AGILE: GIScience Series, 4, 40, 2023. https://doi.org/10.5194/agile-giss-4-40-2023
Proceedings of the 26th AGILE Conference on Geographic Information Science, 2023.
Editors: P. van Oosterom, H. Ploeger, A. Mansourian, S. Scheider, R. Lemmens, and B. van Loenen.
This contribution underwent peer review based on a full paper submission.

Towards dynamically generating immersive video scenes for studying human-environment interactions

Simon Schröder[1], Jan Stenkamp[1], Michael Brüggemann[1], Benjamin Karic[1],
Judith A. Verstegen[2], and Christian Kray[1]

[1]Institute for Geoinformatics, University of Muenster, Muenster, Germany
[2]Department of Human Geography and Spatial Planning, Utrecht University, Utrecht, The Netherlands

Correspondence: Simon Schröder (simon.p.s@uni-muenster.de)

Abstract. Studies where participants interact with their environment are necessary, for example to investigate pedestrian behaviour, in urban planning or for environmental studies. Existing methods such as field studies or 3D virtual reality environments frequently require a lot of effort to create realistic 3D worlds or are subject to real-world interferences. In this paper we propose an alternative approach that uses an immersive video environment (IVE) and a dynamic scene generator (DSG). We use overlays to dynamically generate video scenes to facilitate studying specific combinations of environmental factors with participants. We report on an initial case study in the context of studying pedestrian behaviour, where we show how the approach can facilitate studying how crowds and pedestrian signage affect route choices. Our approach can help researchers studying human-environment interactions in the lab.

Keywords. user studies, pedestrian behaviour, immersive video environment, dynamic scene generation

1 Introduction

There are many scenarios where people interact with their environment to achieve a specific goal, for example making sense of their surroundings and taking decisions, navigation, or extracting information from public signage. Studying these interactions is thus important to understand the behaviours of people in such scenarios, to design technology to potentially steer these behaviours and to test them in simulations.

Common methods used for this purpose include surveys, field studies, observational studies and also studies in virtual reality (Zhao et al., 2020; Feng et al., 2021; Filomena et al., 2022). Each of these methods comes with benefits and drawbacks. For example, surveys are easy to set up and run but lack ecological validity. Field studies can be very realistic but are subject to interference and lack control. Observational studies such as location tracking can provide large amounts of realistic data but raise privacy concerns. Virtual reality can expose participants to realistic environments and provide a high degree of control but the effort to create life-like 3D environments is high and movement can be cumbersome.

Immersive video environments (IVEs) are another approach, where participants are placed in a visualisation environment that surrounds them and displays panoramic footage of real locations. IVEs combine several benefits of the other methods mentioned above: they are comparatively easy to create by recording footage at locations of interest. They can also provide a high degree of visual fidelity (Slater and Wilbur, 1997) and ecological validity (Rossetti and Hurtubia, 2020). Finally, they facilitate full control of the stimuli participants are exposed to so that every participant can experience exactly the same situations. The key drawbacks of using IVEs for user studies are the lack of free movement and the immutable nature of the video scenes. Both stem from the fact that panoramic videos are recorded once at a specific location. Unlike 3D environments, it is not easily possible to move to a slightly different location and still experience the same 3D scene from a different viewpoint. While computational expensive approaches exist to construct 3D scenes from videos recorded according to a rigidly control schema, the resulting 3D worlds still lack realism (Katkere et al., 1997). Alternatively, panoramic footage can be recorded during movement but this does incur additional effort and frequently induces motion sickness (Calogiuri et al., 2018).

While in many cases, being limited to particular viewpoints/decision points does not prevent studying specific questions, the immutable nature of video scenes is more of a problem. While it is possible to record multiple videos at the same location – for example, to provide daytime and nighttime views – other aspects such as the degree of crowdedness are nearly impossible to systematically

