Optimizing Active Vision-Based Policy for Robotic Grasping through 3D Point Cloud Feature Extraction

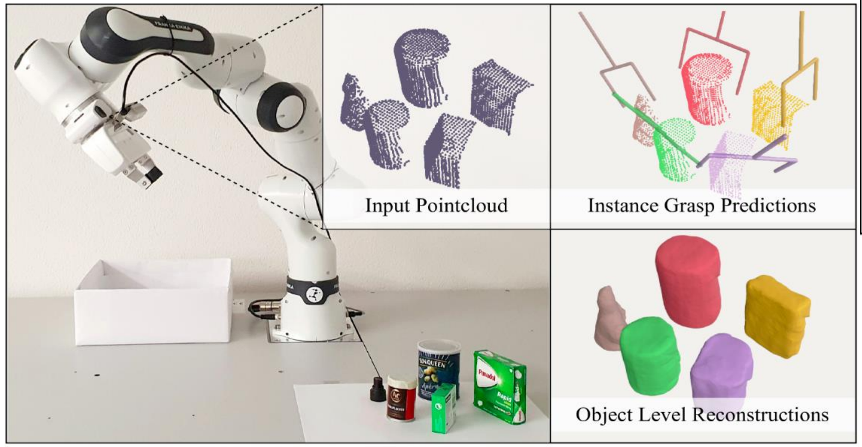
A Comparative Study

Advisor: Prof. Berk Calli

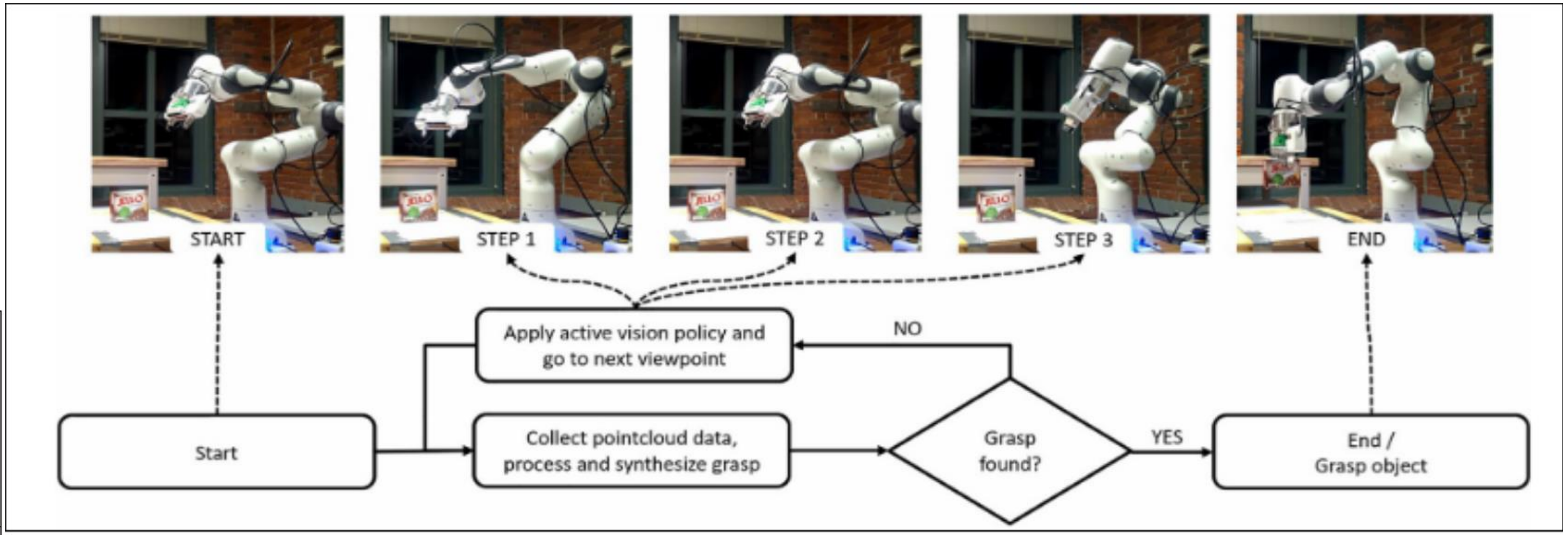
Uthiralakshmi Sivaraman

Vision based Robotic Grasping

- Capture Camera data to find sufficiently good grasp using robotic arms.
- Grasping using Single image vs multiple images vs Real time.



