



Assessing machine learning classifiers for the detection of animals' behavior using depth-based tracking



Patricia Pons^{a,*}, Javier Jaen^a, Alejandro Catala^b

^aISSI Group, Departamento de Sistemas Informáticos y Computación (DSIC), Universitat Politècnica de València, Camino de Vera s/n, 46022 Valencia, Spain

^bHuman Media Interaction, University of Twente, Drienerlolaan 5, 7522 NB Enschede, The Netherlands

ARTICLE INFO

Article history:

Received 20 April 2016

Revised 26 May 2017

Accepted 27 May 2017

Available online 29 May 2017

Keywords:

Tracking system

Animal Computer Interaction

Depth-based tracking

Classification algorithms

Intelligent system

ABSTRACT

There is growing interest in the automatic detection of animals' behaviors and body postures within the field of Animal Computer Interaction, and the benefits this could bring to animal welfare, enabling remote communication, welfare assessment, detection of behavioral patterns, interactive and adaptive systems, etc. Most of the works on animals' behavior recognition rely on wearable sensors to gather information about the animals' postures and movements, which are then processed using machine learning techniques. However, non-wearable mechanisms such as depth-based tracking could also make use of machine learning techniques and classifiers for the automatic detection of animals' behavior. These systems also offer the advantage of working in set-ups in which wearable devices would be difficult to use. This paper presents a depth-based tracking system for the automatic detection of animals' postures and body parts, as well as an exhaustive evaluation on the performance of several classification algorithms based on both a supervised and a knowledge-based approach. The evaluation of the depth-based tracking system and the different classifiers shows that the system proposed is promising for advancing the research on animals' behavior recognition within and outside the field of Animal Computer Interaction.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

Technology is unquestionably changing our world and our lives every day. In this ever-growing digital era, human beings are not the only ones who can make use of technology. With an estimated 75 million households owning at least one pet animal in Europe,¹ and more than 71 million pet dogs and 73 million pet cats in USA,² animals are also a significant population coexisting with our technological surroundings. Domestic dogs and cats have been observed using some of our technological devices, such as smartphones or tablets, in their own way (Baskin, Anavi-goffer, & Zaminsky, 2015; Noz & An, 2011; Westerlaken & Gualeni, 2014) and some zoos are also introducing touchscreens and tablet games for primate enrichment (Carter, Webber, & Sherwen, 2015). All these animals could benefit from the technological advances we have achieved throughout the digital revolution. However, animals have different physical features and mental perceptions of the world, preventing them from fully using and understanding our technol-

ogy and interaction methods, which have been designed with human requirements and characteristics in mind.

Recently, spreading research is addressing the aforementioned concerns within the field of Animal-Computer Interaction (ACI) (Mancini, 2011, 2013). ACI considers animals as the target users of digital systems, and grounds on the development of computer interfaces and digital systems specifically designed for them. Animal-centered technology could improve animal welfare and wellbeing in several scenarios: interactive systems or devices could provide enrichment and stimulation for captive animals in zoos (Carter et al., 2015; French, Mancini, Sharp, & Smith, 2014), digital and/or tangible games could foster physical activity of animals in shelters and even entertain pets alone at home (Hirskyj-Douglas, Luo, & Read, 2014; Pons, Jaen, & Catala, 2014), alleviating stress and isolation.

In the era of computers and with the advances in computer vision and machine learning techniques, a promising research line into ACI would be automatizing behavior recognition on animals. Animals are not verbal communicators; instead, they rely on body postures or sounds to express themselves. For these reasons, ACI studies are taking a big effort in developing technology for the automatic recognition of animal behavior and body postures. The benefits this technology could bring both to animal welfare and ACI research are countless. Firstly, it could provide both objective

* Corresponding author.

E-mail addresses: ppons@dsic.upv.es (P. Pons), fjaen@upv.es (J. Jaen), a.catala@utwente.nl (A. Catala).

¹ http://www.fediaf.org/fileadmin/user_upload/Secretariat/facts_and_figures_2014.pdf.

