legends: Orange (Medium Priority), Red (Critical/High Priority), Blue (Info), Green (Good work!)

Done, Important

# Monday, 15 July 2024

4th Meeting with Dr Hansung Kim @ 10.30am

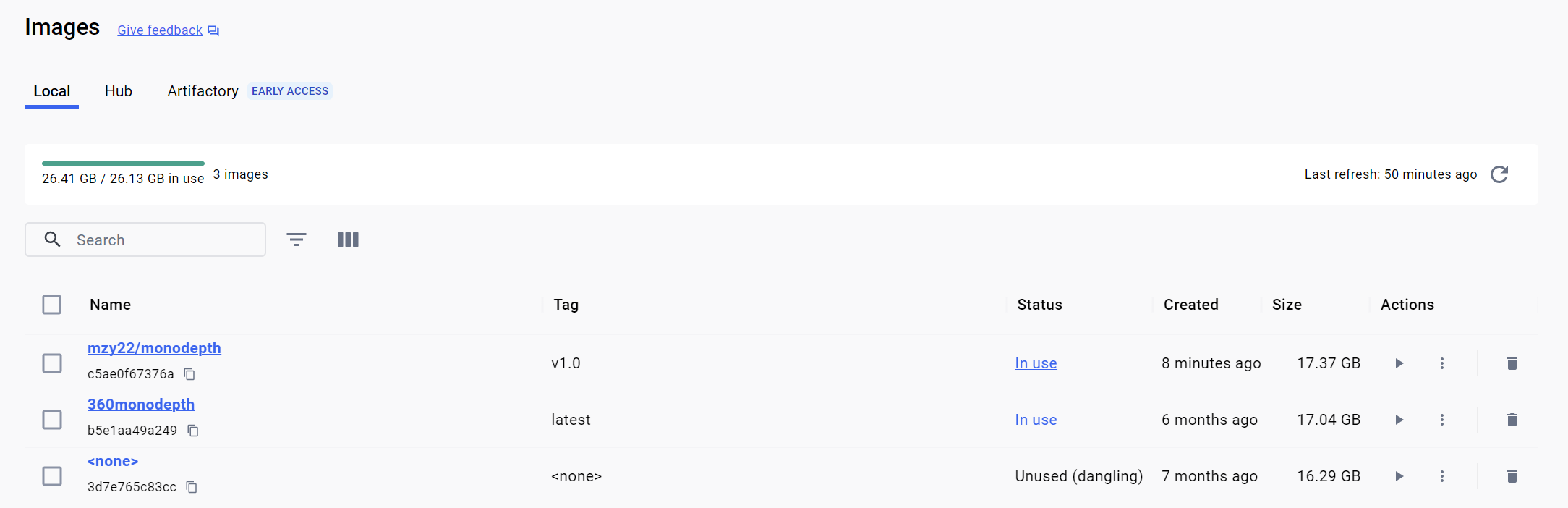
* Keep working in perfecting Unity project to make it as realistic and immersive
* Non VR mode player controller should have collision
* Binaural audio for demonstration
* Mono audio is fine for RIR. (hmm now I think about it, maybe not fine but needed, need to follow how real RIR are calculated, check mic/speaker config for ground truth)
* Mona leaving for WFH starting next week, so ask whatever you need now!
* Email Dr Hansung with recording for feedback.

## **UPDATED TODO:**

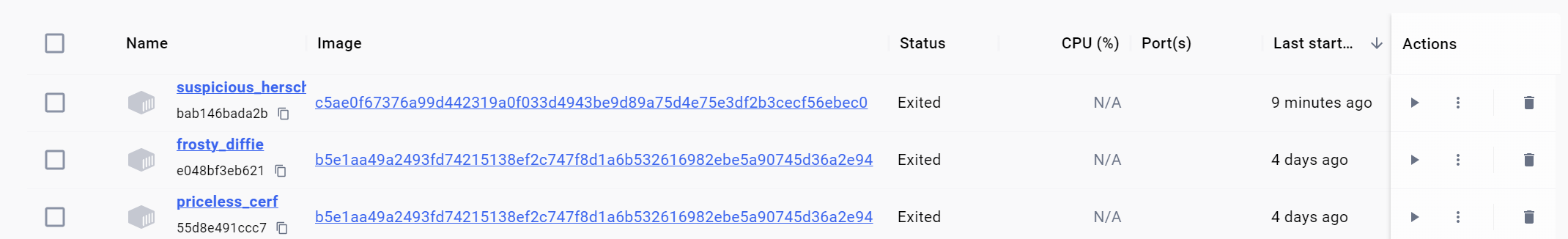
1. Figure out Steam Audio quirks and revamp Unity project
   1. Ask Mona for old unity project for reference, and also for positions of probe/listener/source.
   2. Get MATLAB code for RT60 and EDT calculation.
2. Continue monodepth optimisation ( fix bug on using boostingMonocularDepth on 360monodepth), also check if GDP’s original folder have the submodule or not to make sure.
3. Remove docker clutter
4. Remove LiDAR from LFS, and use separate cloud service to download that, or use MeshLab to optimize it.

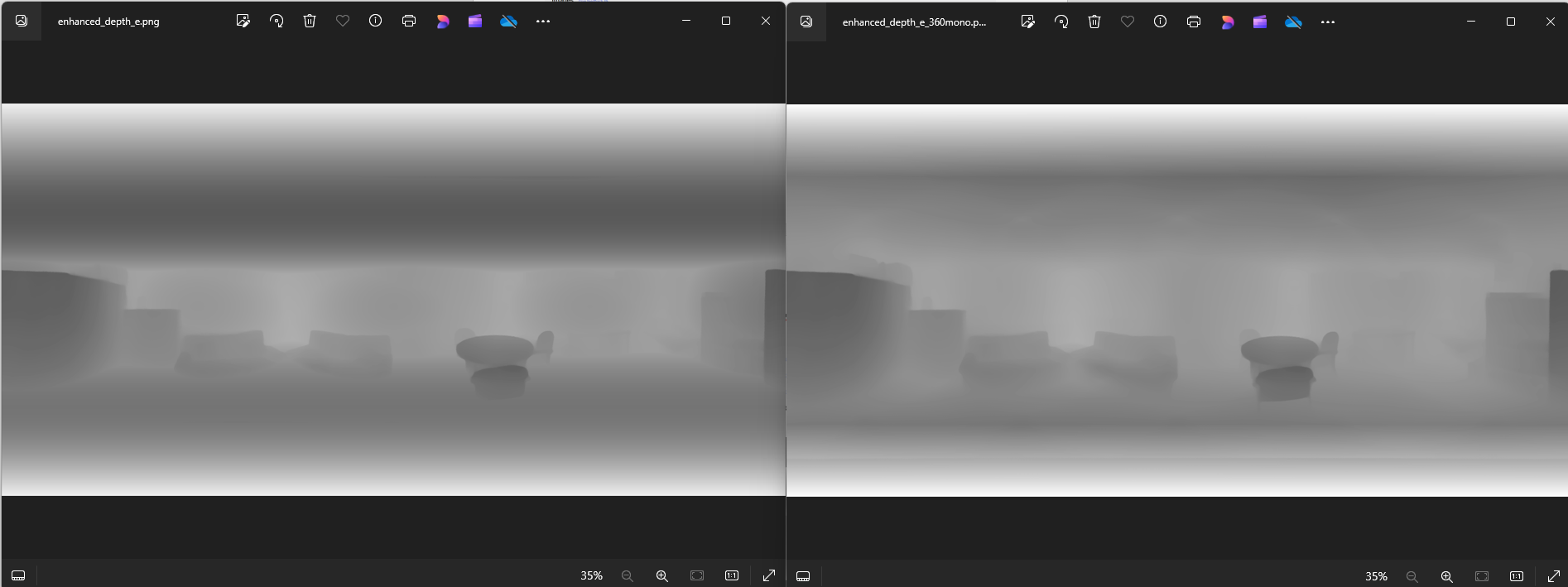
After rereading GDP report on Evaluation part specifically, it also mentions about the limitation from using 360monodepth with edgenet360 which is built and trained with stereo depth map. However, after looking into Future Work section, it seems, they really didn’t implement BoostingMonocularDepth for some reason. This means I should try it. Note to myself: check back the gdp report now and then on specific part to double check their intention and their finding, especially in methodology, evaluation and future work. Don’t be stupid and research same thing and get same finding, but it is good idea to reconfirm their findings!

