Invisible Textures: Comparing Machine and Human Perception of Environment Texture for AR

Tim Scargill
ts352@duke.edu
Department of Electrical and Computer Engineering
Duke University
Durham, NC, USA

Majda Hadziahmetovic
majda.hadziahmetovic@duke.edu
Department of Ophthalmology
Duke University
Durham, NC, USA

Maria Gorlatova
maria.gorlatova@duke.edu
Department of Electrical and Computer Engineering
Duke University
Durham, NC, USA

ABSTRACT

Mobile augmented reality (AR) has a wide range of promising applications, but its efficacy is subject to the impact of environment texture on both machine and human perception. Performance of the machine perception algorithm underlying accurate positioning of virtual content, visual-inertial SLAM (VI-SLAM), is known to degrade in low-texture conditions, but there is a lack of data in realistic scenarios. We address this through extensive experiments using a game engine-based emulator, with 112 textures and over 5000 trials. Conversely, human task performance and response times in AR have been shown to increase in environments perceived as textured. We investigate and provide encouraging evidence for *invisible textures*, which result in good VI-SLAM performance with minimal impact on human perception of virtual content. This arises from fundamental differences between VI-SLAM-based machine perception, and human perception as described by the *contrast sensitivity function*. Our insights open up exciting possibilities for deploying ambient IoT devices that display invisible textures, as part of systems which automatically optimize AR environments.

CCS CONCEPTS

• **Human-centered computing** → **Mixed / augmented reality**.

KEYWORDS

Augmented reality, VI-SLAM, perception, texture

ImmerCom '23, October 6, 2023, Madrid, Spain

ACM ISBN 979-8-4007-0339-3/23/10…$15.00
https://doi.org/10.1145/3615452.3617937

ACM Reference Format:
Tim Scargill, Majda Hadziahmetovic, and Maria Gorlatova. 2023. Invisible Textures: Comparing Machine and Human Perception of Environment Texture for AR. In *The 1st ACM Workshop on Mobile Immersive Computing, Networking, and Systems (ImmerCom '23), October 6, 2023, Madrid, Spain.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3615452.3617937

1 INTRODUCTION

The visual textures in a real-world environment play a prominent role in determining the quality of augmented reality (AR) experiences, because they impact both AR system performance and, as illustrated in Figure 1, user perception of virtual content. These effects are in turn subject to other dynamic factors, including lighting and user motion. Inspired by works that employ IoT devices to optimize environments for comfort [1, 29] and energy efficiency [20, 24], and ambient displays to enhance mental health [9], this motivates the development of IoT-enabled texture optimization systems for AR, in which ambient displays automatically provide environment textures best-suited to current interactions. However, to inform the design of these systems we must first quantify the impact of texture, and examine the potential for textures which satisfy the needs of AR systems and users.

The challenge in identifying optimal environment textures for AR arises from the conflicting requirements of machine and human perception. Modern mobile AR systems rely on visual-inertial simultaneous localization and mapping (VI-SLAM)-based device pose tracking to accurately position virtual content; when sufficient texture is not available in input camera images, VI-SLAM algorithm performance degrades [5, 17], and virtual content appears out of position or unstable [26, 32]. On the other hand, complex background textures are generally detrimental for human perception in AR, because they can be distracting or affect virtual content visibility [30, 37]. Here we compare the effect of texture on machine perception, i.e., VI-SLAM, with human perception of texture, to inform the design of environments which support good tracking performance *and* are perceived as low in complexity by AR users.

(a) (b)

Figure 1: Virtual 3D equipment assembly instructions viewed on a Magic Leap One AR headset, in front of different textures on an IoT ambient display (a smart TV). Interpreting virtual content in front of a texture prioritizing VI-SLAM performance (a) is more challenging than with a texture prioritizing human factors (b).

We start by quantifying the effect of environment texture on VI-SLAM performance, using a game engine-based emulator to test device trajectories with existing ground truth in realistic virtual environments. We then compare VI-SLAM performance with human (subjective) scores of texture complexity to answer two questions: (1) Is human complexity perception a valuable predictor of VI-SLAM performance? and (2) Can we identify what we term *invisible textures*, textures humans perceive as low in complexity, yet which result in near-optimal VI-SLAM performance? To the best of our knowledge, this work is the most extensive evaluation of the effect of texture on VI-SLAM performance, and the first to analyze this in the context of human perception of texture. First we cover related work (Section 2) and detail our VI-SLAM visual texture experiments setup (Section 3). We present our results and analyze the relationships between texture and VI-SLAM performance (Section 4), as well as the relationship between VI-SLAM performance and human perception of texture complexity (Section 5), then finish with conclusions and future work (Section 6). Our key contributions are:

- We perform the most extensive evaluation of the effect of environment texture on VI-SLAM performance to date, with 112 textures, five different environment and trajectory combinations, and over 5000 total trials.
- We demonstrate the impact of texture on VI-SLAM performance, with error for the worst-performing texture ranging from 203–3171% of error for the best-performing texture, and identify key texture characterization metrics.
- We analyze the relationship between VI-SLAM performance and human perception of texture complexity, demonstrating the problem with using subjective complexity measures for environment design, and revealing the potential for deploying *invisible textures* in AR environments.

2 RELATED WORK

Effect of texture on VI-SLAM performance: The vast majority of AR systems use feature-based VI-SLAM, which is less robust than direct methods to low-texture environments [5, 17]. Some VI-SLAM benchmarks provide valuable qualitative descriptions of visual inputs [3, 17], however none quantify environment texture – in general texture in this context is under-studied. In [30] we examined the effect of texture on VI-SLAM performance using two characterization metrics and empty cuboid environments; we build on this with more characterization metrics, environments representative of AR scenarios, and a larger set of textures.

Human perception of texture complexity: In contrast, human perception of texture has received much more attention; we focus on the relationship between texture characteristics and perceived *complexity* (e.g., [8, 13]). This has many applications, including product design [13] and video quality assessment metrics [36]. Qualitatively, a texture's 'regularity', 'understandability' and 'quantity of details' are reported as important in complexity perception, with more regular textures perceived as less complex [8]. A variety of quantitative metrics have been considered in the prediction of human complexity perception (e.g., [8, 13]) and we draw inspiration from this work in our characterization metrics.

Effect of texture on human perception in AR: The cognitive states of human occupants are affected by environment textures. Occupants may report being distracted by textures [30] or complete tasks less effectively in textured environments [10]. Users of AR devices with optical see-through displays (e.g., Microsoft HoloLens and Magic Leap headsets) may also experience impaired virtual content visibility due to background texture. Studies of text legibility reported increased error rates and response times with textured backgrounds [11], and higher background texture contrast results in higher response times and greater perceived virtual content transparency in optical see-through AR [37]. This motivates our search for *invisible textures* with low perceived complexity and low contrast (Section 5).

3 VI-SLAM VISUAL TEXTURE EXPERIMENTS SETUP

In this section we detail the texture datasets and characterization metrics we used for our experiments, the setup of our VI-SLAM visual texture evaluations, and the VI-SLAM sequence configurations we tested in our experiments.

3.1 Visual Texture Datasets and Characterization Metrics

For our VI-SLAM texture experiments we chose the texture images used in [8], for which image files, texture characterization metrics and human subjective scores of complexity are available. This set of 112 textures comprises two datasets: 54 images from the VisTex dataset [25], and 58 images from the RawFooT dataset [16]. We quantified texture properties using

Table 1: Characterization metrics used to quantify the properties of textures in our VI-SLAM visual texture experiments (GLCM = gray-level co-occurrence matrix).

Metric	Description
Brightness	Normalized mean pixel intensity
Contrast_RMS	Root mean square contrast
Entropy	Shannon entropy [34] of pixel intensities
Laplacian	Variance of the Laplacian [12]
Corners	Number of corners detected by the FAST algorithm [28], using OpenCV
Contrast	Contrast derived from GLCM [22]
Correlation	How correlated each pixel is with its neighbors, derived from GLCM [22]
Energy	Sum of squared GLCM elements [22]
Homogeneity	Distribution of GLCM elements with respect to the diagonal [22]
Freq. Factor	Ratio between the frequency corresponding to 99% of the image energy and Nyquist frequency [7]
Edge Density	Density of edge pixels as defined by the Canny edge detector [27]

11 characterization metrics, detailed in Table 1. For each texture we calculated **Brightness**, **Contrast_RMS**, **Entropy**, **Laplacian** and **Corners** using Python. We obtained **Contrast**, **Correlation**, **Energy**, **Homogeneity**, **Freq. Factor** and **Edge Density** from [8]. Higher values indicate greater texture for **Contrast_RMS**, **Entropy**, **Laplacian** and **Corners**, **Contrast**, **Freq. Factor** and **Edge Density**. Lower values indicate greater texture for **Energy** and **Homogeneity**; for **Correlation** lower values indicate randomly arranged pixels rather than recognizable textural elements.