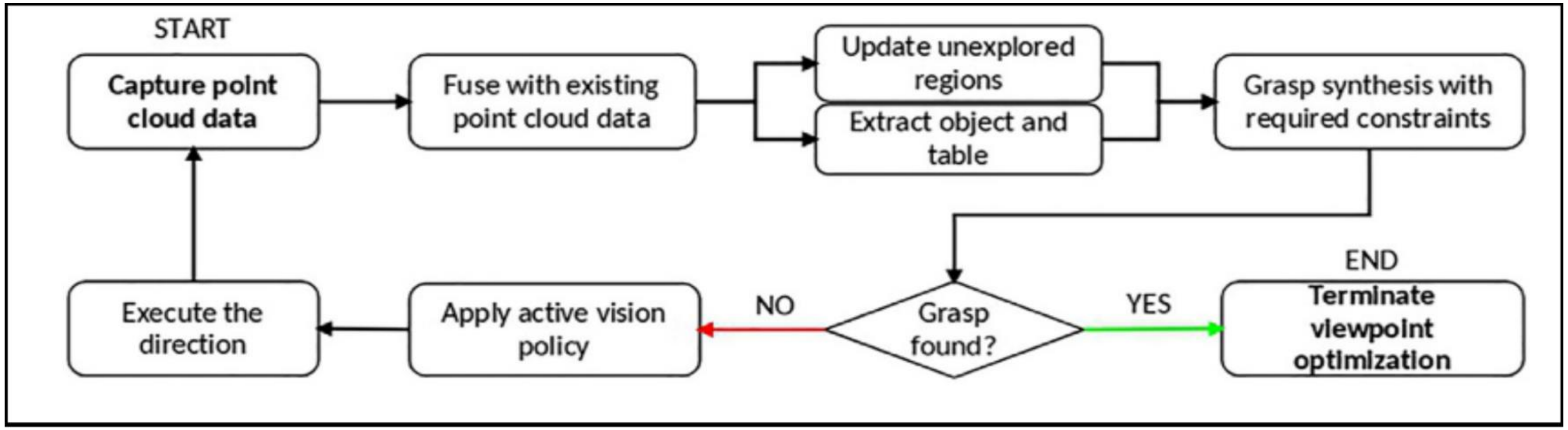
Single Viewpoint Grasping

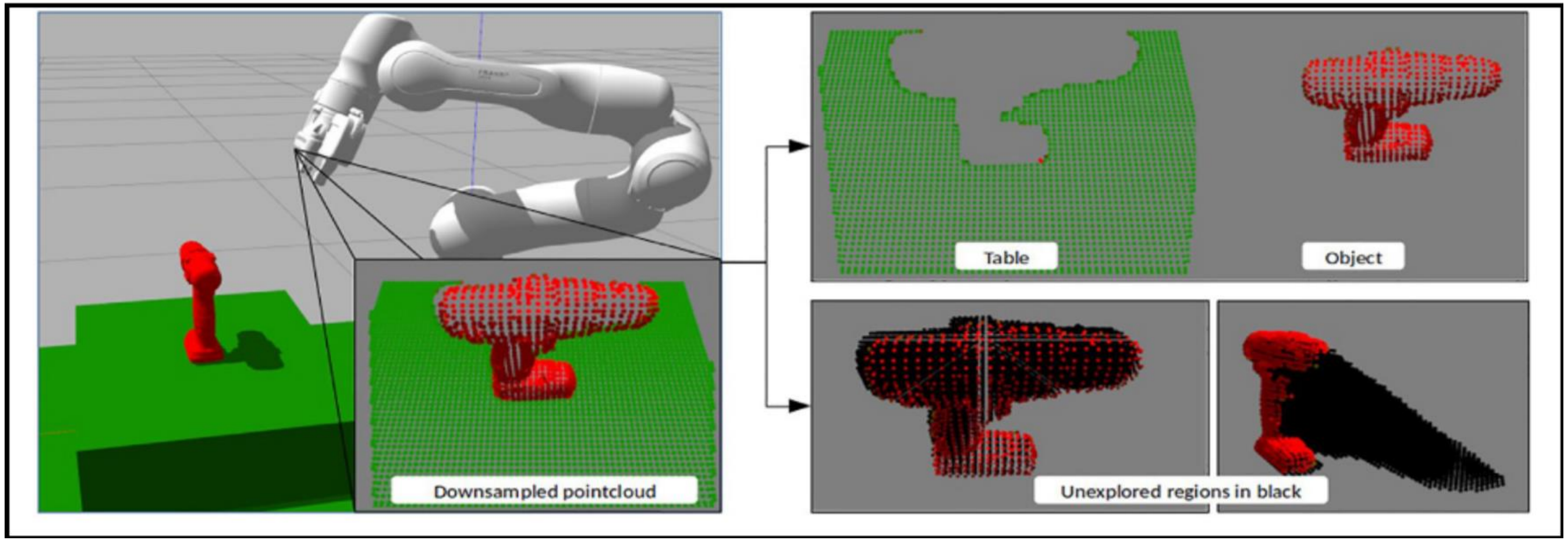


Multiple Viewpoint Grasping

Img:
Natarajan, S., Brown, G., & Calli, B. (2021). Grasp Synthesis for Novel Objects Using Heuristic-based and Data-driven Active Vision Methods.
Zurbrügg, R., Liu, Y., Engelmann, F., Kumar, S., Hutter, M., Patil, V., & Yu, F. (2024). ICGNet: A Unified Approach for Instance-Centric Grasping

