Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 59-11 (2025) 252–257

Creating and Measuring Immersion in Open Public Spaces with *Ariadne's Fibres*

Tim Schneider[1,4], Céline Clavel[2], Gérard Kubryk[1,3], Michèle Gouiffès[2], Emmanuelle Frenoux[2], Matthieu Courgeon[4], Vincent Hulot[5], Guillaume Junot[5], and Xavier Maître[1,5]

[1]*Université Paris-Saclay, CEA, CNRS, Inserm, BioMaps, Orsay, France*
[2]*Université Paris-Saclay, CNRS, LISN, Orsay, France*
[3]*Forum des images, Paris, France*
[3]*Cervval, Brest, France*
[5]*La métonymie, Le sas, Orsay, France*

Abstract: Immersive experiences and art installations are becoming ever more popular. While they are usually developed using specialized technology (such as head-mounted displays) or dedicated, fully controlled environments (often black rooms with video projections on the walls and floor), the implementation of immersive installations in open, public spaces (such as city squares or shopping malls) remains a particularly challenging and interesting task: How to deal with unfavorable conditions on site (noise, stray light etc.) and how to effectively engage passers-by? In this paper, we present the generative evolutive interactive installation *Ariadne's Fibres,* located in one of the busiest shopping malls in Paris, France. With advanced tracking of passers-by, it provides a unique experimental platform for field research on what may catch attention and trigger immersion. In this context, *Ariadne's Fibres* is part of an ongoing study investigating, through trajectory analysis and questionnaires, the key parameters for immersion in open, public spaces.

Keywords: Real-time algorithms, Remote sensor data acquisition, Multi-sensor systems, Multi-modal interaction

1. INTRODUCTION

Immersive experiences in art and entertainment are becoming ubiquitous, whether it be in live performances, installations, or exhibitions. Oftentimes, specific hardware such as head-mounted displays (HMDs) and headphones are used to target the experiencer's senses (e.g. Blanca Li's *Le Bal de Paris*, developed by BackLight[1]) or the experiencer is invited into a controlled environment – typically a dark and isolated room – equipped with video projections on the wall and floor as well as spatialized sound setups (e.g. the exhibition *teamLab Planets* in Tokyo, Japan[2]). However, immersive installations may also be embedded in open, public spaces (such as a city square or a shopping mall) which represent a different kind of unconstrained immersive environment where experiencers, instead of being paying customers filled with expectations and excitement, start as mere passers-by with no intention whatsoever to follow the proposed reality shift. Achieving immersion in such spaces is challenging, both in terms of user engagement and technological development, and, to the authors' knowledge, only comparatively few such installations have been successfully realized with even fewer having been considered for scientific research.

In this work, we present the interactive audio-visual installation *Ariadne's Fibres*, located in a public street in the shopping mall *Forum des Halles* in the centre of Paris which seeks to captivate passers-by and stimulate immersive experiences in open spaces. We describe the technical aspects of the installation and highlight in particular the specific challenges that needed to be overcome. Moreover, we present a custom developed tracking code used for improved real-time interactivity as well as for an automated analysis of passer-by trajectories. The trajectory assessment is part of an ongoing field study investigating the key parameters for immersion in open spaces. While a full presentation of the study goes beyond the scope of this paper, the study design will be briefly described at the end of this article.

Figure 1: *Ariadne's fibres*, a living mirror created by Ikse Maître in 2018 with music by Vincent Hulot and developments by Matthieu Courgeon (Cervval). Users interacting with their avatars and the video bubbles that they generate when moving. ©2024 La métonymie.

[1] https://backlight.fr/portfolio/le-bal-de-paris-de-blanca-li/
[2] https://teamlabplanets.dmm.com/en

Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2025.09.557

2. ARIADNE'S FIBRES

Ariadne's Fibres (AF) is a permanent, digital artwork that can be characterized as a living mirror, a place where images, which move through us, are shared and exchanged. AF is an artistic mirror that, in multiple ways, connects the inside and the outside: first, by using outdoor weather data to control the behaviour of an indoor installation, second, by offering users a new perspective of the insides of their own bodies, and third, by displaying film clips programmed inside the *Forum des images* (the cinema where the artwork is installed) on the outside (Figure 1).

