Patricia Pons, Javier Jaen, Alejandro Catala: *Developing a depth-based tracking system for interactive playful environments with animals,* 2nd International Congress on Animal Computer Interaction – 12th International Conference on Advances in Computer Entertainment Technology (ACE'15), http://dx.doi.org/10.1145/2832932.2837007

Developing a depth-based tracking system for interactive playful environments with animals

Patricia Pons
ISSI Group, Dpto. Sistemas Informáticos y Computación
Universitat Politècnica de Valencia (Spain)
ppons@dsic.upv.es

Javier Jaen
ISSI Group, Dpto. Sistemas Informáticos y Computación
Universitat Politècnica de Valencia (Spain)
fjaen@upv.es

Alejandro Catala
LoUISE Group, Computing Systems Department
University of Castilla-La Mancha (Spain)
alejandro.catala@uclm.es

ABSTRACT
Digital games for animals within Animal Computer Interaction are usually single-device oriented, however richer interactions could be delivered by considering multimodal environments and expanding the number of technological elements involved. In these playful ecosystems, animals could be either alone or accompanied by human beings, but in both cases the system should react properly to the interactions of all the players, creating more engaging and natural games. Technologically-mediated playful scenarios for animals will therefore require contextual information about the game participants, such as their location or body posture, in order to suitably adapt the system reactions. This paper presents a depth-based tracking system for cats capable of detecting their location, body posture and field of view. The proposed system could also be extended to locate and detect human gestures and track small robots, becoming a promising component in the creation of intelligent interspecies playful environments.

Author Keywords
Animal Computer Interaction; tracking system; depth-based tracking; play; interaction design; smart environment.

ACM Classification Keywords
H.5.2. Information Interfaces and Presentation: User interfaces – interaction styles, user-centered design, input devices and strategies. I.4.8. Image Processing and Computer Vision: Scene Analysis – tracking, depth cues.

INTRODUCTION
The development of suitable engaging games for animals is a promising research line within the field of Animal Computer Interaction [13,14]. However, until now, interactive digital games for animals have been tied to a specific device, human participation has usually been limited to a "controller" or "assistant" role and there has been no support for several animal participants playing together as differentiated players [19].

Traditional games with animals rely on a more natural and open interaction, meaning that both animals and humans can move freely during the game. Animals are used to playing by themselves or with humans, and in the latter case the human is an active and essential participant in the activity. In addition, traditional games make use of the elements in the environment to enhance the playful experience, not limiting it to the object itself but to the spontaneous interactions between the players thanks to the mediating object. Moreover, it is not unusual that when a human starts playing with one of his pets other animals in the room join in, forming a multiplayer activity in which the roles of the participants evolve during the game.

Future technologically-mediated games for animals could therefore be conceived as multimodal and multi-device systems, in which animals could play either alone, in a group or with human beings in a natural way. If animals play by themselves, the system should intelligently manage the different devices and objects in the environment in order to adapt the game to the animals' preferences and interactions [18,19]. As an example, we could think of a game in which a cat chases a Sphero® (electronic ball) controlled by the system, but the movements of the ball are not random. Instead, the Sphero® could be programmed to move away from the cat inside its field of view, or suddenly turn left or right if the cat gets too near to it, or even move towards the cat if the tracking system detects that he is crouched in a hunting position, waiting for the ball to approach. If human beings are also participating, their interactions should also be considered an essential part of the game and the digital playful experience should be as natural as traditional games.

The use of non-wearable tracking mechanisms would allow more natural interactions within technologically-mediated environments, not limiting the interaction to a specific device or interaction modality, but instead letting the participants explore their environment and create the dynamics of the game on the go. For this purpose, the tracking system should provide the system with information about the animals' location, body posture and field of view, as well as the location and gestures of the human participants, if applicable. This information would allow the development of playful experiences which improve the animals' wellbeing by introducing new forms of mental and physical stimulation. This paper will describe the development of a prototype for a depth-based tracking system for cats and discuss its future extensions to human gesture recognition and object tracking. Several application scenarios will be outlined, and future work and possible research lines will be given.

RELATED WORKS

Outdoor animal tracking systems are either based on GPS or radio-frequency collars only [12,16,22], or they could also use wearable inertial measurement units (IMUs), such as accelerometers and gyroscopes located on a collar [10] or attached to a harness [3,4]. In the first case, the system only gives information on the animals' location but not about its postures or orientation. In the latter case, classification algorithms are used on the sensors' data to identify the behavior and/or body postures of the animal.