3.2 VI-SLAM Evaluation Setup

To study the effect of texture in controlled experiments we implemented the open-source game engine-based VI-SLAM evaluation methodology we developed in [30]. As such we used the ground truth pose in existing SLAM datasets (e.g., [17, 33]) to create camera images in virtual environments, and combined them with the original inertial data to form new VI-SLAM sequences. We created new virtual environments with our textures in Unity 2020.3.14f1 (see Section 3.1), then generated new sequences for different combinations of environments, trajectories and textures (see Section 3.3).

We executed sequences with a state-of-the-art, open-source monocular VI-SLAM algorithm, ORB-SLAM3 [5] (default parameters). We ran them on a desktop PC (Intel i7-9700K CPU, Nvidia GeForce RTX 2060 GPU), using a virtual machine with 4 CPUs and 8GB RAM – computational resources representative of a mobile AR device. To evaluate trajectories we used the toolbox in [38]. We used the same performance metrics as [30]: *median RE*, the median of the translational component of relative error for each trial, and *robustness*, the mean percentage of input frames tracked over all trials.

3.3 VI-SLAM Sequence Configurations

The aim of our VI-SLAM experiments was to examine the effect of texture across a variety of scenarios. To this end we created five sequence configurations, each with a different combination of virtual environment and device trajectory. We created two virtual environments in Unity, *Room* and *Table*. *Room* was a 6m×6m×6m empty room designed to test textures in isolation; for each new sequence a texture was applied to the walls, floor and ceiling. *Table* was designed to replicate the scenario of an AR-assisted assembly or repair task: an 8m×8m×4m factory environment including a robotic arm and a table with a motorbike engine on it. To create each sequence in *Table* the texture was applied to the 1.5m×1.5m table top. Images of these environments can be found at **https://github.com/timscargill/Invisible-Textures/**.

To generate sequences in our environments we used four trajectories from two VI-SLAM datasets. From TUM VI [33] we used *room5*, a trajectory with rapid motion representative of a dynamic AR user, with camera views scanning almost all environment regions. From SenseTime [17] we used *A3*, *A5* and *A6*, trajectories more representative of an AR user engaged in an assembly or repair task, with slower motion and the camera angled downwards throughout. We ran *room5* and *A6* in *Room*, and *A3*, *A5* and *A6* in *Table*, for five sequence configurations: *Room_room5*, *Room_A6*, *Table_A3*, *Table_A5* and *Table_A6*. *Table_A3* starts by focusing on the wall and floor before focusing on the table top; *Table_A5* starts by focusing on the table top before moving to an area of the floor; *Table_A6* remains focused on the table top throughout.

4 VI-SLAM VISUAL TEXTURE EXPERIMENTS RESULTS

Next we present our VI-SLAM texture experiments results, first covering performance variation with texture, then the relationships between texture properties and performance.

4.1 VI-SLAM Performance Variation with Visual Texture

Figure 2 shows VI-SLAM performance variation across textures, with each data point representing results for one texture over 10 trials. For median RE (Figure 2a–e), the greatest variation was observed with the most challenging inertial data, *Room_room5*, with median RE ranging from 2.1cm to 66.6cm – median RE for the worst-performing texture was 3171% of median RE for the best-performing texture, despite