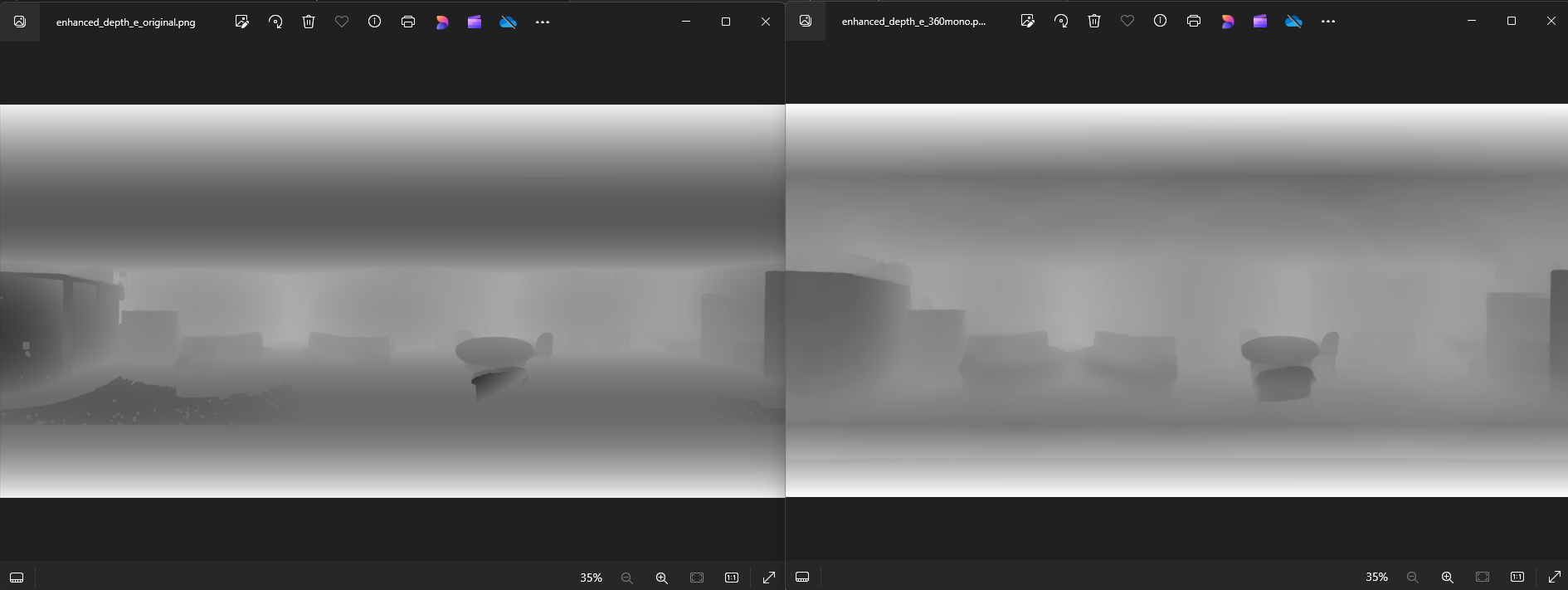
Fixed dockerfile build error on line 39 by replacing it with ‘RUN if [ -d "BoostingMonocularDepth" ]; then echo "Submodule exists"; else echo "Submodule not found" && exit 1; fi’ to check if submodule already exist (manually update before running docker build.)

