

Supplementary material for FlySearch: Exploring how vision-language models explore

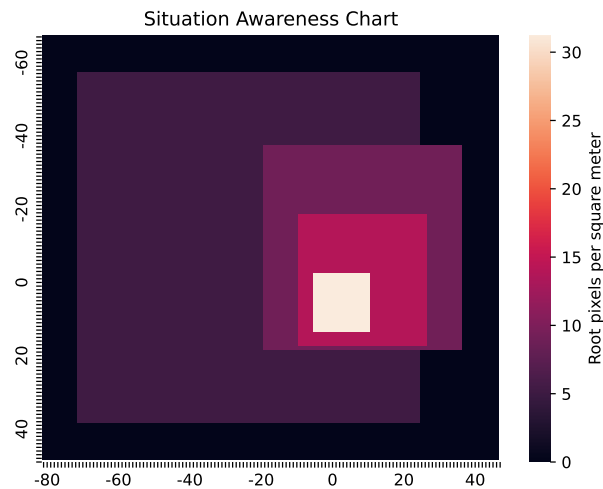
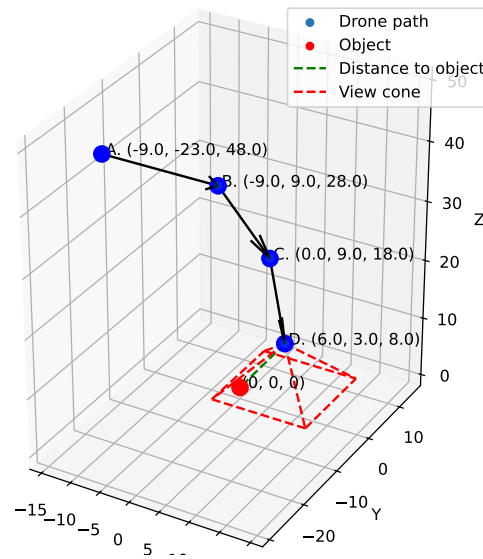
1 Example trajectories

Sample trajectories of different VLMs on FS-1 and FS-2 scenarios.

1.1 GPT-4o in the City; Success; FS-1

1.1.1 Example

- Starting position: (-9.0, -23.0, 48.0)
- End position: (6.0, 3.0, 8.0)
- Euclidean distance from the object: 10.44030650891055
- Object can be seen: True
- Success: True
- Object type: a large trash pile



BENCHMARK : <Context> You are in command of a UAV, tasked with finding a large trash pile.
</Context> (REST OF THE PROMPT)

BENCHMARK *presents GPT-4O with a relevant image*



BENCHMARK : Your current altitude is 48 meters above ground level.

GPT-4O : <Reasoning> The trash pile is not visible in the current field of view. I will start by moving south to get a better view of the area below. I'll also decrease the altitude to get closer to the ground, which will help in identifying the trash pile. </Reasoning> <Action>(0, -32, -20)</Action>

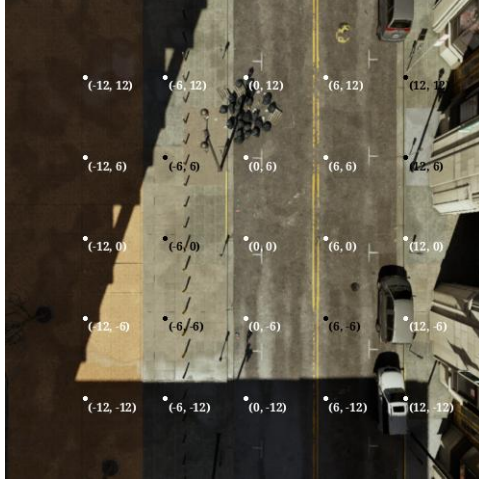
BENCHMARK *presents GPT-4O with a relevant image*



BENCHMARK : Your current altitude is 28 meters above ground level.

GPT-4O : <Reasoning> The trash pile is not visible yet. I will move slightly east and lower the altitude to improve visibility and continue searching for the trash pile. </Reasoning> <Action>(9, 0, -10)</Action>

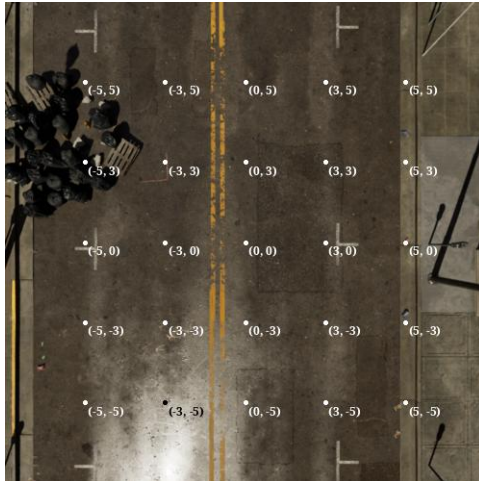
BENCHMARK *presents GPT-4O with a relevant image*



BENCHMARK : Your current altitude is 18 meters above ground level.

GPT-4O : <Reasoning> There is a pile of objects slightly northeast, which might be the trash pile. I will move closer and lower the altitude to inspect it more closely. </Reasoning>
 <Action>(6, 6, -10)</Action>

BENCHMARK *presents GPT-4O with a relevant image*

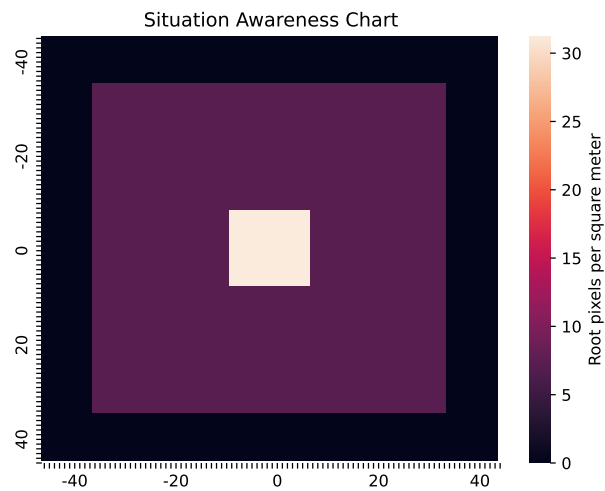
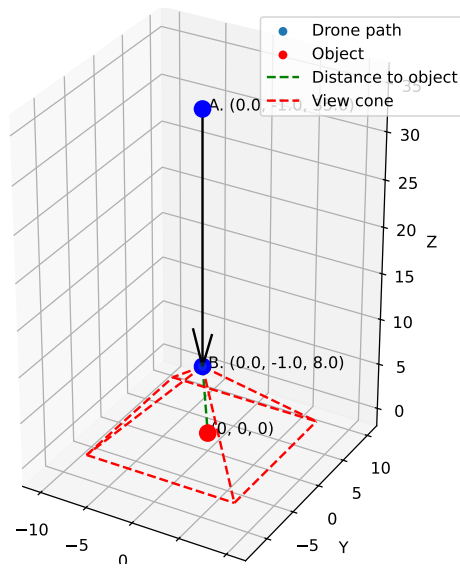


BENCHMARK : Your current altitude is 8 meters above ground level.

GPT-4O : <Reasoning> The trash pile is clearly visible and centered in the view. I'm now at an altitude of 8 meters, which is below the required 10 meters. I will confirm the find. </Reasoning> <Action>FOUND</Action>

1.1.2 Example

- Starting position: (0.0, -1.0, 35.0)
- End position: (0.0, -1.0, 8.0)
- Euclidean distance from the object: 8.06225774829855
- Object can be seen: True
- Success: True
- Object type: a blue sport car



BENCHMARK : <Context> You are in command of a UAV, tasked with finding a blue sport car.
 </Context> (REST OF THE PROMPT)

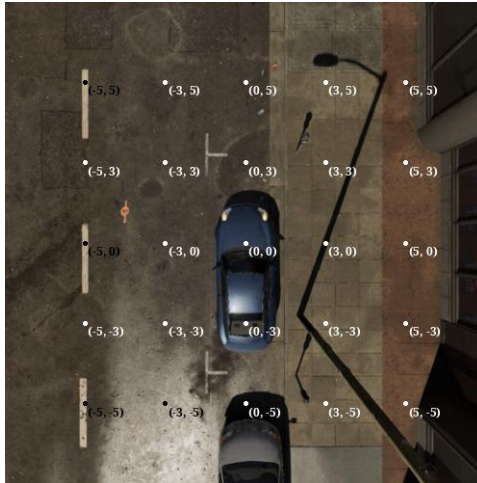
BENCHMARK presents GPT-4O with a relevant image



BENCHMARK : Your current altitude is 35 meters above ground level.

GPT-4o : <Reasoning>The blue car is visible at the center of the grid. I need to descend to be within 10 meters above it to confirm the sighting.</Reasoning> <Action>(0, 0, -27)</Action>

BENCHMARK *presents GPT-4o with a relevant image*



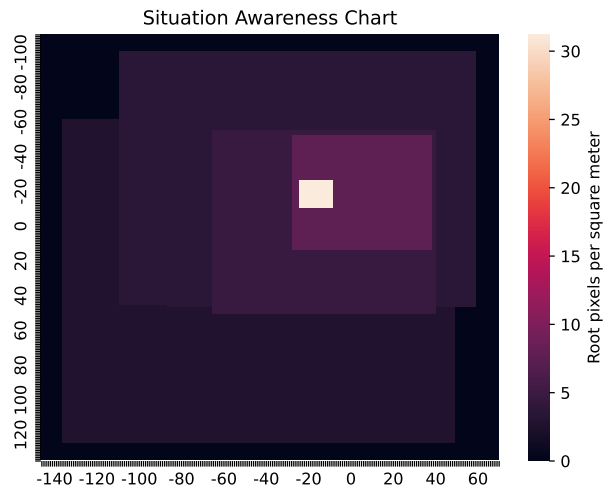
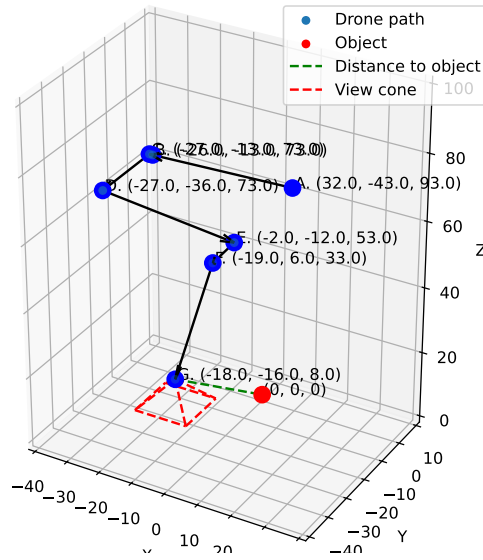
BENCHMARK : Your current altitude is 8 meters above ground level.

GPT-4o : FOUND

1.2 GPT-4o in the City; Failure; FS-1

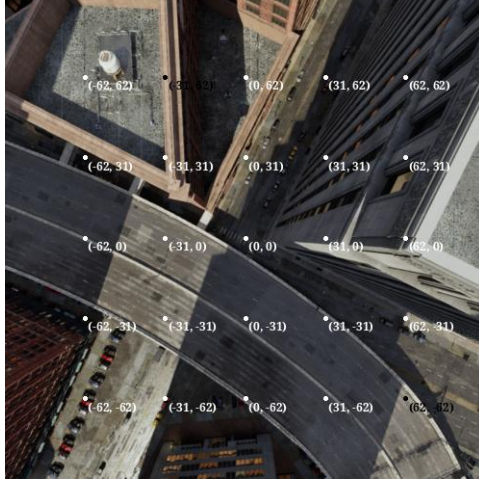
1.2.1 Example

- Starting position: (32.0, -43.0, 93.0)
- End position: (-18.0, -16.0, 8.0)
- Euclidean distance from the object: 25.37715508089904
- Object can be seen: False
- Success: False
- Object type: a black truck



BENCHMARK : <Context> You are in command of a UAV, tasked with finding a black truck.
</Context> (REST OF THE PROMPT)

BENCHMARK presents GPT-4O with a relevant image

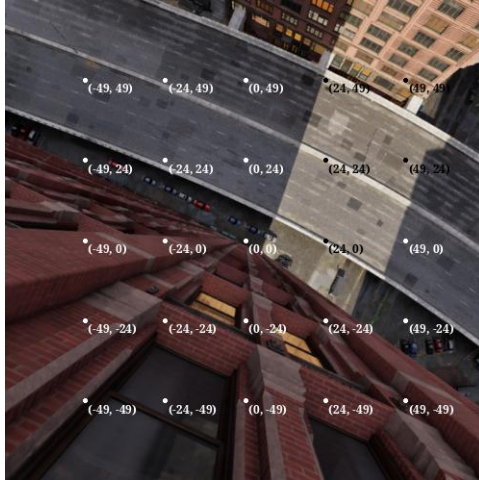


BENCHMARK : Your current altitude is 93 meters above ground level.

GPT-4O : <Reasoning>The black truck might be under the overpass or in the parking areas at the bottom. I'll move southwest toward the (-62, -31) position and decrease altitude to get a better view.</Reasoning> <Action>(-62, -31, -20)</Action>

BENCHMARK : Emergency stop; you've flown too close to something and would have hit it.

BENCHMARK *presents GPT-4O with a relevant image*

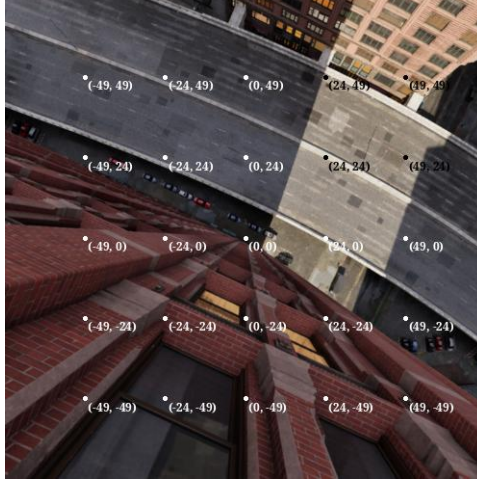


BENCHMARK : Your current altitude is 73 meters above ground level.

GPT-4O : <Reasoning>I'll proceed more carefully and move slightly east toward the (-24, 0) position to inspect the parking area, while maintaining a safe altitude adjustment.</Reasoning> <Action>(-24, 0, -10)</Action>

BENCHMARK : Emergency stop; you've flown too close to something and would have hit it.

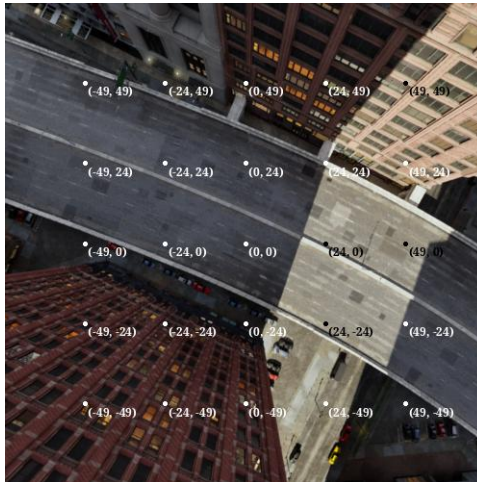
BENCHMARK *presents GPT-4O with a relevant image*



BENCHMARK : Your current altitude is 73 meters above ground level.

GPT-4O : <Reasoning>I'll adjust my approach and move east toward the (0, 24) position, keeping a higher altitude to avoid obstacles.</Reasoning> <Action>(0, 24, 0)</Action>

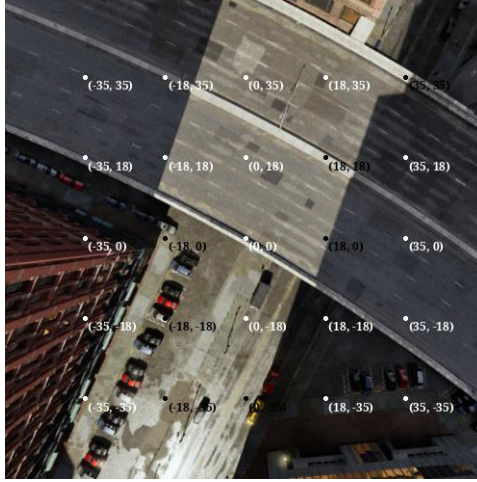
BENCHMARK *presents GPT-4O with a relevant image*



BENCHMARK : Your current altitude is 73 meters above ground level.

GPT-4O : <Reasoning>The black truck might be in the parking area below. I'll move southeast toward the (24, -24) position and decrease the altitude for a closer look.</Reasoning> <Action>(24, -24, -20)</Action>

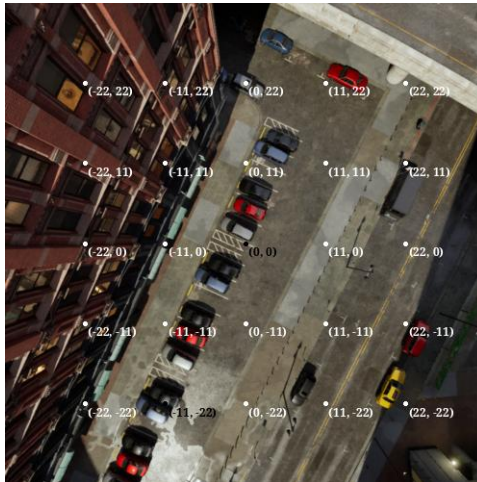
BENCHMARK *presents GPT-4O with a relevant image*



BENCHMARK : Your current altitude is 53 meters above ground level.

GPT-4O : <Reasoning>The black truck might be in the parking area to the southwest. I'll move to the (-18, -18) position and decrease altitude to get a closer view.</Reasoning> <Action>(-18, -18, -20)</Action>

BENCHMARK *presents GPT-4O with a relevant image*



BENCHMARK : Your current altitude is 33 meters above ground level.

GPT-4O : <Reasoning>A black truck is visible at the (0, 22) position. I'll move there and decrease altitude to ensure I'm within 10 meters above it.</Reasoning> <Action>(0, 22, -25)</Action>

BENCHMARK *presents GPT-4O with a relevant image*

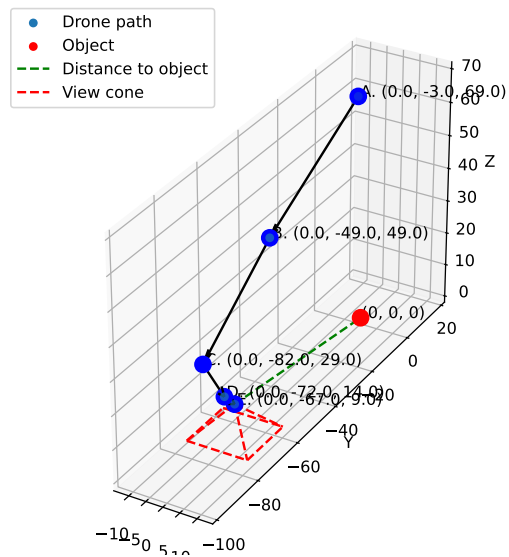


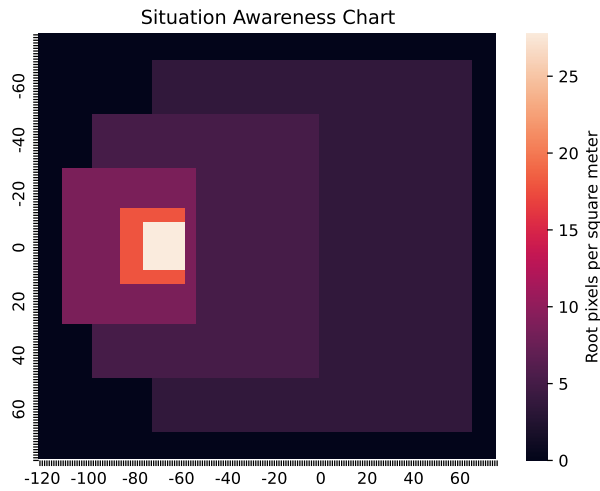
BENCHMARK : Your current altitude is 8 meters above ground level.

GPT-4o : <Reasoning>I am now centered above the black truck and within 10 meters of altitude.
The truck is clearly visible.</Reasoning> <Decision>FOUND</Decision>

1.2.2 Example

- Starting position: (0.0, -3.0, 69.0)
- End position: (0.0, -67.0, 9.0)
- Euclidean distance from the object: 67.60177512462228
- Object can be seen: False
- Success: False
- Object type: a fire





BENCHMARK : <Context> You are in command of a UAV, tasked with finding a fire. </Context>
(REST OF THE PROMPT)

BENCHMARK *presents GPT-4O with a relevant image*



BENCHMARK : Your current altitude is 69 meters above ground level.

GPT-4O : <Reasoning>There is no visible fire in the current image. I will move north to get a better view of more rooftops. I will also lower the altitude to get a closer look.</Reasoning>
<Action>(0, 46, -20)</Action>

BENCHMARK *presents GPT-4O with a relevant image*



BENCHMARK : Your current altitude is 49 meters above ground level.

GPT-4O : <Reasoning>Still no visible fire in the current image. I will move further north and decrease altitude to continue the search.</Reasoning> <Action>(0, 33, -20)</Action>

BENCHMARK *presents GPT-4O with a relevant image*



BENCHMARK : Your current altitude is 29 meters above ground level.

GPT-4O : <Reasoning>I can see smoke coming from the street area. I will move slightly south and lower the altitude to get closer to the possible source of the fire.</Reasoning> <Action>(0, -10, -15)</Action>

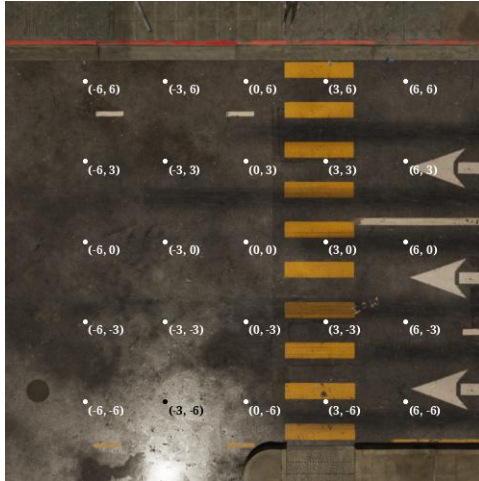
BENCHMARK *presents GPT-4O with a relevant image*



BENCHMARK : Your current altitude is 14 meters above ground level.

