AI Visitor: Tracking and simulating pedestrian trajectories in Machu Picchu

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

Keywords: *pedestrian simulation— machine learning— video tracking— cultural heritage site*

CHNT Reference: Paloma Gonzalez, Takehiko Nagakura, 2020, AI Visitor. Proceedings of the Conference on Cultural Heritage and New Technologies. DOI:TBT

Introduction

This paper presents how a machine learning model gauges the effect that architectural features have on human trajectories. Architectural features refer to spatial aspects of the environment such as peculier buildings and scenery, configured as appealing to visitors. No convenient tools are currently available for designers to simulate how architectural features influence the pedestrian movement.

Contemporary pedestrian simulation tools generally deal with basic aspects of human behavior such as emergency egress and traffic crossing. Most of these tools use procedural models such as the rule-based ones. For instance, Agent Based Models (ABM) enable modeling sophisticated crowds dynamics (Pedica & Vilhjálmsson, 2008). MassMotion is a procedural tool of ABM for egressing pedestrian simulations. SpaceSyntax is another ABM procedural tool that evaluates human sightlines and the accessibility of a space as a graph. While these tools are useful in specific contexts of simulation, they exclude more complex human behaviors such as side tracking the trajectories to approach an interesting view, or freely exploring space during sightseeing.

Including exploratory behavior does not only increase the accuracy in simulating humans in everyday context, but also enables to predict how the expected, deterministic trajectories can vary. Building such a tool requires a new insight for retrieving human behavior in space. The RL model in this paper incorporates exploratory and deterministic navigation. Consequently, it demonstrates the simulation that replicates and predicts the anticipated paths of people in real time, and would be useful, for example, for designing a heritage site circulation, and managing the capacity of the venue with proper social distance. The method for applying the RL Model consists of three parts: data collection, training and simulating.

At the outset, the RL model feeds on real human trajectory data and architecture models, and identifies architectural features on the digital version of the site. A method using aerial videos and photogrammetric models to streamline this potentially laborious and exhausting process is shown in this paper. The field data is captured through the aerial video for tracking human reactions to the architectural settings, that are also captured from aerial photography. Machu Picchu citadel in Peru is the use case scenario for this project.

Fieldwork

The data collection starts with video recording with a drone (Mavic Pro) staying at the same position at 50 to 70 meters height from the target area for 20 to 25 mins at a time, constrained usually by battery duration. The recordings of pedestrians at 4K resolution covered an area of 50 x 50 meters. The collected data comprises 30 mins of video total at five different locations in Machu

Picchu. A typical video recording shows 9 to 10 ‘free walkers’ and 60 to 80 people in large groups with guides. The recorded data was processed using OpenCV toolkit for extraction of the trajectories of the visitors.

Machu Picchu visitors are studied because the place is an ideal architectural attraction and without roofs. The selected zones are where ‘free’ walking is possible. In a cultural heritage site, typical visitors exclusively contemplate the scenes and monuments while traversing the space. Therefore, the assumption is that this is a significant advantage for processing the human trajectory data and assigning ‘intention’ to the paths by means of these attractions on the site.

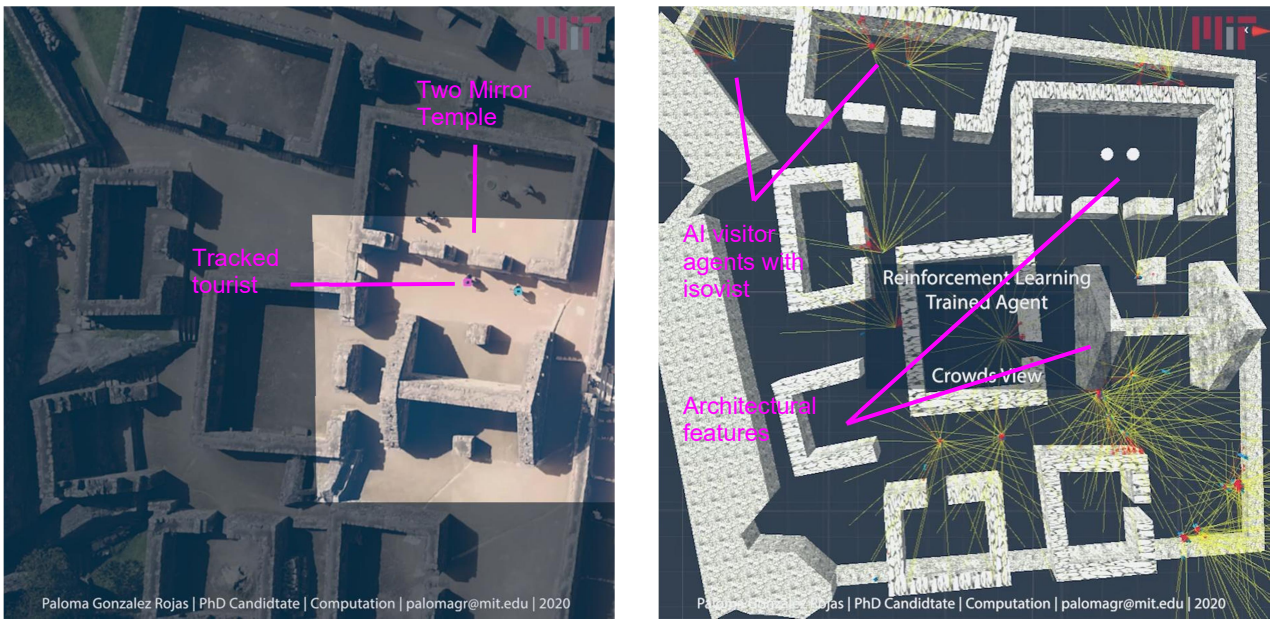

Fig. 1. Machu Picchu, Two Mirror Temple Aerial view. (Paloma Gonzalez, 2018) Left image: the colored pedestrians near the ‘two mirror’ site (located just above the highlighted portion) are being tracked in the video. Right image: In the virtual 3D environment reconstructed from the photogrammetric model, intelligent agents moving through the site are trained through the Machine Learning process, with consideration to their isovist.

