EmotoPet: Exploring Emotion with an Environment-sensing Virtual Pet

Qifan Wu[1], Jason Orlosky[2], Kiyoshi Kiyokawa[3], Haruo Takemura[4]
Cybermedia Center, Osaka University

ABSTRACT

Artificial pets have been used in a number of applications ranging from games to elderly companionship to personal training. However, these pets have always been limited in their capacity to interact with the environment, respond to the user, generate perceptible emotions, and facilitate meaningful interaction.

As a step towards these goals, we present an augmented reality pet that can navigate its immediate 3D environment, respond to a user's voice and physical presence, and interact using a unique emotional model for a more engaging experience. To accomplish this, we have developed a framework that captures a combination of reconstructed environmental information, user gestures, and voice commands. These inputs are then fed into a behavior decision algorithm and emotion model that generates a number of unique emotional states, such as happiness or fear, which affect the actions of the virtual pet in response to the user. Results of two short experiments testing human-pet interaction show that the emotional model facilitates slightly better perception of the pet's emotions and that spatial interaction only demonstrated certain benefits over a stationary monitor.

CR Categories and Subject Descriptors: H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—*Artificial, augmented, and virtual realities*;

Additional Keywords: Augmented reality, virtual pet, spatial interaction, emotion model

1 INTRODUCTION

The potential of a pet to elicit emotional responses in humans is quite amazing. Pets are believed to have been kept as companions since the Stone Age and in present day are even used to help treat physiological medical conditions such as post-traumatic stress disorder [1]. Similarly, it has been shown that virtual pets can provide similar companionship and treatment to humans to some degree. Compared with a real pet, a virtual pet can also have many advantages with regards to life span or care requirements. For example, virtual pets can be easier to look after for disabled individuals or infants, people with allergies don't have to worry about hair or dander, and the risk of property destruction or death are not present. Virtual pets also have tremendous potential for use as a means of treating lifestyle-related problems such as obesity [6]. They also have the potential to be effective for treating depression or mental disorders such as autism [12].

However, up to now virtual pets have been limited in their ability to interact with humans in a meaningful way, and still lack the ability to show emotions and respond to a shared environment. These qualities are essential for generating the same emotional responses we get from real animals, and implementing these characteristics can be very challenging in a dynamic 3D world.

In this paper, as a step towards the realization of these qualities, we have developed an interactive virtual pet that both exhibits emotions and can interact with its environment and owner in real time. In contrast with existing mixed reality agent models used for gaming and health based training [2], [6] our system allows for direct interaction with the pet in the user's immediate environment and generates emotional states that affect the pets interactions and responses to user input.

To do so, the system first computes a 3D reconstruction of both the user and his or her environment. This reconstruction is used for two purposes: to allow the pet to navigate nearby space while accounting for physical objects and to segment out user gestures and movements, as shown in top and bottom left of Fig. 1, respectively.

Next, a decision making algorithm coupled with a multi-level emotion model facilitate both the pet's lone actions and responses to the user. This model is based on two sets of inputs representing the physiological and psychological state of the pet and one set of output states that affects the nature of responses.

Finally, we conduct two experiments to evaluate user experiences with the pet. The first was designed to test participants' perceptions of the emotional model and decision making algorithm as well as to gather feedback on the virtual pet as an interface, and the second compares and contrasts interaction with the pet on the HMD versus a static monitor. Results show that the emotional model allows users to perceive the pet's actions more easily, and that the HMD was only somewhat beneficial.

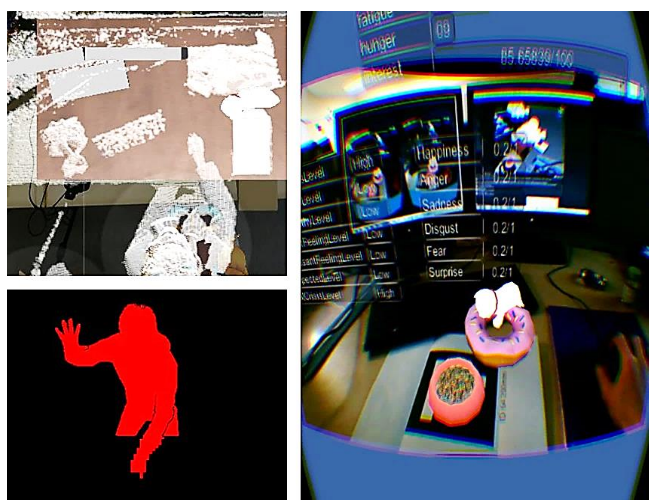

Figure 1: Images of the real-time environment reconstruction, which allows the virtual pet to interact with both the user and nearby physical elements (top left), segmentation of the user's body and hands (lower left), and the right camera view through the immersive display showing an EmotoKitty eating food provided by the user (right), in this case represented by an AR marker.

