Artificial Intelligence in Digital Car Design for Pedestrian Safety

Chen Xu, Katharina Oberst, Shubham Khatri, Willem van Hove
TUM Data Innovation Lab

Mentor: Laure Vuaille, PhD candidate
Co-Mentor: Dr. Nada Sissouno

Project Lead: Dr. Ricardo Acevedo Cabra
Supervisor: Prof. Dr. Massimo Fornasier

TEAM

Chen Xu

Electrical Engineering and Information Technology

Katharina Oberst

Management and Technology

Shubham Khatri

Computational Science and Engineering

Willem van Hove

Computational Science and Engineering

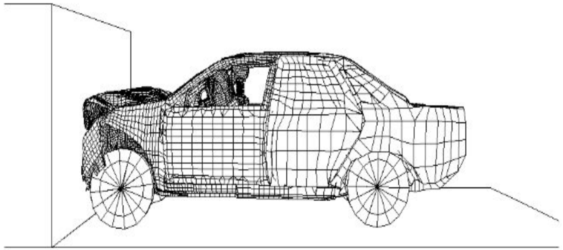
AGENDA

1. Motivation
2. Data
3. Data preprocessing
4. Car matching
5. Car comparison
6. Demo

MOTIVATION

Traditional car design:
- Time-consuming
- Rely on the experience of human expert

Digital car design:
- Computer assisted
- Simulation

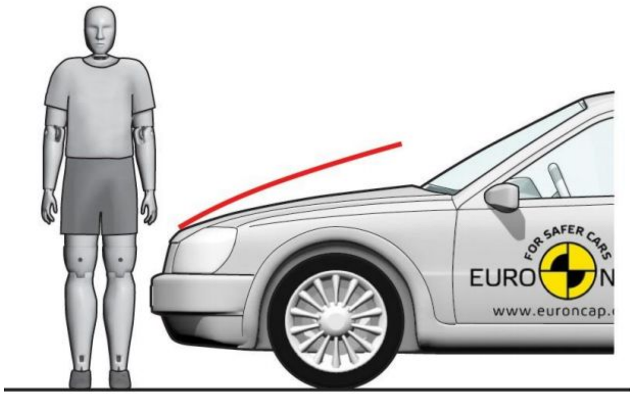


Marzougui D, Kan C D, Bedewi N E. Development and validation of an NCAP simulation using LS-DYNA3D[J]. 1997.

MOTIVATION

Pedestrian safety

- Quantify the influence of the head injury
- Decrease Head Injury Criterion (HIC) Value



Pedestrian Testing Protocol - Euro NCAP

MOTIVATION

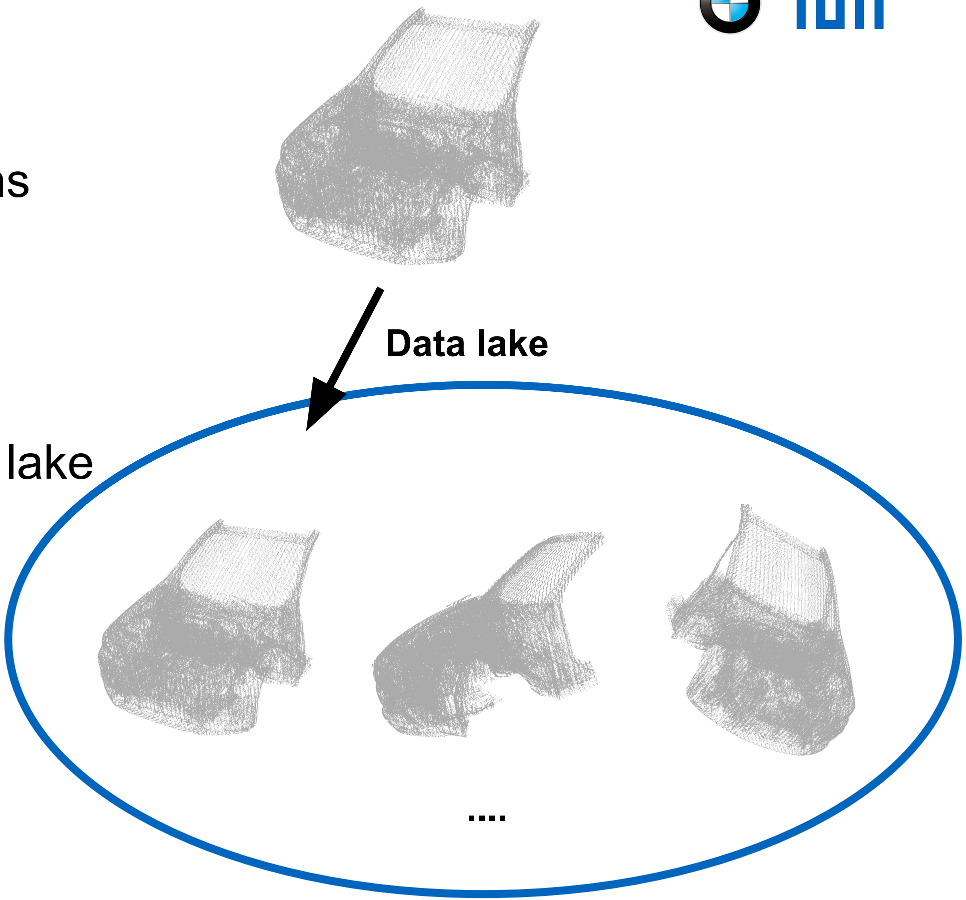
Analysis of data derived from simulations

Project goal:

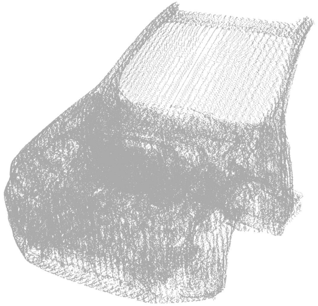
- Given a new car
- Find the most similar car from data lake
- Detect changes between cars

Future goal:

- Prediction of HIC value



DATA

	Difference	Similarity
Data set 1	**Data lake**	● Point clouds ● High dimensional ● Several variables (Material, geometric) 
	173	
	Changes undocumented	
Data set 2	**Validation**	
	10	
	Changes documented	

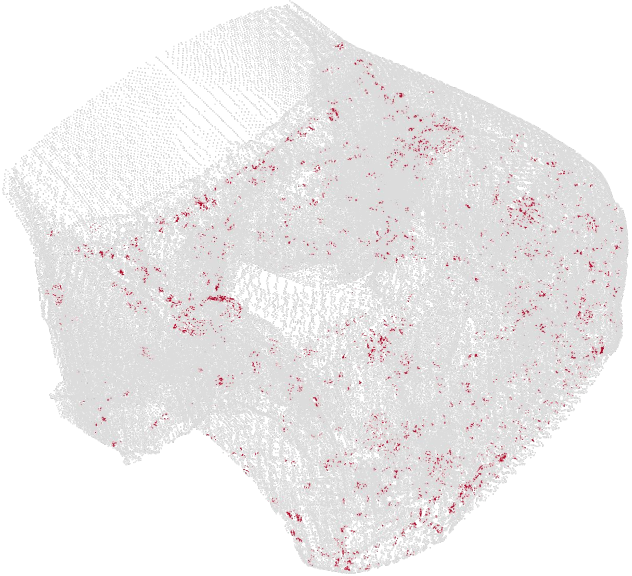
DATA - DESCRIPTION

Definition: dense/sparse variables

- Dense variable: non-zero dominant
- Sparse variable: zero dominant

Ex: Sparse variables

	Number of points	Percentage
Non-Zero	7210	2%
Zero	275580	98%



Non-zero points in red

DATA - DESCRIPTION

- Approximately 3∗10^5 points
- Coordinates: X, Y, Z

Variables	Dense	Sparse	Sparsity
var_0	✗		0.78%
var_1	✗		1.23%
var_2	✗		5.82%
var_3		✗	99.90%
var_4		✗	95.97%
var_5		✗	69.53%
var_6		✗	99.95%

DATA - SET 2

- Validation data set
- Car 1 as the reference
- Changes applied to 9 point clouds

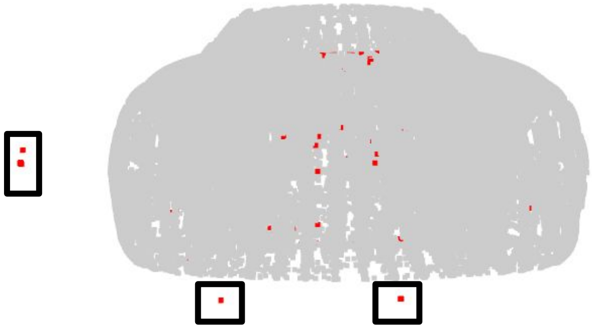
Ex: To check the **material** change ⟶ compare 1 and 9

Car number	Changes applied in comparison to car 1
1	no changes
...	...
9	material of front hood inner reinforcement

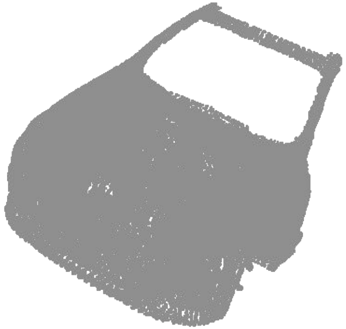
PROGRAM OVERVIEW

- New car
- **Data preprocessing**
  - Data cleaning
  - Data alignment and preparation
- Car matching
- Car comparison
- Changes visualization

DATA PREPROCESSING - CLEANING

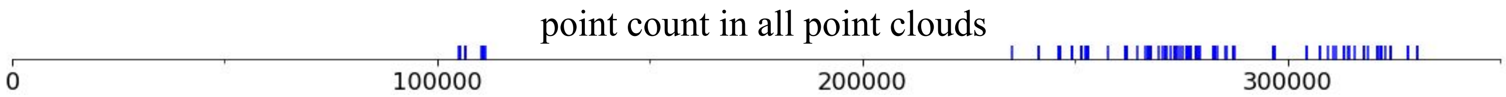


Outlier detection



Faulty data

Undersampled point cloud



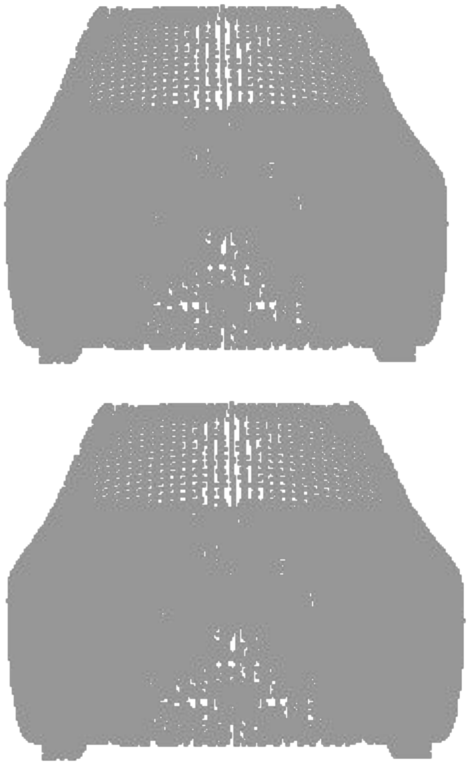
DATA PREPROCESSING - RIGID BODY TRANSFORMATION

