Never Alone: A Video Agents Based Generative Audio-Visual Installation

Matthew Gingold
School of Interactive Art and Technology, Simon Fraser University, Vancouver, Canada.
m@gingold.com.au

Thecla Schiphorst
School of Interactive Art and Technology, Simon Fraser University, Vancouver, Canada.
thecla@sfu.ca

Philippe Pasquier
School of Interactive Art and Technology, Simon Fraser University, Vancouver, Canada.
philippe_pasquier@sfu.ca

CHI'17 Extended Abstracts, May 6–11, 2017, Denver, CO, USA.

ACM ISBN 978-1-4503-4656-6/17/05
http://dx.doi.org/10.1145/3027063.3054949.

Abstract
Never Alone (2016) is a generative large-scale urban screen audio-visual installation, which presents the idea of generative choreographies amongst multiple video agents, or ‘digital performers’. This generative installation questions how we navigate in urban spaces and the ubiquity and disruptive nature of encounters within the cities’ landscapes. The video agents explore precarious movement paths along the façade inhabiting landscapes that are both architectural and emotional.

Author Keywords
Media art; generative art; artificial intelligence; installation; algorithms; video; audio.

ACM Classification Keywords
J.5. Arts and Humanities (Performing arts); H.5.1. Multimedia Information Systems (Artificial, augmented, and virtual realities); I.2.0 Artificial Intelligence (General).

Introduction
Never Alone comments on the overlapping relationships between materiality, poetics, aesthetics and the human experience of interiority and asks how these concepts can express themselves within the emerging field of computational creativity. Computational creativity refers to the applications of machine learning and/or

Figure 2: Twelve physical performers were filmed traversing a climbing wall in specifically enumerated choreographed phrases

artificial intelligence that are simultaneously aesthetic, artistic, computational and material: an inquiry into the capacities of human creativity, interaction, active viewership and the limits to which machines can express creative behavior.

Additionally, an aspect of computational creativity is also how human participants may achieve creative agency through interaction and the application of algorithmic instructions to physical and emotional performance. We are interested in the growing omnipresence of Artificial Intelligence in our societies and lives. From video games to online love, from driving cars, flying planes, controlling drones to taking part in online trading and Google searches - the notion of the 'artificial agent' has become a prevalent yet primarily unconscious or invisible paradigm to embodying machines in virtual spaces. More importantly, a critical approach to revealing the 'black box' of AI agents' potential behaviors also speaks to the need for artistic intervention that takes into account the complex interplay between human agency and AI.

Generative Choreographies

Never Alone uses a simple movement 'alphabet' or motion graph for each of the agents' movement vocabulary. Additionally, we borrow from Laban Movement Analysis (LMA) concepts such as effort, weight, and space in formulating and performing these 'alphabets'. We relied on Laban's concepts for choreographing dancers' movement because Laban Movement Analysis provides a clear formal framework for observing, interpreting, and generating human movement and performance.

The movement 'alphabet' or motion graph of each agent is comprised not only of a direction, but also a mode of movement and an 'intensity'. Each agent can move up, down, left, right or remain stationary. Each directional movement can be achieved in one of a variety of walking, crawling, slithering or crouching. And each of these directions and modes was filmed repeatedly with variation in intention or intensity. The start and end of each 'letter' or 'node' in the graph was choreographed to superimpose with the start/end of one or more other nodes thus creating possible transition frames.

This process resulted in thousands of video clips, corresponding to basic movements, each of which can be near-seamlessly linked to another movement. Thus each 'letter' of the alphabet or 'node' in the motion graph can be treated as a discontinuous primitive, whilst combining several primitives allows the construction of pseudo-continuous phrases.

From the actions of multiple video-agents complex choreography and assemblages emerge. The video agents trace precarious paths through the micro-narratives of our data-saturated world. Unforeseen encounters create momentary connections and disruptions, embracing the temporary, forgettable, nature of data. Unlike with CGI and computer animation, the aesthetic qualities of HD video output is maintained throughout the work and is deeply informed by the studio work with the performers embodying both their own unique qualitative approaches alongside the regimented choreographic phrases that construct the movement vocabularies.

Creation and Computation

One of the core innovations of the project is to implement generative, agent-based movement in the largely non-generative medium of video. AI controlled avatars in CGI and computer game environments is ubiquitous, however the challenge of achieving similar

independence of movement using HD video is, as far as we know a new area of investigation.
In beginning the project we filmed a short test of a limited motion graph – one involving the bare minimum of nodes, or clips, for a video 'agent' to transition from standing still to walking left or right and return to standing still.

