ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XI-4-2026
XXV ISPRS Congress 2026 “From Imagery to Understanding”, Commission IV, 4–11 July 2026, Toronto, Canada

What Features of the Street Influence Visual Walkability? An Innovative Approach Using Cinematic Virtual Reality

Chongan Wang[1,*], Florian Massuyeau[1], Vincent Tourre[1], Thomas Leduc[1], Myriam Servières[1]

[1] Nantes Université, ENSA Nantes, École Centrale Nantes, CNRS, AAU-CRENAU, UMR 1563, F-44000 Nantes, France.
name.surname@crenau.archi.fr

Keywords: Visual Walkability, 360° Video, Semantic Segmentation, Pedestrian, Eye-tracking, CVR.

Abstract

We present a new method for assessing visual walkability using 360° videos and an eye-tracking in Cinematic Virtual Reality (CVR). Visual walkability refers to the walkability perceived by pedestrians through visual stimuli in the urban environment. Our method uses semantic segmentation, viewport exposure, gaze measures, and a custom walkability questionnaire, enabling comparison between scene content, participant’s viewport, and their gaze focus. The 10 videos used, including 2 calibration videos, exhibit distinct semantic characteristics, validated by segmentation analysis. Analysis of the 35 participants’ responses shows that walkability ratings at the video level correlate with several environmental parameters (e.g., road, sidewalk, sky) consistent with previous studies. However, these parameters do not have a similar influence in gaze-based visual attention analysis within the CVR setting, suggesting that CVR attention would requiere further work. Furthermore, our results suggest that unexpected semantic classes may also play a role in perceived walkability and should be considered exploratory pending further validation. This paves the way for further research on using CVR as an assessment tool for visual walkability and for developing methodological guidance on which visual cues are robust across measures (content/viewport).

1. Introduction

Walking daily, whether for work or leisure, is a simple and effective form of exercise that has a positive impact on physical and mental health (Morris and Hardman, 1997, White et al., 2019). Researchers are therefore interested in measuring the walkability of places. In traditional walkability research, field experiments are often used to obtain accurate data, but they are time-consuming and difficult to replicate (Badland et al., 2010). Today, Street View Imagery (SVI) is increasingly used as a helpful tool in studies (Chen and Biljecki, 2023, Huang et al., 2023). Researchers can perform image segmentation with trained Convolutional Neural Network (CNN) models to quantitatively analyze urban environments. Although datasets such as Place Pulse 2.0 (Dubey et al., 2016), which can be used to train these CNN models, provide a practical solution, there are some limitations with these methods : 1/ They use mono-directional images, thus removing the contextual information (Beaucamp et al., 2025) 2/ These datasets were collected mainly from vehicles, so the viewpoints are focused on drivers on the roadway rather than pedestrians on the sidewalks (Ito et al., 2024) 3/ Much of the research has used a single static image, neglecting the pedestrian movement, and has used mono-directional images, further removing contextual information.

Our study attempts to overcome these limitations in measuring the factors that influence visual walkability in urban spaces by using 360° videos of urban walks within a virtual reality environment. We combine the advantages of both on-site and off-site experiments by partially preserving the conditions of on-site experiments while ensuring repeatability in the laboratory. The research questions of this study are:

- **RQ1** What are the most important viewing directions in a 360° Cinematic Virtual Reality (CVR) urban walk video with a linear path?
- **RQ2** What object classes do participants look at?
- **RQ3** What correlations exist between the walkability score and the experimental measurements (video content, viewport direction, and visual attention)?

* Corresponding author

To achieve these goals, we proceeded in four steps: S1) analysis of participants’ gaze directions; S2) semantic segmentation of the videos; S3) analysis of the correlations between the segmentation results and the walkability scores; and S4) analysis of the correlations between the visual attention of participants and the walkability scores. Our contributions to the use of CVR to analyze visual walkability are as follows:

- We developed an experimental protocol to assess visual walkability in CVR and validated it for some parameters through our experiment.
- We demonstrate that the analysis can be limited to three frames (left, front, right), corresponding to a 270° horizontal FOV (**RQ1**).
- The CVR produced the same results when analyzing environmental characteristics, specifically road, sidewalk, and sky, as current state-of-the-art methods (**RQ2**).
- The correlation determined by participants’ viewport direction differed from that determined by their gaze (**RQ3**).
- We found a positive correlation with the sidewalk and a negative correlation with the road in the visual attention analysis (**RQ3**).

The remaining sections of this article are organized as follows: Section 2 reviews recent studies and their limitations; Section 3 describes our experimental methodology; Section 4 presents the experimental results; and Section 5 analyzes these results, discusses our findings, and offers an outlook for future work.

2. State of the Art

2.1 Visual Walkability

The topic of walkability is receiving increasing attention, as promoting walking not only addresses public health concerns but also helps reduce reliance on inactive modes of transportation, which is crucial in the context of climate change (WHO, 2020). However, walkability is a broad concept that describes the potential of a place or urban configuration to encourage or discourage walking. Dovey (Dovey and Pafka, 2020) describes walkability as an interdisciplinary field that encompasses public health, social equity, economic productivity and climate change. Walkability is a multisensory experience; however, since vision is the predominant sense among sighted people (Stokes and Biggs, 2014), our research initially focused on “visual walkability”. Zhou (Zhou et al., 2019) described visual walkability as a subjective concept reflecting environmental characteristics that influence how a person visually perceives a place as walkable.

Several factors can positively or negatively influence walkability. (Lee et al., 2022), (Li et al., 2020) and (Ma et al., 2021) found that sidewalk width and the presence of greenery positively influence visual walkability. Openness is another positive factor identified by the researchers (Li et al., 2020, Liao et al., 2022, Ren et al., 2023), which can be quantified by the proportion of visible sky. In contrast, traffic, vehicles, buildings enclosures, and roads are perceived as negative factors that could reduce the willingness of pedestrians to use a route (Huang et al., 2023, Lee et al., 2022, Ma et al., 2021).

According to the current state of the art, the most influential elements that affect visual walkability are buildings, sky, roads, cars, vegetation, and sidewalks (Huang et al., 2023, Lee et al., 2022, Li et al., 2020, Liao et al., 2022, Ma et al., 2021, Ren et al., 2023), all of which we examined as conditions in our experiment.

2.2 Walkability Assessment in Virtual Reality (VR)

Upon reviewing the state of the art, we found that few papers use virtual reality as a tool to assess walkability. Nakamura (Nakamura, 2021) presented participants with 360° videos on both a head-mounted display (HMD) and a screen to compare the results of these two different media. However, he only concluded that people using HMDs tended to rate the street scene with higher scores. Kim et al. (Kim and Lee, 2022) also performed a comparison between CVR, SVI, and an on-site experiment in 2022. However, since the data they used was from a 2016 experiment, they questioned the reliability of CVR technology, noting that their analysis was based on an “outdated experiment”. Li et al. (Li et al., 2022) and Huang et al. (Huang et al., 2023) collected panoramic photos from Google Street View as well as their own photographs and then presented them to participants using a HMD. They concluded that VR can mitigate the difference between on-site experience and browser-based SVI. The study by Birenboim et al. (Birenboim et al., 2021) tested the use of a walking simulator in VR to assess walkability in a pilot trial. Although the study included only four participants and the virtual environment was less realistic, the authors concluded that immersive VR can be a powerful tool for examining walkability.

