Active Perception

Maren Bennewitz, Rohit Menon
Humanoid Robots Lab, University of Bonn

Exam

- Written exam
- August 18 and September 23 as announced
- Time tbc

Goal of This Chapter

- Definition of active perception and its different constituents
- Understanding active exploration and perception strategies
- Overview of entropy and information gain formulations
- Introduction to active perception planning strategies
- Various active perception applications

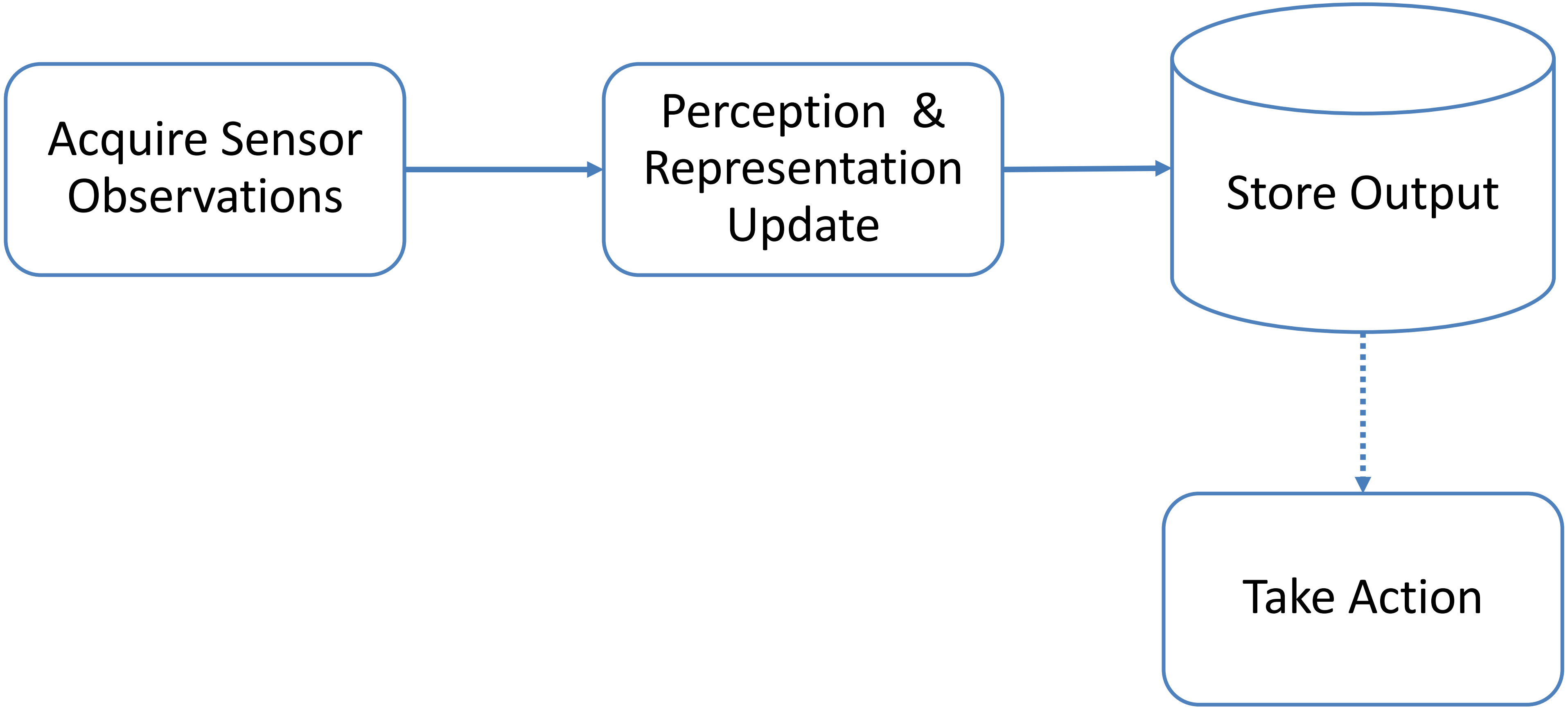
Motivation

[Boston Dynamics, "Atlas Goes Hands On", 2024, www.youtube.com/watch?v=F_7IPm7f1vI]

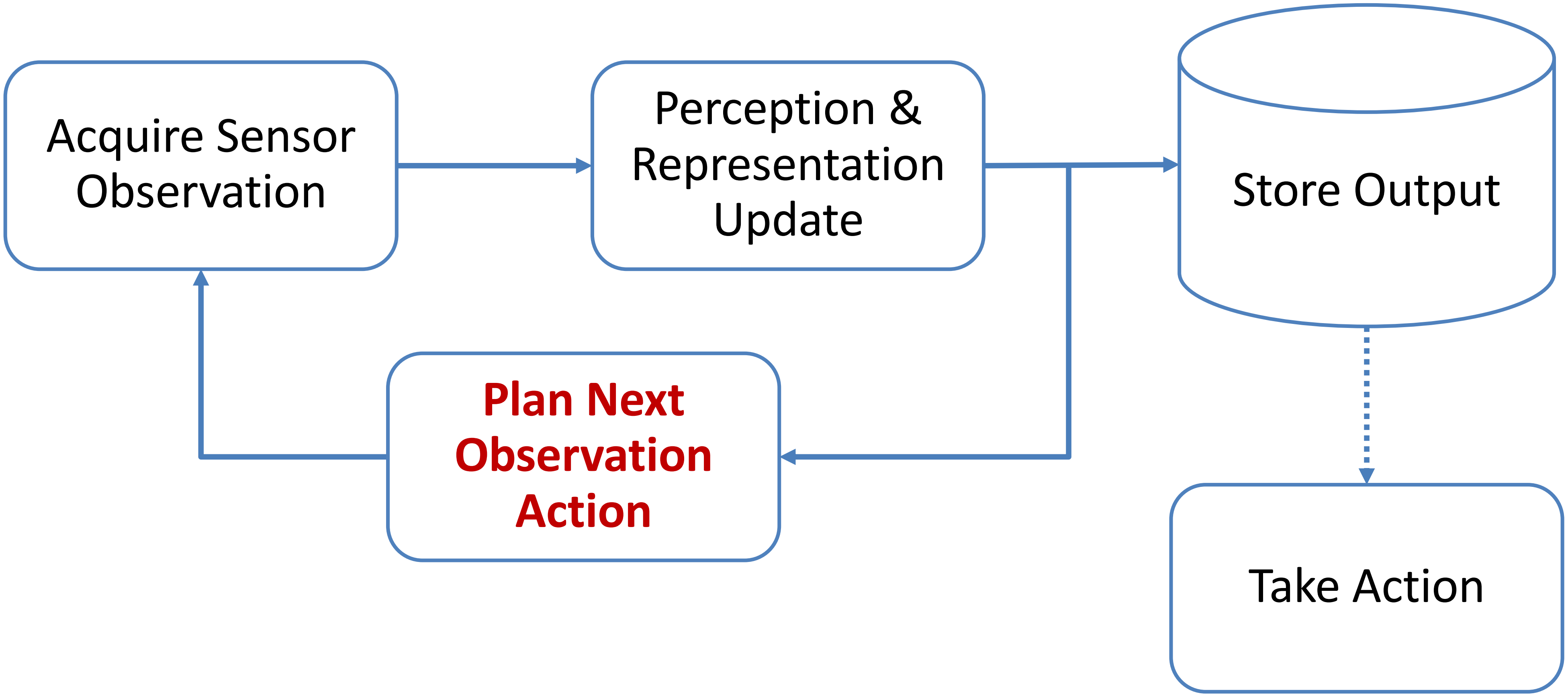
Motivation

- Robots live in unstructured environments
- Human sensorimotor learning shows perception action coupling
- Creating environment representations requires exploration
- Random exploration does not scale well
  - Size
  - Details
- **Active perception enables efficient, detailed, and full coverage of unknown scenes**

