AI Visitor

Tracking Pedestrian Trajectories for Machine Learning Applications in Machu Picchu, Cusco, Peru

Paloma GONZALEZ, Massachusetts Institute of Technology, USA
Takehiko NAGAKURA, Massachusetts Institute of Technology, USA

Abstract: Studying and managing pedestrian movement is essential for cultural heritage sites and public spaces. Much current software, such as evacuation simulators, typically uses ruled-based approaches and visualizes how people move. However, it is difficult for such software to provide critical simulations of visitors that wander through touristic sites. Solving this problem through a data-driven approach, this paper presents the data processing method for Machine Learning applications that simulates how human agents navigate space. This research aims to develop intelligent agents that emulate human pedestrians moving in response to their spatial environments, including architectural features (AF), which refer to spatial aspects of the environment, such as peculiar buildings that attract visitors. AI Visitor is a prototype that includes a pipeline of onsite data collection and the training of agents. This paper introduces the case study of the Machu Picchu citadel, built-in Peru during the 15th century by the Incas, attracting approximately 2000 visitors a day. The raw data was collected through extensive aerial video recordings from drones, providing 0.5 meters of data granularity. Tourists' trajectories were extracted through computer vision, and the AF were identified and scored. Next, two Machine Learning techniques were combined; Reinforcement Learning (RL) used the AF as input for training the agent, and Imitation Learning (IL) deployed human path trajectories as demonstrations for the agent. The combined model helps train complex behaviors of the agent efficiently from a relatively small group of datasets. As a result of applying the data processing method to areas of the Machu Picchu case, the preliminary result indicates the trained agents autonomously trace human trajectories data and move in search of architectural features. The potential use of such data-driven pedestrian modeling includes applications to circulation and facility design of the sites, capacity management of visitors, and administration of social distancing.

Keywords: *Pedestrian Simulation—Machine Learning—Video Tracking—Cultural Heritage Site*

CHNT Reference: Gonzalez, P. and Nagakura, T. (2022). 'AI Visitor. Tracking Pedestrian Trajectories for Machine Learning Applications in Machu Picchu, Cusco, Peru', in Börner, W., Rohland, H., Kral-Börner, C. and Karner, L. (eds.) *Proceedings of the 25th International Conference on Cultural Heritage and New Technologies, held online, November 2020.* Heidelberg: Propylaeum.
doi:10.11588/propylaeum.1045.c14511

Introduction

Contemporary pedestrian simulation tools generally deal with limited aspects of human behavior, such as emergency egress and traffic crossing. Most of these tools use procedural models such as rule-based ones. For instance, Agent-Based Models (ABM) enable the modeling of sophisticated crowd dynamics (Pedica and Vilhjálmsson, 2008). MassMotion is a procedural tool of ABM for egressing pedestrian simulations, among others[1]. While these tools are helpful in specific simulation contexts, they exclude more complex human behaviors. One relevant example is trajectory paths sidetracked to approach an interesting view or freely traversing space sightseeing on cultural heritage sites.

Including these exploratory behaviors in a simulation tool increases the accuracy of simulating humans in an everyday context and enables modeling deviations from expected trajectories. However, building such a tool requires new insight into how to retrieve and embed such human behaviors. The AI Visitor presented in this paper applies a Machine Learning model to exploratory pedestrian movement through heritage sites to simulate and predict the anticipated direction of people in real-time. This would be useful, for example, for designing a circulation path of visitors and managing the capacity of the venue with proper social distance.