Active Vision Policy for Grasping

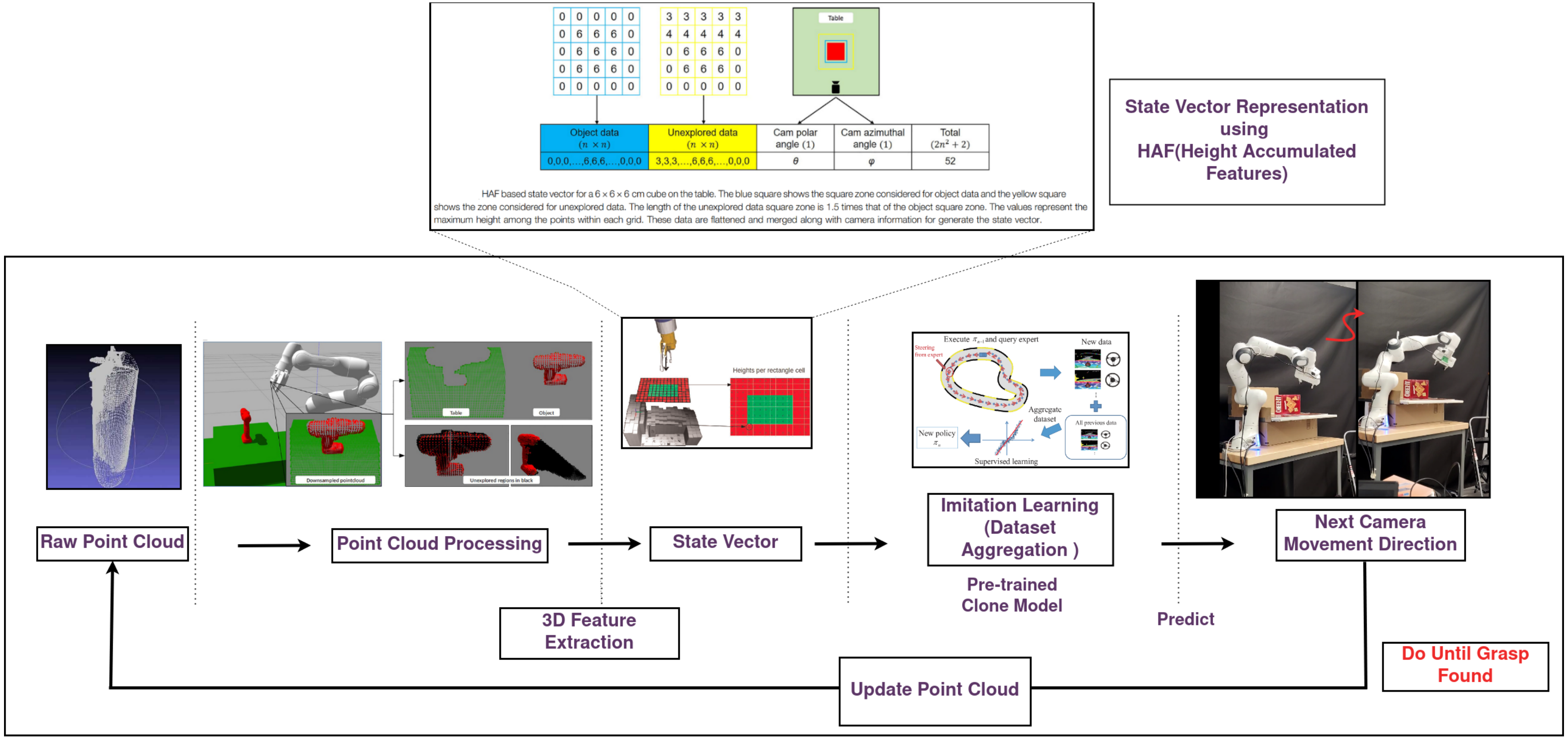




- Franka Emika Panda arm with Intel Real Sense Kinect mounted.
- Collect object data through camera movement.
- Capture and process point cloud data.
- Apply Active Vision Policy to predict next camera movement.

Img :
Natarajan, S., Brown, G., & Calli, B. (2021). Grasp Synthesis for Novel Objects Using Heuristic-based and Data-driven Active Vision Methods.

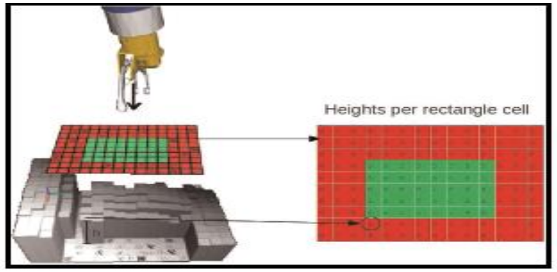
Imitation Learning based Policy Prediction



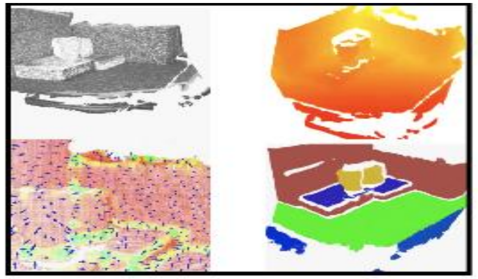
- Imitation learning- Dataset Aggregation (Dagger) to predict the next camera viewpoint direction.
- Epsilon Optimal Expert demonstration for imitation.
- 0° to 360 ° camera movement space with ±5° increment.

Img:
Natarajan, S., Brown, G., & Calli, B. (2021). Grasp Synthesis for Novel Objects Using Heuristic-based and Data-driven Active Vision Methods.
Ross, S., Gordon, G. J., & Bagnell, J. A. (2010). A Reduction of Imitation Learning and Structured Prediction to No-Regret Online Learning.

