, Optical Flow, and
- Scene Flow, Tracking

Deep Learning in Autonomous Driving Perception:

- Convolutional Neural Networks.,
- Detection, Semantic
- Segmentation, Stereo and Optical Flow

Perception in Autonomous Driving: Introduction

- The goal of perception in autonomous driving: Sensing the Dynamic Environment for Safe and Intelligent Driving.

Key Tasks in Autonomous Driving Perception:

- **Obstacle Detection**: Identifying objects or obstacles around the vehicle
- **Road Surface Recognition**: Understanding the type of road (e.g., asphalt, gravel)
- **Lane Divider Recognition**: Detecting lane boundaries for lane-keeping
- **Traffic Signs and Lights Recognition**: Identifying signs and lights for rule compliance
- **Tracking Moving Objects**: Detecting and tracking dynamic objects in 3D space

Perception in Autonomous Driving: Introduction

Go, change the world®

Importance of Accurate Perception

- **Safety**: The foundation of safe driving decisions
- **Intelligent Decision Making**: Perception feeds into planning and control
- **Real-Time Decision Making**: Perception must be fast and reliable

Major Functionalities of Perception

1. **Environment Mapping**: Creating a map of the surroundings in real-time
2. **Object Detection and Classification**: Identifying and categorizing objects
3. **Motion Tracking**: Monitoring and predicting movements of dynamic objects
4. **Scene Understanding**: Recognizing and interpreting road infrastructure

Perception in Autonomous Driving: Introduction

Go, change the world®

Public Datasets in Autonomous Driving Perception

- **KITTI Dataset**: Key dataset for perception tasks, CITYSCAPES
- **Waymo Open Dataset**: Provides diverse data for autonomous vehicles
- **ApolloScape**: Comprehensive dataset for driving scenarios
- **nuScenes**: Multi-modal dataset for autonomous driving research

Problem Definitions in Autonomous Driving Perception

- **Object Detection**: Identify and localize objects in sensor data (e.g., LiDAR, camera)
- **Semantic Segmentation**: Classify each pixel of an image for road understanding
- **Trajectory Prediction**: Predicting the movement of detected objects
- **Scene Understanding**: Recognizing complex scenes for navigation

Typical Algorithms Used in Perception

- **Traditional Methods**:
  - Edge detection
  - Template matching
  - Kalman filters
- **Deep Learning Methods**
  - Convolutional Neural Networks (CNNs)
  - Recurrent Neural Networks (RNNs)
  - 3D object detection

Perception in Autonomous Driving: Datasets

- **Stereo and Optical Flow Data**: The KITTI dataset includes stereo image pairs (194 training and 195 testing) taken by two cameras simultaneously, and optical flow image pairs (194 training and 195 testing) captured by the same camera at consecutive time steps.

- **Data Features**: Approximately 50% of the pixels have ground truth displacement data, with stereo data conveying depth information and optical flow data conveying motion information.

- Visual odometry data: Twenty-two sequences of stereo image pairs, more than 40,000 frames, covering 39.2 km distance.

- Object detection and orientation data: Manually labeled data with 3D frame notating object size and orientation.
- Object types include sedan, van, truck, pedestrian, cyclist, etc.

Perception in Autonomous Driving: Datasets

Go, change the world®

1: KITTI car photo, adapted from the KITTI website.

The car is equipped with:

- 1 Inertial Navigation System (GPS/IMU): OXTS RT 3003,
- 1 Laser scanner: Velodyne HDL-64E,
- 2 Grayscale cameras, 1.4 Megapixels: Point Grey Flea 2 (FL2-14S3M-C), taking snapshots at a 10 Hz rate, and
- 2 Color cameras, 1.4 Megapixels: Point Grey Flea 2 (FL2-14S3C-C), taking snap shots at a 10 Hz rate.

Overview of KITTI Dataset

- KITTI dataset is one of the most widely used benchmarks in autonomous driving research.

- It includes multiple types of data collected using cameras, lidar, and GPS.

- Real-world data captured under varying conditions.

