



# Perception in SUMO: Justification and Experimental Results

Richard Skarbez

3 December 2018

# Scene Understanding and Modeling (SUMO)

The SUMO challenge encourages the development of algorithms for complete understanding of 3D indoor scenes from 360° RGB-D panoramas with the goal of enabling social AR and VR research and experiences. The target 3D models of indoor scenes include all visible layout elements and objects complete with pose, semantic information, and texture. Algorithms submitted are evaluated at 3 levels of complexity, corresponding to 3 tracks of the challenge: oriented 3D bounding boxes, oriented 3D voxel grids, and oriented 3D meshes.

(from SUMOchallenge.org)

# Scene Understanding and Modeling (SUMO)

The SUMO challenge encourages the development of algorithms for **complete understanding** of 3D indoor scenes from 360° RGB-D panoramas with the goal of **enabling social AR and VR research and experiences**. The target 3D models of indoor scenes include all visible layout elements and objects complete with pose, semantic information, and texture. Algorithms submitted are evaluated at 3 levels of complexity, corresponding to 3 tracks of the challenge: oriented 3D bounding boxes, oriented 3D voxel grids, and oriented 3D meshes.

(from SUMOchallenge.org)

# “Complete Understanding”

- I argue that a “complete understanding” of a 3D scene is impossible unless one also considers how users/viewers will **perceive** that 3D scene
  - This informs both the evaluation and the usage of 3D scene representations

# All Errors Are Not Created Equal

- There are infinitely many ways that a scene model can be "wrong"
  - Each object can be in the wrong place, at the wrong orientation, or at the wrong scale; the model itself can contain errors; an object can be missing from the model, or the model can contain spurious objects; etc.
- Given all the possible errors that can take place in a model, how can we decide which ones matter?

# How much should a 1cm error in the position of an object be penalized? 10cm? 100cm?

# It depends.

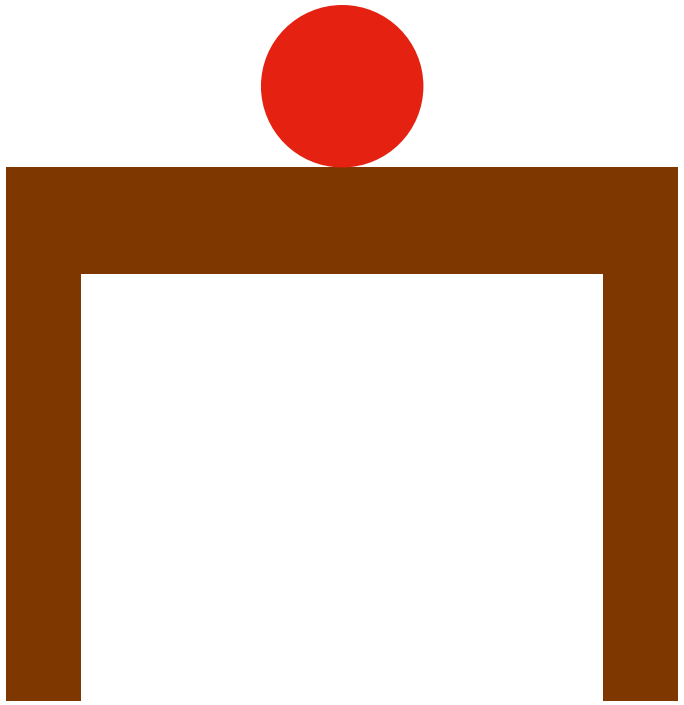
# It depends on *what*?

- It depends, or rather, it **should** depend, on how **noticeable** the error is
- Some factors that can affect noticeability:
  - How big is the object? How big is the scene?
  - How far is the object from the viewer?
  - How semantically important or perceptually salient is the object?