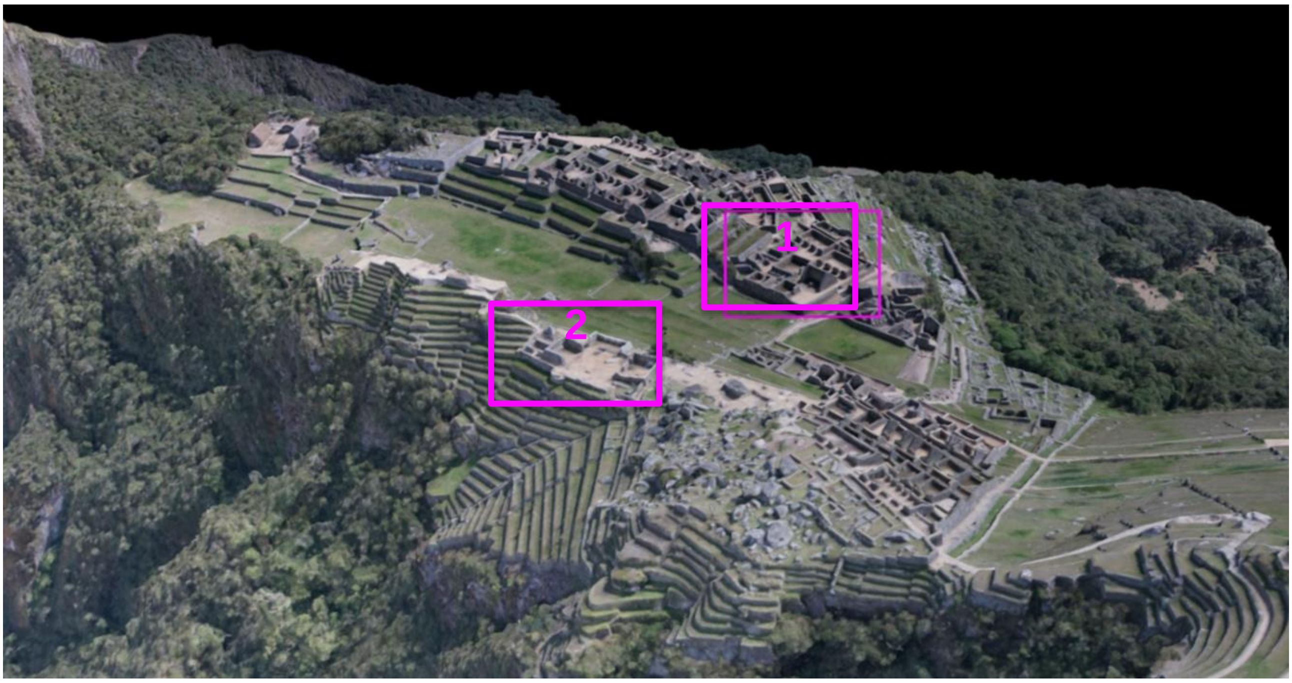

Fig. 1. A photogrammetric model of Machu Picchu (Model by Cesar Medina, 2018). The framed area with the number '1' is the "Two Mirror Temple." The framed site with the number '2' is the "Three Windows Temple." (Model by Cesar Medina, 2018).

AI Visitor's input and data processing pipeline includes onsite data collection, pre-processing, and the training of agents used for simulation. At the outset, a drone flying over the site is used to record the human visitor's movement in video and photograph the site's architectural conditions for 3D photogrammetrically modeling. The raw data then is pre-processed to harvest the human trajectories by tracking the recorded pedestrians through a computer vision software and the locations of architectural features identified and scored as visitor's attractions. The ML model then feeds on the pre-processed data and trains the agents in a simulated, digital site version. After training, agents can finally be tested by comparing their movement back to the recorded human visitors.

[1] *Legion* is one of the most relevant software today, also rule-based, and destined to evaluate the emergency egress of airports and hospitals. Space Syntax is another ABM procedural tool that evaluates human sightlines and the accessibility of space as a graph. Space Syntax also has a module to analyze 'isovists,' the range of vision of a person or agent in a specific location, limited by the shape of the three-dimensional space. These tools are rule-based and depend on the modeler's knowledge, often harvested from experts, regarding the target behavior. Several researchers, such as Ng and Russell (2000), have questioned the limitations of relying on agent designers. They stated that they might have only a rough idea of optimizing a model to generate a desirable behavior.

The Machu Picchu citadel, located in Cusco, Peru, is the case study for this project (Fig. 1). This study is a significant and ideal cultural heritage testing ground with few roofs or coverings to impede observation. The selected areas are where free walking is permitted, and visitors wandering around the site are frequently observed alongside guided tourists. As seen in any specific cultural heritage site, the visitors in these areas actively contemplate various attractions, including peculiar scenes and monuments found during their traversing the space. Therefore, those areas are appropriate and advantageous sampling sites of human trajectory data and their motivating attractions as the source dataset for training the agents.