Traditional Perception Pipeline



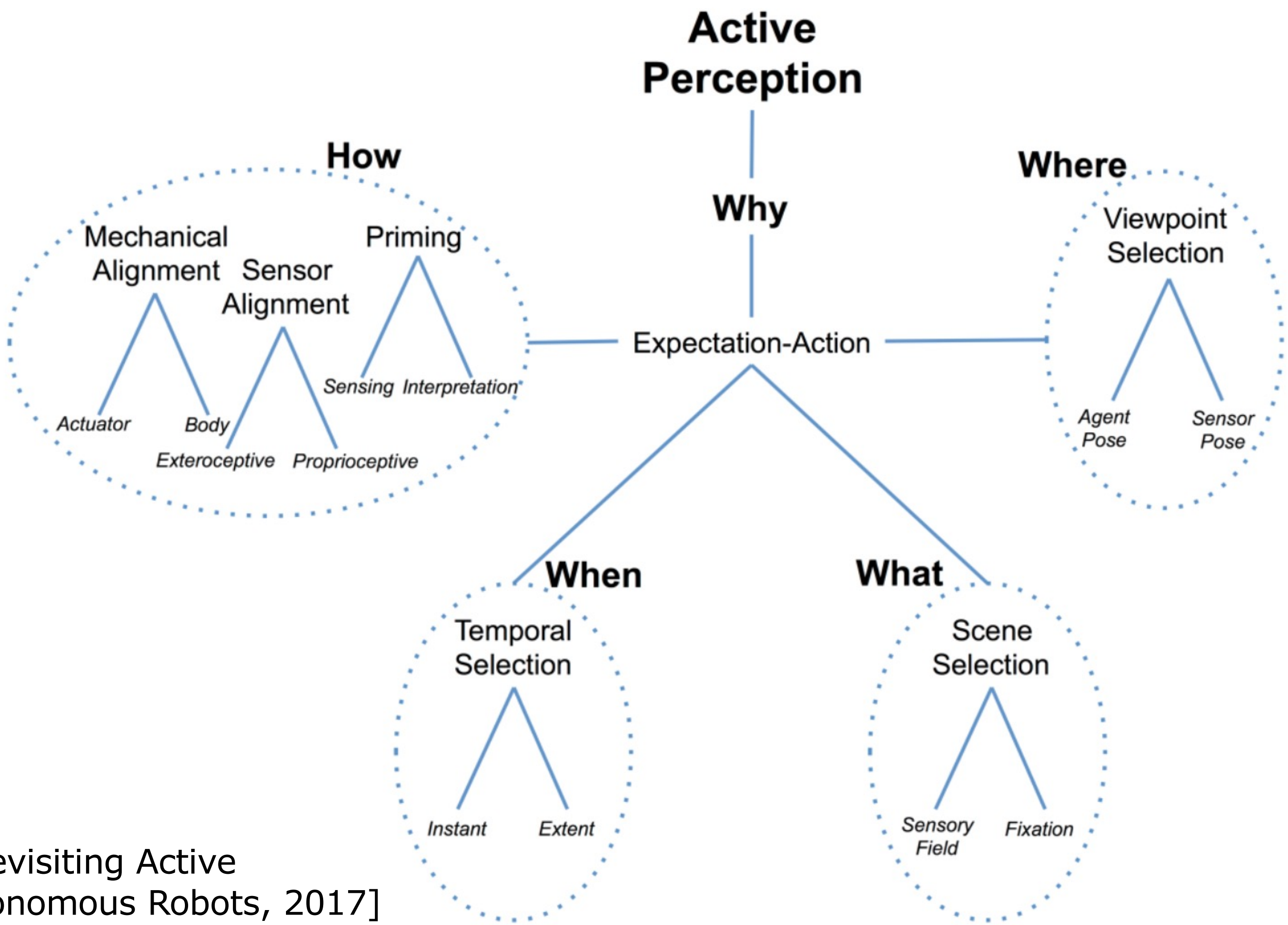
Active Perception Pipeline



Active Perception Definitions

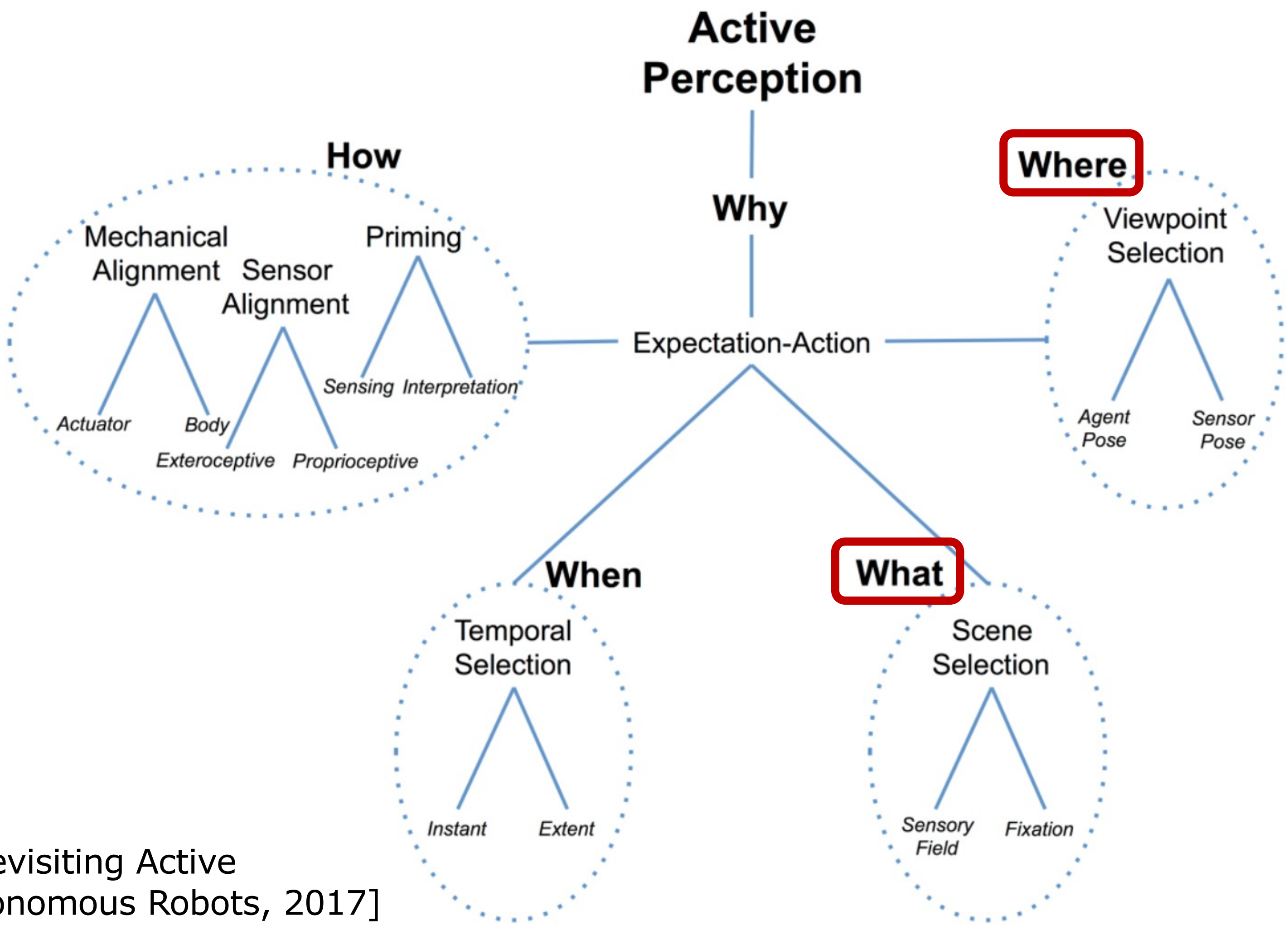
- "[…] *the problem of* ***intelligent control strategies*** *applied to the* ***data acquisition process*** *which will depend on the current state of data interpretation […]"* (Bajcsy, 1988)
- *"An agent is an active perceiver if it knows* ***why*** *it wishes to sense, and then chooses* ***what*** *to perceive, and determines* ***how***, ***when***, *and* ***where*** *to achieve that perception."* (Bajcsy et al., 2017)

Five Main Constituents of Active Perception



[Bajcsy et al., “Revisiting Active Perception”, Autonomous Robots, 2017]

Five Main Constituents of Active Perception



[Bajcsy et al., "Revisiting Active Perception", Autonomous Robots, 2017]

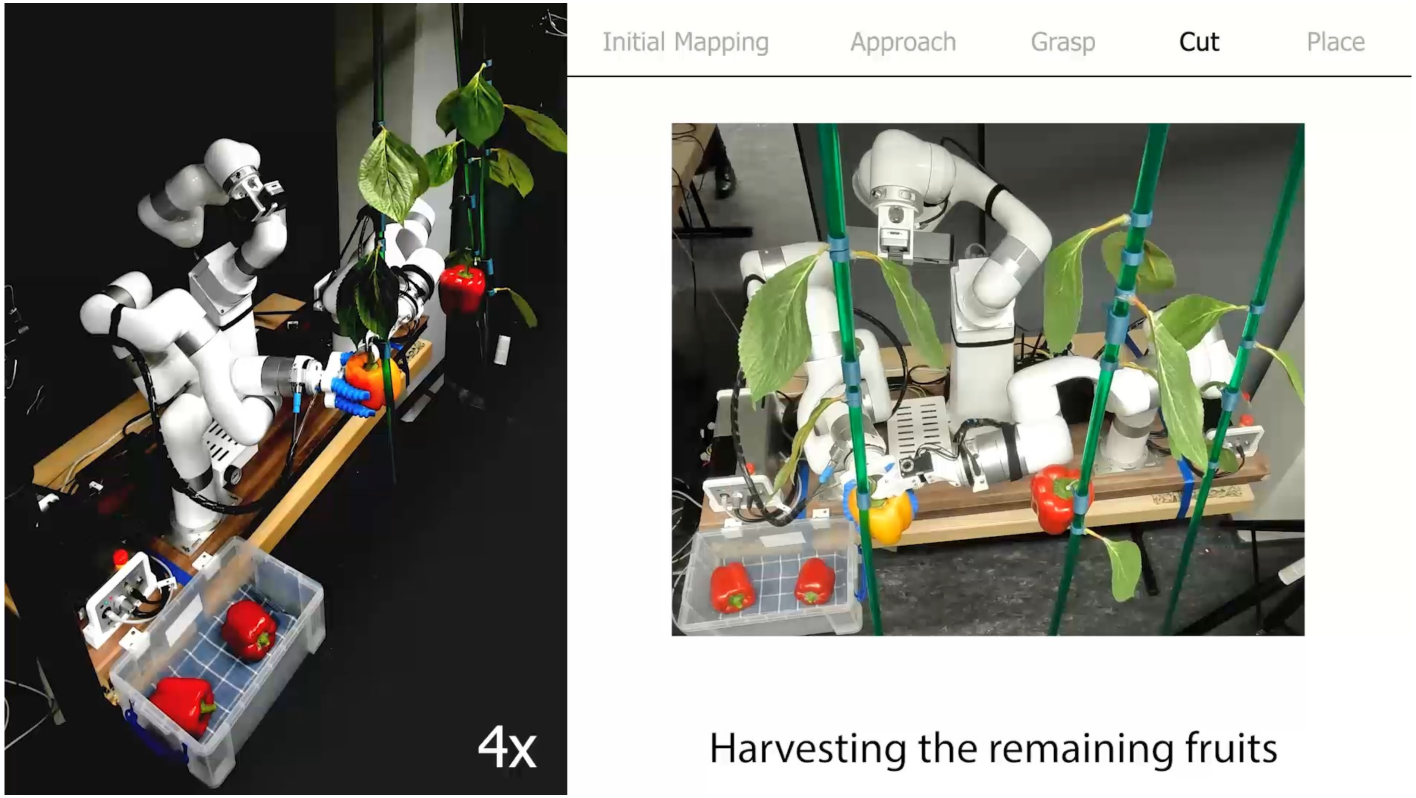
What: Scene Selection

- **Fixation**
  - Active prediction of which part of a real-world scene to view to solve the task

- **Sensory Field**
  - Active prediction of where in a scene a stimulus relevant to the current task may appear, e.g., selection of the subset of an image

[Bajcsy et al., “Revisiting Active Perception”, Autonomous Robots, 2017]

What: Active Peduncle Localization for Harvesting



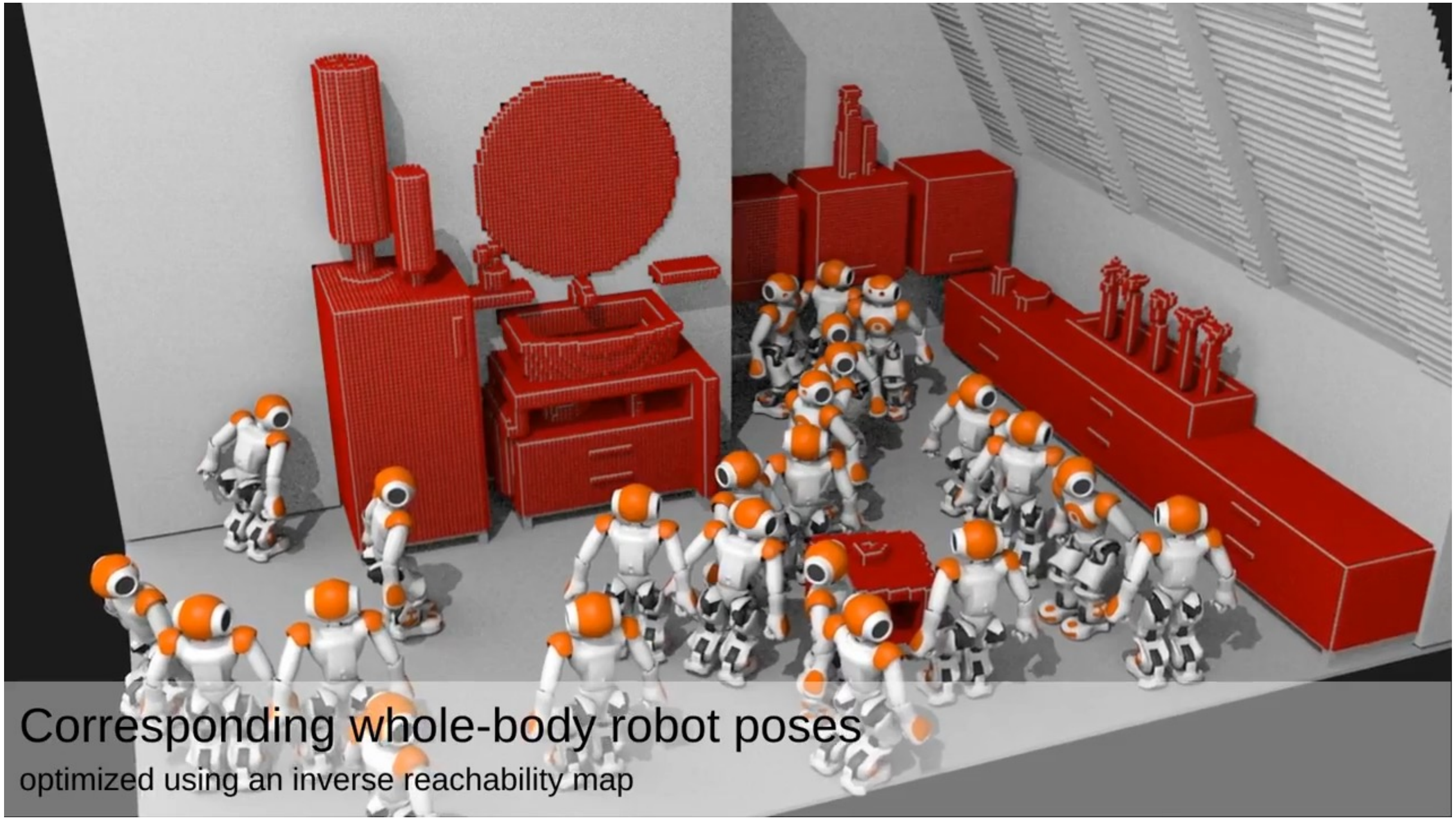
[Lenz et al., "Hortibot: An adaptive multi-arm system for robotic horticulture of sweet peppers", IROS24]

Where: Viewpoint Selection

- **Agent Pose**
  - Active selection of **agent pose** most appropriate for selecting a viewpoint most useful for current task
- **Sensor Pose**
  - Active selection of the **pose of a sensor** most appropriate for the current task, e.g., pointing a camera at a target with the best viewing angle for its recognition

[Bajcsy et al., “Revisiting Active Perception”, Autonomous Robots, 2017]

Where: Agent and Sensor Pose



[Oßwald et al., "Efficient Coverage of 3D Environments with Humanoid Robots Using Inverse Reachability Maps", Humanoids17]

How Do We Decide Where to Look or Move Next?

- Active perception is not just moving sensors—it's about **making informed decisions**
- We need a way to **evaluate** potential actions
- Core idea: **How much** new and useful information will be gained?
- Should the robot move to pose A or pose B?