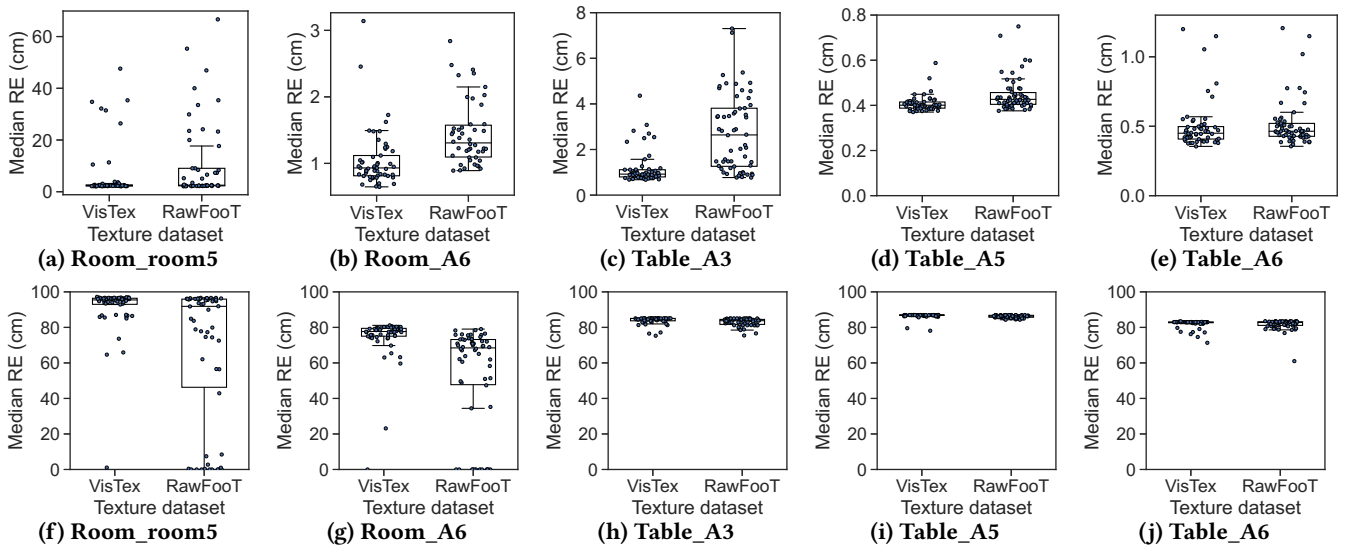


Figure 2: Median RE (a–e) and robustness (f–j) for the five sequence configurations in our VI-SLAM visual texture experiments. We observed notable performance variation across textures in all configurations, with median RE for the worst-performing texture ranging from 200% to 3171% of median RE for the best-performing texture.

high robustness (> 96%) in both cases. In fact, *texture had a large impact on median RE in all configurations*; in the least challenging configuration with the least performance variation, *Table_A5*, median RE for the worst-performing texture (7.5cm) was still 203% of median RE for the best-performing texture (3.7cm). This illustrates the impact of texture selection, and how errors occur even while tracking functions.

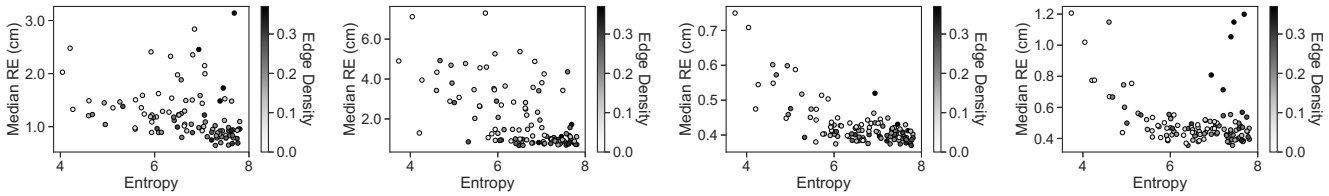
For robustness (Figure 2f–j), performance variation was greater in *Room* than *Table*. In *Room_room5* eight textures failed completely on all trials (robustness = 0%) and seven further textures resulted in robustness < 10%; in *Room_A6* 13 textures failed on all trials. As with median RE, greater performance variation was observed in the RawFooT dataset due to the presence of more low-texture images than in VisTex. Our robustness results demonstrate that *in scenarios where camera views contain a single texture (e.g., a wall, a floor), the choice of that texture is critical for ensuring tracking functions at all.* Environment designers should avoid large low-texture regions to avoid these tracking losses, during which virtual content will be unreliable or disappear completely.

4.2 Relationships between Visual Texture Properties and VI-SLAM Performance

First we consider robustness, i.e., how one selects textures to ensure tracking functions. The most common reason for low robustness in *Room* was low **Brightness**. In *Room_room5* the five textures with the lowest **Brightness** (< 0.14) failed on all trials, while six of the remaining 10 textures with robustness < 10% had **Brightness** < 0.25. In *Room_A6*, the six textures with the lowest **Brightness** (< 0.18) failed on all trials, while five of the remaining seven textures which failed had **Brightness** < 0.25. Because the texture covers all surfaces in *Room*, the light illuminating a texture is largely determined by how much light it reflects (the **Brightness**); hence textures with low **Brightness** result in texture not being visible due to low illuminance. The only low **Brightness** textures with high robustness had high **Laplacian** – strong edges visible even at low illuminance. Other textures which resulted in robustness < 10% had low **Contrast_RMS**. Our results indicate that environments with low illuminance require textures with high edge strength (**Laplacian**), or alternatively, textures with lower edge strength require higher illuminance.

To examine the relationships between texture properties and tracking error we calculated the distance correlation coefficients [35] (chosen to detect non-linear and non-monotonic

Figure 3: Distance correlation coefficients between each of our texture characterization metrics and median RE in each of our five sequence configurations.



