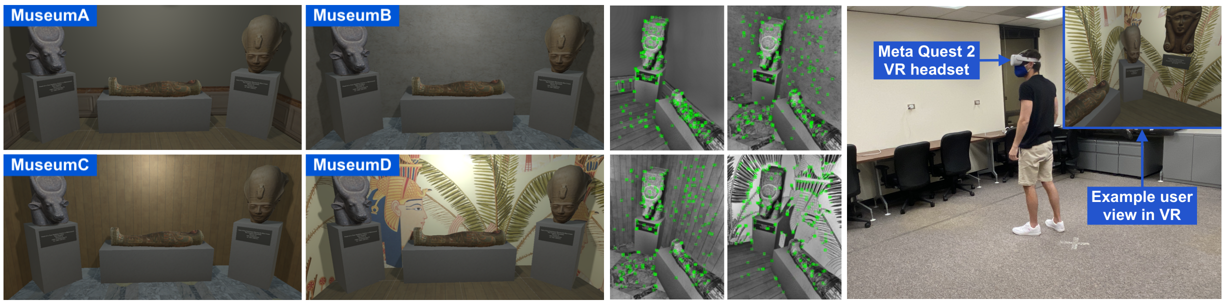
Integrated Design of Augmented Reality Spaces Using Virtual Environments

Tim Scargill*
Department of Electrical and Computer Engineering
Duke University

Ying Chen†
Department of Electrical and Computer Engineering
Duke University

Nathan Marzen‡
Ackland Art Museum
UNC Chapel Hill

Maria Gorlatova§
Department of Electrical and Computer Engineering
Duke University



(a) (b) (c)

Figure 1: A visual overview of our case study, demonstrating our integrated design methodology for augmented reality spaces. Four variations of a museum environment were produced in the Unity game engine (a). VI-SLAM pose tracking performance in these environments was tested using ORB-SLAM3 [6], with different wall and floor textures resulting in different feature point distributions (b). User satisfaction with these environments was evaluated using a VR headset, the Meta Quest 2 [39] (c).

Abstract

Demand is growing for markerless augmented reality (AR) experiences, but designers of the real-world spaces that host them still have to rely on inexact, qualitative guidelines on the visual environment to try and facilitate accurate pose tracking. Furthermore, the need for visual texture to support markerless AR is often at odds with human aesthetic preferences, and understanding how to balance these competing requirements is challenging due to the siloed nature of the relevant research areas. To address this, we present an integrated design methodology for AR spaces, that incorporates both tracking and human factors into the design process. On the tracking side, we develop the first VI-SLAM evaluation technique that combines the flexibility and control of virtual environments with real inertial data. We use it to perform systematic, quantitative experiments on the effect of visual texture on pose estimation accuracy; through 2000 trials in 20 environments, we reveal the impact of both texture complexity and edge strength. On the human side, we show how virtual reality (VR) can be used to evaluate user satisfaction with environments, and highlight how this can be tailored to AR research and use cases. Finally, we demonstrate our integrated design methodology with a case study on AR museum design, in which we conduct both VI-SLAM evaluations and a VR-based user study of four different museum environments.

Index Terms: Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Mixed / augmented reality;

*e-mail: ts352@duke.edu
†e-mail: ying.chen151@duke.edu
‡e-mail: nmarzen@email.unc.edu
§e-mail: maria.gorlatova@duke.edu

1 Introduction

Markerless augmented reality (AR), supported by visual-inertial simultaneous localization and mapping (VI-SLAM), holds great potential thanks to the convenience and flexibility it provides users, and is already commonplace in commercial apps (e.g., [27, 43]). However, this technique places requirements on the spaces that host AR; their visual properties must be conducive to vision-based pose tracking to minimize spatial registration errors. A key challenge is that the requirements of VI-SLAM-based AR systems, in particular the need for recognizable textures, are frequently at odds with human preferences related to comfort and aesthetics. For example, the enduring appeal of minimalism in architecture and interior design has led to indoor visual environments that commonly contain large blank regions, which often result in pose tracking errors [29]. How do we design spaces that work for both AR systems and AR users?

This concept of considering multiple distinct factors in the design process is known as *integrated design*. A common example is building design that takes into account aesthetics, construction methods, energy efficiency, sustainability, accessibility and cost. One specific case of conflicting requirements, analogous to ours, is a building occupant's preference for large windows providing abundant natural light [25], and the greater energy consumption that results from this design choice [63]. In this work we develop a technique that brings together the previously disparate fields of VI-SLAM performance evaluation (e.g., [5,29,54]) and occupant comfort research in immersive virtual environments [1], to better inform the design of AR host spaces — towards a future in which holistic architectural and interior design processes incorporate AR-specific factors as standard.

VI-SLAM performance evaluations are traditionally conducted using established benchmark datasets (e.g., EuRoC [5], TUM VI [54], SenseTime [29]), through comparisons with accurate ground truth pose, obtained through optical tracking systems such as OptiTrack [44] and Vicon [68]. However, due to the practicalities of obtaining accurate ground truth pose data, the visual environments these benchmarks cover are only representative of a tiny fraction of the environments in which AR is designed to be used. While methods have been developed that enable direct measurement of spatial registration errors in a wider range of environments [53, 59], controlling for all possible variables is a near-impossible task. We require an evaluation method which facilitates systematic control of environment conditions, to obtain a quantitative understanding of the relationship between visual environment properties and spatial registration errors, and better inform the design of spaces that host AR.

Critically though, the magnitude of spatial registration errors is not the only way in which the visual properties of a space affect the quality of an AR user's experience. Research from the field of environmental psychology has demonstrated how light [11, 19, 37, 57], color [33,38,47] and visual texture [47,71,73] in an environment, as well as its layout [22,45,72,74], can affect an occupant's comfort, behavior and task performance. These visual properties are a subset of a wider group of factors (including air quality, acoustics and temperature) that determine indoor environment quality (IEQ), and are sometimes referred to as atmospherics [31], particularly in the context of consumer behavior. Visual conditions have been shown to impact other aspects of IEQ, such as thermal comfort perception [51]. Specific to AR, user perception of virtual objects is affected by the properties of the real world background and surroundings [10, 15]. Our evaluation methodology must assess these direct effects of the environment on an AR user, as well as the indirect effects that arise from the pose tracking quality of an AR system.

In this paper we present a methodology that supports the evaluation of both AR device tracking performance, and AR user comfort and perception, in diverse visual environments. *To the best of our knowledge our VI-SLAM evaluation technique is the first to enable accurate measurement of pose tracking performance in virtual environments with real inertial data*. We show how the same virtual environments can be used to conduct virtual reality (VR)-based evaluations of user comfort, towards *our proposed integrated design process that considers both tracking and human factors in the design of AR spaces*. We demonstrate this with a case study on AR museum design, which reveals important trade-offs related to visual environment properties, with the user study illustrating further motivation for the use of virtual environments. While the focus of this work is on simulation of the real-world environment, our methodology also supports the simulation of virtual content, and user interactions with that content, which we will investigate in future work. Our key contributions are summarized as follows:

- We present the first VI-SLAM evaluation method that combines virtual environments with real inertial data, enabling accurate pose tracking performance analysis of handheld or head-mounted AR devices under controlled visual conditions. We release the code for our solution on GitHub as a publicly available tool.

- We conduct the first systematic experiments (2000 trials) on the impact of visual texture on VI-SLAM performance, including quantitative analysis of texture complexity and edge strength. While in general greater edge strength improves performance, sufficient complexity is also key: median relative error on TUM VI room5 was 22.6cm on a low-complexity texture, compared to 3.6cm on a high-complexity texture with lower edge strength.

- We demonstrate how our VI-SLAM evaluation method can be combined with VR-based evaluations of user satisfaction in different visual environments, through a case study on AR museum design (Figure 1). We design and implement a user study, and assess participant responses alongside tracking performance, highlighting the advantages of our integrated design methodology.

The rest of the paper is organized as follows. In Section 2 we cover related work, then in Section 3 present our VI-SLAM evaluation technique, and perform systematic experiments to examine the impact of visual texture on pose error. We present our method of environment occupant comfort evaluation in Section 4, our case study on integrated design of an AR museum in Section 5, and conclusions and future work in Section 6. The code required for implementation of our VI-SLAM evaluation method is publicly available at `https://github.com/timscargill/Virtual-Inertial-SLAM/`.