GPT-4o : <Reasoning>The fire is now clearly visible in the bottom center of the image. I will lower the altitude slightly to be within 10 meters above the fire.</Reasoning> <Action>(0, -5, -5)</Action>

BENCHMARK *presents GPT-4o with a relevant image*



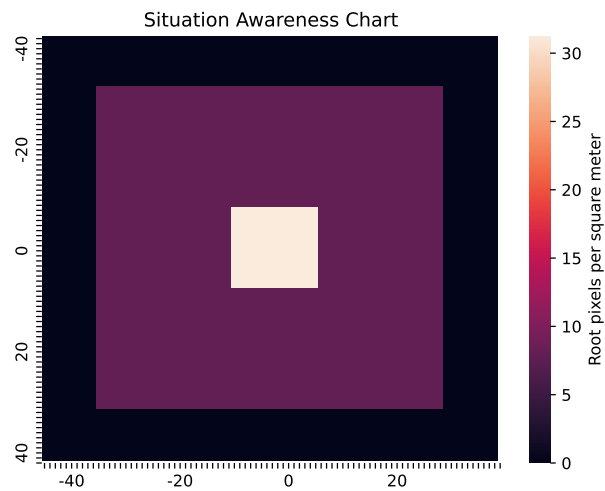
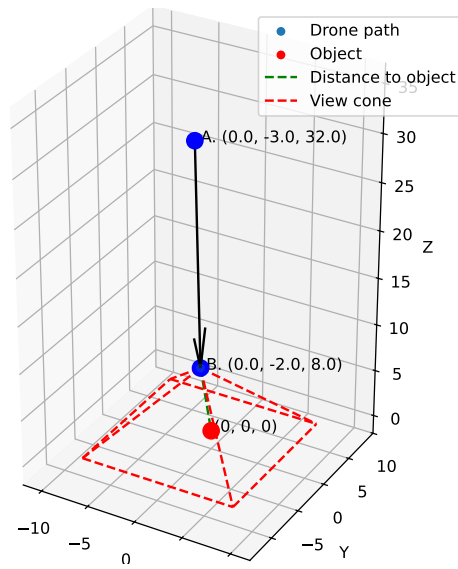
BENCHMARK : Your current altitude is 9 meters above ground level.

GPT-4o : FOUND

1.3 GPT-4o in the Forest; Success; FS-1

1.3.1 Example

- Starting position: (0.0, -4.0, 32.0)
- End position: (0.0, -2.0, 8.0)
- Euclidean distance from the object: 8.246211251235321
- Object can be seen: True
- Success: True
- Object type: a campsite



BENCHMARK : <Context> You are in command of a UAV, tasked with finding a campsite.
</Context> (REST OF THE PROMPT)

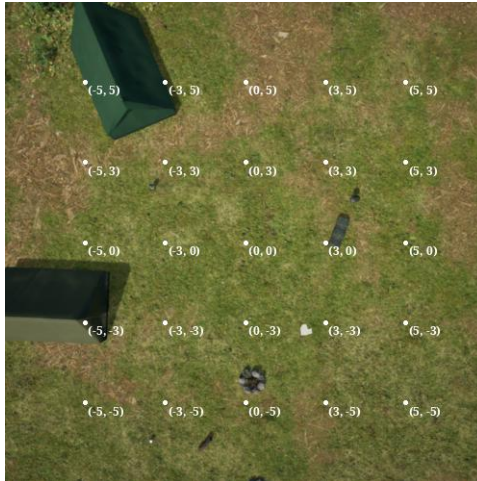
BENCHMARK presents GPT-4O with a relevant image



BENCHMARK : Your current altitude is 32 meters above ground level.

GPT-4O : <Reasoning> The two green tents in the center of the grid at (0, 0) resemble a campsite. I will descend directly to ensure I am below 10 meters above them. </Reasoning>
<Action>(0, 0, -24)</Action>

BENCHMARK *presents GPT-4O with a relevant image*

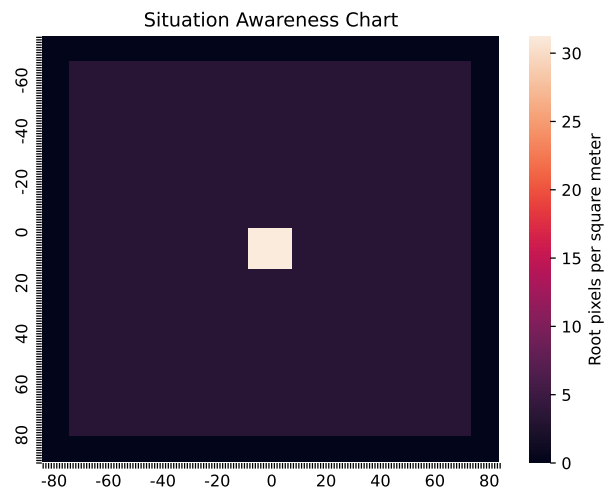
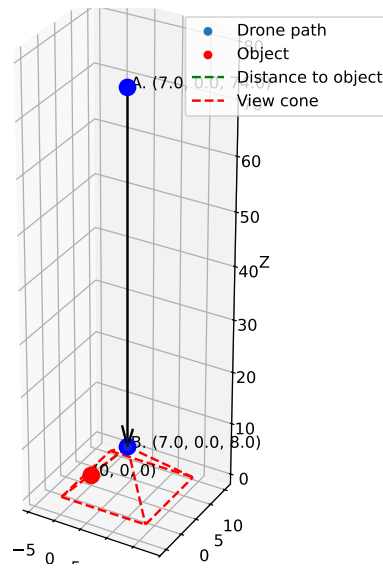


BENCHMARK : Your current altitude is 8 meters above ground level.

GPT-4O : <Reasoning> The two green tents at (0, 0) are clearly visible and centered. I am at 8 meters above them, which is below the required 10 meters. This is the campsite. </Reasoning> <FOUND />

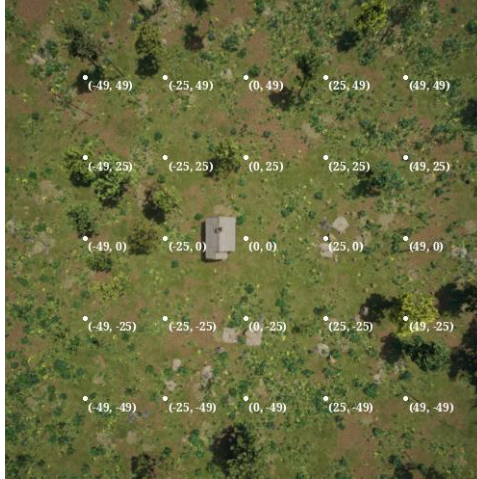
1.3.2 Example

- Starting position: (7.0, -1.0, 74.0)
- End position: (7.0, 0.0, 8.0)
- Euclidean distance from the object: 10.63014581273465
- Object can be seen: True
- Success: True
- Object type: a building



BENCHMARK : <Context> You are in command of a UAV, tasked with finding a building.
</Context> (REST OF THE PROMPT)

BENCHMARK *presents* GPT-4O *with a relevant image*



BENCHMARK : Your current altitude is 74 meters above ground level.

GPT-4O : <Reasoning>The building is clearly visible at the center of the grid. I will lower the altitude to get within the required 10 meters above the building to confirm the objective.</Reasoning> <Action>(0, 0, -65)</Action>

BENCHMARK *presents GPT-4O with a relevant image*



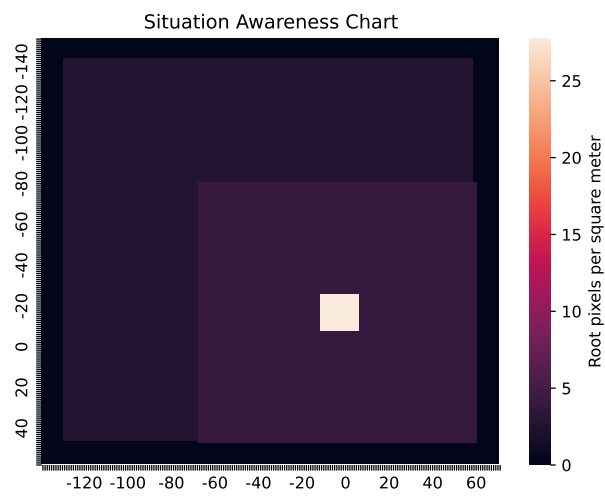
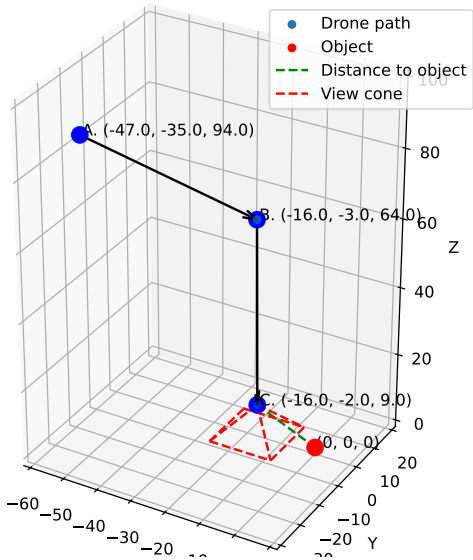
BENCHMARK : Your current altitude is 8 meters above ground level.

GPT-4O : <Reasoning>The building is now in view and I am within 10 meters above it. I will confirm the objective.</Reasoning> <Action>FOUND</Action>

1.4 GPT-4o in the Forest; Failure; FS-1

1.4.1 Example

- Starting position: (-47.0, -36.0, 95.0)
- End position: (-16.0, -2.0, 9.0)
- Euclidean distance from the object: 18.466185312619388
- Object can be seen: False
- Success: False
- Object type: a forest fire



BENCHMARK : <Context> You are in command of a UAV, tasked with finding a forest fire.
 </Context> (REST OF THE PROMPT)

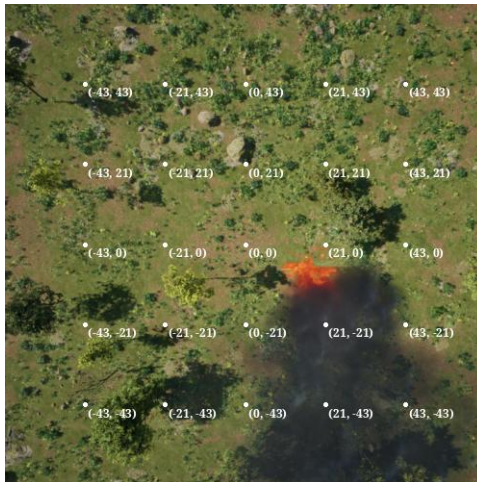
BENCHMARK *presents* GPT-4O *with a relevant image*



BENCHMARK : Your current altitude is 94 meters above ground level.

GPT-4O : <Reasoning>The orange-red area at (31, -31) suggests the presence of a fire. I should move towards it and decrease altitude to confirm.</Reasoning> <Action>(31, -31, -30)</Action>

BENCHMARK presents GPT-4O with a relevant image



BENCHMARK : Your current altitude is 64 meters above ground level.

GPT-4O : <Reasoning>The fire is more centered at (0, 0) now. I need to reduce my altitude further to be within 10 meters of it.</Reasoning> <Action>(0, 0, -55)</Action>

BENCHMARK presents GPT-4O with a relevant image

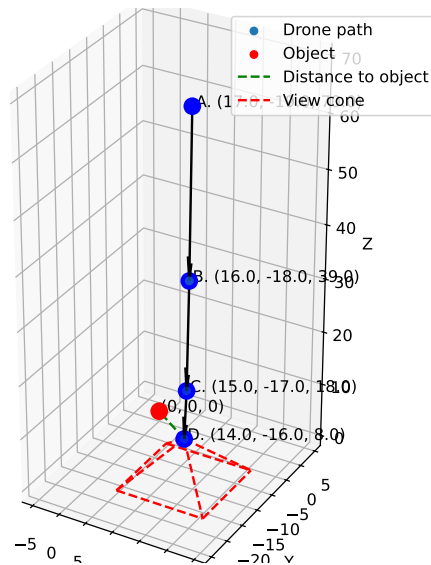


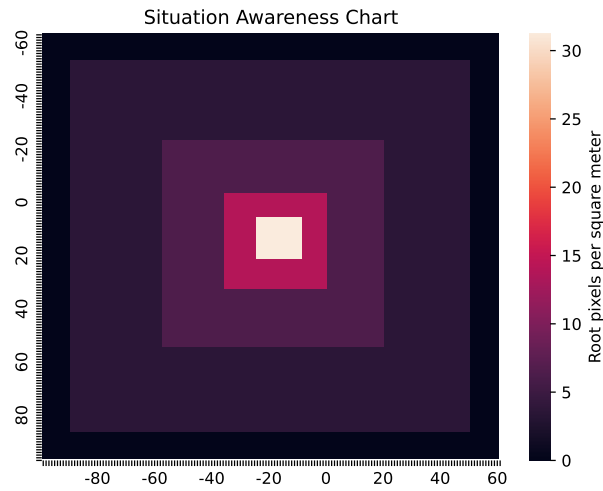
BENCHMARK : Your current altitude is 9 meters above ground level.

GPT-4o : <Reasoning>The fire is directly centered in my view. I am now 9 meters above it, which is within the acceptable altitude.</Reasoning> <Action>FOUND</Action>

1.4.2 Example

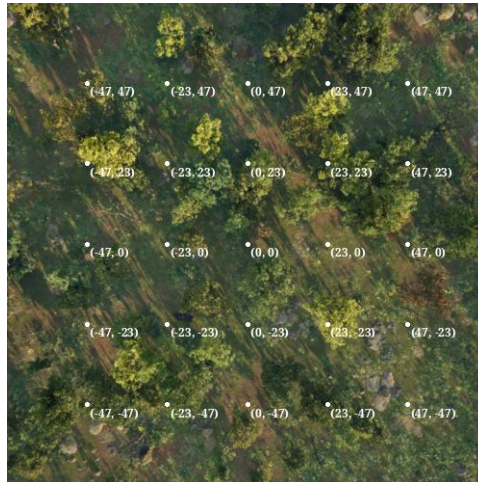
- Starting position: (18.0, -20.0, 71.0)
- End position: (14.0, -16.0, 8.0)
- Euclidean distance from the object: 22.715633383201094
- Object can be seen: False
- Success: False
- Object type: a trash pile





BENCHMARK : <Context> You are in command of a UAV, tasked with finding a trash pile.
</Context> (REST OF THE PROMPT)

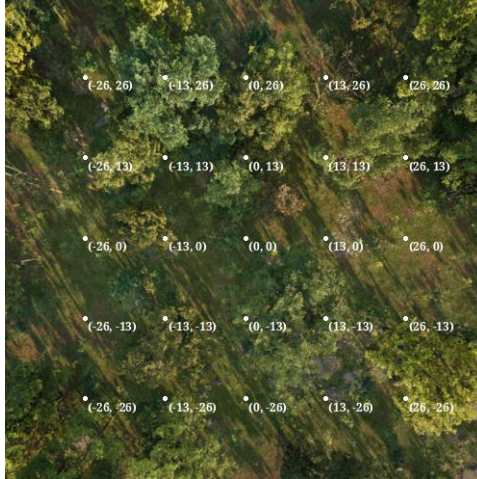
BENCHMARK *presents GPT-4O with a relevant image*



BENCHMARK : Your current altitude is 70 meters above ground level.

GPT-4O : <Reasoning>The area directly in the center (0, 0) appears clear of trees and might potentially contain the trash pile. I will move lower to investigate this area more closely.</Reasoning> <Action>(0, 0, -30)</Action>

BENCHMARK *presents GPT-4O with a relevant image*



BENCHMARK : Your current altitude is 39 meters above ground level.

GPT-4O : <Reasoning>The center area (0, 0) still seems like a viable spot for the trash pile. I will descend further to ensure a closer inspection of the area.</Reasoning> <Action>(0, 0, -20)</Action>

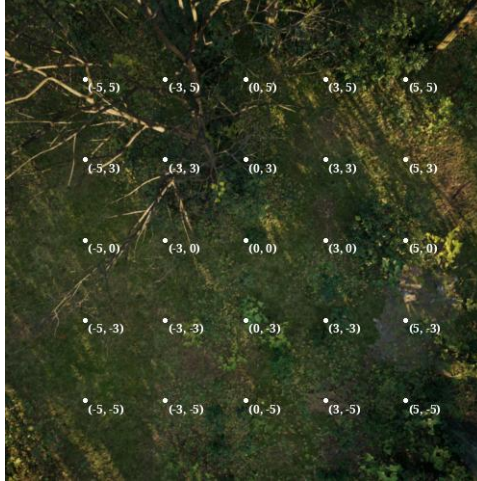
BENCHMARK *presents GPT-4O with a relevant image*



BENCHMARK : Your current altitude is 18 meters above ground level.

GPT-4O : <Reasoning>The area at the center (0, 0) is now more visible, but I need to descend further to ensure I'm within 10 meters of the ground to inspect this potential trash pile clearly.</Reasoning> <Action>(0, 0, -9)</Action>

BENCHMARK *presents GPT-4O with a relevant image*



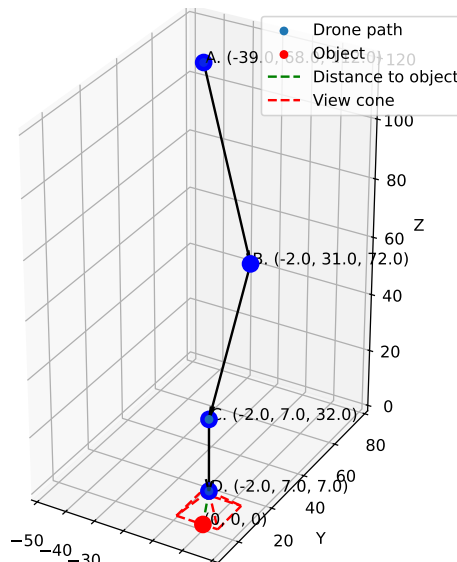
BENCHMARK : Your current altitude is 8 meters above ground level.

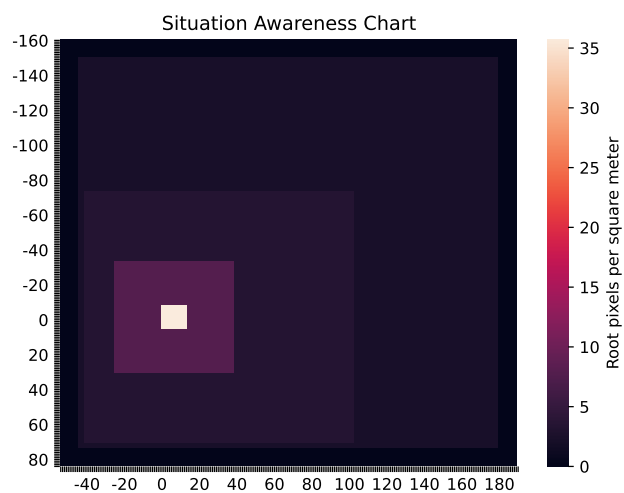
GPT-4o : <Reasoning>The object is now clearly visible and within 10 meters of the drone. It appears to be the trash pile. I will confirm its location.</Reasoning>
<Action>FOUND</Action>

1.5 GPT-4o in the City; Success; FS-2

1.5.1 Example

- Starting position: (-39.0, 68.0, 112.0)
- End position: (-2.0, 7.0, 7.0)
- Euclidean distance from the object: 10.099504938362077
- Object can be seen: True
- Success: True
- Object type: fire





BENCHMARK : <Context> You are in command of a UAV, tasked with finding fire. </Context>
(REST OF THE PROMPT)

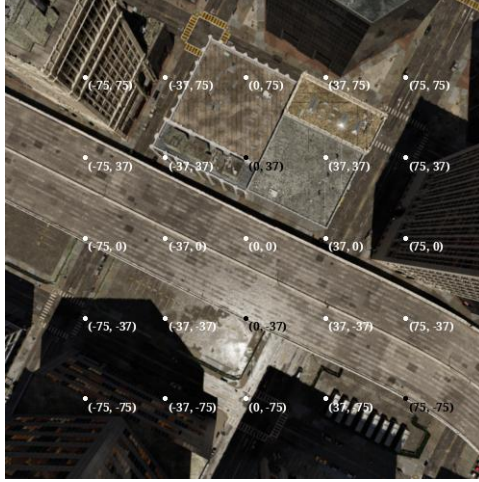
BENCHMARK : The object you're looking for is similar to this. This is NOT the drone's current view.

BENCHMARK *presents GPT-4O with a relevant image*



BENCHMARK : Your current altitude is 112 meters above ground level.

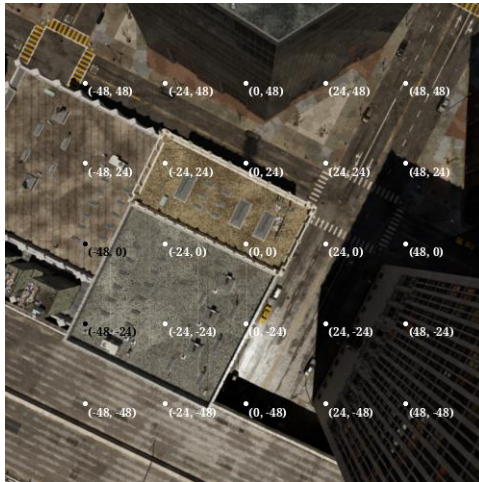
BENCHMARK *presents GPT-4O with a relevant image*



GPT-4O : <Reasoning>Based on the image, I will start by moving towards the open area slightly to the north and to the right to get a better view. I'll also descend to get closer to the ground.</Reasoning> <Action>(37, 37, -40)</Action>

BENCHMARK : Your current altitude is 72 meters above ground level.

BENCHMARK *presents GPT-4O with a relevant image*



GPT-4O : <Reasoning>I will continue moving towards the center of the open area, which might provide a better vantage point. I'll also descend further to improve visibility of potential fire sources.</Reasoning> <Action>(0, 24, -40)</Action>

BENCHMARK : Your current altitude is 32 meters above ground level.

BENCHMARK *presents GPT-4O with a relevant image*



GPT-4O : <Reasoning>There is black smoke and a damaged vehicle directly below. This matches the object I am searching for. I will descend further to ensure I'm within 10 meters of it.</Reasoning> <Action>(0, 0, -25)</Action>

BENCHMARK : Your current altitude is 7 meters above ground level.

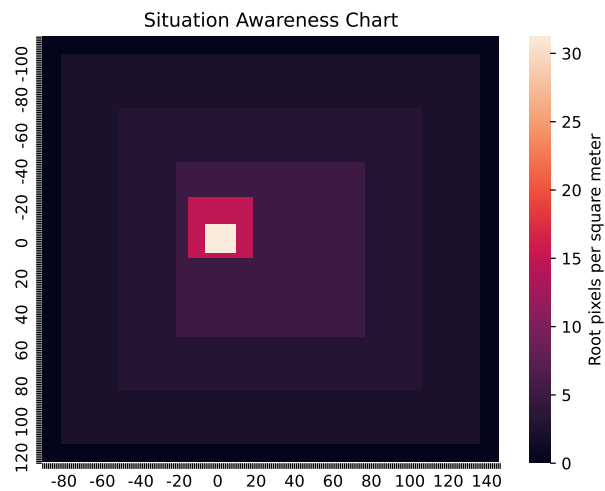
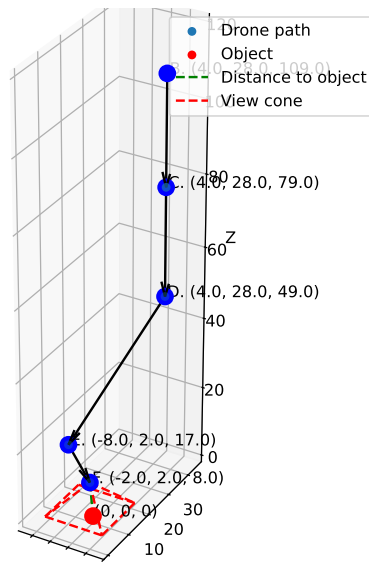
BENCHMARK *presents GPT-4O with a relevant image*



GPT-4O : <Reasoning>The object is now clearly visible and I am centered directly above it at an altitude of 7 meters, which is below the required 10 meters. I have found the object.</Reasoning> <Action>FOUND</Action>

1.5.2 Example

- Starting position: (4.0, 28.0, 109.0)
- End position: (-2.0, 2.0, 8.0)
- Euclidean distance from the object: 8.48528137423857
- Object can be seen: True
- Success: True
- Object type: crowd



BENCHMARK : <Context> You are in command of a UAV, tasked with finding crowd. </Context>
(REST OF THE PROMPT)

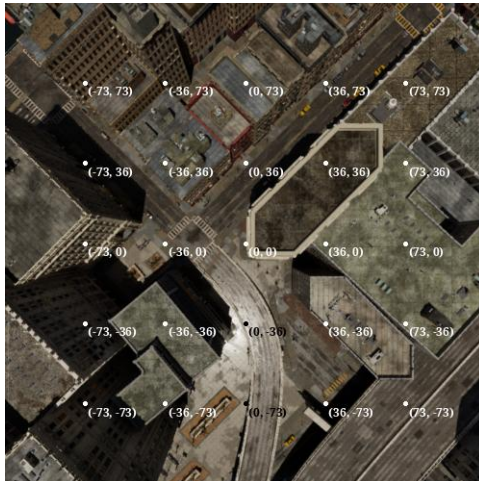
BENCHMARK : The object you're looking for is similar to this. This is NOT the drone's current view.