2.3 Cinematic Virtual Reality for urban space evaluation

As described by Lee et al. (Lee and Kim, 2021), CVR also known as Recorded VR offers a safe, convenient, and realistic way to experience various situations. Cinnamon and Jahui (Cinnamon and Jahiu, 2023) conducted a literature review on the use of 360° videos for virtual location-based research. The review included 69 articles in which 360° videos were used in the field of built environment and land use. However, only one study, by Nakamura (Nakamura, 2021), specifically investigated walkability.

3. Methodology

3.1 Design Rationale

Based on the current state of the art, we used 360° videos as a dynamic medium in a virtual reality environment to provide participants with a more immersive assessment experience and to measure the influence of the urban context on the walkability score. To simulate the experience of walking, participants used the arm-swinging method during the experiment to control the video (Cherni et al., 2020, McCullough et al., 2015). Since our focus is on visual walkability rather than true walkability, the arm-swinging method appears to be one of the most appropriate approaches to facilitate reproducibility by avoiding the use of complex equipment. The explanatory variables we examined included the proportions of buildings, vegetation, sky, sidewalks, cars, and roads in the streetscape (Huang et al., 2023, Lee et al., 2022, Li et al., 2020, Liao et al., 2022, Ma et al., 2021, Ren et al., 2023). These factors are considered the primary influences on walkability. During the VR experiment, we recorded participants’ head and eye movements. In addition, they completed a questionnaire on demographics, walkability, sense of presence (Makransky et al., 2017), and cybersickness (Kim et al., 2018). Our explained variables were the walkability questionnaire results.

3.2 Panoramic Video Dataset

Since we did not have an available dataset, we had to record videos tailored to our specific needs. First, we filmed 10 panoramic videos with the Insta360® Pro 2 camera to create a labeled dataset. Since videos in the field are less controllable than 3D models, we could only attempt to reduce the number of variables by carefully selecting the shooting location, as described in more detail in Section 4.3. The videos were filmed handheld, using software stabilization, at eye level and at a normal walking speed. They were recorded with stereoscopic sound at 60 frames per second and a resolution of $7,680 \times 3,840$. All videos were filmed on straight roads to prevent participants from choosing their walking direction within the video, as sinuosity could influence perceived walkability. Video1 and Video2 were filmed on the same road but in opposite directions. The same applies to Video3 and Video4. Video5 and Video6 were recorded on the same path in a city park, with Video5 filmed in summer and Video6 in winter. Winter scene videos show trees without leaves. Video7 and Video8 were selected primarily for their unique characteristics: Video7 features a sidewalk in the middle of the street with parking lots on both sides, while Video8 was shot at the same location but on the side of the street adjacent to buildings. Video9 and Video10 were recorded on relatively busy streets during rush hour. Figure 1 shows a single frontal frame from each video.

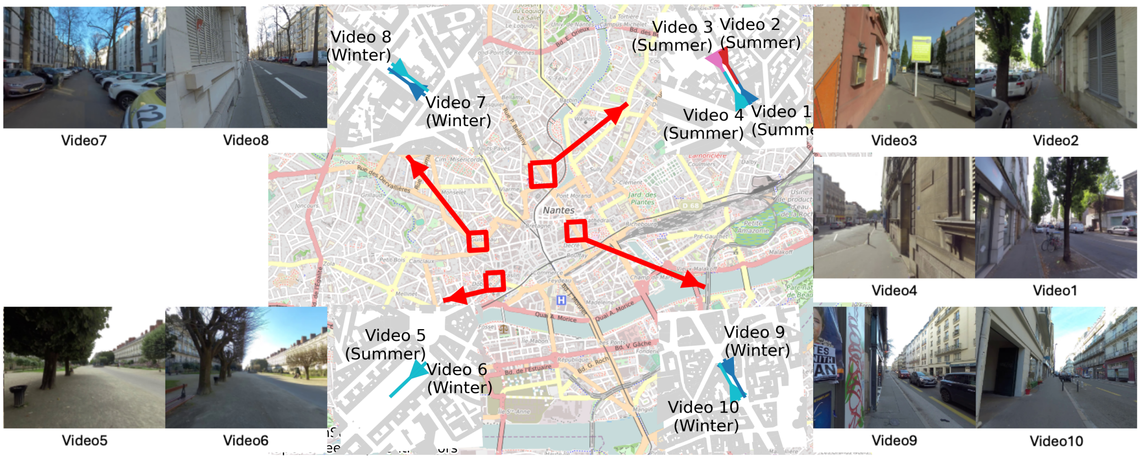

Figure 1. Front-facing frames in the walking direction for each video. Video5 depicts the “High walkability” road and Video9 depicts the “Low walkability” road.

3.3 Pilot Study

A pilot study with seven participants, all with basic knowledge of urban planning, was conducted to determine the optimal video duration and locomotion method for an immersive VR experience. Participants watched 10 full-length videos and paused each when they felt they had enough information to evaluate the environment. The average viewing time was approximately one minute, which was adopted as the fixed duration for the main experiment. Participants also performed arm swinging to simulate walking; the video paused whenever they stopped. This approach aimed to foster a realistic assessment of walkability. Feedback indicated that arm swinging reduced cybersickness and encouraged movement synchronization with the video. Consequently, it was selected as the locomotion method, while playback speed was kept constant, as changing the speed could cause unnatural movements and potentially reduce the sense of presence.

3.4 Experiment

This study has been approved by the ethics committee of Nantes Université, with reference number 04062024-1. The approved protocol included partial disclosure of eye-tracking to limit demand characteristics; participants were fully debriefed at the end of the session and were reminded of their right to withdraw their data.

3.4.1 Participants This experiment involved 37 participants, of whom 35 provided valid data. Among these 35 valid participants, twenty were men (57%), fourteen were women (40%), and one preferred not to specify (3%). Their ages ranged from 19 to 56 years, with an average age of 30.5 years (SD = 8.7). Participants were recruited by email from two different cities. Twenty-one participants lived in the city where the videos were recorded, while the remaining 15 lived in another country and had never visited that city.

3.4.2 Procedure Upon arrival, participants were required to sign a consent form before the experiment began. This form provided general information about the experiment and outlined the potential risks associated with VR. The experimental procedure was then explained to participants in detail, without disclosing that eye-tracking data would be collected.

First, participants began with the calibration of the HMD to familiarize themselves with the headset. The program was then initiated. The first video was a 20-second static demonstration filmed in a different location without movement, serving as a training session. Then, in the main experiment, each walking video was presented for a fixed duration of 60 seconds. The second and third videos were identical for all participants to provide a sort of common baseline for the concept of walkability. Two relatively iconic situations were presented: one (Video5) in which there is consensus that the premises are walkable, and the other (Video9) in which there is consensus that they are not. The two calibration videos presented at the beginning of the session aimed to provide participants with a shared reference for the rating scale. This step was intended to reduce inter-participant differences in scale interpretation. These two calibration videos were also followed by the same five-item Likert questionnaire, using the same wording and scale as the remaining videos.

The remaining eight videos were shown in random order to avoid bias. Each video was followed by a questionnaire containing five questions derived from the walking needs hierarchy, which emphasizes that pleasurable walking experiences can only be achieved once basic needs, such as feasibility, accessibility, safety, and comfort, are fulfilled (Alfonzo, 2005):

- This path is accessible.
- This path is safe.
- This path is comfortable.
- This path is pleasurable.
- I would like to take this path as daily commute.