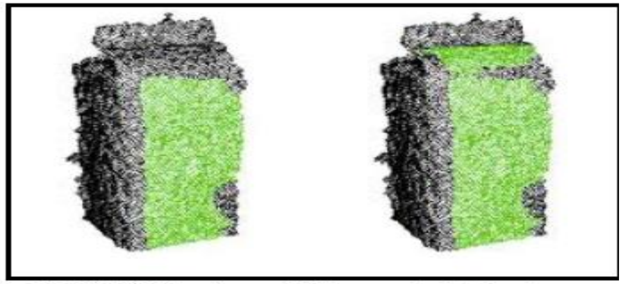
Comparison of 3D Feature Descriptors



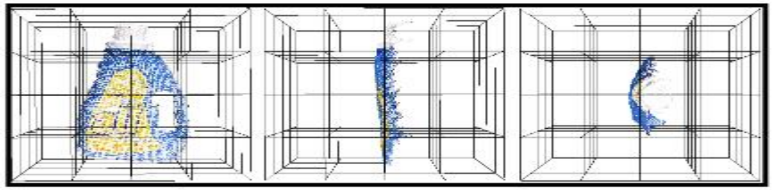
HAF (Height Accumulated Features)



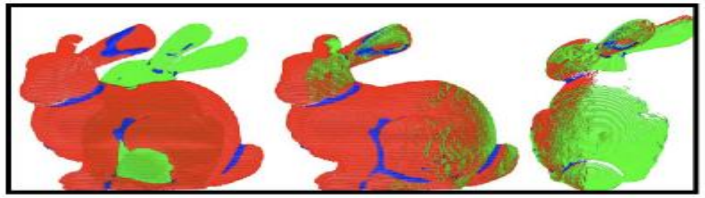
VFH (Viewpoint Feature Histogram)



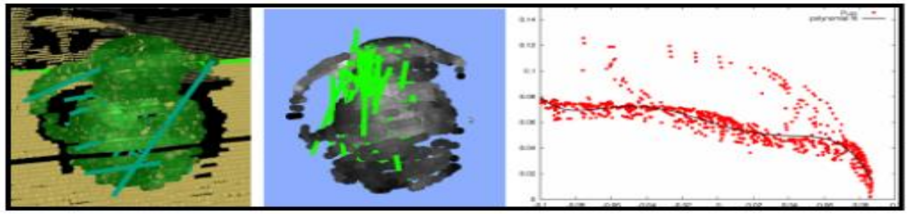
CVFH (Clustered Viewpoint Feature Histogram)



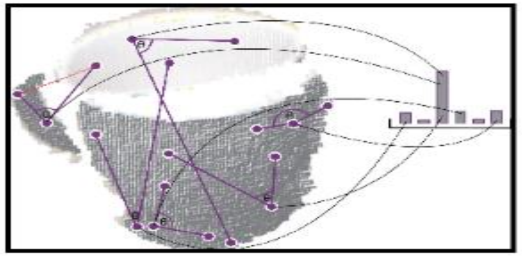
GASD (Global Aligned Spatial Distribution)



FPFH (Fast Point Feature Histogram)



GRSD (Global Radius-based Surface Descriptor)



ESF (Ensemble of Shape Features)

- Process raw RGB point cloud using feature descriptors to enhance object recognition
- 7 different features
- HAF, VFH, CVFH, GASD,FPFH,GRSD, ESF

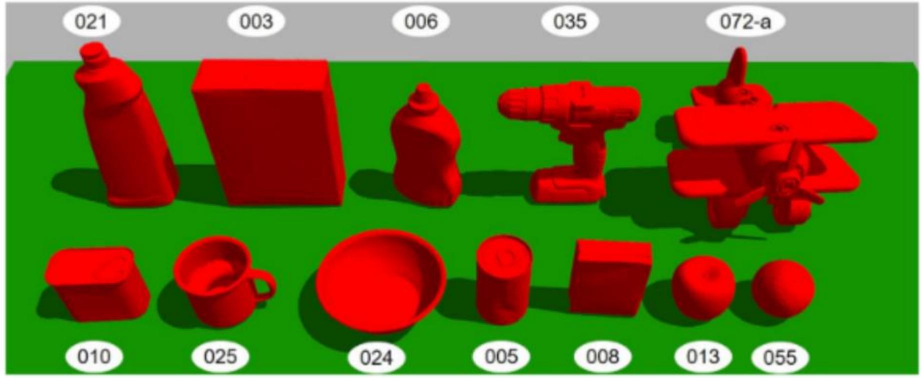
Img: HAF: Fischinger, David & Weiss, Astrid & Vincze, Markus. (2015). Learning grasps with topographic features.
VFH: *Rusu, Radu Bogdan et al. "Fast 3D recognition and pose using the Viewpoint Feature Histogram."*
CVFH: Aldoma, A., Tombari, F., Rusu, R.B., & Vincze, M. (2012). OUR-CVFH - Oriented, Unique and Repeatable Clustered Viewpoint Feature Histogram for Object Recognition and 6DOF Pose Estimation.
GASD: https://pcl.readthedocs.io/projects/tutorials/en/latest/gasd_estimation.html
FPFH: Rusu, Radu Bogdan, Nico Blodow and Michael Beetz. "Fast Point Feature Histograms (FPFH) for 3D registration."
GRSD: Marton, Zoltan & Pangercic, Dejan & Blodow, Nico & Kleinehellefort, J. & Beetz, Michael. (2010). General 3D modelling of novel objects from a single view.
ESF: Wohlkinger, Walter and Markus Vincze. "Ensemble of shape functions for 3D object classification."