Quantifying the Value of Perception

- Information-theoretic decision making
- Actions are chosen to **reduce uncertainty/entropy**
- The aim is to **maximize information gain I**
- Additionally, reduce cost of the action C, e.g.,
  - Motion Cost
  - Energy Cost
- Overall, we aim to **maximize utility U**

$$U = I - \alpha \cdot C$$

Information-Theoretic Entropy (Shannon Entropy)

- **Entropy *H*** of a random variable *X* is the amount of randomness given by

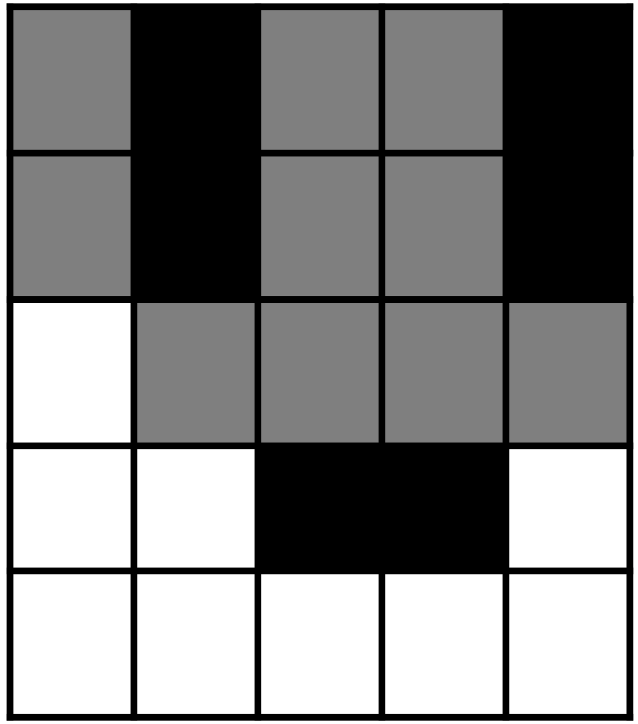
$$H(X) = -\sum p(x) \log p(x)$$

- **Information gain *I*** can be calculated as

$$I = H$$

Binary Occupancy Map Entropy Calculation

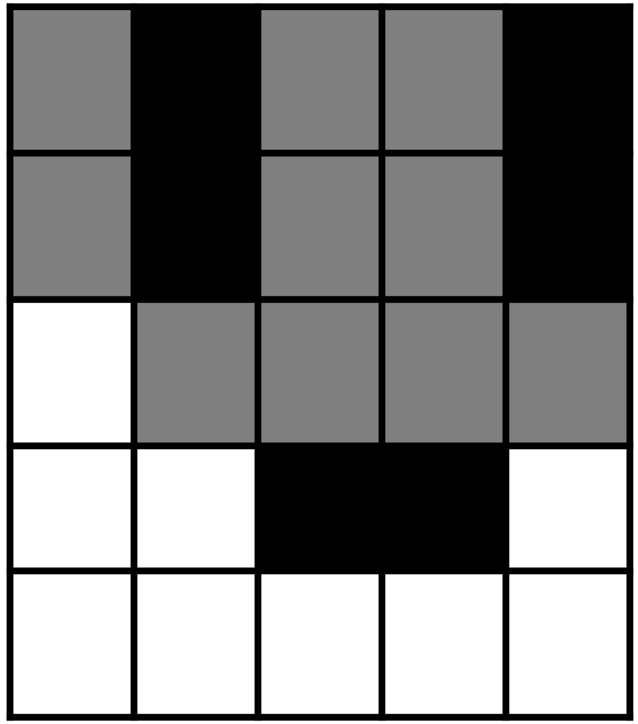
- $p(occ) = 1,\ p(free) = 0,\ p(unknown) = 0.5$
- Black = **occupied**, white = **free**, gray = **unknown**
- What is entropy of the map *m*?
- As occupancy states are binary, we use the binary entropy function



$$H(m) = -\sum p_i * \log_2 p_i + (1 - p_i) * \log_2 p(1 - p_i)$$

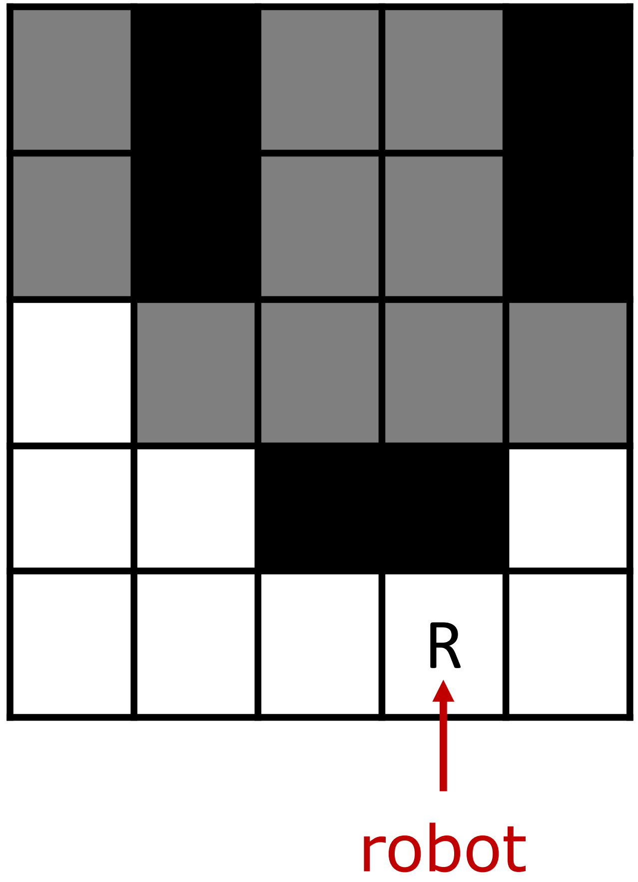
Binary Occupancy Map Entropy Calculation

- Occupied and free cells do not posses any new information/uncertainty
- Hence, their entropy is 0
- Only unknown cells (p=0.5) contribute to entropy in a binary occupancy map
- Hence, map entropy $H(m) = 10$



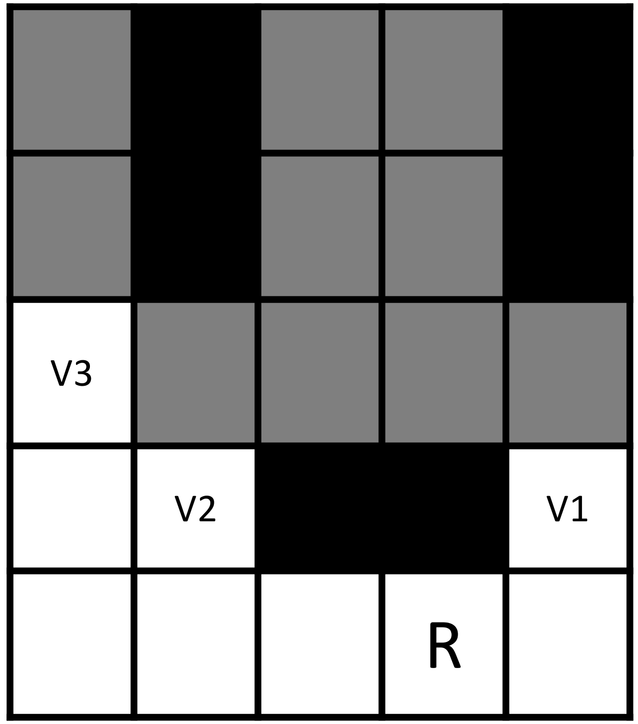
Next Best View for Entropy Reduction

- Robot can move in N, S, W, E directions
- It can only move to free cells and observe the adjacent cells in all four directions at once
- Which is the **next best view (NBV)** for entropy reduction?
- Once a cell is viewed, it leads to unit information gain irrespective of whether it turns out to be free or occupied



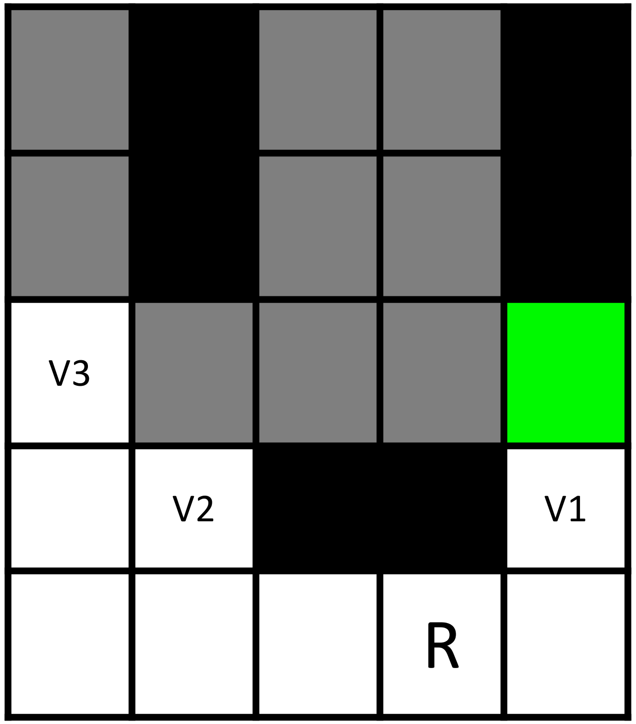
Next Best View for Entropy Reduction

- Three potential candidates for robot goal pose:
  - V1: $I_{v1} = ?$
  - V2: $I_{v2} = ?$
  - V3: $I_{v3} = ?$



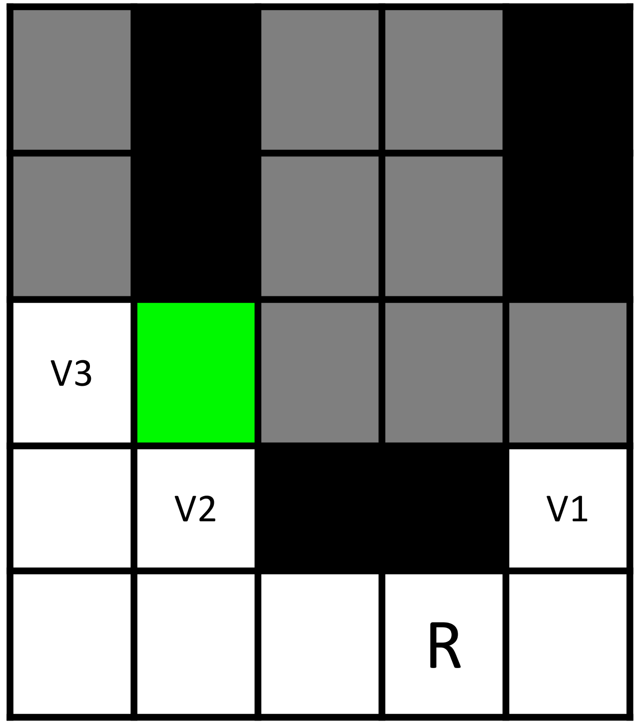
Next Best View for Entropy Reduction

- Three potential candidates for robot goal pose:
  - V1: $I_{v1} = 1$
  - V2: $I_{v2} = ?$
  - V3: $I_{v3} = ?$



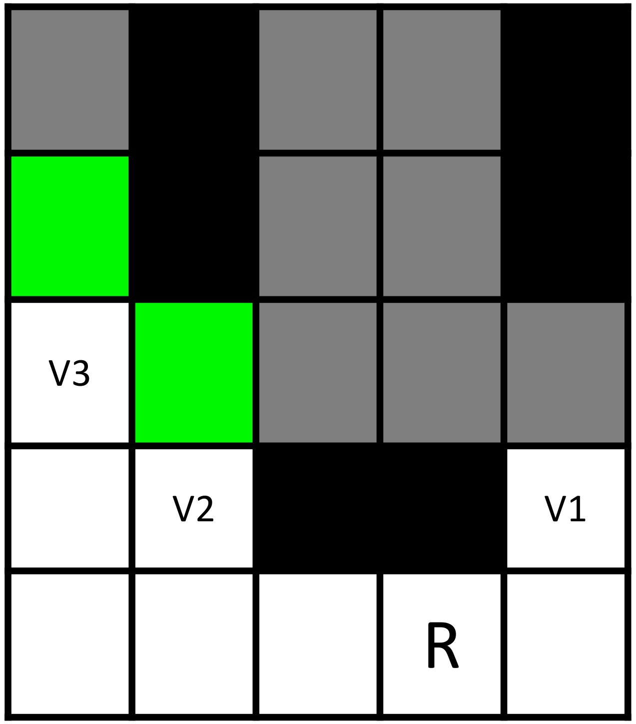
Next Best View for Entropy Reduction

- Three potential candidates for robot goal pose:
  - V1: $I_{v1} = 1$
  - V2: $I_{v2} = 1$
  - V3: $I_{v3} = ?$