BENCHMARK presents GPT-4O with a relevant image



BENCHMARK : Your current altitude is 109 meters above ground level.

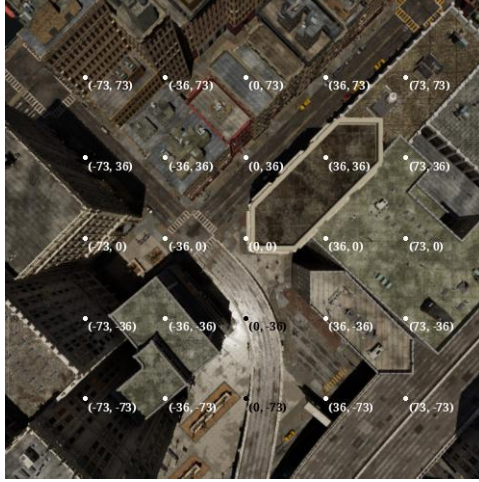
BENCHMARK *presents* GPT-4O with a relevant image



GPT-4O : <Reasoning>The objective is to find the crowd. I will start by moving towards the center of the map to get a better view. The position (0, 0) is a good starting point for broad visibility.</Reasoning> <Action>(0, 0, 0)</Action>

BENCHMARK : Your current altitude is 109 meters above ground level.

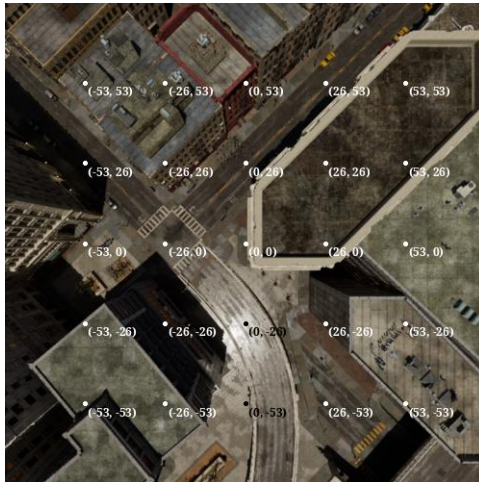
BENCHMARK *presents* GPT-4O with a relevant image



GPT-4O : <Reasoning>I'm currently at the center. I need to lower the altitude to get a clearer view and check for the crowd. I'll descend by 30 meters to improve visibility.</Reasoning>
<Action>(0, 0, -30)</Action>

BENCHMARK : Your current altitude is 79 meters above ground level.

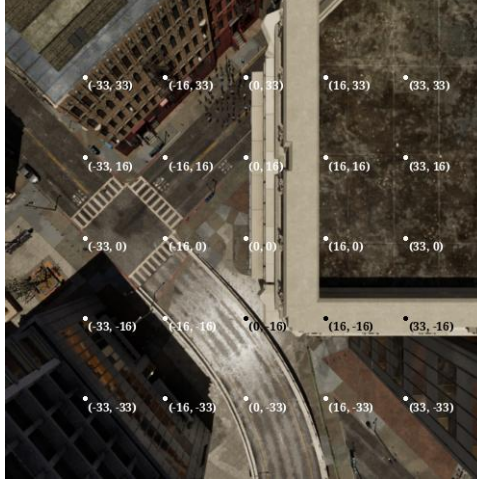
BENCHMARK *presents GPT-4O with a relevant image*



GPT-4O : <Reasoning>I'm reducing altitude further to enhance visibility and check for the crowd. Descending by another 30 meters should help in identifying the object more clearly.</Reasoning> <Action>(0, 0, -30)</Action>

BENCHMARK : Your current altitude is 49 meters above ground level.

BENCHMARK *presents GPT-4O with a relevant image*



GPT-4O : <Reasoning>I've spotted a group of people at (-16, 33). I will move towards that location and descend to get closer to confirm if it matches the objective. This will help center the drone on the crowd and ensure I'm below 10 meters above them.</Reasoning>
<Action>(-16, 33, -40)</Action>

BENCHMARK : Emergency stop; you've flown too close to something and would have hit it.

BENCHMARK : Your current altitude is 17 meters above ground level.

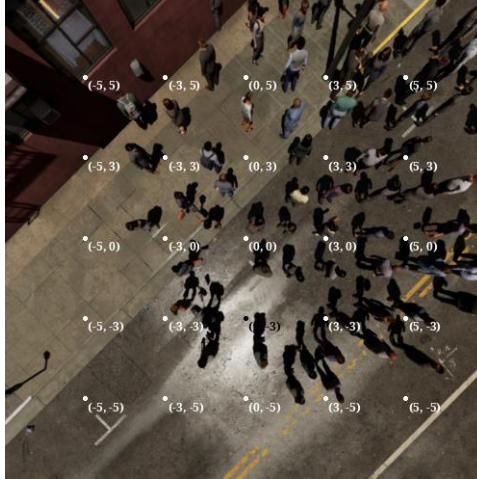
BENCHMARK *presents GPT-4O with a relevant image*



GPT-4O : <Reasoning>The crowd is visible and centered in the field of view at (6, 0). I will descend a bit more to ensure I'm below 10 meters above ground level, which is necessary to confirm the find.</Reasoning> <Action>(6, 0, -9)</Action>

BENCHMARK : Your current altitude is 8 meters above ground level.

BENCHMARK *presents GPT-4O with a relevant image*

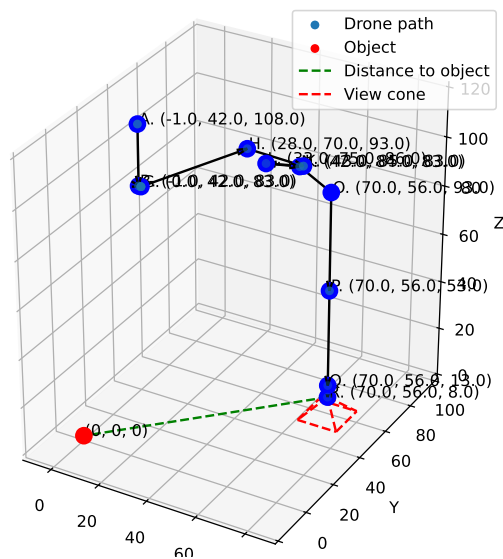


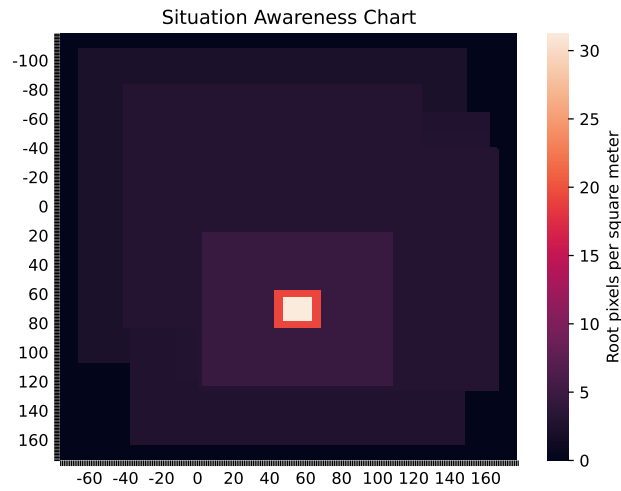
GPT-4o : <Reasoning>The crowd is centered in the view, and I am at an appropriate altitude of 8 meters, which is within the required distance. I can confirm the find.</Reasoning>
 <Action>FOUND</Action>

1.6 GPT-4o in the City; Failure; FS-2

1.6.1 Example

- Starting position: (-1.0, 42.0, 108.0)
- End position: (70.0, 56.0, 8.0)
- Euclidean distance from the object: 90.0
- Object can be seen: False
- Success: False
- Object type: crowd





BENCHMARK : <Context> You are in command of a UAV, tasked with finding crowd. </Context>
(REST OF THE PROMPT)

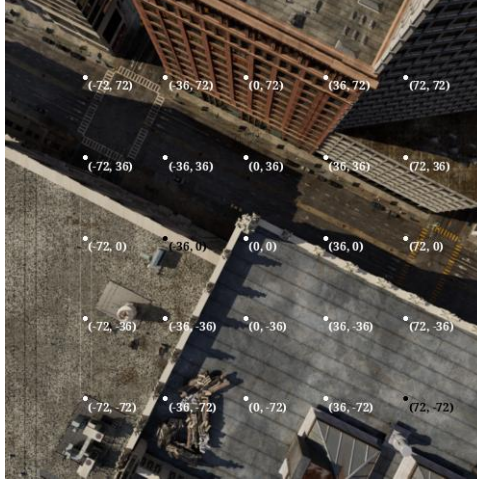
BENCHMARK : The object you're looking for is similar to this. This is NOT the drone's current view.

BENCHMARK *presents GPT-4O with a relevant image*



BENCHMARK : Your current altitude is 108 meters above ground level.

BENCHMARK *presents GPT-4O with a relevant image*

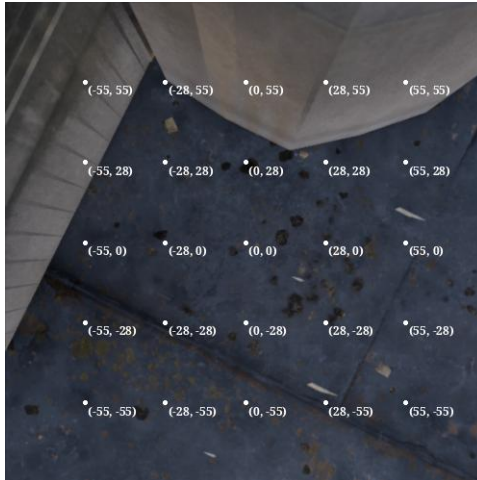


GPT-4O : <Reasoning>From this altitude, I need to lower the drone to get a clearer view and potentially spot the crowd. I will move downward while staying over the central area of the grid for a better observation point.</Reasoning> <Action>(0, 0, -40)</Action>

BENCHMARK : Emergency stop; you've flown too close to something and would have hit it.

BENCHMARK : Your current altitude is 83 meters above ground level.

BENCHMARK *presents GPT-4O with a relevant image*

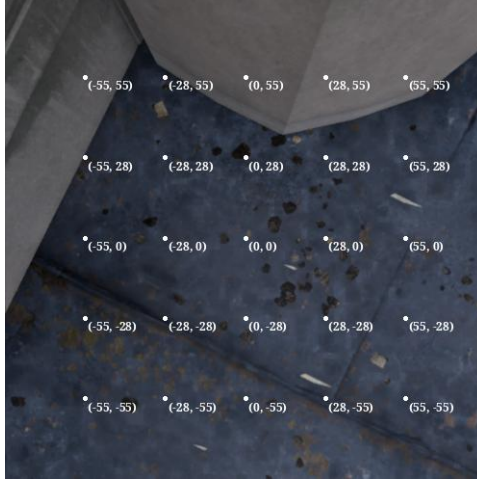


GPT-4O : <Reasoning>I'll move slightly to the open area to the southeast and lower the altitude to continue the search while avoiding obstacles.</Reasoning> <Action>(28, -28, -20)</Action>

BENCHMARK : Emergency stop; you've flown too close to something and would have hit it.

BENCHMARK : Your current altitude is 83 meters above ground level.

BENCHMARK *presents GPT-4O with a relevant image*



GPT-4O : <Reasoning>I'll adjust my position to the open area to the southwest and lower the altitude to enhance my view, avoiding the building close by.</Reasoning> <Action>(-28, -28, -20)</Action>

BENCHMARK : Emergency stop; you've flown too close to something and would have hit it.

BENCHMARK : Your current altitude is 83 meters above ground level.

BENCHMARK *presents GPT-4O with a relevant image*



GPT-4O : <Reasoning>I'll move to a clearer spot to the south and decrease altitude, maintaining a safe distance from obstacles.</Reasoning> <Action>(0, -28, -20)</Action>

BENCHMARK : Emergency stop; you've flown too close to something and would have hit it.

BENCHMARK : Your current altitude is 83 meters above ground level.

BENCHMARK *presents GPT-4O with a relevant image*

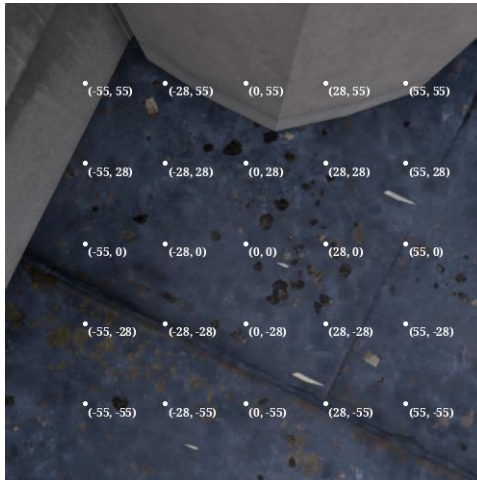


GPT-4O : <Reasoning>I'll move to the southeast this time to avoid obstacles and try to lower the altitude, aiming for an open area.</Reasoning> <Action>(28, -28, -10)</Action>

BENCHMARK : Emergency stop; you've flown too close to something and would have hit it.

BENCHMARK : Your current altitude is 83 meters above ground level.

BENCHMARK *presents GPT-4O with a relevant image*

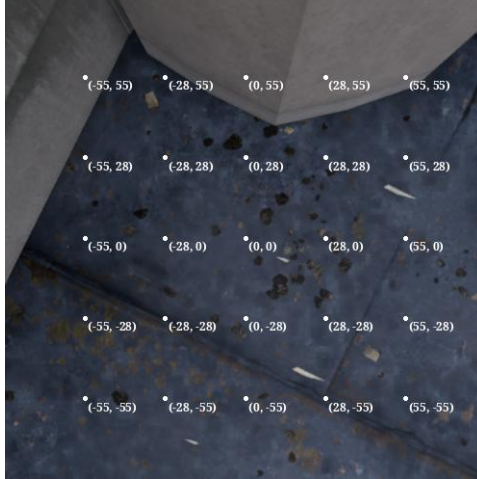


GPT-4O : <Reasoning>I'll move to the east, attempting to find a clear spot and lower the altitude carefully.</Reasoning> <Action>(28, 0, -10)</Action>

BENCHMARK : Emergency stop; you've flown too close to something and would have hit it.

BENCHMARK : Your current altitude is 83 meters above ground level.

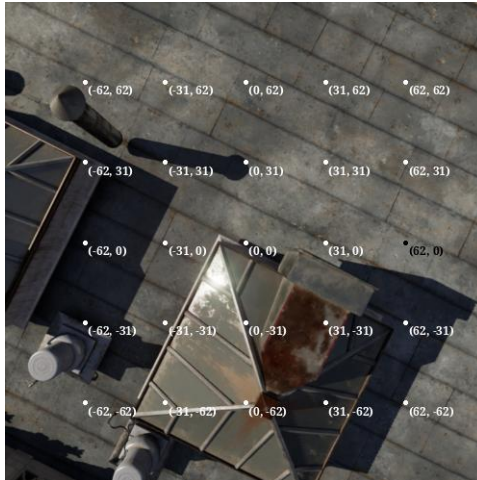
BENCHMARK *presents GPT-4O with a relevant image*



GPT-4O : <Reasoning>I'll ascend slightly to avoid obstacles and move to the southeast to find a clearer view.</Reasoning> <Action>(28, -28, 10)</Action>

BENCHMARK : Your current altitude is 93 meters above ground level.

BENCHMARK *presents GPT-4O with a relevant image*



GPT-4O : <Reasoning>I'll continue to move southeast and descend to get a better view over the open area.</Reasoning> <Action>(31, -31, -20)</Action>

BENCHMARK : Emergency stop; you've flown too close to something and would have hit it.

BENCHMARK : Your current altitude is 83 meters above ground level.

BENCHMARK *presents GPT-4O with a relevant image*

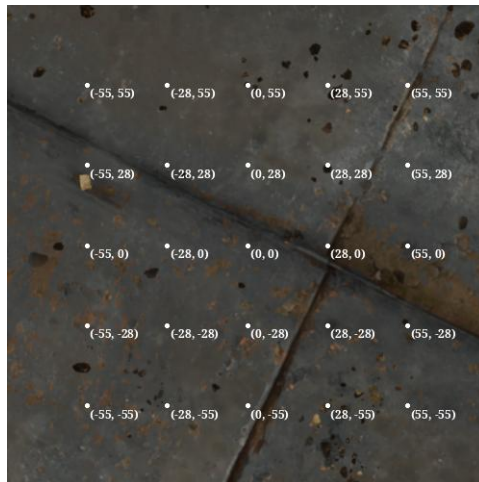


GPT-4O : <Reasoning>I'll move a bit to the east to find a clearer spot and then descend carefully to improve my view.</Reasoning> <Action>(28, 0, -20)</Action>

BENCHMARK : Emergency stop; you've flown too close to something and would have hit it.

BENCHMARK : Your current altitude is 83 meters above ground level.

BENCHMARK *presents GPT-4O with a relevant image*

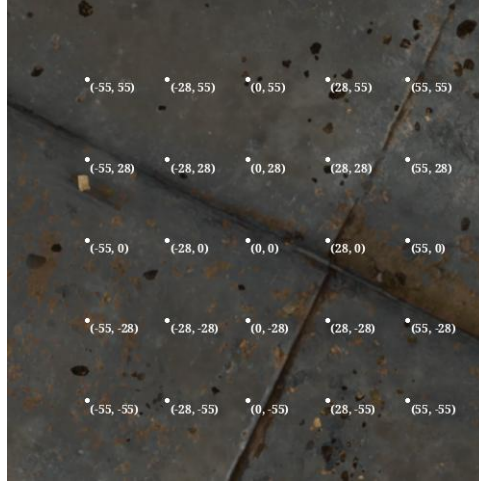


GPT-4O : <Reasoning>I'll move to the north to avoid obstacles and descend slightly to improve visibility.</Reasoning> <Action>(0, 28, -10)</Action>

BENCHMARK : Emergency stop; you've flown too close to something and would have hit it.

BENCHMARK : Your current altitude is 83 meters above ground level.

BENCHMARK *presents GPT-4O with a relevant image*



GPT-4O : <Reasoning>I'll ascend to gain a better view and move to the northwest to locate a clear area.</Reasoning> <Action>(-28, 28, 10)</Action>

BENCHMARK : Emergency stop; you've flown too close to something and would have hit it.

BENCHMARK : Your current altitude is 86 meters above ground level.

BENCHMARK *presents* GPT-4O *with a relevant image*



GPT-4O : <Reasoning>I'll move southeast to avoid structures and try descending slightly to improve visibility of the area below.</Reasoning> <Action>(29, -29, -10)</Action>

BENCHMARK : Emergency stop; you've flown too close to something and would have hit it.

BENCHMARK : Your current altitude is 83 meters above ground level.

BENCHMARK *presents* GPT-4O *with a relevant image*

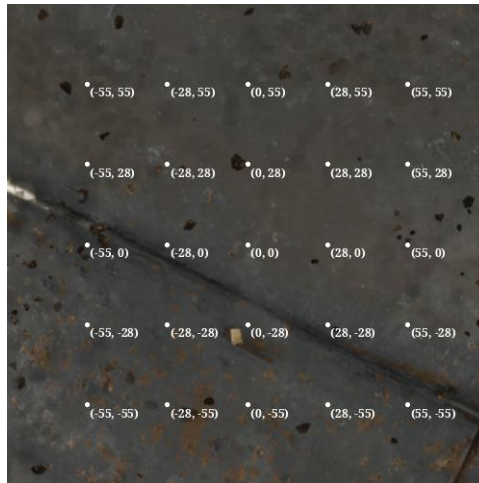


GPT-4O : <Reasoning>I'll move to the east to avoid obstacles and try to descend slightly to get a clearer view.</Reasoning> <Action>(28, 0, -10)</Action>

BENCHMARK : Emergency stop; you've flown too close to something and would have hit it.

BENCHMARK : Your current altitude is 83 meters above ground level.

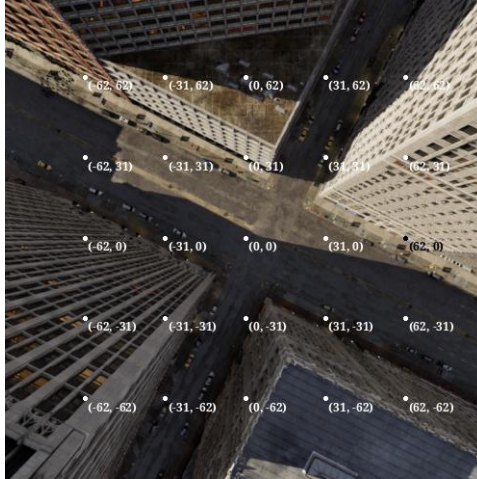
BENCHMARK *presents GPT-4O with a relevant image*



GPT-4O : <Reasoning>I'll ascend to avoid obstacles and move to the northeast to find a clearer area.</Reasoning> <Action>(28, 28, 10)</Action>

BENCHMARK : Your current altitude is 93 meters above ground level.

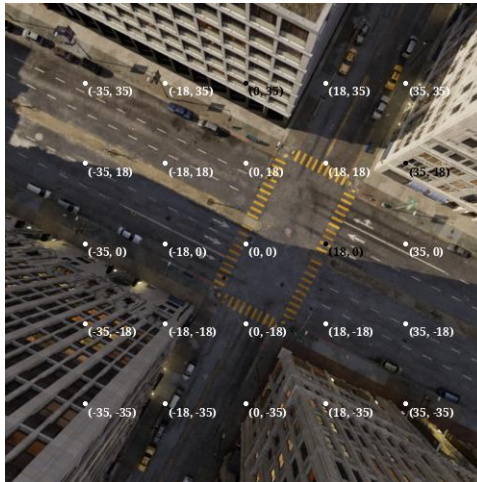
BENCHMARK *presents GPT-4O with a relevant image*



GPT-4O : <Reasoning>The open area to the center of the intersection might provide a good view of the crowd. I'll move directly there and descend to get closer.</Reasoning> <Action>(0, 0, -40)</Action>

BENCHMARK : Your current altitude is 53 meters above ground level.