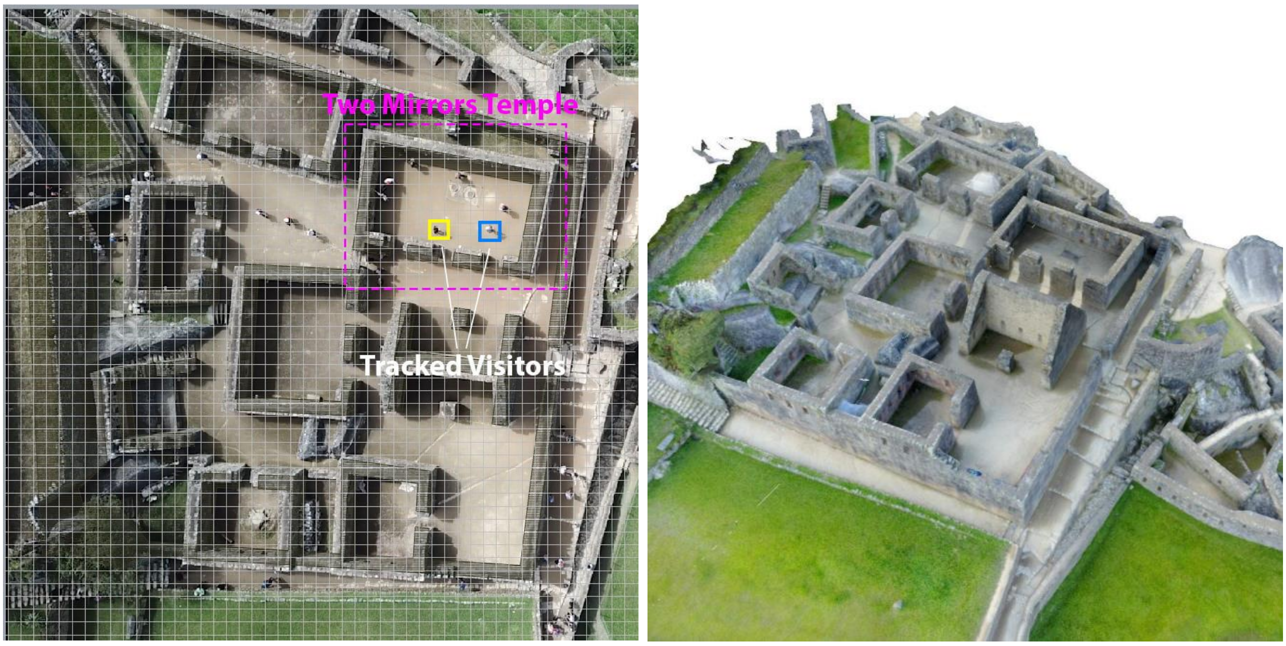

Fig. 2. Right: Two Mirror Temple area in Machu Picchu is captured into a photogrammetric model (Nagakura, 2018). Left: The pedestrians marked with yellow and blue boxes near the entries of the 'two mirrors' site are being tracked in the video recorded from a drone. © Authors.

Field Data Collection and Pre-processing

The field data were collected in Machu Picchu citadel, located on a steep mountain ridge, approximately 2500 meters above sea level. The data collection starts with video recording from a drone (DJI Mavic Pro) hovering at the same position at 50 to 70 meters in height. The flight time was approximately from 20 to 25 mins at a time, constrained by battery duration. The recordings of pedestrians at 4K resolution covered an area of 50 × 50 meters. The collected data comprises 30 mins of pedestrian movement at each of the five different regions of Machu Picchu; only two spots were analyzed, shown in Figure 1. At any moment, a typical video recording includes 9 to 10 'free walkers' and 60 to 80 people in groups with guides. The recorded data were processed using the OpenCV toolkit to extract visitors' trajectories. The data is discretized into steps every one second (Fig. 3).

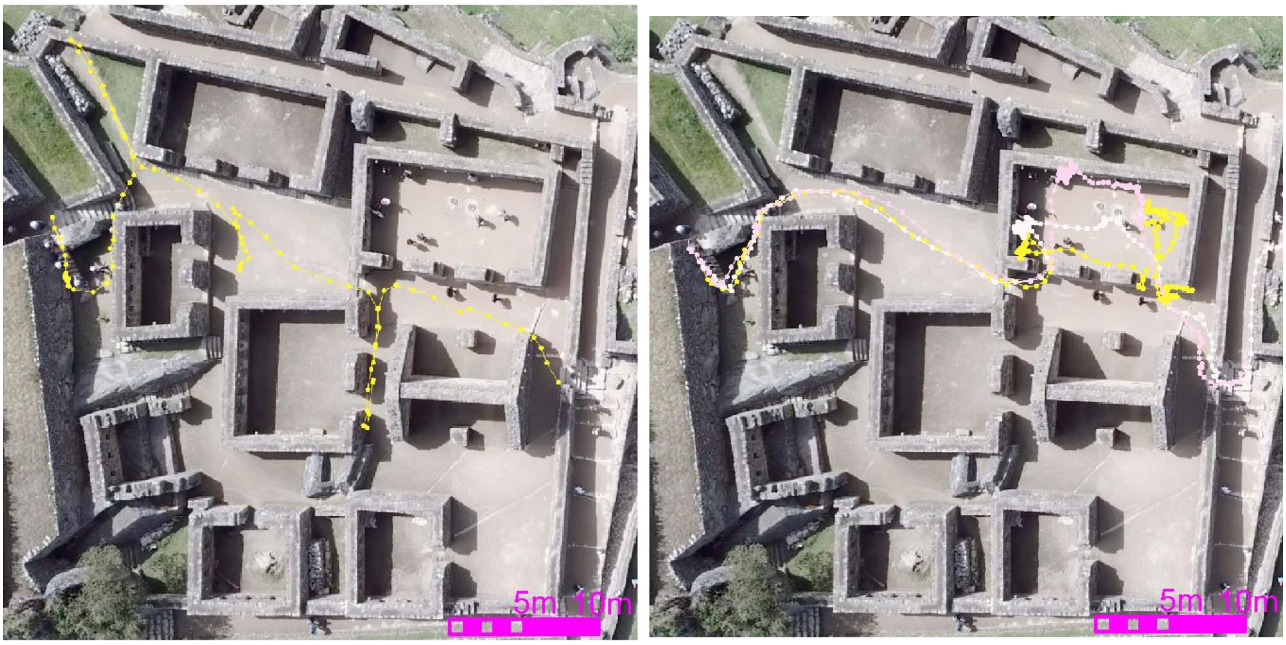

Fig. 3. The discretized route of a 'free walker' visitor (Left) and a guided group of visitors stop at the Two Mirror Temple that traces the path prescribed by the Machu Picchu park (Right). © Authors.