2.1 Technical setup

The installation consists of a video wall (7.7 m wide and 4.5 m high, composed of 4 × 2 4K LCD screens, see Figure 2, left), a floor video projection zone (7.7 m × 4.8 m, illuminated by two M-Vision Laser 23 000 video projectors, see Figure 2, right) and 4 amplified ambient loudspeakers. Interactivity is provided through 11 Kinect One cameras (each connected to a mini-PC for real-time body pose streaming) and one wide-angle RGB security camera mounted to the ceiling above the interaction zone. Two HP workstations (one Z4 and one Z8) process the user inputs (body pose and movement), compute interaction results, render the graphics and control the audio output.

2.2 Installation environment

The artwork was commissioned by *Forum des images* and the City of Paris, France. It is installed in the shopfront of *Forum des images* where it is freely accessible every day from 8 am to midnight. *Forum des images* itself is located in the shopping mall *Forum des Halles* and specifically in a busy passageway that connects (among other places) a large movie theatre complex (the world's leading cinema in terms of admissions with more than 2.2 million spectators a year for 27 cinemas and 3,819 seats), to the entrance of Châtelet-les-Halles, the main Parisian metro and suburb train station (Europe's biggest underground station with more than 33 million travellers a year).

2.3 Interaction and installation scenarios

In its idle state, the visuals of AF, displayed on the video wall and projected onto the floor, consist of three main layers: a bright background of periodically changing colour, a 2D real-time fluid dynamics simulation, and large 3D communication bubbles that slowly grow and sway from left to right, up and down. The communication bubbles, in turn, display images or teaser videos related to the current programme proposed at *Forum des images* (Figure 1 and Figure 3, bottom). Twelve music pieces were composed for AF which, arranged into twelve tracks, are played randomly (only one out of the twelve tracks in the idle state). When passers-by enter the interaction zone, i.e. by walking in front of the video wall or across the floor projection, they are detected by the Kinect cameras and in the top-view camera, causing avatars to appear on the wall and the floor as side and top reflections associated with the visitors. Two versions of the avatars were designed for the ongoing field study: The first avatar (Figure 3, top row) is principally composed of the anatomical fibre systems (nervous, blood, lymphatic and muscular systems) and will sometimes also reveal the bone structures. The avatars resemble the users in shape and size and follow their movements and poses. Their appearance and dynamics (are the fibres stiff or supple) are furthermore impacted by the local weather (live weather data streamed via https://openweathermap.org/) and change with the seasons. The second avatar version (Figure 3, bottom) uses solid, anthropomorphized bodies of different animals (lion, cat, cow, rat, …) which can turn about their vertical axis and move their arms and legs. The avatars follow the movements of the users but are the same size for everyone. Besides the personalized avatars, the users in the interaction zone also introduce a heat source in the real-time fluid simulation which manifests itself in additional disturbances and vortices appearing on the screens and in the floor projections.

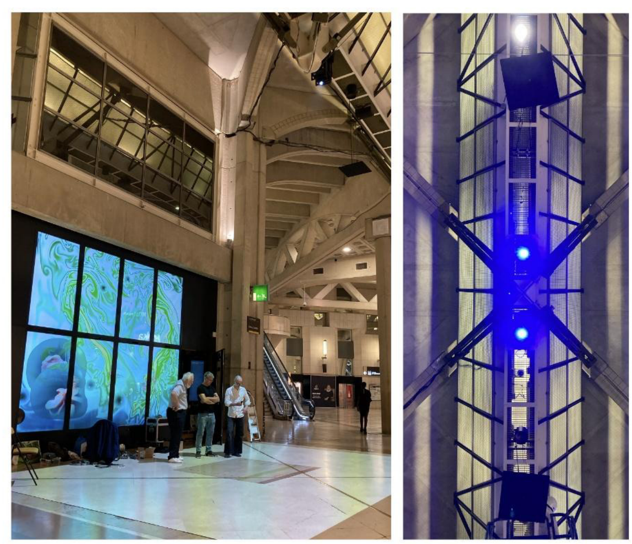

Figure 2: Technical setup of *Ariadne's Fibres*. Left, a video wall composed of 4 × 2 screens around which 11 Kinect One cameras are distributed (too small to be invisible in the image) and an interaction zone defined by video projection on the floor. Right: Two video projectors mounted to the ceiling alongside a wide-angle top-view security camera (just below bottom video projector) and two square ultrasonic directional loudspeakers (not yet in use).

In the ongoing field study, the user experience will be evaluated for multiple scenarios of the installation AF. The scenarios vary in different parameters expected to be relevant for an immersive user experience (e.g. floor video projection on or off, fibrous or animal avatars). Figure 3 shows photos of passers-by interacting in three of these scenarios.