² <http://www.petfoodinstitute.org/?page=PetPopulation>.

measurement mechanisms and more reliable feedback to inform the design and development of animal-centered technology. Second, animal welfare could be evaluated using automatic behavior recognition paired with machine learning techniques. This knowledge could be used to support the detection of abnormal behaviors of animals, allowing early detection of illnesses and other kind of problems which could be derived from an abnormal behavior. Finally, body posture and behavior recognition could also be used to automatically adapt the reactions of a system to the animals' interactions (Pons et al., 2014; Pons, Jaen, & Catala, 2015b). In this way, animals could interact with systems on their own, improving animal wellbeing when they are alone (at home, zoos, shelters, etc.).

Great advances are being done using wearable devices as a way of recognizing animals' body postures and activities. These works make use of the information provided by accelerometers and gyroscopes attached to a wearable device such as a collar or harness that the animal has to put on. The extracted raw information from the sensors is then processed using machine learning techniques to train different classification algorithms to recognize animals' postures and activities. Depending on the activities and/or postures to be recognized, a wide range of classification algorithms have been used, either as standalone algorithms or combining them in a more complex learning process. However, there are animals who are neither used to wear harnesses nor other wearable devices and could find those elements disturbing. Other animals, such as cats, have extreme agility and wide range of movements, and wearable devices could limit their naturalness. There are also animals to whom the use of wearable devices could pose a threat, such as wild animals, zoo animals or protected species. In addition, while wearable tracking systems might offer wider coverage area, they also require maintenance of batteries and sensors and one wearable device can only track one animal at a time. The use of non-wearable tracking mechanisms would be a promising complement for analyzing behavior and body postures in cases in which the use of wearable devices will not be feasible. These systems would allow more natural interactions within technologically-mediated environments as the animals will not be required to wear or carry any device, avoiding stress or affecting their behavior. Although the use of non-wearable systems would be delimited to a specific tracking area, they offer centralized maintenance and a single tracking device could provide information of several animals at a time. This information could also be used to train different classification algorithms for the automatic recognition of animals' postures, therefore providing the same benefits to ACI as the wearable approach in terms of animal wellbeing.

This paper describes the development of a non-wearable depth-based tracking system for cats and the promising results obtained by applying classification algorithms on the obtained depth information for the automatic recognition of the animals' body parts and postures. The use of depth information along with traditional computer vision techniques provides more information about the tracked animal than using solely an RGB camera. By exploiting this information using machine learning techniques and suitable classification algorithms, this approach could be a promising starting point towards the automatic detection and analysis of animals' behavior without requiring the animal to use any wearable device. This article is structured as follows: Section 2 analyzes previous tracking systems for animals based on wearable and non-wearable devices and their purposes, and states the necessity of a new approach grounding on previous research on non-wearable human-tracking systems. Section 3 describes the development of a depth-based tracking system for cats in indoor scenarios. Section 4 reports two experiments carried out to determine the accuracy of the system using both a supervised learning process and a knowledge-based approach for the classification of the cats' body postures and body parts, describing the different classification al-

gorithms used and their accuracy rates. Section 5 explains how this tracking mechanism could be applied in several ACI domains and how it could be coupled with human and object tracking, and conclusions and future work are given in Section 6.

2. Related works

This section provides an overview on existing research about animal tracking focused on animal gesture and body posture recognition for different purposes using both wearable and non-wearable tracking systems.

2.1. Wearable tracking systems

Several works have addressed the necessity of tracking animals in different scenarios. The most common method to gather information about the animal has been using wearable harnesses or collars with attached technological devices providing information to the system in charge of processing the information. One of the most basic methods for animal tracking in outdoor scenarios has been relying on GPS or radio-frequency localization, attaching the emitter devices to a collar or harness. These systems only give information on the animals' location and have been used by pet owners, mostly to assess their dogs' locations and whether or not they are in trouble (Mancini, van der Linden, Bryan, & Stuart, 2012). This technology has also been used during hunting activities with dogs, allowing the human leading the hunting activity to interpret the movements of the dog in the field by following its signal on a handheld display (Paldanius, Kärkkäinen, Väänänen-Vainio-Mattila, Juhlin, & Häkikilä, 2011; Weilenmann & Juhlin, 2011).