Forty free walkers were initially tracked from each of the two selected areas of the site (Fig. 1): "Two Mirrors Temple" (Figures 2 and 3) and "Three Windows Temple." Also, forty guided people in groups of 6 to 9 visitors from both areas were tracked. The walkers were classified into those two groups, free walkers and guided tourists, by applying a Bayesian model classifier called "Bishop" written and developed in Python by Julian Hara-Ettinger (2012). The classifier compares the extracted trajectories to the "most probable route" between two points, assigning a score to the main steps of the extracted trajectories. Next, the trajectories with the lowest scores are deemed unguided explorers and classified as 'free walkers.' This method filtered the explorer trajectories from all the rest, such as those of tourists in guided tours or workers moving on duty, in the dataset.

In addition to the video used for pedestrian tracking, thousands of photos were taken from the drone flying over the site. Using photogrammetric software, those photos were processed to generate the textured mesh model first. Using it as a reference then produced a simplified 3D model, appropriate for agent training sessions later in Machine Learning.

Further along with the data processing, a unique data label, the 'architectural features' (AF), is introduced to support the computation of the agent's interaction with the environment. AF is defined as 'appealing spatial configurations' that attract the trajectories of the visitors, often leading them to explore. Examples of AF are an exciting view such as one commanding the nearby mountains and another looking down on the excavation site from a high point; a building with unique formal or material conditions; its parts such as a pediment, a terrace, and a window; and a peculiar residual space such as a narrow corridor or a large stepped balcony. Particular objects of attractions are included in AF as well. For example, the 'Two Mirrors' in Machu Picchu are two small circular cavities containing reflective water located in the temple-like enclosure within the highlighted area in the left image of Figure 2. The AF analysis explains why people often deviate from the prescribed tourist paths towards specific locations.