[1] wuqifan573 [at] gmail.com
[2] orlosky [at] lab.ime.cmc.osaka-u.ac.jp
[3] kiyo [at] ime.cmc.osaka-u.ac.jp
[4] takemura [at] cmc.osaka-u.ac.jp

2 RELATED WORK

Although the history of a virtual pet concept dates back to the era of 8-bit personal computer, commercialized products only started to appear in the early 1990's. These included titles such as "Petz", released by PF Magic, Inc. in 1995 and the key chain game "Tamagotchi" from Bandai Co., Ltd. These games led to a series of other game based pets like "Pokemon" and "Nintendogs" from Nintendo Co., Ltd. However, interactions were limited to isolated, non-immersive environments such as console, pocket, and PC game systems. Since then, quite a bit of work has been done to help realize a more convincing and personal experience for the user. Related research primarily falls into two categories, including: 1) creation and design of virtual pets and agents and 2) frameworks for achieving more immersive and natural interaction with game agents and the environment.

2.1 Virtual Pets and Agents

Research on virtual pets has existed for quite a long time. A classic example is the "ALIVE" system proposed by Maes et al. [10]. Many researchers have subsequently made contributions in the field focusing on different aspects of pets. For example, "Herding Sheep," proposed by MacWilliams et al., focused on distributed computing [11]. "MonkeyBridge," proposed by Barakonyi et al., demonstrated path finding [2] for realism, and Lee et al. discusses remote interaction [9]. Ushida et al. [14] and Kitamura et al. [7] actually have proposed approaches using an emotion models to simulate emotion and behavior, but these works were still limited to a 2D screen for interaction and viewing.

With the recent progress of display and video technology, it has finally become possible to bring virtual pets into the real world using AR [15]. A virtual entity of this is often classified as a MiRA (Mixed Reality Agent), as defined by Holz et al. [4]. A good example of this AR approach is the PlayStation game "EyePet," released by Sony Computer Entertainment Inc. in 2009.

Though these games and virtual agents represent good steps toward a convincing virtual pet, they still do not adequately address the complex set of interactions between the pet, user, and the surrounding environment. In contrast, our approach takes information from the real environment using an RGB-D camera and microphone-array, which allows for higher autonomy, reactivity, and pro-activity of the pet. Additionally, the wide field of view (FOV) see-through HMD is used to improve sense of presence and facilitate spatial interaction.

2.2 Interactive AR Systems

Approaches for AR and MR have also had a significant effect on the perception of virtual entities and user engagement. To improve the presence and interaction for an object such as a pet, researchers like Unuma et al. have proposed see-through AR systems for 3D interaction [13]. In this work, the user's hands were tracked by a depth camera mounted at the back of the tablet, by which user could interact with the AR object. However, presence is limited and the system requires the user to hold the device using his or her hands. Hilliges et al. proposed a stationary AR display system called "HoloDesk," where AR contents could be shown in the interaction space in the reflection of a LCD display off of beam-splitters [5]. In the system, users could interact in a hands-free manner in natural positions, but the display is relatively bulky and the range in which interaction is possible is relatively limited. Benko et al. proposed a projection based AR display system called "MirageTable," where the 3D AR contents could be shown on a curved screen, projected by a stereo projector [3]. Users needed to wear 3D glasses, and even then, display quality was heavily influenced by the projection plane.

2.3 Further Motivation

In existing research, autonomy and reactivity of virtual pets are relatively limited or programmatic. Just like a real pet uses vision, hearing and other sensations to interact, the same multi-modal interactions are expected from virtual entities. Regarding display medium, a stationary display or a hand-held device is typically used in current approaches. However, with these displays, users may not feel that they are living with the pet in the same physical space, which greatly reduces sense of presence. These traditional display approaches also limit natural interaction, since sense of space and perception of the pet's actions may be distorted.

To help address these issues, we propose an AR virtual pet that incorporates information from the real world, can make autonomous decisions about behavior, and can engage in natural interactions with a user in a real time immersive 3D environment.

3 HARDWARE, FRAMEWORK, AND BEHAVIOR MODEL

The hardware portion of our setup consists of several main components, including a Microsoft Kinect, microphone array, Oculus Rift DK2, Ovrvision stereo camera rig, and a set of AR markers. The hardware data flow, also shown in the leftmost block of Figure 2, starts with the Kinect as input. This facilitates reconstruction of the user's immediate environment, which is later integrated into Unity. The Kinect is also used to segment the user's body and detect gestures and body movements that will influence the pets in the virtual space. These gestures are also combined with speech recognition so that verbal commands will also affect the pet's actions and emotional state. The last form of input is the use of AR markers representing food or toys, which are detected with the Ovrvision stereo camera rig to prevent interference with 3D reconstruction performance.