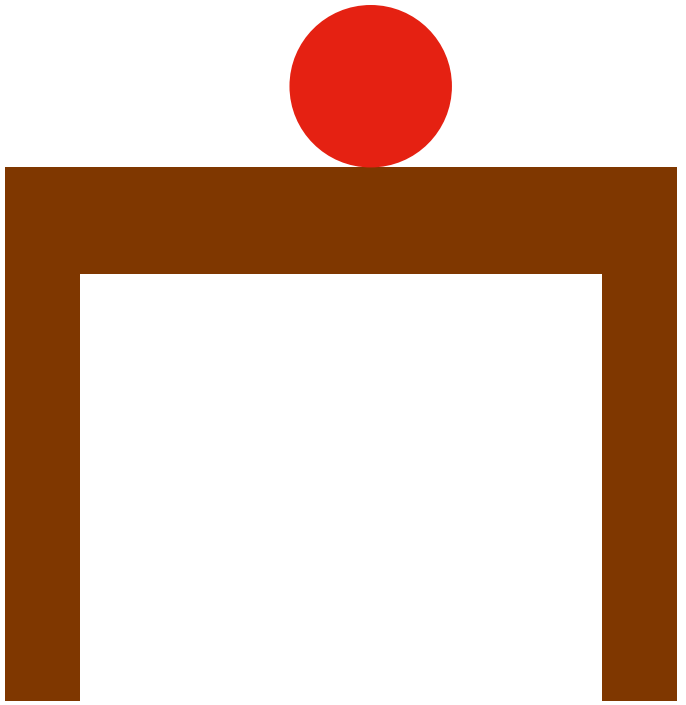
# Coherence

- The **coherence** of a scenario (i.e., a virtual environment) can be roughly described as, “How plausible is the story being told in this scenario?”
  - A more coherent system has fewer glitches/irregularities/ unpredictable behaviors; it is more internally consistent
- Some errors introduce incoherence to the scene, and some do not

# Coherent Model Error



Ground Truth
(Ball sits on table)

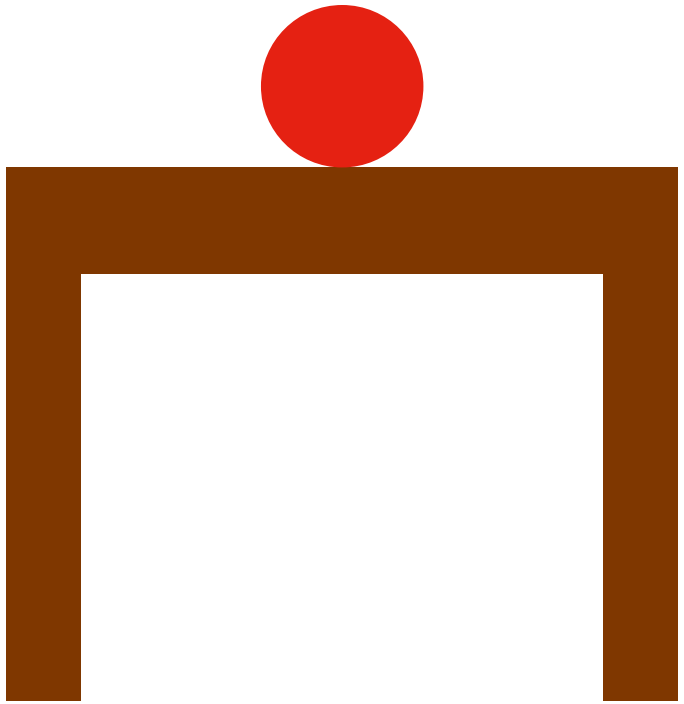


Model contains error
(Ball offset in x direction)