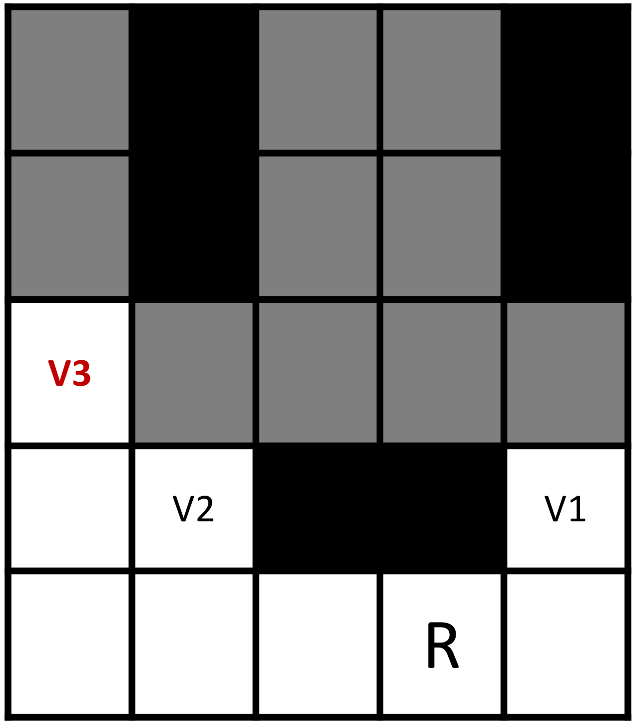
Next Best View for Entropy Reduction

- Three potential candidates for robot goal pose:
  - V1: $I_{v1} = 1$
  - V2: $I_{v2} = 1$
  - V3: $I_{v3} = 2$



Next Best View for Information Gain

- We calculated the information gain for the view cells V1, V2, V3 as follows
  - V1: $I_{v1} = 1$
  - V2: $I_{v2} = 1$
  - V3: $\boldsymbol{I_{v3} = 2}$ most informative view

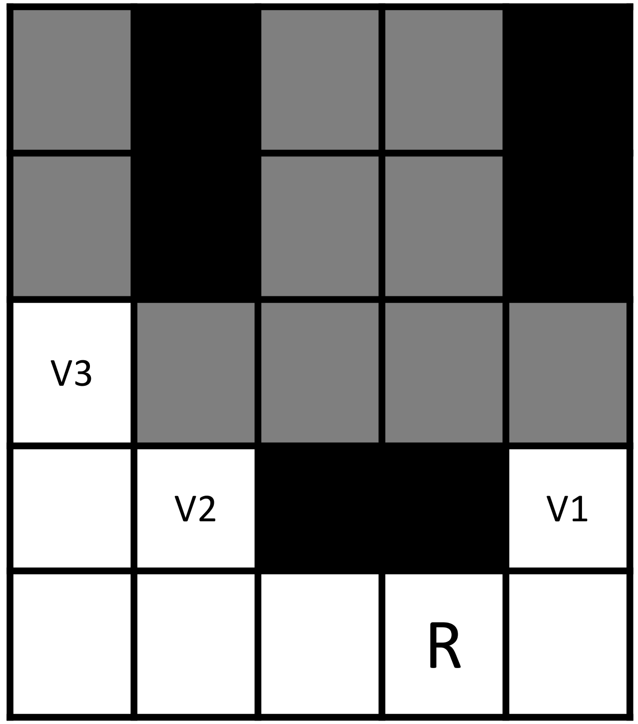


- Hence V3 is the next best view

Next Best View with Motion Cost

- For pure information gain, V3 is the next best view
- However, this evaluation did not account for motion cost
- Assume $\alpha = 0.4$ and motion cost of each traversed cell is 1 in the utility function

$$U = I - \alpha \cdot C$$

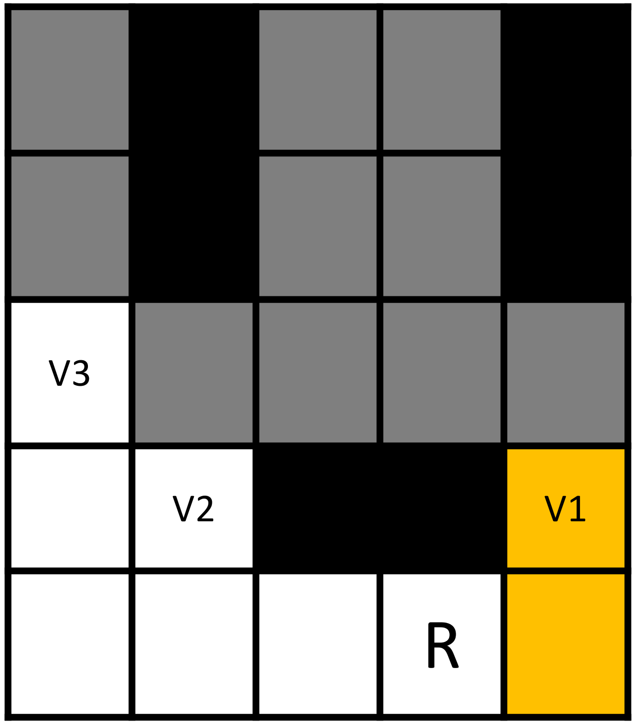


Next Best View with Motion Cost

- Assume $\alpha = 0.4$ and motion cost of each traversed cell is 1 in the utility function

$$U = I - \alpha \cdot C$$

- $U_{v1} = I_{v1} - 0.4 * C_{v1}$
- $U_{v1} = 1 - 0.4 * 2$
- $U_{v1} = 0.2$

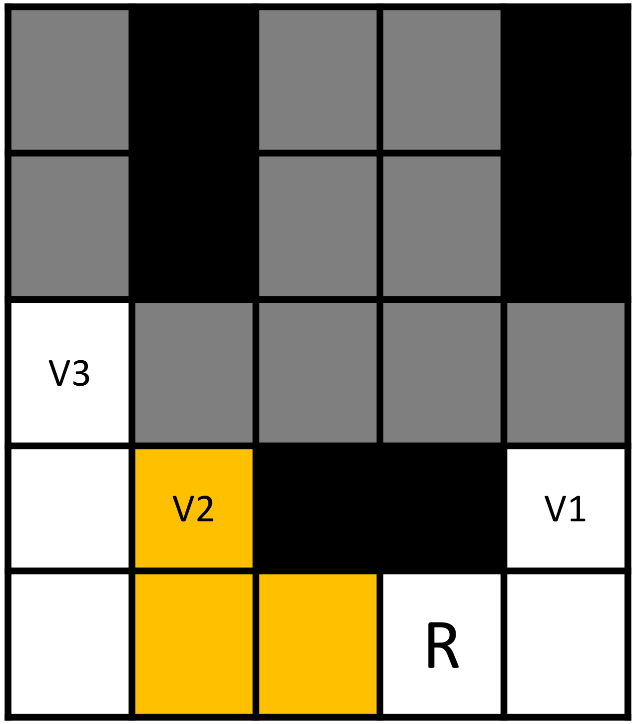


Next Best View with Motion Cost

- Assume $\alpha = 0.4$ and motion cost of each traversed cell is 1 in the utility function

$$U = I - \alpha \cdot C$$

- $U_{v2} = I_{v2} - 0.4 * C_{v2}$
- $U_{v2} = 1 - 0.4 * 3$
- $U_{v2} = -0.2$

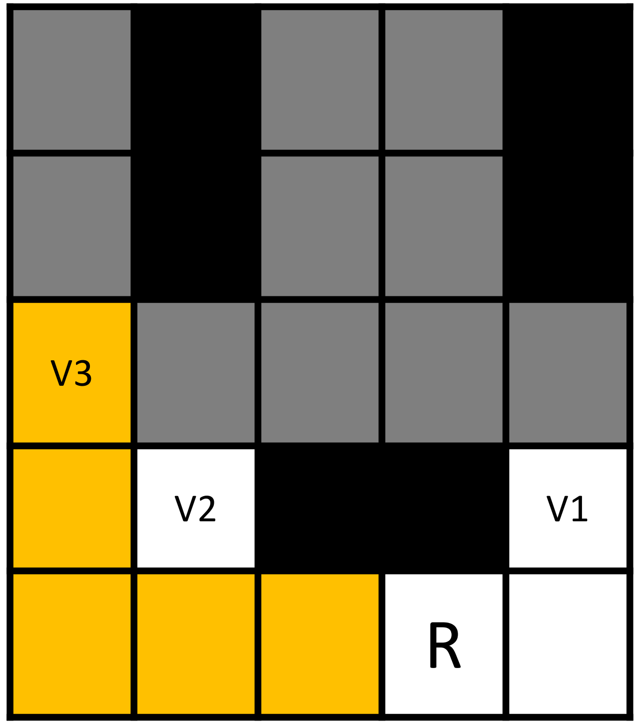


Next Best View with Motion Cost

- Assume $\alpha = 0.4$ and motion cost of each traversed cell is 1 in the utility function

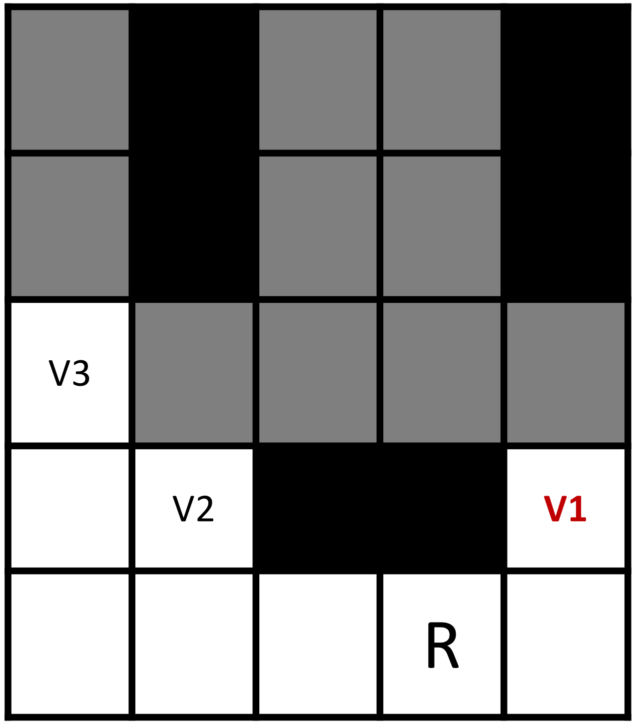
$$U = I - \alpha \cdot C$$

- $U_{v3} = I_{v3} - 0.4 * C_{v3}$
- $U_{v3} = 2 - 0.4 * 5$
- $U_{v3} = 0.0$



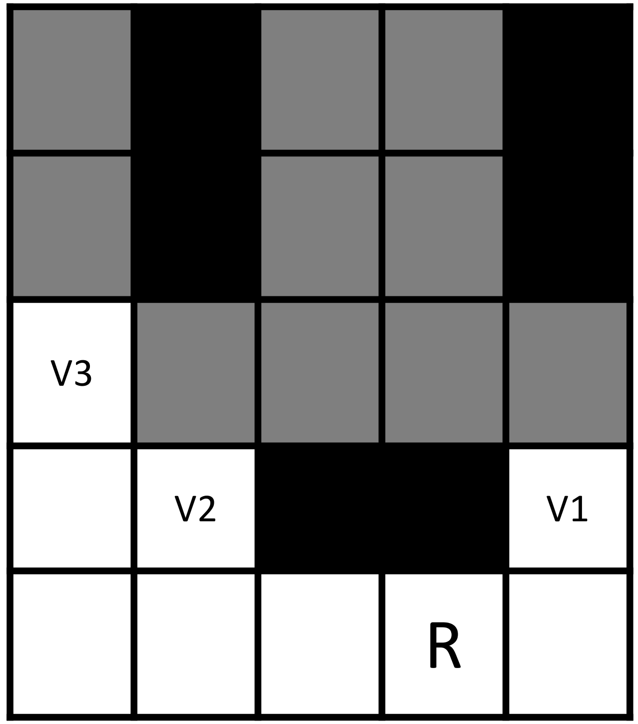
Next Best View with Motion Cost

- As can been seen, with motion cost accounted for
- $\boldsymbol{U_{v1} = 0.2}$
- $U_{v2} = -0.2$
- $U_{v3} = 0.0$
- V1 has highest utility
- Hence, V1 is the next best view



