Recognition: Combining Human Interaction and a Digital Performing Agent

John McCormick[1], Steph Hutchison[1], Adam Nash[2],
Kim Vincs [1,3], Saeid Nahavandi[3], Douglas Creighton[3]

[1] Motion.lab, Deakin University, Australia, 221 Burwood Highway, Burwood 3125, Australia

[2] RMIT University, Australia, Bldg 14, Level 11, Room 12, Melbourne 3001

[3] Centre for intelligent Systems, Research, Deakin University, Australia, 75 Pigdons Road, Warun Ponds 3216, Australia

Editor: Cédric Buche

ABSTRACT — Virtual and augmented environments are often dependent on human intervention for change to occur. However there are times when it would be advantageous for appropriate human-like activity to still occur when there are no humans present. In this paper, we describe the installation art piece Recognition, which uses the movement of human participants to effect change, and the movement of a performing agent when there are no humans present. The agent's Artificial Neural Network has learnt appropriate movements from a dancer and is able to generate suitable movement for the main avatar in the absence of human participants.

Categories and Subject Descriptors — J5 [Arts and humanities]: Arts, fine and performing.

General Terms — Performance, Experimentation, Human Factors.

Index Terms — Performing Software Agent, Machine Learning, Electronic Art,Interactive Installation.

1. INTRODUCTION

Recognition is an interactive installation which uses a performing software agent to effect change in the virtual environment when there are no humans present. Recognition was exhibited at Cube 37, a glass fronted gallery in Frankston, Australia. The exhibition environment used the movement of passers-by to animate a morphing avatar which was projected onto the front of the gallery. As the exhibition ran through the night, there were often times when there were no pedestrians around from which to gather movement information, however there was still a lot of vehicular traffic passing the gallery, which was located on a major road. Rather than have a still screen at these times, a software agent was trained to move using the movement of a dancer, and this agent's movement was used by the projected avatar to animate itself. The use of learned human movement allowed the software agent to quickly acquire the capability to stand in for a human when needed. The projected forms also used pictures of the dancer's iris as a texture for its body, giving the installation a uniquely organic signature. Borrowing appropriate material from a human to quickly generate capacity for the agent was one of the features of Recognition.

Recognition is part of an ongoing investigation into digital performing agents, in particular agents that can learn to dance with a human. We viewed the agent as a performing partner and as such decided to treat the relationship between the agent and dancer in a similar manner to what might occur between two dancers. The dancers might generate dance sequences, perhaps through improvisation initially, they share and learn these movements, then in performance can use this shared movement vocabulary to perform in unison, perform independently or take cues from each other to determine what parts of the movement they might perform (semi – improvised). This is a very simplified structure however it suggested a learning model as the basis for the agent if it were to take the role of a performer. The agent

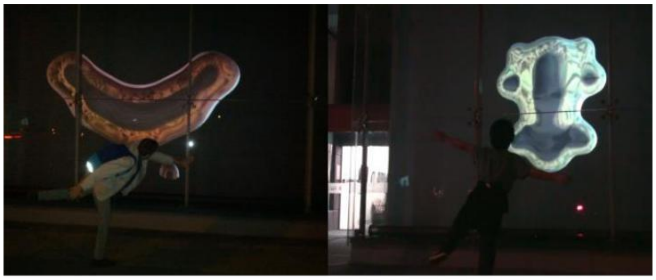

Figure 1. Human participants providing movement data for the visible avatar that is projected onto the glass frontage of the Cube 37 gallery. Image © John McCormick 2013

Laval Virtual VRIC '14, April 9-11, 2014, Laval,
France. Copyright 2014 978-1-4503-2626-1 ...$10.00
http://dx.doi.org/10.1145/2617841.2620707

would also need to be able to both generate movement and recognize what the dancer is doing. These requirements influenced our decision as to the type of structure to use for the agent's learning model. Recognition is both an interactive installation and a performance, for the majority of the time it is available for pedestrians to interact with, however at certain times a dancer may perform within the installation. The installation is essentially the same, the only difference is the familiarity of the dancer with the system and the movement choices available to the dancer due to her flexibility and experience. However even the dancer is improvising as are the pedestrians.