The questions were all presented on a five-point Likert scale. The five items were presented immediately after each video, in the same order as listed below, and participants completed the ratings before proceeding to the next video. Since the videos were filmed in an urban environment, we assumed that feasibility was already established. We replaced the original question at the end with the statement “I would like to take this path as daily commute” to better capture the subjective assessment of walkability.

At the end of the experiment, participants were asked to complete a questionnaire covering demographics, sense of presence, and cybersickness to evaluate the experiment. After all questionnaires were completed, participants were given the opportunity to review their eye-tracking results and provide feedback on these results or on the experiment itself.

4. Results

4.1 Questionnaires Results

According to Pavic (Pavic et al., 2023), physical presence and social presence can influence individual’s emotional responses. We applied the Multimodal Presence Scale (MPS) (Makransky et al., 2017) to assess the degree of presence, obtaining an average score of 3.8/5 (SD=0.89) for physical presence and 3.43/5 (SD=1.12) for social presence. Due to the limitations of the CVR, we only obtained relatively high scores for physical realism (M=4.17/5, SD=0.69) and the sense of being in a natural environment (M=4.03/5, SD=0.50). We also used the Virtual Reality Sickness Questionnaire (VRSQ) (Kim et al., 2018) to measure cybersickness, obtaining a moderately good score of M=23.6/100 (SD=16.4).

Although our primary focus was on the walkability question, the other questions related to the hierarchy of walking needs remain valuable for gaining further insight into how participants rated their experiences. The results of the walkability assessment are shown in Figure 2. There was no significant relationship between gender ($p>0.22$ for all videos) or familiarity ($p>0.26$ for all videos) and walkability ratings. However, we found strong positive correlations between Video1 and Video2 ($r=0.72$, $p<0.01$) and between Video3 and Video4 ($r=0.74$, $p<0.01$), each filmed at the same location. This suggests that walking in opposite directions does not affect pedestrian walkability ratings, at least for our study areas.

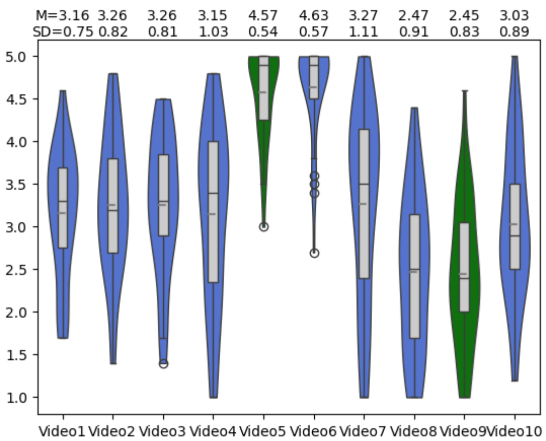

Figure 2. Walkability scores for the different videos. The results for the two calibration videos are highlighted in green.

4.2 Gaze Direction Analysis

To prevent distortion during video analysis, the equirectangular projection was divided into six 90-degree field-of-view frames: front, back, left, right, top, and bottom. The back frame was excluded, and only half of the bottom 90-degree field of view was retained to avoid capturing the camera handler and to reduce bias in the segmentation results for the “Person” class. Figure 3 presented the perspective images that we would obtain from the 5-frame clipping method.

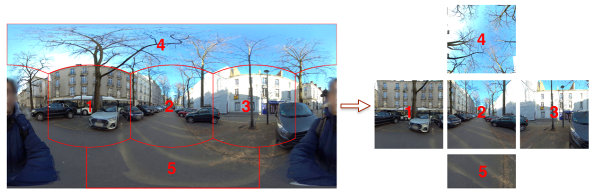

Figure 3. Illustration of 5-frame clipping method.

We analyzed all visual attention data obtained from the eye-tracking phase, which comprised more than 1.3 million gaze points. We then projected these points onto an equirectangular map, as shown in Figure 4. Our results indicated that 93.1% of participants’ visual attention was concentrated within three frames: 90 degrees to the left, directly ahead, and 90 degrees to the right along the walking direction. Notably, 79.1% of visual attention was focused on the front frame, aligning with the typical walking experience. Thus, participants directed their gaze forward more than three quarters of the time, but the sides must also be taken into account. Similar to Figure 4, we have analyzed participants’ viewport coverage (Figure 5), and found that 87.2% of viewport coverage was in the three frames: 10.4% for left, 69.3% for center, 7.5% for right. Based on these findings, we chose to concentrate our analysis on these three frames: front, right, left (**RQ1**).

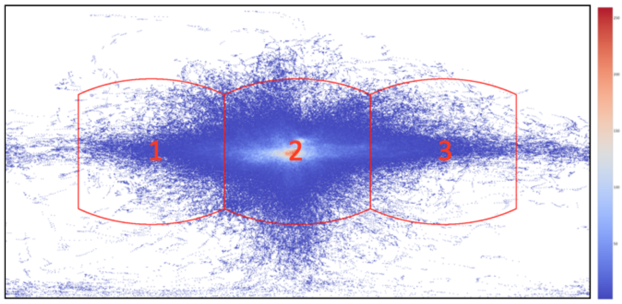

Figure 4. Heatmaps generated from participant’s eye-tracking result represented on an equirectangular frame.

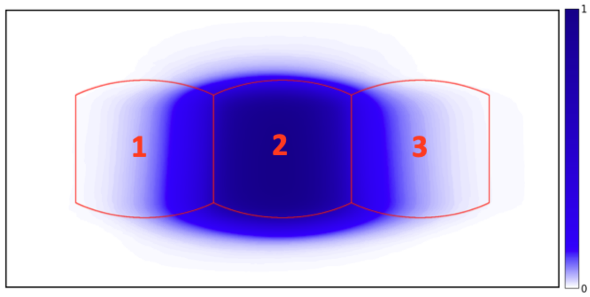

Figure 5. Heatmaps generated with participant’s viewport coverage result represented on an equirectangular frame.

4.3 Video Analysis

For the remainder of the paper, and to avoid biasing our analyses with calibration data, we exclude Video5 and Video9. To correlate the content of the urban environment with participants’ perceived walkability scores, we extracted the characteristic elements of each environment using semantic segmentation. We performed semantic segmentation on every frame of each 60-second video recorded at 60 Hz (3,600 frames per video; 28,800 frames across the 8 videos). As described in Section 4.2, we applied segmentation only to the portions of the videos corresponding to the three main viewing directions. This approach not only reduces computational effort but also improves segmentation accuracy, since current semantic segmentation models are designed for perspective images rather than equirectangular ones. We employed Mask2Former (Cheng et al., 2022) with a pre-trained network on the Mapillary dataset (Neuhold et al., 2017), which demonstrates superior performance in pavement segmentation compared to other solutions. Because our data consisted of videos, we performed segmentation on each frame. We then calculated the average pixel ratio of the identified elements to characterize each video. Details for each video are provided in Table 1. We report the pixel ratios for six classes of interest, selected based on the state of the art discussed in Section 2.