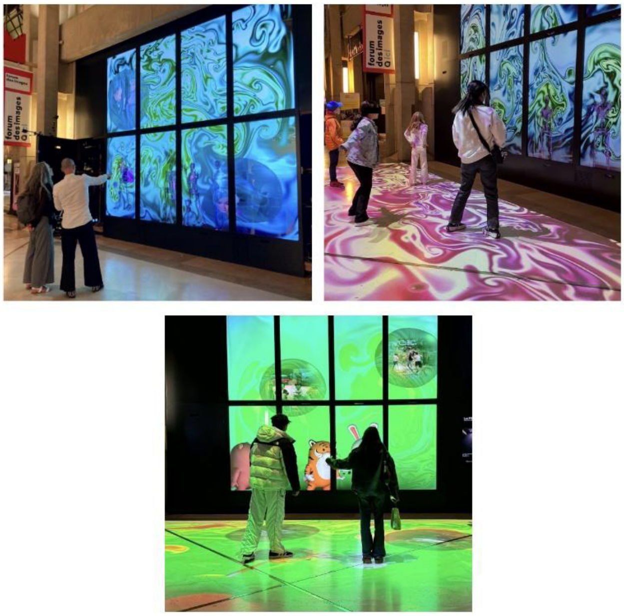

Figure 3: Three different interaction scenarios of AF. Top left: Screen wall only, fibrous avatars. Top right: Screen wall and floor projection, fibrous avatars. Bottom: Screen wall and floor projection, animal avatars.

2.4 Site-specific and control system challenges

With AF, the intention was to create an immersive artwork openly in the passageway of a busy shopping mall, integrated into the existing infrastructure (i.e. into the shopfront of *Forum des images*). For this, a bespoke cabinet with shelves, doors and mounting brackets was created and installed in the outer wall of *Forum des images*. While the installation of the eight screens was relatively straightforward, the same narrow space also needed to house the 13 computers, a network router and the power supplies. This equipment creates a considerable amount of heat which needs to be evacuated through ventilation slits in the shopfront. Moreover, additional fans had to be installed to improve air circulation in the closed cabinet. On the flipside, the ventilation slits also let dust and dirt enter into the cabinet (an effect enhanced by the cooling fans) which makes periodic cleaning and maintenance necessary to prevent excessive dust buildup and its negative impact on cooling and material wear.

From a perspective of control systems, the most important challenges are related to the size of the interaction zone (about 8 m × 6 m, considering margins around the video projection zone), the lighting conditions (bright lights in adjacent shops, quickly varying video projection onto passers-by in interaction zone) and the necessity for real-time interaction (total latencies should ideally be < 40 ms, corresponding to > 25 frames per second, for smooth interaction). The main input channel for interaction is constituted by the Kinect One depth cameras which (via the corresponding SDK) provide real-time person and pose detection as well as person tracking capabilities. The sensitive depth range of these cameras is roughly 0.5 m to 5 m and their field of view is about 70° × 60° (Kurillo et al., 2022). To cover the full width of the interaction zone, 11 cameras are spread out over an area of about 8 m × 2.5 m in and around the screen matrix in two rows (at about 30 cm and 2 m from the ground, respectively). However, this also multiplies the technical complexity of the installation: Firstly, since only a single Kinect per computer can be streamed, requiring the use of 11 extra mini-PCs; secondly, because the 11 sensors, the screen wall and the video projection must be registered to a single virtual and actual frame of reference. This is done using a custom Kinect signal integration application running on the main machine to which the Kinect-streaming mini-PCs are connected via LAN. Sensor calibration is performed using a custom VR application: With the help of two controllers, the operator can rotate and translate the 3D point clouds corresponding to the depth data of each Kinect sensor in order to patch together and align the eleven point clouds in the digital reference frame. The virtual space is then aligned with the actual physical world by selecting with the controllers prominent features of the virtual point clouds displayed in the VR headset (e.g. the corner of a chair or the operator's knee) and matching them with their physical counterparts. While the matrix of 11 Kinect sensors allows for sufficient lateral coverage of the interaction zone, people passing far from the cameras were often poorly detected and tracked. Moreover, in a busy passageway where up to 20 or 30 people may pass by simultaneously, occlusion becomes an important problem for continuous tracking. To address this challenge, we installed a wide-angle RGB camera 9 m above the floor (Figure 2) to provide top-view images of the interaction zone (Figure 4). The camera is connected to a dedicated computer via LAN which continuously reads the camera stream and tracks passers-by in the interaction zone using a custom developed algorithm. The tracking results are integrated in real-time with the data of the 11 Kinect sensors, thereby providing a more robust tracking across the entire interaction zone. The algorithm is explained in more detail in the next section.