Recognition focuses on the agent's ability to generate movement from what is has learnt and perform this in parallel with a human participant in order to maintain a constant flow of movement data for the installation, hence in this paper we will focus on movement generation rather than movement recognition, though both contributed to our model design choices.

2. RELATED WORK

The requirement for human-like movement is very common as seen in a myriad of games, film and animation. Non-Player Characters as agents in a game are often animated with motion capture data to bestow human-like movement qualities on the characters. These libraries of movement are often blended together to form variations on the stock movements. Hsu et al. applied time warping to the learning of motion style from a library of different motion-captured walks in order to generate variations on walking style. [1] Taylor and Hinton applied Conditional Restricted Boltzmann Machines to the same task of learning movement style in walks and generating new variations from the trained network. [2] The latter in particular showed that a learning model could be used instead of a pre-recorded motion model for movement generation. However we were interested in a model that could be used for both movement generation and recognition as well as for learning a substantial movement vocabulary rather than movement style. Han et al. used a combination of Self Organising Map (SOM) and Markov Model for the prediction of human positional movement for location based services. [3] Son et al. used a Recurrent SOM to detect movement patterns in the body of Lumbriculus variegatus (a type of worm) in response to the introduction of toxins to their environment. [4] Caridakis et al. successfully employed SOM and Markov Model to track the 2D point trajectory of the hand for gesture recognition. [5] We applied SOM to full-body 3-dimensional movement as the basis for our agent's learning with a novel approach to its use for both movement generation and recognition. While Recognition illustrates the generation aspect, other examples are available at http://www.johnmccormick.info.

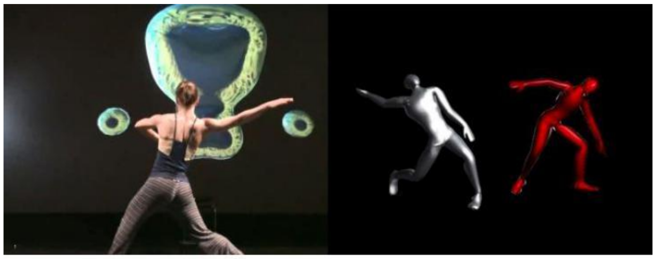

Figure 2. Human movement is tracked by Kinect and used to animate a live avatar (silver avatar on right). At the same time the performing agent (red avatar on right) is continuously generating movement based on what it has learnt from the dancer. The silver and red avatars are not visible to the viewers, their data is used by the morphing avatar (left image behind dancer) to animate itself. Image © John McCormick 2013

3. RECOGNITION DESIGN

The Cube 37 gallery has a glass front onto which imagery is rear projected, allowing it to be viewed by passing pedestrians and people in cars. A Kinect sensor behind the glass screen tracked people's movement on the street in front of the gallery. (Figure 1) The Kinect data is used to extract a skeleton representation of the pedestrian's movement, or at times, the dancer's movement, and this data is used to animate an avatar which is in the background and is unseen by the participants. The joint positions of the unseen avatar are used by the visible morphing eye to change its shape accordingly. There is also another unseen avatar representing the performing agent. When no humans are present the visible eye takes its movement data from the agent's avatar instead. Thus the avatar representing the live human and the avatar representing the performing agent work together to continuously provide movement data to the morphing eye. (Figure 2)

When a person enters the area in front of the gallery they face a giant morphing shape that looks out at them from the gallery window. It changes according to their movement, fluidly transforming like liquid or molten metal. When it solidifies into a single shape it appears like a giant eye, casting its gaze over the pedestrian. Depending on the movement of the pedestrian, it can break into smaller parts or components, leading to extremely varied morphology. It invites improvised participation, either from passers-by or at specific times, from the dancer who provided the agent's movement. It is meant to be playful and engage both participants and onlookers who can appreciate the myriad forms brought about by the participant's movements.