A matrix based on the Jensen–Shannon distance (JSD) was computed for the viewport directions. Each viewport direction is represented as a 65-dimensional vector (corresponding to the 65 semantic classes of the Mapillary Vistas dataset) evolving over time. The Jensen–Shannon distance — defined as the square root of the Jensen–Shannon divergence (Lin, 1991) — is a true metric that quantifies the dissimilarity between two probability distributions. Higher JSD values indicate greater differences between distributions. Hierarchical agglomerative clustering was then performed using the JSD matrix, and the resulting dendrogram is shown in Figure 6. The dendrogram was cut to produce eight clusters (indicated by the horizontal red dashed line), and the branches were colored according to this eight-cluster scheme. Independently, the group labels corresponding to each viewport were colored based on the video labels (cluster1 to cluster8). We observed that the clustering derived from the JSD matrix closely corresponds to the distinct video patterns. This indicates that the eight videos are clearly differentiated and exhibit significantly different semantic segmentation patterns. Consequently, each individual viewport can be unambiguously associated with a specific video group, regardless of the participant's gaze direction.

4.4 Correlation Analysis

Frame Analysis We present the Pearson correlation between the proportion of object classes in the three selected frames aggregated over time and the participants' scores averaged within each of the 8 videos in Figure 7. There is a strong positive correlation with Sidewalk and Sky, and a strong negative correlation with Road for Walkability (0.93, 0.86, and -0.90 respectively), which aligns well with state-of-the-art results. Similar patterns are observed for Comfort (0.90, 0.83, and -0.74 respectively) and Pleasurability (0.92, 0.84, and -0.86 respectively) (**RQ3**).

Viewport Analysis The correlation between viewport orientation and the scores is shown in Figure 8. Correlations are calculated between the walkability scores and the averaged viewports for each video. Only the Sidewalk maintains a significant positive correlation with Walkability (0.95), Pleasurability (0.93), and Comfort (0.92), and additionally shows a positive correlation with Safety (0.87). In contrast, the Road remains important in assessing Accessibility (-0.91) and Safety (-0.92), exhibiting a negative influence (**RQ3**). False Discovery Rate (FDR)-adjusted p-values are reported in italics in the table.

Figure 9 presents a bar plot illustrating the Pearson correlation between the walkability score and the 65 segmentation classes defined in the Mapillary Vistas dataset. As in Figure 8, correlations are computed on aggregated over time and video types values. Bars with a correlation greater than 0.7 are shown in red, while the six labels we focused on are highlighted with a blue outline. Asterisks indicate correlations with FDR-adjusted p-values below 0.05. We observe that some of the focused labels, such as Car, Sky, or Vegetation are less significant compared to others that are rarely addressed in the literature. Additionally, several other classes appear to be strongly correlated with the walkability score. Notably, Bench is unsurprisingly highly positively correlated. Unexpectedly, Bird and Ground Animal also exhibit strong positive correlations and could be grouped into a "moving object" category. We hypothesize that this category may be better captured in dynamic surveys, such as VR experiments, but could be largely underestimated in static surveys (i.e., SVI studies). Furthermore, some of these less-studied classes may have been overlooked due to their relatively small area within the frame compared to the six primary labels.

Gaze Point Analysis Pearson correlation coefficients between aggregated over time gaze points and scores, both subsequently averaged within each of the 8 videos are shown in Figure 10. Although Buildings, Vegetation, Sky, and Sidewalk were present in the gaze focus compared to the Road, participants did not pay much attention to these elements, and none significantly influenced their feedback. Only the Road showed a strong negative correlation with Accessibility (-0.92) and Safety (-0.91) (**RQ3**).

4.5 Reflexive comments

When presented with the eye-tracking results, some participants reported feeling safer in VR than in real life, which encouraged them to explore more confidently without concern for real-world obstacles. However, others still avoided poles or pedestrians in the video, despite knowing that no interaction was possible, though they were uncertain if they would behave the same way in reality. Additionally, some participants noted that their walking choices would vary depending on the season or climate, preferring sun in winter and shade in summer.

5. Discussion

5.1 Viewing directions (RQ1)

Addressing RQ1, which inquired about the important areas of 360° videos to consider in a visual walkability analysis, the examination of viewport and gaze direction revealed that participants' attention was unevenly distributed across the 360° videos. Approximately 79% of visual focus was concentrated on the front frame, while the combined left–front–right views accounted for 93% of attention. This indicates that although forward vision predominates, peripheral context significantly contributes to perception. When preparing the panoramic video dataset, we assumed that participants would primarily focus forward while walking, without strong stimuli in their peripheral vision (Nakamura, 2021, Kwan and Fu, 2021, Nakada et al., 2018). These results confirm our hypothesis but also emphasize the non-negligible contribution of lateral views. Methodologically, this finding provides concrete guidance for CVR-based visual walkability audits: analyses restricted to a forward-facing 270° field of view are likely to capture most viewing behavior while substantially reducing computational costs for semantic segmentation and attention processing. Conversely, limiting analyses to a single front frame would overlook a mean-

	Building	Sidewalk	Sky	Vegetation	Car	Road	Other	Walkability score
Video 5	12.5%	36.8%	8.4%	33.2%	0.0%	0.0%	9.1%	M=4.59 SD=0.53
Video 9	61.8%	4.5%	3.5%	2.2%	7.2%	10.8%	10.0%	M=2.52 SD=0.91
Video 1	52.6%	5.1%	2.6%	15.8%	4.6%	9.0%	10.3%	M=3.19 SD=0.76
Video 2	60.3%	5.1%	1.6%	12.1%	7.3%	6.7%	6.9%	M=3.26 SD=0.80
Video 3	59.7%	7.3%	4.2%	5.1%	2.8%	9.0%	11.9%	M=3.23 SD=0.81
Video 4	61.5%	7.2%	4.6%	4.6%	3.3%	8.2%	10.6%	M=3.18 SD=1.01
Video 6	15.0%	33.8%	11.0%	29.2%	0.0%	0.0%	11.0%	M=4.64 SD=0.56
Video 7	36.7%	13.3%	0.4%	13.3%	24.0%	2.4%	9.9%	M=3.21 SD=1.14
Video 8	50.7%	2.9%	0.9%	12.9%	4.6%	13.7%	14.3%	M=2.46 SD=0.89
Video 10	58.2%	6.4%	3.0%	5.4%	10.0%	7.9%	9.1%	M=3.06 SD=0.88
M: Mean Score		SD: Standard Deviation						

Table 1. Percentage of pixels in each category for 3,600 frames per video across the three selected frames.