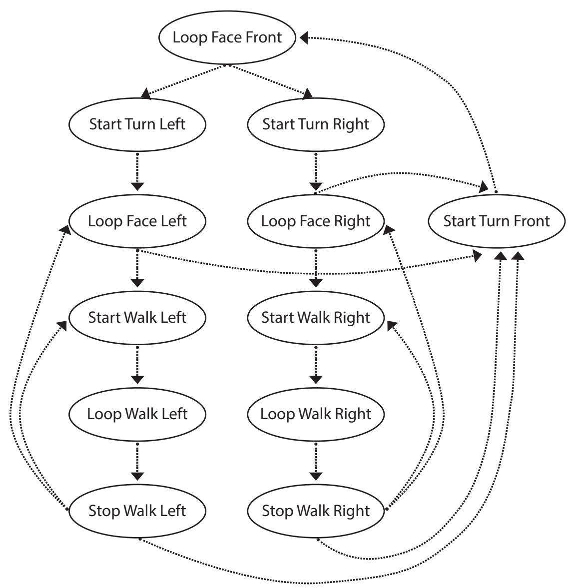

Figure 1: Simple motion graph of a video agent. Each node corresponds to a video (chosen amongst different "takes").

Based on the results of filming all the takes necessary for this 'minimal' motion graph we designed a more complex graph consisting of climbing, crouching, slithering and various static poses. A large stage and climbing wall was constructed to facilitate the process of filming all the primitive movements necessary to fulfill the full motion graph.
We then moved to the movement studio, working with 12 dancer-performers over an extended period. We explored somatic and theatrical techniques of emotional induction based on internal narratives, intertwining each performers' personal experience with performance qualities from Laban Movement Analysis[4].
In order to obtain footage at the required resolution and with a relatively 'flat' perspective (so that cuts between each node of the motion graph appear near-seamless) we filmed all performances on a Red Epic at 5K, using a FOV encompassing the entire stage and climbing frame. Where possible performances of the motion graph were shot in single continuous takes, eg., the motion graph in Figure 1 can be achieved in a single shot, with the performer starting in the facing front position, then turning to the left, walking, stopping, turning back to the front, turning right, walking, stopping, etc.

Figure 2: Production. Filming performers. Surrey BC (2014)

The resulting footage was then hand labeled and motion normalized. Labeling involved encoding XML 'markers' at the start and end of each movement phrase in each video clip. These 16 letter markers can then be used to reconstruct the nodes and edges of the motion graph, directly linking frame, clip and movement type. This was especially important given that each performance varied, both within and between performers. Motion normalization involved 'keyframing' each video so that the performer appears to remain centered within each video clip – effectively providing us with pixel displacement information for each motion in the graph. The keyframe information is then scaled and the inverse displacement is applied during motion graph reconstruction, allowing us to consistently match on-screen position whilst playing back successive clips from the database. This method also minimizes the size of the videos, an important consideration for real-time video playback of between 10 and 100 video agents. For example, instead of encoding a video with a performer walking from left to right, over let's say 2000 pixels, we only need to encode a fixed scale video big enough to fit that performer walking 'on-the-spot' and re-apply their 'walking' motion during playback. In practice we use a maximum square video resolution of 650 x 650 pixels for all performers and motions in the graph.