2 Related Work

VI-SLAM environment characterization: We build on existing VI-SLAM benchmarks [5, 9, 29, 30, 54, 77], some of which provide qualitative descriptions of visual conditions; for example less challenging scenes in EuRoC [5] are labeled 'good texture, bright scene', and more challenging ones 'dark scene', while in SenseTime [29] the authors note that the two most challenging sequences contain 'a glossy wooden floor' and 'a white board and repetitive textured carpets'. However, no existing VI-SLAM benchmarks quantify visual environment properties. There are works which predict pose estimation confidence from the distribution of 2D feature point matches (in part determined by visual texture) [13], study the impact of camera motion (which can affect visual texture through motion blur) [76], and characterize the difficulty of SLAM trajectories [50], but these are for visual-only SLAM. Perhaps closest to our work is [17], which quantifies visual properties (including texture) of visual-only SLAM sequences in forest environments, but does not quantify pose error.

SLAM and AR evaluations using virtual environments: Game engine-based simulations in virtual environments have previously been used as inputs for visual SLAM [3, 16, 46, 58], and drone and autonomous vehicle simulators are available that output visual and inertial sensor data (e.g., [56]). In [50] the authors present a visual-only SLAM dataset in virtual environments that uses realistic handheld and head-mounted trajectories, recorded using a motion capture system. However, to the best of our knowledge our evaluation methodology is the first to support VI-SLAM simulations in virtual environments with real inertial data, along with motion patterns typical of handheld or head-mounted AR devices. Virtual environments have also been employed to evaluate AR user interfaces using VR [34]. We view this work as complementary to our own, in that testing of user interfaces, how users perceive them in different environments, and their effect on user states such as cognitive load, may be incorporated into the VR-based user evaluations of virtual environments we describe in Section 4.

Occupant comfort research using virtual environments: Immersive virtual environments are an established research tool in psychology [35], and over the last decade they have increasingly been used to study occupant comfort and behavior in building environments – for a recent review see [1]. Previous work in this field has reported a high level of perceptual accuracy in virtual environments presented using headset VR compared to physical environments [7, 25]. The vast majority of existing works on user visual comfort in virtual environments have focused on the effects of lighting (e.g., [23, 37]); to the best of our knowledge we are the first to examine the effect of environment textures on occupant visual comfort. In our user evaluations of virtual environments we draw inspiration from [25], which found varying occupant satisfaction with different window sizes in headset VR; we use their experiment design to guide the implementation of our user study (Section 5.2.3).

3 VI-SLAM Evaluations in Virtual Environments

3.1 Motivation

Our motivation for the use of virtual environments for VI-SLAM evaluations is twofold. Firstly, it enables testing in a much wider variety of visual conditions than is currently practical in physical environments, given the logistics associated with setup and calibration of the multi-camera optical tracking system (e.g., OptiTrack [44], Vicon [68]) required for accurate ground truth pose data. This is important because AR is designed to be used in a diverse range of settings with highly heterogeneous visual conditions, from warehouses and operating theaters to classrooms and parks. Secondly,

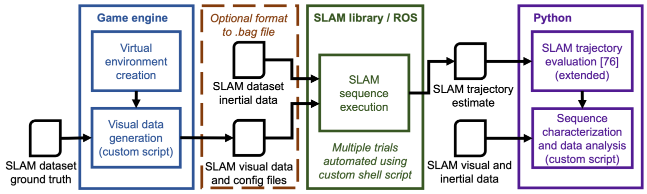


Figure 2: An overview of our VI-SLAM evaluation methodology, using virtual environments with real inertial data, and incorporating environment characterization.

virtual environments facilitate systematic control of environment properties, including light level, visual textures, and room size. This will allow us to obtain a quantitative understanding of the relationship between visual conditions and VI-SLAM performance. Unlike existing VI-SLAM simulation solutions, our technique uses real inertial data, better capturing the challenges of the noisy accelerometer and gyroscope signals available onboard an AR device.

Furthermore, we envision three practical ways in which our VI-SLAM evaluation technique, and the wider integrated design methodology it supports, may be employed. Designers of spaces that host AR, along with AR app designers creating an experience for a specific setting, can use it to inform lighting and visual texture choices in that space, and to better understand the level of pose tracking accuracy a space is likely to support. Researchers can use it to generate sufficient training data for machine learning models that predict VI-SLAM performance from visual and inertial input data, towards pose error prediction models which could be incorporated into real AR systems. Finally, developers of VI-SLAM algorithms have the opportunity to test against a new set of benchmarks, which can also be readily evaluated from human perspectives; we see the potential for our methodology to not only create benchmarks more representative of AR environments, but to introduce more human-centered design into VI-SLAM algorithm development – tracking algorithms designed with AR users in mind, not as an afterthought.

3.2 VI-SLAM Evaluation Methodology

Our VI-SLAM evaluation methodology is based on leveraging the ground truth pose data in existing datasets to generate new camera images, which we can then use with the original inertial data; our solution will work with any dataset that has inertial data and complete ground truth pose data, including existing VI-SLAM datasets (e.g., TUM VI [54], SenseTime [29]), and inertial odometry datasets (e.g., OxIOD [8]). An overview of our solution is shown in Figure 2, and the code required to implement it is publicly available at `https://github.com/timscargill/Virtual-Inertial-SLAM/`. We start by creating a virtual environment in a game engine such as Unity [66] or Unreal Engine [12]. These software facilitate the creation of highly realistic environments with accurate lighting. Given that there are no particular frame rate requirements in this rendering stage, quality settings can usually be maximized, even on machines with fewer computational resources. We use the option to define 'physical' cameras, with specific field of views and sensor sizes, to model the sensor specifications of AR devices, such as a smartphone camera. To the game engine project we add a custom visual data generation script which samples the ground truth pose data from an existing dataset to the desired camera frame rate, then generates new camera images at each of these poses. Our script also outputs a camera intrinsics and extrinsics config file (calculated from the defined camera specifications, the game engine coordinate system and the dataset used), as well as the lists of timestamps and images required by some open-source SLAM algorithms.

Our methodology supports execution of VI-SLAM sequences through either SLAM libraries such as ORB-SLAM [6] or the Robot Operating System (ROS). If ROS is used, then an extra step is required prior to execution, to convert the VI-SLAM input data to *.bag* file format. Once the data has been prepared, sequences can be executed by following the instructions for the open-source VI-SLAM algorithm of choice (e.g., ORB-SLAM3 [6], VINS-Mono [48]). To facilitate batch processing of the large numbers of trials required for systematic experiments, we create a custom shell script, which can be used to execute sequences of multiple types, from multiple different environments. The outputs from this SLAM sequence execution step are the trajectory estimate files, which we evaluate using an existing toolbox [75]; we extend this toolbox to provide not only the standard SLAM error metrics absolute trajectory error (ATE) and relative error (RE), but also data on each individual sub-trajectory used to calculate RE, for more granular analysis. Finally, we create a custom Python script to analyze the visual and inertial input data, and correlations between these data and pose estimation error.

3.3 VI-SLAM Visual Texture Experiments

3.3.1 VI-SLAM Visual Texture Experiments Setup

We now demonstrate our VI-SLAM evaluation methodology through systematic experiments to study the effect of visual texture on pose estimation error. We define two metrics to assess visual texture, calculated from the grayscale image used as visual input to the VI-SLAM algorithm. As a measure of image complexity we use **Entropy** (Shannon entropy), implemented using the Python package scikit-image [55] and defined as:

$$H(X) = -\sum_{x_i \in \chi} P_X(x_i) \log P_X(x_i), \tag{1}$$

where the discrete random variable X is the pixel intensity value and it takes values in a set χ that contains all possible pixel intensity values x_i. $P_X(x_i)$ is the probability that the random variable X takes the value x_i. As a measure of the number and strength of edges in an image we use **Laplacian**, the variance of the Laplacian [17], implemented using the OpenCV Laplacian operator, and defined as:

$$Var(L) = Var\left(\frac{\partial^2 X}{\partial x^2} + \frac{\partial^2 X}{\partial y^2}\right), \tag{2}$$

where x and y are the 2D locations of the pixel with intensity values X, and L denotes the Laplacian of the pixel intensity values. In these experiments we study 2D visual textures by examining empty cuboid environments without 3D objects; the effects of visual textures that arise from 3D objects is a topic for future work, though objects are included in our museum case study (Section 5). In Unity 2020.3.14f1, using the High Definition Render Pipeline [64], we apply each visual texture to the walls, floor and ceiling of 6m×6m×4m rooms (large enough to fit the trajectories in our chosen SLAM datasets), to create different virtual environments. We light these environments using a single point light source, and test each texture at 10 different light levels, by setting the light source to 50, 100, 200, 300, 400, 550, 750, 1000, 2500 and 5000 lumens. Testing at different light levels simulates the scenario in which AR users enter an environment at different times of day. We set the horizontal field of view of our game engine camera to 79°, matching that of a high-end smartphone AR device, the Samsung Galaxy S20 [52].