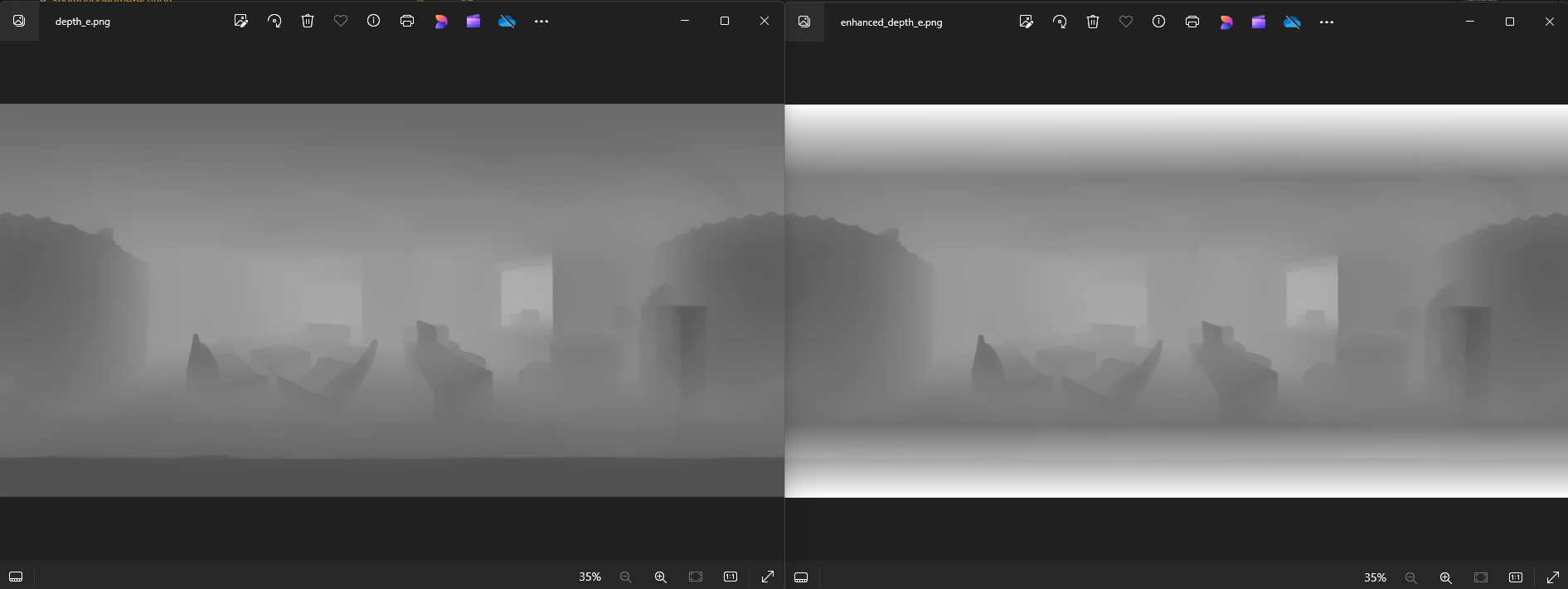


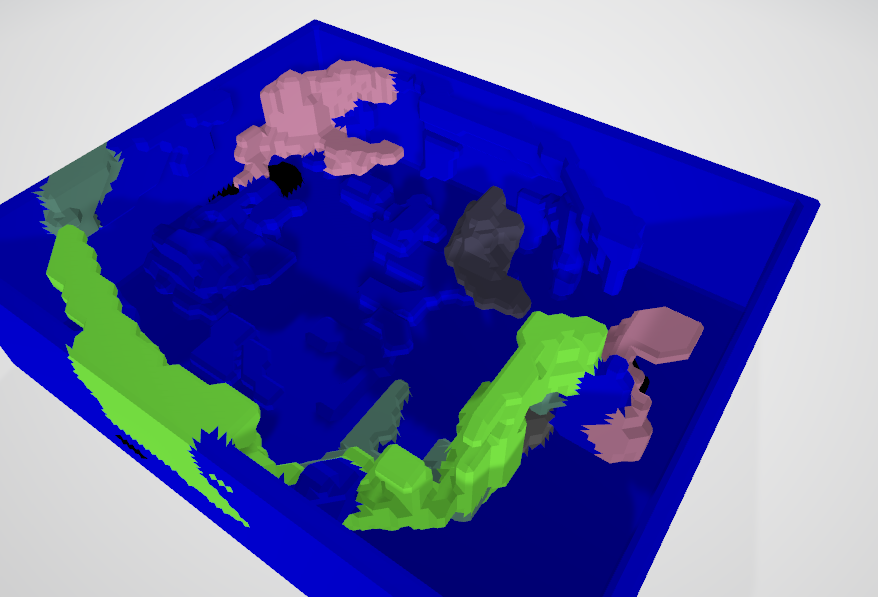
Monodepth v1.0 is the one with BoostingMonocularDepth.

  
it have different Image id as seen above two images, so replace all previous id with new one. (Need to refactor this so only define it in one file instead.

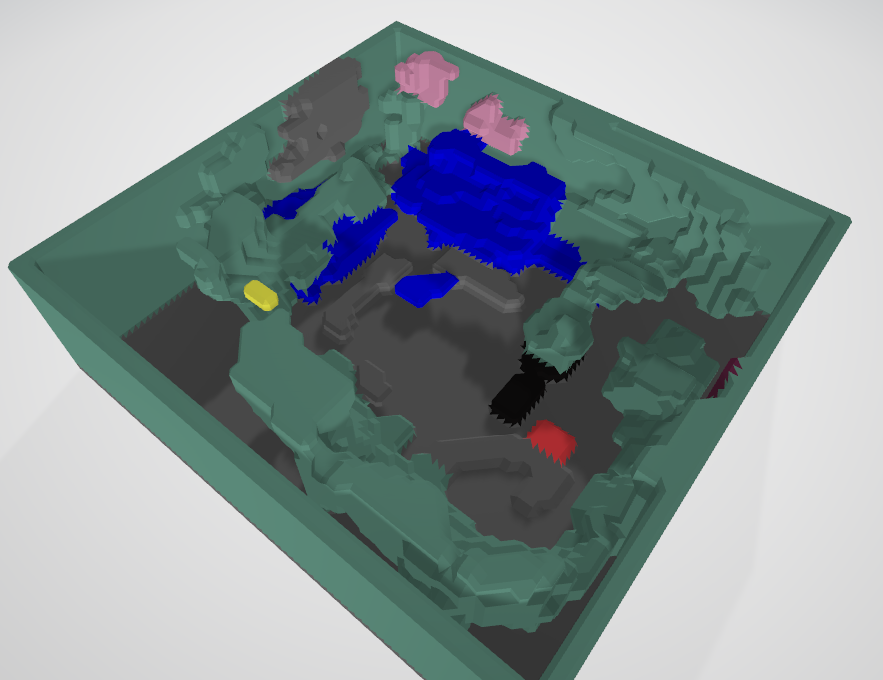


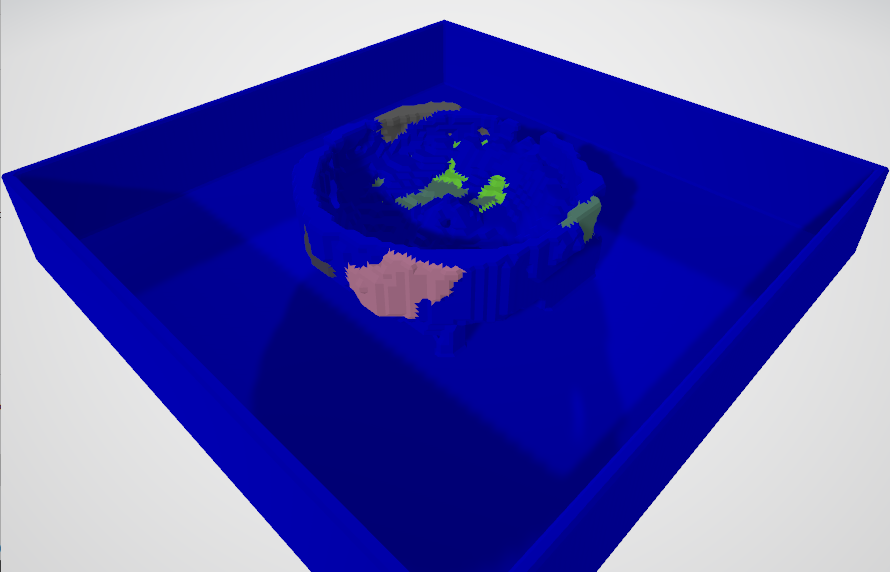
Enhance360.py at –baseline 2.264 vs new enhance360mono.py output that only gradient top and bottom to white, not much different but still can see that right is better in terms of depth details, the room structure might be a lot wrong now as no edge enhancement from rgb specifically for room infer360 made for edgenet and only use MonocularDepthBoosting for that edge enhancement etc. But need to check with mirror (UL scene). Now look at original enhance360.py default setting(baseline) against the new enhance360mono output. Left is sooo bad with lots of missing details and kinda inverted depth for sofa on the left side. But still, we need to see how well this depth map translated to meshes on edgenet, so lets see.

Ran on UL and can confirm MonocularDepthBoosting is still not as good as manual edge optimisation using rgb.png as in enhance360.py 

Left (original depth map), right is depth map output from enhance360mono which don’t have edge optimisation. As seen on mirror on the right, the reflection is still rendered on depth map instead of having only one plane depth.

UL generated mesh. Also causes these artifacts due to presumably unoptimised depth map as compared to enhance360.py, which means I probably need to make a mono version for most function in preproc.py instead.



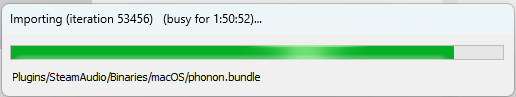
Not as bad for MR especially now we can see the sofa and table more clearly, but still not as ideal as stereo one or previous one in terms of clean mesh imo. Maybe I’ll just generate all scene and ask Dr. Hansung Kim which one he prefer and if I should work on it more etc. 