(a) Room_A6 (b) Table_A3 (c) Table_A5 (d) Table_A6

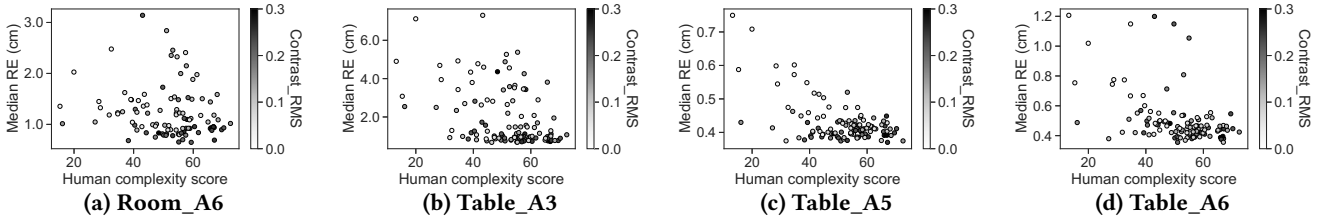
Figure 4: Median RE vs. Entropy for the four SenseTime sequence configurations in our VI-SLAM texture experiments. We observed a negative correlation for all configurations, but notable outliers were textures with low Edge Density (Room_A6 and Table_A3) and the five highest Edge Density values (Room_A6 and Table_A6).

relationships) between our characterization metrics (Section 3.1) and median RE. Figure 3 shows a heatmap of these correlation coefficients for each sequence configuration. In general the coefficients for *Room_room5* are the lowest because the challenging inertial data contributes to error to a greater degree. In trials with high RE in *Room_room5* the pose estimate diverged dramatically from the ground truth at similar points in the trajectory, characterized by rapid motion combined with challenging environment regions, e.g., dark corners or views of a small area with low texture. We note that textures with high **Brightness** and **Edge Density**, such that visible texture covers all environment regions, reduces the likelihood of large errors, but characterization of challenging regions is a topic of our ongoing work.

Taking next configurations with SenseTime trajectories (*Room_A6*, *Table_A3*, *Table_A5* and *Table_A6*), Figure 3 shows higher correlation coefficients between characterization metrics and median RE than for *Room_room5*. For *Room_A6*, the strongest correlation is for **Corners**, and textures with low **Corners** resulted in higher error; in *Room* the lack of texture from other objects means we rely more on the chosen texture for tracking. For *Table_A3* and *Table_A5*, the strongest correlations are with **Entropy** and **Energy** – Figure 4 shows how in general greater **Entropy** (complexity) results in lower median RE. However, these correlation coefficients are lower for *Room_A6* and *Table_A6* due to the effect of other factors. For *Room_A6* a number of RawFooT textures with high **Entropy** but low **Edge Density** (< 0.1) resulted in high error (also the cause of outliers for *Table_A3*); as we observed for *Room_room5*, if camera views contain only a small part of a texture without edges, then tracking quality is degraded. For both *Room_A6* and *Table_A6* the five VisTex textures with the highest **Edge Density** (> 0.32) had high **Entropy** but resulted in high error. Our hypothesis is that limited camera resolution combined with fine texture makes feature matching less reliable, and we are investigating this in our ongoing work. Overall, our experiments show that even when environments contain other textured objects, choosing surface textures with sufficient **Entropy** and **Edge Density** should be priorities for designers of environments that host AR.

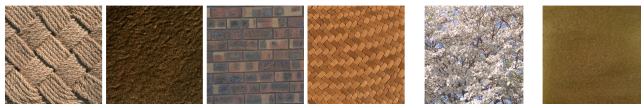
5 VI-SLAM PERFORMANCE AND HUMAN PERCEPTION OF TEXTURE

Our first motivation for studying the relationship between VI-SLAM performance and human perception of texture complexity was to determine whether the latter is a good predictor of the former. AR environment design guidelines (e.g., [23]) are based on the assumption that human perception of complexity is positively correlated with VI-SLAM performance, such that choosing a texture higher in perceived complexity results in more accurate positioning of virtual content. However, we see from Figure 5 that this is not always the case: *textures with high human complexity scores can result in high median RE, while some textures perceived as low in complexity result in near-optimal VI-SLAM performance.* This indicates that we cannot rely on human perception of complexity to predict VI-SLAM performance, and prompts the development of other prediction methods.