When there are no human participants in front of the gallery, the great eye changes in color and texture from that borrowed from the iris of one of the artists, John, to that of the dancer, Steph. The movement behind its morphing forms also changes to the movement of the software agent, which has been learnt from Steph's improvisations. Thus, people in passing cars and pedestrians across the street are able to witness the continued dance of the morphing eye.

4. TEACHING THE AGENT

4.1 Artificial Neural Network

For Recognition we have chosen a particular type of Artificial Neural Network known as a Self Organising Map (SOM). [6] The SOM is an effective means of data mining as it is able to represent high dimensional data in fewer dimensions, allowing the data to be more easily visualized. It is also useful for clustering like segments of data into regions that help elucidate patterns inherent in the data. For Recognition, we are not so interested in the clustering capabilities of the SOM as much as its ability to adjust its internal weights to closely match those of the input data. By doing so it can create a map containing movement postures that can be traversed to generate movement sequences for the agent's avatar and in turn, for the giant eye, in the absence of humans. This feature of SOM is sometimes referred to as Associative Memory.

We have used a modified form of SOM that has multiple layers, the first layer contains information describing postures of the body, 79 weights equating to the position of the agent and its joint rotation angles. The second layer contains temporal information, potential pathways through the first layer to link the postures in order to produce movement. The SOM is an unsupervised form of

ANN, the recorded movement data is presented to it to learn without any labelling or suggested outcome. Poses that are near identical will be encapsulated within the same neuron and neurons with similar poses tend to cluster together in the first layer. A single neuron may encapsulate a number of similar poses from the input data while other neurons may not have any. It is a competitive process accomplished during the learning phase.

In order to provide the training data for the agent we used a motion capture system to record the movement of a dancer while she was interacting live with the projected eye avatar. (Figure 3) The optical motion capture system used provided higher resolution data than the Kinect and resulted in movement that looked closer to the dancer’s original movement. Recording the data as the dancer improvised with the installation resulted in movement that was similar to what a human would perform in the live installation, giving the agent an appropriate movement vocabulary to work with.

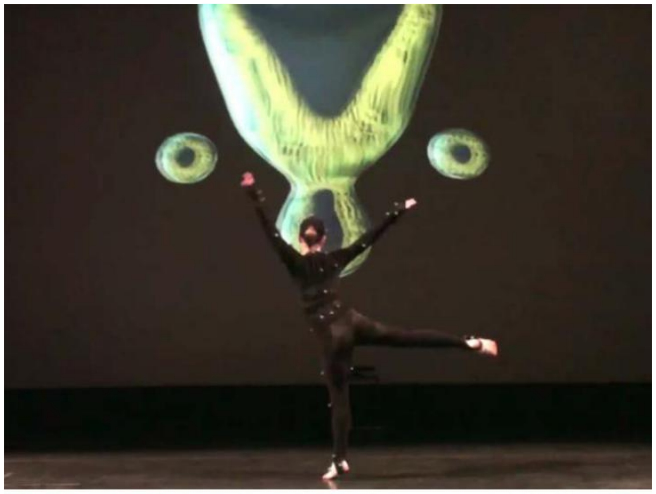

Figure 3. Dancer Steph Hutchison wearing a motion capture suit to capture movement with which to train the agent. Image © John McCormick 2013