Stereo and Optical Flow Data

- **Stereo Image Pair:** Taken by two cameras simultaneously.
- **Optical Flow Pair:** Captured by the same camera at consecutive time steps.
- **Data Details:**
    - 194 training image pairs
    - 195 testing image pairs
    - ~50% of pixels have ground truth displacement data
- **Purpose:** Stereo for depth, Optical flow for motion.

Overview of KITTI Dataset

Visual Odometry Data: Data Details:

- 22 sequences of stereo image pairs
- Over 40,000 frames, and Covers 39.2 km distance

- **Purpose:** Estimating camera motion using stereo images.

Object Detection and Orientation Data: Data Details:

- Manually labelled 3D frame annotations
- Object types: Sedan, Van, Truck, Pedestrian, Cyclist, etc.

- **Challenges:**
  - Occlusion
  - Multiple objects per image

Perception in Autonomous Driving: Datasets

Go, change the world®

Overview of KITTI Dataset

Object Tracking Data: Data Details:

- 21 training sequences and 29 testing sequences

- **Main Targets:** Pedestrians and Cars

- **Purpose:** Tracking moving objects across frames.

Road Parsing Data: Data Details:

- 289 training images and 290 testing images

- **Road Types:**
  - Urban-unmarked, Urban-marked
  - Urban multiple-marked lanes

Perception in Autonomous Driving: Datasets

KITTI and Cityscapes datasets differ from traditionally computer vision datasets in the following ways:

- Due to the use of multiple sensors and 3D scanners, high-precision 3D geometry is available, hence high-quality ground truth;
- They are collected from real world, not synthesized nor collected in a controlled lab setting; and
- They contains data for various perception tasks, such as recognizing different obstacles, in autonomous driving

Perception in Autonomous Driving: Datasets

Several new datasets of autonomous driving haven been released in recent years. These include:

Audi Autonomous Driving Dataset (A2D2): https://www.a2d2.audi/a2d2/en.html,

- nuScenes: http://nuscenes.org,
- Berkeley DeepDrive: http://bdd-data.berkeley.edu,
- Waymo Open Dataset: http://waymo.com/open, and
- Lyft Level 5 Open Data: http://self-driving.lyft.com/level5/data.

Perception in Autonomous Driving: Detection

- Autonomous vehicles interact with diverse traffic participants:
  - Cars, Pedestrians, Lane dividers etc.

- Fast and reliable detection of objects is crucial.

- Object detection is a core problem in computer vision.

Object Detection Pipeline

- Typical Pipeline Steps:
  - **Preprocessing:** Prepare input images.
  - **Region of Interest Detector:** Identify candidate object regions.
  - **Classifier:** Classify detected objects.

- Challenges:
  - Variance in position, size, aspect ratio, orientation.
  - Need for real-time performance.

Feature Extraction and Invariant Representation

- Extract distinctive features for class separation.
- Build invariant representations for reliable detection.
- Real-time speed is critical.

Dalal and Triggs' HOG + SVM Algorithm (2005)

- **Process:**
  - Preprocess input image.
  - Compute HOG features over sliding detection windows.
  - Classify using Linear SVM.
- **Strengths:**
  - Effective appearance modeling.
  - Handles non-linear object articulation.

Perception in Autonomous Driving: Detection

Go, change the world®

Deformable Part Model (DPM)

•Proposed by Felzenszwalb et al. ---- Splits objects into simpler parts.

•**Key Features:**

- HOG feature pyramid.
- Part configuration constraints.
- Latent SVM for part positioning.

•**Strengths:** Reduces training data requirements.

LiDAR-based Object Detection

•**Alternative sensor-based detection.**

•**Strengths:** Accurate car detection.

•**Limitations:** Difficulty detecting pedestrians and cyclists.

•**Conclusion:** Sensor fusion (LiDAR + Camera) improves detection.

Perception in Autonomous Driving: Detection

Go, change the world®

Pedestrian Detection: Pedestrian detection is critical for safety.

- **Challenges:**
  - Unpredictable human behavior.
  - Partial occlusion.
  - Variance in appearance.