Problem: False change detection

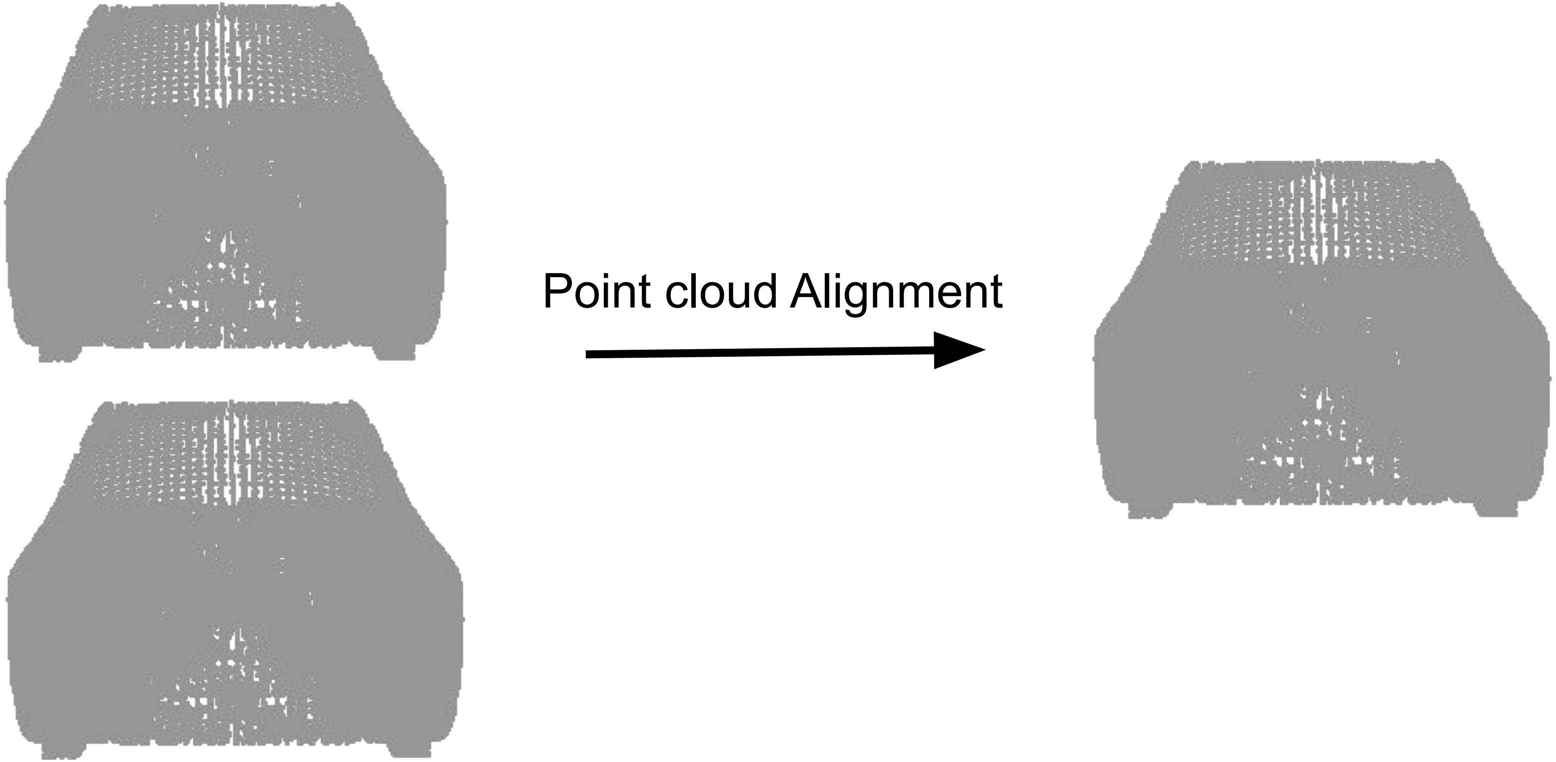
- **Not** a change:
  - Translation
  - Rotation
  - Scaling

Solution: Iterative Closest Point (ICP)

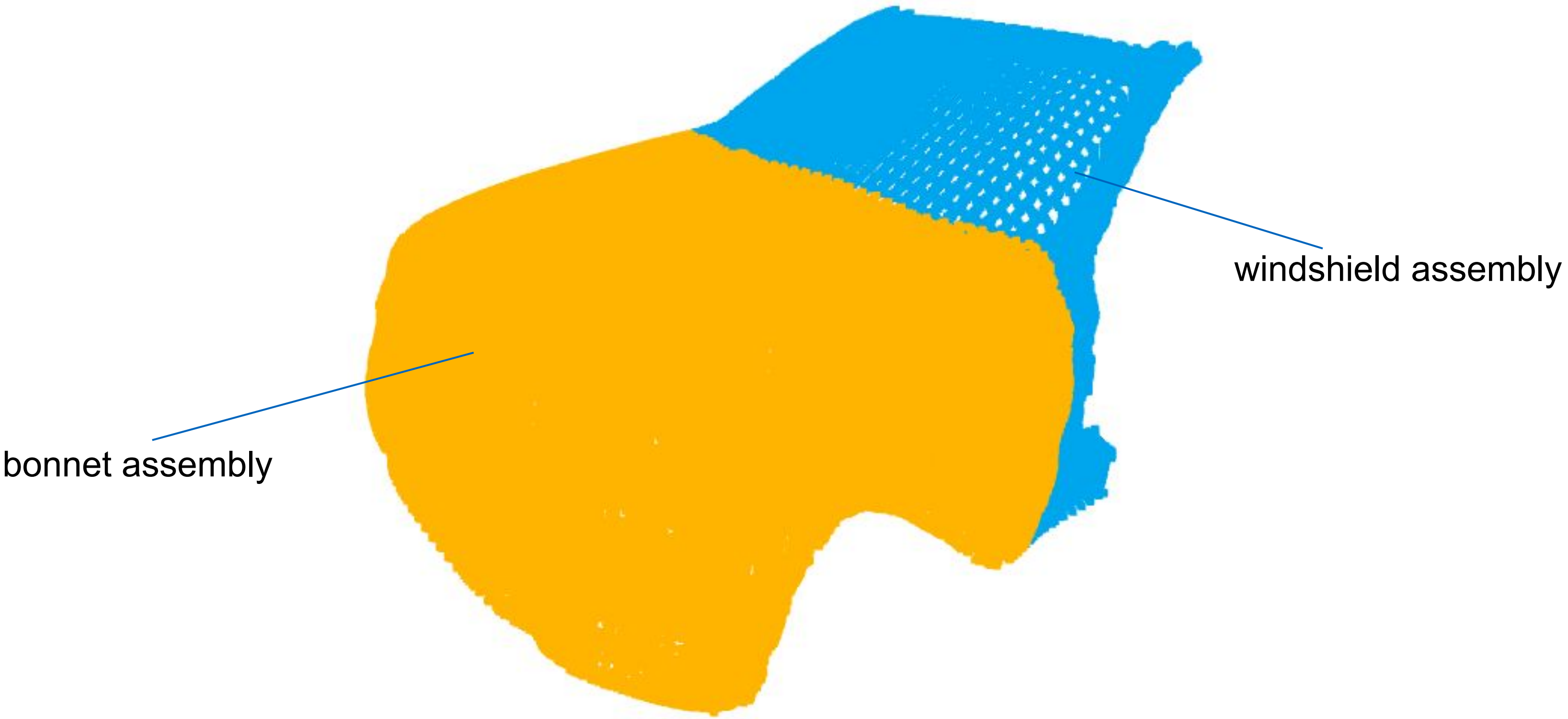
- Find corresponding points
- Minimize distance



DATA PREPROCESSING - ICP

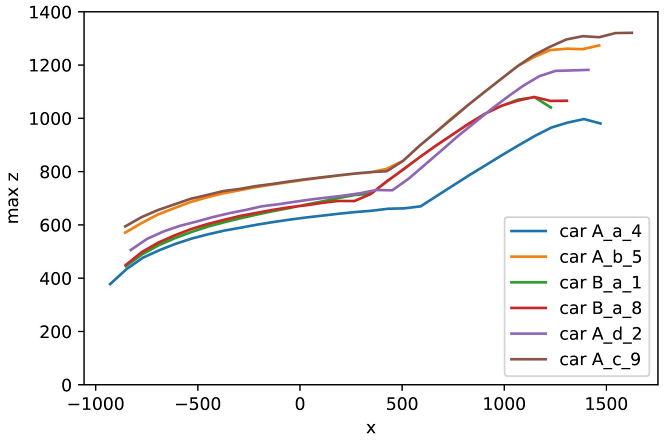


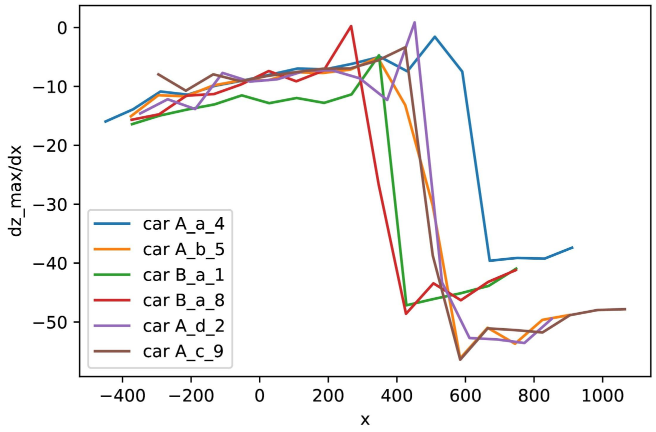
DATA PREPROCESSING - SPLITTING OF CAR



DATA PREPROCESSING - SPLITTING OF CAR

- Split in bonnet- and windshield assembly
- Important for pedestrian safety
- Change of slope of “max z”
- First derivative: Discontinuous





PROGRAM OVERVIEW

New point cloud

Data preprocessing

Car matching

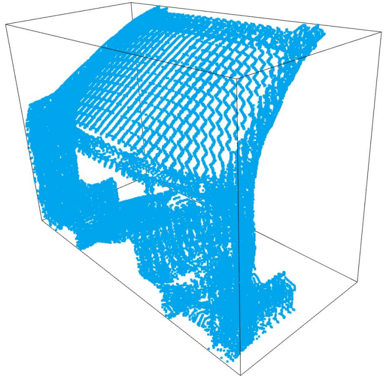
- High level feature extraction
- Prediction in database