However, several outdoor scenarios require more precise information about the animals' movements or body postures during the activity, and even some kind of communication from the animal to the human side. As an example, determining the pose of the animal is of vital importance in the case of Search and Rescue (SAR) dogs. Due to its agility and strong sense of smell, dogs are suited to perform SAR tasks which are not always safe for humans, such as accessing small locations or identifying potential locations of survivors after a catastrophe. Usually SAR dogs have to work away from human sight, and it would be extremely useful for the dog handlers if they could know the location and pose of the dog to determine if the dog is trying to communicate some discovery, and to assess the physical wellbeing of the animal. Recognition of animals' postures and activities in this kind of scenarios are usually performed using accelerometers, gyroscopes or other inertial measurement units, and the majority of works are focused on dogs (Bozkurt et al., 2014; Zeagler et al., 2016). In these works, differentiating clearly between what is activity and what is a posture is usually difficult. A specific activity entails that the animal adopts a specific posture, e.g. walking or jumping are different activities identified by their posture, and eventually some postures are in itself an activity, e.g. sitting.

Most of the works based on wearable devices for activity/posture recognition are based on the use of a tri-axial accelerometer located at the dog's collar and then apply classification techniques to the data obtained from the accelerometer in order to recognize the activity/posture. There are several devices for dogs, some of them even commercial, such as Whistle®,³ FitBark®⁴ or WagTag™ (Weiss, Nathan, Kropp, & Lockhart, 2013) which make use of a tri-axial accelerometer to perform basic activity level recognition. However, these systems are only capable of

³ <http://www.whistle.com/>.

⁴ www.fitbark.com.

Table 1
Existing wearable tracking systems for animals and summary of features.

| Related work | Device | Indoor/Outdoor use | Position detection | Posture recognition | Activity recognition |
|---|--|---|--------------------|---------------------|----------------------|
| (Mancini et al., 2012) | Collar-worn GPS | Outdoor | Yes | No | No |
| (Paldanius et al., 2011) | Collar-worn GPS | Outdoor | Yes | No | No |
| (Weilenmann & Juhlin, 2011) | Collar-worn GPS | Outdoor | Yes | No | No |
| Whistle | Collar-worn GPS and accelerometer | Indoor and outdoor | Yes | No | Basic |
| FitBark | Collar-worn accelerometer | Indoor and outdoor | No | No | Basic |
| WagTag | Collar-worn accelerometer | Indoor and outdoor | No | No | Basic |
| (Ladha et al., 2013) | Collar-worn accelerometer | Indoor (possibly outdoor if the device range of frequencies allow it) | No | Yes | Yes |
| (Valentin, 2014) | Accelerometer on a harness | Indoor and outdoor | No | No | Yes |
| (Valentin et al., 2015) | Collar-worn accelerometer and gyroscope | Indoor and outdoor | No | Yes | Yes |
| (Ribeiro et al., 2009) | Two accelerometers on a harness | Indoor and outdoor | No | Yes | No |
| (Bozkurt et al., 2014; Brugarolas et al., 2013) | Two accelerometers and two gyroscopes on a harness | Indoor and outdoor | No | Yes | Yes |
| (Watanabe et al., 2005) | Two collar-worn accelerometers | Indoor | No | Yes | Yes |
| PawTrack | Collar-worn GPS | Outdoor | Yes | No | No |
| Cat@Log | Collar-worn GPS and accelerometer | Indoor and outdoor | Yes | Yes | Yes |
| (Wingrave et al., 2010) | Harness with IR emitters and Wiimote IR camera | Indoor | Yes | Yes | No |

indicating if the dog was resting or moving and do not differentiate between different activities which involve movement. In Ladha, Hammerla, Hughes, Olivier, and Ploetz (2013), dogs wear a tri-axial accelerometer on the collar and, after being trained with a kNN classifier, the system is able to differentiate between 14 activities and 2 postures.