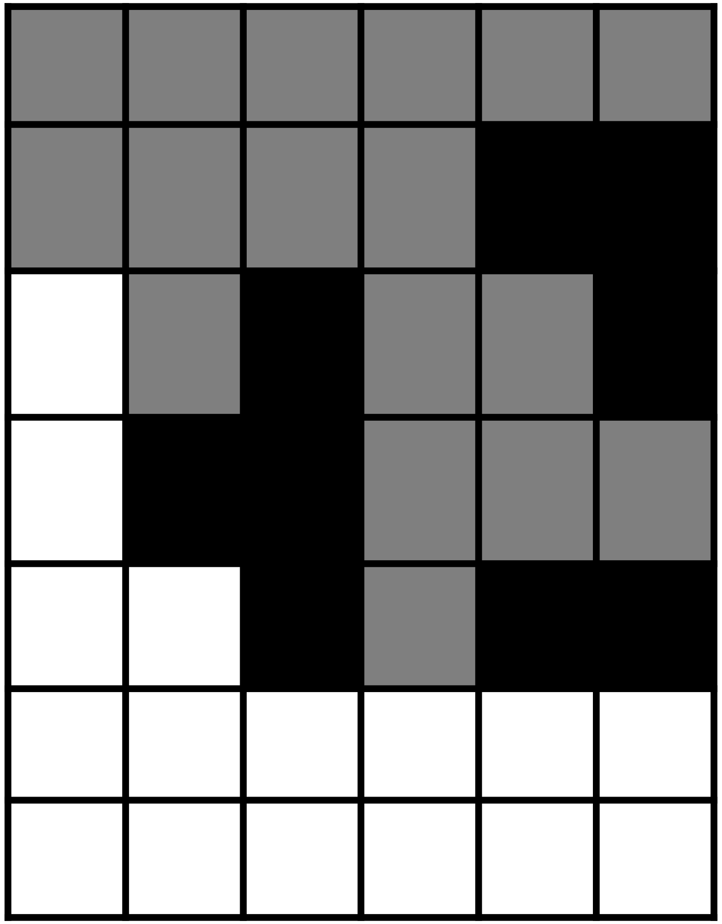
Next Best View with Motion Cost

- Without motion cost, the robot would have visited V3, then V1
- With motion cost considered, the robot visits V1, then V3
- Thus, active perception involves a **trade-off** between **information gain** from perception and **cost** from action



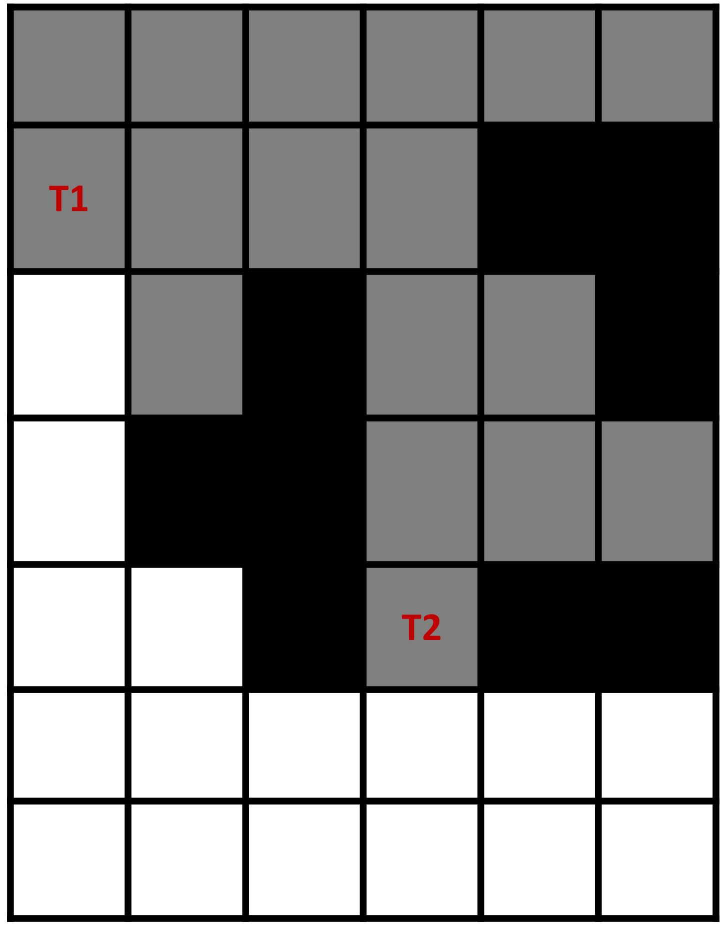
Target and NBV Sampling

- What are informative regions?
- What are candidates for view poses?
- Consider **frontier cells** at the boundary of unknown space



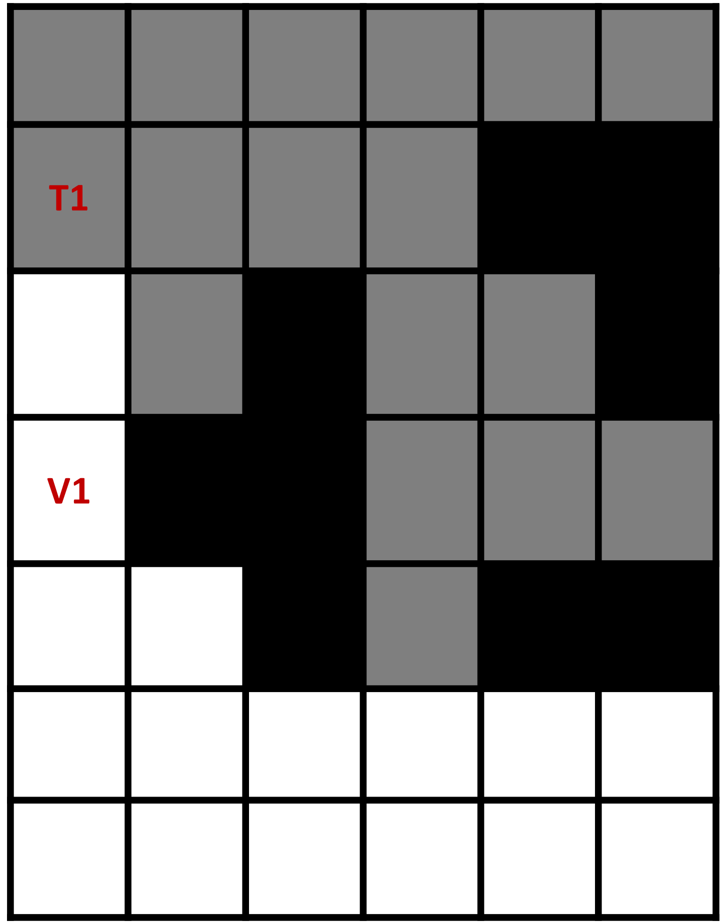
Target Region Sampling

- Depends on active perception objective
  - Active mapping
  - Active object reconstruction
- Assumption for the sensor range:
  - Target cells have to be at least 2 cells away from view cells
  - Free space visibility is up to 3 cells
- T1 and T2 are two potential but different kinds of target cells



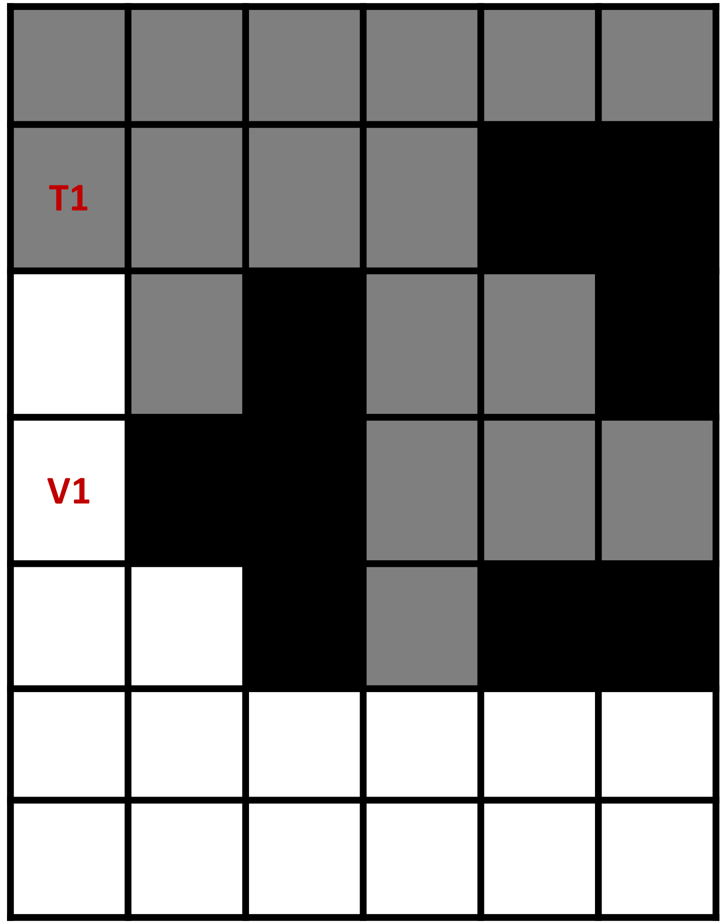
Free-Unknown Border Sampling

- T1 is an unknown cell at the border of free and unknown region
- V1 is a potential view pose for T1



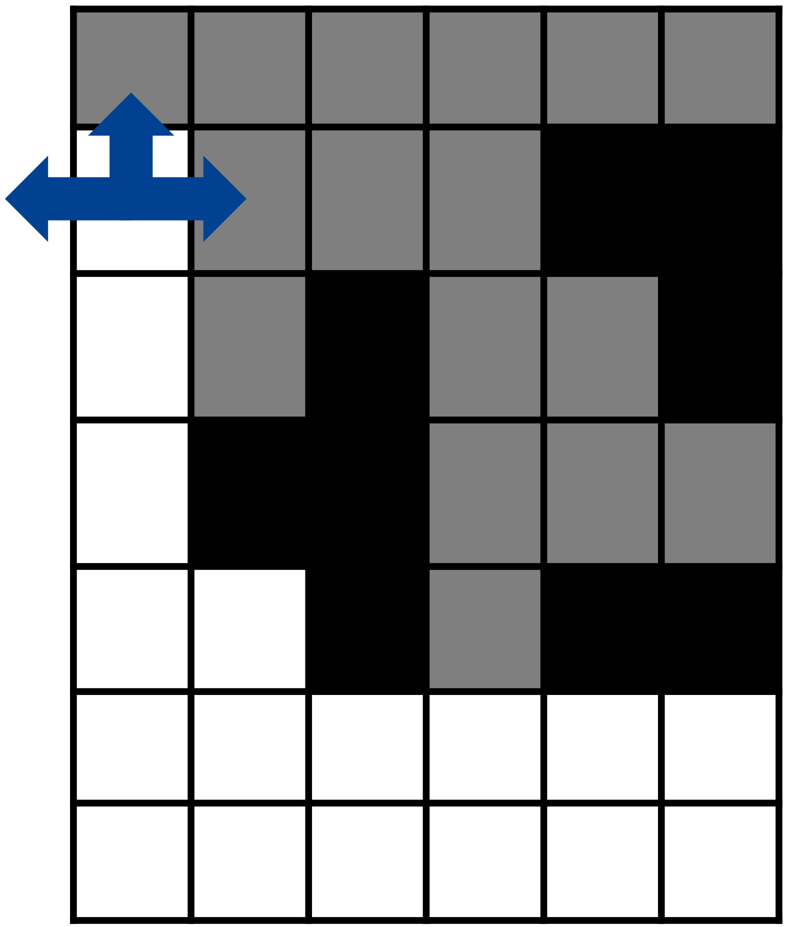
Free-Unknown Border Sampling

- T1 is an unknown cell at the border of free and unknown region
- V1 is a potential view pose for T1
- If T1 is free it enables the robot to uncover new regions by traveling to T1



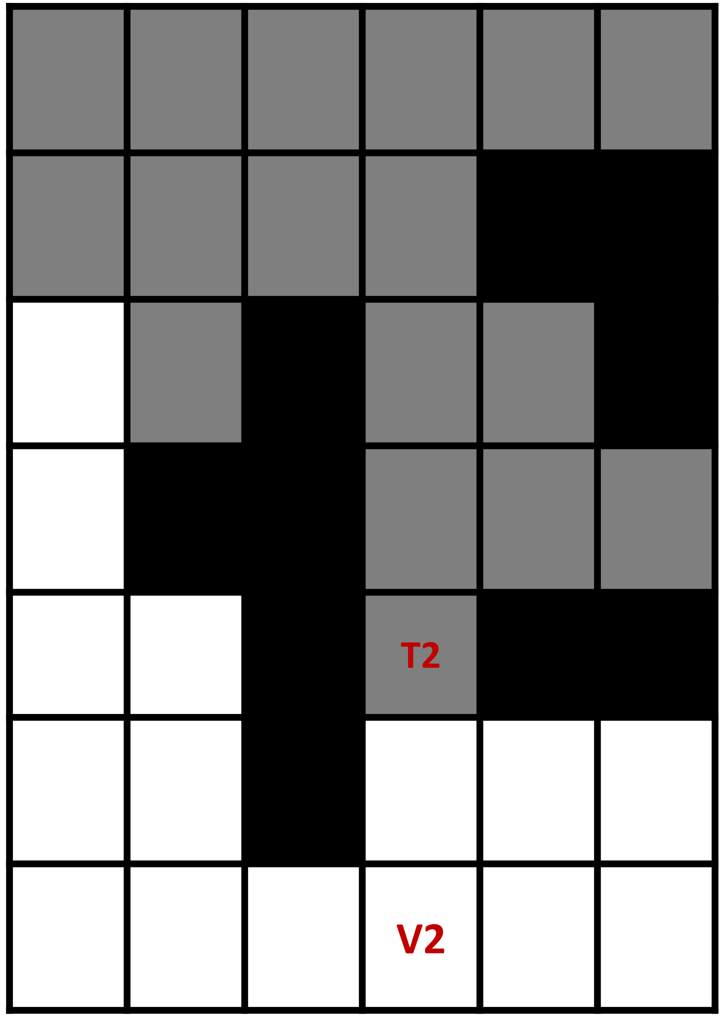
Free-Unknown Border Sampling

- Suppose T1 is free
- Robot travels to T1
- It can explore new map frontiers by looking in 3 directions
- Useful for active exploration of unknown regions