Indoor animal tracking systems usually rely on wearable devices to gather information on the animals' movements. Canine Amusement and Training [23] uses a wearable harness with attached IR emitters to detect the position and posture of the animal. Cat@Log [24] uses a cat collar device with several sensors: a camera, a GPS, an accelerometer, a Bluetooth module, battery and micro SD card. The camera provides videos of the cat's view, while the accelerometer data is used for activity recognition such as sleeping, jumping, walking or scratching. However, it would be difficult to use this activity recognizer to detect specific interactions during a game. Poultry.Internet [11] tracks the movements of a chicken using a camera and an electro-pad located in the chicken's leg to sense its muscle activity. Through camera images they also detect the chicken's head to find the orientation of the animal within the backyard, but no postures are identified. No wearable device is used in Purrfect Crime [20], an interspecies digital game for cats and humans in which a Microsoft Kinect®[1] is used to detect the position of the cats inside the play area. However, the system only locates the central position of the animal, introducing some interactions which were not really intended by the cat.

Within Human Computer Interaction, the arrival of the Microsoft Kinect® sensor introduced new and natural interaction modalities, simplifying indoor human tracking and gesture recognition by using depth information of the scene. This sensor is usually placed in front of the humans interacting. However, multimodal playful scenarios for human players frequently place the depth-sensor at a higher position [2,8,9], providing wider playing areas and avoiding occlusion due to elements in the room. Technological playful environments for animals would require multimodal interactions inside spacious scenarios. It would be preferable to avoid wearable trackers as they could limit the agility of some animals. Therefore, a depth-based tracking system would be a promising way of detecting the animals' location, posture and field of view, taking the development of intelligent playful environments for animals a step forward.

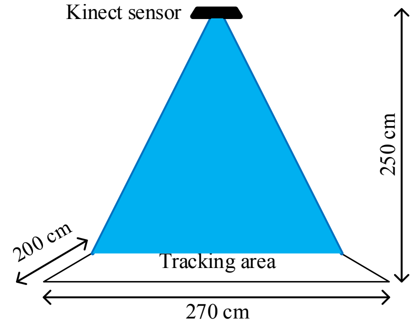

Figure 1. Set up for the tracking system

TRACKING SYSTEM

In order to analyze the suitability of a non-wearable tracking system for animals based on depth information, several sessions with cats were recorded using a Microsoft Kinect® sensor, in which cats were playing with their owners or caretakers, or with small robots. These sessions allowed us to obtain real data on playing postures, behaviors and movements, which were later analyzed and post-processed to develop a preliminary version of the tracking system.

Set-up

The set-up for the tracking system consists of a Microsoft Kinect® sensor looking down from the ceiling, at a height of 250 cm. At this height the sensor covers an area of approximately 200 cm long and 270 cm wide, as shown in Figure 1. The Microsoft Kinect® is capable of recording both color and depth video streams at a rate of 30 frames per second with 640x480 pixel resolution.

Cat tracking

The Microsoft Kinect® sensor provides both color (see Fig. 2a) and depth streams (see Fig. 2b). The first processing step of the algorithm consists of extracting the cat's pixels from the depth frame. Each depth frame provides, for each pixel, the distance in millimeters from the camera plane to the nearest object in that particular pixel (see Fig. 2b). The height of the sensor may change between sessions, however as it remains the same within a single session the pixels corresponding to the ground can be discarded by adjusting the system during the set-up process (see Fig. 2c). However, background segmentation methods, such as plane fitting algorithms, should be used in the future in order to provide more flexible set-up conditions. With the floor removed from the image, computer vision algorithms are applied to extract the cats' contours (see Fig. 2d). In this step, a cat's location within the tracked area can be determined. To detect the cat's posture and orientation, each detected contour is processed

[1] https://msdn.microsoft.com/en-us/library/hh855355.aspx

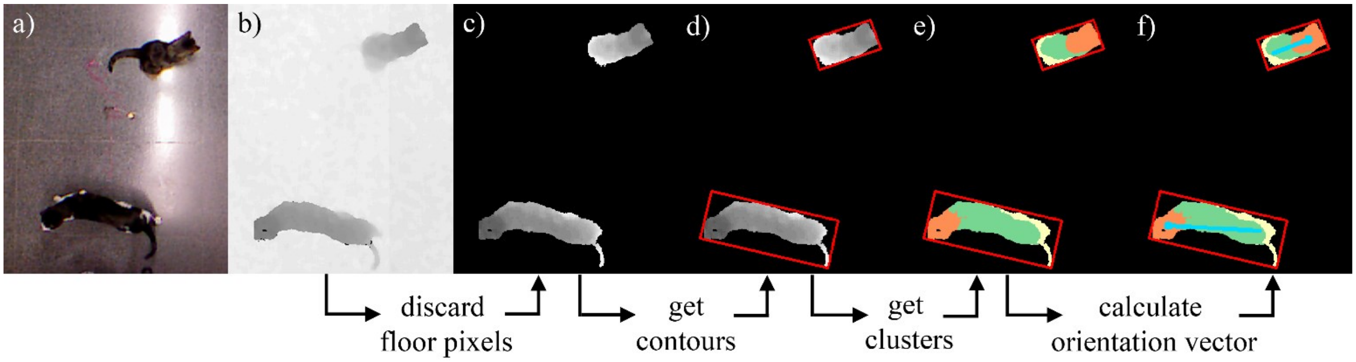

Figure 2. Process of extracting the cat's orientation: (a) color frame (b) depth frame, (c) background segmentation, (d) cat contours, (e) clusters for head, body and tail, (d) orientation vector