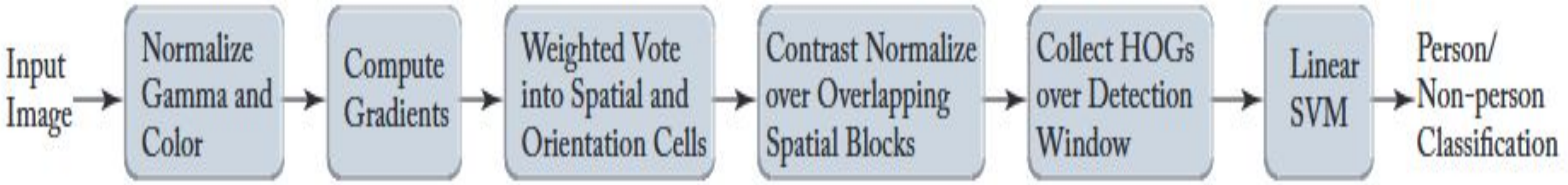
- State-of-the-art detectors use Convolutional Neural Networks (CNNs).

Importance of Sensor Fusion

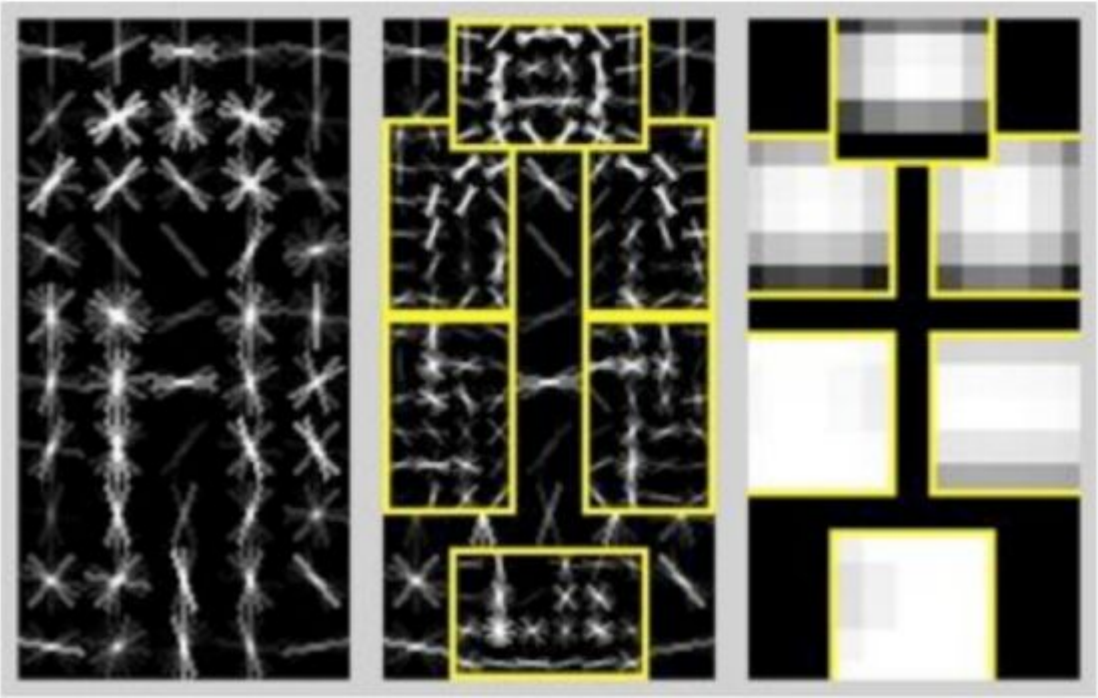
- Combining data from multiple sensors improves object detection accuracy.

- Cameras + LiDAR provide complementary information.

Perception in Autonomous Driving: Detection



HOG+SVM. Adapted from Dalal and Triggs



Deformable part model.

Introduction to Semantic Segmentation

- Enhances object detection for structured environmental understanding.

- Parses images into meaningful segments.

- **Traditional Approach:**
  - Graph labeling problem.
  - Conditional Random Fields (CRFs).

- **Challenges:**
  - Slower inference with larger images and features.
  - Difficulty capturing long-range dependencies.

Advanced Semantic Segmentation Techniques

- **Efficient CRF Algorithms:** Improve speed for dense predictions.
- **Multi-scale Features:** Capture details across different scales.
- **Contextual Reasoning:** Understand relationships between objects.
- Deep learning techniques have revolutionized segmentation.