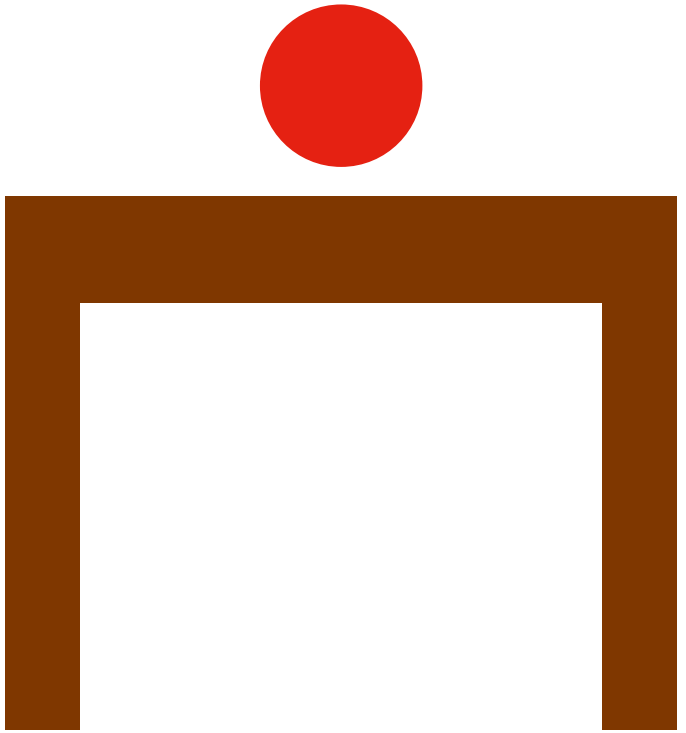
Model is **still coherent**

It may not correspond exactly to the ground truth, but it still makes sense

# Incoherent Model Error



Ground Truth
(Ball sits on table)



Model contains error
(Ball offset in y direction)

Model is **not coherent**

The scale of the difference from the ground truth is the same, but this error is **much worse** from a user's perspective

# Evaluating Scene Understanding

- If our goal is “complete understanding of [scenes]…with the goal of enabling social AR and VR research and experiences,” then we **must** consider user perception (and specifically coherence) as a key part of evaluation
  - A model in which every single object is misplaced but in which coherence is preserved will likely result in a better experience than one in which every single object is correctly placed – but for one which is incoherent

# Evaluating Scene Understanding

- It is clear that if we intend to **use** 3D scene models that we need to understand how users will perceive those scenes
  - If we want to **evaluate** 3D scene models intended for use, we need to build a model that incorporates perception
    - That is, we need to know how different types of errors are perceived and experienced by users

# User Studies

# User Studies Roadmap

- To that end, we (myself and colleagues at Virginia Tech) designed and conducted a series of user studies in order to inform such a model
  - Study 1: Identification of parameters
  - Study 2: Valuation of parameters from Study 1
  - Study 3: Budget-based comparison of parameters

# Pre-Study 1: Real Rooms

- Prior to running these studies, we also prepared several real rooms that could be used both as the basis for 3D scene models, as well as for the literal “ground truth” against which these models could be compared
  - These rooms were then 3D scanned, and these 3D models were then manually edited using Maya to create “perfect” 3D scenes corresponding to the original rooms

# Our Living Room: Real vs. Virtual


Real Room

Scanned + modeled room
(Rendered in Unity)

# Our Living Room: Real vs. Virtual


Real Room

Scanned + modeled room
(Rendered in Unity)

# Our Living Room: Real vs. Virtual

Real Room

Scanned + modeled room
(Rendered in Unity)

# Our Living Room: Real vs. Virtual

Real Room

Scanned + modeled room
(Rendered in Unity)

# Our Living Room: Real vs. Virtual

Real Room


Scanned + modeled room
(Rendered in Unity)

# Pre-Study-1: Parameter generation

- Before running any of the studies, we generated a list of potential types of errors that could appear when creating 3D models of indoor scenes, inspired by the following:

| Room Layout | Furniture | Clutter | Lighting |
|---|---|---|---|
| • Scale (dimensions)<br>• Shape<br>• Discontinuities<br>• Locations of doors & windows | • Existence<br>• Dimensions<br>• Position/orientation<br>• Materials/color | • Existence<br>• Dimensions<br>• Materials & physics<br>• Position/orientation | • Position & Size<br>• Luminance<br>• Type<br>• Whitepoint |