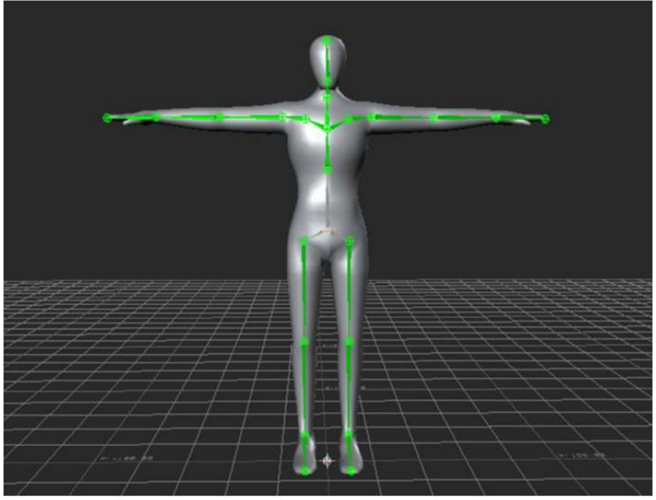

Figure 4. The skeleton used throughout the installation, from recording the movement data from the dancer to animating the agent and human avatars. The common skeleton allows the avatars and their data to be interchangeable within the installation. Image © John McCormick 2013

We used the same skeleton to record the data for the SOM to learn with as well as to animate the avatars. (Figure 4) The skeleton is relatively simple in order to keep the data size as small and optimal as possible so as to reduce the training time and allow the data to perform well in a live performance setting. The same skeleton is used for the avatars representing the live humans and the performing agent allowing their data to be interchangeable. The 19 joints of the skeleton produce 79 input vectors per frame of recorded movement. All rotations are local rotations and the positions are relative changes in position from the previous frame. The local rotations and relative positions allow the agent to generate new movements that follow on from its last position and posture so they are not dependent on where they are in the virtual world. Furthermore, as the agent’s neural network can also be used for recognizing a human’s movement [7] it is able to do so independent of the position of the human’s avatar in the virtual world. The initial tests were done in Matlab and the final installation was developed using the Unity game engine. The trained SOM was imported for the agent to use as its movement memory from which it could generate appropriate sequences to animate its avatar.

The spatial layer of the SOM is visible in (Figure 5). It shows the results of a learning phase with neurons containing weights that describe the postures they have captured. Some neurons have many hits of similar postures, where a movement may have been held or repeated and so re-occurred in the input movement data. When a neuron is stimulated, its weights can be used to animate its avatar. By moving from neuron to neuron, the job of the temporal layer, the avatar is continuously animated by the SOM.

4.2 Clothing the Avatar

Besides providing human movement data to the agent in order to learn how to move, we also used images of Steph and John’s

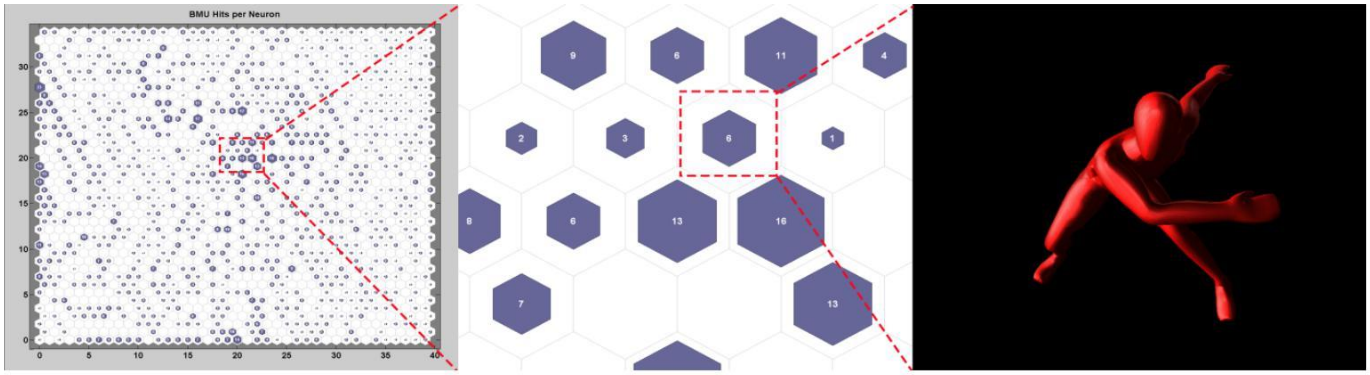

Figure 5. SOM after training showing accumulated postures. The cells represent neurons and the numbers represent how many frames from the input movement data the neuron was able to match. (Best Matching Unit) The image at right shows what the weights of the neuron might look like when translated onto the avatar’s skeleton. Image © John McCormick 2013