We evaluate performance in our environments using four trajectories from two handheld VI-SLAM datasets, one from the commonly used TUM VI dataset [54], and three from an AR-specific dataset, SenseTime [29]. We use (1) TUM VI room5, a trajectory with rapid motion representative of a dynamic AR user, with camera views scanning almost all parts of the environment; (2) SenseTime A1, with slow side-to-side motion facing the wall, as if inspecting a virtual object at head height, followed by repeated walking away and returning with the camera angled more towards the floor (described in [29] as 'inspect+patrol'); (3) SenseTime A4, with slow motion focused on an area of the floor, followed by the same slow side-to-side motion facing the wall ('aiming+inspect'); (4) SenseTime A6, with slow motion focusing on then moving around a small area of floor, as

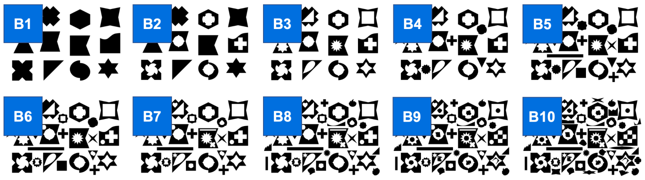


Figure 3: The 10 textures we used in our VI-SLAM experiments with basic textures. More edges are added with increasing texture number, resulting in greater **Laplacian** values (see Table 1).

Table 1: Summary of visual texture properties in our VI-SLAM experiments with basic textures (Section 3.3.2). 'Source' values are calculated directly from the source image texture; sequence values are averages across all input images, across all 10 light levels tested.

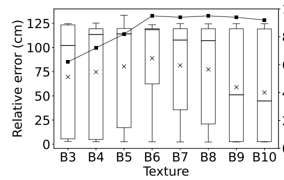
Texture	Entropy	Laplacian				
	Source	Source	room5	A1	A4	A6
B1	1.3	881	210	210	181	79
B2	1.3	1165	270	273	237	101
B3	1.3	1458	329	330	289	124
B4	1.4	1681	336	347	293	139
B5	1.5	1933	375	409	370	172
B6	1.5	2042	394	435	389	184
B7	1.6	2176	420	460	423	198
B8	1.7	2570	459	496	466	210
B9	1.8	2831	490	537	509	230
B10	1.9	3236	526	581	552	249

if inspecting a virtual map placed there ('hold+inspect'). The room5 sequence has the most challenging inertial data, in particular higher rates of rotation: the average magnitude of gyroscope readings for room5 is 1.5rad/s, compared to 0.3rad/s for A1 and A4 and 0.2rad/s for A6. We generate the camera images for each trajectory, in each environment, with a custom C# script, using the method introduced in our VI-SLAM evaluation methodology (Section 3.3.1).

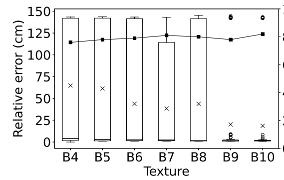
We execute our new sequences using a state-of-the-art open-source monocular VI-SLAM algorithm, ORB-SLAM3 [6], using default settings. We run them on a desktop computer (equipped with an Intel i7-9700K CPU and an Nvidia GeForce RTX 2060 GPU), using a virtual machine with 4 CPUs and 8GB of RAM, representative of the computational resources of a mobile AR device such as a high-end smartphone. To evaluate the estimated trajectories we use the toolbox provided in [75], and focus on the *translational component of RE* of each trajectory as our primary performance metric. We choose this metric because it corresponds to a noticeable issue in markerless AR, virtual object position error; the individual sub-trajectories from which RE is calculated represent a movement a user makes between virtual object views, and the translational component of the pose error at the end of a sub-trajectory is how much a virtual object drifts out of position during that movement. For all results in this paper, we report RE as calculated using a sub-trajectory length of 2m, as a typical length of movement an AR user might make between virtual object views, though in practice RE calculated using different lengths is usually highly correlated (with longer sub-trajectories resulting in greater error). We define an additional performance metric *robustness*, the mean percentage of input sequence frames that are tracked. Robustness is generally slightly higher for the TUM VI sequence, room5, than for SenseTime sequences, due to the latter having a 5-10s stationary period at the start of the trajectory, during which tracking is not initialized. We perform 10 trials at each of 10 light levels, a total of 100 trials for each texture, and 2000 across our experiments described below.

3.3.2 VI-SLAM Experiments: Basic Textures

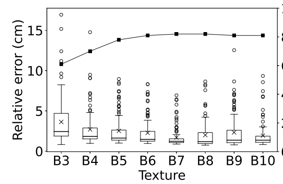
For our first set of visual texture experiments we created 10 low-complexity binary images (**Entropy** < 2 for all), with random shapes added incrementally to gradually increase the number of edges (and thereby the amount of visual texture), as shown in Figure 3. The **Entropy** and **Laplacian** values for these images are shown in Table 1; only **Laplacian** is shown per sequence because different types of motion in different trajectories do not have a large impact on **Entropy**, but do affect edge strength through motion blur. We applied these textures to the walls, floor and ceiling of our virtual environment (each black shape in the first texture, B1, covered approximately 1m$\times$1m), to create 10 different environments, then tested them using the procedure described in our VI-SLAM visual texture experiments setup (Section 3.3.1). We hypothesized that RE would decrease as the number of edges increased (increasing texture number).



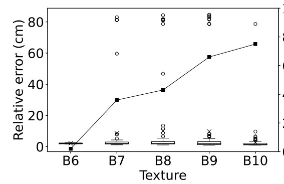
(a) TUM VI: room5



(b) SenseTime: A1



(c) SenseTime: A4



(d) SenseTime: A6

Figure 4: Relative error and robustness for four VI-SLAM sequences run with ORB-SLAM3 in 6m$\times$6m$\times$4m rooms with each basic texture (Figure 3), over 100 trials (10 light levels, 10 trials at each).

Table 2: VI-SLAM performance results for our experiments with basic textures (Section 3.3.2), over 100 trials. Robustness is the mean percentage of frames tracked over all trials.

Texture	Median RE (cm)				Robustness (%)			
	room5	A1	A4	A6	room5	A1	A4	A6
B1	-	-	-	-	0	0	0	0
B2	-	-	-	-	0	0	0	0
B3	102.1	-	2.4	-	62	0	61	0
B4	113.5	4.9	1.8	-	72	76	70	0
B5	114.1	4.8	1.6	-	82	78	78	0
B6	118.4	2.5	1.5	2.0	95	79	81	2
B7	107.7	2.3	1.2	2.1	94	81	82	36
B8	107.1	1.6	1.2	1.9	95	80	82	43
B9	51.1	1.4	1.4	1.8	94	78	81	66
B10	44.7	1.5	1.4	1.4	92	82	81	75

The results for our VI-SLAM experiments with basic textures are shown in Figure 4 and Table 2. Textures B1 and B2 failed on all trials for all sequences, B3 failed for A1, and B3, B4 and B5 failed for A6. Visual texture had the greatest effect on median RE for the more dynamic sequences, room5 and A1; median RE for room5 was 118.4cm with texture B6 compared to 44.7cm with texture B10, while for A1 median RE was 4.9cm with texture B4 but 1.5cm with texture B10. For the A6 sequence a dramatic decline in robustness at low textures was observed: 75% with texture B10, but just 2% with texture B6. This impact on robustness was also observed to a lesser extent for room5 and A4. For room5 textures with low robustness (e.g., B3, B4, B5) actually exhibited slightly lower median RE due to tracking being maintained accurately but only for a small number of frames on some trials. Multiple large RE values greater than 1m occurred with all textures for both room5 and A1.