by a k-means clustering algorithm, which groups the pixels by their depth value and relative position (see Fig. 2e): pixels of similar depth which are located together in the image would be grouped together within the same cluster. The number of clusters was set to three as an initial trade-off between efficiency and accuracy. Further experiments should be conducted to determine the number of clusters which maximizes the trackers' performance. The clusters obtained from each cat's contour are then classified by a customized decision tree algorithm as head, body and tail. Finally, an orientation vector can be defined from the center of the body/tail cluster to the center of the head cluster (see Fig. 2f), roughly estimating the cat's field of view.

The identification of each cluster depends on the posture being analyzed. Different cat postures were seen to generate different cat contours in the processed depth frames. The cat's depth stream contour when sitting showed a smaller, square-shaped bounding rectangle (see Fig. 2d, cat at the top of the image), while the depth stream when standing or walking revealed a larger and rectangular-shaped bounding rectangle (see Fig. 2d, cat at the bottom of the image). In addition, the pixels of the sitting cat's head had a significantly higher average depth than the pixels of the rest of the body. This is observed in Fig. 2d, in which the grey

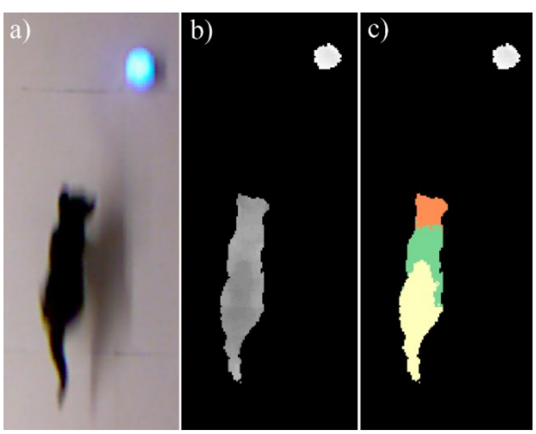

Figure 3. Cat running after a Sphero® robot

pixels of the head are significantly darker than the pixels of the cat's haunch. In contrast, the cat walking in Fig. 3 has an average depth of its head pixels very similar to the average depth of the haunch and tail pixels. This can be observed in Fig. 3b, in which the grey pixels representing depth values are very similar in the whole contour of this cat. In addition, there are some postures, such as the sitting posture in Fig. 2, in which the tail is not detected because of its proximity to the ground. In this case, the head is clearly differentiated from the rest of the body, and in terms of average depth it is undoubtedly the highest cluster detected. Hence, in this posture only the detection of the head and the body clusters would allow the cat's field of view to be determined. However, in other postures the detection of the tail is crucial to determine the orientation of the cat. For instance, determining whether the cat is looking north or south in Fig. 3 would be difficult if the tail was not detected, because the head is not clearly in a higher position than the rest of the body. Therefore, our decision tree algorithm considers the following parameters to identify body postures and classify the clusters: area of the cat's contour, number of pixels for each cluster and average depth for each cluster. This preliminary version of the tracking system is already capable of detecting sitting, semi-sitting, walking, standing, jumping and turning positions of several cats at a time, and classifying the different pixel regions in each posture to detect the head, body and tail of each cat. The algorithm runs with live data at a rate of four frames per second on an Intel Core i5 660, without using GPU processing. The performance of the tracking system would be significantly improved by using GPU computing power as well as more efficient implementations of the clustering algorithm.

An exploratory study on the accuracy of the tracking system has been conducted. For each posture, 200 random frames from the testing set were extracted from the recordings of 2 of the subjects and processed offline by the tracking algorithm to detect the body parts of the cats and calculate the corresponding orientation vector. Table 1 shows a summary of the results indicating the percentage of frames for each posture in which the algorithm correctly identified the body parts and orientation vector. This initial validation

Posture	Sitting/ semi-sitting	Walking/ Standing	Jumping	Turning
Accuracy	86%	74.5%	82%	84%

Table 1. Accuracy of the tracking system identifying the cats' body postures and orientation vector

shows a promising research line and future experiments will be conducted considering different classification algorithms and machine learning techniques in order to improve the accuracy of the system.