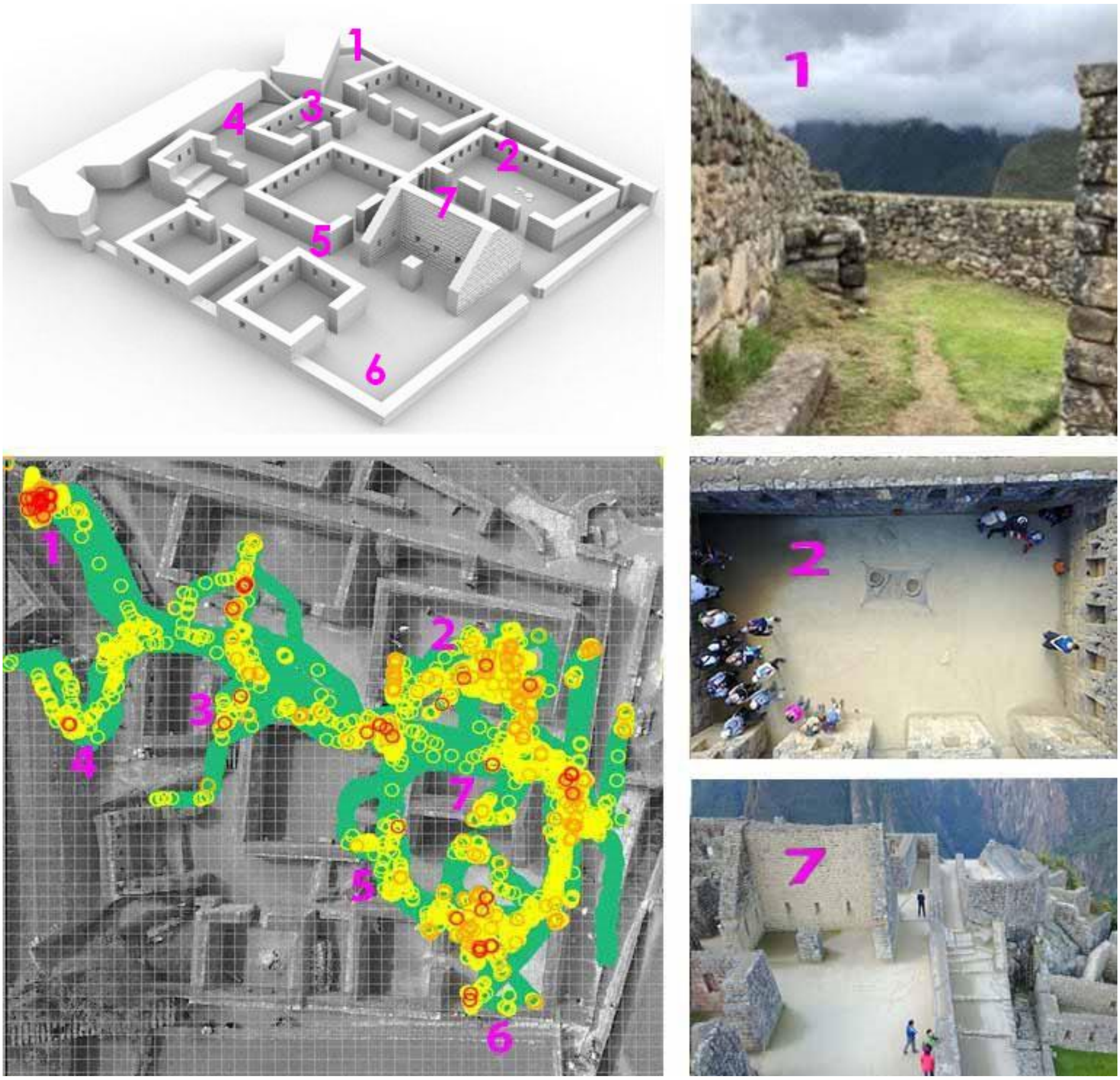

Fig. 4. Two Mirror's temple, showing the scored architectural features (AF): 1. Terrace with a view towards the nearby mountains, 2. Two Mirrors (round pools on the ground), 3. Building with Rock, 4. Large Tiers, 5. Building with windows towards the main grass field, 6. Terrace towards the main grass field, 7. Building with a pediment structure without the roof. The bottom left image shows a heat map with the hotspots of the trajectories. The red circles represent the locations where the visitors spent more than 1.0 seconds, the orange circles represent the locations where the visitors spent more than 0.5 seconds and less than 1.0 seconds. The yellow circles represent the visitors' locations spent less than 0.5 seconds. The green areas are the traces of the paths. The steps are defined every 100 cm distance, approximately 1.0-second intervals—the images on the right show three AF locations. © Image by the authors and photos on the right column courtesy of Eytan Man.

Identifying the AF relies on three sources: the first source is the visitors' behaviors and the amount of time they pause at a location on their trajectories. A shared long pause is evaluated to indicate an attractive feature for visitors. The second source is the official map of Machu Picchu and the information given by the guides, signs, and other onsite authorities, which advises the visitors of the critical and famous locations of the site. Research collaborators with architectural design backgrounds identified additional potential attractions on the site. Figure 4 shows all the AF identified in the "Two Mirrors Temple" site. For example, the "Two Mirrors" is one of the well-known attractions in Machu Picchu and is placed on the official map distributed to all visitors at the main entrance gate into the citadel park. Its location and photo are numbered 2 in Figure 4. On the other hand, places like the one numbered 7 in Figure 4 are not introduced in the Machu Picchu map but are identified as unique architectural features because of their shapes and ruined conditions distinct from others nearby.

Another example is the location numbered 1, a terrace that many people are observed flocking to visit and lookout. In Figure 4, this terrace shows a red hotspot where many visitor trajectories pause. Although the official Machu Picchu map indicates nothing is there, it is a popular location for visitors because of the grand panoramic view towards the nearby mountains. According to the data in Figure 4, there are many such anonymous locations where many people are proven to pause and pay attention.

Each of the AF identified on the site is rated by a scoring system that measures the average time spent by the visitors at the respective location. The highest score is for the most popular site, including the “Two Mirrors,” and is equivalent to 7. The lowest score is for the visitors’ location spent the least time. The place numbered 5 in Figure 4, ‘the building with windows towards the main field,” got this score. In the next step of data processing, these scores are normalized and mapped to the reward system for the RL model, resulting in rewards in the range between 0 and 1.0.

Overall, the limitations of the Data Production for this research were: to collect human trajectory data only from a public space and exclusively from visitors of a tourist destination—the tourist site’s regulations highly constrained the behavior of those visitors. For example, they are allowed to stay for a determined amount of time, can only walk in one direction, and cannot wait for long periods in any location. Furthermore, the visitors are only middle-aged to older adults with good physical shape to endure physical activity. The data does not include people with challenged mobility. Finally, the data was collected during which the tourists visiting Machu Picchu travelled from the North Hemisphere. Therefore, this dissertation dataset was limited to the data collection conditions. Finally, the assumption is that this analysis only includes first-time visitors.

Machine Learning Process

This research is inscribed in modeling intelligent agents to simulate pedestrian behaviors. Agent-Based Models (ABM) is its typical rule-based modeling method, as previously mentioned, and verbal behaviors, such as walking the shortest path to leave a room in an emergency, are formalized into rules, embedded into the agents, and sequenced for simulation. However, how and why humans traverse the environment beyond deterministic paths is often not obvious. Identifying and describing a set of rules for modeling agents in ABM is not a trivial task even for experts. For instance, it is difficult to define rules for visitors exploring the cultural heritage sites, who often seem to move unpredictably by frequently deviating from their prescribed path to various locations, in such a context where machine learning methods augment well-established rule-based models.