3.3.3 VI-SLAM Experiments: Complex Textures

For our next set of visual texture experiments we selected 10 more complex images (**Entropy** > 5 for all), with levels of complexity

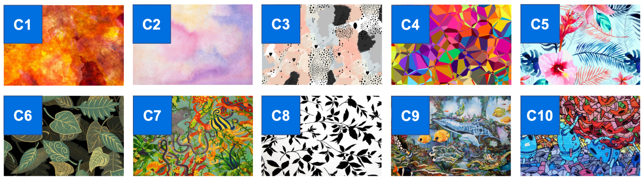


Figure 5: The 10 textures we used in our VI-SLAM experiments with complex textures.

Table 3: Summary of visual texture properties in our experiments with complex textures (Section 3.3.3). 'Source' values are calculated directly from the source image texture; sequence values are averages across all input images, across all 10 light levels tested.

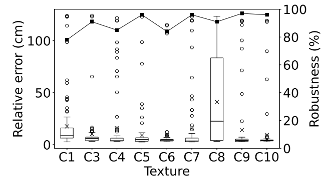
Texture number	Entropy	Laplacian				
	Source	Source	room5	A1	A4	A6
C1	7.1	341	46	59	61	53
C2	6.3	454	23	30	30	28
C3	6.6	498	132	290	301	139
C4	7.0	565	101	117	125	68
C5	6.9	723	128	166	175	117
C6	6.4	879	79	112	101	50
C7	7.4	1276	112	166	173	113
C8	5.1	2916	328	567	513	239
C9	7.8	3966	133	298	262	133
C10	7.8	4077	203	378	328	199

more representative of real-world environments. The selection criteria for these images were that the level of texture should be relatively uniform (i.e., no large blank spaces in which tracking might be lost), they should contain minimal repetitive elements, and that together they should cover a range of **Entropy** and **Laplacian** values. These images are shown in Figure 5, and the respective **Entropy** and **Laplacian** values in Table 3. We applied these textures to create 10 different virtual environments (scaled such that each image in Figure 5 covered a 6m×4m wall), then tested those environments following the procedure described in our VI-SLAM visual texture experiments setup (Section 3.3.1). Following the results we obtained in our experiments with basic textures, we hypothesized that RE would decrease and robustness would increase as the number of edges increased (increasing texture number).

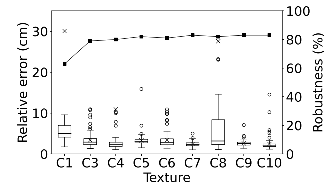
The results of our VI-SLAM experiments with complex textures are shown in Figure 6 and Table 4. Texture C2 failed for all sequences, while for sequence A1, outliers greater than 140cm are excluded from the plot (but included in statistics calculations): 12 for texture C1, 6 for texture C4, and 17 for texture C8. Out of the other textures, the texture with the lowest **Laplacian** values, texture C1, generally resulted in notably lower robustness than others. For room5 robustness was 78% with texture C1, compared to 84% or greater with all other textures; for A1 and A4 respectively robustness was 63% and 68% with texture C1, compared to at least 79% with all other textures. The exception to this was A6 which, as in our experiments with basic textures, exhibited lower levels of robustness across a wider range of textures – only with textures C7, C9, and C10 was robustness 75% or greater. Another standout result was that while edge strength was less impacted by motion blur on texture C8 (sequence **Laplacian** values were higher than for any other textures), performance was worse than with textures with comparable source **Laplacian** values for all sequences, in particular room5: for room5 median RE was 22.6cm with texture C8, compared to 3.6cm with C7 and 4.0cm with C9, and over twice as high as with C1 (8.6cm).

3.3.4 VI-SLAM Visual Texture Experiments Discussion

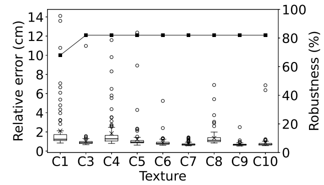
The results of our experiments reveal important insights regarding the impact of visual texture on VI-SLAM performance. The low level of robustness and high RE we observed at lower texture numbers, in particular with textures C1 and C2 in our complex texture



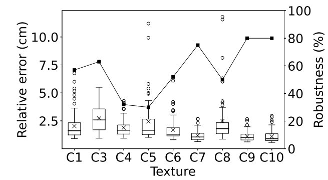
(a) TUM VI: room5



(b) SenseTime: A1



(c) SenseTime: A4



(d) SenseTime: A6

Figure 6: Relative error and robustness for four VI-SLAM sequences run with ORB-SLAM3 in 6m×6m×4m rooms with each complex texture (Figure 5), over 100 trials (10 light levels, 10 trials at each).

Table 4: VI-SLAM performance results for our experiments with complex textures (Section 3.3.3), over 100 trials. Robustness is the mean percentage of frames tracked over all trials.

Texture	Median RE (cm)				Robustness (%)			
	room5	A1	A4	A6	room5	A1	A4	A6
C1	8.6	5.0	1.4	1.6	78	63	68	57
C2	-	-	-	-	0	0	0	0
C3	6.1	2.9	0.9	2.6	91	79	82	63
C4	3.8	2.3	1.3	1.7	85	80	82	32
C5	5.1	3.1	0.9	1.7	96	82	82	30
C6	4.4	2.7	0.8	1.3	84	81	82	52
C7	3.6	2.3	0.7	1.1	96	83	82	75
C8	22.6	3.2	1.1	1.8	91	82	82	50
C9	4.0	2.6	0.7	1.0	97	83	82	80
C10	4.0	2.2	0.7	0.9	96	83	82	80

experiments, clearly demonstrates the challenges of hosting markerless AR in spaces without visual texture; tracking often either fails to initialize or is unavailable for large portions of many trajectories, unacceptable for AR scenarios. We also observed that *both the number and strength of edges (**Laplacian**) and the complexity (**Entropy**) in a visual environment determine pose tracking performance*. Aside from worse performance at lower **Laplacian** values, the complex texture with the lowest **Entropy**, texture C8 (5.1), performed worse than other textures with comparable **Laplacian** values; the texture with the second-lowest **Entropy**, texture C2 (6.3), failed to run at all, despite having a higher **Laplacian** value than the moderately successful texture C1 (**Entropy** = 7.1). Comparing our complex and basic texture results, greater **Entropy** in complex textures resulted in lower median RE for room5 than with basic textures; we posit that this is due to complexity aiding place recognition, as with the basic textures loop closure often did not occur at expected points, resulting in large RE on a dynamic trajectory such as room5.

Furthermore, *the effects of visual texture on pose error were strongest for the sequences with the greatest motion, but certain types of camera views in lower-texture environments also resulted in poor robustness*. The relative performance degradation in low texture environments was greatest with the most challenging inertial data (room5), but in A6 the fact that the camera view only covered a small portion of the environment floor meant that less texture was visible (**Laplacian** values for A6 in Tables 1 and 3 are frequently lower than for room5, despite the lack of motion blur), resulting in more frequent tracking losses. It is clear that to support AR scenarios and applications with dynamic users, for example young children, or animated content that prompts users to move their head rapidly to follow it around a room, we must understand how best to create environments with sufficient texture. However, we must also consider AR interactions which cause users to focus on specific regions of an environment, because they place additional requirements on those regions. The types of motion and camera views that different AR content prompts in users is a vital topic for future work.

3.3.5 Realistic Virtual Environments

The textures we examined in our VI-SLAM experiments were chosen to investigate an extensive distribution of **Laplacian** and **Entropy** values. We have also tested 10 textures designers are more likely to encounter in real-world scenarios, using the same format as with our basic and complex textures. Four of these source textures (speckled marble, brick, stone and wallpaper with a plant texture) had high edge strength (**Laplacian**$>$1800), while the remaining six (paint, concrete, rough concrete, marble with a soft texture, wood and repetitive geometric wallpaper) had low edge strength (**Laplacian**$<$500). The paint and soft marble textures failed to initialize for all sequences, while the geometric wallpaper resulted in a large number of high RE outliers, illustrating the problems posed by insufficient texture and repetitive elements respectively. Results for the remaining textures were consistent with our previous findings, with performance differences greatest in the presence of more challenging inertial data; for room5 three out of four high edge strength textures, brick, stone and the plants wallpaper, achieved median RE $\leq$5cm, compared to $>$20cm for all other textures. The notable exception was the speckled marble (median RE = 95cm): here the fine texture was greatly affected by motion blur, resulting in low edge strength in camera images in dynamic scenarios.