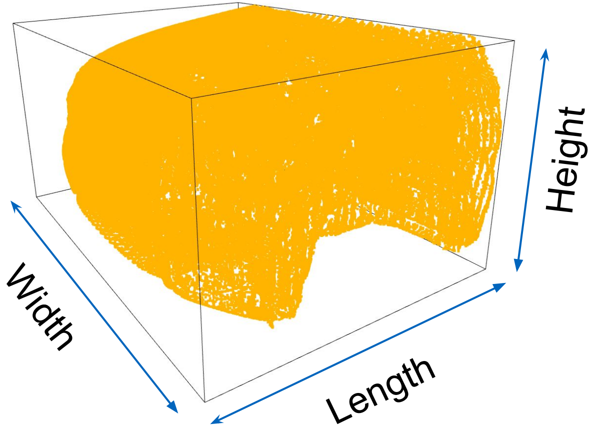
Car comparison

Changes visualization

CAR MATCHING - HIGH LEVEL FEATURES

Total Length	Total Width
Total Height	Hood Length
Hood Width	Hood Height
Windshield Length	Windshield Width
Windshield Height	Mass





CAR MATCHING - ALGORITHM

Aim: Find same model & family

Small dataset: Modified KNN

$$t = \underset{t \in \mathcal{D}}{\arg\min} \; \{a \cdot \|f_t - f_s\| + b \cdot \|N_t - N_s\|\}$$

$$\begin{aligned} f_i &: \; High \;\; level \;\; features \\ N_i &: \; Number \;\; of \;\; points \\ a, b &: \; Weights \end{aligned}$$

CAR MATCHING - ALGORITHM

Large dataset:

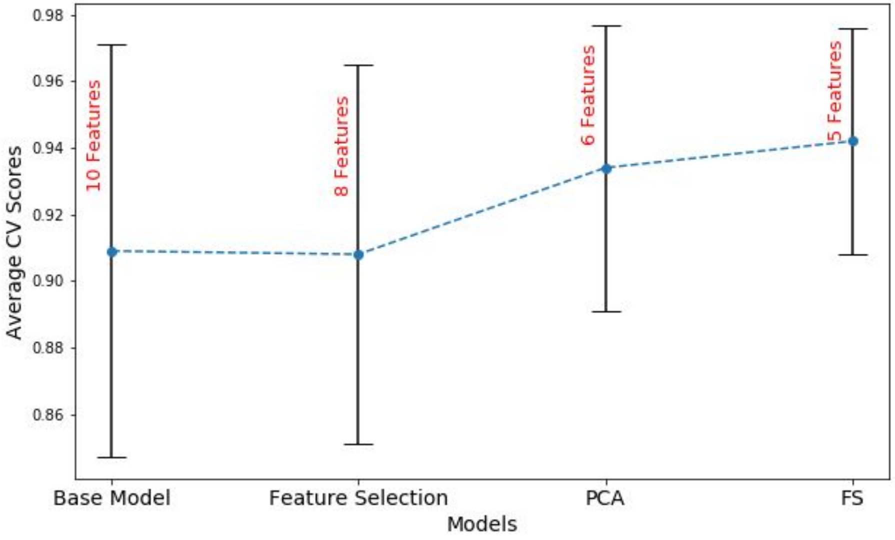
- Issue:
  - Computationally expensive $\mathcal{O}(k \cdot n)$
  - Memory requirement
- Solution: Random Forest Classifier
  - Computationally inexpensive $\mathcal{O}(k \cdot t)$
  - Storage efficient

CAR MATCHING - FEATURE ANALYSIS

Problem: High dimension & correlation

Solution: 3 Step dimensionality reduction

1. Feature Selection
   - Remove correlated feature

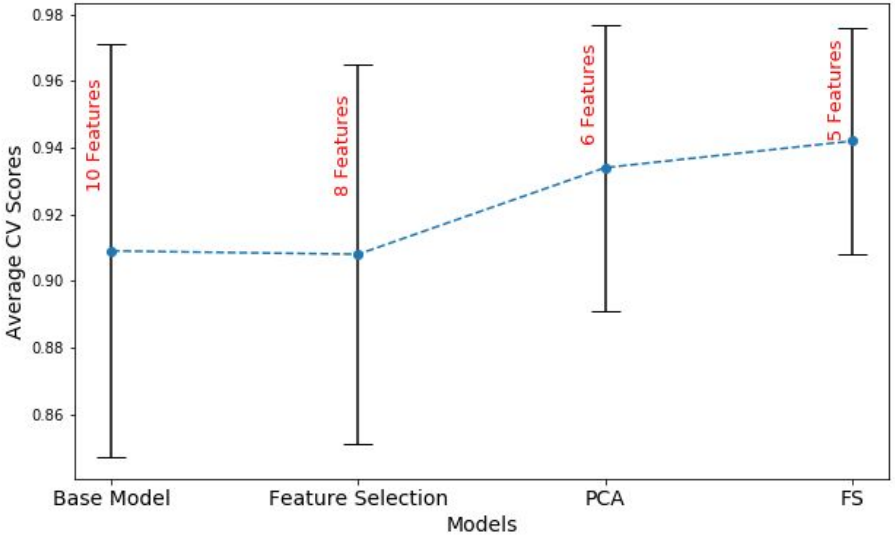


CAR MATCHING - FEATURE ANALYSIS

Problem: High dimension & correlation

Solution: 3 Step dimensionality reduction

1. Feature Selection
   - Remove correlated feature
2. Principal Component Analysis (PCA)
   - Features with 99.9% variance

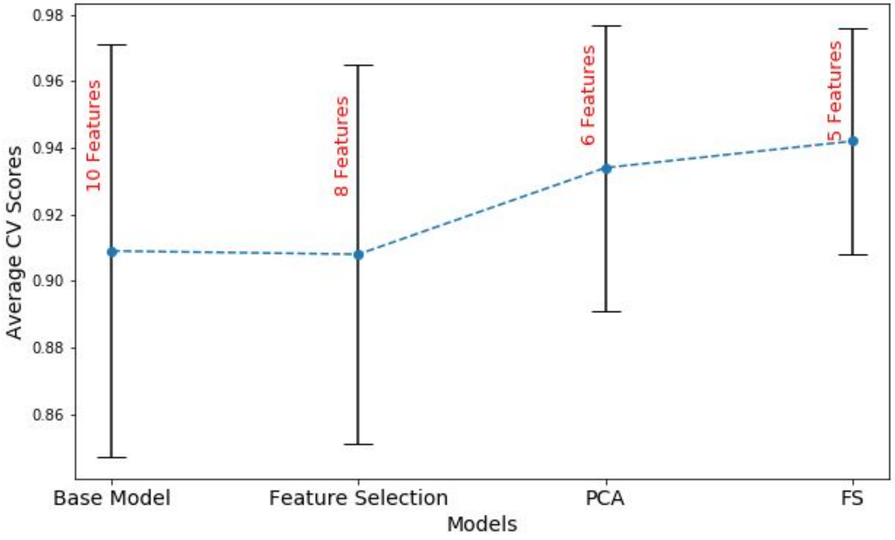


CAR MATCHING - FEATURE ANALYSIS

Problem: High dimension & correlation

Solution: 3 Step dimensionality reduction

1. Feature Selection
   - Remove correlated feature
2. Principal Component Analysis (PCA)
   - Features with 99.9% variance
3. Feature Sensitivity using ANOVA F-Test (FS)
   - Best explaining features



PROGRAM OVERVIEW

New point cloud

Data preprocessing

Car matching

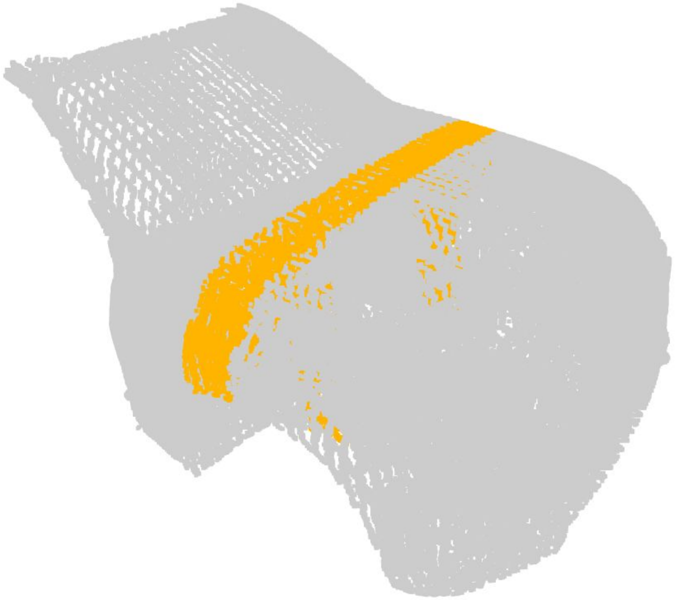
Car comparison

- Dense variable changes
- Sparse and geometric variable changes

Changes visualization

DATA PREPROCESSING - POINT CLOUD SLICING

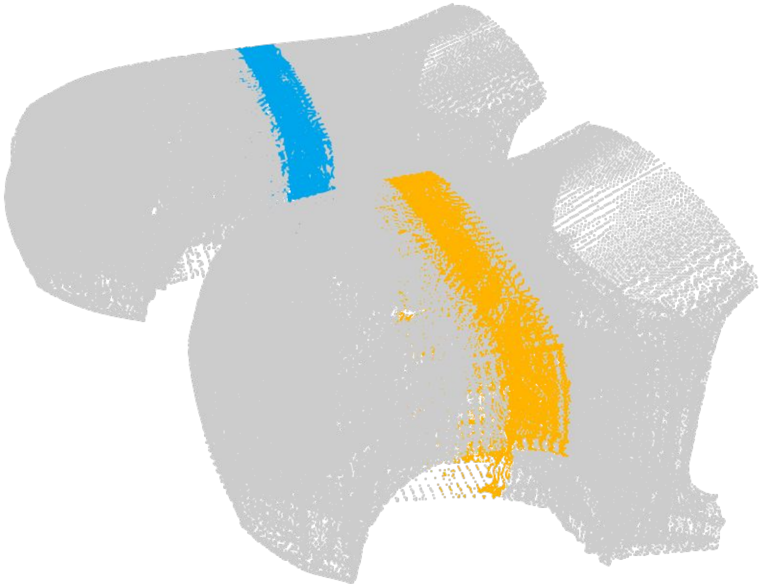
- Bonnet assembly sliced
- Along length
- Enables:
    - Localized view
    - Localized analysis