The motivation of this paper is to present an alternative, data-driven method of agent development that applies Machine Learning. It computationally samples and deploys the likely patterns of human behavior from the field data containing varying degrees of noise and outliers instead of using the researchers’ well-defined expert knowledge or analysis. In Machine Learning terms, agent behaviors are modeled by extracting policies that map policies assigning probabilities to agent’s actions at a given state are initially derived from the recorded human trajectories and updated in the machine learning procedure. In the Machu Picchu project, the onsite human trajectory data and architectural representation of the site were combined to drive the Machine Learning process to train the agents, whose behavior is stipulated by evaluating what architectural elements of the environment prompt visitors' actions to wander and deviate from prescribed paths.

AI Visitor applies two Machine Learning techniques, Reinforcement Learning and Imitation Learning, combined to overcome the shortcomings of each. The idea of using Imitation Learning to complement Reinforcement Learning for training agents has been found in previous research projects such as the ones by Pfeiffer et al. (2018), Youssef et al. (2019) and Juliani et al. (2020). Imitation learning generally helps train the agents efficiently with a relatively small training dataset used as demonstrations. Still, the trained agents often can only replicate the behavior 'taught' by such expert demonstrations. The challenge is to surpass such a limitation by combining this method with another traditional Reinforcement Learning, that uses a more generic but time-consuming reward-driven system for learning.

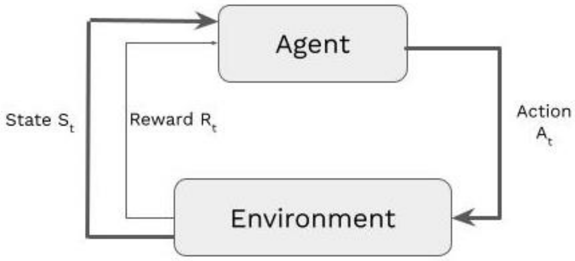

Fig.5 Basic structure of a Markov Decision Process (MDP) for Reinforcement Learning Agents. Adapted from Sutton and Barto (2018).

Reinforcement Learning (RL) is a method used to program intelligent agents to make valid sequences of decisions/actions. The structure of this agent model consists of state, actions, rewards, and policy (Fig. 5). The state corresponds to the current position of the visitor agent, and the action is its step-by-step movement. The reward is the cumulative amount of stimulus that the visitor agent collects from the environment, for instance, when an attraction is found on the site. The policy is a strategy for finding the rewards developed by the agent after training. The policy guides the agent about what action it should take in the next step which way to move or pause next. As the RL training sessions continue, the agent maps its training experience into a policy described in a probabilistic format. And as the policy develops, it maps agents' states to actions, so the visitor agents take advantage of the policy in deciding their next move.

In detail, the RL in AI Visitor lets an agent collect a reward when it encounters an instance of Architectural Features (AF). In the current implementation, the pre-processing stage before the Machine Learning training sessions identifies all the AF instances on the site. It adds each AF instance a score quantified from the ground data. As previously described, this score derives from the average time visitors spend at each AF location and establishes the rated rewards in the RL environment. During the training, the agent then maximizes the prize by visiting the AF locations, with each defined as a bounding box of 3 × 3 × 3 meters for the efficiency of RL computation. However, this bounding box can represent an attraction physically found beyond its location, such as a distant mountain view that a visitor can command from that location. When the agent approaches an AF instance, the collision detection is activated, and a reward is obtained.

The other Machine Learning technique used is Imitation Learning (IL), in which the agent emulates the ‘experiences of others by using them as the training input for learning. In AI Visitor, the experience of others is the experience of human visitors exploring the site and signified by the trajectories recorded in the field. Using these human movements as demonstrations, the agent learns which way to move or pause. In other words, the expert demonstrations are trajectories, with each instance of them represented in the form of: $\mathrm{t}_i = (s0, \mathrm{a}1, \mathrm{a}2, \mathrm{a}3, \ldots)$

The first term is the original state; a sequence of actions is taken during runtime. Only the data from the free walkers were used for Imitation Learning in the Machu Picchu project. The tourists in guided groups are excluded.

The Imitation Learning process assumes policies extracted from the experience of human visitors are appropriate for training agents. Therefore, if it is used for training a game competitor, the training dataset would require a careful selection to include only good demonstrations provided by solid players. For AI Visitor, there are no better or worse tourists to play the role of demonstrators. But identifying free walkers from guided tourists and others such as onsite workers is essential, and this classification method was discussed earlier in this paper. The Imitation Learning model works efficiently with a relatively small dataset once appropriately selected. At the same time, the Reinforcement Learning process often requires very intensive training cycles to achieve a similar result. The agent in the Imitation Learning process possibly learns and traces the demonstrations, including the nuanced behaviors human tourists make. However, because AI Visitor’s Imitation Learning process only considers the sequence of expert actions without explicitly utilizing the AF in the environment, it alone does not train the agent to properly behave in a situation much different from the demonstrated cases. The more robust, time-consuming Reinforcement Learning process using AF for rewards complements this problem in AI Visitor.