Our VI-SLAM evaluation methodology facilitates the creation of highly realistic virtual environments that incorporate these types of textures and natural 3D objects. Guidance on how to do this can be found, along with examples, on our GitHub page: `https://github.com/timscargill/Virtual-Inertial-SLAM/`.

4 User Evaluations of Virtual Environments

4.1 Motivation

In order to inform the design of AR spaces we require a method of testing the effect of environments on occupants, including both AR users and non-AR users. Physical environments are often impractical for this because (1) they require the space to already have been constructed, (2) they do not facilitate easy adjustment of visual conditions other than lighting, and (3) human participants have to be physically present to experience them. Specific to AR, the study of how different visual conditions affect the perception of virtual content is constrained by the limited range of physical environments available to researchers. Inspired by the recent use of head-mounted VR to study occupant satisfaction and behavior in environmental psychology [1], and for AR user interface design [34], we propose combining and extending these two approaches, to incorporate VR-based user evaluations into our integrated design of AR spaces. These user evaluations provide additional motivation for the use of virtual environments, and are a powerful tool for investigating human factors in both AR environment design and VI-SLAM algorithm design. Just as crowdsourcing can be employed for web quality of experience assessments [70], we envision the increasing availability of VR headsets facilitating crowdsourced, remote testing of simulated AR environments, in both academia and industry.

4.2 User Evaluation Methodology

An overview of our user evaluation methodology for simulated AR environments is shown in Figure 7. We create a virtual environment in a game engine such as Unity [66] or Unreal Engine [12], which can be evaluated from a tracking perspective using our VI-SLAM evaluation methodology (Section 3.2), in parallel to the process described here. For user evaluations, this virtual environment may simulate either only the real elements of an AR space or both real

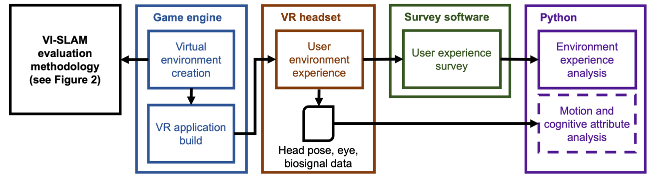


Figure 7: An overview of our user evaluation methodology to study the effect of different environments on occupants. In our integrated design methodology the virtual environments that are created can also be evaluated in terms of tracking performance (see Figure 2).

and virtual content. We then build a VR app tailored to the nature of the AR experience to be tested, which can be run on a VR headset such as the Meta Quest 2 [39] or Vive Pro [69]. This app may simply facilitate user movement in an environment, or incorporate interactions with simulated virtual content as well. It is also configured to record user head pose, and depending on available hardware, eye tracking data and biosignals such as heart rate (the HP Reverb G2 Omnicept VR headset contains both an eye tracker and a heart rate sensor [26]). This app can be packaged for different headset models and distributed online in the case of remote testing.

A human participant wears a headset running the VR app to experience the virtual environment (possibly performing a predefined task), while head pose, eye and any other available biosignal data are captured and saved to a file. After experiencing the environment, the participant answers survey questions regarding their experience; in this version of our methodology we use external survey software, but questions could also be incorporated into the VR app itself. These survey questions cover both general comfort and feelings about an environment (see [25] for examples), as well as application or use case-specific topics – for example, how easy it was to find or focus on AR content, or information learned from an AR experience.

Finally, we develop custom Python scripts to analyze the data collected in our user evaluations. Here we extract and examine user survey responses, analyzing them alongside the sequence characterization step in our VI-SLAM evaluation methodology, to enable quantitative analysis of the effect of visual conditions on human responses, and side-by-side comparison of performance from tracking and human perspectives. Although not implemented here, our methodology can also be extended to include analysis of user motion from head pose data (we provide an example of recorded head pose data in Figure 10), and cognitive attribute detection from eye and biosignal data, which has recently been demonstrated in VR for stress [24] and mental workload [36]. By gathering data from multiple onboard sensors in VR, it will be possible to gain additional insights on the impact of different virtual environments on users, further enhancing the efficacy of our integrated design methodology.

5 Case Study: Integrated Design of an AR Museum

Having developed our methodology for integrated design of AR spaces, we now demonstrate it using a case study that replicates a real-world use case. We consider the scenario in which an institution wishes to host AR applications in a specific physical space, and wishes to design that space to have optimal visual properties, to support high quality user experiences. Here we focus on the example of a small museum exhibit, featuring real historical artifacts that would be supplemented by virtual content. However, this scenario covers a wide range of possible AR host spaces, from art galleries and classrooms to conference rooms and design studios.

5.1 Motivation: AR Museum Design

Educational spaces such as museums and art galleries are a prominent use case for AR. AR has the ability to provide engaging, interactive experiences for visitors, and a number of institutions have already started to adopt it for specific exhibits, including the

Smithsonian Natural History Museum [60, 61], the Hamburger Kunsthalle [20], the Muséum national d'Histoire naturelle in Paris [42], and the Dalí Museum in St Petersburg, Florida [62]. In their AR-supported exhibit 'Invasive Species', the Perez Art Museum in Miami noted how visitors not only interacted with the technology, but that AR encouraged interactions between people, and that positive experiences were reported across age groups [41]. From a research perspective, study of these spaces involves many interesting challenges and opportunities, such as dynamic, multi-user environments, heterogeneous users, and a wide variety of possible virtual content.

In our work we extend an existing body of research on the effect of museum environments on visitor comfort, engagement, and behavior [4, 28, 32, 45, 72]. Now, given the recent interest in using AR in museums, there is a clear need to understand how the environmental requirements of the AR system and the museum visitor trade off against one another. The spatial registration errors that result from poor AR device pose tracking limit, and can even destroy, the sense of immersion and engagement that a visitor has with an exhibit; therefore it is vital that we can design spaces that minimize these errors, even for dynamic users, while maintaining a pleasant environment for visitors. Only through an integrated design methodology, which considers both tracking and human factors, can we do that.

5.2 Case Study Experiments Design

5.2.1 Virtual Museum Environment

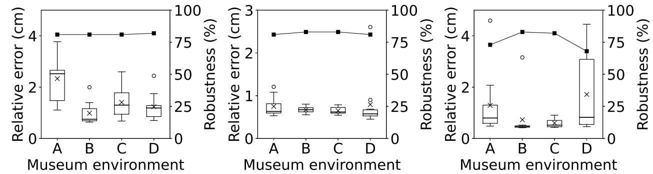
We created a virtual museum environment in the cross-platform game engine Unity [66], using 3D models and textures from the 'Museum VR Complete Edition' Unity asset [65]. The environment dimensions were 4m×3m×2.4m, chosen to match the size of the physical environment in which we conducted our VR-based experiments. We created an exhibit on Ancient Egypt, containing five 3D models of real artifacts, sourced from the aforementioned Unity asset. To study the effects of different visual textures in our museum space we created four versions of the environment, each with a different wall and floor texture combination. Images of these environments are shown in Figure 1a: *MuseumA* (top left) had blank gray walls with a wooden paneled baseboard and a wooden floor; *MuseumB* (top right) had gray marble walls and a patterned marble floor; *MuseumC* (bottom left) had wooden walls and a patterned marble floor; *MuseumD* (bottom right) had commercially available wallpaper with an Ancient Egyptian design [67] and a wooden floor.

5.2.2 VI-SLAM Performance Experiments Design

We evaluated VI-SLAM performance for each of the four museum environments (Figure 1a) with ORB-SLAM3 (default settings), using the methodology we established in Section 3, on the SenseTime A1, A4 and A6 sequences (the TUM VI sequence trajectory we used in our previous experiments was too large to fit within our museum environment). The **Entropy** and **Laplacian** values for these environments, averages across the input images for each sequence, are shown in Table 5. We performed 10 trials for each sequence in each environment, with a fixed, 'normal' light level. We hypothesized that the well-defined textures of the Ancient Egyptian wallpaper (*MuseumD*) would result in the lowest RE, the less clear marble and wood textures (*MuseumB* and *MuseumC*) would result in higher RE, and the blank walls (*MuseumA*) would result in the highest RE.