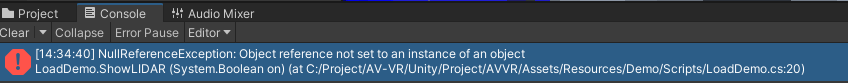
Bruh, I think the V2 have too white thus why give out such crazy room dimension (too large, tbh I should probably just normalise it to be lower, maybe have scaling factor to affect the output depth map as well.)

Rn, worst case I’ll just go back to normal enhance360.py with –baseline 2.264. ( still so bad esp in some scenes with stuff closer to camera… like UL, but idk, lets figure out monoenhance tomorrow.

# Tuesday, 16 July 2024

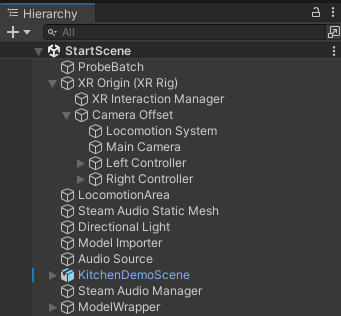
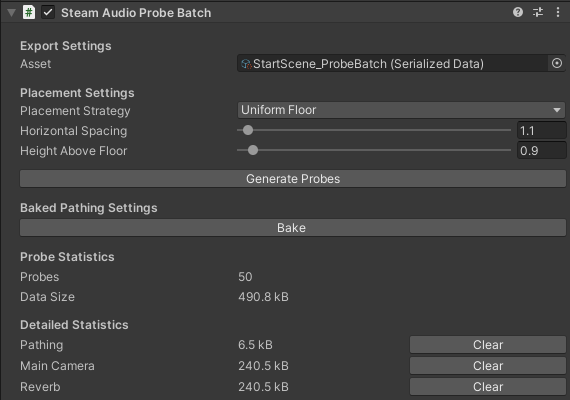
Now reconfiguring the StartScene to include Steam Audio Probe Batch and adding baked source and listener component to audio source and listener respectively. Also changed the behaviour from realtime to baked. This is per Mona’s recommendation. Apparently, the way to export audio is via external application such as Wavepad (recommended by Mona) but I think Audacity would suffice as that is what GDP’s group used. Just make sure to record in stereo mix/channel.

Its so joever

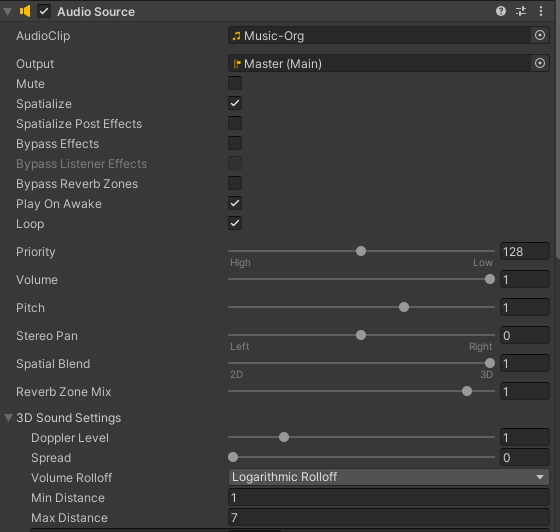
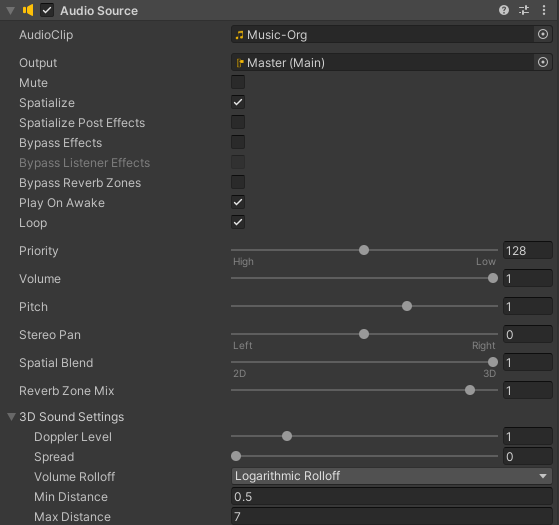
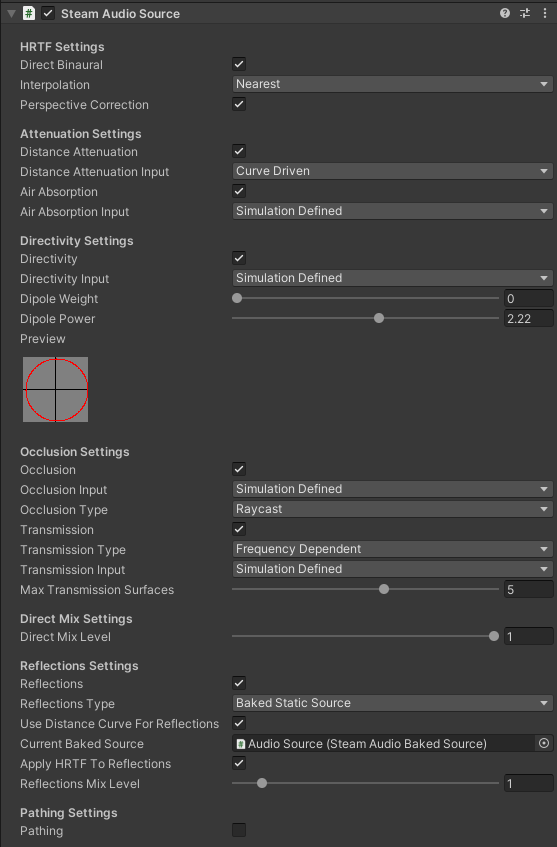
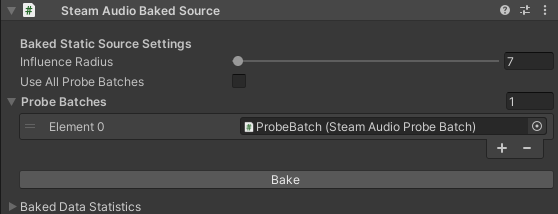
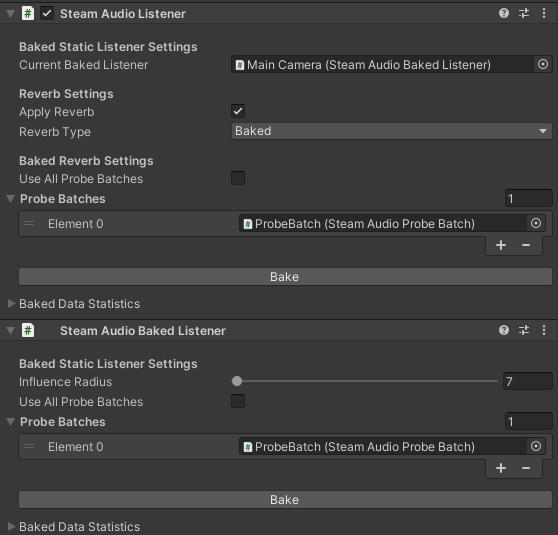