Within the FIDO project (Jackson et al., 2013), researchers have been studying how wearable devices could mediate the communication between working dogs and their handlers. They have undertaken extensive work on providing dogs with suitable wearable activators (Jackson et al., 2015). In addition, they have also considered to mediate this communication by recognizing motion-based dog gestures – sit, spin, roll, jump, etc. – using a three-axis accelerometer attached to the front of a service dog harness (Valentin, 2014). More recently, they have studied the use of a dog collar with an accelerometer and gyroscope for the recognition of head gestures on dogs (Valentin, Howard, & Jackson, 2015).

The effectiveness of wearable harnesses with several inertial measurement units located along the harness has also been studied. The work of Ribeiro, Ferworn, Denko, and Tran (2009) uses the angles of two accelerometers on different locations on the dogs' harness to develop an algorithm capable of estimating four poses including: standing, lying down, sitting, and walking. Other works (Bozkurt et al., 2014; Brugarolas et al., 2013) extend this idea by using more inertial measurement units located on the optimal locations of a dog's body, which have been determined attending to the algorithm's performance and the dogs' comfort and physiology. Using the information provided by these units and applying machine learning techniques, five static postures and three dynamic behaviors can be identified. They have also compared the performance of the classification algorithm using supervised against unsupervised classification methods (Winters et al., 2015).

Acceleration data-loggers are also a common and efficient way of detecting cats' body postures and frequent behaviors based on movement (Watanabe, Izawa, Kato, Ropert-Coudert, & Naito, 2005). Commercial devices for cat activity recognition are also available, such as PawTrack®,⁵ which detects whether the cat is at home or outside, and offers GPS geolocation for outdoor walks. However,

it neither monitors any activity nor gesture. Cat@Log (Yonezawa, Miyaki, & Rekimoto, 2009) is a non-commercial but more complete device. It consists of 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.

Canine Amusement and Training (Wingrave, Rose, Langston, & LaViola, 2010) presents a wearable tracking system for dogs not based on accelerometers. It consists of IR emitters attached to the dog's harness, and a Wiimote's IR camera placed on the ceiling. The system detects the location and posture of the animal by tracking the IR emissions of the harness using the Wiimote. The detected postures and location are used by the system to determine whether the dog is performing correctly the proposed training activities offered by the system. Table 1 provides a summary of the existing approaches for wearable tracking systems for animals and their most distinctive features: the device being used for tracking, whether it works in indoor or outdoor locations, and whether or not it detects the position of the animal within the tracking area, its posture and/or its activity.

2.2. Non-wearable tracking systems

Although wearable devices are the most common way of tracking an animals' position and posture, a number of non-wearable solutions have also been proposed. Poultry.Internet (Lee et al., 2006) illustrates a remote communication system between a pet and its owner, in which the owner is able to remotely obtain real-time information on the location and orientation of its poultry inside the house backyard. This system tracks the movements of the 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. In Karlsson, Ren, and Li (2010) computer vision methods have been used to track the movements of animals inside a zoo environment using multiple cameras, but again neither body postures nor gestures are identified. No wearable device is either used in Purrfect Crime (Trindade et al., 2015), an interspecies digital game for cats and

⁵ <http://pawtrack.com/>.

humans. In this case a Microsoft Kinect®⁶ is used to detect the position of the cats inside the play area, using the depth information provided by this sensor. However, the system only detects the central position of the animal, and neither posture nor orientation is detected. Therefore, the interactive responses of the system were sometimes erroneous, as not being able to identify where the cat was looking introduced some interactions which were not really intended by the animal.