3.1 Environment reconstruction

For the pet to navigate in the physical world, the environment 3D reconstruction coordinate system must be matched with the user's perspective. To accomplish this, data is first mapped from the Kinect output to the Unity Pro Engine, and the virtual coordinate system is matched with the coordinate system of a desktop space for experimentation.

This data is then segmented into two parts, the environment and the user, which are then used for different functions as outlined in the input and processing blocks in Figure 2. The environment map is first bound to Unity's particle system, which allows for collision detection (since we don't want the pet passing through physical objects), occlusion culling, and pathfinding algorithms for navigating the space. Changes to the physical space are also accounted for by updating the reconstructed environment when a new entity enters the scene.

3.2 User Input

The user's presence is filtered out using the Kinect's body index data. This body index provides us with gestures such as petting or caressing the pet, and also allows for sudden movements that may startle or surprise the pet, resulting in a change in emotional state or escaping action.

When users interact with the virtual pet, information such as luminosity distribution, direction and volume of sound, and obstacles are calculated from raw input data using the Kinect and microphone. This raw input is then sent to the behavior decision module, and a reflex processing unit first determines immediate behavior such as running away when startled. If no reflexive behavior is triggered, raw data is then passed to the emotion model, where other behaviors are simulated based on input

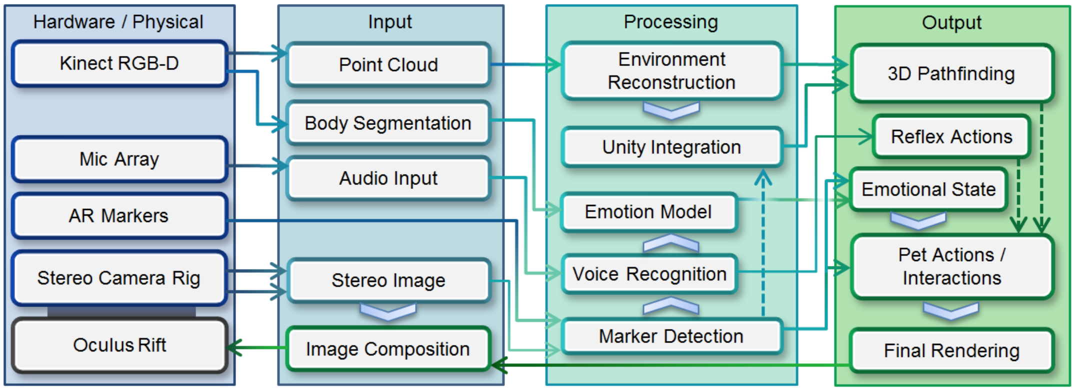

Figure 2: Flow of data within the system, including hardware, input, processing, and output stages.

parameters and current emotional state. Finally, integrated functions from the Unity Pro Engine such as pathfinding and rendering were utilized to output the mixed reality scene to the HMD on a frame by frame basis. It should be noted that no explicit gestures are defined so that the user is free to interact in any way he or she pleases. One reason for doing this is to focus on the pet's autonomy so that user interaction will be more natural.

AR markers are also available to dispense food to the pet. These markers are tracked through Ovrvision rather than the Kinect's camera in order to maximize the efficiency of the reconstruction API and for more accurate calibration.

3.3 Decision Making Algorithm and Emotion Model

As previously mentioned, the pet's behavior is dependent on two sets of input parameters, including the following:

- Physiology: *hunger*, *fatigue*, and *interest*, and
- Psychology: *goal success, goal failure, guilt, pleasure, unpleasantness, unexpectedness*, and *crisis*.

These parameters continuously influence six emotion states:

- *happy, angry, sad, disgusted, afraid,* and *surprised.*

The current emotion affects the nature of responses to the user. Parameters such as hunger or pleasure are affected by the introduction of food or caressing of the pet, and parameters like unexpectedness or interest are affected by events such as sudden user movements or sounds. Goal success and failure are affected by the completion of voice commands. These inputs are all fed into a rule set that determines changes of the emotional states of the pet, similar to that of Ushida et al. [14]. The changes in emotion intensity E are described by

$$E_i(t) = \frac{1}{1 + e^{-\left(X(t-1)+\delta-\gamma+\sum_{i=0}^{j} W_{ji}E_j(t-1)\right)}},$$

where E represents a particular emotion at time t. δ is the input from the rule set described above, γ is the decay coefficient of an emotion over time, and W_{ji} is excitatory or inhibitory gain from the previous emotion state to the next. When the intensity of an emotion state crosses the threshold and is higher than the values of other emotion states, it is promoted to the active emotion.