Need to fix this, either add checks to skip when not relevant

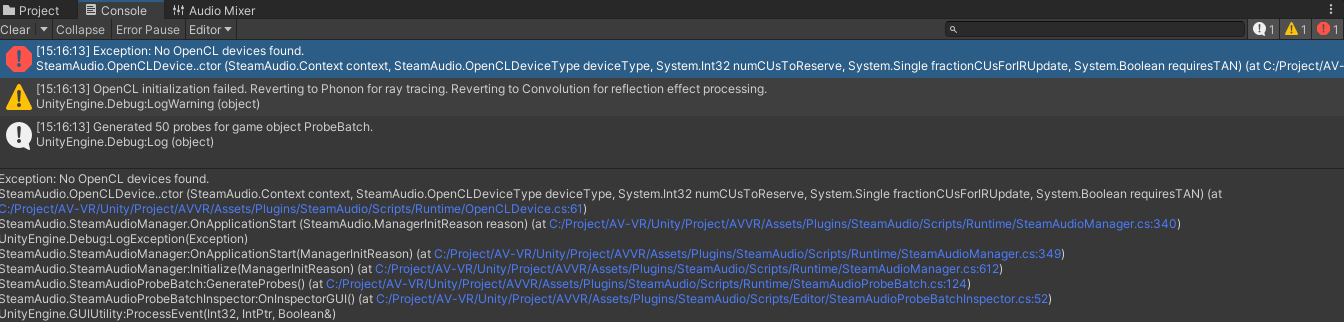
Ok added checks fixes and also for import scenery, added dynamic project dir instead of hardcoded.

Project hierarchy so far.

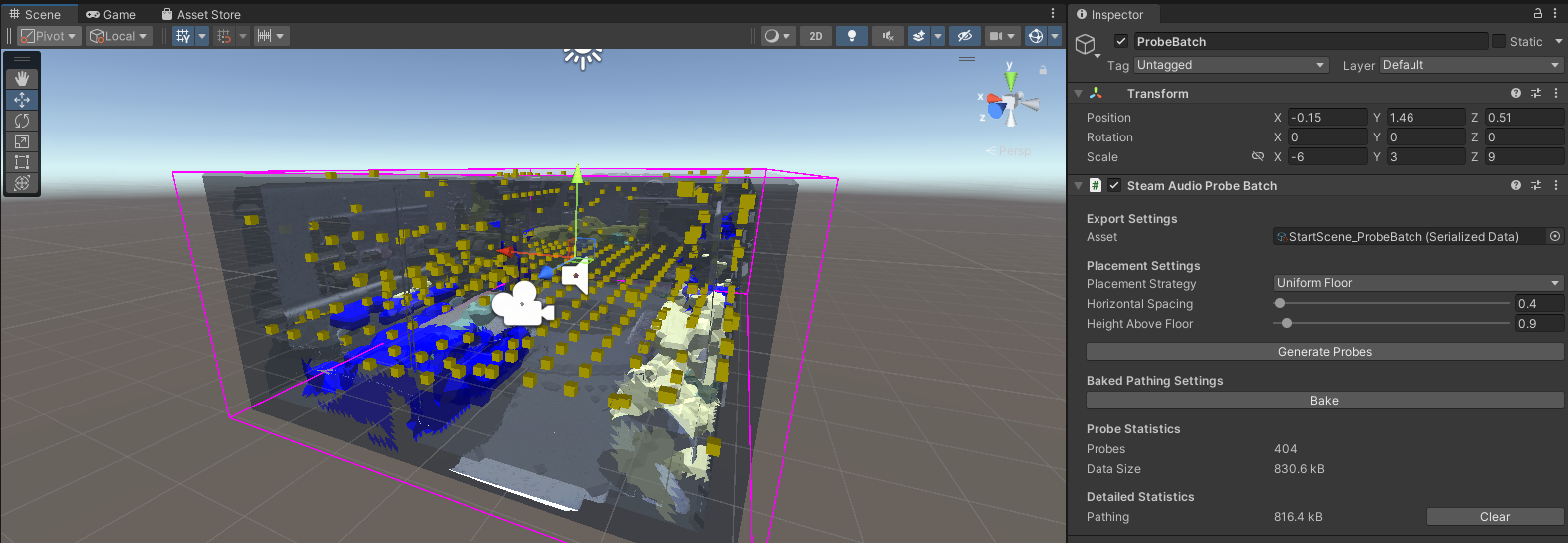
Probe Batch settings follow S3A equivalent scene (in this case, KT)

The same for all other settings except some are new so followed Mona’s suggestion. Apparently, based on my evaluation so far, Baked setting does sound clearer directivity and reverb wise compared to realtime, though that might just be because some other settings like Mix and rolloff also changes.

Remember to always export active scenes after every bake etc.

Weird bug, why is it looking at ‘AV-VR’ (old project directory) instead of ‘AVVR-Pipeline-Internship’ (new and current project directory). I’ll try reimport steam audio and see if this fixes it.

Ok, reimport hangs, so simply delete steam audio folder with its .met in plugin folder then reimport latest one fixes it.

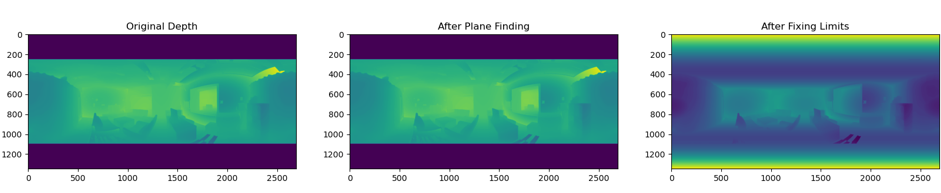


Note that for baked probe, it only have one level/layer thus why there’s height above floor for uniform floor setting, so using freecam mode is not realistic, thus why a player controller is needed with collision.

For some reason Radeon Rays mode give OpenCL error, also im unsure whether I should use TrueNext or not in reflection mode/type in Global Steam Audio setting.

Note: If baking take very little time especially on Listener, then there must be something wrong somewhere.  
There is weird occlusion+ distance attenuation bug I think which causes the sound to keep playing at normal/loud volume even at outrageous distance when leaving mesh at certain angle. Freecam related bug basically…

Gave up on SteamAudio and unity rn, I feel like its so buggy and hard to determine when its actually working or not, maybe I should have a log/debug or even ingame indicator to see if steam audio running/working as expected.

Now lets continue enhance360mono thingy, creates a visualiser script to see what each major step of enhance360.py did 

And from there, we can see that fix limits is the issue here, it remove the mirrored thing as its out of limit (far) but also remove near object (close to center), I assume this is because it thought it was part of floor. Maybe I can fix this by faking the closer object to have stronger depth or fake depth line. Or fix limit while taking into account edges(this most likely will bring back mirror problem tho).