(a) Room_A6 (b) Table_A3 (c) Table_A5 (d) Table_A6

Figure 5: Median RE vs. human complexity scores (from [8]) for textures in our experiments. If an environment designer chooses a texture with a high human complexity score (> 60), it can still result in sub-optimal VI-SLAM performance, while some textures with low human complexity scores (< 40) provide near-optimal performance.



(a) (b) (c)

Figure 6: (a) Textures with low median RE and low perceived complexity for *Room_A6, Table_A3, Table_A5* and *Table_A6*; (b) best *Table* texture prioritizing VI-SLAM; (c) best *Table* texture prioritizing human factors.

Our second motivation was the study of what we term *invisible textures*, textures with near-optimal VI-SLAM performance (median RE < 110% of the best-performing texture) but perceived by humans as low in complexity. The textures with the lowest human complexity score among those with near-optimal VI-SLAM performance for *Room_A6*, *Table_A3*, *Table_A5* and *Table_A6* are shown in Figure 6a. These textures had sufficient **Entropy** (> 6.0) and **Edge Density** (> 0.1), and for the *Table* configurations, low **Contrast_RMS** (< 0.11). No single texture had near-optimal VI-SLAM performance for all *Table* configurations; the texture in Figure 6b, with a somewhat high complexity score (51) and **Contrast_RMS** (0.20) was closest to meeting these criteria. This illustrates the challenge of selecting 'simple' textures that result in near-optimal VI-SLAM performance for a variety of trajectories.

Indeed we note from [37] that the contrast in the textures in Figure 6a and 6b may still cause visibility problems on AR devices with optical see-through displays. We therefore further restrict our definition of invisible textures to those with lower **Contrast_RMS** (< 0.05), as well as a human complexity score < 50. Prioritizing these human factors, the texture with the best VI-SLAM performance is shown in Figure 6c; this fine, low-contrast texture results in a maximum 0.2cm increase in median RE compared to the best-performing texture in the *Table* configurations, and a 1.0cm increase in the *Room* configurations, acceptable in many scenarios.

This ability to achieve good VI-SLAM performance with low contrast textures less visible to humans is explained by a fundamental difference between machine and human perception. Texture visibility for humans is described by the *contrast sensitivity function* [4], with decreasing pixel intensity difference (contrast) resulting in decreasing visibility. Detection of recognizable features in VI-SLAM on the other hand is based on a fixed and relatively low threshold of pixel intensity difference [28]. This insight, along with our experimental results, indicate that *as long as we select a texture with sufficient **Entropy** and **Edge Density**, we can minimize human perception of that texture by reducing contrast down to the threshold required for feature extraction in VI-SLAM, with relatively little impact on VI-SLAM performance.* Moreover, human contrast sensitivity decreases both at high spatial frequencies (i.e., high **Edge Density**), and in certain areas of the visual field [14, 15]. This indicates that we may be able to further reduce the perceptual impact of low-contrast textures to the point where they are truly invisible to the naked eye, either by increasing **Edge Density** or placing texture away from the center of a user's gaze. We are investigating this and the influence of other factors such as illuminance, viewing angle [2] and camera image resolution in our ongoing work.

The promise of even near-invisible textures raises exciting prospects for optimizing AR environments using distributed systems. Just as IoT light bulbs can be used to optimize environment illuminance [31], we envision employing ambient displays to optimize environment texture. As shown in Figure 1, displaying a texture which prioritizes human factors (Figure 6c) makes interpreting virtual content easier than a texture prioritizing VI-SLAM performance (Figure 6b). Motivated by the asymmetries of human contrast sensitivity across the visual field, these textures should adapt in real time to the pose and eye movements of AR users (similar to [6]), as well as their cognitive states, current task, and virtual content properties. This in turn prompts the development of systems and networks capable of consolidating and acting upon contextual data with sufficiently low latency (e.g., edge architectures [21, 39]), and which adapt texture optimization techniques to current network conditions.

6 CONCLUSIONS AND FUTURE WORK

In this paper we compared machine and human perception of environment texture for AR, revealing evidence for *invisible textures*, that result in good VI-SLAM performance with minimal impact on human perception of virtual content. Arising from the differences between VI-SLAM feature detection and the human *contrast sensitivity function*, the characteristics of these textures inform environment design for AR, including IoT-enabled texture optimization systems using ambient displays. In future work we will develop and test these systems, which will automatically display textures according to the current context of AR interactions. As part of this work, we will create and evaluate invisible textures manufactured to our specifications, investigate methods of adapting them in real time, and explore other types of ambient displays, e.g., light projections [18, 19]. We will also examine the effect of texture for other VI-SLAM algorithms, and determine optimal VI-SLAM parameters for performance and resource management with different textures.