The training process in AI Visitor applies Reinforcement Learning and Imitation Learning simultaneously and is conducted in a simulated digital environment. In the Machu Picchu project, a simplified 3D model was reconstructed from the photogrammetric capture of the site first. The agent in the training software was forced to move on the recorded human trajectories. At the same time, it is presented with the isovist (i.e., the steering view around the agent’s head) of the reconstructed model (Fig. 6). At any location, the isovist commands the spatial configuration of the site from the vantage of the agent, with indications of all the architectural features (AF) in it. Through this training, the agent learns how to decide its movement in response to the surrounding spatial environments, including AF.

The ML model of AI Visitor was implemented using the Unity 3D ML-Agents Toolkit (Juliani et al., 2020). This toolkit implements ML on the widely used game engine software and generates intelligent agents. The ML process is defined as “model-free,” using only the experience extracted from the training data to achieve the optimal behavior of the agent. The Toolkit includes Proximal Policy Optimization (PPO) that optimizes the policy during the training sessions. The setup of the action space consists of 4 actions, forward, backward, rotate left and right. There are two types of rewards: a) Discrete rewards from following the demonstrations accurately, and b) AF rewards at specific locations, with a penalty of -0.001 at each training update, discounting from the cumulative reward if the agent is not changing location. Overall, the learning rate of the agent improved significantly when RL and IL were combined, as expected from previous work and the toolkit documentation.

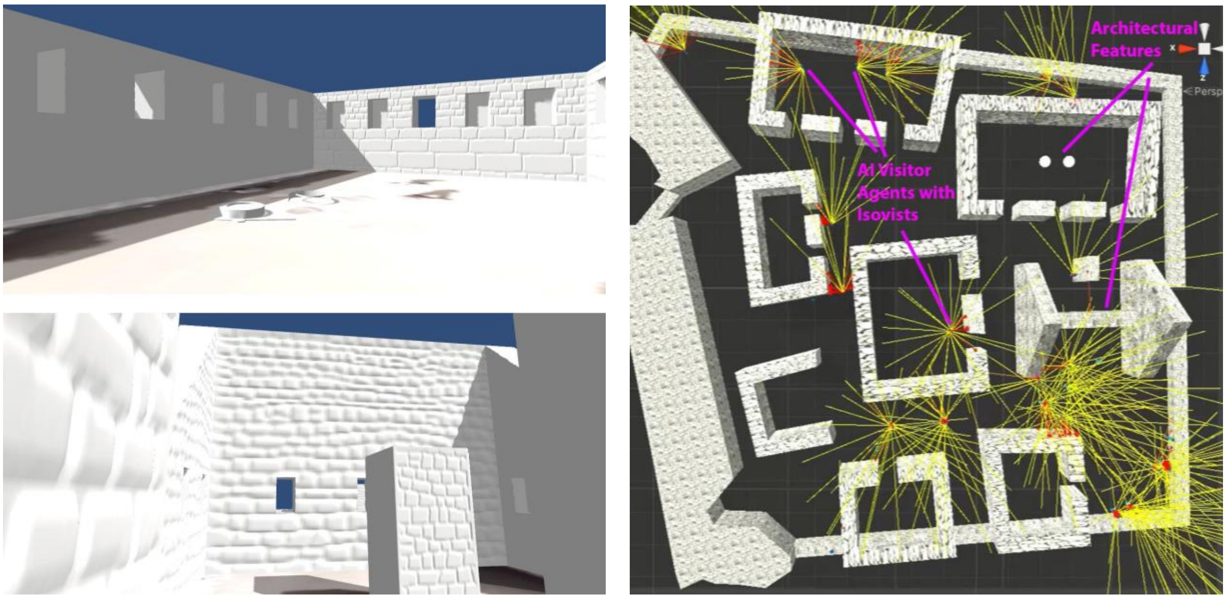

Fig. 6. Machu Picchu digital model reconstructed from aerial captures and used as the simulated environment for Reinforcement Learning. Right: A plan view of the model shows the agent's visitors (red dots) exploring the site with indications of their isovists (yellow lines). Left: Example views correspond to the isovists at two locations. © Authors.

Preliminary Result and Future Applications

The trained agent in the Machu Picchu project has been tested by putting it back to the Two Mirror Temple model, where the field data was sampled from the human visitors. Its behavior was validated when its movement was observed compared to the human trajectories (the ground truth) in the source training set. This result indicates a strong potential of the proposed training method as a means of adequately embedding human explorers' behaviors on cultural heritage sites and the use of such an agent for effectively simulating how people deviate from predetermined routes.

The main contribution of this paper is the framework for developing a new pedestrian simulation tool for cultural heritage sites that uses architectural features as the motivation for visitors' exploratory movements and applies a Machine Learning method to train the intelligent agents for simulation. It described the methods for field data collection, extraction and classification of human trajectories, identification, and scoring of the architectural features, and applying the combined Imitation Learning and Reinforcement Learning model. Using these methods in the case of the Two Mirror Temple area in Machu Picchu, the trained agent was produced from a dataset of human trajectories and architectural features of the site.