Reflex processing is also included outside of the emotional model function to deal with strong, spontaneous stimuli such as a sudden gesture from the user, and a memory function allows the pet to remember a prior emotion states associated with specific events, like its name being called when it was happy.

4 Experiments

For experimentation, we implemented *EmotoKitty*, an EmotoPet with ten unique behavior patterns triggered both by user input and the emotion/behavior model above. Behaviors and their triggers were designed to mimic typical behavior of a pet cat as follows:

- Run from loud noises or from gestures that are too sudden.
- Beg for food if very hungry and if no food is present.
- Fall asleep if fatigued.
- Move towards food if hunger is moderate or greater.
- Move to a bright location for rest when fatigue is moderate.
- Self-clean when being caressed if no emotion is present.
- Jump playfully when caressed if happy is the current emotion.
- Wander the environment if no actions are pending.
- Respond to the name called by the user if happiness is present.
- Approach if called again.

Figure 3: A captured (top left) and segmented (top right) pet stroking action. An avatar representing the user's body is overlaid for reference. Image of the video-see through Oculus Rift DK2 (with detachable stereo cameras, bottom left), and two examples of actions and responses from the virtual pet from the view of the Rift, including begging and playful jumping (bottom middle and right).

4.1 Setup

To evaluate how the emotion model affected user experiences and to see if an HMD would improve the user's sense of spatial interaction, we set up a short experiment with 8 participants (Average age 24.1). Users were first asked to interact with the system freely for 2x 2 minute intervals, with and without the emotion model. A second analysis was conducted with the emotion model on, but interacting for 2x 4 minute sessions first with the HMD and then a desktop monitor, as shown on the upper left of Figure 3. Participants were instructed to carefully observe both the autonomous actions of the pet and human-pet interactions, and following the interactions participants were presented with a survey. Participants rated both the pet's emotional behavior and their experience with the HMD vs. the monitor on a 7 point Likert scale. Abbreviations representing the nature of each question are displayed on the *x* axes of the graphs in Figure 4.

4.2 Results

Graphs representing the differences with the emotion model on/off and the HMD/monitor conditions are summarized in the top and bottom graphs of Figure 4, respectively. Kruskal-Wallis tests for each question revealed a near statistical significance for perceiving changes in the emotion of the pet using the emotion model ($\chi^2 = 3.378,\ \chi_U^2 \le 3.841, \mathrm{p} = 0.0661 \ge .05$) and for ease of interaction with the HMD when compared to the monitor ($\chi^2 = 2.482,\ \chi_U^2 \le 3.841, \mathrm{p} = 0.11 \ge .05$. Several participants also commented that they found the HMD to be more natural in general, though no similar comments were reported for the monitor. Other feedback included that the voice recognition was very useful, that the system would be more usable if the HMD were more comfortable, and that size and perceived distance of the EmotoKitty were somewhat unnatural.

5 Discussion and Conclusion

Combining a 3D interface with an emotion model proved to be beneficial in several ways, and we learned quite a bit from user comments. For example, many comments were about both the perceived and physical distance of the cat to either their hand or body. This suggests that spatial presence is very important for the pet to appear as if it lives and interacts in the same physical environment as the user. More natural interactions such hand gestures or playing with elements in a real environment were also possible since the virtual pet navigates the physical world and can respond accordingly. Based on experimental results and comments, participants were able to perceive some of the benefits of the 3D interface, though further experiments with gesture and voice interaction would be beneficial. As future work, we plan to use object recognition so that the pet can both recognize and understand individual elements in the environment.

In conclusion, we presented an AR virtual pet that can interact with both its surroundings and a human present in the same environment by sensing light, sound, obstacles, and motion. The pet then decides on appropriate behavior in real time using an emotion model. Experiments showed that the virtual pet has potential for eliciting a higher level of emotion perception with the model, though spatial perception may even be more important.