Human tracking

ACI playful experiences should be conceived for both animal players only as well as for animals and humans playing together, enhancing interspecies communication and interactions. The development of tracking systems for humans using depth-based sensors placed on the ceiling would help to create new forms of natural interactions between humans and animals based on gestures and body postures rather than using actions on a specific device such as a screen or tablet. The first step in developing a human tracking system for this purpose would be extracting common human gestures, postures and behaviors during playful activities with animals. For instance, playing with our animals at home is a relaxed and informal activity, and humans usually adopt comfortable postures which allow proximity and confidence with their pets. It is likely that the human sits on the couch or on the ground, near to the pet, or even kneels on the floor, so both the human and the animal share the same play area and can interact directly with each other. In this situation, a top-down depth sensor will provide better contextual information of the human's relative location and interactions with the animal. A preliminary session was recorded to give some hints on the types of human postures and gestures to expect during these playful activities (see Figure 4).

In the same way that several sessions with cats were recorded to assist in the development of the cat tracking algorithm, further sessions should be conducted to record pets and humans playing together, creating a knowledge base to which we could apply computer vision algorithms to detect and learn human gestures during play. There is already an extensive research area in the field of computer vision which studies human tracking and gesture recognition. Approaches like the ones in [1,7] would be a good starting point to recognize common body gestures adopted by humans playing with their pets. These gestures could later be used inside the game to perform actions, or to adapt the game to the interactions and postures of both humans and animals.

Object tracking

With the current rise in the use of personal drones and robots, it is likely that games with animals will also evolve to incorporate such objects into the playful experience. These devices can be controlled by humans using tablets or smartphones, or they can be programmed to perform different movements or actions. In the latter case, the robots are not usually aware of their environment and pre-programming behaviors to be performed during the game could represent a threat to the animals' welfare.

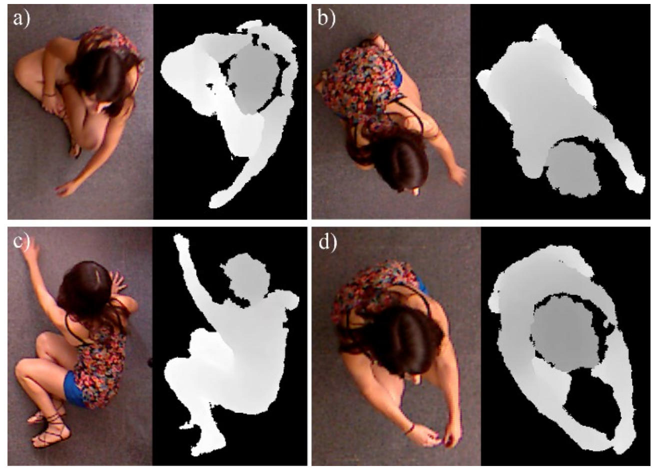

Figure 4. Human postures and gestures during play: a) sitting b) kneeling c) sideways d) crouching down

These robots could also be tracked using the same approach: discarding the floor pixels from the depth stream and extracting the robots' contours. For instance, in Fig. 3, a cat is chasing a Sphero® and this robot appears as a small blob in the image, which is easy to identify and track. Knowing the exact location of the robot and the cat inside the play area would allow us to define safety rules in the system to ensure the robot is not going to approach the animal in a dangerous way or perform risky actions. In addition, tracking the position of the robots together with the cats' location, orientation and body posture would allow to create novel and engaging playful experiences, as described later in this paper.

Challenges

At this development stage of the algorithm we are capable of tracking cats, humans and objects separately. The next step in the development process will be the integration of the three tracking mechanisms into a single system to allow collaborative and interspecies interactions with the intelligent objects in the game.

Currently, the tracking system only relies on single frame information in order to classify the pixel regions and detect the cat's body parts. However, temporal information of past frames would be very useful in the detection and classification of ambiguous contours and postures, as well as in improving the accuracy of the already detected postures. For instance, a probability orientation vector could be defined using the last n-vectors calculated. This vector could be used in cases such as those in Fig. 5c and 5d in which the cat bent down and the tail could not be detected, causing the orientation vector to take a significantly different direction from the one in the previous frame (see Fig. 5a and 5b).

Another challenging issue to be solved is related to the definition of the orientation vector: there are some cases in

Patricia Pons, Javier Jaen, Alejandro Catala: *Developing a depth-based tracking system for interactive playful environments with animals,* 2nd International Congress on Animal Computer Interaction – 12th International Conference on Advances in Computer Entertainment Technology (ACE'15), http://dx.doi.org/10.1145/2832932.2837007

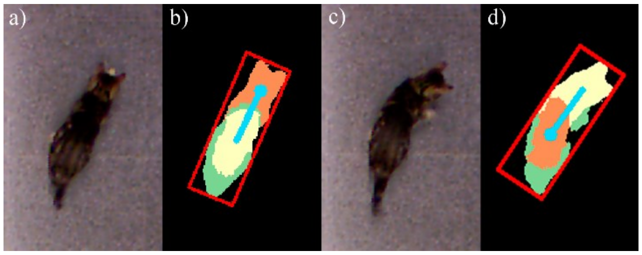

Figure 5. a) and b) Correct classification c) and d) erroneous classification due to ambiguous posture