To the best of the authors' knowledge, none of the existent non-wearable systems intended for animal tracking is capable of detecting body postures. In order to address the aforementioned limitations that wearable devices present in several application domains, there is a need for the development of tracking mechanisms for animals not based on wearable devices which allow to detect body postures as well as locating the animal within the tracked area. The following section describes a promising non-wearable tracking system for animals based on depth information. For the recognition of the animals' body parts and postures using the data provided by this depth-based tracking system, it has been essential to analyze how different classification algorithms perform in this domain. Both supervised and knowledge-based classification techniques have been tested, and the promising results obtained will be described in Section 4.

3. Tracking system

Within Human Computer Interaction, depth sensors have been successfully used for gesture detection and posture recognition on human beings. A depth frame of an image provides, for each pixel, the distance in millimeters from the sensor plane to the nearest object in that particular pixel. Depth sensors, such as the Microsoft Kinect®, have been very useful to detect and recognize volumes and 3D shapes from 2D images. Usually, these sensors are located in a vertical plane, either in front of the user who is interacting or facing the scene to be analyzed. However, recent works have located this kind of sensors on the ceiling to track an open space area (Benko, Wilson, Zannier, & Benko, 2014; Jones et al., 2014; Jones, Benko, Ofek, & Wilson, 2013; Moreno, Van Delden, Poppe, Reidsma, & Heylen, 2015), providing wider tracking areas and avoiding occlusion due to elements in the room. Human gestures can be detected using this set-up (Bednarik & Herman, 2015; Hu, Reilly, Alnusayri, Swinden, & Gao, 2014; Lin, Liu, Hsu, & Fu, 2015), and this configuration of the depth-sensor could also be useful for animal tracking applications. The skeleton of animals such as cats or dogs should be easily recognizable from above in several postures, e.g. sitting or walking, if volumetric information of the image such as depth analysis is used: the head could be identified as a volumetric shape different from the body and/or tail, and depth information would also allow to locate the position of each body part in a tridimensional representation space. In the case of other animals more similar to human beings in terms of skeletal characteristics, such as orangutans, similar approaches as the ones used for human gesture detection could be used. Therefore, a depth-based tracking system would be a promising way of detecting the animals' location, posture and field of view (Pons, Jaen, & Catala, 2015a, 2017). In the following sections, a depth-based tracking system for the detection of cats' body postures, location and orientation is described. The tracking system has been developed and tested with cats as target users, but it could be adapted to work with other animal species using a similar approach.

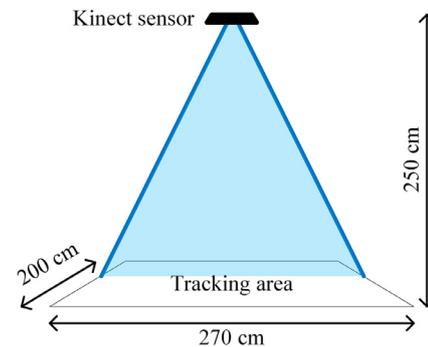


Fig. 1. Set-up for the tracking system.

3.1. Equipment and procedure

Several sessions with cats were carried out in which a Microsoft Kinect® v1.0 sensor was used to record video streams of depth and color information from the cats' natural movements during a period of time. During the sessions, cats moved freely and were also encouraged to play with their owners, caretakers or small interactive robotic toys. The Microsoft Kinect® was placed looking down from the ceiling at a height of 250 cm, where it covered an area of approximately 200 cm long and 270 cm wide, as shown in Fig. 1. The tracking area was a clear space with neither furniture nor objects besides the toys/robots used for the games. The Microsoft Kinect® recorded both color and depth video streams at a rate of 30 frames per second with 640×480 pixel resolution.

Ethical guidelines for ACI studies were considered (Väätäjä & Pesonen, 2013), and therefore the subjects were not forced to interact and they could walk freely around the room. Cats were encouraged to interact within the tracking area of the sensor by means of their owners/caretakers drawing their attention to this area with toys or by calling them. However, as the cat could move freely inside the room where the interaction took place, only the moments in which the cats were within the tracking area were valid recordings. These sessions allowed to obtain real data on common and spontaneous postures, behaviors and movements, which were later analyzed and processed to develop the depth-based tracking system. The tracking system has been developed using C#, Microsoft Kinect® SDK 1.8, and EmguCV, an OpenCV framework for .NET systems which has been used for image processing.