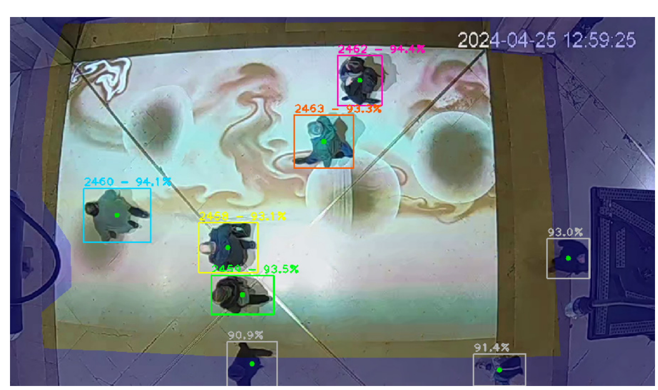

Figure 4: An analysed top-view image of passers-by in front of *Ariadne's Fibres*. Detected persons are indicated by a rectangular bounding box with coloured/grey boxes corresponding to tracked/untracked persons, respectively. The text over the boxes states the tracking ID (whence in the tracking zone) and the detection confidence in percent. The blue area around the edges of the image corresponds to the non-tracking zone.

3. PASSERS-BY DETECTION AND TRACKING

A custom, real-time person detection and tracking algorithm was developed for the top-view images streamed by the RGB camera installed above the interaction zone. This algorithm complements the tracking provided by the 11 Kinect sensors and enables an automated analysis of the trajectories of passers-by which can be used for research purposes (see section 4). The algorithm can be broken down into two main stages: detection and tracking.

3.1 Person detection

A custom machine learning algorithm was used for the detection of passers-by in the top-view camera images. The model is based on the *nano* variant of the object detection model YOLOv8 (Varghese & M., 2024) which is able to provide real-time inference (25-30 FPS) on consumer-grade hardware (NVIDIA Quadro P4000 graphics card). While YOLOv8 provides well-performing pre-trained models, the specific conditions of the installation (top-view, video projection onto persons) require a custom refinement training which was carried out on 1432 manually labelled images captured with the top-view camera under various lighting conditions. The training process yielded the best model after 350 epochs (training patience 100 epochs) with the following performance metrics[3]:

Precision:	0.9694
Recall:	0.9661
mAP50:	0.9819
mAP50-95:	0.8151

As a detection output, the model predicts a rectangular bounding box for each detected person as well as a confidence score of the detection (Figure 4).

The detection of persons using top-view camera images is challenging since most common detector networks have been pre-trained on frontal or side views of people and usually only a comparatively small portion of the body is visible in top-view images. Yet, an important advantage of these top-view images is that passers-by are continuously visible by the camera, i.e. no occlusions can occur by one person passing in front of another. A particularity of the installation *Ariadne's fibres* is the use of floor video projections which were often visible in the camera images on and around the bodies (Figure 4). Person detection in such conditions is especially challenging but could be achieved by including such images in the training data for the detection model.

3.2 Person tracking

A custom tracking algorithm was developed to group together individual detections across different video frames and thereby establish a track for each unique passer-by. The algorithm associates the detections in a new frame with the detections in previous frames by comparing the positions (centres of the bounding boxes) and applying a proximity criterion. The general concept goes as follows:

- For each detection in the current video frame, check if the last position of any of the previous tracks is within a threshold radius (in this case 80 px for an image size of 1280 px × 720 px).
- If yes, then choose the closest previous track and add this new detection to it.
- If not, then start a new track with a new tracking ID.
- A track will terminate if no new detections were added to it for a given number of frames (in this case 30 frames, corresponding to 1 second).

This approach is in principle sufficient for the given setup, as persons cannot be occluded in the top-view camera images. However, the predictions of the detector model are naturally imperfect and failure to detect persons in certain frames as well as double detections of a single person can occur. To address these issues, the tracking algorithm includes further filters:

1. Before getting to the tracking stage, detections are passed through a filter that removes double detections of a single person. The filter considers all detections in pairs and checks if the *intersection over union* of their bounding boxes is greater than 0.9. If so, one of the detections from the pair is discarded.
2. The tracking threshold radius is adjusted for each track depending on the number of frames since the last detection was added to the track. Concretely, the threshold is allowed to increase from 80 px (last detection one frame ago) to 120 px (last detection 30 frames ago).
3. A *tracking zone* was defined in the centre of the image to mitigate detection issues near the image edges. The principle of this tracking zone is as follows: While persons can be detected anywhere in the image, only detections where the centre of the bounding box is inside the tracking zone are considered for the tracking stage. This helps to reduce the recounting of people moving close to the edges of the camera image where they will frequently oscillate between a detected and an undetected state. The complement of this zone (i.e. the *non-tracking zone*) is represented in Figure 4 by a blue overlay.