Dataset

A subset of **Yale-CMU-Berkeley (YCB)** Object and Model Set

- Objects of daily life with different shapes, sizes, textures, weight and rigidity
- Widely used for robotic manipulation benchmark

Yale-CMU-Berkeley (YCB) Dataset



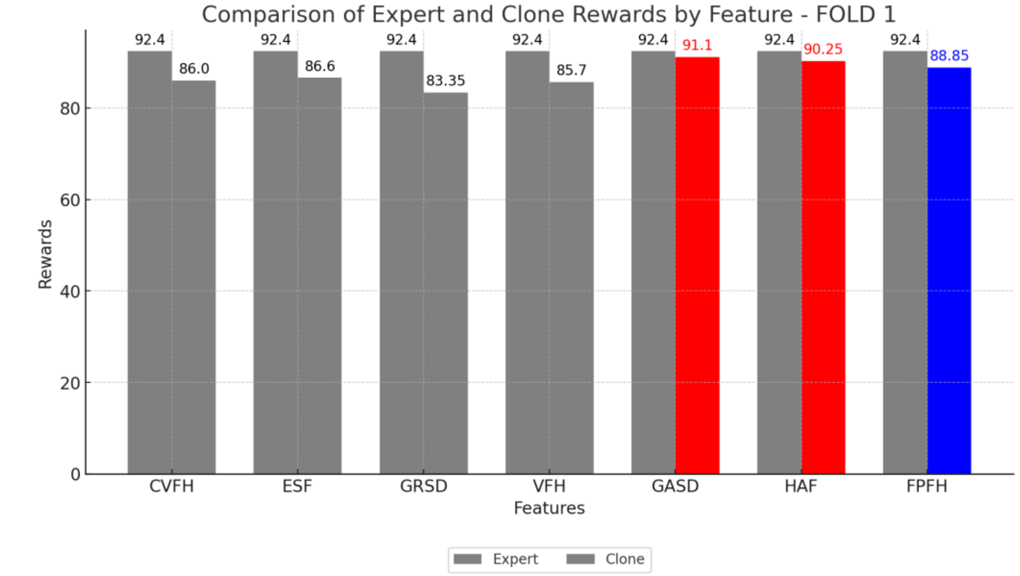
Img:https://www.ycbbenchmarks.com/
Natarajan, S., Brown, G., & Calli, B. (2021). Grasp Synthesis for Novel Objects Using Heuristic-based and Data-driven Active Vision Methods.

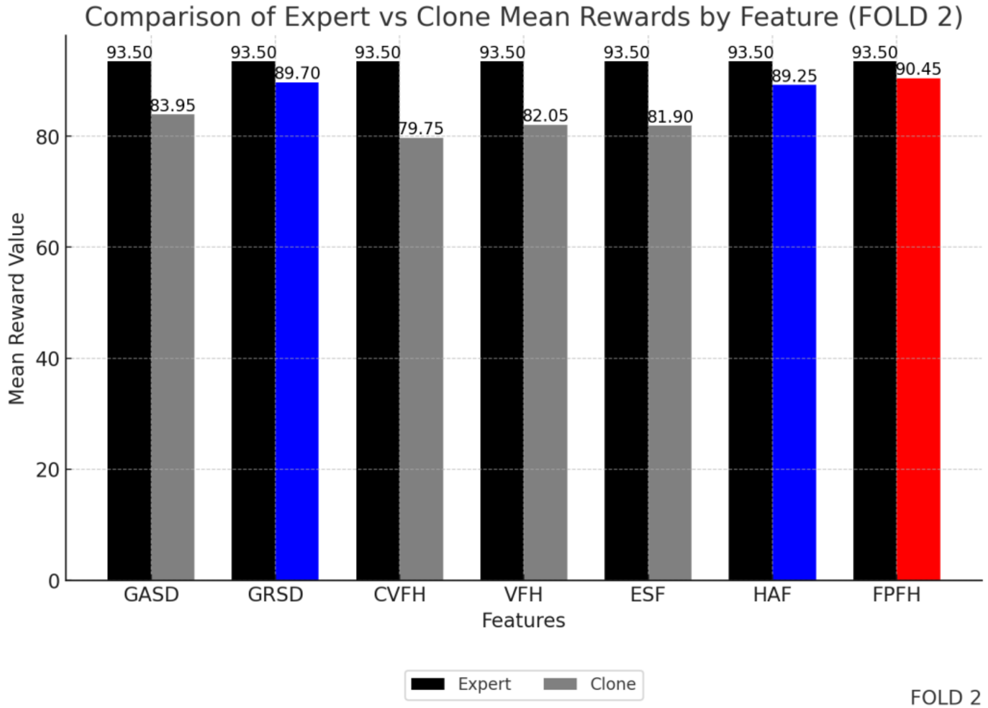
Methodology and Evalution Metrics

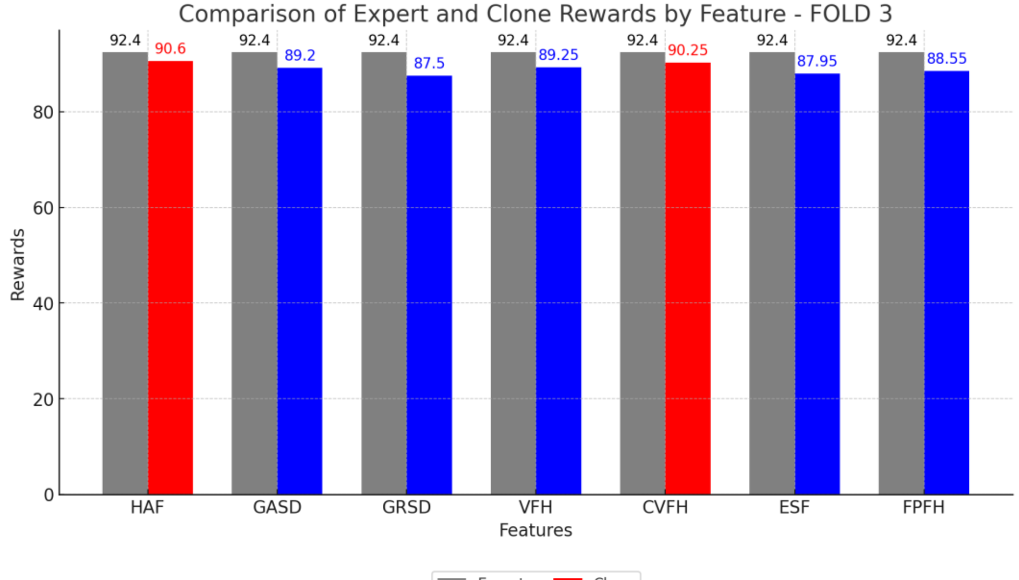
- Number of camera movement steps taken to find Successful Grasp
- Mean cumulative reward and standard deviation
- Model training time vs evaluation time
- Percentage of Successful Grasps
- 3 cross-validation folds and test on simulation