3.2. Processing depth-based information

The Microsoft Kinect® v1.0 sensor provides both color (see Fig. 2a) and depth streams (see Fig. 2b). The tracking system will only use depth information as input, so the color streams are discarded. 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). In the obtained depth images, the contours of the cats can be clearly observed as their depth values are greater than the floor's depth, which is constant. Instead, the depth pixels of a cat vary along its body, allowing a human eye to differentiate between the different parts of the cat's body just by looking at the depth frame. In the same way, different cat postures can be observed to generate different cat contours in the processed depth frames.

The first processing step of the algorithm consists of extracting the cat's pixels from the depth frame (see Fig. 2b). In the current implementation of the algorithm, which has been tested with pre-recorded data, background segmentation has been done by simply discarding from the image those pixels corresponding to the ground, as the sensor was placed in a fixed position during the

⁶ <https://msdn.microsoft.com/en-us/library/hh855355.aspx>.

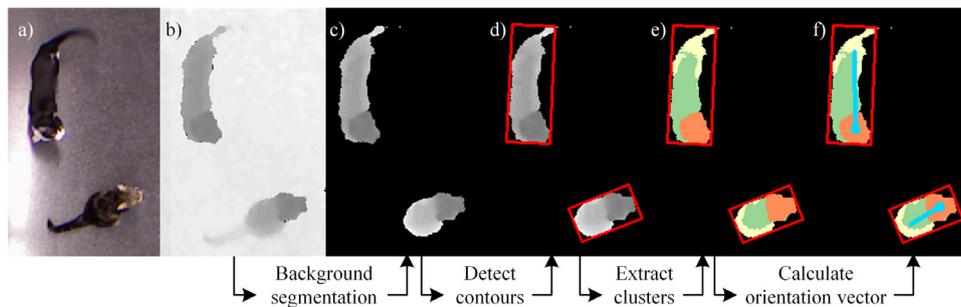


Fig. 2. Process of extracting the cat's orientation: (a) color frame (b) depth frame, (c) background segmentation, (d) cat contours, (e) clusters for head, body and tail, (f) orientation vector. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

sessions and therefore the distance to the floor was known (see Fig. 2c). However, this approach can unintentionally remove the cats' tail (as observed in Fig. 2c, cat on the lower-right corner of the image), which in some cases could be a source of valuable information when creating richer posture descriptors. Thus, more elaborated approaches for background segmentation should be incorporated in order to overcome this current limitation and to allow for more flexible set-up conditions. With the floor removed from the image, computer vision algorithms are applied to the depth image in order to extract the cats' contours, which now appear as grey-scale blobs on the image (see Fig. 2d). In this step, a cat's location within the tracked area can be determined by using the centroid of the extracted contours as a 2D coordinate. Each detected contour is then processed 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 in order to divide the cat contour into the three most noticeable parts of the cat's body, i.e. head, body and tail. The following step of the tracking system would be the recognition of the cats' postures and body parts. For this purpose, the obtained clusters of each cat have to be classified into either head, body or tail. Once the head is detected, its position in relation to the body/tail clusters allows an orientation vector to 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. Moreover, not only the cats' body parts can be detected but also different body postures can be identified. A pseudocode description of the tracking, classification and clustering process is shown in Algorithm 1 and Algorithm 2.

Algorithm 1 Processing pipeline.

```

BEGIN
1. Capture depth image
2. Carry out background segmentation
3. Obtain contours
4. Call Algorithm 2
5. Calculate orientation vector
END

```

Algorithm 2 Cat clustering algorithm.

```

Input: feature vector
Output: posture, head, body, tail
BEGIN
1. Obtain unclassified clusters with k-means (k=3)
2. Match body parts to clusters and obtain posture
END

```

Step 2 in our clustering algorithm (Algorithm 2) can be carried out in different ways. In this work, a supervised training method