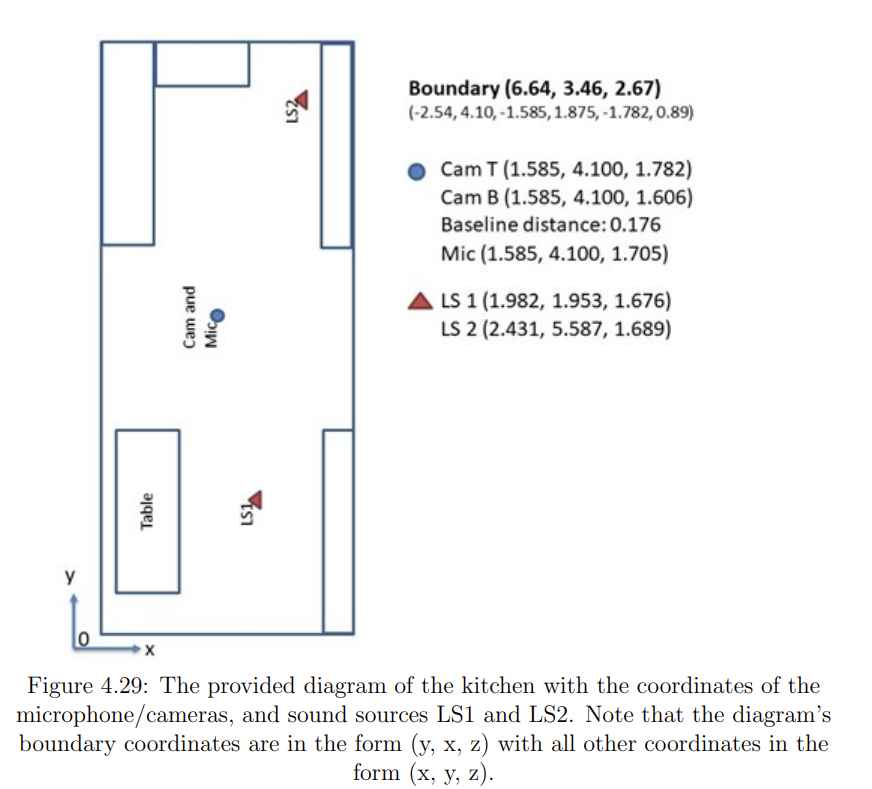
Looks promising now, fix\_limits\_mono from claude doing pretty well (the code version after I prompted patches of black/white) [Depth Map Enhancement for 360-Degree Imaging - Claude](https://claude.ai/chat/9fe50efd-9b07-4d45-95bf-39ad6a2db238)

Gonna sign off for today.

# Wednesday, 17 July 2024

Got MATLAB and python code for audio evaluation from Mona. So ig for now I should play around and get some results to compare.

Now there’s also another model with the best reproduction (stereo apparently).  
After looking back at GDP’s KT LS1/LS2 position images (note: where can I get these for other scenes? Or did they redraw it themsleve?) and old S3A unity project config in KT\_Eval

In this, LS2 is muted when checked in S3A.

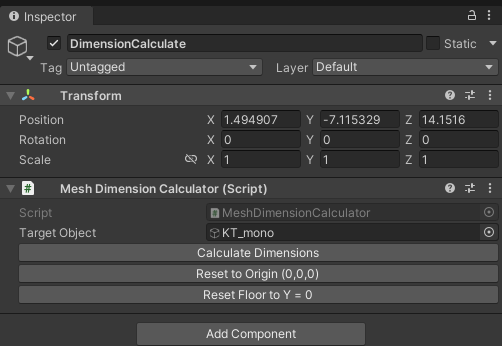
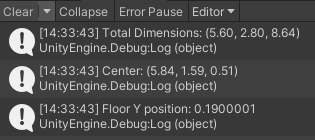
Im convinced that this is the correct position for cam/listener and 2 audio source. Another thing, because the unit is in meter, I should also just scale the boundary/scene size to follow so its easier to determine where each point should be (standardise it basically), and considering we have separate different scene saves for each thing (demo/eval/dimension comparison), might as well reset all to origin for setup at least.

**LAST TODO:**

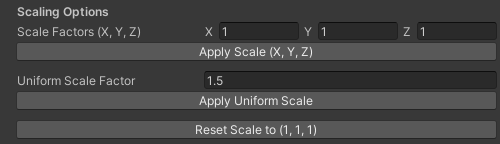
1. Figure out Steam Audio quirks and revamp Unity project
   1. Ask Mona for old unity project for reference, done and also for positions of probe/listener/source.
   2. Get MATLAB code for RT60 and EDT calculation. done
2. Continue monodepth optimisation otw, updated to use enhance360debug.py
3. Remove docker clutter
4. Remove LiDAR from LFS, and use separate cloud service to download that, or use MeshLab to optimize it.

## **UPDATED TODO:**

1. Confirm steam audio setting and standardize/clean project directory to be less convoluted and organised
   1. Ask Mona for positions of probe/listener/source as pictured above for each scene.
   2. Confirm Steam Audio setting with Mona for each scene.
   3. Reorganise project directory done
   4. Add lighting indicator for when steam audio is working
2. Add tools/script to streamline evaluation audio source/listener positioning done
3. Add Rachel and Joao for Demo for more immersive VR demo, and probably have more audio cue.
4. Continue monodepth optimisation using enhance360debug.py
5. Remove docker clutter automatically
6. Remove LiDAR from LFS, and use separate cloud service to download that, or use MeshLab to optimize it.

Created a helper scripts/gameobject to calculate dimension of model and recenter/refloor when needed. It’s pretty nice and can work with any imported via AVVR script as long as they have same hierarchy. This is important so we can center it to origin and get an idea of the ratio for positioning the probes/listener and audio source for evaluation.

For now, it shows on console logger which is a bit annoying imo cus it keeps stacking up but it work for now, better than nothing.

Also added scaling options so its easier to scale and check dimension in one menu!

Hmm I just realised the problem, idk what theboundary here means, are its using 0,0,0 as origin even? The diagram showing position I mean. Ah just realised the 0,0,0 is at the edge of kitchen bruh.

Added button options to reposition and rotate so that the model edge at 0,0,0 and extrude towards +ve x and +ve z axis. Ok this gonna be harder to code/prompt than expected. Idea so far is first to rotate to correct orientation THEN bring the corner to 0,0,0.

Nice, it works perfectly. Only problem is the dimension of monodepth is so wrong, so I need to find ratio to scale them to correct dimension first, or else I cant really use the coords provided unless I scale them first, which is more annoying so might as well find ratio first ( I need to find it anyway)

Also, changing the coord to be x,z,y because that’s how it defined in the diagram. OK NVM, this is confusing, and make it harder, lets just use normal XYZ or unity, and map the diagram to proper XYZ lmao. Y axis is height in unity, yea this is so confusing. And import sometime on wrong direction of width/length so need to visual inspect and rotate accordingly.

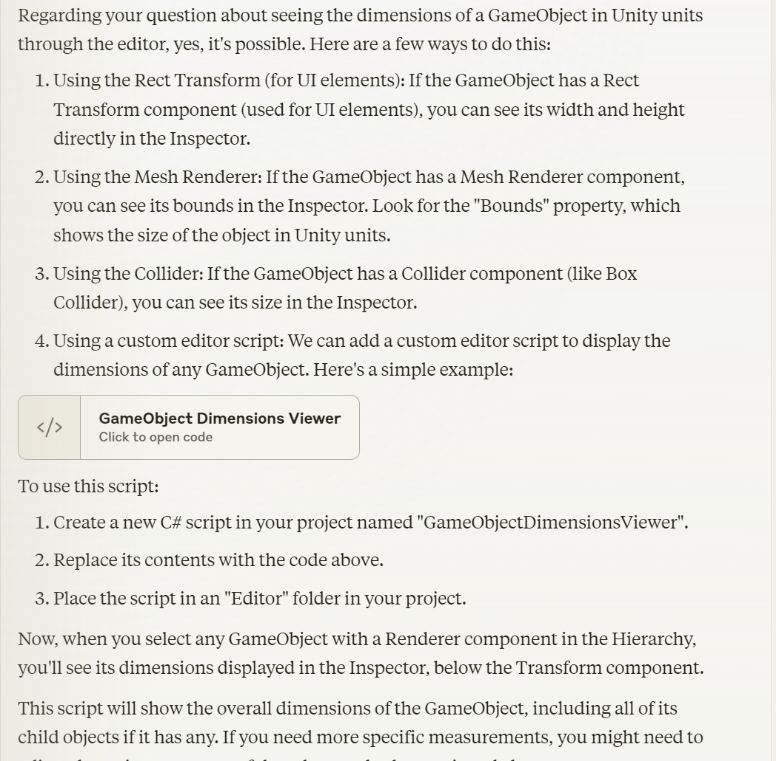
## Dimension (x, y, z) of scene meshes from original enhance360.py and 360monodepth to calculate scale ratio needed as follows:

1. KT – (4.16, 2.08, 5.44) enter scale (0.83, 1.28, 1.22) , align corner 2,
2. MR – (4.64, 1.84, 5.28), first rotate -90, enter scale (1.06, 1.27, 0.92), align corner 3
3. LR – (5.12, 2.08, 5.60), enter scale in table, align corner 0, swap X and Z for coords place.
4. ST – (6.08, 2.08, 6.40), first rotate -90, enter scale (2.67, 3.125, 2.39)
5. UL – (4.48, 1.92, 4.24), enter scale in table, align corner 0

|  |  |  |  |
| --- | --- | --- | --- |
| Scene and ratio (x, y, z) | Cam/Listener coords | Audio/Source coords | Ground Truth Dimen |
| KT (0.83, 1.28, 1.22) | (4.100, 1.705, 1.585) | (1.953, 1.676, 1.982) | (3.46, 2.67, 6.64) |
| MR (0.92, 1.27, 1.06) | (2.12, 1.00, 0.33) | (2.12, 1.00, 3.00) | (4.28, 2.33, 5.61) |
| LR (0.99, 1.39, 1.01) | (2.55, 1.08, 2.79) | (0.51, 1.20, 2.80) | (5.05, 2.90, 5.64) |
| ST (2.39, 3.125, 2.67) | (6.94, 1.50, 5.00) | (4.94, 1.50, 5.00) | (14.55, 6.50, 17.08) |
| UL (1.16, 1.52, 1.31) | (3.58, 1.07, 2.27) | (1.32, 1.07, 2.24) | (5.20, 2.91, 5.57) |



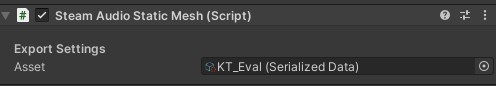
From <https://doi.org/10.1007/s10055-021-00594-3>, in Z, X, Y ( Length, Width, Height).

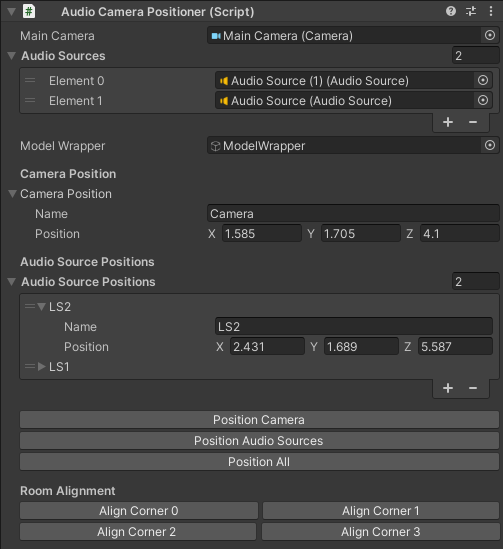
Unrelated but just realised I could just use a simpler way to get dimension of the modelwrapper, either instantiate it with trigger/disabled collider or put transparent mesh renderer over it.

This require me to modify some part of ImportScenery.cs though. Ok nvm, idk if instantiating collider or renderer through script would even inherit the whole meshes dimension, and current script works well anyway so oh well.

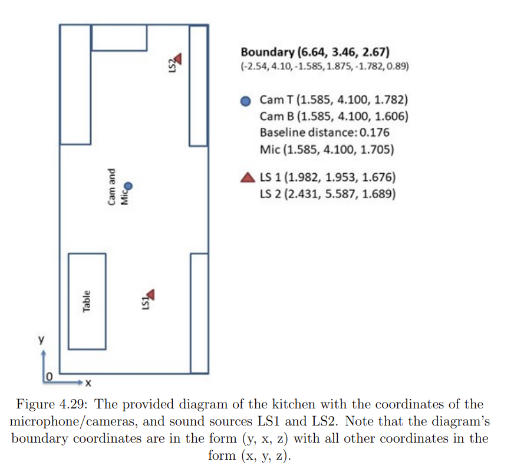
Did it anyway so can double check/confirm. And can confirm it is correct calculation for both

Found a bug with

Remember to set this to None and create new one for new scene.

YAHOOO FIXED AND IT WORKS PERFECTLY (AudioCameraPositioner editor menu helper script)

Having all corner options make it easy to rotate no matter how the diagram is made as long as it follows positive XZ quadrant and start at 0,0,0 on corner of mesh!

For ref, KT diagram given:

Done for today ngl.

# Thursday, 18 July 2024

Next, lets remind ourselves of our tasks

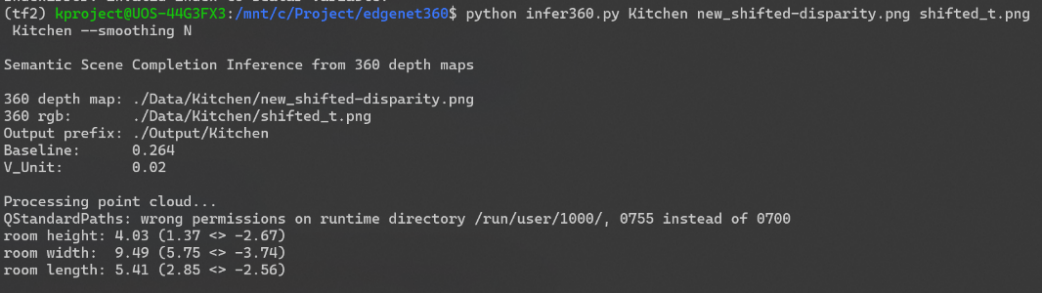
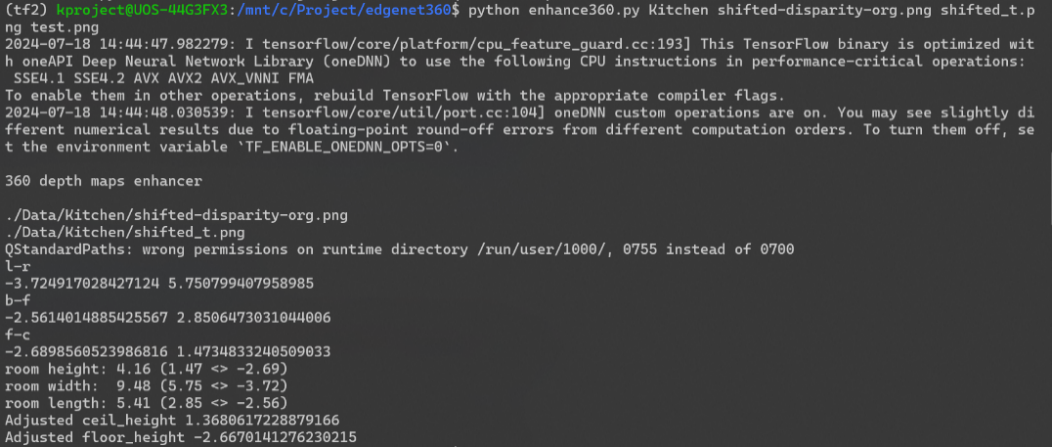
## **UPDATED TODO:**