# Study 1

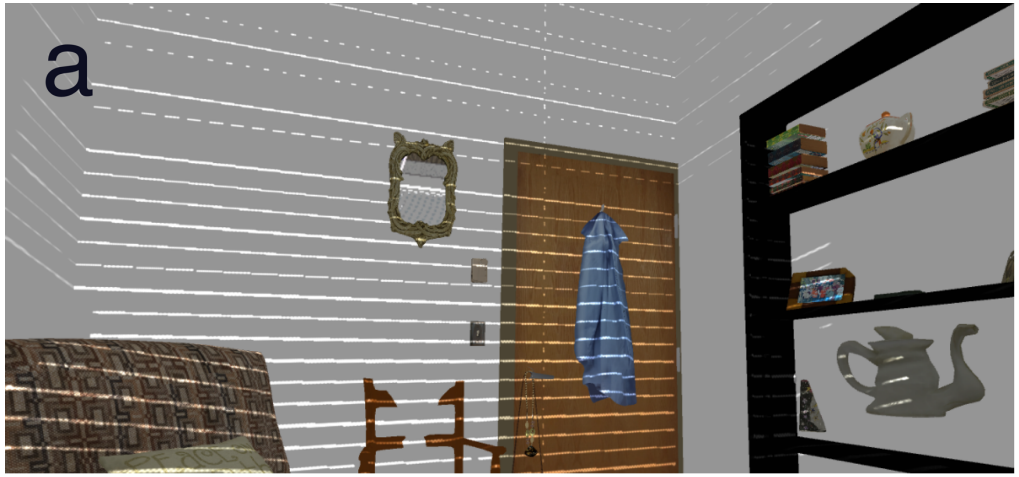
# Error Types Considered in Study 1

- Room Layout:
  - Ceiling height wrong
  - Room width wrong
  - Room length wrong
  - Door missing
  - Window missing
- Furniture
  - Missing
  - Mismatched (different model)
  - Quality reduced (using poly count as proxy)
  - Wrong position
  - Wrong scale
- “Clutter”/Small objects
  - Missing
  - Quality reduced (using poly count as proxy)
  - Wrong position
  - Wrong scale
- Lighting
  - Missing/off
  - Brightness changed
  - Hue changed

# Study 1: Modified “Error Rooms”

- Using the list of errors on the previous slide, we edited the “perfect” 3D model of the living room to create a set of 10 “error rooms”, each of which demonstrated several of the errors from that list

# Study 1: “Error Room” Examples






# Study 1: Participants & Procedure

- 6 participants each experienced each of the 10 error rooms, interleaved with experiences of the perfect room
  - In this study, participants did not experience the real ground truth room
- Participants were asked to think aloud, and comment on anything that they noticed in the error rooms

# Study 1: Results

- Some errors (such as missing furniture, or light hue) were always noticed almost immediately by all participants
  - We decided that these errors would be important components of the evaluation model, but did not need to be further investigated in Studies 2 and 3

# Study 1: Results

- Some errors (such as furniture quality) were almost never noticed by any participant
  - We decided that this meant that these errors are not perceptually relevant, and that they neither merited further investigation nor would be included in the final evaluation model

# Study 1: Results

- And finally, there were some errors that were noticed by participants some of the time
  - These are the errors that were brought forward for Studies 2 and 3
- These were:
  - Room length
  - Room width
  - Furniture elevation
  - Furniture scale
  - Clutter missing
  - Clutter elevation
  - Lights missing/off

# Study 2

# Study 2: Purpose

- Identify points of subjective equivalence between the errors identified for further study in Study 1
  - Necessary to establish the relative costs of each parameter for Study 3
- In the interests of time, I will not fully discuss the design and results from Study 2
  - Instead, we'll skip to the budget-based Study 3

# Study 3

# Study 3: Purpose

- Generate a rank ordering of the studied parameters, in terms of how important they are to participants
- Generate a measure of “how correct” each parameter needed to be in order to satisfy participants

# Study 3: Participants

- Participants were 40 university students (19 female)
- Each participant experienced a short training session, followed by 7 experimental trials
- The study lasted approximately one hour

# Study 3: Parameters

- Following Study 2, we made some changes to the parameter list
  - Room width and room length were combined into a single parameter, Room Scale
  - Lighting was split into three parameters, Sun light, Ceiling lights, and Lamp lights