BENCHMARK *presents GPT-4O with a relevant image*



GPT-4O : <Reasoning>I'll descend further and remain centered over the intersection to look for the crowd. This position should provide a clear view of the area.</Reasoning> <Action>(0, 0, -40)</Action>

BENCHMARK : Your current altitude is 13 meters above ground level.

BENCHMARK *presents GPT-4O with a relevant image*



GPT-4o : <Reasoning>I'm now close to the ground and centered over the intersection. I'll descend further to confirm the presence of the crowd.</Reasoning> <Action>(0, 0, -5)</Action>

BENCHMARK : Your current altitude is 8 meters above ground level.

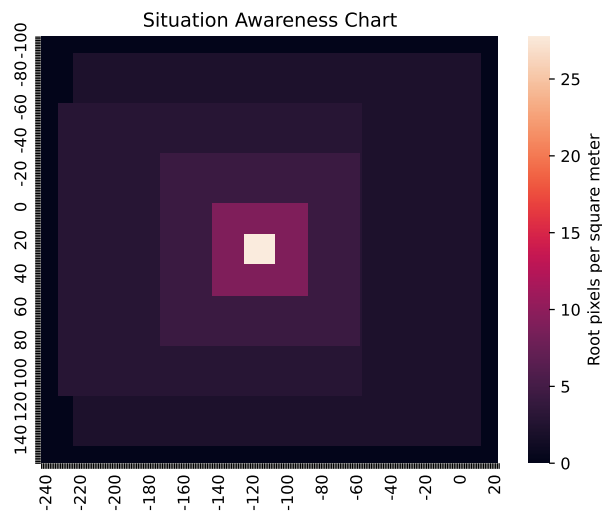
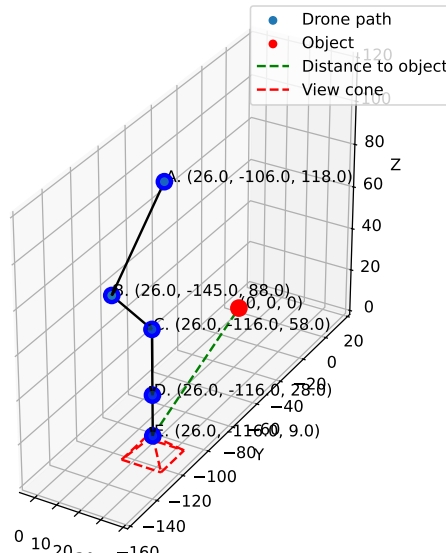
BENCHMARK *presents GPT-4o with a relevant image*



GPT-4o : <Reasoning>I am now at an appropriate altitude and centered on the intersection, which matches the objective. I should confirm the presence of the crowd.</Reasoning> <Action>FOUND</Action>

1.6.2 Example

- Starting position: (26.0, -106.0, 118.0)
- End position: (26.0, -116.0, 9.0)
- Euclidean distance from the object: 119.21828718783037
- Object can be seen: False
- Success: False
- Object type: road construction site



BENCHMARK : <Context> You are in command of a UAV, tasked with finding road construction site. </Context> (REST OF THE PROMPT)

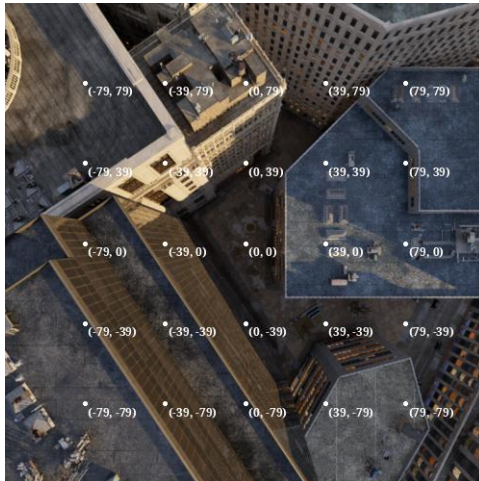
BENCHMARK : The object you're looking for is similar to this. This is NOT the drone's current view.

BENCHMARK presents GPT-4O with a relevant image



BENCHMARK : Your current altitude is 118 meters above ground level.

BENCHMARK *presents* GPT-4O *with a relevant image*



GPT-4O : <Reasoning>The search area includes several grid points. I will start by moving to the nearest point that allows a better view of the area directly ahead. I'll descend to get closer to the potential object.</Reasoning> <Action>(0, 39, -30)</Action>

BENCHMARK : Your current altitude is 88 meters above ground level.

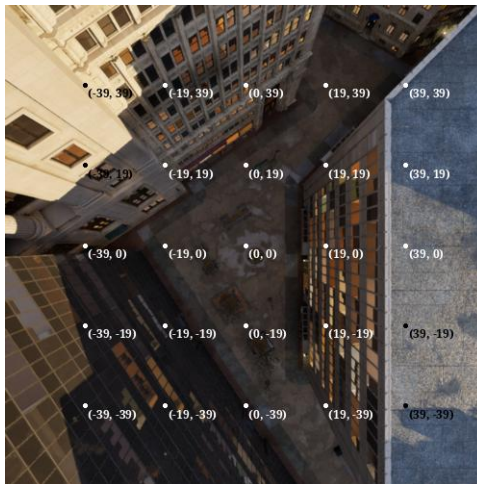
BENCHMARK *presents* GPT-4O *with a relevant image*



GPT-4O : <Reasoning>I'll continue moving south to explore the lower area of the grid for better visibility, while also descending to get closer to the ground.</Reasoning> <Action>(0, -29, -30)</Action>

BENCHMARK : Your current altitude is 58 meters above ground level.

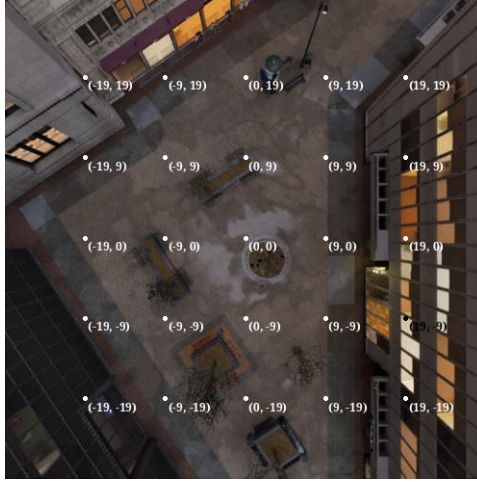
BENCHMARK *presents GPT-4O with a relevant image*



GPT-4O : <Reasoning>The object might be in the open area below. I will move directly over the center of this area while continuing to descend to get a clearer view.</Reasoning> <Action>(0, 0, -30)</Action>

BENCHMARK : Your current altitude is 28 meters above ground level.

BENCHMARK *presents GPT-4O with a relevant image*



GPT-4O : <Reasoning>The construction site is likely near the center of the open area. I'll continue to descend directly above it to ensure I'm within 10 meters.</Reasoning> <Action>(0, 0, -19)</Action>

BENCHMARK : Your current altitude is 9 meters above ground level.

BENCHMARK *presents* GPT-4O with a relevant image

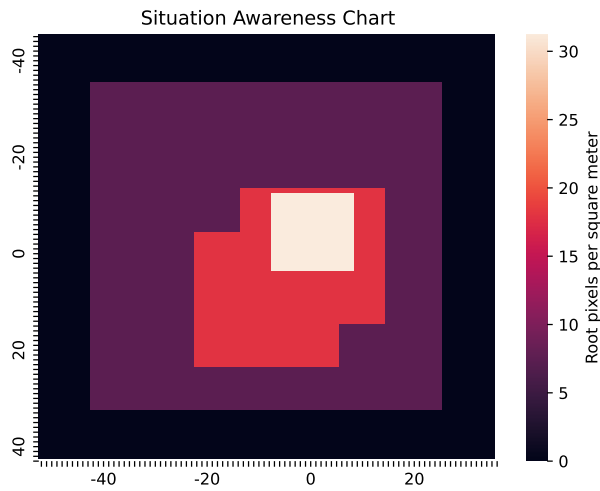
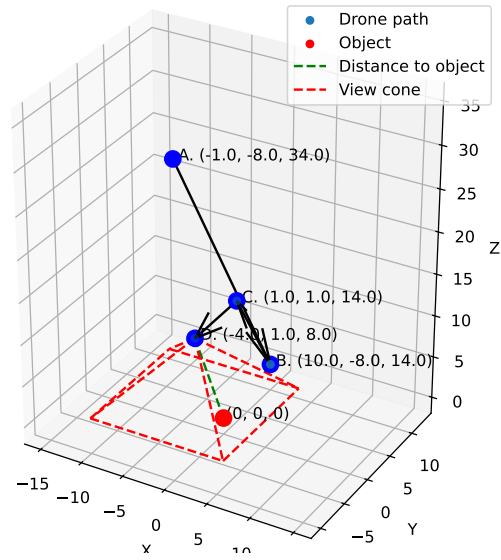


GPT-4O : <Reasoning>I am now 9 meters above the object in the center of the image, which resembles the construction site. I am within 10 meters of altitude and centered over it.</Reasoning> <Answer>FOUND</Answer>

1.7 Claude 3.5 Sonnet in the City; Success; FS-1

1.7.1 Example

- Starting position: (-1.0, -8.0, 34.0)
- End position: (-4.0, 1.0, 8.0)
- Euclidean distance from the object: 9.0
- Object can be seen: True
- Success: True
- Object type: a large trash pile



BENCHMARK : <Context> You are in command of a UAV, tasked with finding a large trash pile.
</Context> (REST OF THE PROMPT)

BENCHMARK : Image 1:

BENCHMARK presents CLAUDE 3.5 SONNET with a relevant image



BENCHMARK : Your current altitude is 34 meters above ground level.

CLAUDE 3.5 SONNET : <Reasoning>I need to systematically search the area for a large trash pile. From my current altitude of 34 meters, I can see a good portion of the area, but I need to get lower to better identify objects. I'll start by moving towards the center-right area (11, 0) and descending to get a better view, as this area seems to have some potential objects that need closer inspection.</Reasoning> <Action>(11, 0, -20)</Action>

BENCHMARK : Image 2:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*



BENCHMARK : Your current altitude is 14 meters above ground level.

CLAUDE 3.5 SONNET : <Reasoning>I can see what appears to be some debris or potential trash in the lower left corner of the image around coordinates (-9, -9). I'll move towards that location while maintaining my current low altitude to get a better look at what appears to be the target.</Reasoning> <Action>(-9, -9, 0)</Action>

BENCHMARK : Image 3:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*



BENCHMARK : Your current altitude is 14 meters above ground level.

CLAUDE 3.5 SONNET : <Reasoning>I can now clearly see a pile of debris/trash in my field of view, but I need to get a bit lower to ensure I'm within 10 meters of it as required by the objective. The pile appears to be centered around (-5, 0). I'll move slightly to better center on it and descend to about 8 meters altitude to ensure I'm meeting the requirement of being below 10 meters.</Reasoning> <Action>(-5, 0, -6)</Action>

BENCHMARK : Image 4:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*



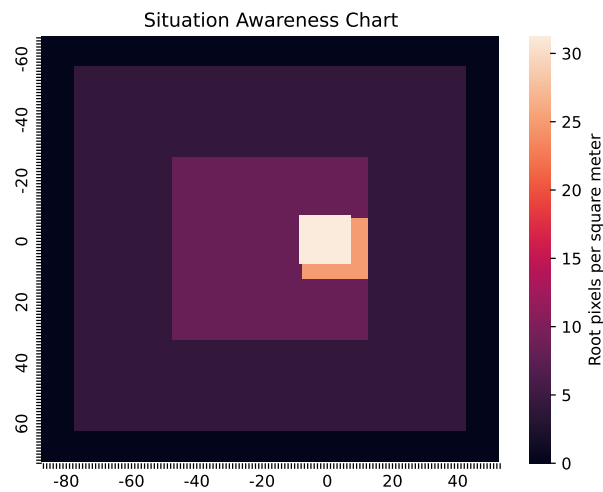
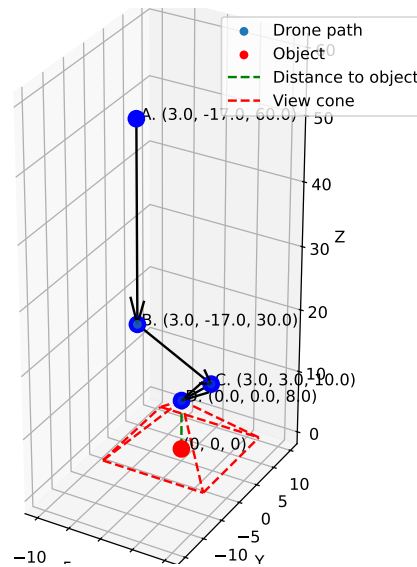
BENCHMARK : Your current altitude is 8 meters above ground level.

CLAUDE 3.5 SONNET : <Reasoning>I can now clearly see the trash pile in my field of view, and I am positioned at 8 meters altitude, which is below the required 10 meters. The object is centered in my view, and I am at an appropriate height to confirm this is indeed the target.</Reasoning> FOUND

1.7.2 Example

- Starting position: (3.0, -17.0, 60.0)
- End position: (0.0, 0.0, 8.0)

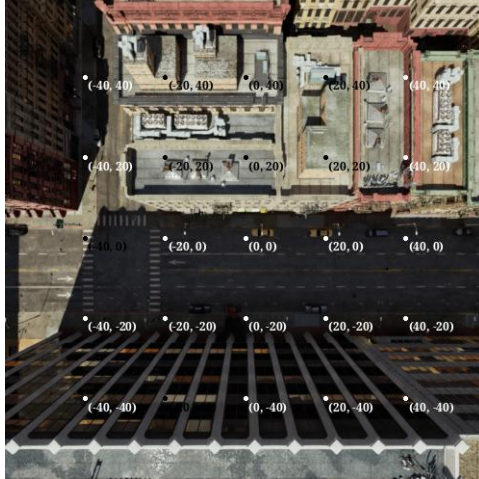
- Euclidean distance from the object: 8.0
- Object can be seen: True
- Success: True
- Object type: a fire



BENCHMARK : <Context> You are in command of a UAV, tasked with finding a fire. </Context>
(REST OF THE PROMPT)

BENCHMARK : Image 1:

BENCHMARK presents CLAUDE 3.5 SONNET with a relevant image

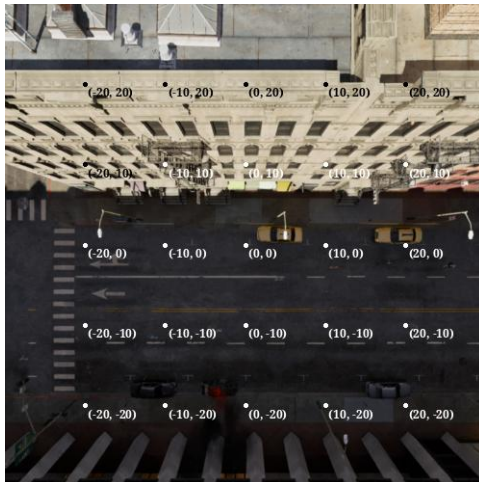


BENCHMARK : Your current altitude is 60 meters above ground level.

CLAUDE 3.5 SONNET : <Reasoning> I can see this is an urban environment viewed from above. Since I'm looking for a fire and need to search carefully, I should start by scanning the area systematically. I notice some open areas in the center of the image. I'll move to a central position and lower my altitude to get a better view of potential fire signs. </Reasoning>
<Action>(0, 0, -30)</Action>

BENCHMARK : Image 2:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*

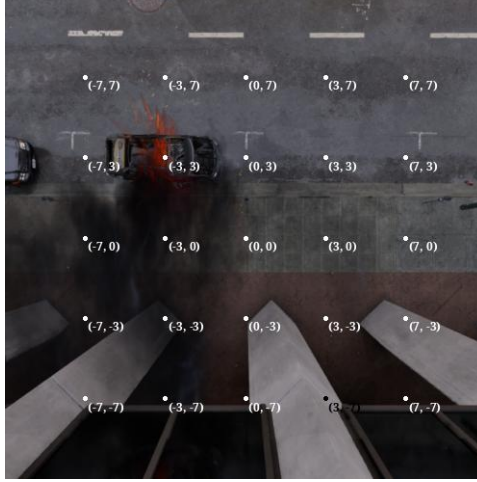


BENCHMARK : Your current altitude is 30 meters above ground level.

CLAUDE 3.5 SONNET : <Reasoning> From this lower altitude, I can see better details of the street. I notice what appears to be a small red glow or potential fire near the coordinates (0, -20). I should move closer to that location and lower my altitude further to confirm if this is the fire we're looking for. </Reasoning> <Action>(0, -20, -20)</Action>

BENCHMARK : Image 3:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*



BENCHMARK : Your current altitude is 10 meters above ground level.

CLAUDE 3.5 SONNET : <Reasoning> I can now clearly see a car on fire near coordinates (-3, 3). I need to position myself directly above it at a height less than 10 meters to meet the objective requirements. I'll move to those coordinates and lower my altitude slightly to ensure I'm below 10 meters. </Reasoning> <Action>(-3, 3, -2)</Action>

BENCHMARK : Image 4:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*



BENCHMARK : Your current altitude is 8 meters above ground level.

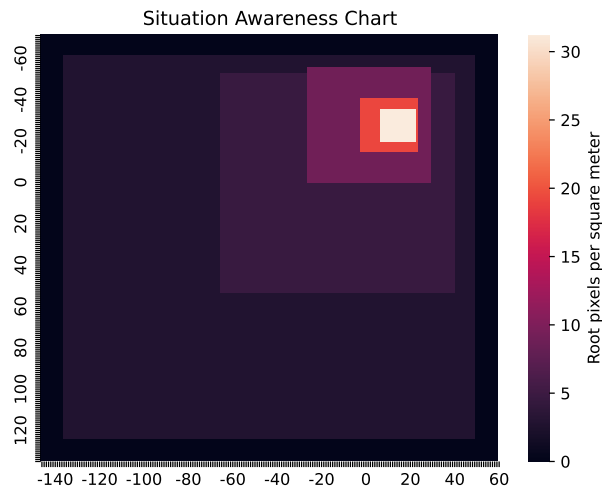
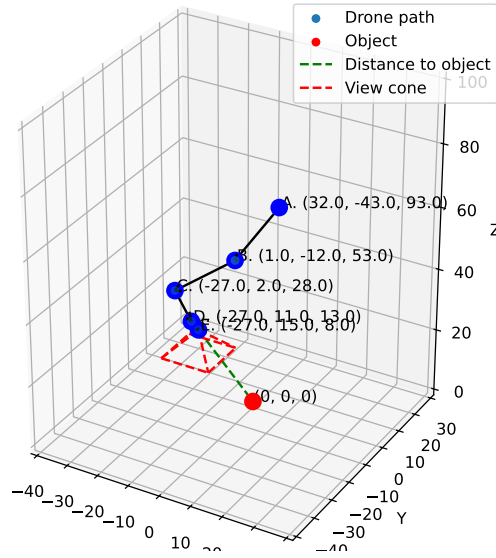
CLAUDE 3.5 SONNET : <Reasoning> I am now positioned at 8 meters above the burning car, which is clearly visible in my field of view at the center of the image. This meets all the requirements: the fire is in view, I am centered above it, and my altitude is below 10 meters (specifically at 8 meters). </Reasoning> FOUND

1.8 Claude 3.5 Sonnet in the City; Failure; FS-1

1.8.1 Example

- Starting position: (32.0, -43.0, 93.0)

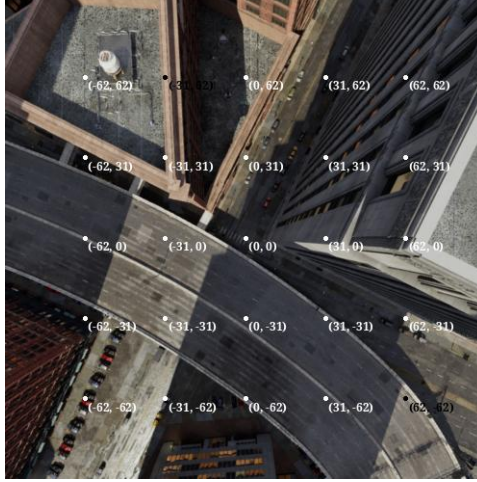
- End position: (-27.0, 15.0, 8.0)
- Euclidean distance from the object: 31.906112267087632
- Object can be seen: False
- Success: False
- Object type: a black truck



BENCHMARK : <Context> You are in command of a UAV, tasked with finding a black truck.
</Context> (REST OF THE PROMPT)

BENCHMARK : Image 1:

BENCHMARK presents CLAUDE 3.5 SONNET with a relevant image

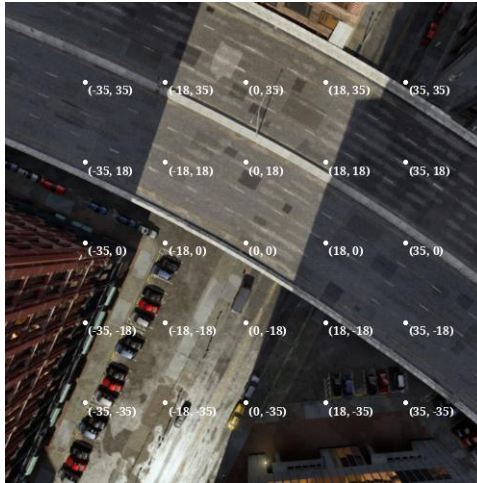


BENCHMARK : Your current altitude is 93 meters above ground level.

CLAUDE 3.5 SONNET : <Reasoning>I can see several vehicles in the image but from this height and position, I need to get lower and explore different areas to find the specific black truck. I'll start by moving to a position that gives me a better view of the area with parked vehicles while reducing altitude significantly.</Reasoning> <Action>(-31, -31, -40)</Action>

BENCHMARK : Image 2:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*



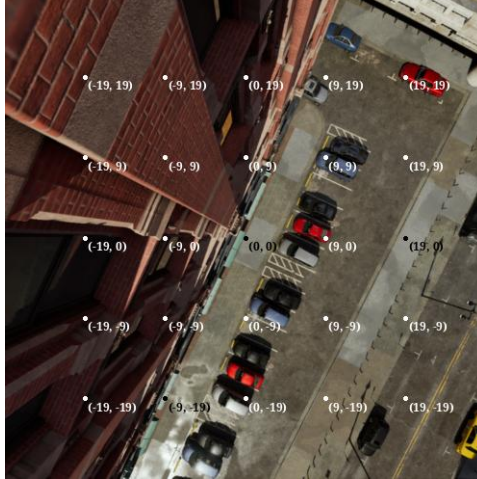
BENCHMARK : Your current altitude is 53 meters above ground level.

CLAUDE 3.5 SONNET : <Reasoning>I can see a parking area with several vehicles on the left side of the image. I'll move closer to that area and continue descending to get a better look at the vehicles there, as one of them could be the black truck we're looking for.</Reasoning> <Action>(-35, -18, -30)</Action>

BENCHMARK : Emergency stop; you've flown too close to something and would have hit it.

BENCHMARK : Image 3:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*



BENCHMARK : Your current altitude is 28 meters above ground level.