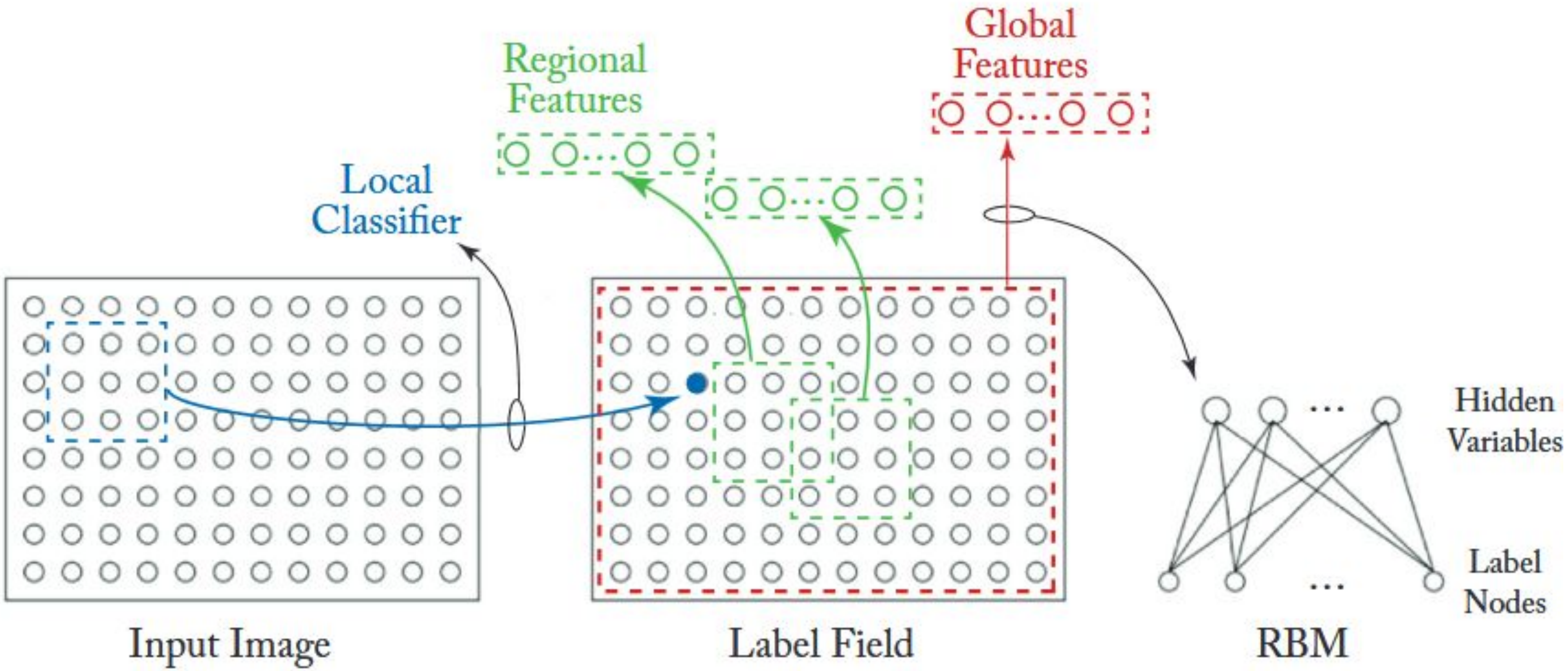
Perception in Autonomous Driving: Segmentation

Go, change the world®

Semantic segmentation of a scene in Zurich. Courtesy of Cityscapes Dataset

Perception in Autonomous Driving: Segmentation

Go. change the world®



Graphical model representation

Perception in Autonomous Driving: STEREO, OPTICAL FLOW, AND SCENE FLOW

Go, change the world®

- Autonomous vehicles navigate in a **3D world.**
- Accurate **3D spatial information** is essential for safe and efficient operation

- Key perception components:
  - **Depth perception**
  - **Color and texture recognition** – possible with normal **image (single) – unable to capture depth accurately**

Lidar Technology: High-precision depth data

- Outputs **sparse 3D point clouds**

- Effective in detecting obstacles and distances

- **Limitation:** Lacks spatial density

Stereo Camera Perception

- Mimics **human binocular vision**
- Two cameras capture images from **slightly different angles**
- Combines images to produce **depth information**
- Effective in understanding **3D spatial structure**

Method	Strengths	Limitations
Lidar	Accurate depth	Sparse point cloud
Single Camera	Dense color/texture	No depth data
Stereo Camera	Depth + color data	Computationally intensive

Perception in Autonomous Driving: STEREO, OPTICAL FLOW, AND SCENE FLOW

Go, change the world®

- Given an image pair from stereo camera (I_l, I_r), extracting stereo information is essentially a correspondence problem where pixels in the left image I_l are matched to pixels in the right image I_r based on a cost function.