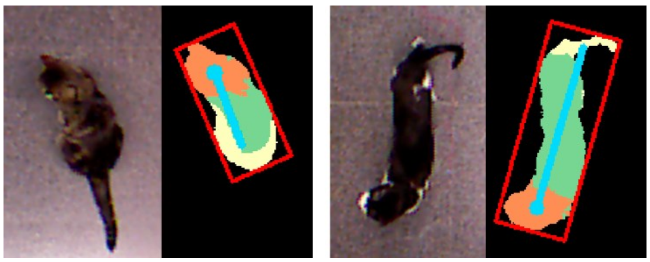

Figure 6. Orientation vectors differ from the heads' current orientation

which the orientation vector does not always point to the area the cat is looking at, as can be seen in Figure 6. Relying on the center of the clusters to define the orientation vector leads to these situations. A solution to determine the real area the cat is looking at could be applying computer vision algorithms to the detected head contour in order to learn and detect possible head shapes. Another solution would be to calculate the major axis of the detected head cluster, which in most cases corresponds to the real orientation of the cat's head.

APPLICATIONS

Intelligent playful environments based on gesture and posture recognition

Tracking systems such as the one presented in this paper have an essential role in the development of future intelligent playful environments for animals. A previous study on cats' interest in different kinds of stimuli [17] showed that in order to keep the cats' interest in the technological artifacts deployed in the study, some kind of "intelligence" is required. Random movements or actions of a digital element would cause the animal to eventually lose interest in the activity. Hence, being able to interpret an animal's body posture and interactions in a similar manner humans do, especially those suggesting either engagement or disengagement as well as distress signals during the game, is very important for providing proper actuations in the playful environment. Therefore, a tracking system capable of detecting a cat's location, posture and orientation would allow us to create engaging and realistic games using technological artifacts which adapt to the detected cat's postures and field of view.

Detecting the location of the cat inside the play area is essential to creating any kind of game which includes technological elements that appear and interact with the animal inside this area, such as robots or digital projections. This would mean, for example, the game could start whenever the cat approached a technologically-mediated toy inside the play area, and end when the cat runs away from the toy. The automatic detection of the cat's orientation would allow enriching and realistic playful activities to be created in which the technological artifacts could use the cat's area of interest, understood as its field of view, as the region on which to deploy the necessary stimuli to attract and maintain its attention. For example, a floor projection which passes in front of the cat's field of view and then moves away from him, encouraging the beginning of a chasing game. Posture detection is also a very promising feature because, together with the cat's field of view, it is an indicator of the cat's intentions towards the game. For instance, if the cat lies down and stops looking at the object involved in the playful activity, it is likely that he is not interested any more in the game. On the other hand, it is possible that the cat slowly approaches the object, e.g. a small robot, to jump over it when it is not moving, and the system could detect the cat's playful behavior by the position of his tail and the smoothness of his movements. In this situation, the system could make the robot move at the very last moment, surprising the cat and engaging him in the activity by imitating human-like playful interactions.

The automatic recognition of human gestures by a top-down depth-based tracking system would allow humans to participate in the game in a natural way for both the animal and the human, not being tied to any specific device. Human gestures could be used to control the movements or features of the digital elements in the play area, e.g. the human player points at an element with his hand and then points to another place, and the element moves in the direction indicated by the human. Some games could introduce specific gesture-based interactions, e.g. a competitive game between the animal and the human in which a clapping gesture makes the system move a toy/cord for 10 seconds to distract the cat so the human can take advantage.

Learning behavioral habits

Pet owners as well as animal caretakers in zoos or shelters easily identify the mental and physical state of the animals they are in charge of by analyzing their posture and movements. They are also capable of anticipating what the animal will do in some situations, as they have learned his behavioral habits and routines. For example, a human knows the time of day each of his pets prefers to play, the behaviors or movements that indicate that the animal is willing to start a playful activity and the meaning of the pets' interactions during the game. In a similar way, zookeepers know the routines of many of the animals they look after: the path that the animal will follow inside its habitat, the routine in which the animal prefers to drink, eat, sleep, play or walk, and even the animals' preferred spot for each interaction. This knowledge is acquired through daily observation and coexistence and it is difficult to transfer to another person who does not share the same context. In addition, it is also difficult to integrate this specific knowledge into a playful digital system, as each animal's preferences and routines would differ: generic knowledge would not cover the singularity of distinct individuals, while defining specific knowledge for each individual is not feasible.

Complete animal tracking systems such as the one presented in this paper coupled with machine learning algorithms would allow us to learn behaviors, habits and even playful dynamics of individual animals. In this way, a personalized knowledge base could be obtained for each animal, similar to the knowledge their human companions have about them. Through the tracking system we could learn, for example, the habitual location and movements an animal performs in a specific context, the amount of time a day it spends doing physical activity such as walking or playing, and the intensity of this activity. This knowledge base could be used to detect changes in behavioral patterns, such as increasing/decreasing physical activity or resting time, therefore supporting early detection of illnesses or other problems. On the other hand, not all animals perform the same interactions in play: during a chasing game, a cat might prefer to wait patiently until an object approaches him to catch it, whereas another cat might be more eager and prefers to run after the object. Therefore, the tracking system could help to learn the specific play dynamics of an animal during a game. An intelligent playful environment could use this information in order to adapt the game to the animal's play preferences.