CLAUDE 3.5 SONNET : <Reasoning>I can now see several parked vehicles including what appears to be a black vehicle in the parking area. I need to move closer to the center of the parking area and continue descending to verify if it's the black truck we're looking for. I'll move more carefully this time to avoid any obstacles.</Reasoning> <Action>(0, -9, -15)</Action>

BENCHMARK : Image 4:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*



BENCHMARK : Your current altitude is 13 meters above ground level.

CLAUDE 3.5 SONNET : <Reasoning>I can now clearly see a black vehicle in the parking area and I'm at a good height to verify it. I'll move slightly closer to center directly over it and descend just a bit more to get within the required 10 meters while keeping it in view.</Reasoning> <Action>(0, -4, -5)</Action>

BENCHMARK : Image 5:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*

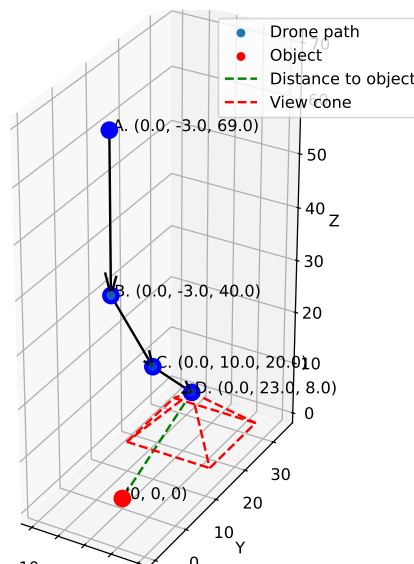


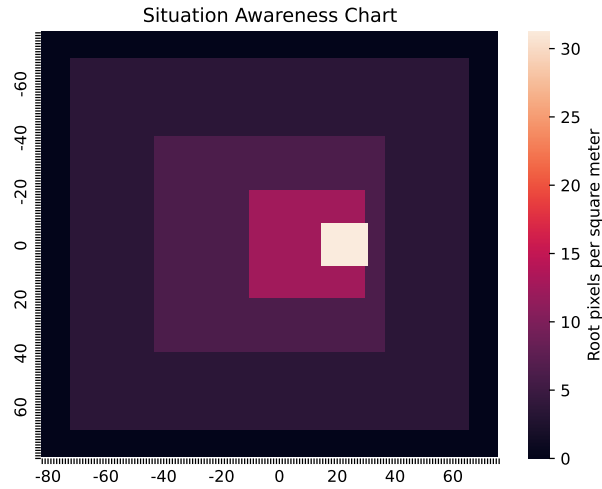
BENCHMARK : Your current altitude is 8 meters above ground level.

CLAUDE 3.5 SONNET : <Reasoning>I now have a clear view of a black vehicle that appears to be a truck, and I am positioned at 8 meters above it, which is within our required range of being below 10 meters. The vehicle is centered in my field of view.</Reasoning> FOUND

1.8.2 Example

- Starting position: (0.0, -3.0, 69.0)
- End position: (0.0, 23.0, 8.0)
- Euclidean distance from the object: 24.351591323771842
- Object can be seen: False
- Success: False
- Object type: a fire





BENCHMARK : <Context> You are in command of a UAV, tasked with finding a fire. </Context>
(REST OF THE PROMPT)

BENCHMARK : Image 1:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*

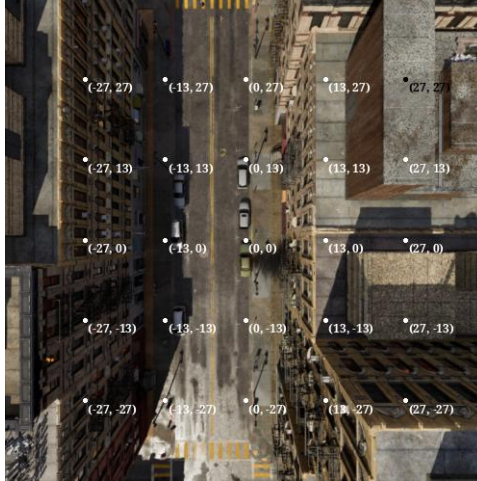


BENCHMARK : Your current altitude is 69 meters above ground level.

CLAUDE 3.5 SONNET : <Reasoning> From this aerial view of an urban area at 69 meters altitude, I need to conduct a systematic search for any signs of fire. I notice several buildings and streets, but no immediate signs of fire. I should first try to get a lower altitude to better observe any potential smoke or flames, while maintaining a central position to maximize my field of view. I'll descend to about 40 meters first while staying in the center position to get a better overview.</Reasoning> <Action>(0, 0, -29)</Action>

BENCHMARK : Image 2:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*

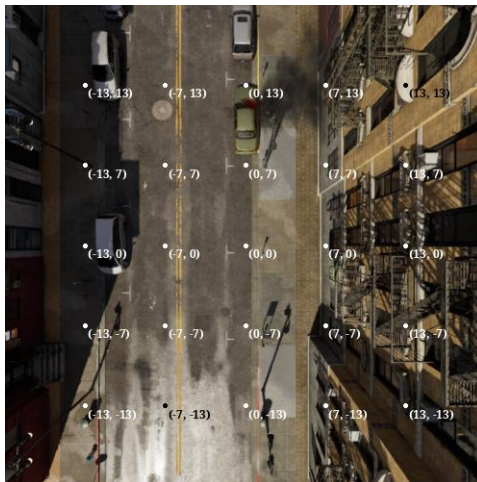


BENCHMARK : Your current altitude is 40 meters above ground level.

CLAUDE 3.5 SONNET : <Reasoning> At 40 meters altitude, I have a clearer view but still no obvious signs of fire. However, I notice what appears to be some activity or potential smoke in the lower part of the image, around coordinates (0, -13). I should investigate this area more closely by moving slightly south and descending further to get a better look.</Reasoning>
<Action>(0, -13, -20)</Action>

BENCHMARK : Image 3:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*



BENCHMARK : Your current altitude is 20 meters above ground level.

CLAUDE 3.5 SONNET : <Reasoning> I can now see what appears to be smoke or fire activity more clearly in the lower portion of the image, around coordinates (0, -13). I need to descend further to get within 10 meters of the target while staying centered on it. I'll maintain my horizontal position but decrease altitude to 8 meters to ensure I'm within the required range.</Reasoning> <Action>(0, -13, -12)</Action> FOUND

BENCHMARK : Image 4:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*



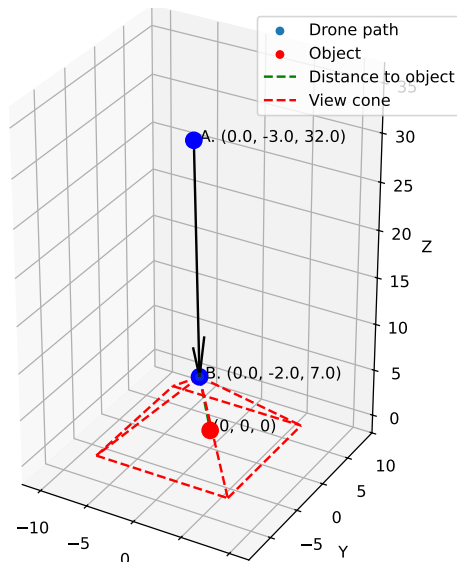
BENCHMARK : Your current altitude is 8 meters above ground level.

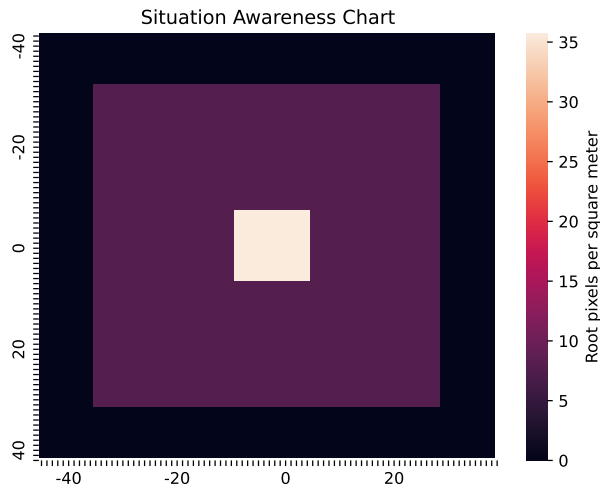
CLAUDE 3.5 SONNET : <Reasoning> I can now clearly see what appears to be smoke on the ground. I'm positioned correctly at 8 meters altitude, which is below the required 10 meters, and I'm centered on the target. The smoke/fire is clearly visible in my field of view, so this satisfies all the mission requirements.</Reasoning> FOUND

1.9 Claude 3.5 Sonnet in the Forest; Success; FS-1

1.9.1 Example

- Starting position: (0.0, -4.0, 32.0)
- End position: (0.0, -2.0, 7.0)
- Euclidean distance from the object: 7.280109889280518
- Object can be seen: True
- Success: True
- Object type: a campsite





BENCHMARK : <Context> You are in command of a UAV, tasked with finding a campsite.
</Context> (REST OF THE PROMPT)

BENCHMARK : Image 1:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*

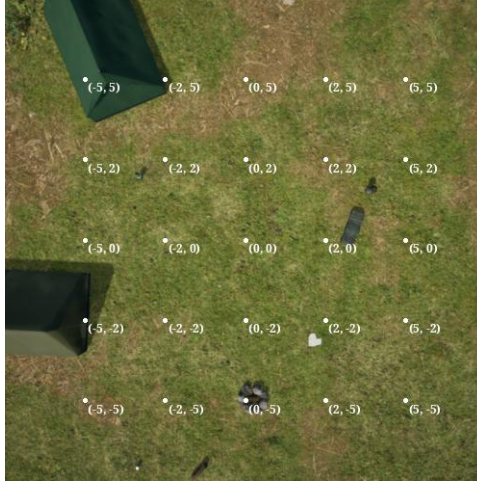


BENCHMARK : Your current altitude is 32 meters above ground level.

CLAUDE 3.5 SONNET : <Reasoning>I can see what appears to be a tent or small campsite near the center of the image. To get a better view and position myself correctly, I'll move to coordinates (0, 0) which appear to be directly above the campsite. Since I need to be below 10 meters to confirm the target, I'll also decrease my altitude by 25 meters, bringing me to approximately 7 meters above ground level.</Reasoning> <Action>(0, 0, -25)</Action>

BENCHMARK : Image 2:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*

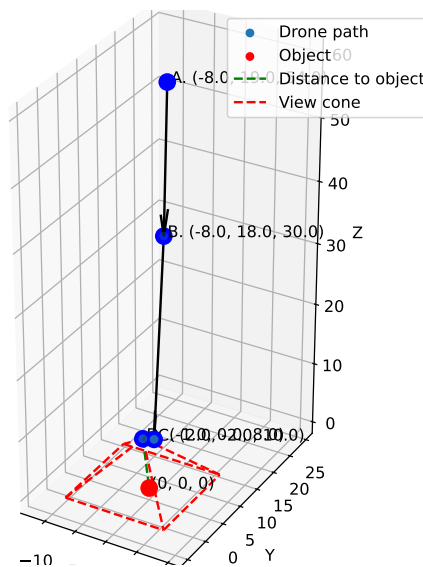


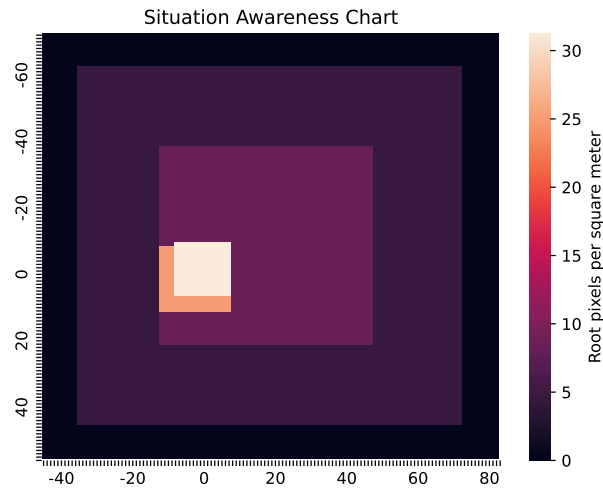
BENCHMARK : Your current altitude is 7 meters above ground level.

CLAUDE 3.5 SONNET : <Reasoning>I can now clearly see the campsite. I'm at the correct altitude (7 meters, which is below 10 meters) and I'm centered on the target area. The campsite is clearly visible in my field of view.</Reasoning> FOUND

1.9.2 Example

- Starting position: $(-8.0, 20.0, 54.0)$
- End position: $(-1.0, 0.0, 8.0)$
- Euclidean distance from the object: 8.06225774829855
- Object can be seen: True
- Success: True
- Object type: a trash pile





BENCHMARK : <Context> You are in command of a UAV, tasked with finding a trash pile.
</Context> (REST OF THE PROMPT)

BENCHMARK : Image 1:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*



BENCHMARK : Your current altitude is 54 meters above ground level.

CLAUDE 3.5 SONNET : <Reasoning> From this initial altitude of 54 meters, I can see a wide area but need to conduct a systematic search for the trash pile. I notice some variations in the terrain and need to get lower to better identify potential targets. I'll start by moving to the center of the visible area and descending to get a better view of any potential trash piles.
</Reasoning> <Action>(0, 0, -24)</Action>

BENCHMARK : Image 2:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*



BENCHMARK : Your current altitude is 30 meters above ground level.

CLAUDE 3.5 SONNET : <Reasoning> At 30 meters altitude, I can now see more detail but still need to search systematically. I notice what appears to be some irregular objects or potential debris near the coordinates (10, 20). I'll move in that direction and descend further to investigate more closely. </Reasoning> <Action>(10, 20, -20)</Action>

BENCHMARK : Image 3:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*



BENCHMARK : Your current altitude is 10 meters above ground level.

CLAUDE 3.5 SONNET : <Reasoning> I can now clearly see a trash pile consisting of various debris near coordinates (-3, -3). I need to position myself directly above it and ensure I'm at the correct altitude to meet the mission requirements. I'm currently at 10 meters which is just at the maximum allowable height, but I'll descend slightly to ensure I'm definitively under 10 meters while staying centered over the trash pile. </Reasoning> <Action>(-3, -3, -2)</Action>

BENCHMARK : Image 4:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*



BENCHMARK : Your current altitude is 8 meters above ground level.

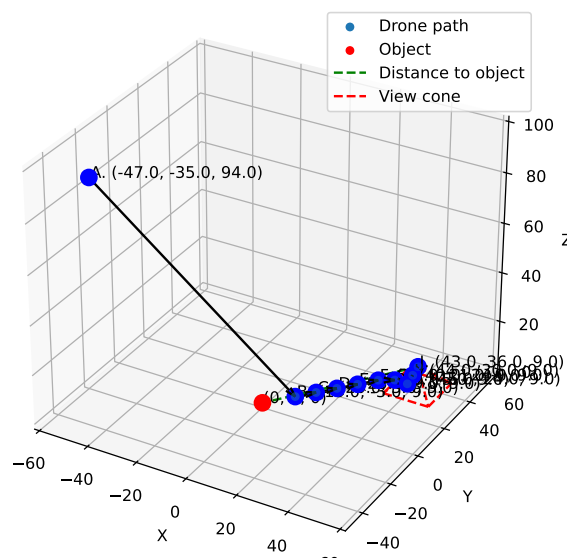
CLAUDE 3.5 SONNET : FOUND

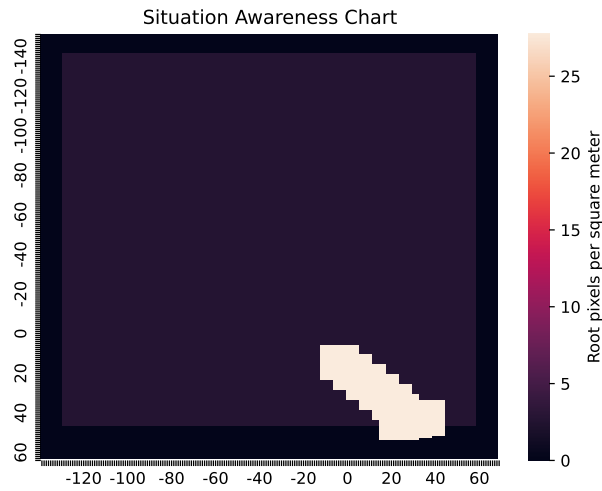
<Reasoning> I am now perfectly positioned above the trash pile at 8 meters altitude, which is below the required 10 meters. The trash pile is clearly visible in my field of view and I am centered above it. This meets all the mission requirements: 1. Below 10 meters (at 8 meters) 2. Object in field of view 3. Centered on the object </Reasoning>

1.10 Claude 3.5 Sonnet in the Forest; Failure; FS-1

1.10.1 Example

- Starting position: (-47.0, -36.0, 95.0)
- End position: (43.0, 36.0, 9.0)
- Euclidean distance from the object: 56.79788728465171
- Object can be seen: False
- Success: False
- Object type: a forest fire

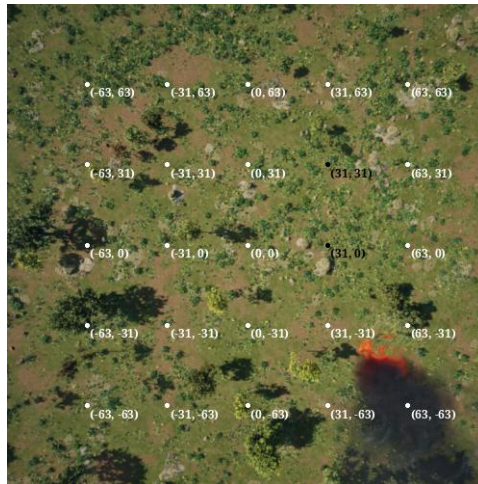




BENCHMARK : <Context> You are in command of a UAV, tasked with finding a forest fire.
</Context> (REST OF THE PROMPT)

BENCHMARK : Image 1:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*



BENCHMARK : Your current altitude is 94 meters above ground level.

CLAUDE 3.5 SONNET : <Reasoning>I can see what appears to be smoke and fire in the south-eastern portion of the image, around coordinates (63, -31). I should move towards that location and decrease altitude to get a better view and get within the required 10-meter height requirement.</Reasoning> <Action>(63, -31, -85)</Action>

BENCHMARK : Image 2:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*



BENCHMARK : Your current altitude is 9 meters above ground level.

CLAUDE 3.5 SONNET : <Reasoning>I'm now at a good altitude (9 meters) which is below the required 10 meters, but I don't see the fire in my current view. I need to move back towards the fire's location. Based on the previous image, I should move slightly southeast.</Reasoning>
<Action>(6, -6, 0)</Action>

BENCHMARK : Image 3:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*



BENCHMARK : Your current altitude is 9 meters above ground level.

CLAUDE 3.5 SONNET : <Reasoning>I still don't see the fire in my current view. From the first image, I know the fire was in the southeastern direction. I need to continue moving in that direction while maintaining my current good altitude of 9 meters.</Reasoning>
<Action>(6, -6, 0)</Action>

BENCHMARK : Image 4:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*



BENCHMARK : Your current altitude is 9 meters above ground level.

CLAUDE 3.5 SONNET : <Reasoning>I still need to continue moving towards the southeastern direction where I initially spotted the fire. I'm maintaining my good altitude of 9 meters, which is within the required range, but I need to keep moving until I have the fire in my field of view.</Reasoning> <Action>(6, -6, 0)</Action>

BENCHMARK : Image 5:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*



BENCHMARK : Your current altitude is 9 meters above ground level.

CLAUDE 3.5 SONNET : <Reasoning>I need to adjust my strategy. While I'm at the correct altitude of 9 meters, I've been moving but haven't located the fire yet. From the first image, the fire was in the southeastern quadrant. Let me make a larger movement to the east to try to locate it.</Reasoning> <Action>(63, -31, 0)</Action>

BENCHMARK : This command would endanger the drone, as you would fly out of bounds of the last seen image, possibly flying into unknown territories, recklessly. Please adjust your command so that you don't fly out of bounds of the last glimpse.

CLAUDE 3.5 SONNET : <Reasoning>I apologize for the unsafe command. You're right - I should only move within the visible grid coordinates to ensure safe navigation. Let me make a smaller, safer movement continuing towards the southeast direction while staying within the visible grid.</Reasoning> <Action>(6, -6, 0)</Action>

BENCHMARK : Image 6:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*



BENCHMARK : Your current altitude is 9 meters above ground level.

CLAUDE 3.5 SONNET : <Reasoning>I need to reconsider my search pattern. While I'm at the correct altitude of 9 meters, I seem to be moving in a direction that's not helping me locate the fire. Let me try moving more directly east since that's where we initially saw the fire in the first image.</Reasoning> <Action>(6, -3, 0)</Action>

BENCHMARK : Image 7:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*



BENCHMARK : Your current altitude is 9 meters above ground level.

CLAUDE 3.5 SONNET : <Reasoning>I still don't see the fire in my current view. From the first image, we know it was further east. Let me continue moving east while maintaining my current good altitude of 9 meters.</Reasoning> <Action>(6, 0, 0)</Action>

BENCHMARK : Image 8:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*



BENCHMARK : Your current altitude is 9 meters above ground level.

CLAUDE 3.5 SONNET : <Reasoning>I need to rethink my strategy. We only have one move left after this one, and I haven't located the fire yet. Looking back at the first image, the fire was in the southeastern quadrant, but I might have moved in the wrong direction. Let me try moving back towards where we started and then southeast.</Reasoning> <Action>(0, -6, 0)</Action>

BENCHMARK : Image 9:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*



BENCHMARK : Your current altitude is 9 meters above ground level.

CLAUDE 3.5 SONNET : <Reasoning>Unfortunately, I have not been successful in locating the fire. While I maintained the correct altitude of 9 meters throughout the search, my navigation strategy was not effective in reaching the fire location that we initially saw in the southeastern portion of the first image. I've used all 9 allowed moves without successfully centering on the fire.</Reasoning> <Action>(0, -6, 0)</Action>

BENCHMARK : Image 10:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*

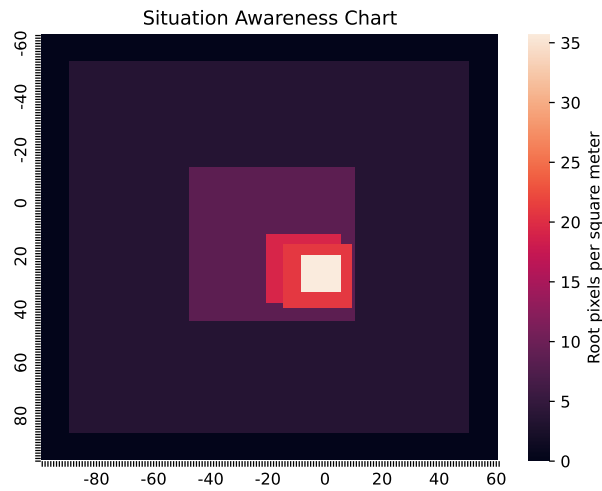
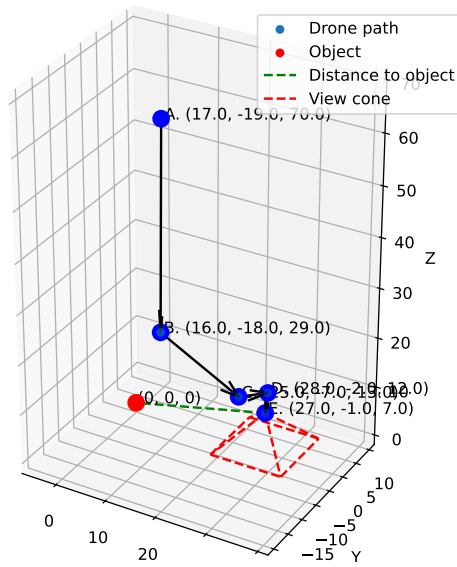


BENCHMARK : Your current altitude is 9 meters above ground level.