- The assumption is that corresponding pixels map to the same physical point, and thus have the same appearance:

$$I_l(p) = I_r(p+d),$$

where p is a location in left image and d is the disparity.

Feature-based methods replace pixel values with more distinctive features ranging from simple ones like edge and corner to sophisticated manually designed features

- Area-based methods exploits spatial smoothness according to the following equation:

$$d(x, y) \approx d(x + \alpha, y + \beta)$$

for fairly small (α, β). Solving for d thus becomes a minimization problem:

$$min_d\, D(p, d) = min_d \sum_{q \in N(p)} ||I_r\,(q + d) - I_l\,(q)||.$$

This can generate dense outputs with higher computation cost.

The Correspondence Problem

- **Definition:** Matching corresponding points between two images.

- **Formulation:** Can be approached as an **Optimization Problem**.

- **Key Methods:** Feature-Based Methods and Area-Based Methods

Local vs Global Methods: Local Methods:

  - Compute disparity (d) using local information.
  - Examples: Feature-based, Area-based.

- **Global Methods:**
  - Approach as an **Energy Minimization Problem**.
  - Based on:
    - Constant Appearance Assumption
    - Spatial Smoothness Constraints

Techniques for Global Optimization

- **Variational Methods:** Solve energy minimization equations.
- **Dynamic Programming:** Break problem into simpler sub-problems.
- **Belief Propagation:** Probabilistic message-passing algorithm.

Semi-Global Matching (SGM): A well-known stereo matching algorithm.

- **Key Features:**
  - Energy function calculated along multiple 1D paths.
  - Smoothness terms enforced across pixels.
- **Strengths:**
  - Theoretically supported and Computationally efficient.

Semi-Global Matching (SGM): SGM aims to calculate the **disparity map** (difference in pixel positions between two stereo images) efficiently while maintaining global consistency in results.

Energy Minimization Framework:

SGM formulates the stereo matching problem as an **energy minimization problem** with two main components:

- **Data Term:** Measures how well a pixel in one image matches a corresponding pixel in the other image based on intensity or feature similarity.

- **Smoothness Term:** Penalizes large disparity differences between neighboring pixels to ensure spatial smoothness.

The **energy function** is defined as:

$$E(d) = \sum_{p} C(p, d_p) + \sum_{(p,q)\in N} P_1 \cdot |d_p - d_q| = 1 + P_2 \cdot |d_p - d_q| > 1$$

Where:

- $C(p, d_p)$: Matching cost for pixel p with disparity d_p
- P_1: Penalty for small disparity differences (encourages smooth transitions)
- P_2: Penalty for larger disparity jumps (preserves edges)

Semi-Global Optimization

Instead of fully global optimization, which is computationally expensive, SGM performs energy minimization **along multiple 1D paths** across the image.

- Disparities are aggregated along paths in **multiple directions** (typically 8 or 16 paths).

- This reduces computational overhead while maintaining high accuracy.

Advantages of SGM:

- Accuracy: Produces high-quality disparity maps with preserved edges.

- Efficiency: Significantly faster than traditional global optimization methods.

- Flexibility: Works well with real-world data, including noise and textureless regions.

Perception in Autonomous Driving: OPTICAL FLOW,

What is Optical Flow?

- **Definition:** 2D motion of intensities between two images.

- **Key Difference from Stereo Matching:**
  - Stereo: Geometry dominates disparity.
  - Optical Flow: Captures temporal changes.

$$I_t(p) = I_{t+1}(p + d)$$

- **Optical Flow:**
  - Image pairs captured at different times.
  - Motion, lighting, reflections, and transparency affect results.
  - Appearance constancy often violated.