ACKNOWLEDGMENTS

This work was supported in part by NSF grants CSR-1903136, CNS-1908051, and CNS-2112562, NSF CAREER Award IIS-2046072, a Meta Research Award, and a CISCO Research Award.

REFERENCES

[1] Ashrant Aryal, Burcin Becerik-Gerber, Francesco Anselmo, Shawn C Roll, and Gale M Lucas. 2019. Smart desks to promote comfort, health, and productivity in offices: A vision for future workplaces. *Frontiers in Built Environment* (2019), 76.

[2] Peter GJ Barten. 2003. Formula for the contrast sensitivity of the human eye. In *Proceedings of SPIE Image Quality and System Performance*.

[3] Michael Burri, Janosch Nikolic, Pascal Gohl, Thomas Schneider, Joern Rehder, Sammy Omari, Markus W Achtelik, and Roland Siegwart. 2016. The EuRoC micro aerial vehicle datasets. *The International Journal of Robotics Research* 35, 10 (2016), 1157–1163.

[4] Fergus W Campbell and John G Robson. 1968. Application of Fourier analysis to the visibility of gratings. *The Journal of Physiology* 197, 3 (1968), 551.

[5] Carlos Campos, Richard Elvira, Juan J Gómez Rodríguez, José MM Montiel, and Juan D Tardós. 2021. ORB-SLAM3: An accurate open-source library for visual, visual–inertial, and multimap SLAM. *IEEE Transactions on Robotics* 37, 6 (2021), 1874–1890.

[6] Jaewon Choi, HyeonJung Park, Jeongyeup Paek, Rajesh Krishna Balan, and JeongGil Ko. 2019. LpGL: Low-power graphics library for mobile AR headsets. In *Proceedings of ACM MobiSys*.

[7] Silvia Corchs, Francesca Gasparini, and Raimondo Schettini. 2013. Grouping strategies to improve the correlation between subjective and objective image quality data. *IS&T/SPIE Electronic Imaging* 8653 (2013), 104–111.

[8] Silvia Elena Corchs, Gianluigi Ciocca, Emanuela Bricolo, and Francesca Gasparini. 2016. Predicting complexity perception of real world images. *PloS One* 11, 6 (2016), e0157986.

[9] Yuan Feng, Suihuai Yu, Dirk Van De Mortel, Emilia Barakova, Jun Hu, and Matthias Rauterberg. 2019. LiveNature: Ambient display and social robot-facilitated multi-sensory engagement for people with dementia. In *Proceedings of ACM DIS*.

[10] Anna V Fisher, Karrie E Godwin, and Howard Seltman. 2014. Visual environment, attention allocation, and learning in young children: When too much of a good thing may be bad. *Psychological Science* 25, 7 (2014), 1362–1370.

[11] Joseph L Gabbard, J Edward Swan, Deborah Hix, Si-Jung Kim, and Greg Fitch. 2007. Active text drawing styles for outdoor augmented reality: A user-based study and design implications. In *Proceedings of IEEE VR*.

[12] James Garforth and Barbara Webb. 2019. Visual appearance analysis of forest scenes for monocular SLAM. In *Proceedings of IEEE ICRA*.

[13] Xiaoying Guo, Chie Muraki Asano, Akira Asano, Takio Kurita, and Liang Li. 2012. Analysis of texture characteristics associated with visual complexity perception. *Optical Review* 19 (2012), 306–314.

[14] Nina M Hanning, Marc M Himmelberg, and Marisa Carrasco. 2022. Presaccadic attention enhances contrast sensitivity, but not at the upper vertical meridian. *Iscience* 25, 2 (2022), 103851.

[15] Marc M Himmelberg, Jonathan Winawer, and Marisa Carrasco. 2020. Stimulus-dependent contrast sensitivity asymmetries around the visual field. *Journal of Vision* 20, 9 (2020), 18–18.

[16] Imaging and Vision Laboratory, Department of Informatics, Systems and Communication, University of Milano-Bicocca. 2023. RawFooT DB: Raw Food Texture Database. http://www.ivl.disco.unimib.it/minisites/rawfoot/.

[17] Li Jinyu, Yang Bangbang, Chen Danpeng, Wang Nan, Zhang Guofeng, and Bao Hujun. 2019. Survey and evaluation of monocular visual-inertial SLAM algorithms for augmented reality. *Virtual Reality & Intelligent Hardware* 1, 4 (2019), 386–410.