Table 5: Summary of visual properties for the four museum environments in our case study. Values for each sequence are averages across all sequence input images.

Museum	Entropy			Laplacian		
	A1	A4	A6	A1	A4	A6
A	5.8	5.8	5.4	74	67	24
B	6.0	6.0	5.8	102	262	119
C	6.2	6.1	5.8	112	269	119
D	7.1	6.4	5.4	228	145	23



(a) SenseTime: A1 (b) SenseTime: A4 (c) SenseTime: A6

Figure 8: Relative error and robustness for three VI-SLAM sequences run with ORB-SLAM3 in our four museum environments, over 10 trials.

Table 6: VI-SLAM performance results for the four museum environments in our case study, over 10 trials. Robustness is the mean percentage of frames tracked over all trials.

Museum	Median RE (cm)			Robustness (%)		
	A1	A4	A6	A1	A4	A6
A	2.5	0.6	0.8	81	81	73
B	0.7	0.7	0.5	81	83	83
C	1.3	0.6	0.5	81	83	82
D	1.2	0.6	0.8	82	81	68

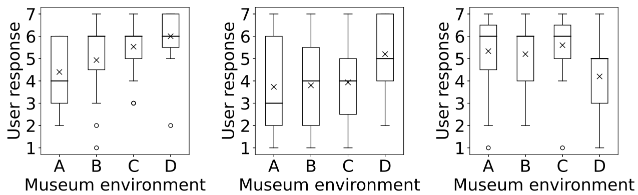
5.2.3 User Satisfaction Experiments Design

We evaluated user satisfaction with each museum environment through an IRB-approved user study, using the process described in our user evaluation methodology (Section 4.2). We chose to use the Meta Quest 2 VR headset because it is a standalone (non-tethered) headset, allowing natural user motion in our museum environment. All survey questions were administered through Qualtrics [49]. We recruited 15 participants aged 18 to 34 (6 female) for this pilot study, from our personal and professional networks. Our experiment design and surveys were based on an existing study of user comfort in VR environments [25], which measured satisfaction with four different window sizes. After an initial pre-experiment survey to determine eligibility, previous VR experience, and gather demographic data, participants freely explored each environment in VR for two minutes, with the order the environments were presented in randomized for each participant. After each environment, participants answered eight survey questions on their experience, in which they rated the extent to which they agreed or disagreed with statements on a 7-point Likert scale. These questions covered the participant's overall visual comfort in and satisfaction with a space, how pleasant, comfortable and open the space felt, how easy it was to focus on the museum objects, as well as eye, mind and body fatigue. Participants had the option to provide additional comments about each environment, then had a two-minute break or 'washout' period before the next environment. After experiencing all four environments they completed a short post-experiment survey to gather feedback on the experiment protocol and any resulting fatigue or motion sickness.

5.3 Case Study Results

5.3.1 VI-SLAM Performance Experiments Results

The results of our case study VI-SLAM experiments are shown in Figure 8 and Table 6. As in our VI-SLAM visual texture experiments (Section 3.3), tracking performance varied more between environments for A1 and A6 than for the less challenging A4 sequence. For A1, robustness was consistent across environments, but the lowest median RE was achieved in the marble walls and floor of *MuseumB* (0.7cm), followed by *MuseumD* (1.2cm) and *MuseumC* (1.3cm), with *MuseumA*, which had blank walls, resulting in the highest median RE (2.5cm). For A6, with camera views focused on the floor, the two environments with the marble floor, *MuseumB* and *MuseumC*, resulted in greater robustness and lower median RE (83% and 82% robustness respectively, median RE = 0.5cm for both) than those with wooden floors, *MuseumA* (73% robustness, median RE = 0.8cm) and *MuseumD* (68% robustness, median RE = 0.8cm).



(a) "This space was pleasant"

(b) "This space felt open"

(c) "It was easy to focus on the museum objects"

Figure 9: User opinions of our four museum environments, rated on a 7-point Likert scale.

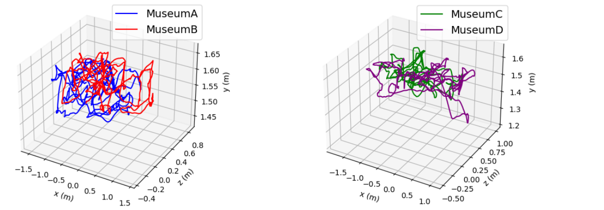
5.3.2 User Satisfaction Experiment Results

The results of our user satisfaction experiment for our case study are shown in Figure 9. We assign the following numeric values to the responses on our 7-point Likert scale: 1 = Strongly disagree, 2 = Disagree, 3 = Somewhat disagree, 4 = Neither agree nor disagree, 5 = Somewhat agree, 6 = Agree, 7 = Strongly agree. Participants found *MuseumD*, with the Egyptian wallpaper pattern, the most pleasant environment (Figure 9a), with a mean response of 6.0 to the statement "This space was pleasant" (*MuseumA* = 4.4, *MuseumB* = 4.9, *MuseumC* = 5.5). This may be due to the fact that *MuseumD* also felt most open to our participants (Figure 9b). We hypothesize that wall texture brightness caused this effect; one participant commented that in *MuseumD* "*The wallpapers felt back-lit, which made the space feel less dark and gloomy*." However, participants also found that the texture in *MuseumD* made it more difficult to focus on the museum objects (Figure 9c); the mean response for the statement "It was easy to focus on the museum objects" was 4.2 for *MuseumD*, compared to 5.2 or greater for the other museum environments. Comments about *MuseumD* included "*The room felt brighter, but the wallpaper was distracting*," "*The mural on the wall felt a bit distracting*," and "*...it was hard for me to focus on the artifacts as I was focusing on the walls*," while one participant commented that in *MuseumB*, "*The simpler background seems to make it easier to focus on the objects*." Examples of the user head poses we captured during this experiment are shown in Figure 10.

5.4 Case Study Discussion

The results of our user satisfaction experiment in our case study highlight the potential pitfalls of adding visual texture to real-world environments to improve AR tracking performance, without considering the impact on users. Commercial AR platform guidelines [2, 18, 40] call for surfaces with distinct, recognizable textures, which might lead a designer of a space hosting AR to choose a texture such as the wallpaper in *MuseumD*. But as our user study illustrated, these textures can be distracting and interfere with an occupant's activities.

In fact, our VI-SLAM experiments results revealed that *MuseumB*, with the marble wall and floor textures, actually provided better tracking performance than *MuseumD*. The reason for the greater tracking error observed in A6 was clear: the wooden floors provided less visual texture than the marble floors, both in terms of complexity (**Entropy**) and edge strength (**Laplacian**), as shown in Table 5. We hypothesized that this was also the cause of greater error in A1, while the camera was pointed towards the floor. However, when we used our extended trajectory evaluation tool (see Section 3.2) to analyze the pose error in different parts of the sequence, we found that the difference in error between *MuseumB* and *MuseumD* was greatest when the camera was focused on the wall. This unexpected result, that a soft marble wall texture provided better tracking performance than the distinct Egyptian wallpaper (at least when combined with other 3D objects), indicates both that further or refined metrics will be required to fully characterize environments, and that human perception of visual texture differs from the degree to which those textures support VI-SLAM. This opens up exciting avenues for future research to develop and characterize environment textures that facilitate high quality tracking, with minimal impact on AR user perception of real and virtual objects.



(a) *MuseumA* and *MuseumB*

(b) *MuseumC* and *MuseumD*

Figure 10: Examples of the head pose trajectories captured (at approximately 72Hz), during the two minute free exploration of each virtual museum environment, from one participant. Analyzing these trajectories can help inform selection of the most appropriate VI-SLAM dataset sequences in our integrated design methodology.

6 Conclusion and Future Work

In this paper we presented and demonstrated an integrated design methodology for real-world spaces which host AR experiences, which takes into account both tracking and human factors. We developed the first VI-SLAM evaluation method that enables controlled testing in virtual environments while using real inertial data from handheld or head-mounted AR devices, and employed it to conduct systematic experiments with different visual textures. We then presented a method to evaluate user satisfaction with those same virtual environments using VR headsets. Combining both of these elements, we demonstrated our integrated design methodology through a case study on AR museum design, which highlighted the complex and sometimes unexpected effects of applying different visual textures to an environment, and thereby the need for our solution.