slice of point clouds

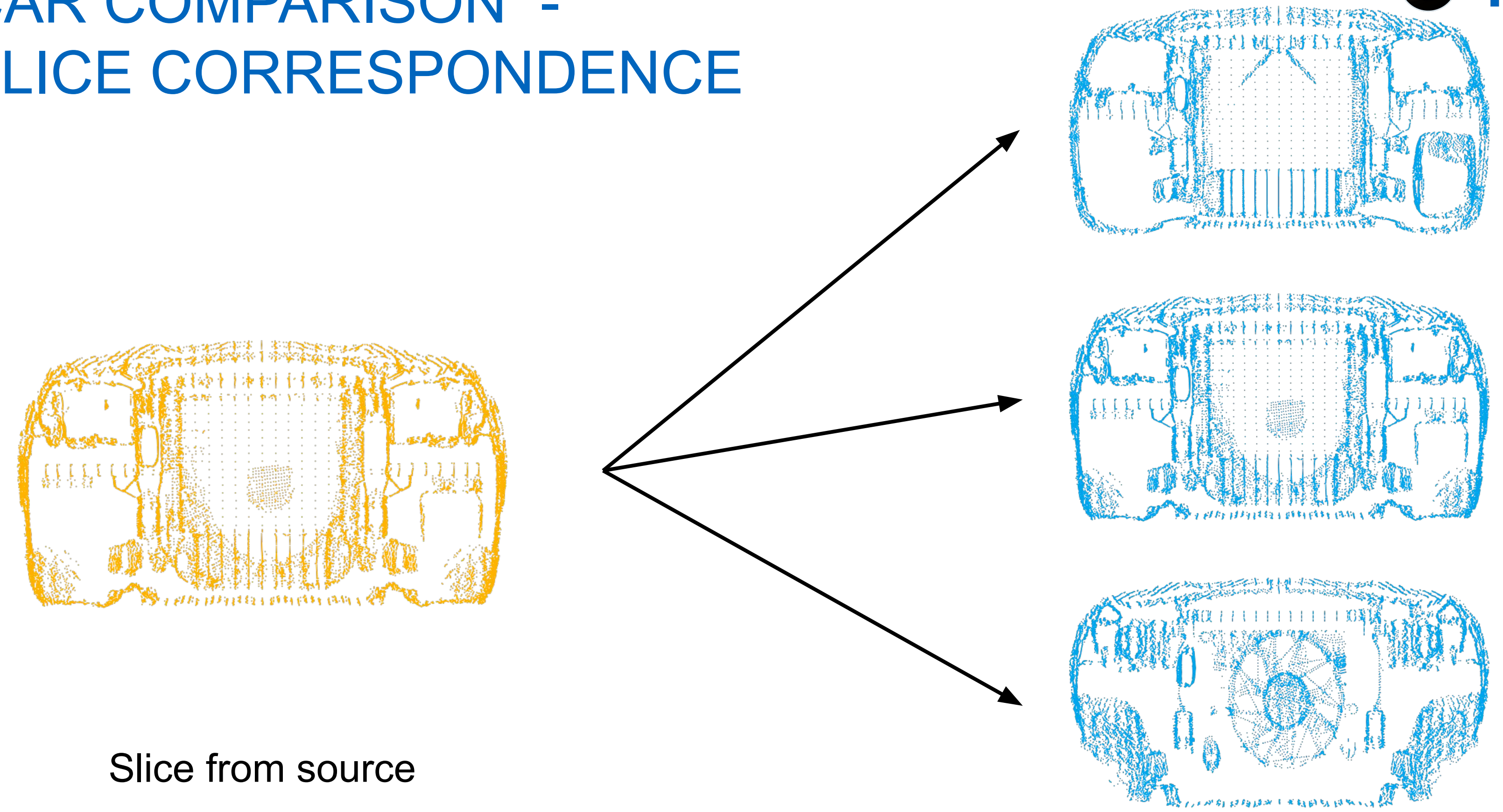
CAR COMPARISON - SLICE CORRESPONDENCE

- Match source and target slices
- Slices with similar geometry are matched
- Useful when car has elongated bonnet structure



Elongated point cloud

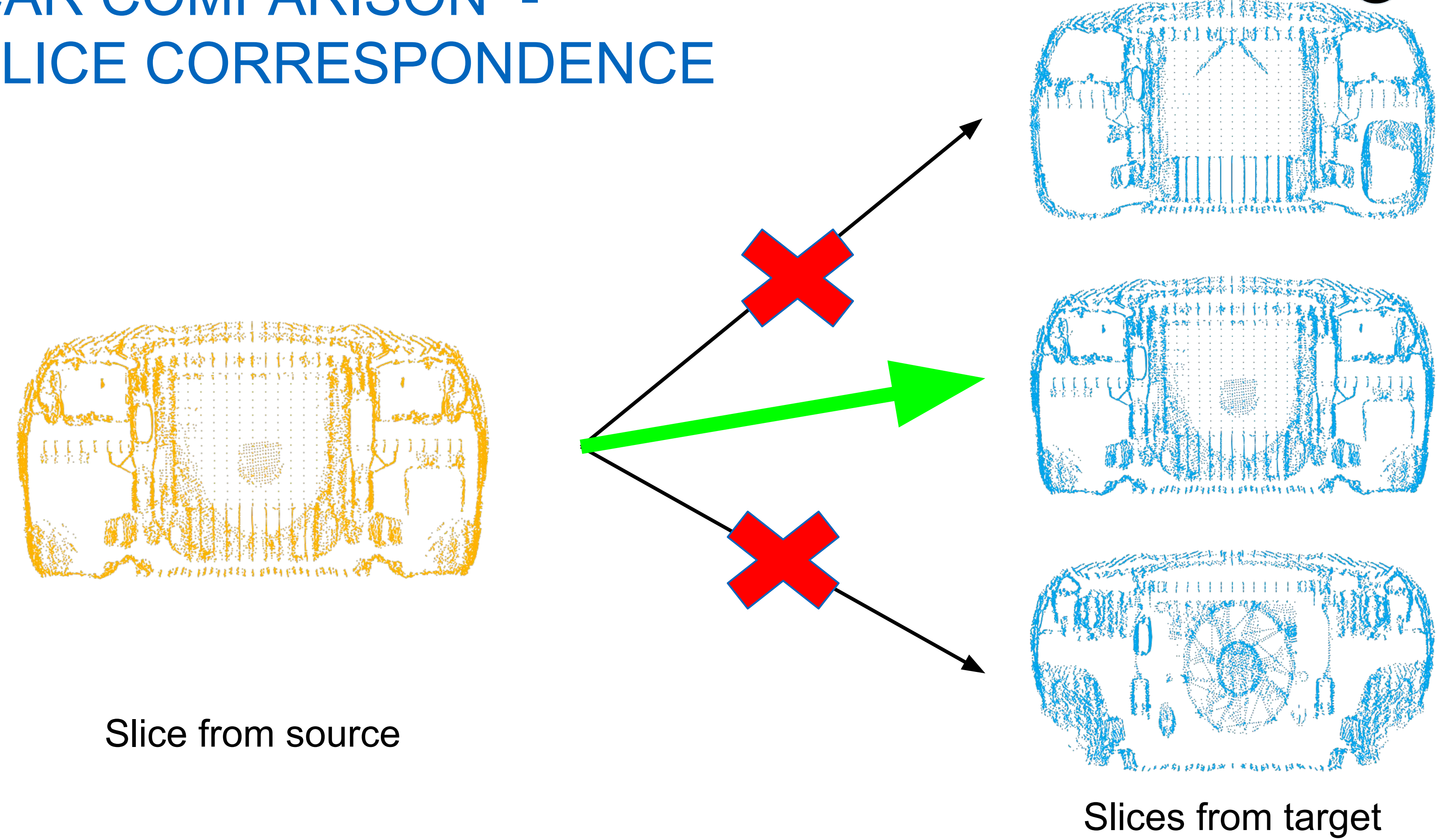
CAR COMPARISON - SLICE CORRESPONDENCE



Slice from source

Slices from target

CAR COMPARISON - SLICE CORRESPONDENCE

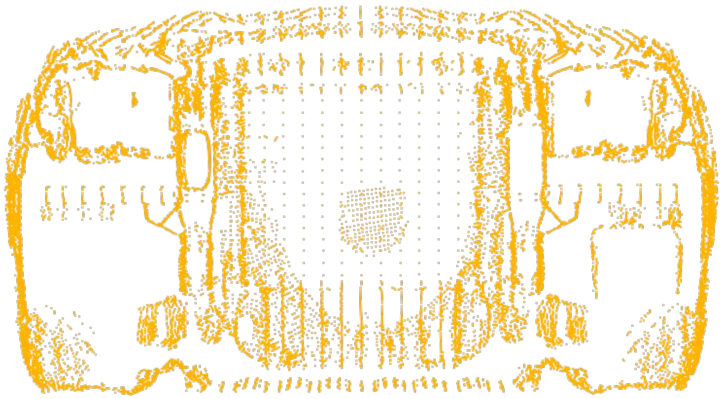


CAR COMPARISON

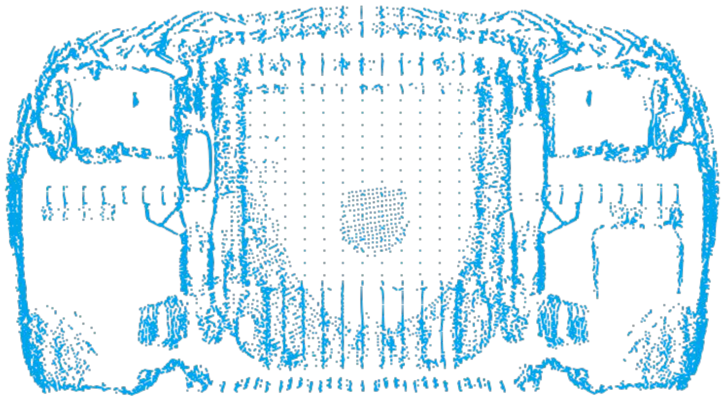
- Similarity assumption
- Two methods:
  - Dense variables: point mapping
  - Sparse variables and geometry: quadtree

CAR COMPARISON - DENSE VARIABLES

- Aim: detect material change
- Visible in dense variable data



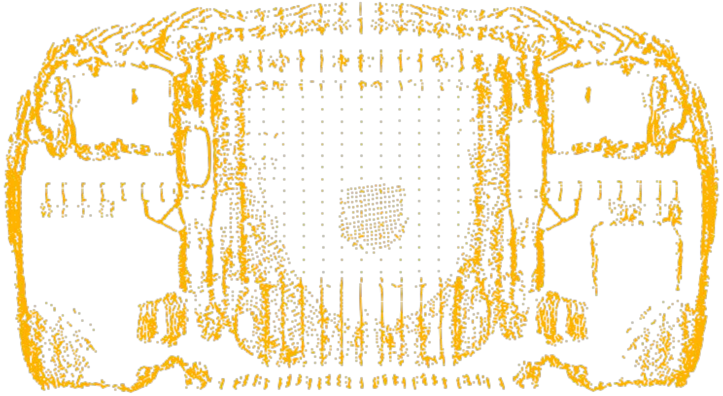
Source slice



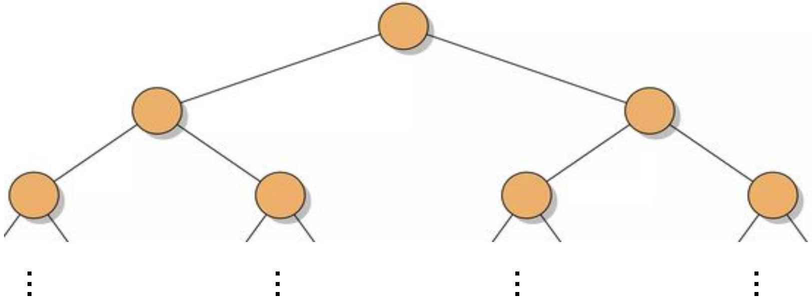
Target slice

CAR COMPARISON - DENSE VARIABLES

- Mapping: source point $\longrightarrow$ nearest neighbor target point
- K dimensional tree