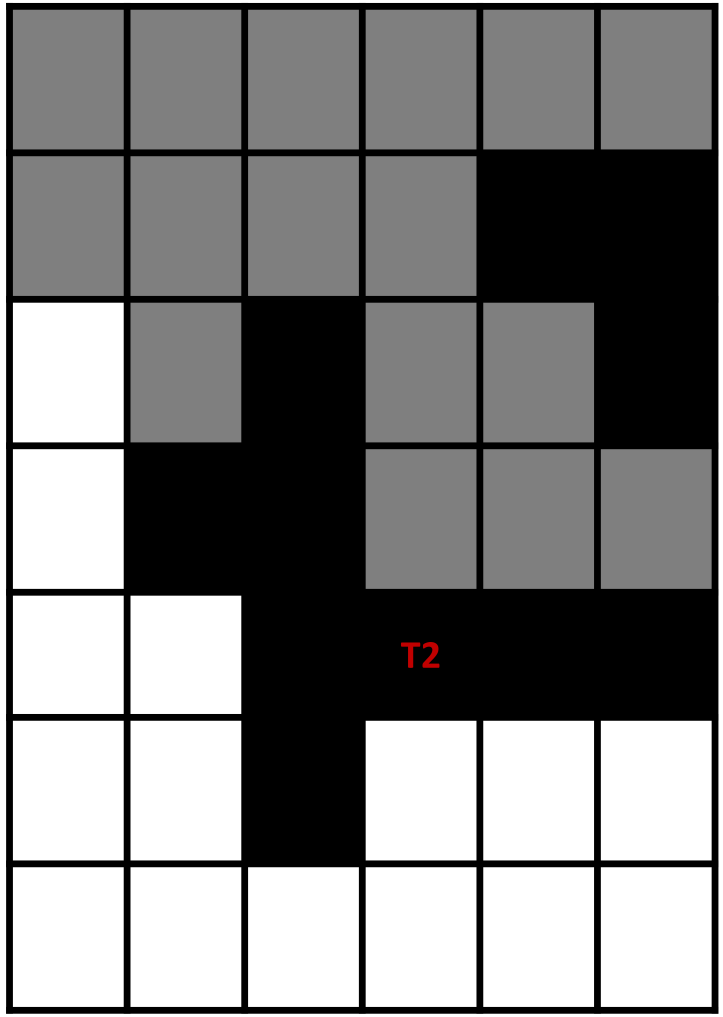
Occupied-Unknown Border Sampling

- T2 is at the border between occupied and unknown
- V2 is a view pose for T2
- High chances T2 is also occupied



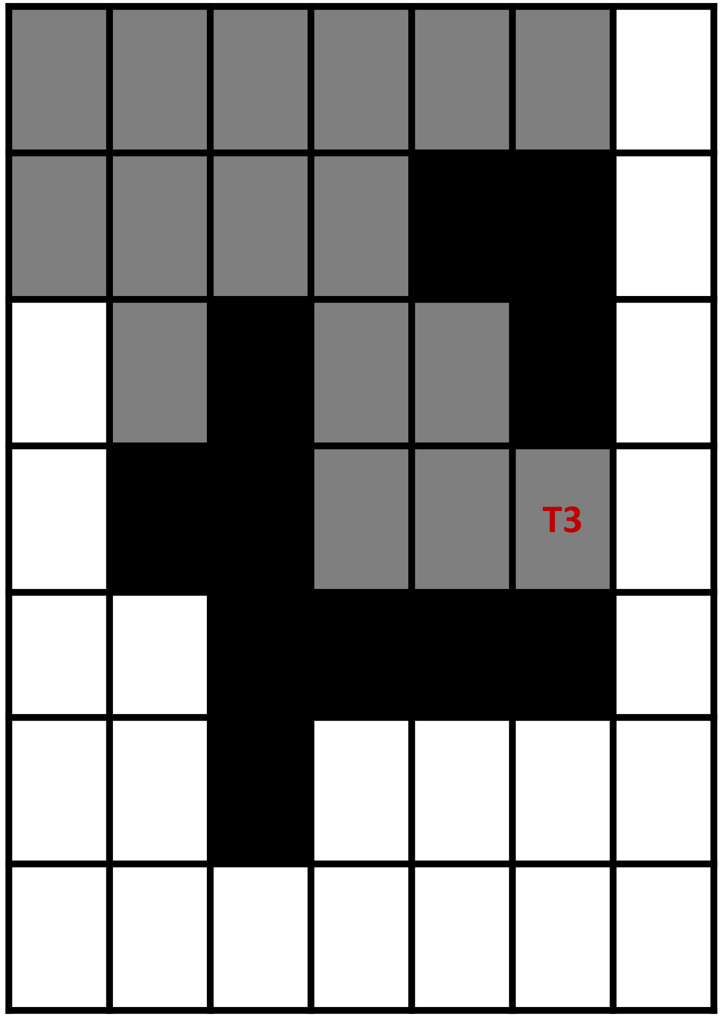
Occupied-Unknown Border Sampling

- If T2 occupied, it probably represents a wall/object surface
- Enables to create map of occupied regions/obstacles for navigation
- Used also for active object reconstruction
  - Aim is to uncover occluded regions of target object



Occupied-Unknown Border Sampling

- If T2 occupied, it probably represents a wall/object surface
- Enables to create map of occupied regions/obstacles for navigation
- Used also for active object reconstruction
  - Aim is to uncover occluded regions of target object
  - Next view potentially target T3



Active Vision for Closed-Loop Grasping

[Breyer et al., "Closed-loop next-best-view planning for target-driven grasping”, IROS22]

Binary to Continuous Maps

- In practice, we use maps with continuous occupancy probabilities
- $$\mathrm{s(x)} = \begin{cases} occupied, & if\ p(x) > 0.7 \\ unknown, & if\ 0.3 \leq p(x) \leq 0.7 \\ free, & if\ p(x) < 0.3 \end{cases}$$
- Hence, entropy calculation is more involved

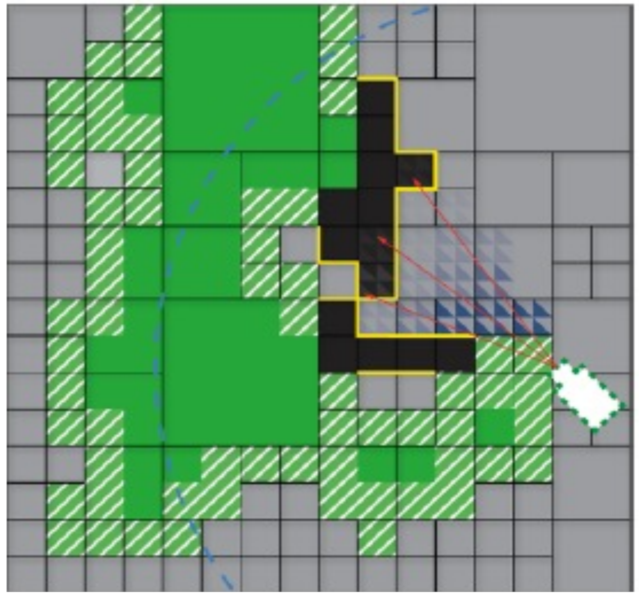
Extension to 3D

- Information gain for 3D volumetric map
- Consider sensor field of view and sensor range to estimate information gain of observation
- Weigh each observed voxel's entropy by its visibility likelihood from candidate view
- Different metrics exist to calculate the volumetric information gain (VI)

Explanation for Visualization

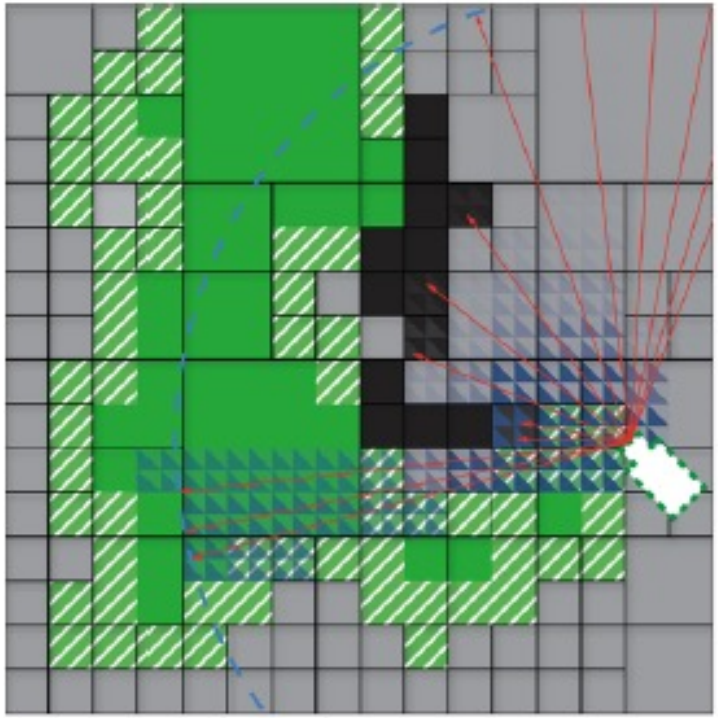
Shown in 2D on an exemplary state of the map

- Likely occupied (black)
- Unknown (grey)
- Likely free (green)
- Frontier voxels (striped white)
- Unknown object sides (yellow)
- View candidate (white camera)
- Sensor rays (red)
- Maximal ray length (dashed blue circle)
- VI weights (opacity of blue triangles)



[Delmerico et al., "A comparison of volumetric information gain metrics for active 3D object reconstruction", Autonomous Robots, 2018]

Occlusion-Aware VI



- Consider likelihood P_v of a voxel x_n **being visible** from a particular view, instead of simply integrating entropy over all traversed voxels
- $P_v(x_n) = \prod_{i=1}^{n-1}(1 - P_o(x_i)\ ),$

  where $\mathrm{P_o(x_i)}$: *occupancy probability of voxel* $\mathrm{x_i}$
- Occlusion-aware VI

$$I_v = P_v(x)H(x)$$

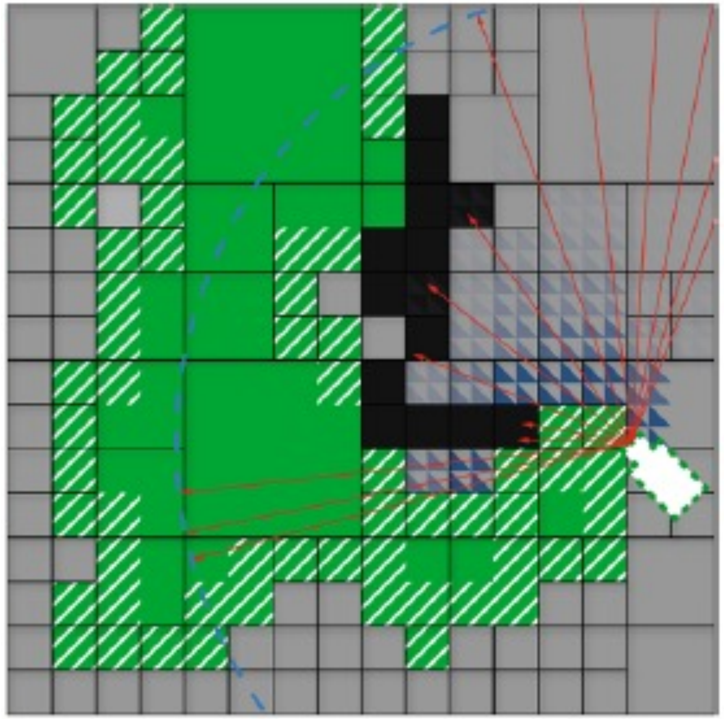
[Delmerico et al., "A comparison of volumetric information gain metrics for active 3D object reconstruction", Autonomous Robots, 2018]