DESIGNING FUTURE TRACKING-BASED GAMING EXPERIENCES

Tracking systems for open spaces

The tracking system described in this paper is only suitable for indoor detection, due to the sensor restrictions in terms of light conditions, field of view and connectivity. Other methods would have to be considered if we wanted to carry out animal tracking outdoors without using a wearable device, such as dogs in their yards, zoo animals in their ecosystems or wild animals in their habitats.

Currently, there is a wide range of commercial or even home-made flying drones with built-in cameras or even powerful infrared sensors. This type of device could be used not only to observe the animals and collect images from the flights for post-processing [6,21], but to track the animals autonomously and in real-time using the drones without a human controller. Computer vision algorithms could be applied to the extracted frames when the drone is flying over the animal's open ecosystem to detect its shape. Once detected, the drone could be linked to that specific animal and follow it throughout the area, locating its position using the drone's GPS or relative coordinates, and analyzing the extracted images to detect the animals' body posture.

Non-wearable tracking systems in open spaces using drones would make it possible to study the animals' natural interactions in less restricted ecosystems, learn their behavioral habits and also detect abnormal behaviors or possible illnesses. There are already studies within ACI that propose automated behavior recognition and animal tracking in the wild [15]. Autonomous outdoor tracking systems with drones could help these studies in the fulfillment of their goals. In addition, intelligent playful environments such as the ones described in this paper for indoor spaces could also be created outdoors using drone tracking mechanisms. For instance, a dog could be playing inside the living room with his electronic ball, which is controlled by the system and reacts to his interactions. During the game, the dog eventually runs towards the garden and expects the ball to follow him, as a human would do if they were playing together. The system would detect that the dog has gone outdoors and would transfer the playful activity as well as the mediating digital elements to the new play area. In this case, a drone would be activated and the first step of the system would be to send the drone to the garden in order to detect the dog's position. Once it is detected, the drone would keep track of the dog and communicate its movements, position and posture to the system, as the indoor depth-based tracking system would do. In this way, the playful experience would be seamless and more natural for the animal, not being restricted to only a specific area in the house.

Mixed-reality games and embodied interactions

Developing digital playful experiences exclusively for animal players is as important as developing interspecies digital environments in which humans are also active participants in the game. Playful digital environments with pets and humans are not difficult to envision, as the interaction with them is natural and familiar to us and does not entail any risk to our safety. However, playful experiences involving animals in zoos imply that the human and the animal have to be in separate spaces for safety reasons. In addition, as we are not used to interacting with these animal species, our playful interactions might not make sense in the animals' mental perception of the game, and vice versa. How could we create a joint playful experience for users of different species with different perceptions of their interactions and who cannot share the same physical space?

Mixed-reality games could come to the rescue in order to create a common game with separate playful spaces in which the human interactions in its own space are transferred to the animal space in a meaningful way, and vice versa [5]. For example, humans could play using natural body interactions and movements in their play area, which could be recognized by a Microsoft Kinect® tracking system as described above. Human actions on their side of the game could be transferred to the animals' play area, for example to a digital device the animals like to use during their games without human participants. A very promising interaction modality on the human side of the game would be to use virtual or augmented reality devices, such as the Oculus Rift[2] or the Microsoft

[2] https://www.oculus.com/en-us/rift/

Patricia Pons, Javier Jaen, Alejandro Catala: *Developing a depth-based tracking system for interactive playful environments with animals,* 2nd International Congress on Animal Computer Interaction – 12th International Conference on Advances in Computer Entertainment Technology (ACE'15), http://dx.doi.org/10.1145/2832932.2837007

Figure 7. Tiger playing with giant plastic ball (image courtesy of Boomer Ball)

Figure 8. Human using the Virtuix Omni platform (image courtesy of Virtuix)