Source slice

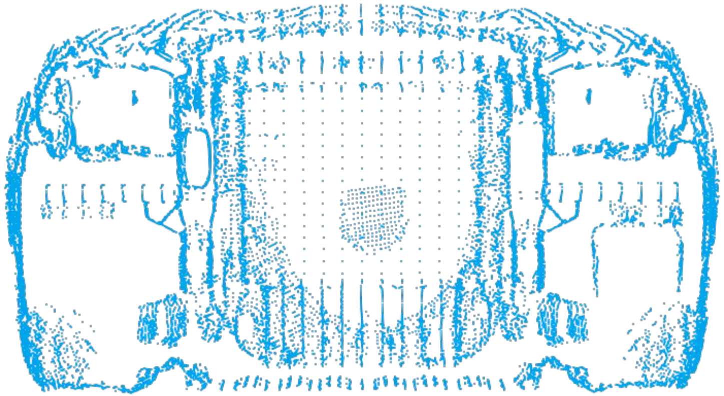


Target slice

CAR COMPARISON - QUADTREE

- Localization:
  - Sparse
  - Geometric
- Project x coordinates to plane: 3D → 2D
- Recursive subdivision into four quadrants
- Tree nodes have zero or four children
- **Traversal** of the tree:

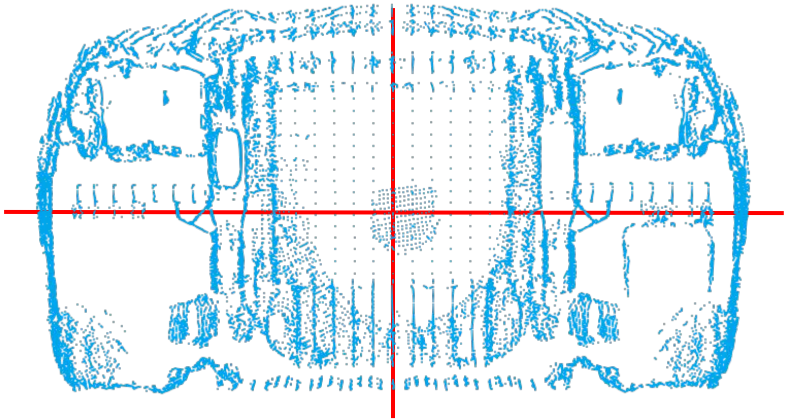
  subdivision of space



CAR COMPARISON - QUADTREE

- Localization:
  - Sparse
  - Geometric
- Project x coordinates to plane: 3D → 2D
- Recursive subdivision into four quadrants
- Tree nodes have zero or four children
- **Traversal** of the tree:

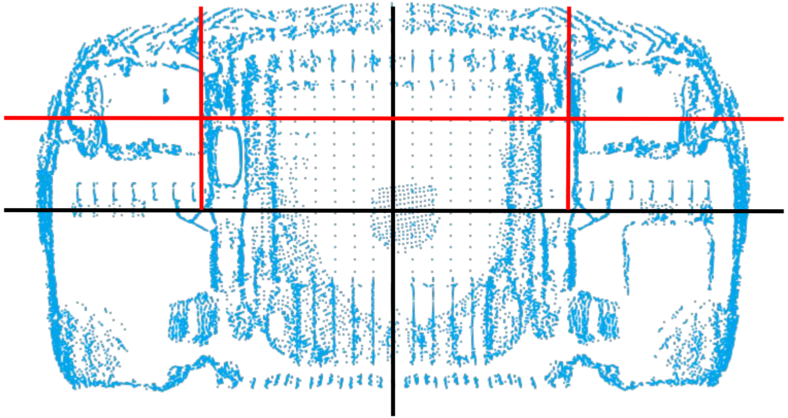
  subdivision of space



CAR COMPARISON - QUADTREE

- Localization:
  - Sparse
  - Geometric
- Project x coordinates to plane: 3D → 2D
- Recursive subdivision into four quadrants
- Tree nodes have zero or four children
- **Traversal** of the tree:

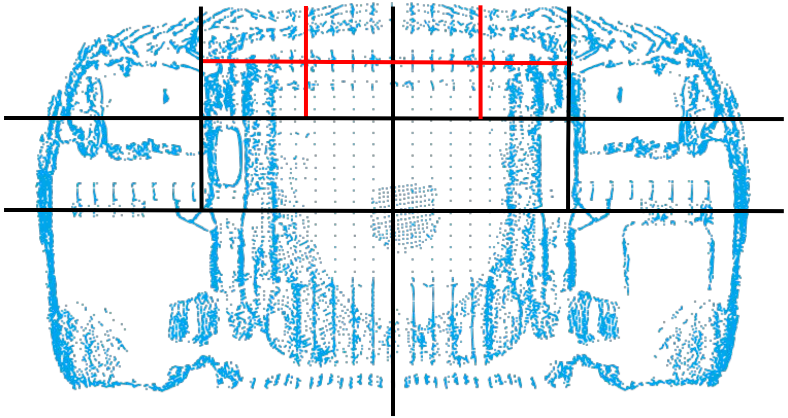
  subdivision of space



CAR COMPARISON - QUADTREE

- Localization:
  - Sparse
  - Geometric
- Project x coordinates to plane: 3D → 2D
- Recursive subdivision into four quadrants
- Tree nodes have zero or four children
- **Traversal** of the tree:

  subdivision of space



CAR COMPARISON - SPARSE AND GEOMETRIC VARIABLES

Point cloud A, Point cloud B:

- **Sparse distance**:

$$d_S(A,B) = \sum_{a\in A} a - \sum_{b \in B} b$$

- **Hausdorff distance**:

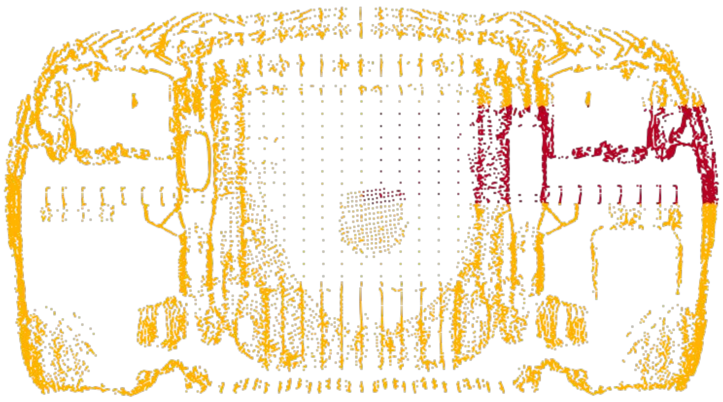
$$d_H(A,B) = \max\left\{ \sup_{a\in A} \inf_{b\in B} \mathrm{d}(a,b), \sup_{b\in B} \inf_{a\in A} \mathrm{d}(a,b) \right\}$$

- **Split criterion**

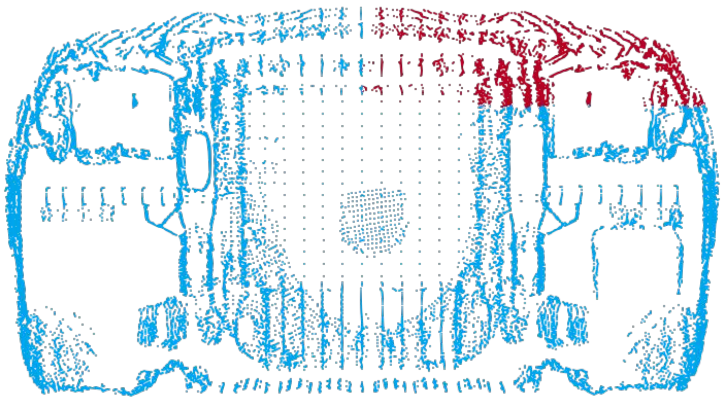
$$\exists_{\mathrm{S}} \; (d_{\mathrm{S}} \geq \epsilon_{\mathrm{S}}) \vee (d_{\mathrm{H}} \geq \epsilon_{\mathrm{H}}) \vee (\mathrm{depth} \geq \mathrm{max.\ depth})$$

CAR COMPARISON - SPARSE VARIABLES

- Quadtree limitation
- Minimize: minimum traversal depth, traversal with all distances



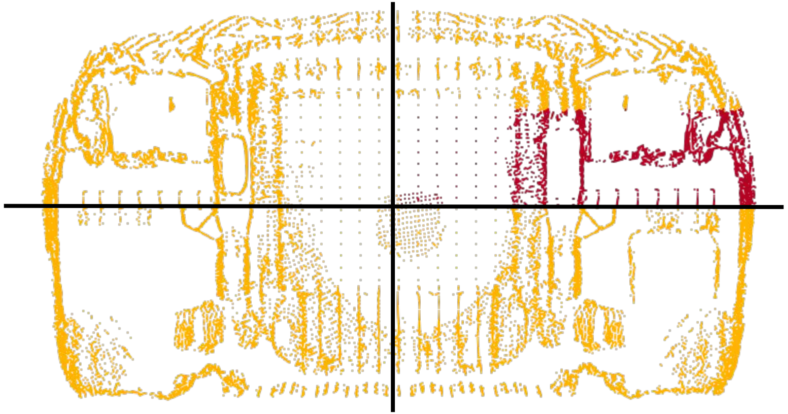
Source slice



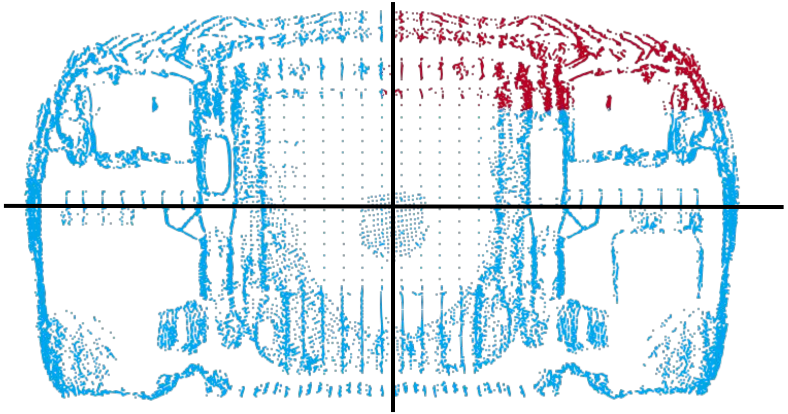
Target slice

CAR COMPARISON - SPARSE VARIABLES

- Quadtree limitation
- Minimize: minimum traversal depth, traversal with all distances