# Study 3: Parameters

| Parameter | Initial Value | Max Value | Increment | Cost (in points/increment) | Max Cost |
|---|---|---|---|---|---|
| Room Scale | 0.5x | 1x | 0.01x | 1.5 | 75 |
| Furniture Elevation | -25cm | 0cm | 1cm | 2 | 50 |
| Furniture Scale | 0.5x | 1x | 0.01x | 2 | 100 |
| # of Small Objects | 0 | 72 | 1 | 0.5 | 36 |
| Small Objects Elevation | -25cm | 0cm | 1cm | 1 | 25 |
| Lamp Lights | 0 (off) | 1 (on) | 1 | 10 | 10 |
| Sun Light | 0 (off) | 1 (on) | 1 | 10 | 10 |
| Ceiling Lights | 0 (off) | 1 (on) | 1 | 10 | 10 |
| | | | | **Cost for all upgrades:** | **316** |

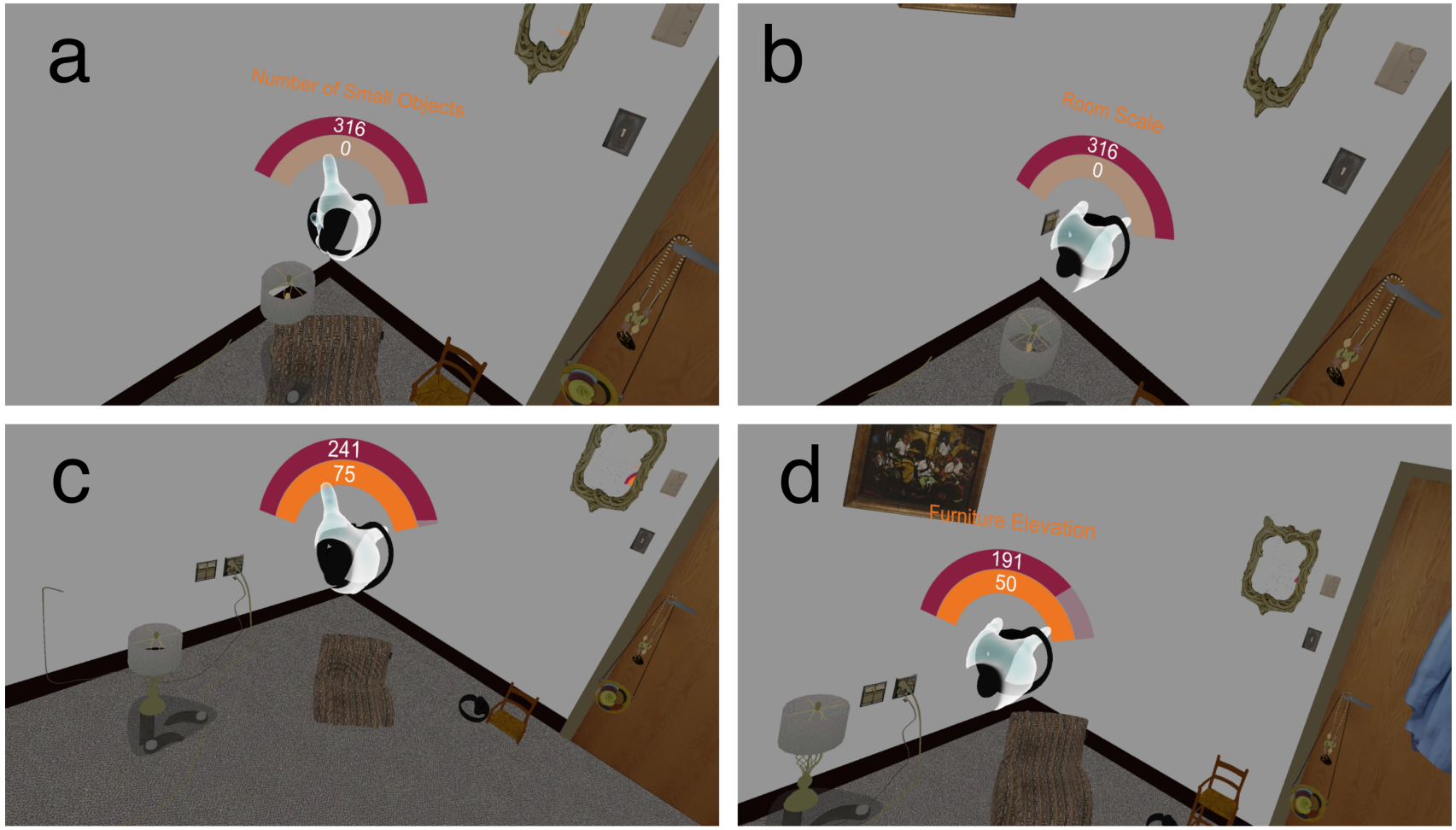
# Study 3: Procedure

- In each trial, the environment started with most or all of the parameters at the minimum level, and the participant was given a points budget to improve the environment