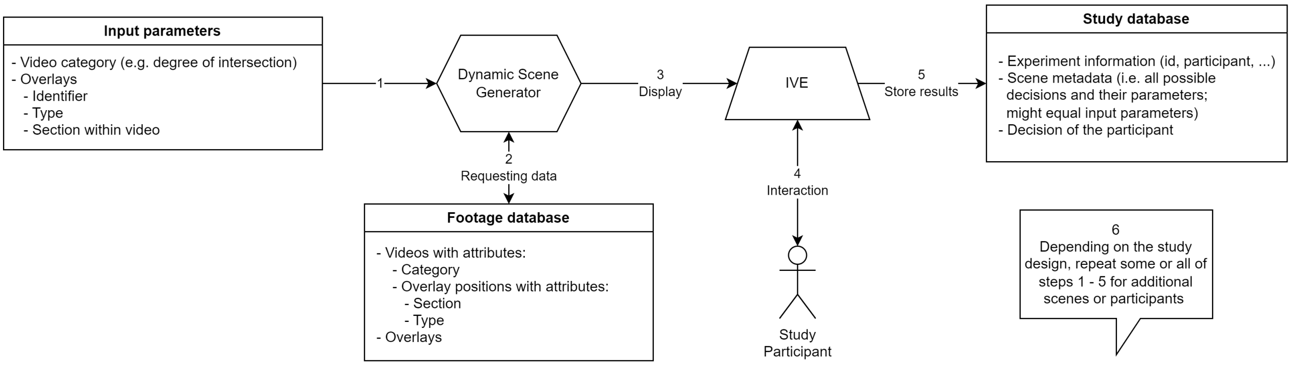
Input parameters
- Video category (e.g. degree of intersection)
- Overlays
- Identifier
- Type
- Section within video
1
Dynamic Scene Generator
2
Requesting data
Footage database
- Videos with attributes:
- Category
- Overlay positions with attributes:
- Section
- Type
- Overlays
3
Display
IVE
4
Interaction
Study Participant
5
Store results
Study database
- Experiment information (id, participant, ...)
- Scene metadata (i.e. all possible decisions and their parameters; might equal input parameters)
- Decision of the participant
6
Depending on the study design, repeat some or all of steps 1 - 5 for additional scenes or participants

Figure 1. Generic conceptual model of how a user study in an IVE supported by the dynamic scene generator proceeds.

vary using just recorded footage. Previous work has proposed to augment recorded panoramic footage with rectangular overlays to simulate the presence of public displays (Ostkamp and Kray, 2014). In this paper, we report on an extension of this approach to overcome the issue of immutable footage: we dynamically overlay prerecorded panoramic footage with different visual overlays according to a set of parameters. We created a prototypical implementation and tested it in the context of studying pedestrian behaviour in the presence of different crowd densities. In the following, we describe our approach in detail, outline an initial case study and summarise our findings and future plans.

2 Approach

In our approach, we use an IVE, which consists of three big screens and custom software, to immerse the participants of a user study in a real-world location (subsequent *location*). Participants stand in front of these screens to experience a 180 degree view of the panoramic footage supplemented by surround sound that was recorded together with the video. Since their entire field of view is covered by the screens, this setup creates a realistic simulation of the displayed location (Bowman and McMahan, 2007). Images and videos can be added to the panoramic footage as overlays, e.g. public displays can be added to footage of a streetscape. We denote the combination of panoramic footage and overlays as *scene*. The generated scenes are then shown to the participants of a study so that they can react to or interact with it. Their feedback, interactions and behaviour are collected for later analysis. In studies conducted so far with such systems, e.g. by Stenkamp et al. (2023), scenes shown in the IVE corresponded directly to real-world locations: when locations were adjacent in the real-world, the corresponding scenes were presented as adjacent in the IVE as well.

Building on the process described above, the new approach we present here foregoes this direct mapping between real-world locations and immersive video scenes. Instead, it is based on dynamically generating a scene by combining a panoramic video with overlays in such a way that the requirements of a user study are met.