1. Confirm steam audio setting and standardize/clean project directory to be less convoluted and organised
   1. Ask Mona for positions of probe/listener/source as pictured above for each scene. done
   2. Confirm Steam Audio setting with Mona for each scene. (done ish? For now I’ll just follow settings in S3A)
   3. Figure out the proper steps to bake/export steam audio (check git changes), and automate baking with 1 button to streamline
   4. Add lighting indicator for when steam audio is working
2. Add Rachel and Joao for Demo for more immersive VR demo , and probably have more audio cue.
3. Continue monodepth optimisation using enhance360debug.py
4. Remove docker clutter automatically
5. Remove LiDAR from LFS, and use separate cloud service to download that, or use MeshLab to optimize it.

To further organise, created 3 folder in Scenes Folder: DemoNoVR, DemoVR and Eval. Only difference of Eval and DemoVR is that in Eval, Main camera should be fixed, ie disable both XR input and KeyboardMove script. However, for DemoNoVR, all VR related thing should be removed for clarity and also add collision (also to be added for VR mode but later)

Note, we got a problem with the ModelDimension calculation of dimension, its obviously buggy with the way the yellow outline not really following the mesh (maybe it doesn’t work with voxel meshes?), because interestingly enough it was accurate with mono KT mesh before lol.

Interesting to note, if you drag and drop both .obj and .mtl into assets area, u can then just drag drop the mesh directly without use of AVVR/Import Scene. Good to know that the script doesn’t mess with scale and just help with steam audio material assignment etc.

Apparently the edgnet library especially one related to enhance and pointcloud for dimension estimation could be not updated cus that’s where we got wrong dimension. Question is where did KT21 model come from because it is non reproducible rn. (atleast for me and my setup)

Interesting dimension note (explanation by Mona: due to limitation of how edgenet works, voxel point cloud etc),its impossible to get correct mesh output dimension for ST (which is very large) so they scale it in Unity, which then ig it means that the dimensions comparison value they got are from enhance360/infer360.py CLI output instead.

Todo, add .sofa to gitignore and publish Internship-tools git repo, reinit it to remove first commit.

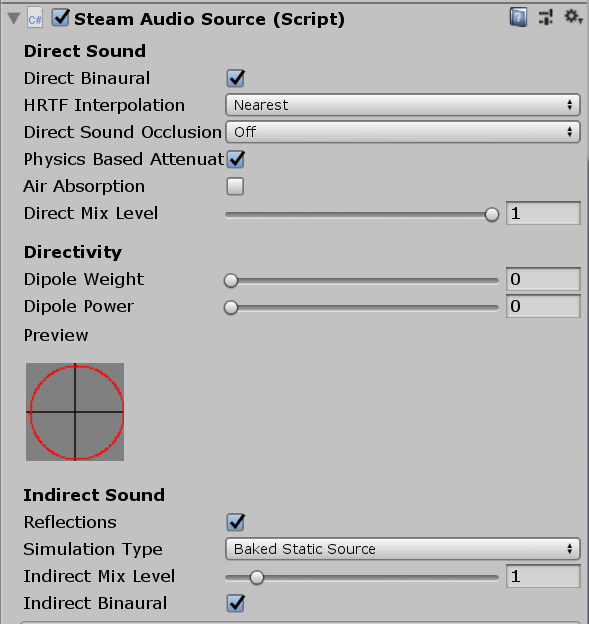
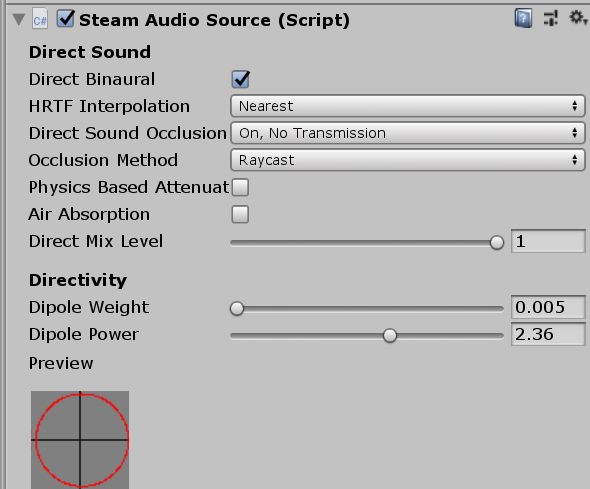
Sign off

# Friday, 19 July 2024

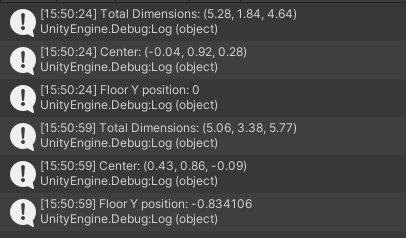
Added .sofa to gitignore and pushed successfully. Next, lets save all scenes for audio eval and generate all audio files for analysis.

## **UPDATED TODO:**

1. Confirm steam audio setting and standardize/clean project directory to be less convoluted and organised
   1. Figure out the proper steps to bake/export steam audio (check git changes), and automate baking with 1 button to streamline
   2. Add lighting indicator for when steam audio is working
2. Create unity scenes for all scenes and its category (VRDemo, nonVRDemo, Eval)
3. Add Rachel and Joao for Demo for more immersive VR demo , and probably have more audio cue. Done ish
4. Continue monodepth optimisation using enhance360debug.py
5. Remove docker clutter automatically
6. Remove LiDAR from LFS, and use separate cloud service to download that, or use MeshLab to optimize it.

Hmm, curious for some reason, MR (left) audio sources have physics based attenuation enabled while others doesn’t. Another note is in the version of Steam Audio im using, indirect binaural is HRTF as they are the same.

Ok, maybe its just kinda random and set as tried because on LR (right), its like shown lmao. I have no clue

!!! Big bug/oversight with Mesh Dimension Calculator script!!!

Because the way it is implemented to calculate room dimension based on global coord of meshes, it doesn’t really tell which way is correct ie which is width or length, so rotating it would give different value instead. Should use local coord or something instead!!! Not really critical issue if we know the orientation first and orientate before measuring!

After ruminating about it, the solution is not so simple because we need reference point so we can’t really automate this. Best way is to visually examine and confirm the features of the room (using voxel mesh when needed because it have better details and info about voxel category) to orient the room first!

I mean, its only 2 option, so if the default is wrong, then it’s the other way around.

In my defense, the provided diagram for measurement is such a mess….

Done generating and aligning, scaling 5 scenes for evaluation. Audio source also setup according to S3A settings. Gonna generate audio next week and analyse it next week as well. Can confirm export only for meshes and only need to bake afterward (and save) and export at the end only (confirmed via git tracking staging check)

Filled the dimension table for reference and pasted to position in rooms pptx

So step is export active scene -> generate probe -> bake probe -> bake audio -> bake baked listener -> bake listener -> export active scene -> save.