Our integrated design methodology sets the stage for exciting opportunities for future work. Firstly, the promising results we obtained regarding the relationship between visual texture and tracking performance invite further experiments using our VI-SLAM evaluation method, towards accurate prediction of performance from visual and inertial input data. The study of non-uniform textures, in particular the impact of 'challenging regions' such as blank walls will be informative. Secondly, we focus here on the simulation of the real-world environment, but our methodology also supports the simulation of AR content in VR (e.g., user interfaces [34]), to study how real-world environments affect perception of AR content, as well as overall user comfort. Methods to accurately replicate the color-blending effects that occur in optical see-through AR, guided by recent research [14, 21], will be important for realistic AR content simulation. Our ultimate goal is to also simulate the AR content registration errors that arise from inaccurate tracking, thus fully integrating the study of tracking and human factors; this is a complex and open challenge however, requiring the implementation of an open-source VI-SLAM algorithm on a VR headset, running in real time on the headset images of the virtual environment and inertial data. As we noted in our user evaluation methodology, the collection and analysis of eye and biosignal data in user studies will enable additional insights into the cognitive states of participants in different environments. Finally, no existing studies have examined occupant comfort in multi-user virtual environments [1] – this is highly relevant to AR, and forms another part of our vision for remote, collaborative design and evaluation of real-world AR scenarios using VR.

Acknowledgments

We thank our user study participants for their invaluable assistance in this research. This work was supported in part by NSF grants CSR-1903136, CNS-1908051, and CNS-2112562, NSF CAREER Award IIS-2046072, and a Facebook Research Award.

REFERENCES

[1] H. Alamirah, M. Schweiker, and E. Azar. Immersive virtual environments for occupant comfort and adaptive behavior research–a comprehensive review of tools and applications. *Building and Environment*, 207:108396, 2022.

[2] Apple. Human Interface Guidelines: Augmented Reality. `https://developer.apple.com/design/human-interface-guidelines/ios/system-capabilities/augmented-reality/`, 2022.

[3] A. M. Bettens, B. Morrell, M. Coen, N. McHenry, X. Wu, P. Gibbens, and G. Chamitoff. UnrealNavigation: Simulation software for testing SLAM in virtual reality. In *Proceedings of AIAA Scitech 2020 Forum*, 2020.

[4] M. A. Bonn, S. M. Joseph-Mathews, M. Dai, S. Hayes, and J. Cave. Heritage/cultural attraction atmospherics: Creating the right environment for the heritage/cultural visitor. *Journal of Travel Research*, 45(3):345–354, 2007.

[5] M. Burri, J. Nikolic, P. Gohl, T. Schneider, J. Rehder, S. Omari, M. W. Achtelik, and R. Siegwart. The EuRoC micro aerial vehicle datasets. *The International Journal of Robotics Research*, 35(10):1157–1163, 2016.

[6] C. Campos, R. Elvira, J. J. G. Rodríguez, J. M. Montiel, and J. D. Tardós. ORB-SLAM3: An accurate open-source library for visual, visual–inertial, and multimap SLAM. *IEEE Transactions on Robotics*, pp. 1–17, 2021.

[7] K. Chamilothori, J. Wienold, and M. Andersen. Adequacy of immersive virtual reality for the perception of daylit spaces: Comparison of real and virtual environments. *Leukos*, 15(2-3):203–226, 2019.

[8] C. Chen, P. Zhao, C. X. Lu, W. Wang, A. Markham, and N. Trigoni. Deep-learning-based pedestrian inertial navigation: Methods, data set, and on-device inference. *IEEE Internet of Things Journal*, 7(5):4431–4441, 2020.

[9] S. Cortés, A. Solin, E. Rahtu, and J. Kannala. ADVIO: An authentic dataset for visual-inertial odometry. In *Proceedings of ECCV*, 2018.

[10] S. Debernardis, M. Fiorentino, M. Gattullo, G. Monno, and A. E. Uva. Text readability in head-worn displays: Color and style optimization in video versus optical see-through devices. *IEEE Transactions on Visualization and Computer Graphics*, 20(1):125–139, 2013.

[11] L. Edwards and P. Torcellini. *Literature review of the effects of natural light on building occupants*. National Renewable Energy Laboratory, Golden, CO (US), 2002.

[12] Epic Games. Unreal Engine. `https://www.unrealengine.com/`, 2022.

[13] L. Ferranti, X. Li, J. Boutellier, and J. Kannala. Can you trust your pose? Confidence estimation in visual localization. In *Proceedings of IEEE ICPR*, 2020.

[14] J. L. Gabbard, M. Smith, C. Merenda, G. Burnett, and D. R. Large. A perceptual color-matching method for examining color blending in augmented reality head-up display graphics. *IEEE Transactions on Visualization and Computer Graphics*, 28:2834–2851, 2022.

[15] J. L. Gabbard, J. E. Swan, and A. Zarger. Color blending in outdoor optical see-through AR: The effect of real-world backgrounds on user interface color. In *Proceedings of IEEE VR*, 2013.

[16] O. Ganoni and R. Mukundan. A framework for visually realistic multi-robot simulation in natural environment. *arXiv preprint arXiv:1708.01938*, 2017.

[17] J. Garforth and B. Webb. Visual appearance analysis of forest scenes for monocular SLAM. In *Proceedings of IEEE ICRA*, 2019.

[18] Google. ARCore design guidelines: Environment. `https://developers.google.com/ar/design/environment/definition/`, 2022.

[19] A. Guide et al. Environmental design. *Chartered Institute of Building Services Engineers (CIBSE)*, 2006.

[20] Hamburger Kunsthalle. Augmented reality application on Emil Nolde's painting technique. `https://www.hamburger-kunsthalle.de/en/augmented-reality-application-emil-noldes-painting-technique-0`, 2022.

[21] N. Hassani and M. J. Murdoch. Investigating color appearance in optical see-through augmented reality. *Color Research & Application*, 44(4):492–507, 2019.

[22] B. P. Haynes. The impact of office layout on productivity. *Journal of Facilities Management*, 2008.

[23] A. Heydarian, E. Pantazis, J. P. Carneiro, D. Gerber, and B. Becerik-Gerber. Lights, building, action: Impact of default lighting settings on occupant behaviour. *Journal of Environmental Psychology*, 48:212–223, 2016.

[24] C. Hirt, M. Eckard, and A. Kunz. Stress generation and non-intrusive measurement in virtual environments using eye tracking. *Journal of Ambient Intelligence and Humanized Computing*, 11(12):5977–5989, 2020.

[25] T. Hong, M. Lee, S. Yeom, and K. Jeong. Occupant responses on satisfaction with window size in physical and virtual built environments. *Building and Environment*, 166:106409, 2019.

[26] HP. HP Reverb G2 Omnicept Edition. `https://www.hp.com/us-en/vr/reverb-g2-vr-headset-omnicept-edition.html`, 2022.

[27] IKEA. IKEA Place app. `https://www.ikea.com/us/en/customer-service/mobile-apps/`, 2022.

[28] J.-H. Jeong and K.-H. Lee. The physical environment in museums and its effects on visitors' satisfaction. *Building and Environment*, 41(7):963–969, 2006.

[29] L. Jinyu, Y. Bangbang, C. Danpeng, W. Nan, Z. Guofeng, and B. Hujun. Survey and evaluation of monocular visual-inertial SLAM algorithms for augmented reality. *Virtual Reality & Intelligent Hardware*, 1(4):386–410, 2019.

[30] M. Kasper, S. McGuire, and C. Heckman. A benchmark for visual-inertial odometry systems employing onboard illumination. In *Proceedings of IEEE/RSJ IROS*, 2019.

[31] P. Kotler. Atmospherics as a marketing tool. *Journal of retailing*, 49(4):48–64, 1973.

[32] R. Kottasz. Understanding the influences of atmospheric cues on the emotional responses and behaviours of museum visitors. *Journal of Nonprofit & Public Sector Marketing*, 16(1-2):95–121, 2006.