irises as the textures for the main avatar. (Figure 6) This provided further connection to the human "donors" and a quick means of giving the main avatar a unique identity. When a human participant was present the main avatar used John's iris to clothe its body. When no humans were present and it was drawing on the agent's movement, it used Steph's iris. (Figure 7)

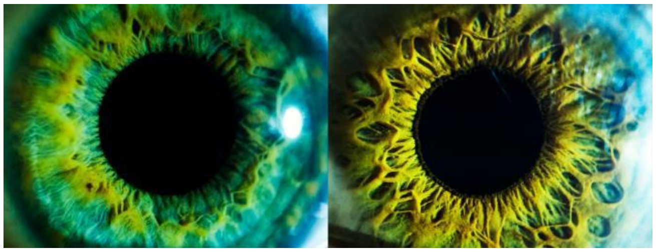

Figure 6. (Left) Steph's iris used by the main avatar when no humans are present and the agent's movement is in use, (Right) John's iris used when the human avatar movement is active. Image © John McCormick 2013

The main avatar's body was produced using a marching cubes algorithm and tables courtesy of Paul Bourke. [8] This algorithm allowed the main avatar to constantly reform according to the movement data from the agent or live human.

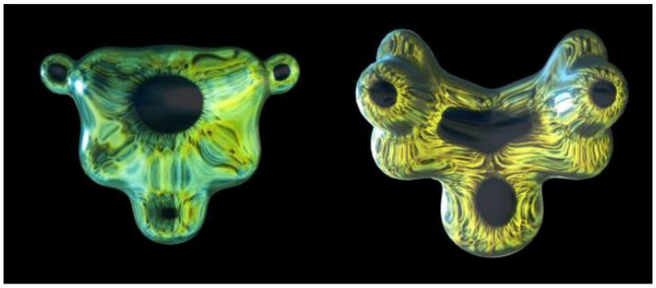

Figure 7. (Left) The main avatar using Steph's iris and the agent's movement, (Right) using John's iris and live human (Kinect) movement. Image © John McCormick 2013

4.3 Sound

Extending upon the manner in which the installation drew upon existing human data for its movement and visual sources, so too the sound for the installation drew upon the environment it was occurring in. The traffic passing by the front of the building created a peculiar undertone of noise with strong Doppler characteristics. At dusk huge flocks of birds flew into the area and their calls were almost overwhelming. The sound emitted from the installation had similar qualities with a deep undercurrent of dopplered noise and short higher pitched overtones reflecting the calls of the birds. The sounds palette changed depending on whether the human or agent's movement data was being used and moved spatially according to the movement of the visible avatar.

5. CONCLUSION

One of the goals in developing the performing agent was to borrow what we could from a human in order to rapidly develop the agent's capability. Using movement and biological data from the dancer allowed the agent to develop quickly. Allowing the agent to learn from the dancer's movement which was captured while interacting with the installation gave the agent an appropriate vocabulary to use when there were no humans present to interact. Having the agent and human avatars co-existing and ready to provide movement data when necessary proved a reliable and seamless solution to the problem of the main avatar having no data to animate itself with if there were no humans present. The movement generated by the agent was visually similar to what a human would have produced.

The SOM can also be used for movement recognition, the neuron containing the closest match to the human's current live movement will fire, and we can act upon this recognition with appropriate events. Going beyond Recognition we are currently investigating the development of artworks utilizing both the movement synthesis and recognition capabilities of the ANN to allow the agent to actively engage with human participants.