The performance of the tracking algorithm was evaluated using a set of videos recorded at AF, ranging from 7 to 30 min. The number of unique passer-by IDs predicted by the tracking algorithm was compared to a manual count, showing an over-estimation by the algorithm by about 10% to 30%. This means that a single person passing in front of AF may be temporarily lost by the tracking algorithm before being redetected and assigned a new ID. This results in small discontinuities of the avatars that are assigned to interacting

---

[3] A perfect model would have a score of 1.0 for each metric. More details can be found at: https://docs.ultralytics.com/guides/yolo-performance-metrics/?q=#how-to-calculate-metrics-for-yolo11-model

passers-by. In practice, however, the tracking accuracy was found to be acceptable.

4. USER EXPERIENCE STUDY

The unique location of the installation *Ariadne's Fibres* in the passageway of a shopping mall and the possibility to track passers-by using the top-view camera constitute an ideal experimental platform for the investigation of immersive experiences in open, public spaces. A corresponding field study is ongoing at the time of the writing of this article. In the following, the two-part design of the study shall be briefly sketched out.

4.1 Trajectory analysis

The custom detection and tracking algorithm described in section 3 can also be used to assess the trajectories of people passing in front of and interacting with the installation *Ariadne's Fibres*. As stated at the end of section 3.2, the accuracy of the real-time tracking algorithm is about 10-30%. However, by applying a set of post-processing filters to recorded video and tracking data, the accuracy may be improved to about 1-3%. Then, entire trajectories can be reconstructed from which the position, walking speed and direction, and dwelling time of individual passers-by can be retrieved[4]. This data may be used for an automated classification of the trajectories (hereafter called *tracks* for brevity):

- *Passing track*: The track of a person who crosses the tracking zone without stopping or interacting.
- *Interaction track*: The track of a person who interacts with the installation AF inside the tracking zone.
- *Stopping track*: The track of a person who stops within the tracking zone but does not engage in explicit interaction with the installation AF.
- *Other track*: A track that does not fall into any of the above classes.

While the specific classification criteria are being validated at the time of the writing of this article, it appears that the track length, the track duration, variations in local track speed, as well as the linearity/directedness of a track all seem to be useful discriminators. This is illustrated by Figure 5 which plots the position and instantaneous speed data for manually selected examples of passing, interaction and stopping classes.

The automated classification can then be used to assess the fractions of passing, interaction and stopping tracks at different times of the day and for different days of the week. Crucially, the assessment of these fractions also provides a quantitative measure for the comparison of the different installation scenarios of AF (see section 2.3 and Figure 3) and thus a means to investigate what installation parameters stimulate or inhibit engagement and interaction among passers-by.