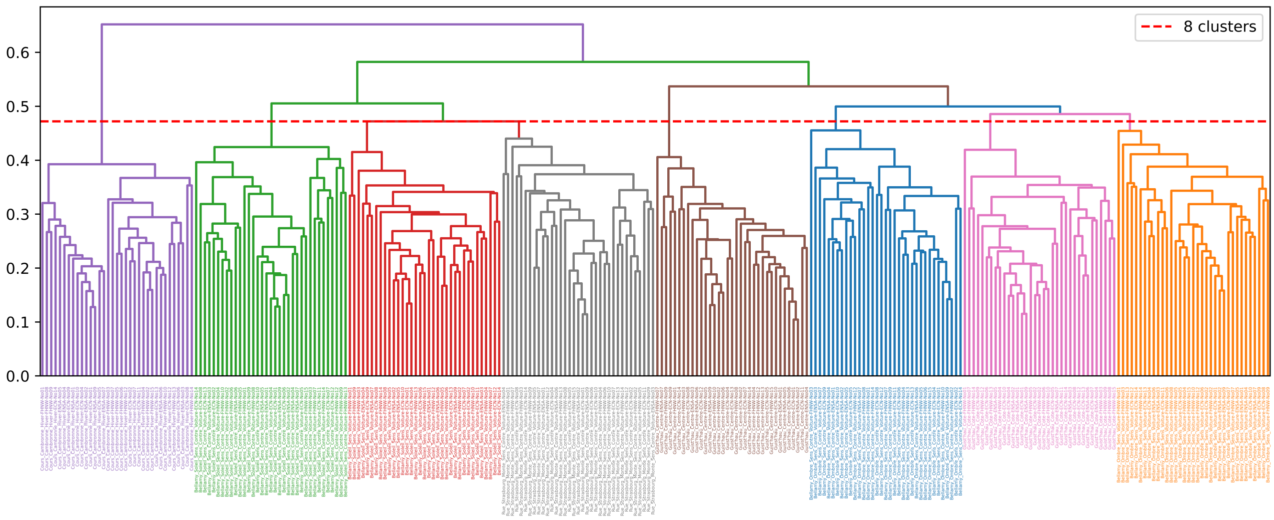

Figure 6. Hierarchical agglomerative clustering dendrogram derived from the Jensen–Shannon distance (JSD) matrix computed across 3,600 segmented viewports consisting of 65 values each.

	Building	Vegetation	Road	Sidewalk	Car	Sky
Accessibility	-0.56 0.32 (0.15)	0.34 0.12 (0.41)	-0.94 0.88 (0.00)**	0.75 0.56 (0.03)*	-0.02 0.00 (0.95)	0.69 0.47 (0.06)
Safety	-0.66 0.44 (0.07)	0.44 0.19 (0.28)	-0.95 0.90 (0.00)**	0.81 0.66 (0.01)*	-0.05 0.00 (0.91)	0.73 0.53 (0.04)*
Comfort	-0.74 0.54 (0.04)*	0.56 0.32 (0.15)	-0.92 0.84 (0.00)**	0.90 0.81 (0.00)**	-0.22 0.05 (0.61)	0.83 0.69 (0.01)*
Pleasurability	-0.80 0.64 (0.02)*	0.71 0.50 (0.05)	-0.86 0.74 (0.01)**	0.92 0.85 (0.00)**	-0.31 0.09 (0.46)	0.84 0.71 (0.01)**
Walkability	-0.78 0.61 (0.02)*	0.62 0.38 (0.10)	-0.90 0.81 (0.00)**	0.93 0.87 (0.00)**	-0.26 0.07 (0.53)	0.86 0.74 (0.01)**

Figure 7. Correlations between the walkability questionnaire and variables across the three frames (front, left, right). R^2 in small font, p-values in parentheses. Significance levels are indicated as follows: * $p < 0.05$; ** $p < 0.01$.

ingful portion of lateral context that may contribute to judgments, and should therefore be avoided when the aim is to approximate perceived walkability from omnidirectional stimuli.

5.2 Explanatory variables

Given that the forward-facing 270° covers most viewing behavior, we further examine how semantic exposure within the corresponding frames (front, left, and right) relates to walkability ratings. Our initial analysis of these three frames aligns with the general approach of other walkability studies utilizing SVI (Huang et al., 2022, Huang et al., 2023, Li et al., 2022). The results closely correspond with some of their findings, specifically

	Building	Vegetation	Road	Sidewalk	Car	Sky
Accessibility	-0.56 0.32 (0.14) *0.40*	0.33 0.11 (0.43) *0.81*	-0.91 0.84 (0.00) *0.03**	0.82 0.68 (0.01) *0.11*	-0.07 0.00 (0.87) *0.96*	0.51 0.26 (0.19) *0.50*
Safety	-0.65 0.43 (0.08) *0.24*	0.41 0.17 (0.32) *0.65*	-0.92 0.84 (0.00) *0.04**	0.87 0.76 (0.00) *0.08*	-0.06 0.00 (0.89) *0.99*	0.51 0.26 (0.20) *0.47*
Comfort	-0.72 0.51 (0.05) *0.16*	0.44 0.19 (0.27) *0.58*	-0.84 0.70 (0.01) *0.06*	0.93 0.87 (0.00) *0.05**	-0.23 0.05 (0.59) *1.00*	0.61 0.37 (0.11) *0.29*
Pleasurability	-0.80 0.64 (0.02) *0.08*	0.57 0.32 (0.14) *0.43*	-0.72 0.53 (0.04) *0.17*	0.94 0.88 (0.00) *0.01***	-0.32 0.10 (0.43) *0.83*	0.58 0.34 (0.13) *0.41*
Walkability	-0.78 0.62 (0.02) *0.10*	0.51 0.26 (0.20) *0.45*	-0.79 0.63 (0.02) *0.09*	0.96 0.93 (0.00) *0.01***	-0.27 0.07 (0.52) *0.92*	0.63 0.40 (0.09) *0.30*

Figure 8. Correlations between the walkability questionnaire scores and variables within across participant’s viewport. R^2 in small font, p-values in parentheses, FDR-adjusted p-values in italic. Significance levels are indicated as follows: * $p < 0.05$; ** $p < 0.01$.

that roads negatively impact walkability, while sidewalks and sky contribute positively. Although some videos yielded similar walkability scores, correlation analysis remains valid because they differ in their segmented classes. For instance, in Table 1, we observe that Video4 and Video7 exhibit complementary features compared to the two calibration videos. Despite this apparent contradiction, their nearly identical walkability scores of approximately 3.2 demonstrate that their actual implement-

https://doi.org/10.5194/isprs-annals-XI-4-2026-241-2026

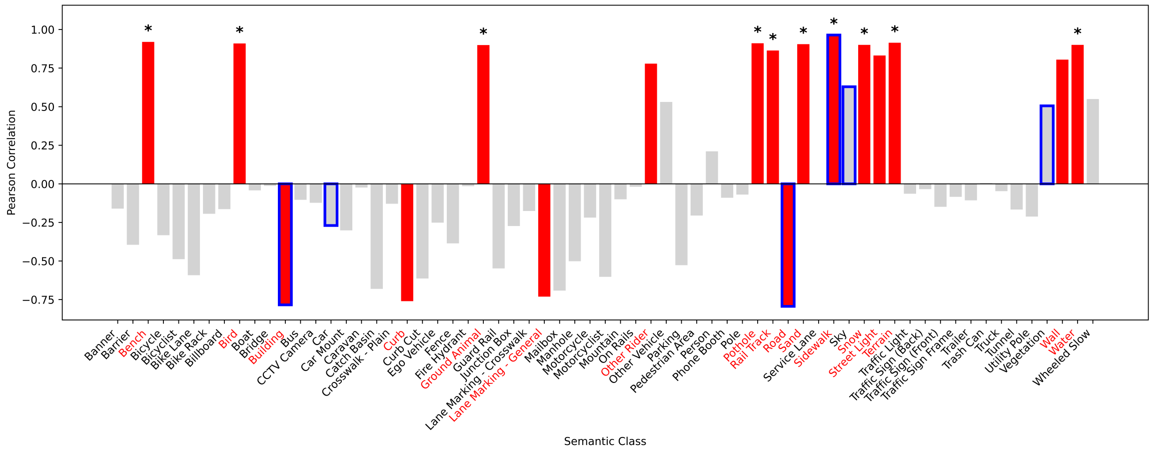

Figure 9. Bar plot of Pearson correlations between the walkability score and the aggregated over time and video type viewports for the 65 segmentation classes. Correlations above 0.7 are shown in red, the six selected labels are highlighted with a blue outline and asterisks indicate correlation with FDR-adjusted p-values below 0.05.