- The participant could upgrade parameters to whatever amount and in whatever order they saw fit
  - No backtracking was allowed – The participant could explore the effects of multiple parameters, but once a parameter was confirmed to be modified, it could not be revisited

# Study 3: Starting Configurations

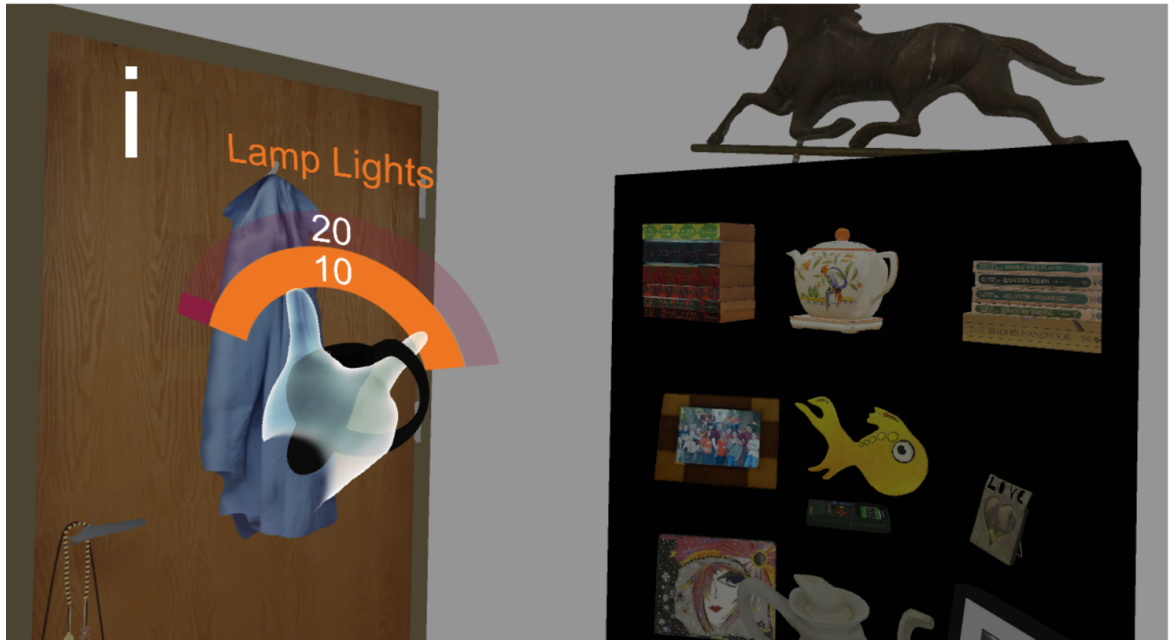
| Starting Configuration | Improvements pre-assigned | Points available to participant |
|---|---|---|
| {0, 0, 0, 0, 0, 0, 0, 0} | None | 250 |
| {1, 0, 0, 0, 0, 0, 0, 0} | Room Scale (75) | 175 |
| {0, 1, 0, 0, 0, 0, 0, 0} | Furniture Elevation (50) | 200 |
| {0, 0, 1, 0, 0, 0, 0, 0} | Furniture Scale (100) | 150 |
| {0, 0, 0, 1, 0, 0, 0, 0} | Small Objects (36) | 214 |
| {0, 0, 0, 0, 1, 0, 0, 0} | Small Objects Elevation (25) | 225 |
| {0, 0, 0, 0, 0, 1, 1, 1} | All lights turned on (30) | 220 |