HoloLens[3], to immerse the human into the animals' ecosystem. One step further would be to introduce more immersive and embodied interactions for humans using omnidirectional treadmill peripherals, such as the Virtuix Omni[4], together with these virtual reality devices. With these embodied interactions, humans could control elements within the animals' ecosystem and also experience feedback from the animals' interactions. For illustration purposes, let's assume that the favorite toy of orangutans in a zoo is a huge plastic ball, as several zoo animals have been observed playing enthusiastically with these elements [25] (see **Figure 7**). The ball could have an electronic inner rotor which allows it to be controlled remotely, like the Sphero®. A human patient in a hospital could be using the Virtuix Omni for rehabilitation purposes, moving inside a digital recreation of the orangutans' ecosystem that he is watching through the Oculus Rift glasses (see **Figure 8**). Similar to what *Metazoa Ludens* [5] proposed with hamsters, the real movements of the human could be transferred to the huge ball within the orangutans' ecosystem, so the orangutans could really chase and interact with the ball as they would usually do. The movements of the orangutans inside their real ecosystem could then be tracked by depth-based sensors or cameras mounted on flying drones, depending on whether it is an indoor or an outdoor scenario. These movements would be transferred to the avatar representation of the orangutans within the digital world the human is observing through the Oculus Rift device. In this context, the goal of the human within the game could vary depending on the welfare needs of the animal playing: if the orangutan is physically active, the goal of the human could be not letting the orangutan catch him (which means that the orangutan could not catch the ball); if the orangutan participating in the game is detected to have lost his appetite, the goal could be to lead the animal towards the food area, i.e., to move the ball towards the food. Regardless of the scenario, animals' mental and physical wellbeing must be a priority. As in indoor domestic playful environments, the system would always ensure the participants' welfare by monitoring their interactions and body postures, introducing safety rules to impede risky behaviors, stopping the activity if the animals show distress and keeping the playing time within healthy boundaries.

CONCLUSIONS

This paper describes the development of a tracking system for cats, humans and objects for indoor playful experiences, based on the analysis of depth information captured with a Microsoft Kinect® sensor placed top-down on the ceiling. The depth-based tracking system is capable of detecting a cat's location, body posture and orientation. It can also track humans and physical objects moving inside the tracking area.

Our immediate future work will be the integration of the three tracking modes into a single tracker. We are also working on improving the accuracy of the cat tracker by including temporal information from previous frames as well as studying its performance using different classification algorithms. After this, we will focus on the detection of human orientation, postures and gestures from a top-down point of view. This would allow us to create an intelligent playful environment in which the digital devices controlled by the system respond to the interactions of both humans and animals, attending to the contextual information of their bodies. We will also explore how this tracking system could be adapted to working with other pets, such as dogs, and zoo animals, such as orangutans.

We have also outlined future research directions in the area of playful environments for outdoor scenarios as well as for animals beyond pets. On one hand, we propose to explore outdoor tracking systems using drones and computer vision algorithms for different purposes, such as creating playful outdoor activities or learning animals' behaviors in open spaces. On the other hand, we believe that mixed-reality games for animals in restricted environments such as zoos could provide benefits for both animals and humans. On the human side, these games could play an educational role to raise awareness of the importance of wildlife preservation, as well as introducing more amusing ways of cognitive or physical rehabilitation. On the animals' side, these playful experiences would improve their wellbeing by introducing new forms of mental and physical stimulation adapted to each animal's context and preferences.

ACKNOWLEDGMENTS

Work supported by the Spanish Ministry of Economy and Competitiveness and funded by the EDRF-FEDER (TIN2014-60077-R). The work of Patricia Pons has been supported by a national grant from the Spanish MECD (FPU13/03831). Alejandro Catalá also received support from a VALi+d fellowship from the GVA (APOSTD/2013/013). Special thanks to our cat participants, their owners, and our feline caretakers and therapists.

REFERENCES

1. Jan Bednarik and David Herman. 2015. Human gesture

[3] https://www.microsoft.com/microsoft-hololens/en-us

[4] http://www.virtuix.com/

recognition using top view depth data obtained from Kinect sensor. *Excel. - Student Conf. Innov. Technol. Sci. IT*, 1–8.