Unobserved Voxel VI

- Remove all voxels already observed with a high degree of certainty

$$\mathcal{I}_u(x) = \begin{cases} 1 & x \text{ is unobserved} \\ 0 & x \text{ is already observed} \end{cases}$$

$$\mathcal{I}_k(x) = \mathcal{I}_u(x)\,\mathcal{I}_v(x)$$



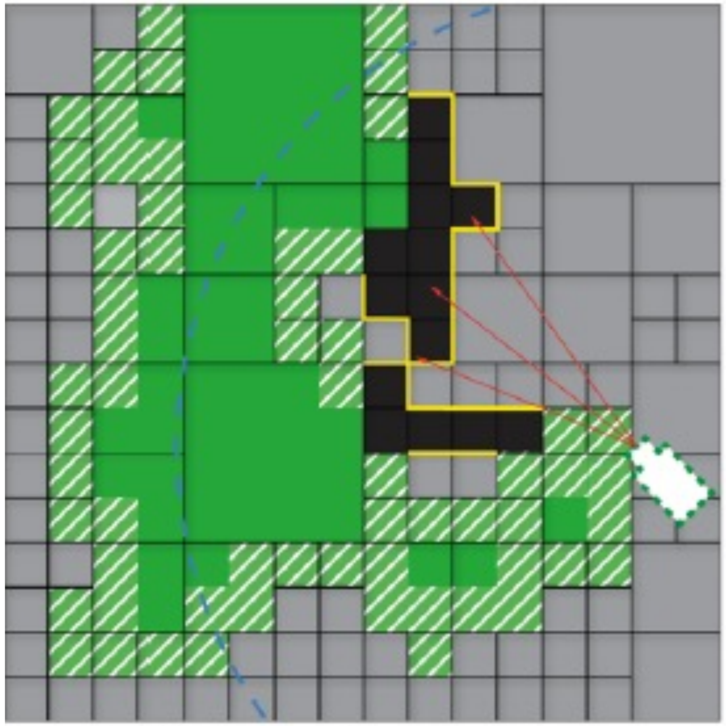
[Delmerico et al., "A comparison of volumetric information gain metrics for active 3D object reconstruction", Autonomous Robots, 2018]

Rear Side Voxel VI

- For object reconstruction, consider unobserved voxels at the border of occupied regions

$$\mathcal{I}_b(x) = \begin{cases} 1 & x \in \mathcal{S}_o \\ 0 & x \notin \mathcal{S}_o \end{cases}$$

- S_o: **unobserved voxels** such that the **next voxel on their ray** is estimated to be **occupied**



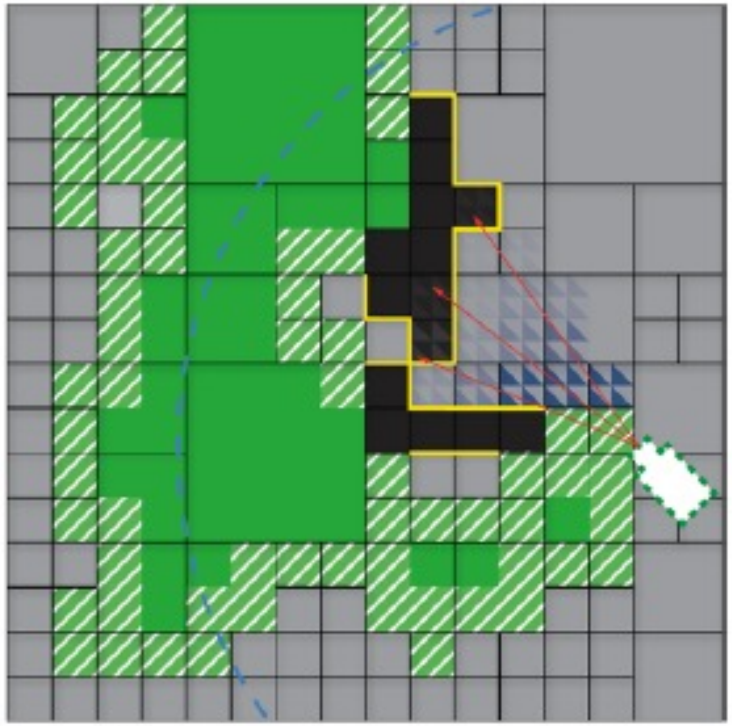
[Delmerico et al., "A comparison of volumetric information gain metrics for active 3D object reconstruction", Autonomous Robots, 2018]

Rear Side Voxel VI

- Combined with occlusion-aware VI:

$$I_n(x) = I_b(x) \cdot I_v(x)$$

- Focuses on unknown voxels between sensor and occupied voxels



[Delmerico et al., "A comparison of volumetric information gain metrics for active 3D object reconstruction", Autonomous Robots, 2018]

Next Best View Planning with Occlusion-Aware VI

NBV-SC:
Next Best View Planning Based on Shape Completion for Fruit Mapping and Reconstruction

Rohit Menon, Tobias Zaenker, Nils Dengler, and Maren Bennewitz
Humanoid Robots Lab, University of Bonn

[Menon et al., "Next Best View Planning Based on Shape Completion for Fruit Mapping and Reconstruction, IROS23]

NBV Planning: Good Enough?

- Selects the view that maximizes immediate entropy reduction
- **Single-Step Lookahead**: Decisions are made based solely on the next best candidate
- Does not account for future views or overlapping information
- Can lead to redundant or myopic decisions if similar areas are repeatedly chosen

Next-Best View Traversal

Submodular Information Gain

- Recognizes that additional views yield less new information as overlap increases
- A set function f is **submodular** if it exhibits diminishing returns: for any sets $A \subseteq B$ and any candidate view s

$$f(A \cup \{s\}) - f(A) \geq f(B \cup \{s\}) - f(B)$$

- Thus, the **incremental benefit of adding a new view decreases** as the **set of views grows**
- Overall information gain of a map via additional observations exhibits submodular behaviour

N-Step Greedy Planning

- Instead of single step NBV, *n*-step greedy planning
- Evaluate **sequences of *n* actions** to estimate cumulative information gain
- Compute the **total expected gain** over *n* steps and choose the view sequence that maximizes this sum
- Greedy selection provides **strong theoretical guarantees with low *n***
- Reason: submodular property of information gain, i.e., the incremental benefit of an extra view diminishes as more views are added

Receding Horizon Planner

- **Planning horizon** (*n* steps): Compute an optimal sequence over *n* steps
- **Execution window** (*m* steps, $m < n$):
  - Execute only the first *m* actions
  - Replan after *m* steps with updated state information
- **Continuous replanning**: Adapt to dynamic changes and new observations

Receding Horizon Planner for Active Perception for Mobile Manipulation

[Jauhri et al., "Active-perceptive motion generation for mobile manipulation", ICRA24]

NBV vs. One-Shot Global Planners

- Next-best view (NBV)
  - Adaptive view placement
  - Suboptimal path

Next-best view paths

NBV vs. One-Shot Global Planners

- Next-best view (NBV)
  - Adaptive view placement
  - Suboptimal path

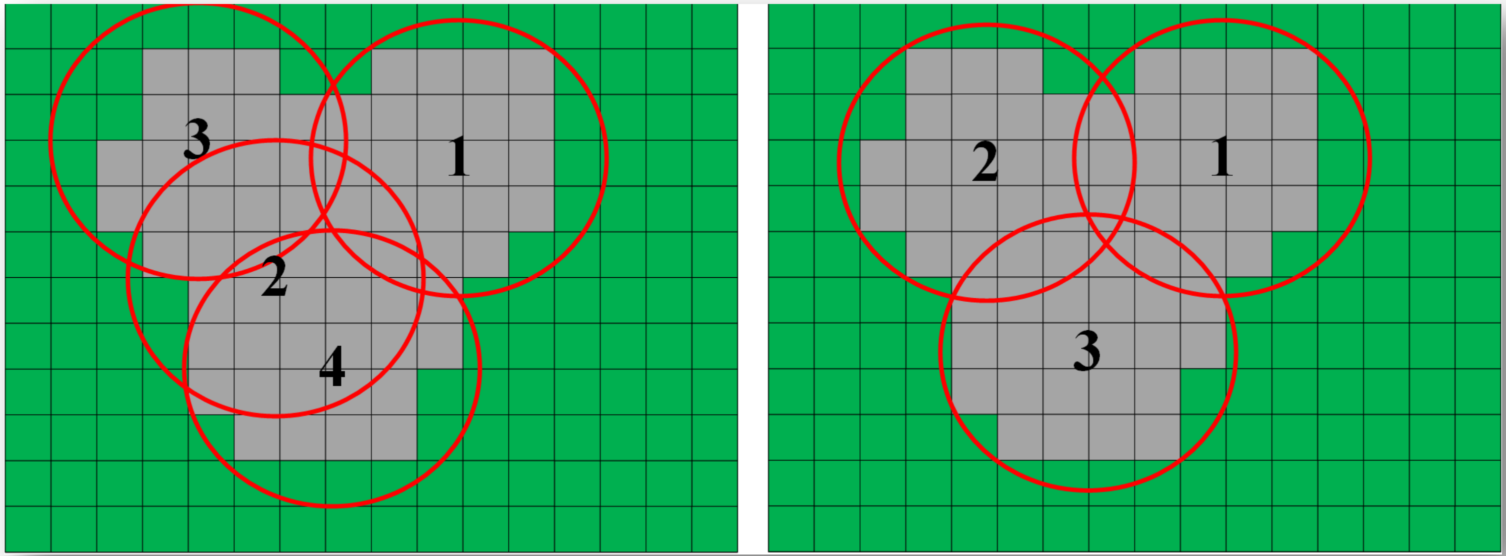
Next-best view paths

- One-shot view path
  - Fixed view configuration
  - Globally shortest path

One-shot view paths

Coverage Maximization

- Set Covering Optimization Problem: cover all surfaces with the smallest set of views

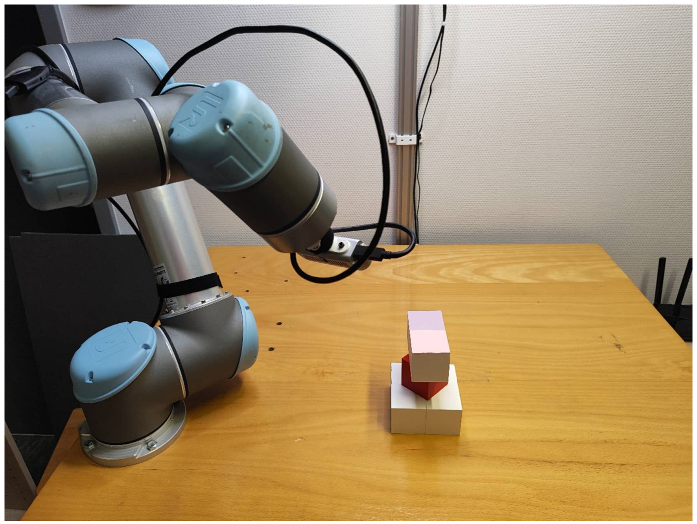


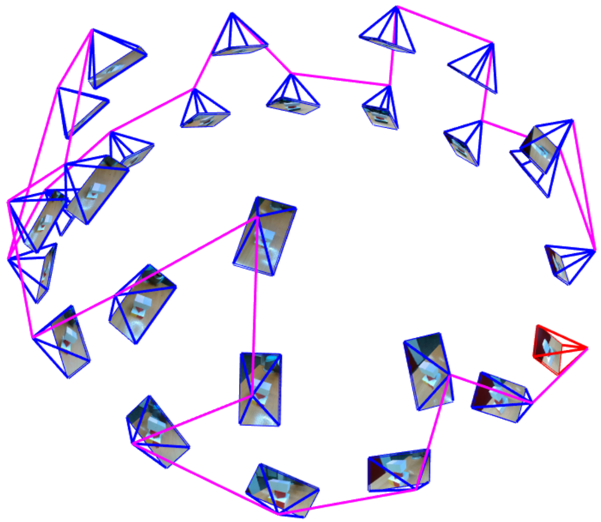
[Pan and Wei, "A global generalized maximum coverage-based solution to the non-model-based view planning problem for object reconstruction", Computer Vision and Image Understanding, 2023]