To achieve this, we use a *dynamic scene generator* (DSG), which creates scenes according to a set of input parameters by combining overlays with panoramic videos fulfilling specific criteria. In practice, the DSG works as follows (see Figure 1): In the first step, *input parameters* for the scene to be generated are passed to the generator. These contain the category of the video to be selected (e.g. type of street intersection displayed) and all the overlays that should be placed in the video. Overlay parameters include information about the type of overlay and, if applicable, the section within a video where the overlay has to be placed (e.g. a certain street at an intersection). An overlay can occur one or more times in the same video and can also be used for different videos (e.g. a specific pedestrian sign can be overlaid at two positions in video A and also be used for video B). In the second step, the DSG retrieves the actual data for the specified input parameters from a database. From a set of stored videos that meet the criteria specified in the input parameters, it retrieves one, e.g. at random or according to a systematic variation such as Latin Square, depending on the implementation. The database also stores data about the anchor points where the respective overlays should be placed in each video. The overlays themselves are also retrieved from the database. In the third step, the DSG combines all the data and generates a scene that is then displayed in the IVE and shown to the participant of a user study. The fourth step consists of participants experiencing the generated scene and reacting to it in a way appropriate for the given study. In a last step, this reaction is stored together with the other information describing the scene, i.e. the input parameters and an identifier, so it can later be analysed easily.

Our approach thus relies on three key elements: an annotated collection of panoramic videos, an annotated collection of overlays, and a systematic description of the user study, i.e. the underlying experimental plan. The DSG then uses these elements to generate immersive scenes for participants.

The *annotated collection of panoramic videos* contains a set of recorded footage that has been annotated with at-

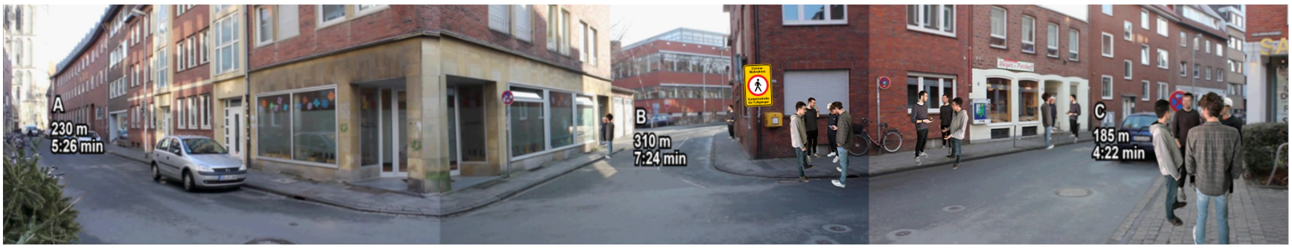

Figure 2. Dynamically generated immersive video scene showing recorded panoramic footage of a location overlaid with pedestrian signage (middle), added people and routing information; position and type of signage and routing information as well as number and position of people were parameters for the generation process.

tributes that are relevant in the context of the study and with anchor points for overlays. For example, for a navigation study, the annotations might include how many streets lead away from the recorded location as an attribute. Anchor points correspond to the absolute pixel-coordinate in the footage, where a specific type of overlay can be placed. In Figure 2, the sign in the middle of the figure is placed at such an anchor point as are the added people.

The *annotated collection of overlays* consists of all static pictures and videos that can be overlaid over the footage as well as annotations that describe the type of overlay. For the navigation example, this can be pictures of different types of signage, routing information, or videos of people (see Figure 2).

Finally, the *description of the user study* contains information about which overlays should be added to which panoramic video. In the navigation example, this part describes which signage to add where and how many people to include where.

For an actual study, several scenes would usually be generated by creating sets of parameters of interest and passing them successively to the DSG. The DSG then creates the respective scenes. During the study, each scene is shown to participants, their reaction is recorded and the next scene is shown. Since the scenes are automatically generated for the purposes of the study without the need to assemble each one manually, the effort required is comparatively low while the degree of control over the content of each scene is high.

3 Initial case study

As a case study for the approach introduced in the previous section, we implemented it for a user study on pedestrian behaviour. The study follows the same protocol as a previous study (Stenkamp et al., 2023) but includes additional overlays to investigate the impact of crowd sizes on route choices. The procedure underlying the study is as follows. The participants are instructed to navigate from a given starting point to a specified destination. At each intersection, participants have to decide which of the shown streets they wanted to follow, given the information they are seeing in the IVE (see Figure 3). In order to support their decision process, the IVE also displays the time and distance of the shortest path from each street at the intersection to the destination (see Figure 2). This process is repeated until they reach the specified destination.