[33] N. Kwallek, C. Lewis, J. Lin-Hsiao, and H. Woodson. Effects of nine monochromatic office interior colors on clerical tasks and worker mood. *Color Research & Application*, 21(6):448–458, 1996.

[34] D. Lindlbauer, A. M. Feit, and O. Hilliges. Context-aware online adaptation of mixed reality interfaces. In *Proceedings of ACM UIST*, 2019.

[35] J. M. Loomis, J. J. Blascovich, and A. C. Beall. Immersive virtual environment technology as a basic research tool in psychology. *Behavior Research Methods, Instruments, & Computers*, 31(4):557–564, 1999.

[36] T. Luong, N. Martin, A. Raison, F. Argelaguet, J.-M. Diverrez, and A. Lécuyer. Towards real-time recognition of users mental workload using integrated physiological sensors into a VR HMD. In *IEEE ISMAR*, 2020.

[37] J. H. Ma, J. K. Lee, and S. H. Cha. Effects of lighting CCT and illuminance on visual perception and task performance in immersive virtual environments. *Building and Environment*, 209:108678, 2022.

[38] F. H. Mahnke. *Color, environment, and human response: an interdisciplinary understanding of color and its use as a beneficial element in the design of the architectural environment*. John Wiley & Sons, 1996.

[39] Meta. Meta Quest 2. `https://store.facebook.com/quest/products/quest-2/`, 2022.

[40] Microsoft. Guidelines for effective anchor experiences. `https://docs.microsoft.com/en-us/azure/spatial-anchors/concepts/guidelines-effective-anchor-experiences`, 2022.

[41] MuseumNext. Augmented reality art at Pérez Art Museum Miami. `https://vimeo.com/313237380`, 2022.

[42] Muséum national d'Histoire naturelle. Revivre, extinct animals in augmented reality. `https://www.mnhn.fr/en/experience/revivre-extinct-animals-in-augmented-reality`, 2022.

[43] Niantic. Pokemon GO app. `https://www.pokemongo.com/`, 2022.

[44] OptiTrack. OptiTrack. `https://optitrack.com/`, 2022.

[45] J. Peponis, R. C. Dalton, J. Wineman, and N. Dalton. Measuring the effects of layout upon visitors' spatial behaviors in open plan exhibition settings. *Environment and Planning B: Planning and Design*, 31(3):453–473, 2004.

[46] T. Pollok, L. Junglas, B. Ruf, and A. Schumann. UnrealGT: Using

Unreal engine to generate ground truth datasets. In *Proceedings of ISCV*, 2019.
[47] P. K. Pressly and M. Heesacker. The physical environment and counseling: A review of theory and research. *Journal of Counseling & Development*, 79(2):148–160, 2001.
[48] T. Qin, P. Li, and S. Shen. VINS-Mono: A robust and versatile monocular visual-inertial state estimator. *IEEE Transactions on Robotics*, 34(4):1004–1020, 2018.
[49] Qualtrics. Qualtrics. `https://www.qualtrics.com`, 2022.
[50] S. Saeedi, E. D. Carvalho, W. Li, D. Tzoumanikas, S. Leutenegger, P. H. Kelly, and A. J. Davison. Characterizing visual localization and mapping datasets. In *Proceedings of IEEE ICRA*, 2019.
[51] F. Salamone, A. Bellazzi, L. Belussi, G. Damato, L. Danza, F. Dell'Aquila, M. Ghellere, V. Megale, I. Meroni, and W. Vitaletti. Evaluation of the visual stimuli on personal thermal comfort perception in real and virtual environments using machine learning approaches. *Sensors*, 20(6):1627, 2020.
[52] Samsung. Galaxy S20 specifications. `https://www.samsung.com/us/mobile/galaxy-s20-5g/specs/`, 2022.
[53] T. Scargill, G. Premsankar, J. Chen, and M. Gorlatova. Here to stay: A quantitative comparison of virtual object stability in markerless mobile AR. In *Proceedings of IEEE/ACM CPHS Workshop (co-located with CPS-IoT week 2022)*, 2022.
[54] D. Schubert, T. Goll, N. Demmel, V. Usenko, J. Stückler, and D. Cremers. The TUM VI benchmark for evaluating visual-inertial odometry. In *Proceedings of IEEE/RSJ IROS*, 2018.
[55] scikit image. scikit-image: Image processing in Python. `https://scikit-image.org/`, 2022.
[56] S. Shah, D. Dey, C. Lovett, and A. Kapoor. AirSim: High-fidelity visual and physical simulation for autonomous vehicles. In *Field and service robotics*, pp. 621–635. Springer, 2018.
[57] A. Sivaji, S. Shopian, Z. M. Nor, N.-K. Chuan, and S. Bahri. Lighting does matter: Preliminary assessment on office workers. *Procedia-Social and Behavioral Sciences*, 97:638–647, 2013.
[58] J. Skinner, S. Garg, N. Sünderhauf, P. Corke, B. Upcroft, and M. Milford. High-fidelity simulation for evaluating robotic vision performance. In *Proceedings of IEEE/RSJ IROS*, 2016.
[59] C. Slocum, X. Ran, and J. Chen. RealityCheck: A tool to evaluate spatial inconsistency in augmented reality. In *Proceedings of IEEE ISM*, 2021.
[60] Smithsonian. Critical distance. `https://naturalhistory.si.edu/exhibits/critical-distance`, 2022.
[61] Smithsonian. Skin and bones. `https://www.si.edu/newsdesk/releases/smithsonian-brings-historic-specimens-life-free-skin-and-bones-mobile-app`, 2022.
[62] The Dalí Museum. Visual magic: Dalí's masterworks in AR. `https://thedali.org/exhibit/visual-magic-dalis-masterworks-in-augmented-reality/`, 2022.
[63] L. Troup, R. Phillips, M. J. Eckelman, and D. Fannon. Effect of window-to-wall ratio on measured energy consumption in US office buildings. *Energy and Buildings*, 203:109434, 2019.
[64] Unity. High definition render pipeline. `https://unity.com/srp/High-Definition-Render-Pipeline`, 2022.
[65] Unity. Museum VR complete edition. `https://assetstore.unity.com/packages/3d/environments/museum-vr-complete-edition-89652`, 2022.
[66] Unity. Unity. `https://unity.com/`, 2022.
[67] Vertigo Home. Egyptian queens wallpaper. `https://www.vertigohome.us/products/egyptian-queens-wallpaper-by-mind-the-gap?variant=39363204513873`, 2022.
[68] Vicon. Vicon. `https://www.vicon.com/`, 2022.
[69] Vive. Vive Pro. `https://www.vive.com/us/product/vive-pro/`, 2022.
[70] P. Wang, M. Varvello, and A. Kuzmanovic. Kaleidoscope: A crowdsourcing testing tool for web quality of experience. In *Proceedings of IEEE ICDCS*, 2019.
[71] M. Watchman, A. Potvin, and C. M. Demers. Wood and comfort: a comparative case study of two multifunctional rooms. *BioResources*, 12(1):168–182, 2017.
[72] J. D. Wineman and J. Peponis. Constructing spatial meaning: Spatial affordances in museum design. *Environment and Behavior*, 42(1):86–109, 2010.
[73] D. D. Wollin and M. Montagne. College classroom environment: Effects of sterility versus amiability on student and teacher performance. *Environment and Behavior*, 13(6):707–716, 1981.
[74] K. Yildirim, A. Akalin-Baskaya, and M. Celebi. The effects of window proximity, partition height, and gender on perceptions of open-plan offices. *Journal of Environmental Psychology*, 27(2):154–165, 2007.
[75] Z. Zhang and D. Scaramuzza. A tutorial on quantitative trajectory evaluation for visual (-inertial) odometry. In *Proceedings of IEEE/RSJ IROS*, 2018.
[76] Z. Zhu, F. Xu, M. Li, Z. Wang, and C. Yan. Challenges from fast camera motion and image blur: Dataset and evaluation. In *Proceedings of ECCV*, 2020.
[77] D. Zuñiga-Noël, A. Jaenal, R. Gomez-Ojeda, and J. Gonzalez-Jimenez. The UMA-VI dataset: Visual–inertial odometry in low-textured and dynamic illumination environments. *The International Journal of Robotics Research*, 39(9):1052–1060, 2020.