Global Multi-View Planning

Active object reconstruction with NeRFs

- A small number of informative views
- Minimize robot movement cost



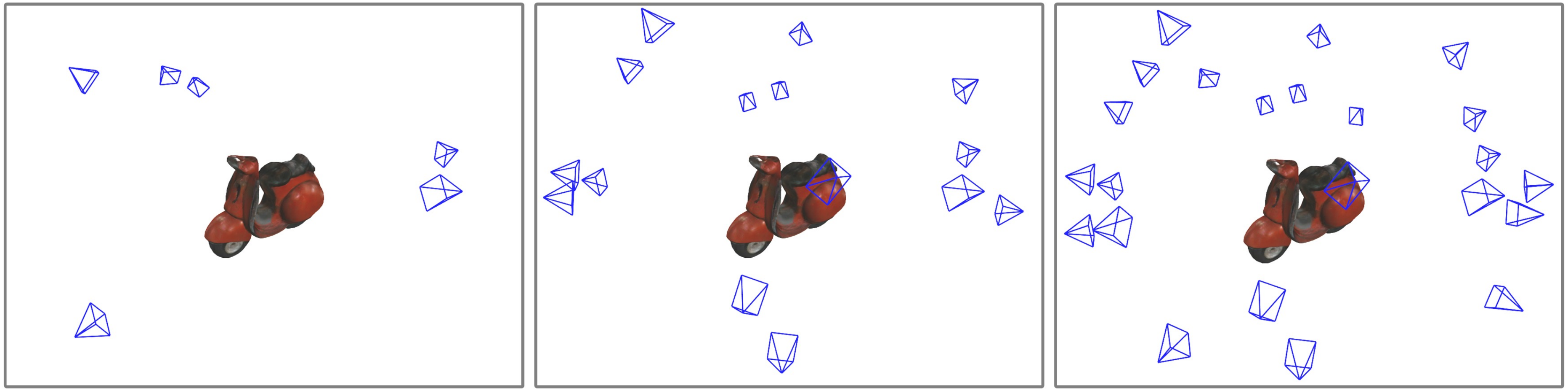


Customized Set Covering Optimization

- Customized multi-view constraint
  - NeRF representation learning is achieved by minimizing the photometric loss

Customized Set Covering Optimization

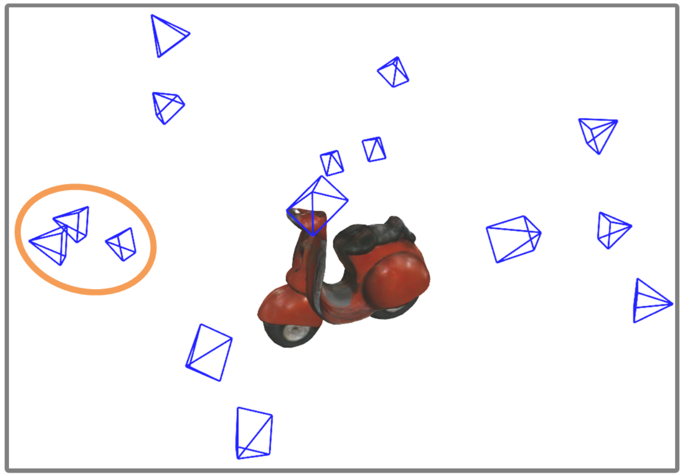
- Customized multi-view constraint
  - NeRF representation learning is achieved by minimizing the photometric loss
  - Cover each surface point by at least α view



$\alpha = 1$ $\alpha = 3$ $\alpha = 5$

Customized Set Covering Optimization

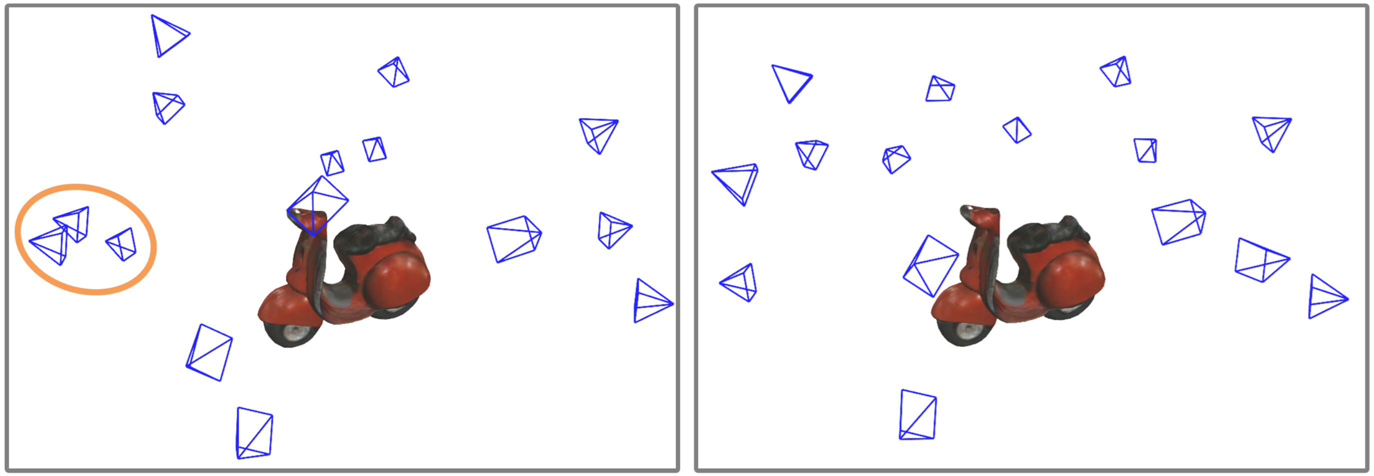
- Customized distance constraint
  - Feasible solutions of spatially clustered views with redundant information



w/o Distance Constraints

Customized Set Covering Optimization

- Customized distance constraint
  - Feasible solutions of spatially clustered views with redundant information
  - Find the most spatially uniform views



w/o Distance Constraints　　w/ Distance Constraints

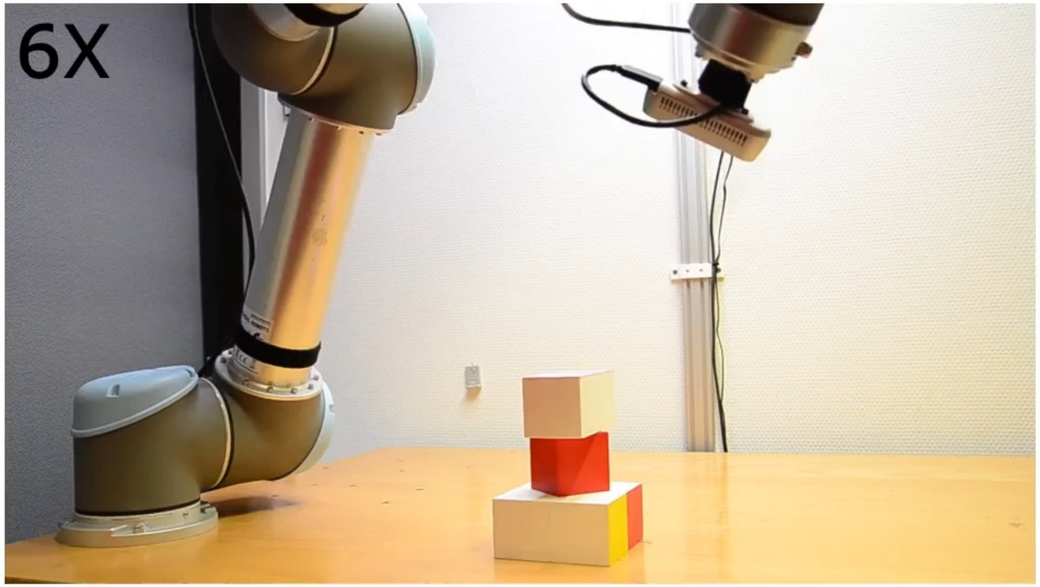
Optimization via Constrained Integer Linear Programming

- Minimize the total number of selected views
- Subject to multi-view and distance constraints

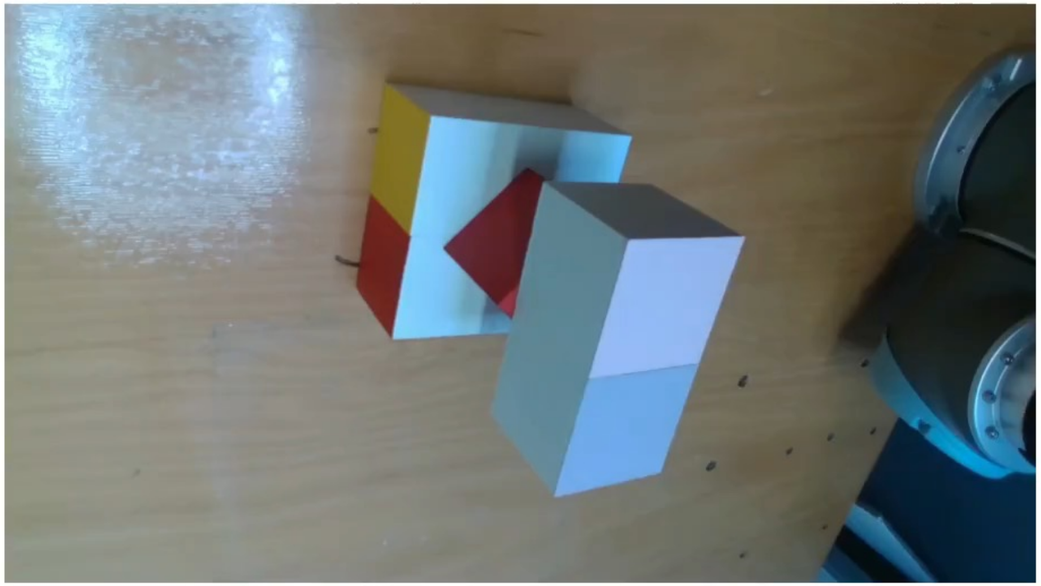
$$
\begin{aligned}
\min : \quad & \sum_{v \in \mathcal{V}} x_v \,, & & \\
\text{s.t.} : \quad & (a) \quad x_v \in \{0, 1\} & \forall v \in \mathcal{V} \quad & \text{decision variables} \\
& (b) \quad \sum_{v \in \mathcal{V}} I(p, v)\, x_v \geq \alpha & \forall p \in \mathcal{P}_{surf} \quad & \text{multi-view constraint} \\
& (c) \quad x_v + x_{v'} \leq 1 & \forall d_v^{v'} \leq D(v) \quad & \text{distance constraint}
\end{aligned}
$$

[Pan et al., IROS24]

Real World Environment



Real World Scene



In-Hand View

[Pan et al., IROS24]

Different IG Formulations

- Entropy is not the only information gain metric
- Other IG formulations are:
  - TSDF reconstruction based IG
  - NeRF uncertainty based IG
  - Fisher mutual information
  - Predicted variance for Bayesian neural networks

Active Perception for Different Objectives

- Active perception can also be used for other tasks
  - Semantic mapping
  - Object search
  - Localization
- Also including knowledge from LLMs
- The basic principles, however, remain the same

Summary

- Active perception is needed to **efficiently gain relevant information** about the environment
- Uses the **expected information gain**
- Different strategies to gather information exist
- The **costs of acquiring new sensor** data have to be taken into account
- Various applications exist, e.g., mapping, object search, 3D reconstruction etc.

Literature

- *Active perception*,
  Bajcsy R., Proceedings of the IEEE, 1988
- *Revisiting active perception,*
  Bajcsy, R., Aloimonos, Y., & Tsotsos, J. K., Autonomous Robots, 2018
- *Hortibot: An adaptive multi-arm system for robotic horticulture of sweet peppers,*
  Lenz, C., Menon, R., Schreiber, M., Jacob, M.P., Behnke, S. and Bennewitz, M., IEEE/RSJ Int. Conf. on Int. Robots and Systems (IROS), 2024
- *Efficient coverage of 3D environments with humanoid robots using inverse reachability maps,*
  Oßwald, S., Karkowski, P., & Bennewitz, M., IEEE/RAS Int. Conf. on Humanoid Robotics (Humanoids), 2017
- *Closed-loop next-best-view planning for target-driven grasping,*
  Breyer, M., Ott, L., Siegwart, R., & Chung, J. J., IEEE/ RSJ Int. Conf. on Int. Robots and Systems (IROS), 2022
- *A comparison of volumetric information gain metrics for active 3D object reconstruction,*
  Delmerico, J., Isler, S., Sabzevari, R., & Scaramuzza, D., Autonomous Robots 2018

Literature

- *NBV-SC: Next best view planning based on shape completion for fruit mapping and reconstruction*,
  Menon, R., Zaenker, T., Dengler, N., & Bennewitz M., IEEE/RSJ Int. Conf. on Int. Robots and Systems (IROS), 2023
- *Active-perceptive motion generation for mobile manipulation,*
  Jauhri, S., Lueth, S., & Chalvatzaki, G., IEEE/RAS Int. Conf. on Robotics and Automation (ICRA), 2024
- *A global generalized maximum coverage-based solution to the non-model-based view planning problem for object reconstruction,*
  Pan, S., & Wei, H., Computer Vision and Image Understanding, 2023
- *SCVP: Learning one-shot view planning via set covering for unknown object reconstruction,*
  Pan, S., Hu, H., & Wei, H., IEEE Robotics and Automation Letters (R-AL), 2022
- *How many views are needed to reconstruct an unknown object using NeRF?,*
  Pan, S., Jin, L., Hu, H., Popović, M., & Bennewitz, M., IEEE/RAS Int. Conf. on Robotics and Automation (ICRA), 2024