For our case study, we implemented the generic approach in the following way. The videos were categorised according to the number of streets present at the filmed intersection. For each shown street, anchor points for the same three types of overlays were defined. These three types of overlays consisted of one-way-street signage for pedestrians, crowds of different density and routing information, i.e. time and distance to a destination. There were two types of one-way-street signage. One indicating pedestrians were allowed to enter the street from the current direction, and a second one indicating entry is prohibited (see Figure 2, middle). Crowd overlays were included for three densities: empty street, moderately crowded, very crowded. The crowd overlays were generated from footage taken in front of a green screen (see Figure 2, people shown). Routing information consisted of the shortest distance in meters from the current location to the destination plus the needed travel time.

The main database in this example contained the different types of signage and crowd overlays as well as several videos for the categories two-street intersection and three-street intersection. The videos were reused from previous studies and we added further attributes describing the anchor points. Videos showing two streets had six anchor points and videos with three streets nine anchor points (one anchor point for each type of overlay per street). Routing information overlays were created automatically as texts by the DSG based on the input parameters. The DSG was implemented in JavaScript as a RESTful API and the main database was based on MongoDB. The process of dynamically generating scenes has the following steps (see Figure 1):

1. Input parameters are passed to the DSG according to the study needs, for example containing the following elements:
    - video category: three-street intersection
    - signage overlay A: None
    - crowd overlay A: None / empty

- routing information A: 230 m
- signage overlay B: Entry forbidden
- crowd overlay B: Moderately crowded
- routing information B: 310 m
- signage overlay C: None
- crowd overlay C: Very crowded
- routing information C: 185 m

2. The DSG ingests these parameters and then randomly picks a video from the database showing three streets. It also retrieves the overlays *entry forbidden*, *moderately crowded* and *very crowded*.

3. The DSG generates the scene by placing the overlays onto the video, on areas indicated by the anchor points: at street A, the 230 m as routing information is added, at street B, the *entry forbidden* sign, the *very crowded* overlay and 310 m as routing information are injected and at street C, the *very crowded* overlay and 185 m as routing information are overlaid. The generated scene is stored in the IVE system for playback.

4. The scene is shown on the IVE screens to a participant as part of a study. The participant reacts to the scene by choosing one of the streets, A, B or C.

5. In the study database, all information about the scene is time-stamped and logged including the input parameters that were used to generate the scene and the reaction of the participant.

6. This process repeats for each participant until the study is completed. The data collected in the study database can then be used to analyse participant behaviour in detail.

4 Discussion & Conclusions

For our case study, we were interested in the impact of crowd densities and pedestrian signage on navigation behaviour. Recording footage with all possible combinations of route intersections, pedestrian signage and crowd densities was unfeasible. Using the DSG that implemented our approach for this specific case study, we only needed to do green-screen recordings of different crowd sizes and to specify input parameters as described in Section 3 to generate all scenes for a user-study. This also enabled us to systematically vary particular independent variables (such as crowd density) to ensure results lend themselves to specific statistical analyses. For this approach to work, the videos need to be annotated with anchor points for the overlays and they must be categorised to facilitate retrieval by the DSG. The anchor points have to be chosen carefully to ensure a high level of immersion, e.g. a street sign should be overlaid at a wall of a building and not float in the air.

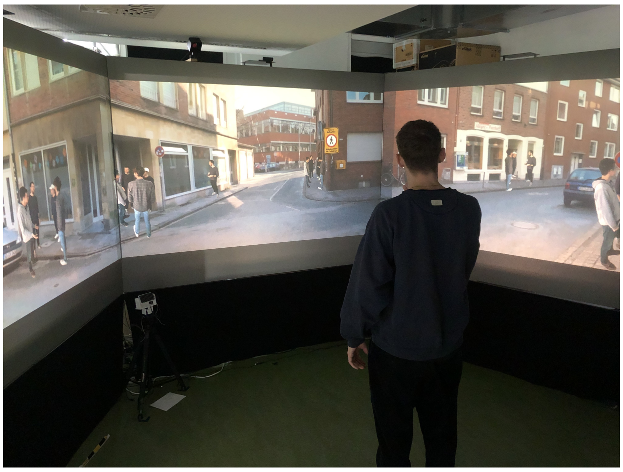

Figure 3. Immersive video environment with participant experiencing dynamically generated video scene, similarly to that shown in Figure 2.