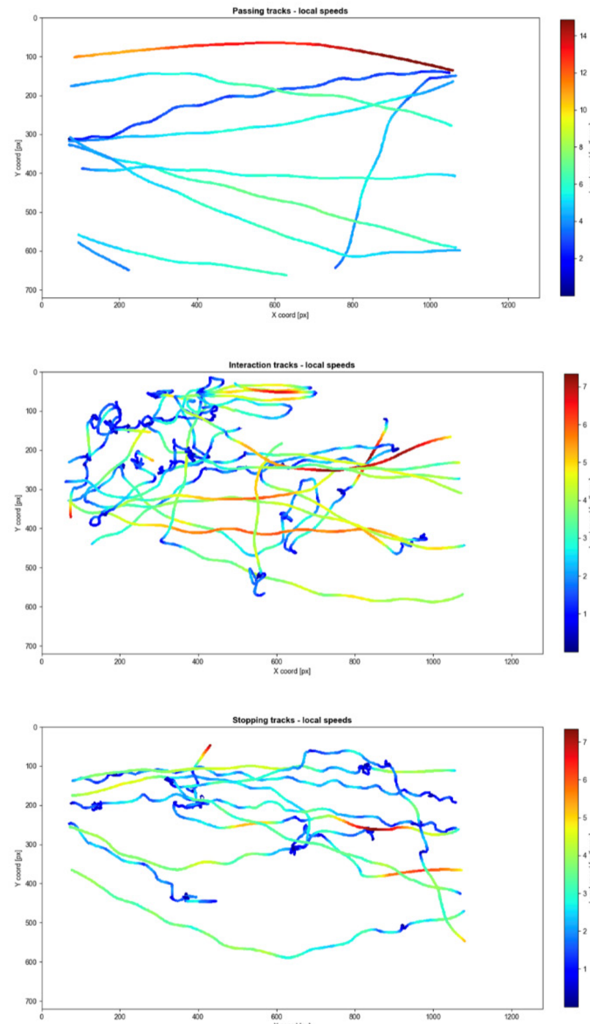

Figure 5: Maps showing the instantaneous speed along the 10 manually selected tracks for each of the three track classes: passing tracks (top), interaction tracks (centre) and stopping tracks (bottom). The X and Y coordinates correspond to the pixel positions in the top-view camera images (cf. Figure 4).

4.2 Subjective user experience

The second part of the field study uses the *Spatial Presence for Immersive Environments* (SPIE) questionnaire developed by Khenak et al. (Khenak, 2019) to assess the subjective user experience. This questionnaire was specifically designed for immersive systems and measures a participants' experience along seven dimensions: spatial presence, affordance of the interaction with the installation, enjoyment, perceived realism of the installation, attention paid to the activity, cybersickness, and perceived social presence of other users' presence when avatars are present. These measurements will again be carried for the different installation scenarios, thereby providing complementary data to the track classification.

[4] Note that this data is anonymous and cannot be connected to individual, physical persons.

4.3 Comparison with other immersive experiences

The field study is carried out as part of the European project *Artcast4D* (https://www.artcast4d.eu). In this context, the authors already employed the SPIE questionnaire with participants of three other interactive experiences of varying degrees of immersivity: (i) a virtual reality experience, (ii) a 360-degree video projection setup, and (iii) a touch- and motion-sensitive spherical sculpture with inner video projection. These three installations/experiences were exhibited in different public environments, namely an art gallery, a cultural and shopping center, and a museum. The comparison of the questionnaire results obtained with participants of AF with the results of the three other experiences will allow it to further widen the investigation of and consider other parameters that may impact immersion (such as the device/screen setup, interaction mode or exhibition environment).

5. CONCLUSIONS

In this paper, we presented the interactive, multi-user art installation *Ariadne's Fibres*. The installation is located inside one of the busiest shopping malls in Paris, France, which poses specific technical challenges for robust multi-user interaction, namely coverage of a large interaction area, simultaneous tracking of many persons, and mitigation of person occlusions. These challenges are addressed by a custom developed composite tracking system, combining 11 Kinect One depth cameras and real-time machine learning detection and tracking in top-view images of a wide-angle RGB camera.

Due to its special location and the advanced tracking system, the installation provides a unique experimental platform for field research on immersive and interactive experiences in open, public spaces. In the context of the *Artcast4D* project, a field study on the keys for immersion in public spaces is ongoing where the user experiences in different scenarios of *Ariadne's Fibres* are analysed and different types of immersive installations are compared. Here, we briefly sketched out the study design which involves an automated assessment and classification of passer-by trajectories as well as a questionnaire to measure the subjective experience of interacting passers-by.

ACKNOWLEDGEMENTS

Ariadne's Fibres was commissioned by *Forum des images* and the City of Paris. It was included in a comparative field study carried out in the context of the project Artcast4D which received funding from European Union's Horizon Europe research and innovation programme (grant agreement No 101061163).

REFERENCES

Khenak, N., Vézien, J.-M., & Bourdot, P. (2019). The Construction and Validation of the SP-IE Questionnaire: An Instrument for Measuring Spatial Presence in Immersive Environments. In P. Bourdot, V. Interrante, L. Nedel, N. Magnenat-Thalmann, & G. Zachmann (Eds.), *Virtual Reality and Augmented Reality* (Vol. 11883, pp. 201–225). Springer International Publishing. https://doi.org/10.1007/978-3-030-31908-3_13

Kurillo, G., Hemingway, E., Cheng, M.-L., & Cheng, L. (2022). Evaluating the Accuracy of the Azure Kinect and Kinect v2. *Sensors*, 22(7), 2469. https://doi.org/10.3390/s22072469

Varghese, R., & M., S. (2024). YOLOv8: A Novel Object Detection Algorithm with Enhanced Performance and Robustness. *2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS)*, 1–6. https://doi.org/10.1109/ADICS58448.2024.10533619