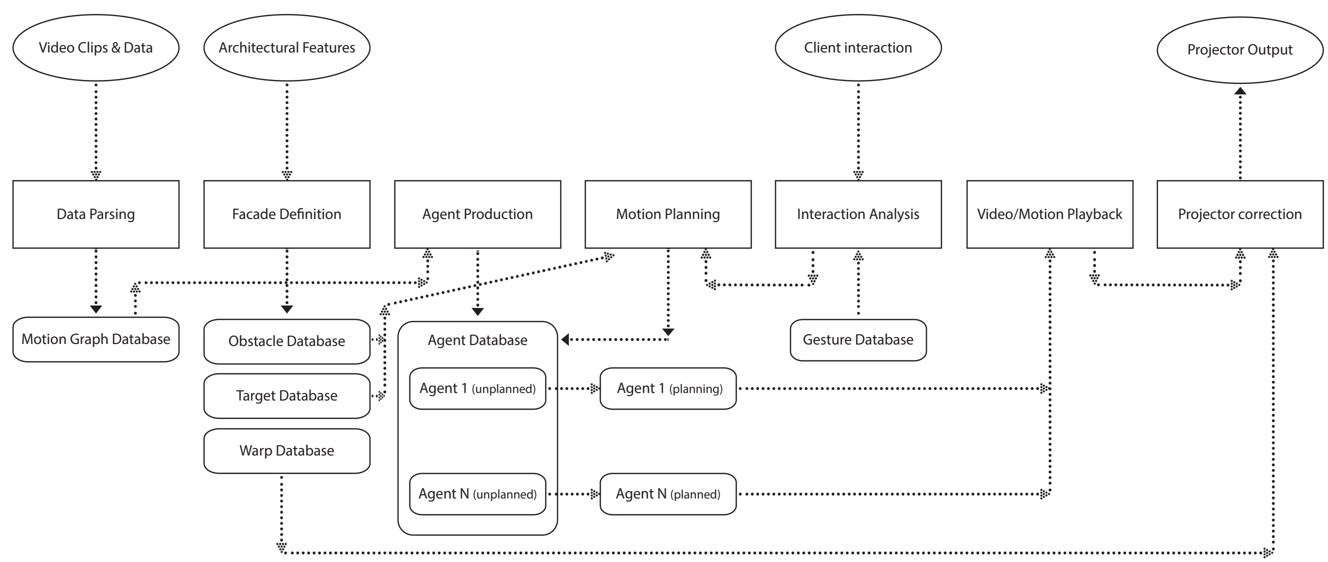

Figure 3: Software architecture overview

While a full overview of the software developed for each iteration of the project is beyond the scope of this article, it is worth noting the core components. Firstly, we construct database models of all possible motion graphs for each performer from the labeled and keyframed video data. Secondly, we use custom tools to specify the façade – windows, ledges and other features are drawn into the computer and defined as masks, obstacles or targets for movement. We also specify any warping necessary to rectify the projected image. Meta-strategies for producing agents are defined (often several strategies are randomly or user selected). Each agent then begins motion planning, often communicating with other agents to synchronize or avoid paths. In situations where user interaction is required (usually via mobile device) we filter and then analyze the input, using this to trigger re-planning. Once an agent has completed a valid motion plan, and/or playback is triggered, each agents' motion plan is played back frame by frame with the correct keyframe position applied as the video's play. The final composited image is rectified for any optical distortion and projector overlap.

Motion Planning: A*

Of key interest is the motion planning component of the installation software. We have implemented several versions of this based on variations of the A* (pronounced "A star") search algorithm. This algorithm has a long history of application in both robotic and computer game path planning. The most common approach is to divide up the search space into cells of fixed size, and assign tokens (or nodes) to movement between adjacent cells [1]. Our first iterations of the installation used this approach in a three-stage process. First we solved for paths using abstract rectilinear tokens (ie., moving left, right, up or down). We then fitted our data to these tokens by searching for valid edge connections in our motion graphs to 'transition' between these tokens. Finally, we fitted our directional displacment by foreshortening or repeating the token directions (ie., walking left, or climbing down etc). Whilst this method is the more traditional approach to applying A* to motion planning problems, it was ill-suited to our data. We have an asymmetrical graph, discontinuous and varying data. Unlike computer game characters, or even most robots, our agents do not all move the same; do not move continuously or even consistently in any direction. As such we have now implemented a modified A* based approach.

In the current iteration of the software we do not impose a grid on the topology to be explored. Instead we directly search the motion graph. That is, the current node is expanded based directly on all valid next-nodes in the actual motion graph of each agent. Cost is calculated as the sum of both the distance travelled and the total number of nodes visited. Some random noise is also added to the cost to increase the exploration of the agents. Whilst this increases search time, as well as increasing the probability that no valid path may be found, we mitigate these factors by relaxing the final goal test: instead of requiring equality we either terminate when the current node is within a set distance threshold and/or when a local minima in distance to the goal has been achieved. This results in a much higher accuracy of path planning, automatically taking into account even small displacements in 'transition' clips (eg., turning left, not just walking left).

Conclusion and Future Work

Never Alone is the next stage in a historical exploration that was originally conceived as a generative and interactive, large-scale urban video installation entitled Longing + Forgetting (2014), exploring the concept of artificial video agents colonizing architectural spaces. Based on the initial prototype, we have applied unique methods towards designing an adaptive generative audio-visual system, that allows the work to be succesfully re-configured for indoor and outdoor locations with a variety of architectural features. Future developments include:

- Expanding the movement vocabulary, with new movement interaction, including existential movements such as falling, leaping, drowning or sliding, and dialogical movements such as embracing, pursuing, hiding and avoiding;
- Improving motion planning by implementing a modified version of the RRT* algorithm, which improves exploration, is probabilistically complete and allows for anytime motion planning [2];
- Increasing the coupling between the video generation and the audio generation engine (based on the engine of Auditory Tactics [3]).

While Never Alone contributes to the state of the art on generative video through the notion of video agents, there is much more to be done.

Acknowledgements

This project benefited from the financial support of the Social Science and Humanities Research Council of Canada (SSHRC). The team would also like to express our special thanks to performers and graduate students who we involved in the production and without whom this project would not have been possible.

References

1. Hart, Peter E., Nils J. Nilsson, and Bertram Raphael. "A formal basis for the heuristic determination of minimum cost paths." IEEE transactions on Systems Science and Cybernetics 4.2 (1968): 100-107.
2. Karaman, Sertac, et al. "Anytime motion planning using the RRT." Robotics and Automation (ICRA), 2011 IEEE International Conference on. IEEE, 2011.
3. Philippe-Aubert Gauthier, Philippe Pasquier Auditory Tactics: A Sound Installation in Public Space Using Beamforming Technology. Leonardo, Vol. 43, Num. 5 - The MIT Press, 43, 426-433, 2010
4. Schiphorst, Thecla. Self-Evidence: Applying Somatic Connoisseurship to Design, CHI'11 EA on Human Factors in Computing Systems, ACM Press, 145-160.

Figure 4: top, *Longing+Forgetting* as an outdoor large-scale facade projection at the Surrey Urban Screen, BC (Surrey Art Gallery, January – April 2014); bottom, *Never Alone*, as a generative façade projection, Zentrum für Kunst und Medientechnologie, Karlsuhe, Germany (2016).