	Building	Vegetation	Road	Sidewalk	Car	Sky
Accessibility	-0.60 0.36 (0.11)	0.41 0.17 (0.32)	-0.92 0.85 (0.00)**	0.22 0.05 (0.60)	-0.32 0.10 (0.44)	0.27 0.07 (0.52)
Safety	-0.65 0.43 (0.08)	0.47 0.22 (0.25)	-0.91 0.83 (0.00)**	0.25 0.06 (0.55)	-0.33 0.11 (0.42)	0.24 0.06 (0.57)
Comfort	-0.70 0.49 (0.05)	0.48 0.23 (0.23)	-0.78 0.61 (0.02)*	0.46 0.21 (0.25)	-0.47 0.22 (0.24)	0.29 0.08 (0.48)
Pleasurability	-0.77 0.59 (0.03)*	0.55 0.31 (0.16)	-0.64 0.40 (0.09)	0.60 0.36 (0.12)	-0.51 0.26 (0.20)	0.26 0.07 (0.53)
Walkability	-0.76 0.58 (0.03)*	0.49 0.24 (0.21)	-0.71 0.50 (0.05)*	0.57 0.32 (0.14)	-0.50 0.25 (0.21)	0.34 0.12 (0.40)

Figure 10. Correlations between walkability questionnaire and variables of gaze points. R^2 in small font, p-values in parentheses. Significance levels are indicated as follows: * $p < 0.05$; ** $p < 0.01$.

ations produce the same outcome. Since we have only these three frames and no depth map, we cannot determine the distance between the viewpoint and other objects. Therefore, we analyzed the field of view and gaze point to better understand the origin of the scores (**RQ1**).

Building Indeed, analyzing the pixel ratio alone cannot capture all the characteristics of a street. Since this ratio is analogous to a solid angle, a large, distant building can occupy the same number of pixels as a small, nearby building. These two scenarios can lead to different perceptions of walkability. Therefore, the distance to the building also influences pedestrian perception. However, there is no direct method to represent this variable. As shown in Figure 1, Video8 is actually very close to the building due to the narrow sidewalk. This proximity not only resulted in a low walkability score but also caused some participants to report a stronger sensation of cybersickness (**RQ2**).

Vegetation We were surprised to find no significant correlation between vegetation and walkability, as vegetation is one of the elements most frequently highlighted in previous studies (**RQ3**). The decision to film along a straight path may have influenced this outcome, as we aimed to minimize the number of variables in the video, with sinuosity being one of them. Because we used a straight road, participants did not have the opportunity to observe the roadside before turning, which may have diminished the role of vegetation in our analysis (**RQ1**). In addition, three videos classified with a high vegetation ratio by semantic segmentation (Video6, Video7 and Video8, see Table 1) were recorded in winter, thus showing trees without leaves to the participants. This lack of leaves could have negatively influenced the ratings, consequently lowering the expected correlation between vegetation detection and walkability ratings (**RQ2**).

Road and Car Video7 showed 24% of the vehicles, a significantly higher proportion compared to the other videos. However, Video7 also received one of the highest walkability scores (median=3.5/5, the second highest), as the vehicles were parked on both sides of the sidewalk. It is not merely the presence of parked vehicles but the traffic flow, i.e. moving cars, that influences pedestrians’ perception of walkability. Our approach analyzes individual images and does not consider the temporal sequence of events in the video. Consequently, we cannot distinguish the movement of objects using the segmentation technique employed (**RQ2**).

Sidewalk As shown in Figures 7 and 8, sidewalks have a significant influence on participants’ assessment of visual walkability. However, in Figure 10, this influence disappears. This may be due to differences between VR and reality, as discussed in Section 4.5. Participants did not need to pay attention to the sidewalks they were walking on because they were aware of their presence; this may also be attributed to the limited freedom within VR environment. Nevertheless, the sidewalk must be present and sufficiently large within the viewport (**RQ3**).

Sky The sky is one of the classes that participants did not need to focus on, but the natural lighting of the scene implicitly indicates its presence. The sky attracts less visual attention and seemed to be less important in Figure 10 as its sense is conveyed by the ambient light.

Gaze points In our analysis of participants’ visual attention, we identified variables that were not initially considered according to the current state of the art. Figure 11 presents two analyses of visual attention for Video7 as an example. The left panel shows the absolute percentage of participants’ visual attention, indicating how many times the 35 participants looked at each object. We observe that the Person class has a very low percentage (0.32%). However, when considering the proportion of the Person class present in Video7, the relative visual attention directed toward the Person class rises to second place, and the Billboard class shows a similar pattern in other videos such as in Figure 12. Instances of these two classes hold a unique status. Advertisements are inherently designed to attract attention; unsurprisingly, their ability to capture attention exceeds the visual space they occupy. Regarding the individuals depicted in the video, it is likely that, as part of our species' natural sociability, we feel compelled to identify the people we encounter. Nevertheless, when these two variables are subjected to correlation analysis, no significant influence is found (p=0.465 for Person, p=0.700 for Billboard) (**RQ2**).

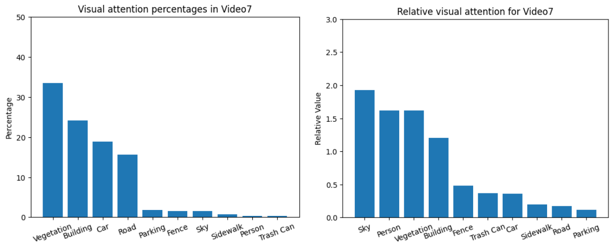

Figure 11. (Left) The percentages of top 10 objects visually focused by participant in Video7. (Right) The ratio of visual attention percentage over object class percentage in Video7.

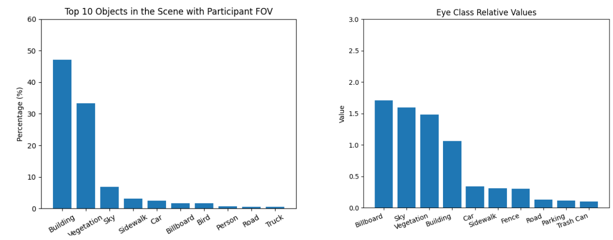

Figure 12. (Left) The percentages of top 10 objects visually focused by participant in Video2. (Right) The ratio of visual attention percentage over object class percentage in Video2.

Visual Attention According to Wang et al. (Wang et al., 2023), we should analyze not only a single viewpoint, but also the surrounding context, i.e. the entire viewport. A closer examination of Figures 7, 8, and 10 reveals that the correlations with the individual segments of walkability generally align well with the hierarchy of walking needs (Alfonzo, 2005), with the relationship strengthening toward the top of the pyramid (e.g. comfort and pleasurability), except for the Road variable. Notably, Road is the only variable that demonstrates a significant p-value for Accessibility and Safety when analyzing visual attention (viewport and gaze). This raises a new question: If our results using gaze points and viewport data cannot demonstrate reliable correlations for other factors, as shown in previous studies, then the results obtained from the state-of-the-art based on a frame analysis may also be questionable, since global environmental characteristics alone cannot accurately reflect pedestrians' true experiences (**RQ3**).