[18] Brett Jones, Rajinder Sodhi, Michael Murdock, Ravish Mehra, Hrvoje Benko, Andrew Wilson, Eyal Ofek, Blair MacIntyre, Nikunj Raghuvanshi, and Lior Shapira. 2014. RoomAlive: Magical experiences enabled by scalable, adaptive projector-camera units. In *Proceedings of ACM UIST*.

[19] Brett R Jones, Hrvoje Benko, Eyal Ofek, and Andrew D Wilson. 2013. IllumiRoom: Peripheral projected illusions for interactive experiences. In *Proceedings of CHI*.

[20] Murad Khan, Bhagya Nathali Silva, and Kijun Han. 2016. Internet of Things based energy aware smart home control system. *IEEE Access* 4 (2016), 7556–7566.

[21] Luyang Liu, Hongyu Li, and Marco Gruteser. 2019. Edge assisted real-time object detection for mobile augmented reality. In *Proceedings of ACM MobiCom*.

[22] Mathworks. 2023. Graycoprops: Properties of gray-level co-occurrence matrix. https://www.mathworks.com/help/images/ref/graycoprops.html.

[23] Microsoft. 2023. HoloLens environment considerations. https://learn.microsoft.com/en-us/hololens/hololens-environment-considerations.

[24] Daniel Minoli, Kazem Sohraby, and Benedict Occhiogrosso. 2017. IoT considerations, requirements, and architectures for smart buildings—Energy optimization and next-generation building management systems. *IEEE Internet of Things Journal* 4, 1 (2017), 269–283.

[25] MIT Media Lab. 2023. Vision texture homepage. https://vismod.media.mit.edu/vismod/imagery/VisionTexture/.

[26] Xukan Ran, Carter Slocum, Maria Gorlatova, and Jiasi Chen. 2019. ShareAR: Communication-efficient multi-user mobile augmented reality. In *Proceedings of ACM HotNets*.

[27] Ruth Rosenholtz, Yuanzhen Li, and Lisa Nakano. 2007. Measuring visual clutter. *Journal of Vision* 7, 2 (2007), 17–17.

[28] Edward Rosten and Tom Drummond. 2006. Machine learning for high-speed corner detection. In *Proceedings of ECCV*.

[29] Francesco Salamone, Lorenzo Belussi, Cristian Currò, Ludovico Danza, Matteo Ghellere, Giulia Guazzi, Bruno Lenzi, Valentino Megale, and Italo Meroni. 2018. Integrated method for personal thermal comfort assessment and optimization through users' feedback, IoT and machine learning: A case study. *Sensors* 18, 5 (2018), 1602.

[30] Tim Scargill, Ying Chen, Nathan Marzen, and Maria Gorlatova. 2022. Integrated design of augmented reality spaces using virtual environments. In *Proceedings of IEEE ISMAR*.

[31] Tim Scargill, Sangjun Eom, Ying Chen, and Maria Gorlatova. 2023. Ambient intelligence for next-generation AR (to appear in the Springer Handbook of the Metaverse). arXiv:2303.12968

[32] Tim Scargill, Gopika Premsankar, Jiasi Chen, and Maria Gorlatova. 2022. 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)*.

[33] David Schubert, Thore Goll, Nikolaus Demmel, Vladyslav Usenko, Jörg Stückler, and Daniel Cremers. 2018. The TUM VI benchmark for evaluating visual-inertial odometry. In *Proceedings of IEEE/RSJ IROS*.

[34] C. E. Shannon. 1948. A mathematical theory of communication. *The Bell System Technical Journal* 27, 3 (1948), 379–423.

[35] Gábor J Székely, Maria L Rizzo, and Nail K Bakirov. 2007. Measuring and testing dependence by correlation of distances. *Annals of Statistics* 35, 6 (2007), 2769–2794.

[36] Juncai Yao Yao and Guizhong Liu. 2018. Bitrate-based no-reference video quality assessment combining the visual perception of video contents. *IEEE Transactions on Broadcasting* 65, 3 (2018), 546–557.

[37] Lili Zhang and Michael J Murdoch. 2021. Perceived transparency in optical see-through augmented reality. In *Proceedings of IEEE ISMAR-Adjunct*.

[38] Zichao Zhang and Davide Scaramuzza. 2018. A tutorial on quantitative trajectory evaluation for visual (-inertial) odometry. In *Proceedings of IEEE/RSJ IROS*.

[39] Yiqin Zhao and Tian Guo. 2021. Xihe: a 3D vision-based lighting estimation framework for mobile augmented reality. In *Proceedings of ACM MobiSys*.