6. ACKNOWLEDGMENTS

The artists and authors would like to thank the following people for their much appreciated assistance:

Peter Divers and Daniel Skovli from the Deakin Motion.lab, Chee Peng Lim and Matthew Watson from the Centre for Intelligent System Research and Paul Bourke. Thanks to Cube 37 for their support in presenting Recognition.

Additional information and video is available at:

http://www.johnmccormick.info/?s=recognition

7. REFERENCES

[1] Hsu, E.; Puli, K.; Popovic, J., *Style translation for human motion*. ACM Transactions on Graphics (TOG) - Proceedings of ACM SIGGRAPH 2005 2005**,** *24* (3), 1082 - 1089.

[2] Taylor, G.; Hinton, G. *Factored Conditional Restricted Boltzmann Machines for Modeling Motion Style*, Proceedings of the 26th International Conference on Machine Learning,, Montreal, Canada, 2009; Montreal, Canada, 2009.

[3] Han, S.-J.; Cho, S.-B., *Predicting User's Movement with a Combination of Self-Organizing Map and Markov Model*. Artificial Neural Networks – ICANN 2006 Lecture Notes in Computer Science 2006**,** *4132*, 884 - 893.

[4] Son, K. H.; Ji, C. W.; Park, Y. M.; Cui, Y.; Wang, H. Z.; Chon, T. S.; Cha, E. Y., *Recurrent Self-Organizing Map implemented to detection of temporal line-movement patterns of Lumbriculus variegatus (Oligochaeta: Lumbriculidae) in response to the treatments of heavy metal*. WIT Transaction on Biomedicine and Health 2006**,** *Vol. 10*, pp. 77-91.

[5] Caridakis, G.; Karpouzis, K.; Drosopoulos, A.; Kollias, S., SOMM: *Self organizing Markov map for gesture recognition*. Pattern Recognition Letters 2010**,** *31* (1), 52-59.

[6] Kohonen ,T. 1997 *Self-organizing maps*, 2nd ed, Springer, Berlin, New York.

[7] McCormick, J. Vincs, K. Nahavandi, S. Creighton,D. *Learning to Dance with a Human*, in Cleland, K., Fisher, L. & Harley, R. (2013) Proceedings of the 19th International Symposium of Electronic Art, ISEA2013, Sydney.

[8] Bourke, P. 1994 *Polygonising a Scalar Field*, http://paulbourke.net/geometry/polygonise/ (accessed December 13 2013)

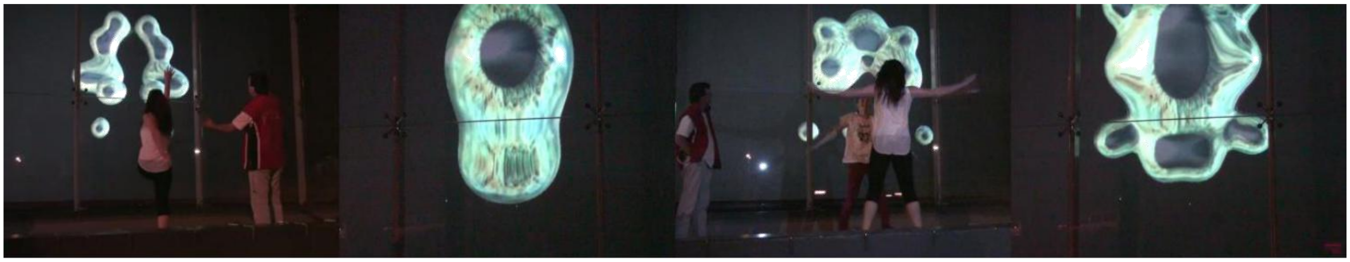

Figure 8. The visible avatar alternating between using the human movement to animate itself and the agent's movement when the humans leave the area in front of the gallery. Image © John McCormick 2013