Fold	Training Objects and IDs	Testing Objects and IDs
1	009 gelatin box (8) 055 baseball (41) 072-a toy airplane (51) 010 potted meat can (9) 003 cracker box (2) 035 power drill (28) 006 mustard bottle (5) 021 bleach cleanser (19) 013 apple (12) Weisshai Great White Shark (65)	004 sugar box (3) 005 tomato soup can (4)
2	055 baseball (41) 072-a toy airplane (51) 010 potted meat can (9) 003 cracker box (2) 005 tomato soup can (4) 006 mustard bottle (5) 021 bleach cleanser (19)	009 gelatin box (8) 035 power drill (28) 013 apple (12)
3	009 gelatin box (8) 055 baseball (41) 010 potted meat can (9) 003 cracker box (2) 035 power drill (28) 005 tomato soup can (4) 006 mustard bottle (5) 021 bleach cleanser (19) 013 apple (12)	072-a toy airplane (51) Weisshai Great White Shark (65)

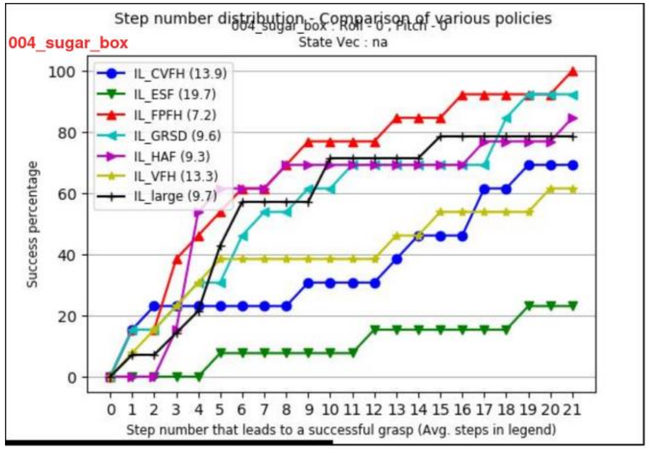
Train Results

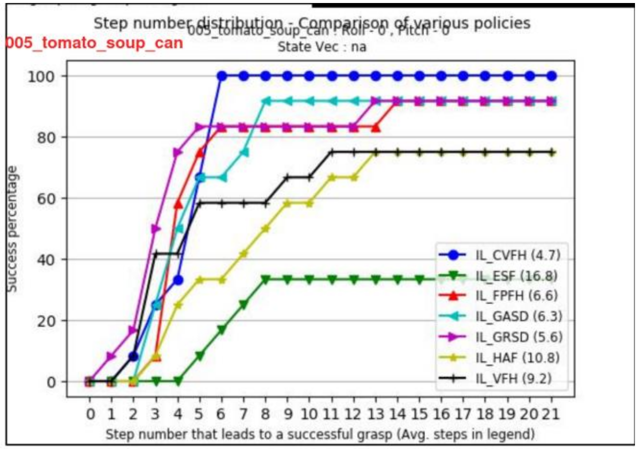


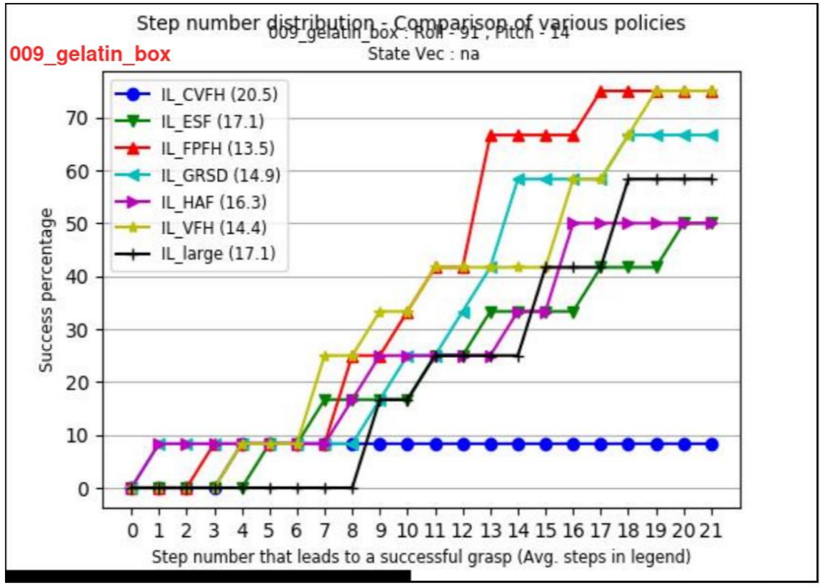




Test Results

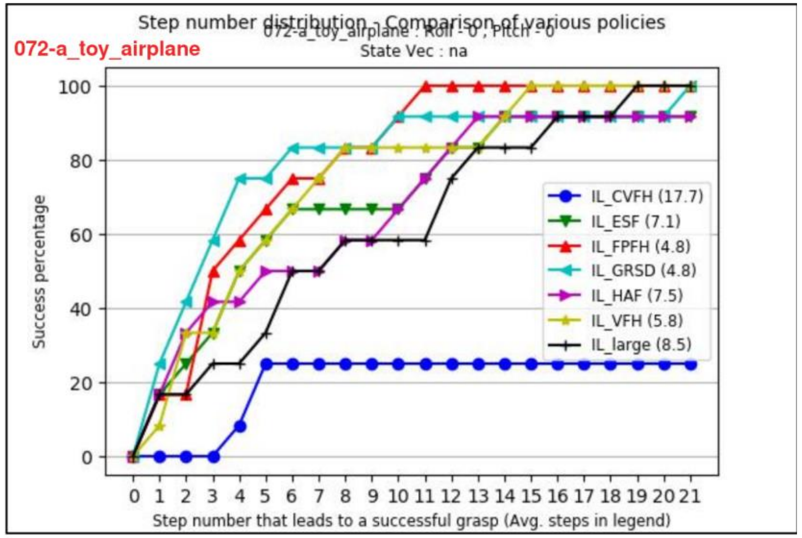






FOLD2

FOLD1



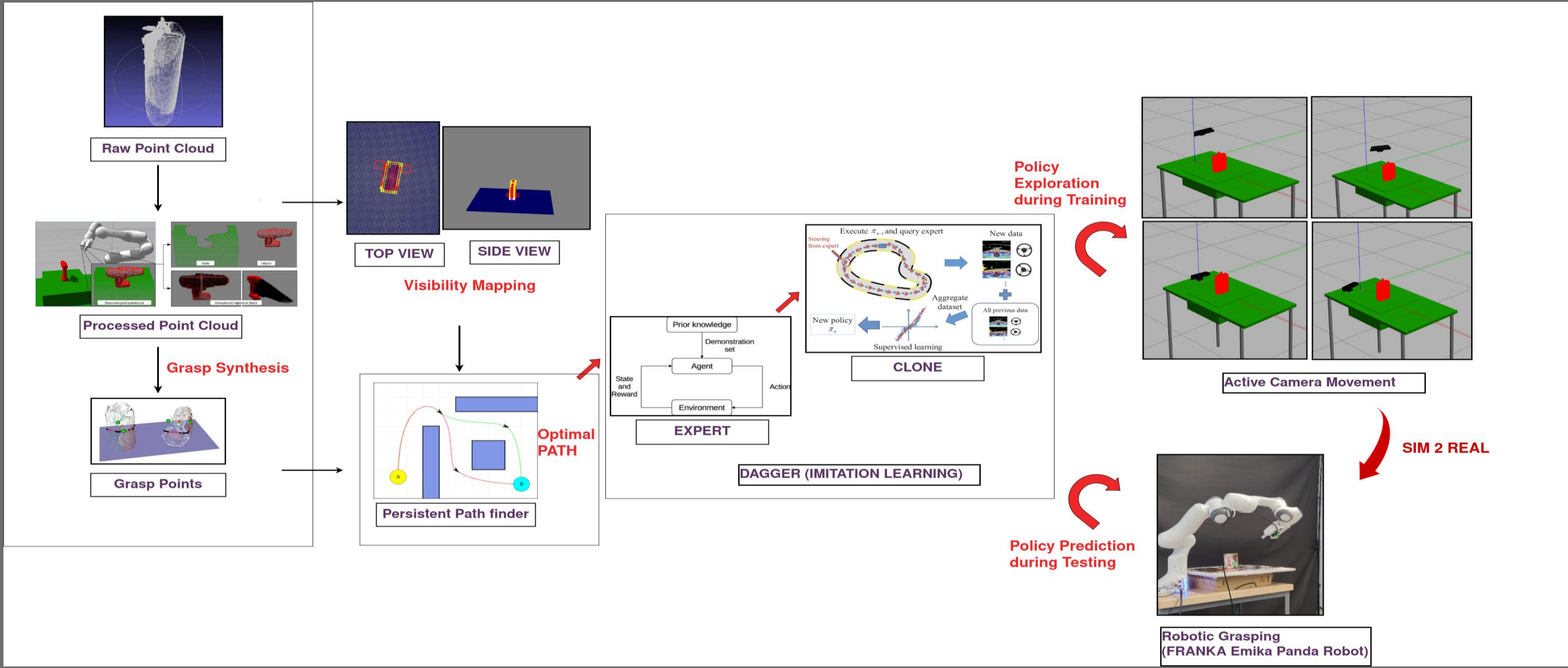
FOLD3

- FPFH showing optimal performance and efficiency in training time.
- FPFH excelled in learning curves and efficiency across all folds.