According to the discussion above, by applying the CVR method, we can evaluate at least some aspects of visual walkability, such as the influence of roads, sidewalks, and sky, as well as investigate visual attractions for pedestrians.

6. Limitations and Perspectives

6.1 Experimental protocol

VR The simulation of walking using the arm-swinging method can be challenged by others approaches, such as walk-in-place, using a treadmill, or allowing the user to walk freely; especially if the experiment involves adding turns (Templeman et al., 1999). Participants reported that VR may influence their behavior because they feel safer and lack real interactions. Although this bias in visual attention did not affect our results compared to the state of the art, further studies, such as field experiments, are needed to validate these findings in real-world settings. To further validate our approach, future work could include a real-world field trial using Tobii© eye-tracking glasses, which enable the collection of gaze data under real-world conditions.

Priming videos The protocol included two “consensus” calibration videos shown at the beginning of the experiment to establish a common interpretation of the walkability rating scale, it may also prime an implicit framework that affects later judgments. Therefore, absolute ratings should be interpreted with caution, and the calibration step may have influenced the magnitude of observed differences between videos. Future studies should explicitly test the effect of this calibration, for example, by using a between-subjects design with and without calibration. Such designs would allow researchers to quantify whether calibration improves rating reliability without systematically shifting subsequent walkability judgments.

Participants Although the final sample of 35 valid participants is typical for CVR and eye-tracking protocols, the participants were recruited from the university community, which may limit demographic diversity and thus the generalizability of absolute walkability ratings. Future work should replicate the study with broader and more diverse samples (e.g., age ranges, mobility profiles, familiarity with the study area and VR) to assess robustness across populations.

6.2 Data analysis

Semantic segmentation One challenge in semantic segmentation is that object classes do not all behave uniformly in real life. For example, we do not need to look directly at the sky to perceive its extent, as ambient light already conveys this information. This aspect is often overlooked by correlation algorithms. Furthermore, segmentation is imperfect and can misclassify objects, as illustrated in Figure 1, where Video7's winter scene detected vegetation but failed to identify the sky obscured by tree branches. Finally, image analysis lacks depth perception, preventing the proper weighting of the visual influence of distant objects.

Motion detection To enhance the results of our study, we could incorporate motion detection techniques into our post-processing algorithm to identify moving objects in our videos. Since we treated videos as individual frames for semantic segmentation, we were unable to automatically analyze the impact of differences between static and dynamic objects. Integrating motion information would help capture these distinctions and could lead to more accurate interpretations—especially considering the findings presented in Figure 9, which suggest a potential influence of object motion on perceived walkability.

https://doi.org/10.5194/isprs-annals-XI-4-2026-241-2026 | © Author(s) 2026. CC BY 4.0 License.

Gaze and viewport Gaze and viewport were analyzed using aggregated semantic exposure and class-level correlations. We did not explicitly model the spatial distribution of gaze within the viewport. As a result, we cannot determine whether the gaze–viewport mismatch reflects a limitation of gaze-based inference in CVR or indicates that walkability judgments rely on ambient or peripheral cues not captured by focal fixations. Future work should quantify gaze–viewport spatial relationships and test whether incorporating these features improves the alignment between attention metrics and walkability ratings.

6.3 Walkability

Multisensory Some participants mentioned that the soundscape of the street influences their perception of walkability. Even when efforts are made to minimize the impact of urban background noise, the sounds on a street remain distinct from those in a park. Our experiment was intentionally designed to assess visual walkability under controlled and reproducible conditions using CVR. As such, it does not capture other determinants of walking experience, including soundscape, physical effort, microclimate, and social interactions. Therefore, our findings should be interpreted as evidence about visual cues and about the methodological behavior of CVR-based measures, rather than as a comprehensive account of walkability in real environments. Field studies and multisensory VR setups will be required to test whether the observed relationships hold when non-visual factors are present and interacting with visual perception.

Motivation Participants’ route choices may be influenced by their travel motivations (Salazar Miranda et al., 2021). In our experiment, we instructed participants to evaluate the road as part of their daily commute, thereby excluding other options driven by different motivations. Different results might have emerged if the scenario had involved a leisure walk instead.

Application Ultimately, our method, which relied solely on 360° videos, proved to be a relatively inexpensive and straightforward approach in terms of both time and cost. This method could be a valuable option for urban professionals, such as developers or planning agencies, who wish to assess walkability without resorting to more complex multisensory setups like treadmills or wind tunnels.

7. Conclusion

We examined the relationships between participants’ perceived walkability, the characteristics of urban environments depicted in 360° videos, and participants’ viewport and gaze data. Our findings suggest that the environmental factors influencing visual walkability assessments generally align with those identified in previous real-world studies. However, discrepancies emerged in the viewport and gaze analyses, indicating that participants’ visual attention in CVR may not fully replicate real-world experiences. Therefore, although our findings imply that CVR can capture high-level constructs such as perceived walkability, further research is needed to determine whether it is a reliable method for assessing lower-level walking needs. Methodologically, our results also suggest that CVR-based visual walkability audits can primarily rely on scene content and viewport exposure, while gaze-based indicators should be interpreted cautiously. Additionally, restricting analyses to a forward-facing 270° field of view is likely to capture most viewing behavior while reducing processing costs.

We argue that current approaches to assess urban walkability based on street-level image segmentation and the quantification of a limited set of visual classes are sensitive to methodological choices and may provide an incomplete account of perceived walkability. We identify at least three main limitations. First, the selection of classes is typically restricted to a small subset, often no more than half a dozen (e.g., building, sky, road, car, vegetation, sidewalk). Yet, we show that several other classes among the 65 categories analyzed in our study appear either to influence, or at least to correlate with, the quality indicators derived from the survey. Second, relying on the proportion of pixels occupied by each class in the image is an overly simplistic metric. Some elements that are visually small in terms of pixel area, such as a bird, may nonetheless strongly attract attention and shape observers’ preferences, positively or negatively. Pixel ratios alone therefore fail to capture the perceptual salience of urban features. Third, these approaches generally ignore the spatial position of classes within the visual field. For certain elements, being located in the central field of view may have a different perceptual impact than being situated in the periphery. This limitation further justifies the implementation of eye-tracking procedures to account for the spatial distribution of attention and its relationship to the content of the viewport in future work.

References

Alfonzo, M. A., 2005. To Walk or Not to Walk? The Hierarchy of Walking Needs. *Environment and Behavior*, 37(6).

Badland, H. M., Opit, S., Witten, K., Kearns, R. A., Mavoa, S., 2010. Can virtual streetscape audits reliably replace physical streetscape audits? *Journal of Urban Health*, 87(6).

Beaucamp, B., Leduc, T., Tourre, V., Servières, M., 2025. Beyond the frame: evaluating panoramic vs. perspective images for assessing place perception. *International Journal of Geographical Information Science*, 39(10), 2300–2332.

Birenboim, A., Ben-Nun Bloom, P., Levit, H., Omer, I., 2021. The Study of Walking, Walkability and Wellbeing in Immersive Virtual Environments. *International Journal of Environmental Research and Public Health*, 18(2), 364.

Chen, S., Biljecki, F., 2023. Automatic assessment of public open spaces using street view imagery. *Cities*, 137, 104329.