CLAUDE 3.5 SONNET : <Reasoning>I have now used all available moves and was not successful in my mission. While I maintained the correct altitude of 9 meters, which would have been appropriate for identifying the fire, my search pattern was ineffective. I should have made larger initial movements towards the southeastern direction where we first spotted the fire, while still staying within the safe visible boundaries. This would have been a better strategy than the smaller incremental movements I chose to make.</Reasoning>

1.10.2 Example

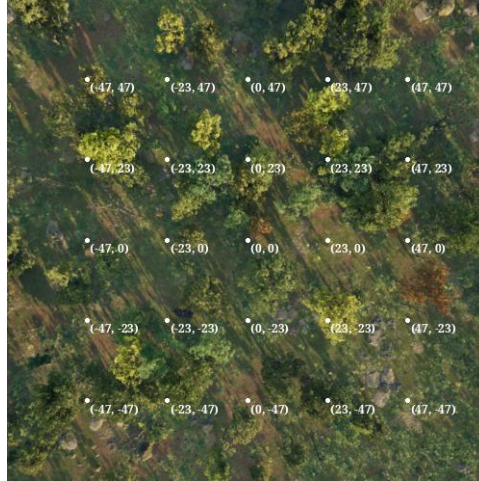
- Starting position: (18.0, -20.0, 71.0)
- End position: (27.0, -1.0, 7.0)
- Euclidean distance from the object: 27.910571473905726
- Object can be seen: False
- Success: False
- Object type: a trash pile



BENCHMARK : <Context> You are in command of a UAV, tasked with finding a trash pile.
</Context> (REST OF THE PROMPT)

BENCHMARK : Image 1:

BENCHMARK presents CLAUDE 3.5 SONNET with a relevant image



BENCHMARK : Your current altitude is 70 meters above ground level.

CLAUDE 3.5 SONNET : <Reasoning> I need to start searching for the trash pile systematically. From this altitude (70m), it's difficult to spot small details clearly. I notice some variations in the terrain, but I need to get lower to better identify potential targets. I'll move to a central position and decrease altitude significantly to get a better view of the area.

I'll move towards coordinates (0, 0) which is the center of the grid, and simultaneously decrease altitude by 40 meters to get a better view while maintaining enough height to continue searching if needed. </Reasoning> <Action>(0, 0, -40)</Action>

BENCHMARK : Image 2:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*

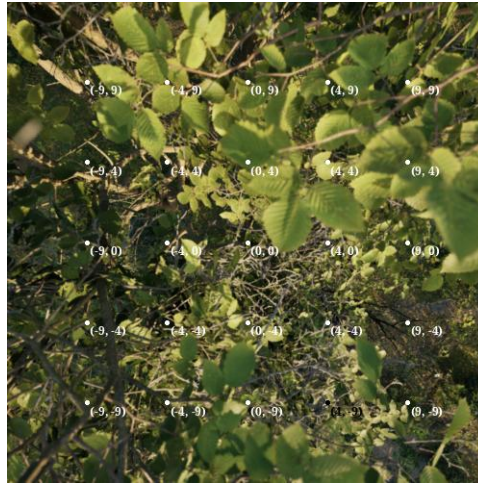


BENCHMARK : Your current altitude is 29 meters above ground level.

CLAUDE 3.5 SONNET : <Reasoning> From this lower altitude (29m), I can see more details but I still don't clearly spot the trash pile. I notice some variations in the terrain and coloring in the southeast quadrant of the visible area. I'll move towards that direction (10, -10) and lower the altitude a bit more to get an even better view of that area. </Reasoning> <Action>(10, -10, -15)</Action>

BENCHMARK : Image 3:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*

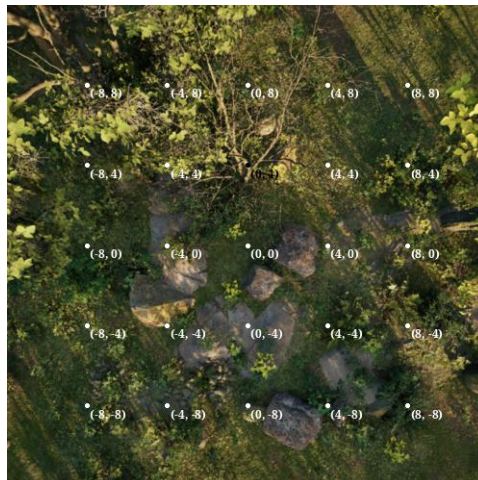


BENCHMARK : Your current altitude is 13 meters above ground level.

CLAUDE 3.5 SONNET : <Reasoning> At this altitude (13m), I'm quite close to the vegetation, but I still need to continue my search. I can see mostly leaves and foliage in this area, but no trash pile. I'll move slightly east and south where there seems to be a bit more open space between the vegetation, and maintain this altitude to get a good view of that area.
</Reasoning> <Action>(4, -4, 0)</Action>

BENCHMARK : Image 4:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*



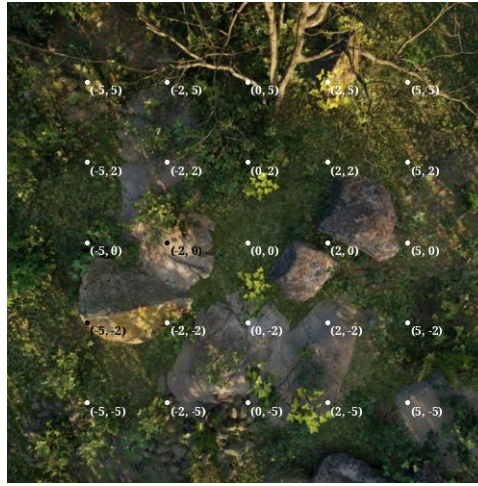
BENCHMARK : Your current altitude is 12 meters above ground level.

CLAUDE 3.5 SONNET : <Reasoning> I can now see what appears to be some debris or trash (dark objects) near the center of the image. Since I need to be below 10 meters and centered on the object, I'll move to position (0, 0) and descend a few more meters to get below the 10-meter

requirement while staying centered over what appears to be the trash pile. </Reasoning>
<Action>(0, 0, -4)</Action>

BENCHMARK : Image 5:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*



BENCHMARK : Your current altitude is 7 meters above ground level.

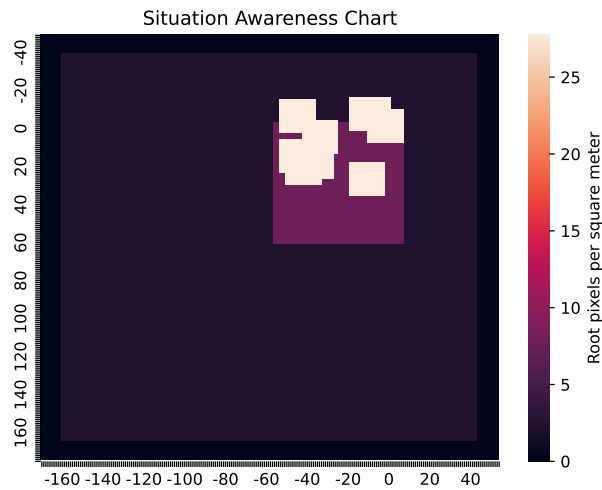
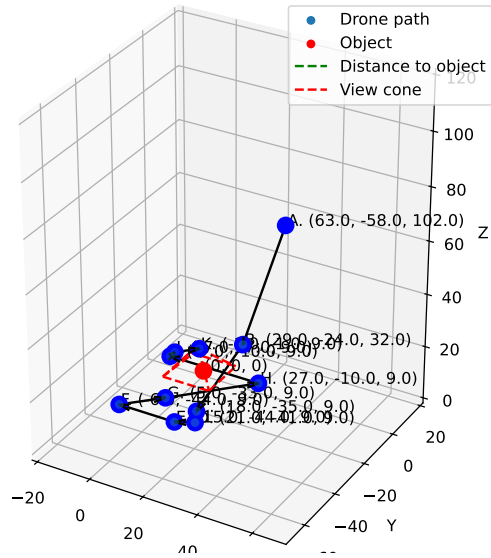
CLAUDE 3.5 SONNET : FOUND

<Reasoning> I am now perfectly positioned at 7 meters altitude, which is below the required 10 meters, and I am centered over what appears to be a collection of debris/trash (dark objects scattered on the ground). The trash pile is clearly visible in my field of view and I am centered on it at coordinates (0,0). </Reasoning>

1.11 Claude 3.5 Sonnet in the City; Success; FS-2

1.11.1 Example

- Starting position: (63.0, -58.0, 102.0)
- End position: (-1.0, -1.0, 9.0)
- Euclidean distance from the object: 9.1104335791443
- Object can be seen: True
- Success: True
- Object type: road construction site



BENCHMARK : <Context> You are in command of a UAV, tasked with finding road construction site. </Context> (REST OF THE PROMPT)

BENCHMARK : The object you're looking for is similar to this. This is NOT the drone's current view.

BENCHMARK : Image 1:

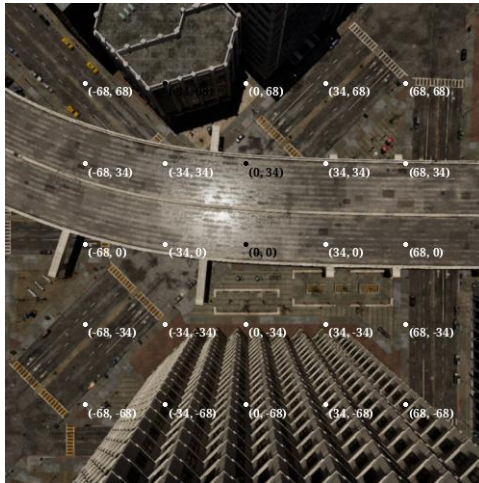
BENCHMARK presents CLAUDE 3.5 SONNET with a relevant image



BENCHMARK : Your current altitude is 102 meters above ground level.

BENCHMARK : Image 2:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*

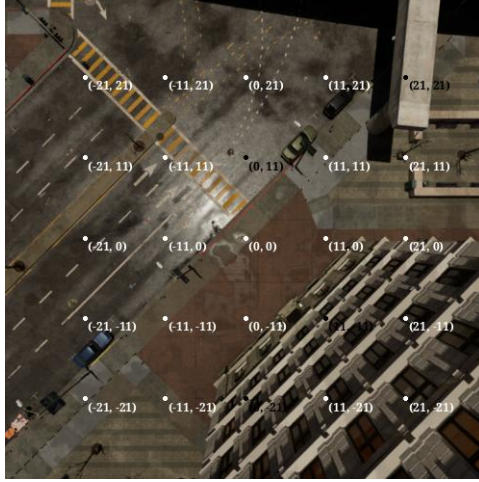


CLAUDE 3.5 SONNET : <Reasoning>I can see a wide urban area from my current position at 102 meters. Given that I'm looking for road construction equipment similar to what was shown in the reference image, I need to get lower to better identify potential targets. I notice some construction-like activity near the coordinates (-34, -34). I'll move there and decrease altitude significantly to get a better view.</Reasoning> <Action>(-34, -34, -70)</Action>

BENCHMARK : Your current altitude is 32 meters above ground level.

BENCHMARK : Image 3:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*



CLAUDE 3.5 SONNET : <Reasoning>I can see some construction-related activity in this area, but I need to get closer to confirm. I notice something that might be construction equipment near (-11, 11). I'll move there and decrease altitude to get under 10 meters for a better view.</Reasoning> <Action>(-11, 11, -23)</Action>

BENCHMARK : Your current altitude is 9 meters above ground level.

BENCHMARK : Image 4:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*



CLAUDE 3.5 SONNET : <Reasoning>From this position, I can see road markings and some construction work, but I don't see the specific type of construction equipment I'm looking for (like what was shown in the reference image). I'll move towards (3, 6) where I can see some yellow markings that might be related to construction activity.</Reasoning> <Action>(3, 6, 0)</Action>

BENCHMARK : Your current altitude is 9 meters above ground level.

BENCHMARK : Image 5:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*

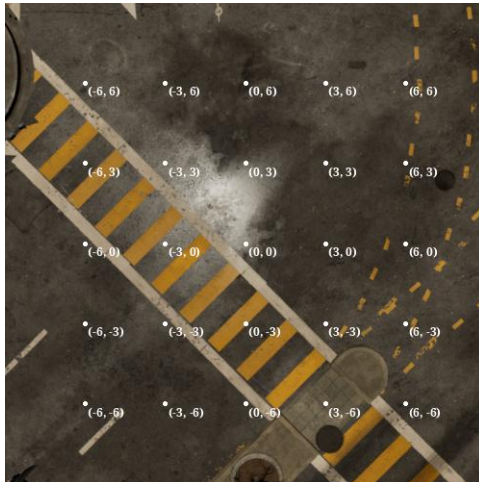


CLAUDE 3.5 SONNET : <Reasoning>I still don't see the specific construction equipment I'm looking for in this area. I need to explore more of the search area. I'll move towards (-6, 3) to check the area with some different markings on the road.</Reasoning> <Action>(-6, 3, 0)</Action>

BENCHMARK : Your current altitude is 9 meters above ground level.

BENCHMARK : Image 6:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*

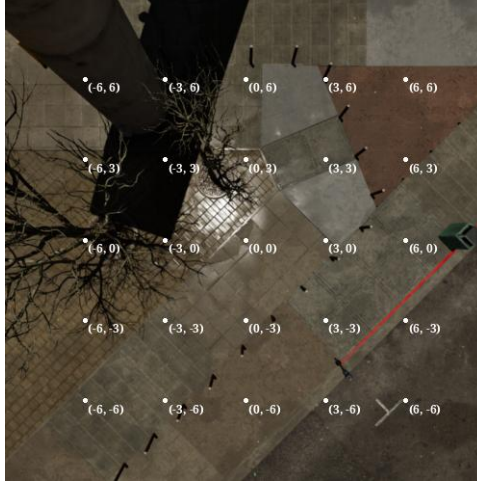


CLAUDE 3.5 SONNET : <Reasoning>The construction equipment isn't visible in this section either. I need to explore more of the initial search area. Let's move back to a different section where there might be construction activity. I'll head towards (-21, 0) to check another part of the road.</Reasoning> <Action>(-21, 0, 0)</Action>

BENCHMARK : Your current altitude is 9 meters above ground level.

BENCHMARK : Image 7:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*



CLAUDE 3.5 SONNET : <Reasoning>This area shows different pavement patterns but no construction equipment. I need to check another section of the search area. I'll move towards (11, -11) to investigate an area we haven't explored yet.</Reasoning> <Action>(11, -11, 0)</Action>

BENCHMARK : Your current altitude is 9 meters above ground level.

BENCHMARK : Image 8:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*



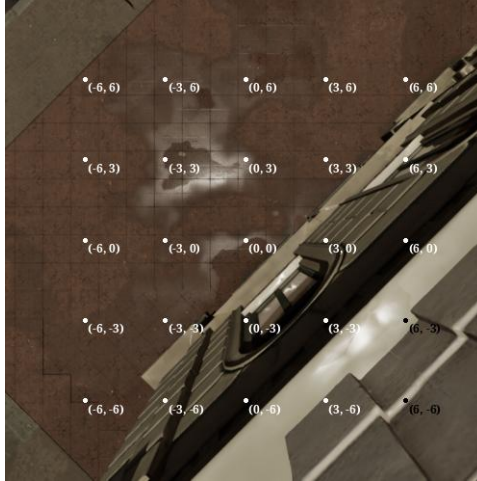
CLAUDE 3.5 SONNET : <Reasoning>Still no sign of the construction equipment we're looking for. Let's check the area near (34, -34) where there might be more construction activity. We haven't explored that section yet.</Reasoning> <Action>(34, -34, 0)</Action>

BENCHMARK : Emergency stop; you've flown too close to something and would have hit it.

BENCHMARK : Your current altitude is 9 meters above ground level.

BENCHMARK : Image 9:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*



CLAUDE 3.5 SONNET : <Reasoning>I apologize for the unsafe move. I need to be more careful about obstacles. Let's move to a safer location and continue our search. I'll head towards (-34, 0) where there appears to be more open space.</Reasoning> <Action>(-34, 0, 0)</Action>

BENCHMARK : Your current altitude is 9 meters above ground level.

BENCHMARK : Image 10:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*



CLAUDE 3.5 SONNET : <Reasoning>I see something that looks like construction equipment at coordinates (0, -3). Let me center on it to confirm if it matches what we're looking for.</Reasoning> <Action>(0, -3, 0)</Action>

BENCHMARK : Your current altitude is 9 meters above ground level.

BENCHMARK : Image 11:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*



CLAUDE 3.5 SONNET : <Reasoning>I can see what appears to be road construction equipment in the lower right corner of the image, matching the reference image shown earlier. Let me move closer to it at coordinates (6, -6) to confirm.</Reasoning> <Action>(6, -6, 0)</Action>

BENCHMARK : Your current altitude is 9 meters above ground level.

BENCHMARK : Image 12:

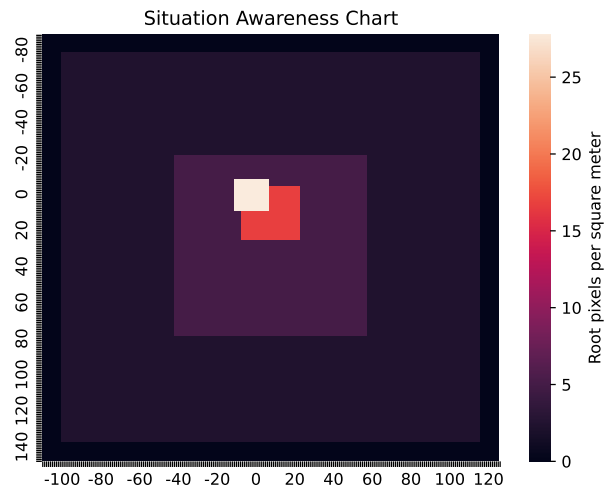
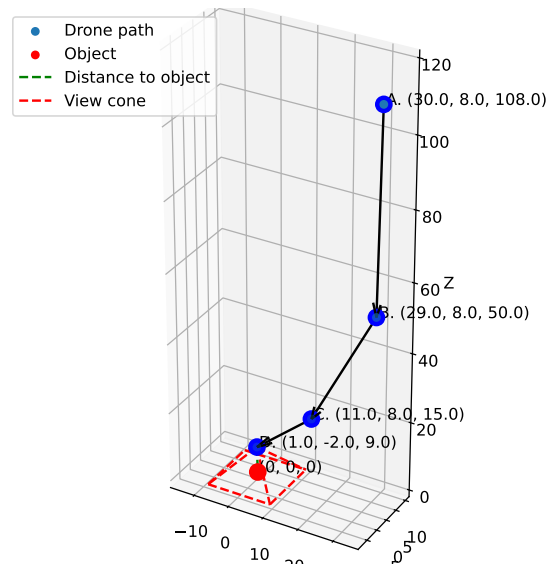
BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*



CLAUDE 3.5 SONNET : FOUND

1.11.2 Example

- Starting position: (31.0, 8.0, 108.0)
- End position: (1.0, -2.0, 9.0)
- Euclidean distance from the object: 9.273618495495704
- Object can be seen: True
- Success: True
- Object type: fire



BENCHMARK : <Context> You are in command of a UAV, tasked with finding fire. </Context>
(REST OF THE PROMPT)

BENCHMARK : The object you're looking for is similar to this. This is NOT the drone's current view.

BENCHMARK : Image 1:

BENCHMARK presents CLAUDE 3.5 SONNET with a relevant image



BENCHMARK : Your current altitude is 108 meters above ground level.

BENCHMARK : Image 2:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*

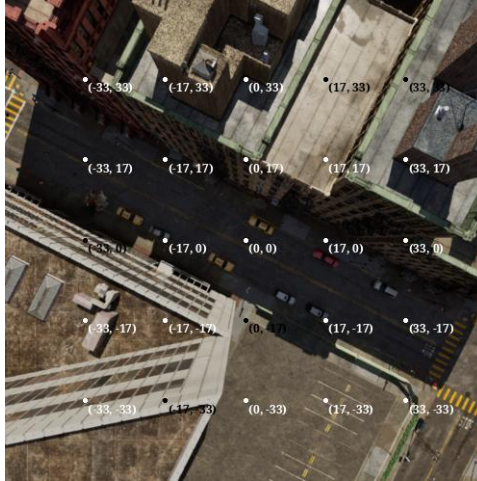


CLAUDE 3.5 SONNET : <Reasoning> I need to start a systematic search for any signs of fire. Given the example images show objects on fire at ground level, I should begin by lowering my altitude to get a better view while maintaining a safe height initially. I'll start by moving to the center of the visible area (0,0) and descending to about 50 meters to get a better overview while still maintaining good visibility. </Reasoning> <Action>(0, 0, -58)</Action>

BENCHMARK : Your current altitude is 50 meters above ground level.

BENCHMARK : Image 3:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*

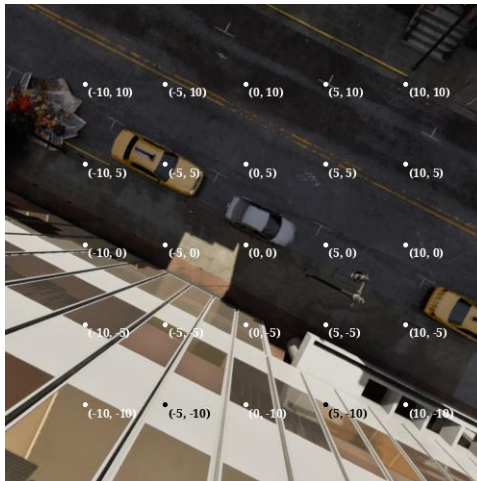


CLAUDE 3.5 SONNET : <Reasoning> From this lower altitude, I can see more detail. I notice what appears to be some objects on the street around (-17, 0). I should move closer to that area and descend further to get a better look at what might be potential fire hazards or burning objects. </Reasoning> <Action>(-17, 0, -35)</Action>

BENCHMARK : Your current altitude is 15 meters above ground level.

BENCHMARK : Image 4:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*

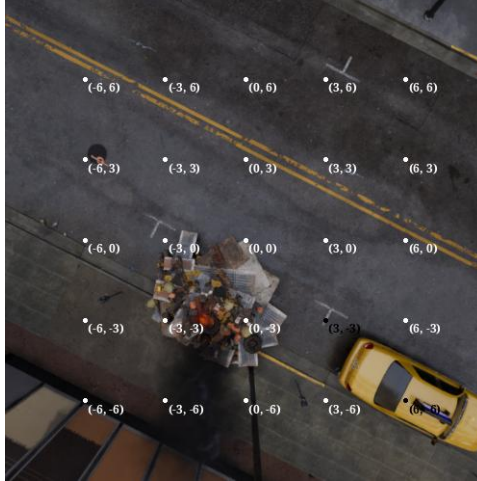


CLAUDE 3.5 SONNET : <Reasoning> I can see what appears to be flames or fire at approximately (-10, 10). I should move directly over it and descend slightly to get within 10 meters while keeping it in my field of view. </Reasoning> <Action>(-10, 10, -6)</Action>

BENCHMARK : Your current altitude is 9 meters above ground level.