Our approach is not limited to studying pedestrian behaviour. Since the DSG only requires a database with categorised and tagged videos as well as annotated overlays, it can generate all kinds of scenes (e.g. using videos of natural landscape instead of a streetscape and using overlays of wind turbines to study perceptions of sustainable energy generation). With sufficient numbers of videos, repetitions can be avoided and videos can be reused in different studies that investigate similar locations. It is also possible to dynamically generate scenes in response to participants' reactions to previously generated scenes.

Our approach is subject to a number of limitations. While it does not overcome the lack of free movement of video footage, it addresses the immutable footage problem to some degree: it allows for a wide variety of image and video overlays to study different subjects. However, since the overlays are placed on top of the footage, visual inconsistencies can occur, e.g. when a car passes *behind* an overlaid sign that looks as if it was further away than the car. Overlays are also not suitable for certain scenarios, e.g. to simulate different degrees of flooding. Furthermore, our approach also cannot facilitate physical human-environment interaction. Finally, there is some additional preparation effort for each study (video tagging, overlay creation, parameterising scene generation) but this could be reduced in the future by establishing an open database for video material and overlays as well as an annotation standard.

In summary, we proposed an approach to dynamically generate video scenes for user studies in an immersive video environment. It requires comparatively low effort (compared to field studies and full 3D environments) and combines realistic experiences with full control over what stimuli participants are exposed to. We demonstrated its usefulness in the context of a case study investigating navigation behaviour when encountering pedestrian signage. While the approach is subject to some limitations (no free movement, overlays only suitable for some types of content), it can enable researchers studying human-environment interactions to carry out lab-based studies that expose participants to realistic stimuli. In the future, we plan to extend and apply this approach in a number of different contexts, e.g. in studies exploring sense of place or focusing on specific situations where currently there is a lack of knowledge of what factors affect decisions.

Acknowledgements. We acknowledge the Volkswagenstiftung for providing funding for our project "Distance-Keeping: Influence of the StreetScape (Dist-KISS)" in the call "Corona Crisis and Beyond – Perspectives for Science, Scholarship and Society".

References

Bowman, D. A. and McMahan, R. P.: Virtual Reality: How Much Immersion Is Enough?, Computer, 40, 36–43, https://doi.org/10.1109/MC.2007.257, 2007.

Calogiuri, G., Litleskare, S., Fagerheim, K. A., Rydgren, T. L., Brambilla, E., and Thurston, M.: Experiencing nature through immersive virtual environments: Environmental perceptions, physical engagement, and affective responses during a simulated nature walk, Frontiers in psychology, 8, 2321, 2018.

Feng, Y., Duives, D., Daamen, W., and Hoogendoorn, S.: Data collection methods for studying pedestrian behaviour: A systematic review, Building and Environment, 187, 107 329, 2021.

Filomena, G., Kirsch, L., Schwering, A., and Verstegen, J. A.: Empirical characterisation of agents' spatial behaviour in pedestrian movement simulation, Journal of Environmental Psychology, 82, 101 807, 2022.

Katkere, A., Moezzi, S., Kuramura, D. Y., Kelly, P., and Jain, R.: Towards video-based immersive environments, Multimedia Systems, 5, 69–85, 1997.

Ostkamp, M. and Kray, C.: Supporting design, prototyping, and evaluation of public display systems, in: Proceedings of the 2014 ACM SIGCHI symposium on Engineering interactive computing systems, pp. 263–272, 2014.

Rossetti, T. and Hurtubia, R.: An assessment of the ecological validity of immersive videos in stated preference surveys, Journal of choice modelling, 34, 100 198, 2020.

Slater, M. and Wilbur, S.: A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments, Presence: Teleoperators & Virtual Environments, 6, 603–616, 1997.

Stenkamp, J., Karic, B., Scharf, P., Verstegen, J., and Kray, C.: Using an immersive video environment to assess pedestrians' compliance with COVID distance keeping interventions, Interacting with Computers, https://doi.org/10.1093/iwc/iwad021, 2023.

Zhao, H., Thrash, T., Kapadia, M., Wolff, K., Hölscher, C., Helbing, D., and Schinazi, V. R.: Assessing crowd management strategies for the 2010 Love Parade disaster using computer simulations and virtual reality, Journal of the Royal Society Interface, 17, 20200 116, 2020.