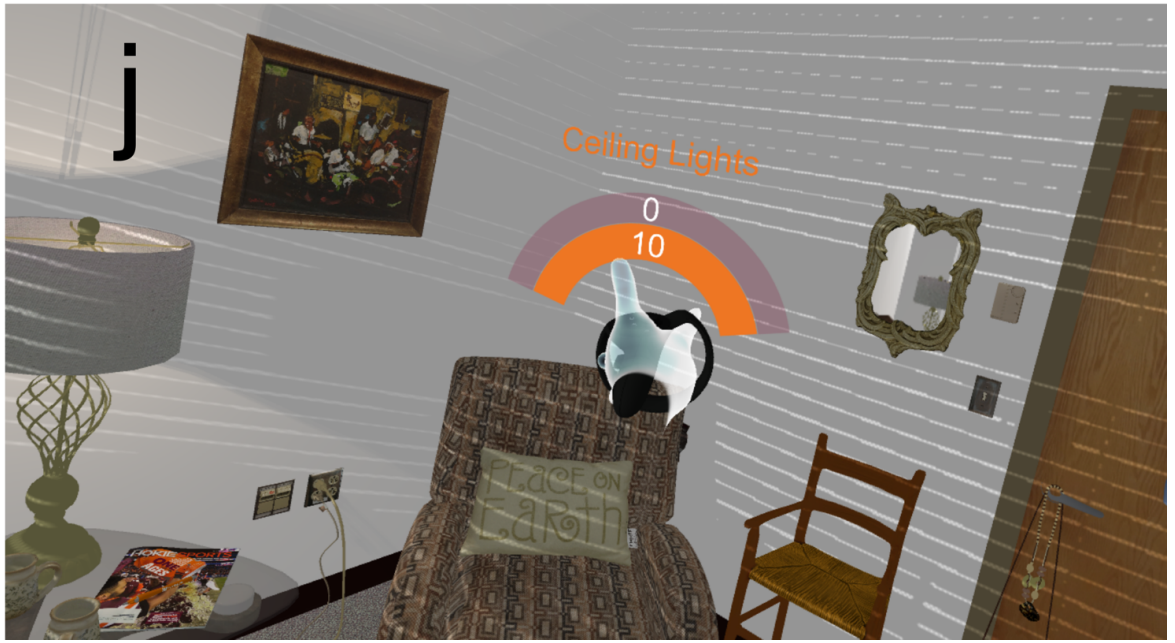
# Study 3: Sample



# Study 3: Sample

# Study 3: Sample



# Study 3: Measures

- There were three types of dependent variable
  - For each trial, we recorded the sequence in which the participant chose to improve room parameters
  - For each trial, we recorded the amount of their budget that the participant spent on each upgrade
  - At the end of the study, each participant filled out a questionnaire indicating the order in which they valued the parameters

# Study 3 Results: Most Likely Transitions

| Configuration | Most likely next configuration (%) | Most likely action |
|---|---|---|
| {0, 0, 0, 0, 0, 0, 0, 0} | {1, 0, 0, 0, 0, 0, 0, 0} (75%) | Improve Room Scale |
| {1, 0, 0, 0, 0, 0, 0, 0} | {1, 0, 1, 0, 0, 0, 0, 0} (55.7%) | Improve Furniture Scale |
| {1, 0, 1, 0, 0, 0, 0, 0} | {1, 1, 1, 0, 0, 0, 0, 0} (85.9%) | Improve Furniture Elevation |
| {1, 1, 1, 0, 0, 0, 0, 0} | {1, 1, 1, 1, 0, 0, 0, 0} (76.4%) | Improve # of Small Objects |
| {1, 1, 1, 1, 0, 0, 0, 0} | {1, 1, 1, 1, 1, 0, 0, 0} (87.5%) | Improve Small Objects Elevation |
| {1, 1, 1, 1, 1, 0, 0, 0} | {1, 1, 1, 1, 1, 1, 0, 0} (43.1%) | Turn on Lamp Light |
| {1, 1, 1, 1, 1, 1, 0, 0} | {1, 1, 1, 1, 1, 1, 0, 1} (50%)<br>{1, 1, 1, 1, 1, 1, 1, 0} (50%) | Turn on remaining lights in either order |