Source slice



Target slice

PROGRAM OVERVIEW

New point cloud

Data preprocessing

Car matching

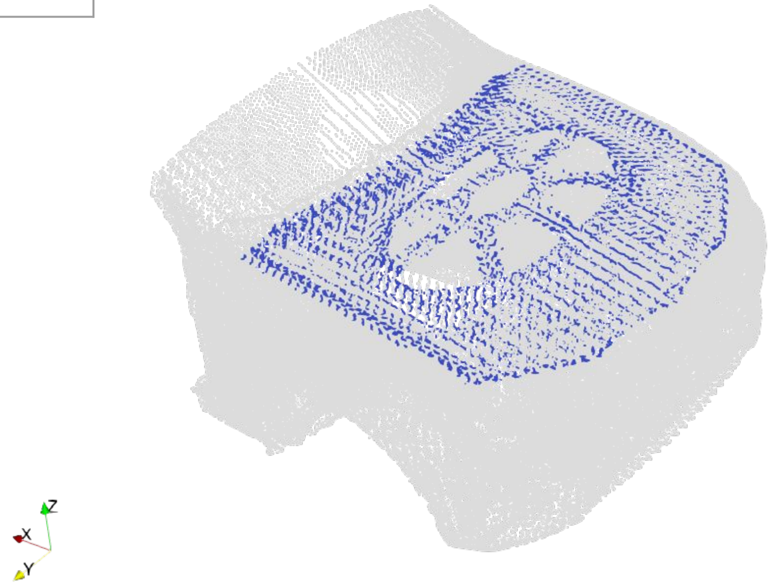
Car comparison

Changes visualization

- Results
- Demo

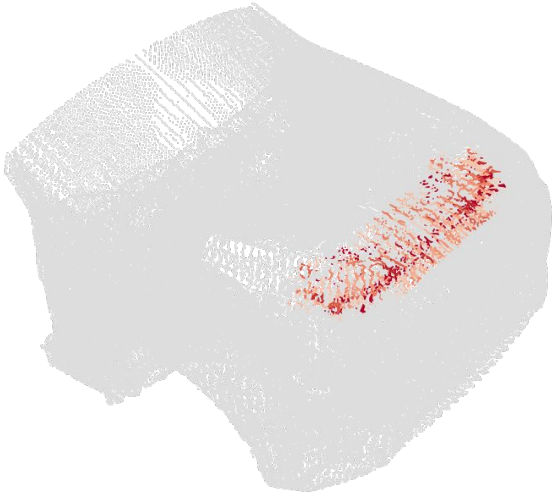
CAR COMPARISON - DENSE VARIABLES

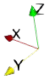
Car number	Changes applied in comparison to car 1
6	Material of front hood innerskin; location/level of *var_6*

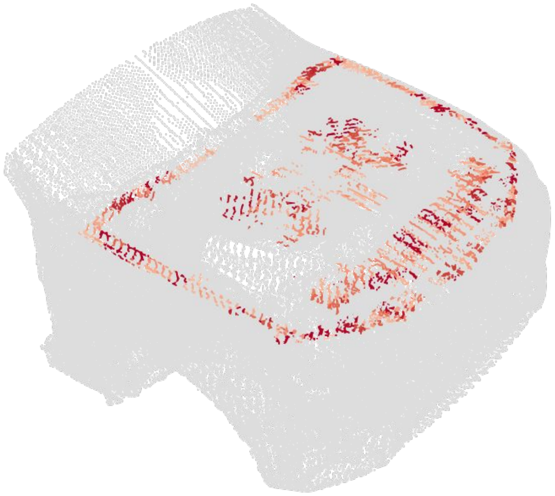


CAR COMPARISON - GEOMETRY

Car number	Changes applied in comparison to car 1
2	Material and thickness of front hood inner reinforcement
4	Thickness of front hood outer skin; subtraction of new items for *var_6*





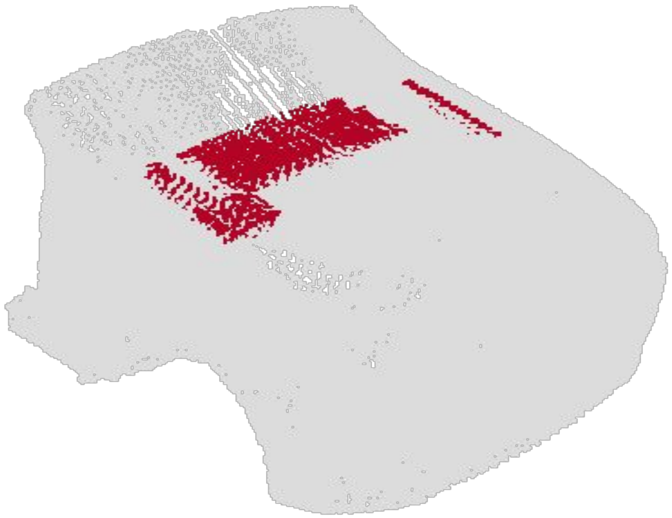


car 2

car 4

CAR COMPARISON - SPARSE VARIABLES

Car number	Changes applied in comparison to car 1
6	Material of front hood innerskin; location/level of *var_6*
4	Thickness of front hood outer skin; subtraction of new items for *var_3*



car 6

car 4

CONCLUSION

Pro

- Algorithm detects changes between two point clouds
- Local and global changes
- Geometric and non-geometric changes
- Saves time

Limitations

- Not all changes are being detected (sparse)
- Changes are correlated

OUTLOOK

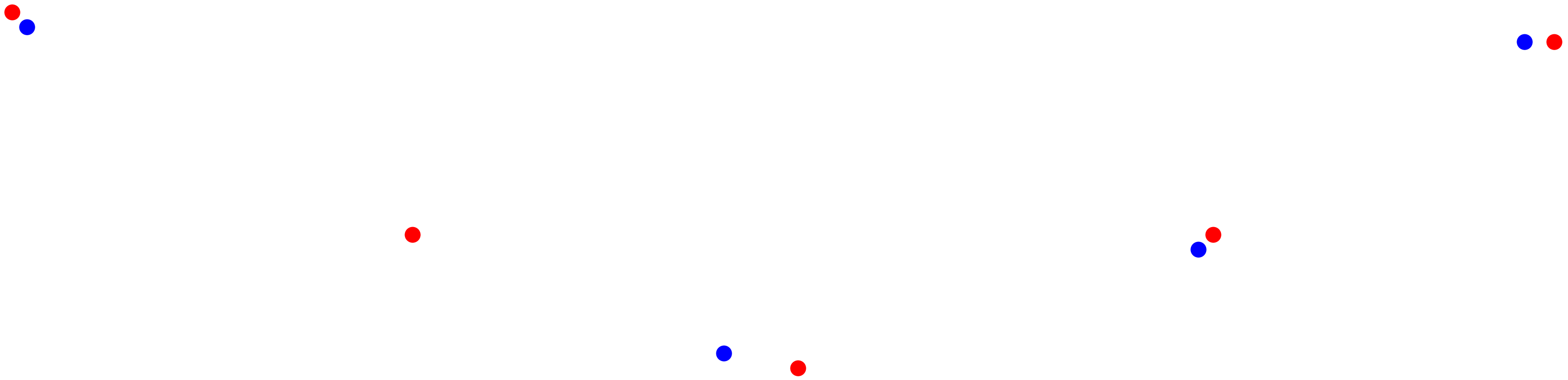
Improve geometric change detection	Predict HIC value	Recommendation during design process
Multi dimensional change measures Coherent Point Drift	Regression model Change classification	Extend regression model Suggest changes

DEMO

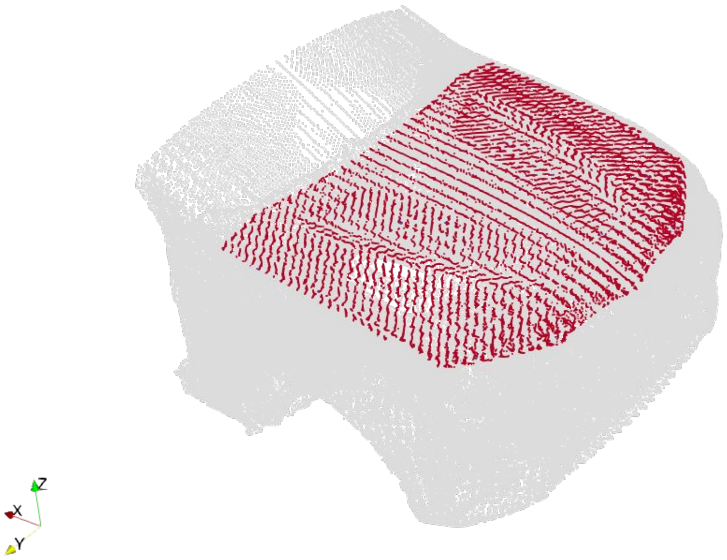
Thank you.
Questions?