3

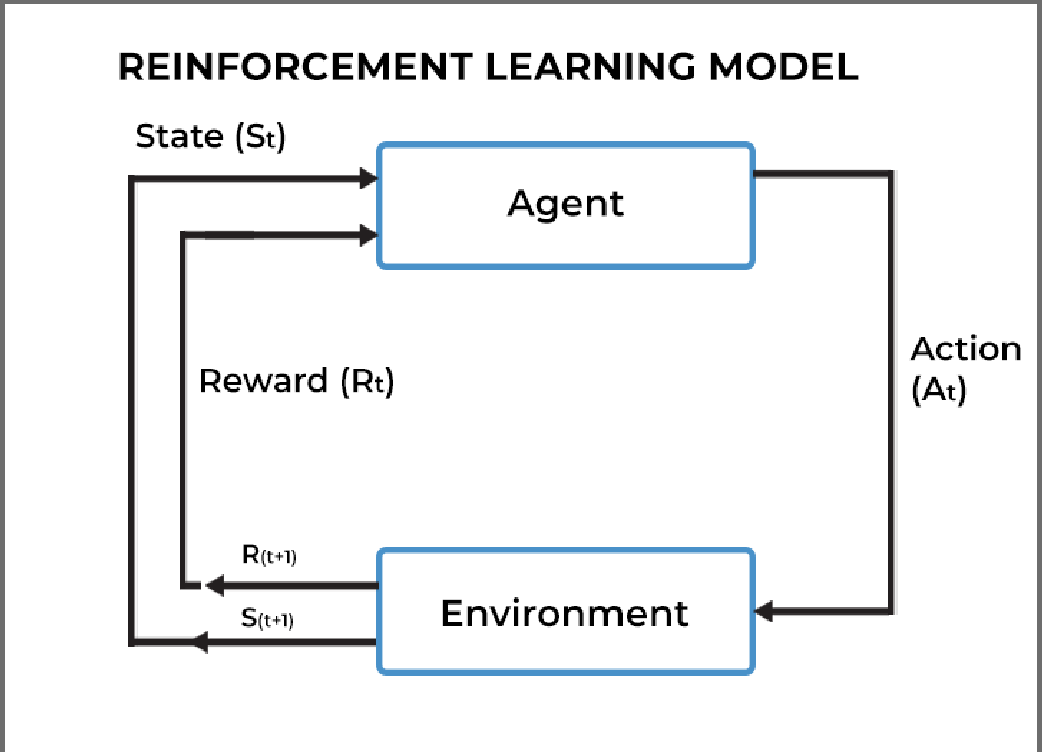
Thank You!
Any Questions?

Imitation Learning with Expert Demonstration



Img:
Natarajan, S., Brown, G., & Calli, B. (2021). Grasp Synthesis for Novel Objects Using Heuristic-based and Data-driven Active Vision Methods.
Ross, S., Gordon, G. J., & Bagnell, J. A. (2010). A Reduction of Imitation Learning and Structured Prediction to No-Regret Online Learning.

Imitation Learning Reward Calculation



Formula

$$\text{Reward} = \left(\frac{\text{maxPath} - \text{cStep}}{\text{maxPath}}\right)$$

maxPath- The maximum number of steps allowed for an agent to achieve its goal within an episode.
cStep- The current number of steps the agent has taken in the current episode.

- During training, the imitation learning policy explores all possible camera movements.
- Clone imitates expert to achieve maximum cumulative reward

Img: https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-reinforcement-learning/

VFH, CVFH, FPFH

Step	VFH	CVFH	FPFH
Surface Normal Estimation	Calculate the normals for each point	Calculate the normals for each point	Calculate the normals for each point
Region Growing Segmentation	Not used	Segment cloud into clusters based on curvature	Not used
Centroid Computation	Calculate the centroid of the cloud	Compute centroid for each cluster	Not used
Viewpoint Direction Check	Determine the viewpoint direction to the centroid	Normals not pointing to viewpoint are flipped	Not used
Histogram Calculation	Compute histogram of angles between normals and viewpoint direction	Compute histogram for each cluster including angles and distances	Compute local histograms (SPFH) within the neighborhood
Histogram Concatenation	Combine histograms with centroid distances into a large histogram	Concatenate histograms from clusters, weighted by cluster size	Aggregate SPFHs of neighbors to create the FPFH
Normalization	Optional to make descriptor size invariant	Histogram is normalized for invariance to scale	Not typically normalized as it is a local feature

HAF, GASD, GRSD, ESF

Step	HAF (Height Accumulated Features)	GRSD (Global Radius-based Surface Descriptor)	GASD (Global Aligned Spatial Distribution)	ESF (Ensemble of Shape Functions)
Surface Normal Estimation	Not specifically used for height accumulation	Calculate the normals for each point	Not explicitly required for spatial distribution calculations	Not directly involved in shape function ensemble
Region Growing Segmentation	Not used	Not used	Not used	Not used
Centroid Computation	Calculates the highest point within each grid cell	Global computation not centered on centroids	Computes a spatially-aligned global centroid to orient the point cloud	Not used
Viewpoint Direction Check	Not applicable as it uses a grid-based approach	Not used	Not used as orientation is handled globally	Not used
Histogram Calculation	Accumulates maximum heights within each grid cell to form a feature map	Quantifies distribution of surface normals globally	Captures the global distribution of points, enhancing orientation robustness	Aggregates multiple shape functions to describe the overall shape globally
Histogram Concatenation	Not applicable, as the method does not use histogram concatenation	Not used	Not used	Not used
Normalization	Optional, to scale the height values relative to the scene for invariance	Typically normalized to enhance comparison across different scales and orientations	Normalization to ensure scale invariance and improve robustness against variations in orientation	Generally not normalized as it focuses on global shape description

Dagger (Dataset Aggregation)