BENCHMARK : Image 5:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*

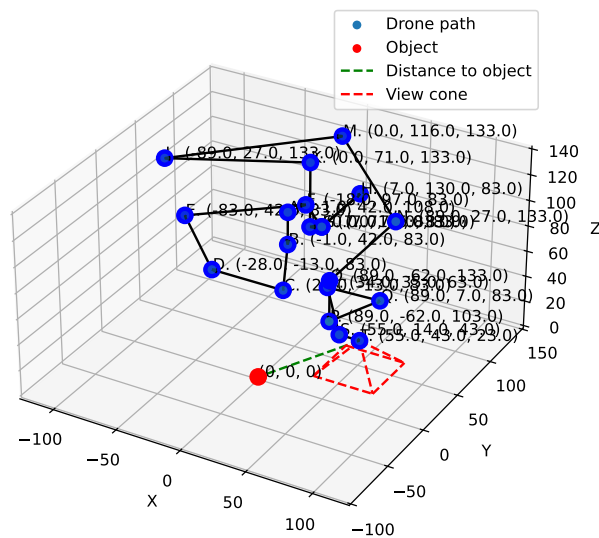


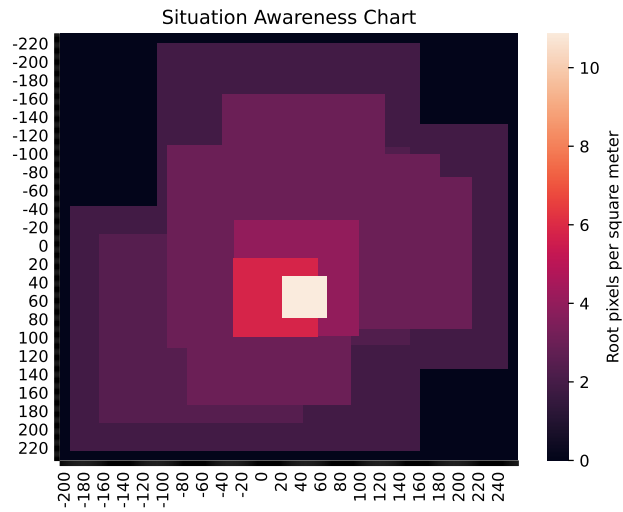
CLAUDE 3.5 SONNET : <Reasoning> I can now clearly see the burning debris/fire at my current position. I am at 9 meters altitude which is below the required 10 meters, and the fire is directly in my field of view. I am properly centered over the target. </Reasoning> FOUND

1.12 Claude 3.5 Sonnet in the City; Failure; FS-2

1.12.1 Example

- Starting position: (-1.0, 42.0, 108.0)
- End position: (55.0, 43.0, 23.0)
- Euclidean distance from the object: 73.50510186374821
- Object can be seen: False
- Success: False
- Object type: crowd





BENCHMARK : <Context> You are in command of a UAV, tasked with finding crowd. </Context>
(REST OF THE PROMPT)

BENCHMARK : The object you're looking for is similar to this. This is NOT the drone's current view.

BENCHMARK : Image 1:

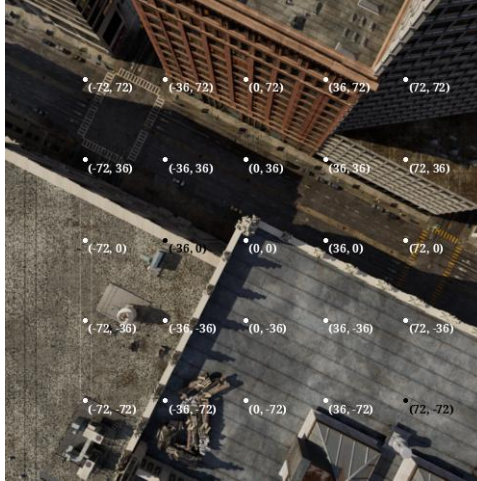
BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*



BENCHMARK : Your current altitude is 108 meters above ground level.

BENCHMARK : Image 2:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*



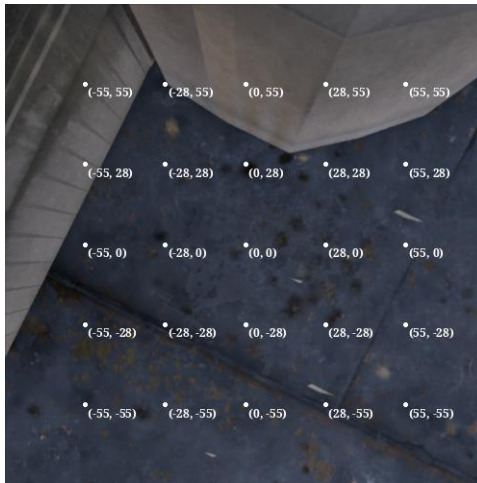
CLAUDE 3.5 SONNET : <Reasoning>From this height, I need to conduct a systematic search for a crowd gathering. I notice some potential movement near the (0, 0) coordinate. I'll move closer to that area and reduce altitude significantly to get a better view. The example shows a crowd from above, so I need to be methodical in my search while maintaining a safe distance.</Reasoning> <Action>(0, 0, -50)</Action>

BENCHMARK : Emergency stop; you've flown too close to something and would have hit it.

BENCHMARK : Your current altitude is 83 meters above ground level.

BENCHMARK : Image 3:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*

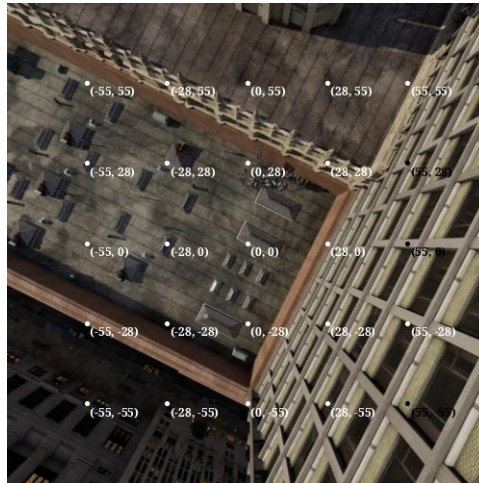


CLAUDE 3.5 SONNET : <Reasoning>I need to be more careful about altitude changes. I see that this area appears to be indoors or covered. I should move to a different area to search for the crowd, as they're more likely to be in an open space. Let's move towards (28, 55) where there might be more open area, and maintain the current altitude for safety.</Reasoning> <Action>(28, 55, 0)</Action>

BENCHMARK : Your current altitude is 83 meters above ground level.

BENCHMARK : Image 4:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*

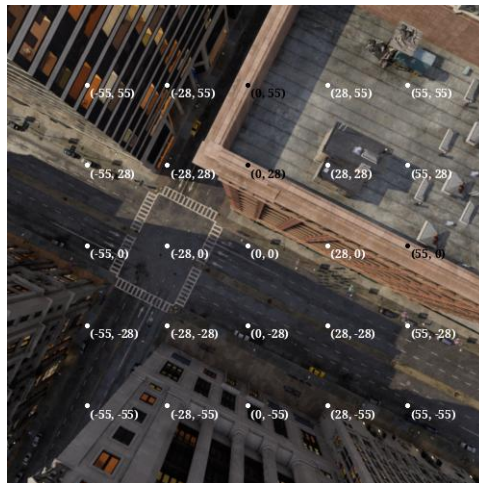


CLAUDE 3.5 SONNET : <Reasoning>I need to expand my search area as this location shows mostly building structures. Let's move to (-55, 0) to explore the western area where there might be more open spaces that could accommodate a crowd. I'll maintain the current altitude for now until I spot a potential gathering.</Reasoning> <Action>(-55, 0, 0)</Action>

BENCHMARK : Your current altitude is 83 meters above ground level.

BENCHMARK : Image 5:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*

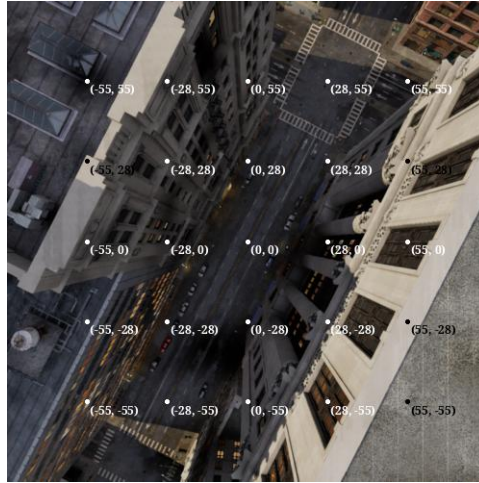


CLAUDE 3.5 SONNET : <Reasoning>I can see an intersection and open streets below. There appears to be more visible area to the south that might be suitable for a crowd gathering. Let's move to (-55, -55) to explore that direction while maintaining our current altitude until we spot any potential gatherings.</Reasoning> <Action>(-55, -55, 0)</Action>

BENCHMARK : Your current altitude is 83 meters above ground level.

BENCHMARK : Image 6:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*



CLAUDE 3.5 SONNET : *<Reasoning>*I see mostly streets and buildings in this area. Let's explore the eastern side of the search area where there might be open spaces or plazas. I'll move to (65, -55) to continue our systematic search while maintaining the current altitude until we spot any gatherings.*</Reasoning>* *<Action>*(65, -55, 0)*</Action>*

BENCHMARK : Your current altitude is 83 meters above ground level.

BENCHMARK : Image 7:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*



CLAUDE 3.5 SONNET : *<Reasoning>*This area shows mostly rooftops with mechanical equipment. Let's check the northern area that we haven't explored yet. I'll move to (65, 55) to continue our systematic search while maintaining altitude until we find signs of a crowd gathering.*</Reasoning>* *<Action>*(65, 55, 0)*</Action>*

BENCHMARK : Emergency stop; you've flown too close to something and would have hit it.

BENCHMARK : Your current altitude is 83 meters above ground level.

BENCHMARK : Image 8:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*



CLAUDE 3.5 SONNET : <Reasoning>I need to be more careful about obstacles. This appears to be an indoor or covered area, which isn't suitable for crowd gatherings. Let's move back towards (0, -55) where there might be more open space, and stay at our current altitude for safety.</Reasoning> <Action>(0, -55, 0)</Action>

BENCHMARK : Your current altitude is 83 meters above ground level.

BENCHMARK : Image 9:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*



CLAUDE 3.5 SONNET : <Reasoning>I'm still seeing mostly rooftop structures. Since we've explored several areas without finding the crowd, let's try moving to (0, 55) to check the northern open areas. I'll maintain the current altitude until we spot potential gathering spaces.</Reasoning> <Action>(0, 55, 0)</Action>

BENCHMARK : Your current altitude is 83 meters above ground level.

BENCHMARK : Image 10:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*



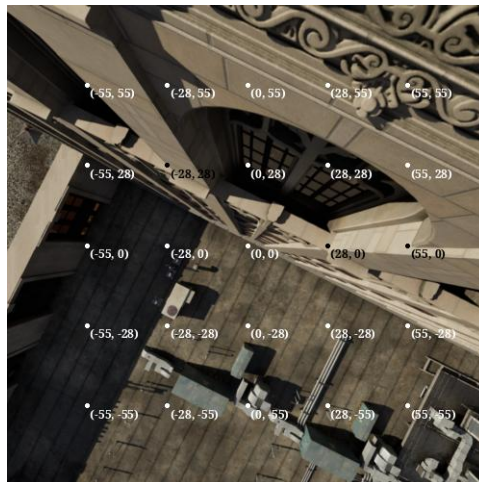
CLAUDE 3.5 SONNET : <Reasoning>We seem to be in some kind of covered or indoor structure. We need to get to an open area where crowds would typically gather. Let's try moving to (-55, 28) where there might be a plaza or open space, while maintaining our current altitude for safety.</Reasoning> <Action>(-55, 28, 0)</Action>

BENCHMARK : Emergency stop; you've flown too close to something and would have hit it.

BENCHMARK : Your current altitude is 83 meters above ground level.

BENCHMARK : Image 11:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*

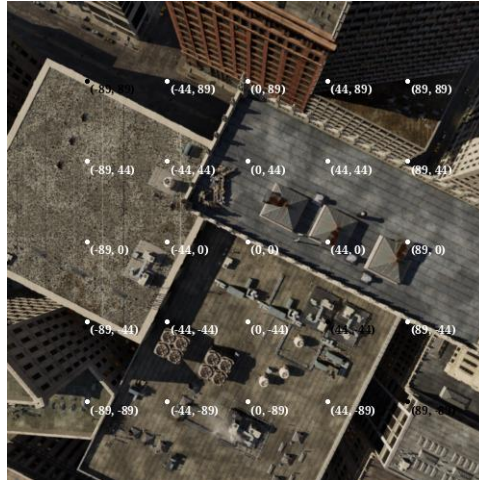


CLAUDE 3.5 SONNET : <Reasoning>I need to get out of this covered structure. Since we're having trouble navigating at this level, let's first gain some altitude to get a better overview of the area and avoid obstacles. I'll increase altitude by 50 meters to help identify open spaces where crowds might gather.</Reasoning> <Action>(0, 0, 50)</Action>

BENCHMARK : Your current altitude is 133 meters above ground level.

BENCHMARK : Image 12:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*

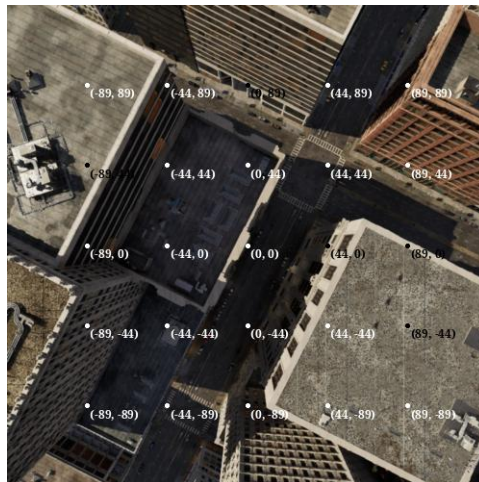


CLAUDE 3.5 SONNET : <Reasoning>From this higher vantage point, I can see what appears to be an open area to the northwest (-89, 44). That looks like the kind of space where a crowd might gather. Let's move there to investigate while maintaining this safer altitude before making any descent.</Reasoning> <Action>(-89, 44, 0)</Action>

BENCHMARK : Your current altitude is 133 meters above ground level.

BENCHMARK : Image 13:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*

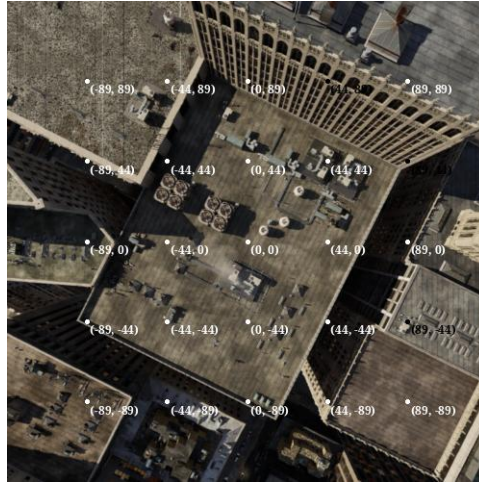


CLAUDE 3.5 SONNET : <Reasoning>I can see a large open area in the southeast quadrant of the image. Let's move to (89, -89) to investigate that space. From this altitude, we have a good overview but we'll need to descend once we spot any potential crowd gatherings.</Reasoning> <Action>(89, -89, 0)</Action>

BENCHMARK : Your current altitude is 133 meters above ground level.

BENCHMARK : Image 14:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*

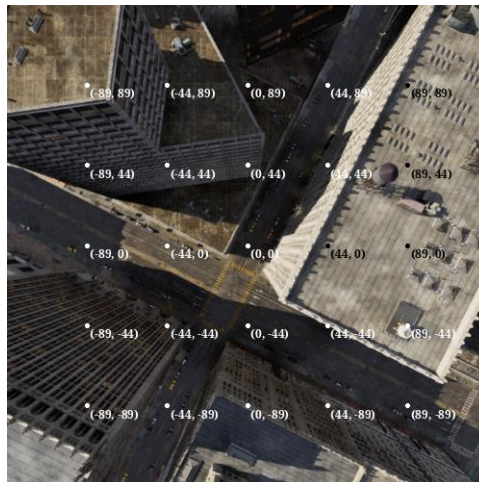


CLAUDE 3.5 SONNET : <Reasoning>I'm seeing what appears to be the rooftop of a building with mechanical equipment. Looking at the entire search pattern so far, we haven't thoroughly checked the northeastern area. Let's move to (89, 89) to explore that sector while maintaining this altitude for a good overview.</Reasoning> <Action>(89, 89, 0)</Action>

BENCHMARK : Your current altitude is 133 meters above ground level.

BENCHMARK : Image 15:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*

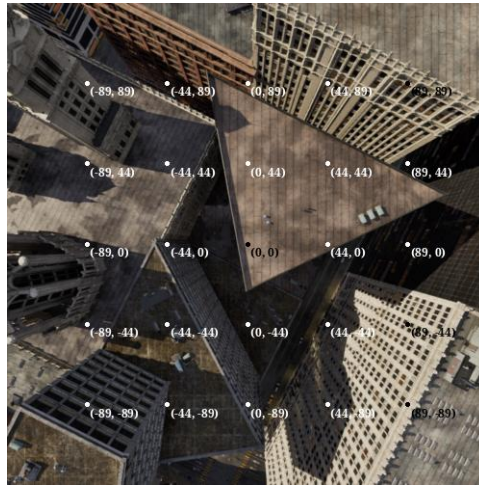


CLAUDE 3.5 SONNET : <Reasoning>I see streets and buildings but no crowd gatherings yet. Let's move to (0, 89) to explore the northern area more thoroughly. We've covered most other directions, so this is one of our remaining unexplored areas that could potentially have open spaces suitable for crowds.</Reasoning> <Action>(0, 89, 0)</Action>

BENCHMARK : Your current altitude is 133 meters above ground level.

BENCHMARK : Image 16:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*



CLAUDE 3.5 SONNET : <Reasoning>From this vantage point, I can see what appears to be a triangular open space or plaza. Let's move to (-44, 44) to get a better view of this area, and then we can consider descending if we spot any crowd gatherings. We need to be methodical as we haven't found our target yet.</Reasoning> <Action>(-44, 44, 0)</Action>

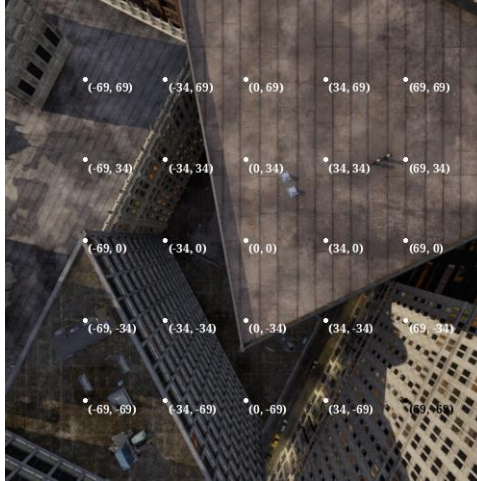
BENCHMARK : This command would cause you to fly out of the search area's bounds. You can't fly further than 108.0 meters from the starting point in any axis.

CLAUDE 3.5 SONNET : <Reasoning>I apologize for trying to exceed the search bounds. Let's adjust our strategy. Since we're at 133 meters height and haven't found the crowd yet, we should try descending a bit in a central location to get a clearer view. I'll move to (0, 0) and reduce altitude to get a better look at the ground level while staying within bounds.</Reasoning> <Action>(0, 0, -30)</Action>

BENCHMARK : Your current altitude is 103 meters above ground level.

BENCHMARK : Image 17:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*

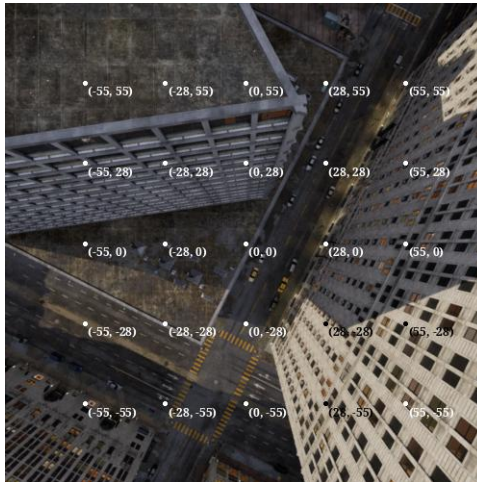


CLAUDE 3.5 SONNET : <Reasoning>I see we're still quite high up and looking at mostly building structures. Let's check the southern area at (0, -69) and continue descending to get closer to ground level where crowds would be. We need to find open spaces at street level.</Reasoning> <Action>(0, -69, -20)</Action>

BENCHMARK : Your current altitude is 83 meters above ground level.

BENCHMARK : Image 18:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*

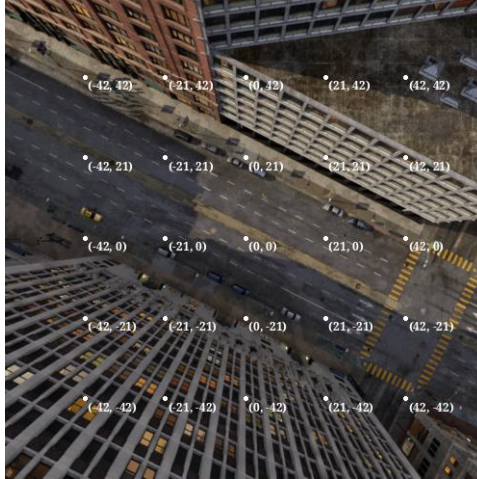


CLAUDE 3.5 SONNET : <Reasoning>I can now see an intersection with streets below. Let's check the western open area at (-55, -28) where there appears to be more space that could accommodate a crowd, and continue to descend carefully to get a better view of ground level activity.</Reasoning> <Action>(-55, -28, -20)</Action>

BENCHMARK : Your current altitude is 63 meters above ground level.

BENCHMARK : Image 19:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*

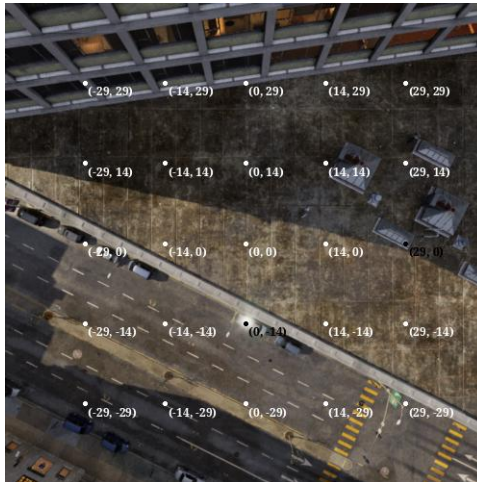


CLAUDE 3.5 SONNET : <Reasoning>I see a wide street with intersections but still no crowd. Let's check the northeast area at (21, 21) where there might be a plaza or open space. We're getting closer to ground level which is good, but we need to continue our search methodically.</Reasoning> <Action>(21, 21, -20)</Action>

BENCHMARK : Your current altitude is 43 meters above ground level.