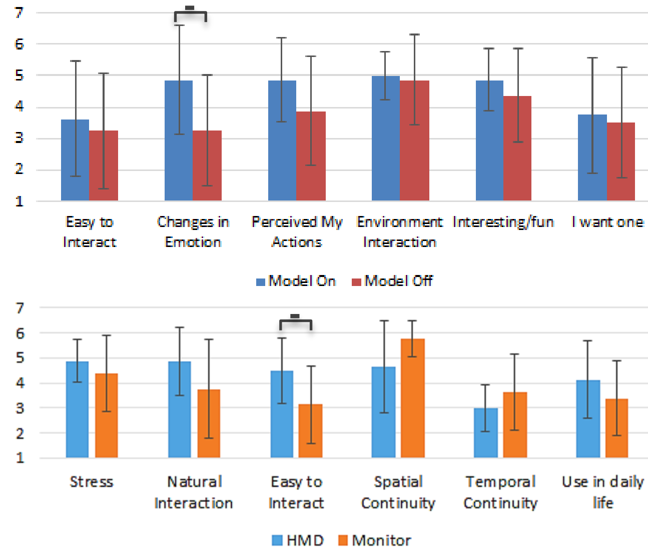

Figure 4: Graphs showing differences in emotional factors (top), and perceptual differences between HMD and monitor (bottom).

References

[1] Altschuler, E. L. (1999). Pet-facilitated therapy for posttraumatic stress disorder. Annals of Clinical Psychiatry, 11(1), 29-30.

[2] Barakonyi, I., Weilguny, M., Psik, T., & Schmalstieg, D. (2005, June). MonkeyBridge: autonomous agents in augmented reality games. In Proceedings of the 2005 ACM SIGCHI International Conference on Advances in computer entertainment technology (pp. 172-175). ACM.

[3] Benko, H., Jota, R., & Wilson, A. (2012, May). MirageTable: freehand interaction on a projected augmented reality tabletop. In Proceedings of the SIGCHI conference on human factors in computing systems (pp. 199-208). ACM.

[4] Holz, T., Campbell, A. G., O'Hare, G. M., Stafford, J. W., Martin, A., & Dragone, M. (2011). MiRA—mixed reality agents. International journal of human-computer studies, 69(4), 251-268.

[5] Hilliges, O., Kim, D., Izadi, S., Weiss, M., & Wilson, A. (2012, May). HoloDesk: direct 3d interactions with a situated see-through display. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 2421-2430). ACM.

[6] Johnsen, K., Ahn, S. J., Moore, J., Brown, S., Robertson, T. P., Marable, A., & Basu, A. (2014). Mixed reality virtual pets to reduce childhood obesity. Visualization and Computer Graphics, IEEE Transactions on, 20(4), 523-530.

[7] Kitamura, Y., Rong, R., Hirano, Y., Asai, K., & Kishino, F. (2008, October). Video agent: interactive autonomous agents generated from real-world creatures. In Proceedings of the 2008 ACM symposium on Virtual reality software and technology (pp. 30-38).

[8] Kriglstein, S., & Wallner, G. (2005, April). HOMIE: an artificial companion for elderly people. In CHI'05 extended abstracts on Human factors in computing systems (pp. 2094-2098). ACM.

[9] Lee, P., Cheok, D., James, S., Debra, L., Jie, W., Chuang, W., & Farbiz, F. (2006). A mobile pet wearable computer and mixed reality system for human–poultry interaction through the internet. Personal and Ubiquitous Computing, 10(5), 301-317.

[10] Maes, P., Darrell, T., Blumberg, B., & Pentland, A. (1997). The ALIVE system: Wireless, full-body interaction with autonomous agents. Multimedia Systems, 5(2), 105-112.

[11] MacWilliams, A., Sandor, C., Wagner, M., Bauer, M., Klinker, G., & Bruegge, B. (2003, October). Herding sheep: live system for distributed augmented reality. In Mixed and Augmented Reality, 2003. Proceedings. The Second IEEE and ACM International Symposium on (pp. 123-132). IEEE.

[12] Sams, M. J., Fortney, E. V., & Willenbring, S. Occupational therapy incorporating animals for children with autism: A pilot investigation. American Journal of Occupational Therapy, 60(3), 268-274.

[13] Unuma, Y., Niikura, T., & Komuro, T. (2014, February). See-through mobile ar system for natural 3d interaction. In Proceedings of the companion publication of the 19th international conference on Intelligent User Interfaces (pp. 17-20). ACM.

[14] Ushida, H., Hirayama, Y., Nakajima, H. (1998, September). Emotion model for life-like agent and its evaluation. In Proceedings AAAI-98 (pp. 62-70). Aaai Press.

[15] Wu, Q., Orlosky, J., Takemura H., and Kiyokawa, K. (2015, October) Poster: "Multi-modal Interaction with a Context-aware AR Virtual Pet," Proc. of the 25th International Conference on Artificial Reality and Telexistence and the 20th Eurographics Workshop on Virtual Environments.