HAUSDORFF DISTANCE

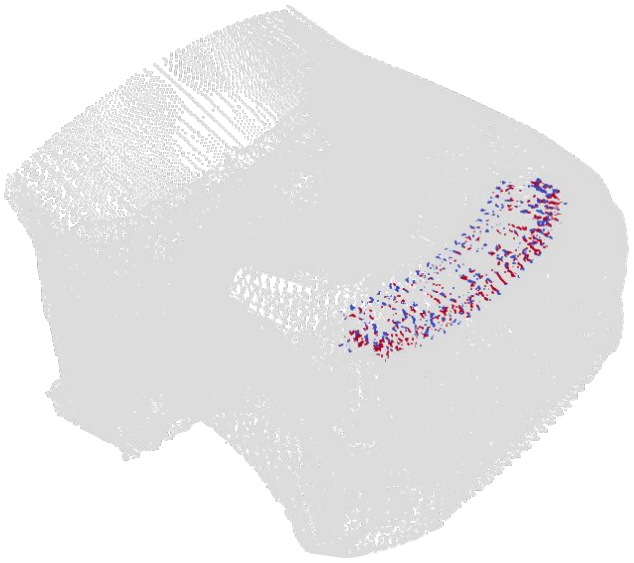
$$d_H(A,B) = \max\left\{ \sup_{a\in A} \inf_{b\in B} \mathrm{d}(a,b), \sup_{b\in B} \inf_{a\in A} \mathrm{d}(a,b) \right\}$$



COMPARISON DENSE POINTS



Outer hood material change

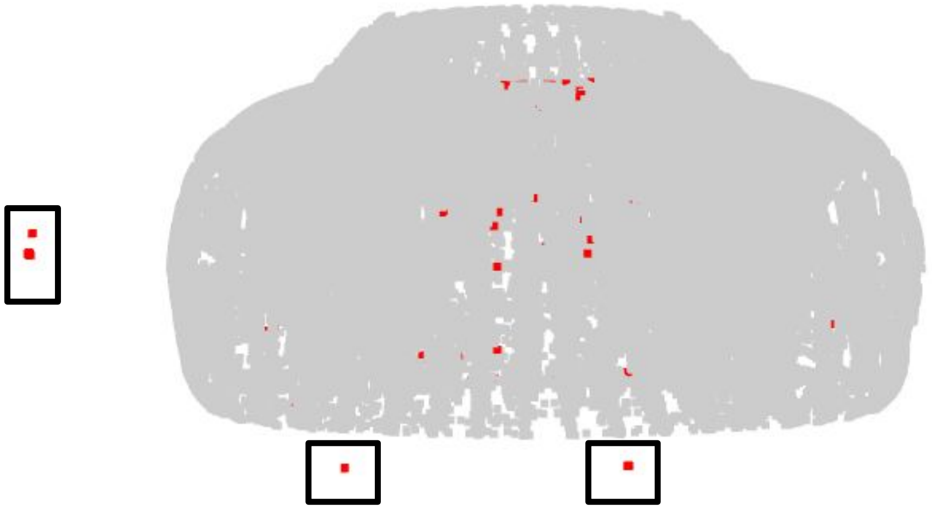


Front hood reinforcement material change

Data Cleaning - Outlier Treatment

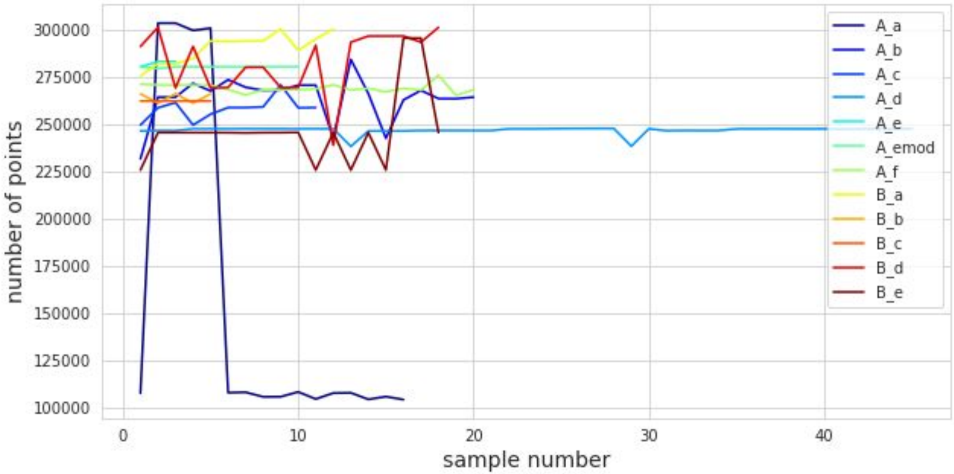
- Removal of points outside the pointcloud.
- For every point take a set of K points
- Compute standard deviation of distance std_K.

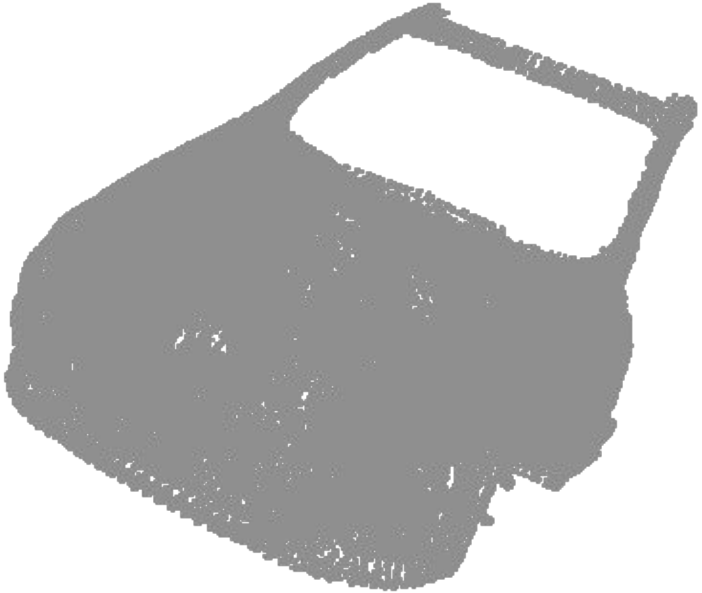
$$\frac{std_K}{std_{average}} \geq threshold \implies Outlier$$



Data Cleaning - Faulty Data Removal

- Point Clouds missing windshield.
- Undersampled Point Clouds
- Help in tackling Data Bias.





Outlier Computation

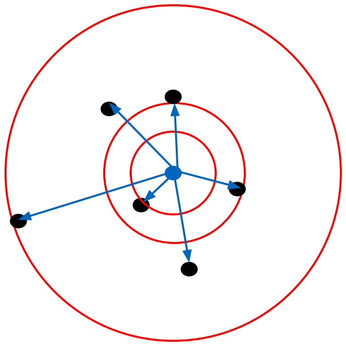
- For every point we make a group of N_L points.
- Perform following computations

$$d_{avg} = \frac{1}{N_L} \sum_{i \in L} d_i$$

$$d_{std} = \sqrt{\frac{1}{N_L} \sum_{i \in L} (d_i - d_{avg})^2}$$

$$std_{average} = \frac{1}{N} \sum d_{std}$$

$$if \quad \frac{d_{std}}{std_{average}} \geq theshold \implies Outlier$$



ICP

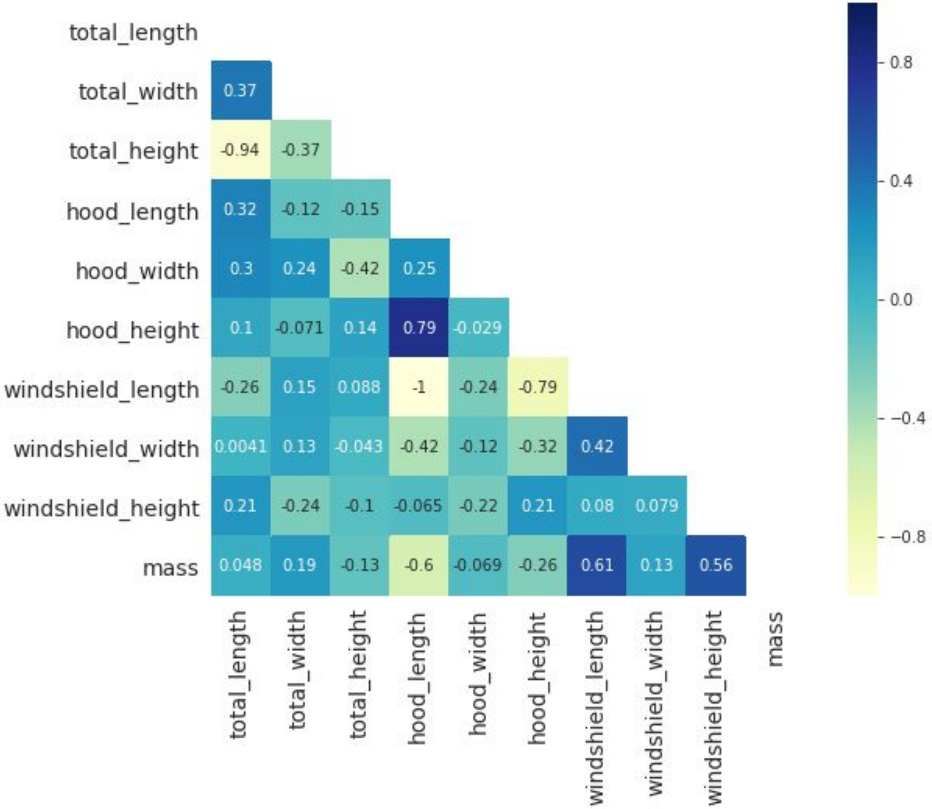
- Find correspondence set $\mathcal{K} = (\mathbf{p}, \mathbf{q})$ from the target point cloud $\mathbf{P}$ and the source point cloud $\mathbf{Q}$ using KDtree.
- Update the transformation matrix $\mathbf{T}$ by minimizing the objective function $E(\mathbf{T})$ described below

$$E(\mathbf{T}) = \sum_{(\mathbf{p},\mathbf{q})\in\mathcal{K}} \|\mathbf{p} - \mathbf{T}\mathbf{q}\|^2$$

- Apply transformation $\mathbf{T}$ on the source point cloud

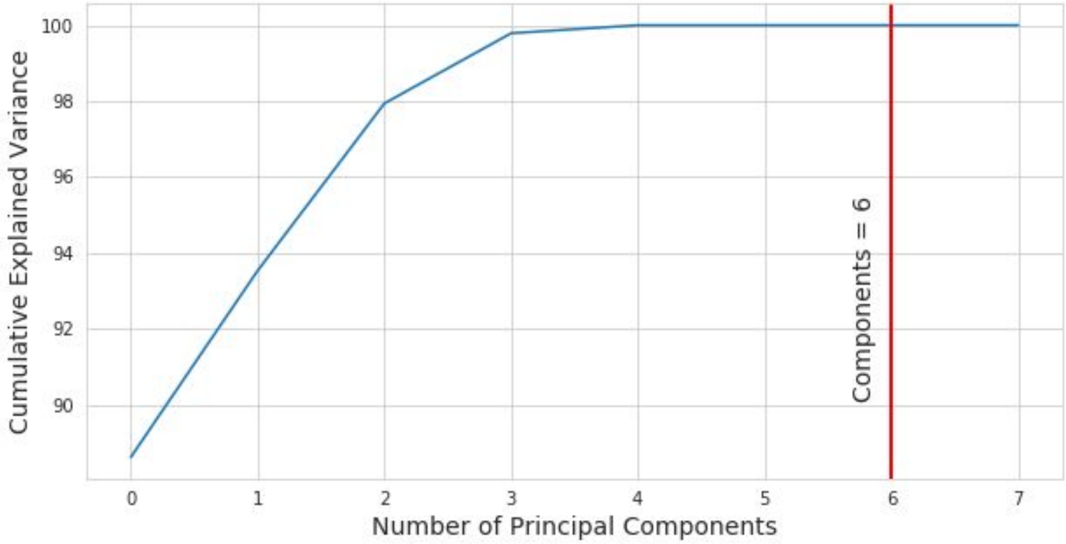
Car Matching - Feature Analysis

- Strong Correlation:
  - Total Length and Total Height
  - Hood Length and Windshield Length
- Remove correlated features:
  - Simplifies model
  - Better generalization