BENCHMARK : Image 20:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*

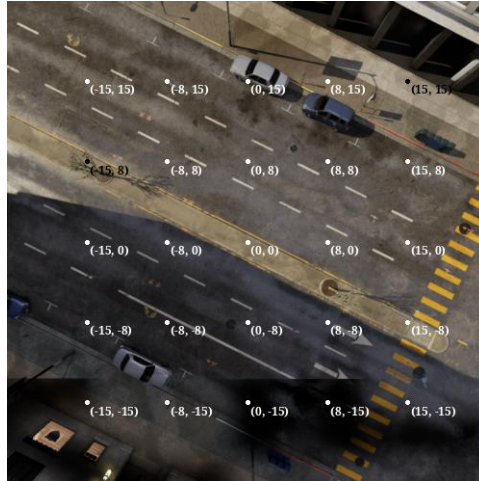


CLAUDE 3.5 SONNET : <Reasoning>We're getting close to our 19-move limit and haven't found the crowd yet. I notice we haven't thoroughly checked the southern area. Let's make a final move to (0, -29) and descend further to get a better view of any potential gatherings at street level.</Reasoning> <Action>(0, -29, -20)</Action>

BENCHMARK : Your current altitude is 23 meters above ground level.

BENCHMARK : Image 21:

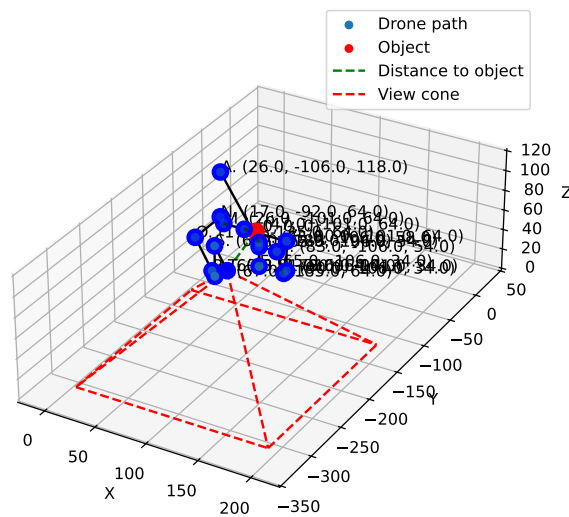
BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*

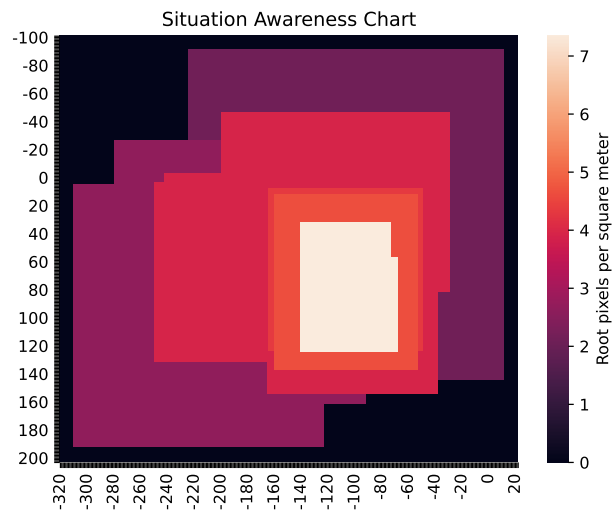


CLAUDE 3.5 SONNET : <Reasoning>We're now much closer to ground level but I still don't see the crowd we're looking for. With only one move left, let's check the last remaining area at (-15, 8) where there appears to be some open street space, and continue descending to get even closer to ground level.</Reasoning> <Action>(-15, 8, -10)</Action>

1.12.2 Example

- Starting position: (26.0, -106.0, 118.0)
- End position: (98.0, -216.0, 94.0)
- Euclidean distance from the object: 255.1391777050322
- Object can be seen: False
- Success: False
- Object type: road construction site





BENCHMARK : <Context> You are in command of a UAV, tasked with finding road construction site. </Context> (REST OF THE PROMPT)

BENCHMARK : The object you're looking for is similar to this. This is NOT the drone's current view.

BENCHMARK : Image 1:

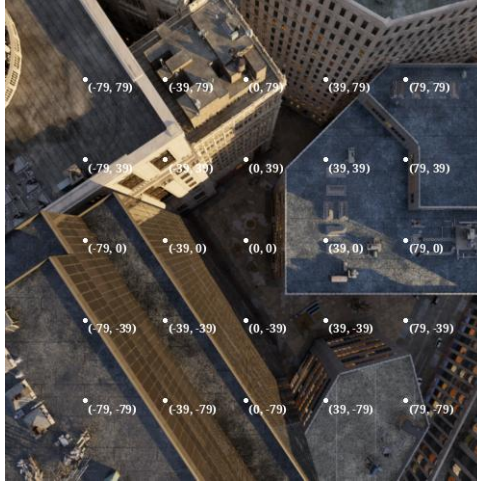
BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*



BENCHMARK : Your current altitude is 118 meters above ground level.

BENCHMARK : Image 2:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*



CLAUDE 3.5 SONNET : <Reasoning>From the high altitude of 118 meters, I need to conduct a systematic search for the road construction equipment. I can see several areas that need closer inspection. I'll start by moving to coordinates (39, 0) and reducing altitude by 60 meters to get a better view of the central area. This will give me a good vantage point to spot any construction equipment while maintaining enough height to keep a broad field of view.</Reasoning> <Action>(39, 0, -60)</Action>

BENCHMARK : Your current altitude is 58 meters above ground level.

BENCHMARK : Image 3:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*



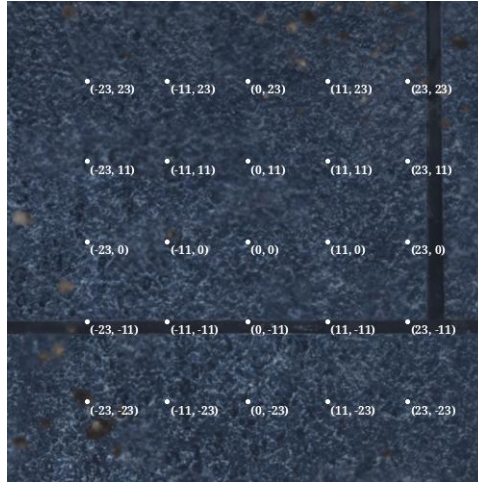
CLAUDE 3.5 SONNET : <Reasoning>I can see what appears to be construction equipment near coordinates (0, 0). I'll move closer to that position and reduce altitude further by 40 meters to get a better look and ensure I'm within the required 10-meter height requirement.</Reasoning> <Action>(0, 0, -40)</Action>

BENCHMARK : Emergency stop; you've flown too close to something and would have hit it.

BENCHMARK : Your current altitude is 34 meters above ground level.

BENCHMARK : Image 4:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*



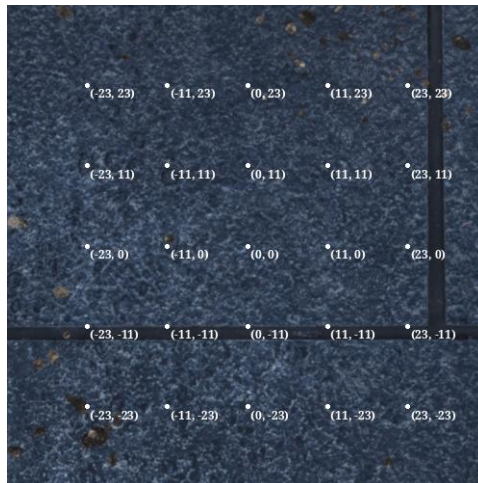
CLAUDE 3.5 SONNET : <Reasoning>I need to be more careful with my movements. I can see construction equipment near my current position, but I need to get closer while maintaining a safe distance. I'll make a smaller adjustment to my position and reduce altitude by 25 meters to get within the required 10-meter height while staying safe.</Reasoning> <Action>(11, 0, -25)</Action>

BENCHMARK : Emergency stop; you've flown too close to something and would have hit it.

BENCHMARK : Your current altitude is 34 meters above ground level.

BENCHMARK : Image 5:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*



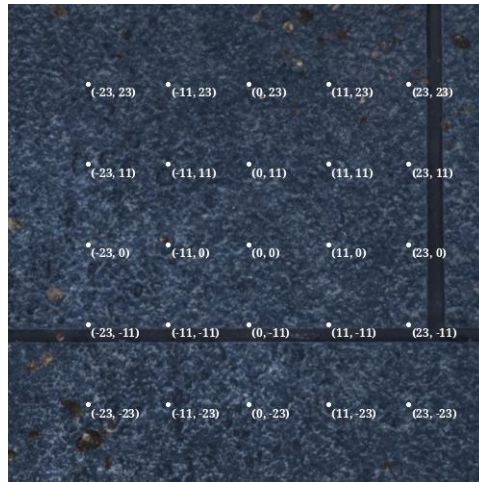
CLAUDE 3.5 SONNET : <Reasoning>I need to adjust my strategy. Let me try moving to a slightly different position and reduce altitude by a smaller amount to avoid any potential collisions. I'll move to (19, 11) and descend 15 meters to get a safer angle on the construction site.</Reasoning> <Action>(19, 11, -15)</Action>

BENCHMARK : Emergency stop; you've flown too close to something and would have hit it.

BENCHMARK : Your current altitude is 34 meters above ground level.

BENCHMARK : Image 6:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*



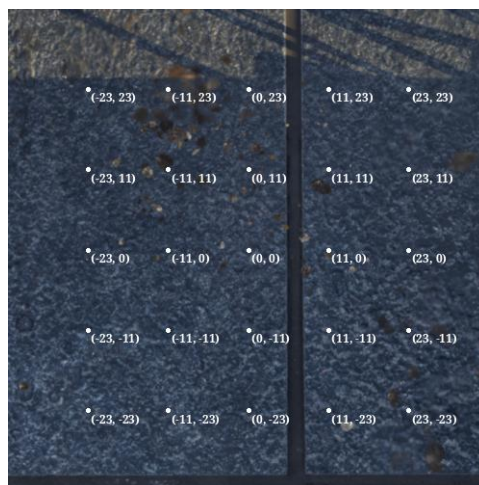
CLAUDE 3.5 SONNET : <Reasoning>I'm having difficulty getting closer safely. Let me try moving to a position further out first and then reassess. I'll move to (23, 23) to get a different perspective while maintaining the current altitude.</Reasoning> <Action>(23, 23, 0)</Action>

BENCHMARK : Emergency stop; you've flown too close to something and would have hit it.

BENCHMARK : Your current altitude is 34 meters above ground level.

BENCHMARK : Image 7:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*



CLAUDE 3.5 SONNET : <Reasoning>I need to change my approach. Let me first gain some altitude to get a better overview and then plan a safer descent path. I'll move up by 20 meters first while staying in my current position.</Reasoning> <Action>(0, 0, 20)</Action>

BENCHMARK : Your current altitude is 54 meters above ground level.

BENCHMARK : Image 8:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*

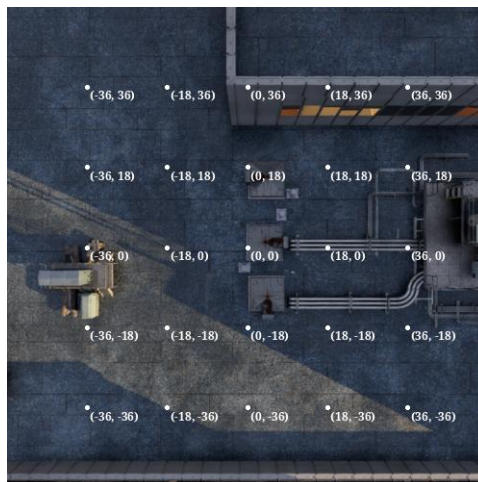


CLAUDE 3.5 SONNET : <Reasoning>Now I have a better view of the area. I can see the construction equipment more clearly. I'll move to (18, 0) to get a better position near the target while keeping my current altitude. This should give me a safer approach angle.</Reasoning>
<Action>(18, 0, 0)</Action>

BENCHMARK : Your current altitude is 54 meters above ground level.

BENCHMARK : Image 9:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*



CLAUDE 3.5 SONNET : <Reasoning>From this position, I can see the construction equipment more clearly. I'll now try to descend gradually to get within 10 meters of it. I'll reduce altitude by 45 meters to get close to the required height while staying in the same horizontal position.</Reasoning> <Action>(18, 0, -45)</Action>

BENCHMARK : Emergency stop; you've flown too close to something and would have hit it.

BENCHMARK : Your current altitude is 34 meters above ground level.

BENCHMARK : Image 10:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*



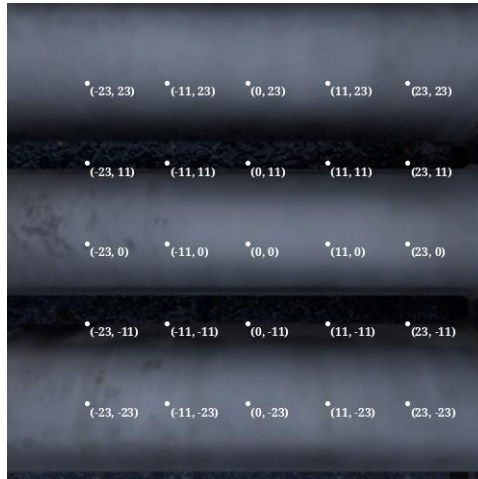
CLAUDE 3.5 SONNET : <Reasoning>I need to move to a position where I have better visibility and can safely descend. I'll move to (0, -23) which should give me a clearer view from a different angle while maintaining my current altitude.</Reasoning> <Action>(0, -23, 0)</Action>

BENCHMARK : Emergency stop; you've flown too close to something and would have hit it.

BENCHMARK : Your current altitude is 34 meters above ground level.

BENCHMARK : Image 11:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*

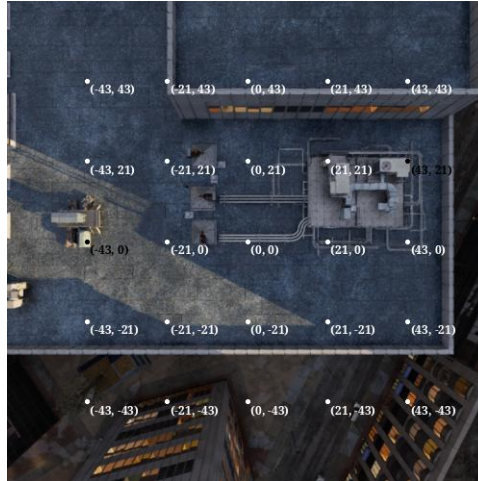


CLAUDE 3.5 SONNET : <Reasoning>I seem to be in a difficult position with limited visibility and movement options. Let me first gain some altitude to get a better overview of the situation. I'll move up 30 meters while staying in my current position.</Reasoning> <Action>(0, 0, 30)</Action>

BENCHMARK : Your current altitude is 64 meters above ground level.

BENCHMARK : Image 12:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*



CLAUDE 3.5 SONNET : <Reasoning>Now I have a better view. I can see the construction equipment near (-43, 0). I'll move to that position while maintaining this safer altitude, which should give me a better approach angle for my final descent.</Reasoning> <Action>(-43, 0, 0)</Action>

BENCHMARK : Your current altitude is 64 meters above ground level.

BENCHMARK : Image 13:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*

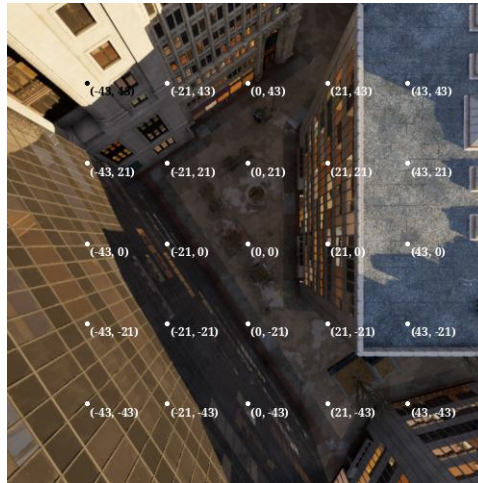


CLAUDE 3.5 SONNET : <Reasoning>From this position, I can see the construction site more clearly. I'll try to move to (-21, 0) which appears to be a safer position closer to the target, while maintaining my current altitude. This should give me a better vantage point for the final approach.</Reasoning> <Action>(-21, 0, 0)</Action>

BENCHMARK : Your current altitude is 64 meters above ground level.

BENCHMARK : Image 14:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*



CLAUDE 3.5 SONNET : <Reasoning>I need to get to a position where I have a clearer view of the construction site. I'll move to (-21, -21) to get away from the buildings and have a better angle, while maintaining my current altitude.</Reasoning> <Action>(-21, -21, 0)</Action>

BENCHMARK : Emergency stop; you've flown too close to something and would have hit it.

BENCHMARK : Your current altitude is 64 meters above ground level.

BENCHMARK : Image 15:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*

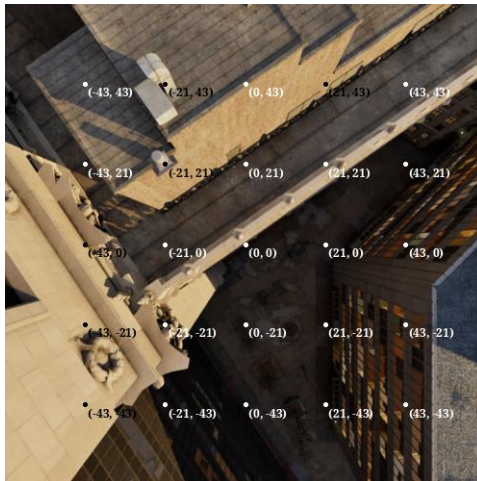


CLAUDE 3.5 SONNET : <Reasoning>I seem to be in a challenging position near tall buildings. I need to move to a more open area. Let me try moving to (0, 43) which should take me away from the buildings while maintaining the current altitude.</Reasoning> <Action>(0, 43, 0)</Action>

BENCHMARK : Your current altitude is 64 meters above ground level.

BENCHMARK : Image 16:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*

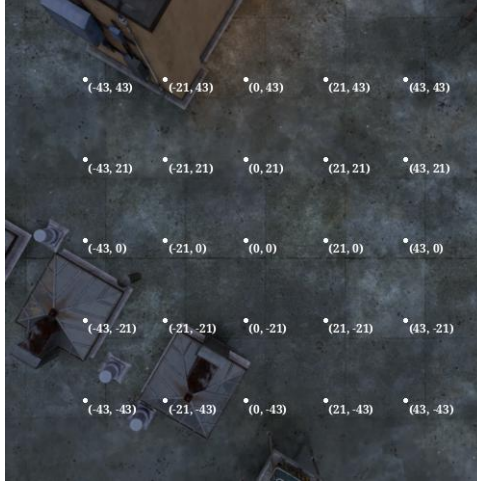


CLAUDE 3.5 SONNET : <Reasoning>I need to get away from these buildings and find a clearer path to the construction site. Let me move to (43, 43) which appears to be a more open area, while maintaining altitude.</Reasoning> <Action>(43, 43, 0)</Action>

BENCHMARK : Your current altitude is 64 meters above ground level.

BENCHMARK : Image 17:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*



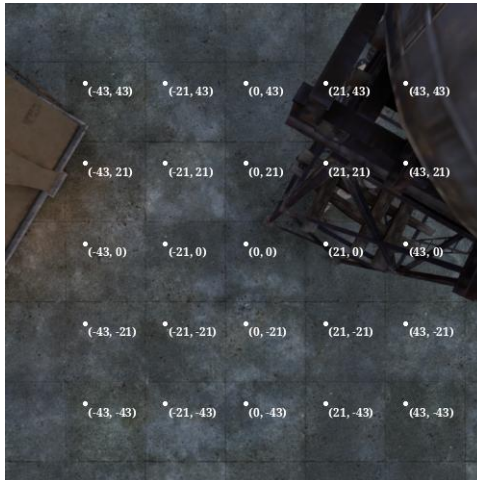
CLAUDE 3.5 SONNET : <Reasoning>Now I have a clearer view of the area. I'll move to (21, 21) which should give me a better position to approach the construction site area, while keeping the current altitude.</Reasoning> <Action>(21, 21, 0)</Action>

BENCHMARK : Emergency stop; you've flown too close to something and would have hit it.

BENCHMARK : Your current altitude is 64 meters above ground level.

BENCHMARK : Image 18:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*



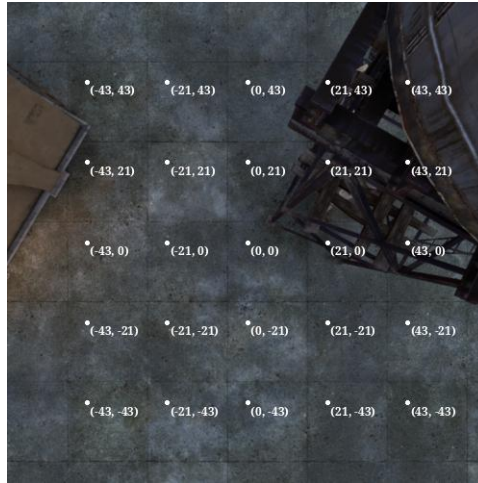
CLAUDE 3.5 SONNET : <Reasoning>I'm still encountering obstacles. Let me try moving to (0, 21) which should be a clearer path while maintaining the current altitude. This position might give me a better angle to approach the target area.</Reasoning> <Action>(0, 21, 0)</Action>

BENCHMARK : Emergency stop; you've flown too close to something and would have hit it.

BENCHMARK : Your current altitude is 64 meters above ground level.

BENCHMARK : Image 19:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*



CLAUDE 3.5 SONNET : $\langle \text{Reasoning} \rangle$ I need to try a different strategy. Let me first gain more altitude to get a better overview and avoid obstacles. I'll increase height by 30 meters while staying in my current position. $\langle / \text{Reasoning} \rangle$ $\langle \text{Action} \rangle (0, 0, 30) \langle / \text{Action} \rangle$

BENCHMARK : Your current altitude is 94 meters above ground level.

BENCHMARK : Image 20:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*



CLAUDE 3.5 SONNET : $\langle \text{Reasoning} \rangle$ From this higher altitude, I have a better view of the area. I can see the road and various structures more clearly. I'll move to (31, 31) which appears to be a clearer area while maintaining this altitude, and then I can plan a safer descent path. $\langle / \text{Reasoning} \rangle$ $\langle \text{Action} \rangle (31, 31, 0) \langle / \text{Action} \rangle$

BENCHMARK : Your current altitude is 94 meters above ground level.

BENCHMARK : Image 21:

BENCHMARK *presents* CLAUDE 3.5 SONNET *with a relevant image*



CLAUDE 3.5 SONNET : <Reasoning>I can see more of the area now. I'll move to (-31, 0) to get closer to where I saw the construction equipment earlier, while maintaining this safer altitude before attempting a final approach.</Reasoning> <Action>(-31, 0, 0)</Action>