The critical future step for this research is to test how the agent trained through the field data extracted from the Two Mirror Temple area moves when it is placed on other resembling areas of Machu Picchu. The evaluation should be made by comparing the simulated behavior of the agents to that of the human visitors recorded on those sites. If they are reasonably similar, pedestrian simulations created by the agents elsewhere would likewise be trusted. The measurement of similarity includes how the agent deviates from the prescribed paths and the locations and durations of the pause, indicating its reaction to architectural features.

The Machine Learning method described in this paper intends to make the agents replicate the behaviors extracted from the human visitors included in the training set. The factors for the optimization of the agent have the number of the sampled trajectories and the selection of one or more areas for recording the human trajectories and architectural features in them. While many portions around the center of Machu Picchu citadel include architectural formation analogous to the Two Mirror Temple area, other nearby spatial environments have long stairs, large stepped terraces, or open fields not found in the Two Mirror Temple area. The agent probably would need more training sets from those areas to simulate visitors of Machu Picchu in a comprehensive scope.

As for the architectural features, representations in Building Information Modeling (BIM) will be considered in the future development to move away from the manual process of AF identification described earlier in this paper. Suppose the target cultural heritage site model is available in BIM format. In that case, it is described as a composition of architectural elements classified by the types such as walls, floors, stairs, trees, terraces, and various subtypes of these. Computational tools for discretizing the mesh model from 3D scanning into a BIM model representation are emerging (Braun and Borrmann, 2019). The architectural features for AI Visitor can be prepared by systematically extracting relevant architectural elements in BIM data and interpreting them for features meaningful to visitors of the cultural heritage sites.

The current AI Visitor prototype is built around the Machu Picchu case study to prototype visitor agents there. Once this data-driven foundation is coupled with a good interface for applying data from any cultural heritage site, valuable tools for specific simulation purposes such as circulation and facility planning, capacity management, and social distancing administration can be developed to study, design, and manage the site efficiently.

Acknowledgments

A part of the fieldwork included in this paper was derived from the Cusco and Machu Picchu expeditions funded by the MISTI MIT-Peru program directed by Eduardo Rivera and conducted in collaboration with UNSAAC, the Cusco Culture Directorate of Peru. The authors thank Director Cesar Medina, Prof. Rene Navarro, and Prof. Wilbert Salazar for all the support provided during the fieldwork sessions and the MIT students who participated in the operation. Part of the field data was processed by Kimberly McPherson, MIT Undergraduate researcher, in 2020. Special thanks to Dr. Daniel Tsai for his valuable comments and assistance.

Funding

Please provide information if and by which institutions the research was funded. If applicable, provide the grant number of the funding agency.

Conflict of Interests Disclosure

Please disclose any financial or personal relationships with other individuals or organisations, such as sponsors, that could make your work appear biased or influenced.

Author Contributions

The authors contributed equally to the research presented in the paper.

References

Braun, A. and Borrmann, A. (2019). 'Combining inverse photogrammetry and BIM for automated labeling of construction site images for machine learning', *Automation in Construction*, 106, 102879. doi:10.1016/j.autcon.2019.102879

Jara-Ettinger, J., Baker, C.L., and Tenenbaum, J.B. (2012). 'Learning What is Where from Social Observations', *Proceedings of the 34th Annual Conference of the Cognitive Science Society*, pp. 515–520.

Juliani, A., Berges, V., Teng, E., Cohen, A., Harper, J., Elion, C., Goy, C., Gao, Y., Henry, H., Mattar, M., and Lange, D. (2020). 'Unity: A General Platform for Intelligent Agents', arXiv preprint arXiv:1809.02627. https://github.com/Unity-Technologies/ml-agents.

Narahara, T. (2007). *The Space Re-Actor: Walking a Synthetic Man through Architectural Space*, Thesis. Massachusetts Institute of Technology.

Pedica, C., and Vilhjálmsson, H. (2008). 'Social Perception and Steering for Online Avatars'. In Prendinger, H., Lester, J., and M. Ishizuka (eds.), *Intelligent Virtual Agents*, pp. 104–116. Springer. doi:10.1007/978-3-540-85483-8_11

Pfeiffer, M., Shukla, S., Turchetta, M., Cadena, C., Krause, A., Siegwart, R., and Nieto, J. (2018). 'Reinforced Imitation: Sample Efficient Deep Reinforcement Learning for Mapless Navigation by Leveraging Prior Demonstrations', IEEE Robotics and Automation Letters, 3(4), pp. 4423–4430.doi:10.1109/LRA.2018.2869644

Sutton, R. S. and Barto, A. G. (2018). *Reinforcement Learning: an Introduction*. Second edition. Cambridge, Massachusetts: The MIT Press.

Unity-Technologies/ml-agents. (2021). [C#]. Unity Technologies. https://github.com/Unity-Technologies/ml-agents (Original work published 2017).

Youssef, A., Missiry, S. E., El-gaafary, I. N., ElMosalami, J. S., Awad, K. M., and Yasser, K. (2019). 'Building your kingdom Imitation Learning for a Custom Gameplay Using Unity ML agents', *2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, pp. 0509–0514.