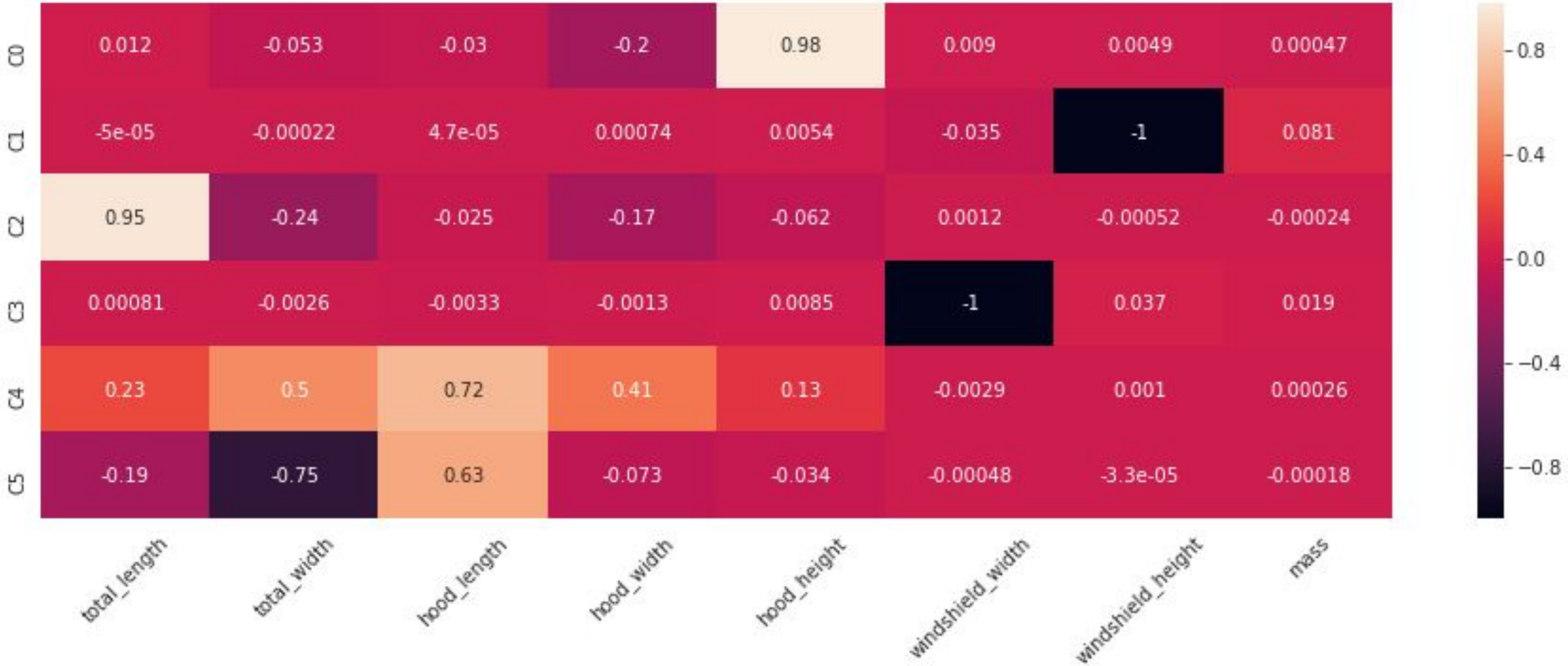


Car Matching - Feature Analysis

- Data only 6 Dimension:
    - We get total 6 features now.
- Better generalization



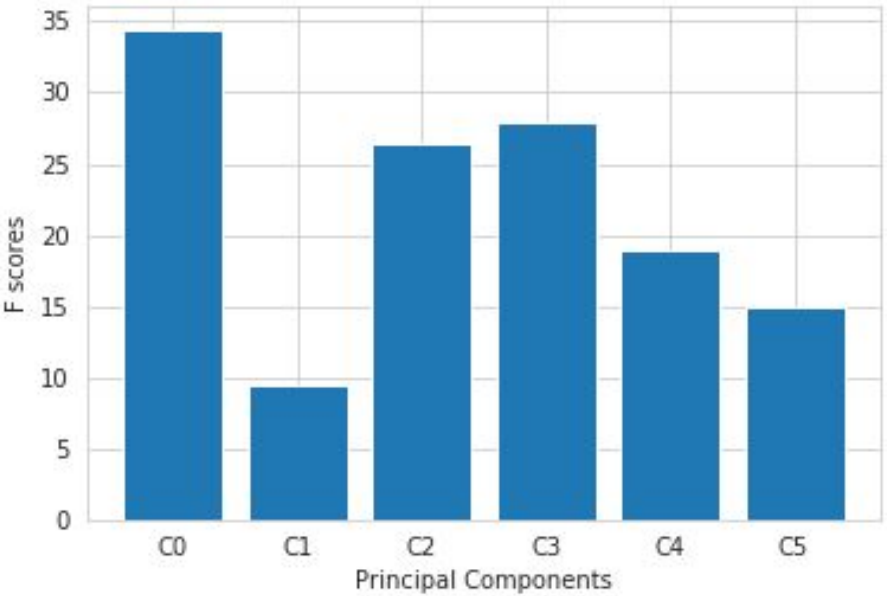
Car Matching - Feature Analysis



Car Matching - Feature Analysis

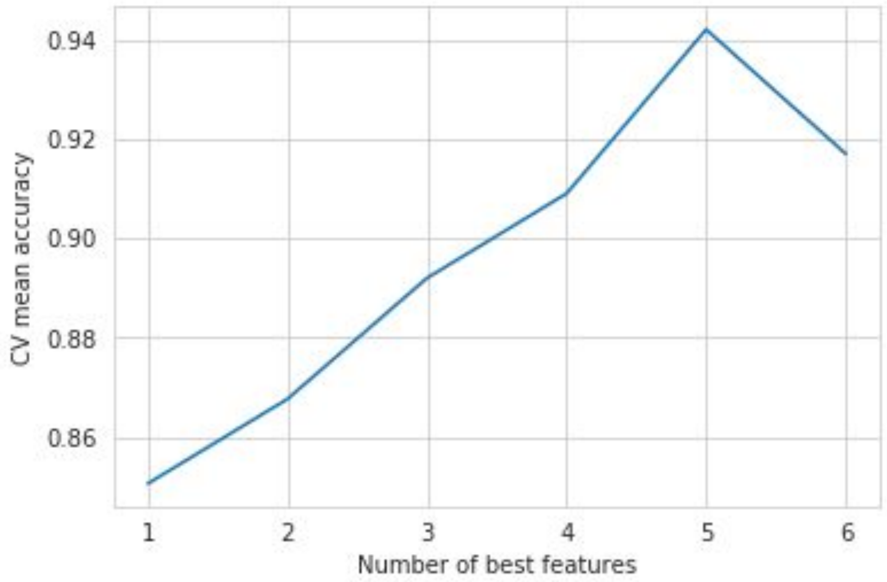
- Feature ranking based on F scores
  - C0
  - C3
  - C2
  - C4
  - C5
  - C1
- ANOVA F-Test

$$F_{score} = \frac{variance\ between\ cars}{variance\ within\ cars}$$



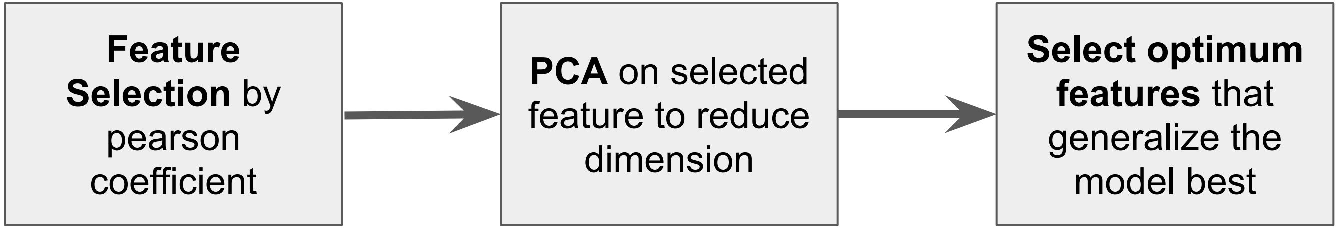
Car Matching - Feature Analysis

- Only 5 features improves the model.
- Sixth feature is not very informative for prediction
- Selected features:
    - C0
    - C3
    - C2
    - C4
    - C5



Car Matching - Feature Analysis

- Feature Selection: Simplify the model, hence the misclassification goes down.
- PCA: Reduces Dimension and simplifies more, hence improves generalization.
- Feature sensitivity: Selects the best components, denoising.

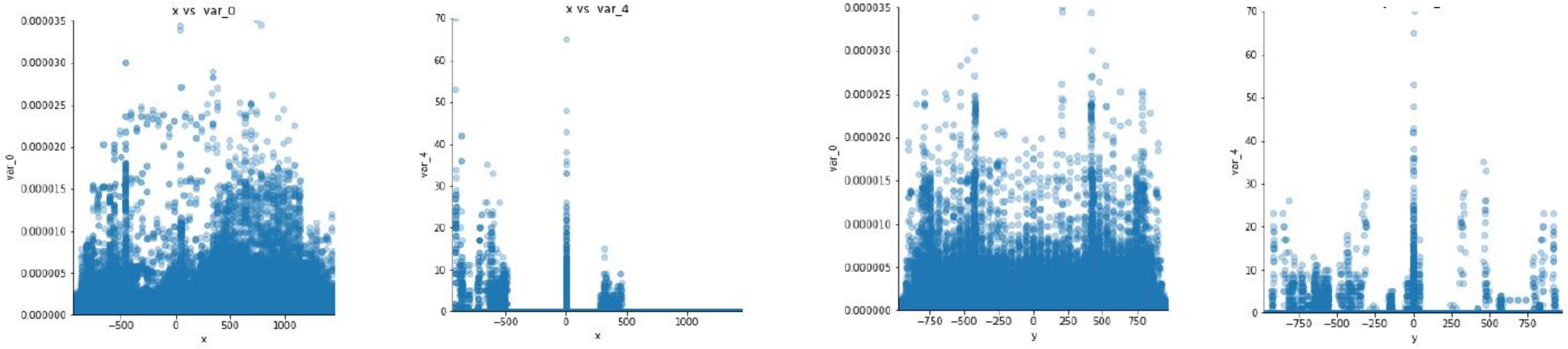


Car Matching

	Base Model	Feature Selection	PCA	Feature Sensitivity
Misclassification Rate	0.05769	0.03846	0.05769	0.05769
Cross Validation Score	0.909 ± 0.062	0.908 ± 0.057	0.934 ± 0.043	0.942 ± 0.034

- Total 161 car models
- 80:20 Train:Test
- 5 fold Cross Validation

DATA - SCATTER PLOTS



(a) Variation of feature *var_0* across *X* (b) Variation of feature *var_4* across *X* (c) Variation of feature *var_0* across (d) Variation of feature *var_4* across *X*.