The data extracted from the video is discretized to be read by the RL model and get indicators to score and label the architectural features of the site. A group of 40 free walkers’ were initially tracked from each of the two selected areas of the site; “Two Mirrors Temple”, corresponding to the aerial image in Fig. 1 and “Three windows temple”. 40 guided people in large groups of 6 to 9 visitors from both areas as well. The walkers were classified in those two groups using a Bayesian model classifier, developed by Julian Hara-Ettinger called “Bishop”. The classifier compares the trajectories to the “most probable route” between two points, assigning a score to the main steps of the trajectories. The data is also processed in a ‘time spent’ heatmap for labeling of the architectural features.

Machine Learning and Analysis

Predetermined behaviors are easy to model into a finite number of rules and recombined to recreate the original behavior in a procedural simulation. However, determining a model of accurate human behavior in space is very difficult without the data from human movement on real sites or just through analytical approach. For the Machu Picchu case, the onsite human trajectory data and reinforcement and imitation learning techniques combined in a model is deployed to train

agents, and simulate how architectural features motivate humans to explore the cultural heritage site.

Architectural features (AF) are ‘appealing spatial configurations’ that attract the trajectories of the visitors, many times leading them to explore. As for the Machine Learning application, the first key idea of this research is the application of reinforcement learning (RL) to analyze the AF. RL is a machine learning method used to program intelligent agents to make a good sequence of decisions/actions. The agent maps its experiences to decisions, to make a policy. In the case of the AI Visitor, the RL model is used for the identification and labeling of AF, and association of AF and pedestrian trajectories. Examples of architectural features are: an interesting view, a peculiar building or its parts such as a temple or stone wall formations, or special objects such as the ‘two mirrors’ at Machu Picchu, the small circular cavities containing reflective water, located just above the highlighted area in Fig.1. The AF analysis explains why people deviate from the prescribed tourist paths towards certain locations. The AI Visitor agent is trained through the statistical indicators from the field data with the labeled AF and assigned scores. In this case both the data from 'free walkers' and deterministic pedestrians are included.

The second idea is to use imitation learning, a particular flavor of reinforcement learning that emulates the experience of others. Instead of using algorithmic exploration, this process assumes good policies extracted from experience of other people as input for training. It involves optimization, generalization and delayed consequences. Imitation learning is a suitable option to reinforcement learning when the behavior to model is very nuanced, and therefore, it is exceedingly challenging setting up the rules of the behavior, or when it would take excessive training to achieve the same result through simple RL exploration. In the case of the AI Visitor model, the human trajectory data is used as ‘the experience of others’, to show the agent how a human tourist traverses space when exploring a site. The training process considers what people see at various points of the site and relies on their isovist. Only the data from the ‘free walkers’ in Machu Picchu is used here for the training set and the tourists in guided groups are excluded.

For AI Visitor modeling, both machine learning techniques, reinforcement learning and imitation learning, are combined in the end to overcome the shortcomings of each.

Current Result and Prospect

The main contribution of this paper is to present the machine learning logic of using architectural features as the motivation for people’s movement. The first step for such realization is to produce accurate training data for the model. A credible dataset for the AI Visitor model was developed with a method replicable for further expansion. A main inquiry for the second step is to test if the data from one location is applicable to other locations. The research investigated this through the discretization and the classification of the architecture features with labeling. The current result indicates a trained model properly embedded with human behavior effectively simulates how people deviate from predetermined routes, shown in Fig. 2. The model is validated when contrasted with the human trajectory data used for the source training set as well as other field-born sets from Machu Picchu.

For future development, it is worth mentioning the details of the imitation learning in that the agent is generally only able to replicate the behavior ‘taught’ by the expert data, and the challenge is to surpass such limitation by combining it with other traditional RL means. Once the foundation of the

AI visitor model is reliably implemented, building various applications on it is possible for purposes of designing and managing the cultural heritage sites, including its circulation planning, congestion relief and capacity control, and social distancing.

References

Jara-Ettinger, J. (2012). Learning What is Where from Social Observations

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

Pedica, C., & Vilhjálmsson, H. (2008). Social Perception and Steering for Online Avatars. In Intelligent Virtual Agents (pp. 104–116).

Acknowledgements

A part of the fieldwork included in this paper derived from the Cusco and Machu Picchu expeditions funded by MISTI MIT-Peru program and conducted in collaboration with UNSAAC and Cusco Culture Directorate of Peru. Authors thank Director Cesar Medina, Prof. Rene Navarro and Prof. Wilbert Salazar for all the support provided during the fieldwork sessions as well as the group of MIT students who participated in the operation. Part of the field data was processed by Kimberly McPherson, MIT Undergraduate researcher in 2020.