# Study 3: Summary Stats

| Parameter | *Expenditure* Mean | *Expenditure* Median | *Expenditure* Std. Dev. | *Final Value* Mean | *Final Value* Median | *Final Value* Std. Dev. |
|---|---|---|---|---|---|---|
| Room Scale | 61.2 | 63 | 11.4 | 0.908 | 0.92 | 0.076 |
| Furniture Elevation | 39.7 | 43 | 11.2 | -0.051 | -0.035 | 0.056 |
| Furniture Scale | 69.3 | 70 | 18.3 | 0.847 | 0.85 | 0.091 |
| # of Small Objects | 26.5 | 28.75 | 9.58 | 53.0 | 57.5 | 19.2 |
| Small Objects Elevation | 20.3 | 25 | 7.93 | -0.047 | 0 | 0.079 |
| Lamp Lights | 4.02 | 0 | 4.92 | 0.402 | 0 | 0.492 |
| Sun Light | 6.65 | 10 | 4.73 | 0.665 | 1 | 0.473 |
| Ceiling Lights | 4.31 | 0 | 4.97 | 0.431 | 0 | 0.497 |

# Perceptual Scoring Discussion

# Penalized Errors

- In the current version of the perceptual scoring system, the following errors are penalized:
  - Room Scale (area of “floor” elements)
  - Object Scale (bounding volume)
  - Object position (translation)
  - Object position (elevation)
  - Missing objects
  - Extra objects

# Gaussian Scoring

- Most parameters are weighted according to a Gaussian function whose parameters are based on the results of Studies 2 and 3

$$G(x, A, \mu, \sigma) = Ae^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

# Gaussian Scoring

$$G(x, A, \mu, \sigma) = Ae^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- $A$ : How important a given parameter is
- $\mu$ : Where the center of a given parameter is
  - For example, objects being higher than they should be is generally worse than objects being lower than they should be, so the center of the Gaussian could be negative
- $\sigma$ : How much tolerance is in the given parameter
  - For example, object elevation has very narrow tolerance, while object scale is relatively wide

# Object Size Matters

- There is an additional weight accounting for object size (volume)
  - One could argue that this should be visual angle subtended, rather than volume, but in a VR scene, there is no way to predetermine which viewpoint(s) need to be accounted for
- This is based on the observation in Studies 2 and 3 that the room shell and furniture were the most important objects

# Relative and Absolute Scale

- Object scale is checked both absolutely (against the object scale in the ground truth data) and relatively (scaled by the overall room scale)
- In VR, both are relevant – coherence demands that objects be appropriately scaled to the room, but also that all objects are appropriately scaled relative to the human viewer

# Elevation Matters More than Translation

- The Gaussians used to score an object's position are much wider for translation than for elevation
  - That is, even small errors in object elevation are penalized, while some errors in object translation are allowed with minimal or no penalty

# Potential Future Modifications

- Penalty for bounding box orientation errors
  - Need to determine what “object front” means in the general case, especially for objects with rotational symmetry, etc.
- Penalize object scale per axis rather than by volume
  - Avoids potential cases where the object shape is significantly deformed, but the overall volume is the same

# Potential Future Modifications

- Add explicit object relationships to ground truth data and scoring system
  - I.e., the painting is ON the wall, the cup is ABOVE the table, the fruit is IN the bowl
  - Generalizes the idea of penalizing incoherence-inducing errors
    - More precise than just penalizing elevation more than translation

# Conclusion

- Not all errors are equally bad
  - Need to consider how the scene is actually perceived by users
- Ran user studies to rank/cost errors
  - Can use same methodology to evaluate more/different errors
- User studies informed SUMO perceptual scoring metric

LA TROBE UNIVERSITY

CENTER FOR HUMAN-COMPUTER INTERACTION
VIRGINIA TECH

# Thank you