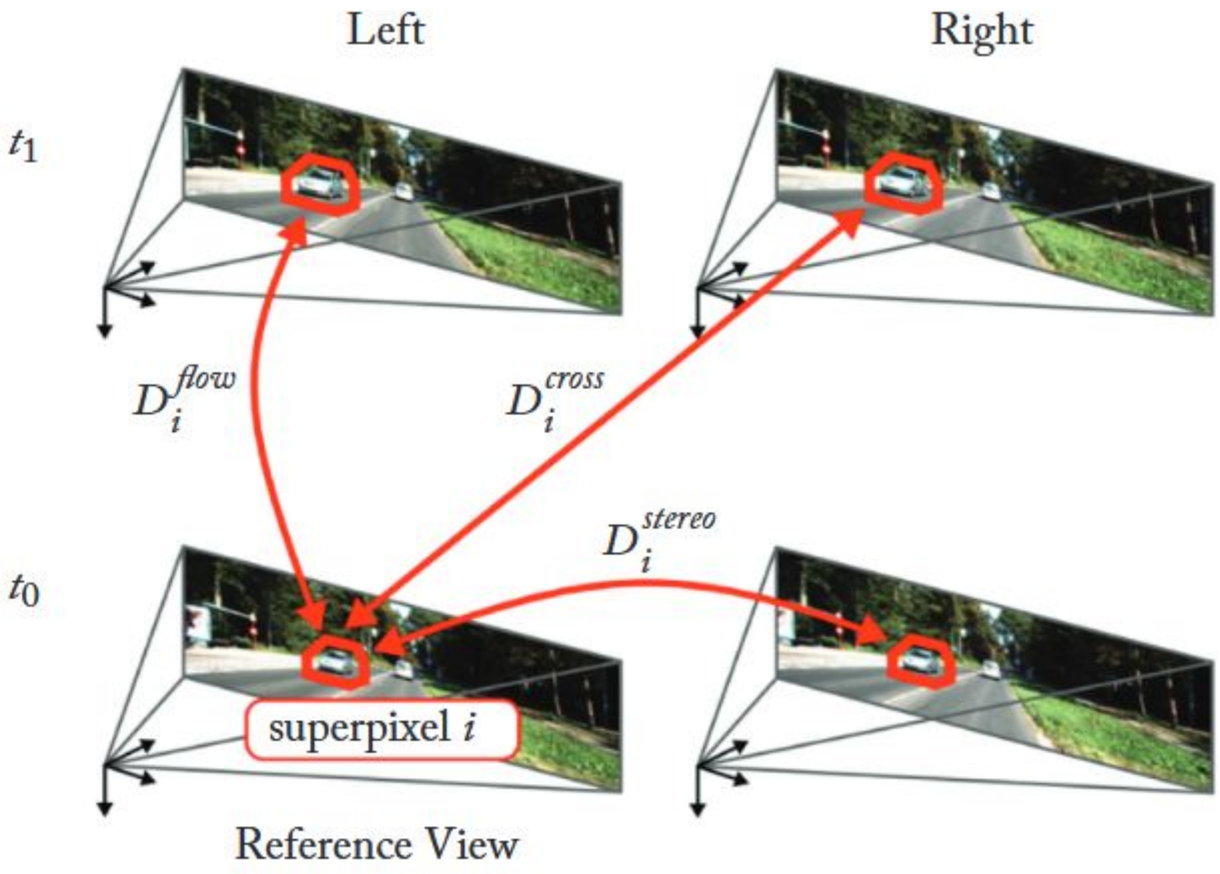
Challenges in Optical Flow

- **Appearance Constancy Violation:**
  - Changes in lighting, reflections, and transparency.
- **Aperture Problem:**
  - One constraint with two unknowns.
  - Ambiguity in motion direction.
- **Smoothness Constraint:**
  - Helps resolve ambiguity.

Perception in Autonomous Driving: SCENE FLOW

Go, change the world®

- Scene flow estimation is based on two consecutive stereo image pairs - where the correspondence produces not only the 3D position of points but also their 3D motion between time intervals.



Perception in Autonomous Driving: SCENE FLOW

Go, change the world®