Cheng, B., Misra, I., Schwing, A. G., Kirillov, A., Girdhar, R., 2022. Masked-attention mask transformer for universal image segmentation. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 1280–1289.

Cherni, H., Métayer, N., Souliman, N., 2020. Literature review of locomotion techniques in virtual reality. *International Journal of Virtual Reality*, 20(1), 1–20.

Cinnamon, J., Jahiu, L., 2023. 360-degree video for virtual place-based research: A review and research agenda. *Computers, Environment and Urban Systems*, 106, 102044.

Dovey, K., Pafka, E., 2020. What is walkability? The urban DMA. *Urban Studies*, 57(1), 93–108.

Dubey, A., Naik, N., Parikh, D., Raskar, R., Hidalgo, C. A., 2016. Deep Learning the City: Quantifying Urban Perception at a Global Scale. *Computer Vision – ECCV 2016*, Lecture Notes in Computer Science, Springer International Publishing, 196–212.

Huang, G., Yu, Y., Lyu, M., Sun, D., Zeng, Q., Bart, D., 2023. Using google street view panoramas to investigate the influence of urban coastal street environment on visual walkability. *Environmental Research Communications*, 5(6), 065017.

Huang, J., Liang, J., Yang, M., Li, Y., 2022. Visual Preference Analysis and Planning Responses Based on Street View Images: A Case Study of Gulangyu Island, China. *Land*, 12(1), 129.

Ito, K., Quintana, M., Han, X., Zimmermann, R., Biljecki, F., 2024. Translating street view imagery to correct perspectives to enhance bikeability and walkability studies. *International Journal of Geographical Information Science*, 0(0), 1–31.

Kim, H. K., Park, J., Choi, Y., Choe, M., 2018. Virtual reality sickness questionnaire (VRSQ): Motion sickness measurement index in a virtual reality environment. *Applied Ergonomics*, 69, 66–73.

Kim, S. N., Lee, H., 2022. Capturing reality: Validation of omnidirectional video-based immersive virtual reality as a streetscape quality auditing method. *Landscape and Urban Planning*, 218. Publisher: Elsevier B.V.

Kwan, K. C., Fu, H., 2021. Automatic Image Checkpoint Selection for Guider-Follower Pedestrian Navigation. *Computer Graphics Forum*, 40(1), 357–368.

Lee, H., Kim, S.-N., 2021. Perceived Safety and Pedestrian Performance in Pedestrian Priority Streets (PPSs) in Seoul, Korea: A Virtual Reality Experiment and Trace Mapping. *International Journal of Environmental Research and Public Health*, 18(5), 2501.

Lee, J., Kim, D., Park, J., 2022. A Machine Learning and Computer Vision Study of the Environmental Characteristics of Streetscapes That Affect Pedestrian Satisfaction. *Sustainability*, 14(9), 5730.

Li, Y., Yabuki, N., Fukuda, T., 2022. Measuring visual walkability perception using panoramic street view images, virtual reality, and deep learning. *Sustainable Cities and Society*, 86, 104140.

Li, Y., Yabuki, N., Fukuda, T., Zhang, J., 2020. A big data evaluation of urban street walkability using deep learning and environmental sensors a case study around Osaka University Suita campus. *eCAADe 2020*.

Liao, B., van den Berg, P. E., van Wesemael, P., Arentze, T. A., 2022. Individuals' perception of walkability: Results of a conjoint experiment using videos of virtual environments. *Cities*, 125, 103650.

Lin, J., 1991. Divergence measures based on the Shannon entropy. *IEEE Transactions on Information Theory*, 37(1), 145-151.

Ma, X., Ma, C., Wu, C., Xi, Y., Yang, R., Peng, N., Zhang, C., Ren, F., 2021. Measuring human perceptions of streetscapes to better inform urban renewal: A perspective of scene semantic parsing. *Cities*, 110, 103086.

Makransky, G., Lilleholt, L., Aaby, A., 2017. Development and validation of the Multimodal Presence Scale for virtual reality environments: A confirmatory factor analysis and item response theory approach. *Computers in Human Behavior*, 72, 276–285.

McCullough, M., Xu, H., Michelson, J., Jackoski, M., Pease, W., Cobb, W., Kalescky, W., Ladd, J., Williams, B., 2015. Myo arm: swinging to explore a VE. *Proceedings of the ACM SIGGRAPH Symposium on Applied Perception*, New York, NY, USA, 107–113.

Morris, J. N., Hardman, A. E., 1997. Walking to Health. *Sports Medicine*, 23(5), 306–332.

Nakada, M., Chen, H., Terzopoulos, D., 2018. Learning Biomimetic Perception for Human Sensorimotor Control. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 1998–19985.

Nakamura, K., 2021. Experimental analysis of walkability evaluation using virtual reality application. *Environment and Planning B: Urban Analytics and City Science*, 48(8), 2481–2496. Publisher: SAGE Publications Ltd.

Neuhold, G., Ollmann, T., Bulò, S. R., Kontschieder, P., 2017. The mapillary vistas dataset for semantic understanding of street scenes. *2017 IEEE International Conference on Computer Vision (ICCV)*, 5000–5009.

Pavic, K., Chaby, L., Gricourt, T., Vergilino-Perez, D., 2023. Feeling Virtually Present Makes Me Happier: The Influence of Immersion, Sense of Presence, and Video Contents on Positive Emotion Induction. *Cyberpsychology, Behavior and Social Networking*, 26(4), 238–245.

Ren, X., Wei, P., Wang, Q., Sun, W., Yuan, M., Shao, S., Zhu, D., Xue, Y., 2023. The effects of audio-visual perceptual characteristics on environmental health of pedestrian streets with traffic noise: A case study in Dalian, China. *Frontiers in Psychology*, 14.

Salazar Miranda, A., Fan, Z., Duarte, F., Ratti, C., 2021. Desirable streets: Using deviations in pedestrian trajectories to measure the value of the built environment. *Computers, Environment and Urban Systems*, 86, 101563.

Stokes, D., Biggs, S., 2014. The Dominance of the Visual. *Perception and Its Modalities*, Oxford University Press, 0.

Templeman, J. N., Denbrook, P. S., Sibert, L. E., 1999. Virtual Locomotion: Walking in Place through Virtual Environments. *Presence: Teleoperators and Virtual Environments*, 8(6), 598-617.

Wang, F. S., Lohmeyer, Q., Duchowski, A., Meboldt, M., 2023. Gaze is more than just a point: Rethinking visual attention analysis using peripheral vision-based gaze mapping. *Proceedings of the 2023 Symposium on Eye Tracking Research and Applications*, Association for Computing Machinery, 1–7.

White, M. P., Alcock, I., Grellier, J., Wheeler, B. W., Hartig, T., Warber, S. L., Bone, A., Depledge, M. H., Fleming, L. E., 2019. Spending at least 120 minutes a week in nature is associated with good health and wellbeing. *Scientific Reports*, 9(1), 7730.

WHO, 2020. *WHO guidelines on physical activity and sedentary behaviour*. 1st ed edn, World Health Organization, Geneva. World Health Organization.

Zhou, H., He, S., Cai, Y., Wang, M., Su, S., 2019. Social inequalities in neighborhood visual walkability: Using street view imagery and deep learning technologies to facilitate healthy city planning. *Sustainable Cities and Society*, 50.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-4-2026-241-2026