2. Hrvoje Benko, Andrew D. Wilson, Federico Zannier, and Hrvoje Benko. 2014. Dyadic projected spatial augmented reality. *Proc. 27th Annu. ACM Symp. User interface Softw. Technol. - UIST '14*, 645–655.
3. Alper Bozkurt, David L Roberts, Barbara L Sherman, et al. 2014. Toward Cyber-Enhanced Working Dogs for Search and Rescue. *IEEE Intell. Syst.* 29, 6, 32–39.
4. Rita Brugarolas, Robert T. Loftin, Pu Yang, David L. Roberts, Barbara Sherman, and Alper Bozkurt. 2013. Behavior recognition based on machine learning algorithms for a wireless canine machine interface. *2013 IEEE Int. Conf. Body Sens. Networks*, 1–5.
5. Adrian David Cheok, Roger Thomas K C Tan, R. L. Peiris, et al. 2011. Metazoa Ludens: Mixed-Reality Interaction and Play for Small Pets and Humans. *IEEE Trans. Syst. Man, Cybern. - Part A Syst. Humans* 41, 5, 876–891.
6. Amanda Hodgson, Natalie Kelly, and David Peel. 2013. Unmanned aerial vehicles (UAVs) for surveying Marine Fauna: A dugong case study. *PLoS One* 8, 11, 1–15.
7. Gang Hu, Derek Reilly, Mohammed Alnusayri, Ben Swinden, and Qigang Gao. 2014. DT-DT: Top-down Human Activity Analysis for Interactive Surface Applications. *Proc. Ninth ACM Int. Conf. Interact. Tabletops Surfaces - ITS '14*, 167–176.
8. Brett R Jones, Hrvoje Benko, Eyal Ofek, and Andrew D. Wilson. 2013. IllumiRoom: Peripheral Projected Illusions for Interactive Experiences. *Proc. SIGCHI Conf. Hum. Factors Comput. Syst. - CHI '13*, 869–878.
9. Brett Jones, Lior Shapira, Rajinder Sodhi, et al. 2014. RoomAlive: magical experiences enabled by scalable, adaptive projector-camera units. *Proc. 27th Annu. ACM Symp. User Interface Softw. Technol. - UIST '14*, 637–644.
10. Cassim Ladha, Nils Hammerla, Emma Hughes, Patrick Olivier, and Thomas Ploetz. 2013. Dog's Life: Wearable Activity Recognition for Dogs. *Proc. 2013 ACM Int. Jt. Conf. Pervasive Ubiquitous Comput. - UbiComp'13*, 415.
11. Shang Ping Lee, Adrian David Cheok, Teh Keng Soon James, et al. 2006. A mobile pet wearable computer and mixed reality system for human–poultry interaction through the internet. *Pers. Ubiquitous Comput.* 10, 5, 301–317.
12. Clara Mancini, Janet van der Linden, Jon Bryan, and Andrew Stuart. 2012. Exploring interspecies sensemaking: Dog Tracking Semiotics and Multispecies Ethnography. *Proc. 2012 ACM Conf. Ubiquitous Comput. - UbiComp '12*, 143–152.
13. Clara Mancini. 2011. Animal-computer interaction: a manifesto. *Mag. Interact.* 18, 4, 69–73.
14. Clara Mancini. 2013. Animal-computer interaction (ACI): changing perspective on HCI, participation and sustainability. *CHI '13 Ext. Abstr. Hum. Factors Comput. Syst.*, 2227–2236.
15. Steve North, Carol Hall, Amanda Roshier, and Clara Mancini. 2015. HABIT: Horse Automated Behaviour Identification Tool – A Position Paper. *Proc. Br. Hum. Comput. Interact. Conf. - Anim. Comput. Interact. Work.*, 1–4.
16. Mikko Paldanius, Tuula Kärkkäinen, Kaisa Väänänen-Vainio-Mattila, Oskar Juhlin, and Jonna Häkkilä. 2011. Communication technology for human-dog interaction: exploration of dog owners' experiences and expectations. *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, 2641–2650.
17. Patricia Pons, Javier Jaen, and Alejandro Catala. Multimodality and Interest Grabbing: Are Cats Ready for the Game? *Submitt. to Int. J. Human-Computer Stud. Spec. Issue Anim. Comput. Interact. (under Rev).*
18. Patricia Pons, Javier Jaen, and Alejandro Catala. 2014. Animal Ludens: Building Intelligent Playful Environments for Animals. *Proc. 2014 Work. Adv. Comput. Entertain. Conf. - ACE '14 Work.*, 1–6.
19. Patricia Pons, Javier Jaen, and Alejandro Catala. 2015. Envisioning Future Playful Interactive Environments for Animals. In *More Playful User Interfaces*, Anton Nijholt (ed.). Springer, 121–150.
20. Rui Trindade, Micaela Sousa, Cristina Hart, Nádia Vieira, Roberto Rodrigues, and João França. 2015. Purrfect Crime. *Proc. 33rd Annu. ACM Conf. Ext. Abstr. Hum. Factors Comput. Syst. - CHI EA '15*, 93–96.
21. Jessica van Vonderen. 2015. Drones with heat-tracking cameras used to monitor koala population. Retrieved July 1, 2015 from http://www.abc.net.au/news/2015-02-24/drones-to-help-threatened-species-koalas-qut/6256558
22. Alexandra Weilenmann and Oskar Juhlin. 2011. Understanding people and animals: the use of a positioning system in ordinary human-canine interaction. *Proc. 2011 Annu. Conf. Hum. factors Comput. Syst. - CHI '11*, 2631–2640.
23. Chadwick A. Wingrave, J. Rose, Todd Langston, and Joseph J. Jr. LaViola. 2010. Early explorations of CAT: canine amusement and training. *CHI '10 Ext. Abstr. Hum. Factors Comput. Syst.*, 2661–2669.
24. Kyoko Yonezawa, Takashi Miyaki, and Jun Rekimoto. 2009. Cat@Log: sensing device attachable to pet cats for supporting human-pet interaction. *Proc. Int. Conf. Adv. Comput. Enterntainment Technol. - ACE '09*, 149–156.
25. 2013. ZOO Boomer balls. Retrieved July 1, 2015 from https://www.youtube.com/watch?v=Od_Lm8U5W44