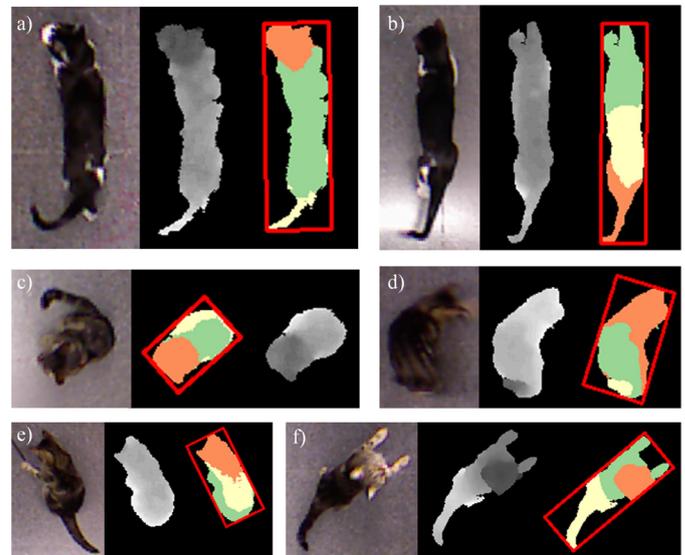


Fig. 3. Cat's postures and the corresponding depth and clustered image (a) standing (b) walking (c) sitting (d) turning (e) semisitting (f) jumping.

and a knowledge-based classification system have been evaluated for the detection of both postures and body parts. Both approaches will be explained in Sections 4.1 and 4.2, respectively. A dataset was created from the recordings in order to train and evaluate these classification mechanisms. It was observed that the tracking device was providing better images for one of the cats, probably due to its color, size and density of the coat. However, with the current data it was difficult to quantify to which extent each factor was affecting the tracking accuracy. To avoid introducing errors due to such factors in the experiment, only images in which the cat contour was correctly tracked and extracted were used for the experimental data. The dataset is comprised of 1422 contours of cats in different postures, which were manually labeled indicating the posture and the different body parts of the cat.

4. Classification results and discussion

A set of experiments have been conducted in order to determine the accuracy of the system in the detection of a cat's posture and the classification of its body parts. The following postures were considered, and representative color and clustered images of each of them can be observed in Fig. 3:

- Standing: the cat is standing on its four legs, head slightly higher than the body.
- Walking: the cat is standing, head slightly bent forward and legs moving forward.

Table 2

Accuracy rates for base learners when classifying a cluster belonging to a cat's contour.

| | Decision tree | Random tree | Random forest | Rule induction* | Support vector machine* | kNN ($k = 4$) | Naïve Bayes | Logistic regression* |
|---------|---------------|-------------|---------------|-----------------|-------------------------|-----------------|-------------|----------------------|
| Head | 97.57% | 42.98% | 77.55% | 84.63% | 64.90% | 61.46% | 34.16% | 49.23% |
| Body | 47.44% | 40.43% | 50.82% | 74.70% | 59.37% | 54.58% | 29.63% | 59.34% |
| Tail | 93.00% | 63.57% | 85.29% | 81.18% | 73.29% | 66.23% | 59.65% | 81.46% |
| Average | 62.87% | 47.17% | 62.28% | 82.20% | 66.67% | 60.64% | 34.81% | 60.33% |

Table 3

Accuracy rates for base learners when classifying a cat's posture.

| | Decision tree | Random tree | Random forest | Rule induction | Support vector machine | kNN ($k = 4$) | Naïve Bayes | Logistic regression |
|-------------|---------------|-------------|---------------|----------------|------------------------|-----------------|---------------|---------------------|
| Standing | 72.62% | 50.65% | 55.17% | 85.41% | 83.44% | 85.32% | 84.62% | 84.62% |
| Walking | 75.12% | 51.35% | 76.39% | 77.19% | 74.89% | 74.55% | 63.31% | 71.43% |
| Sitting | 89.77% | 67.27% | 85.71% | 92.90% | 89.08% | 87.63% | 90.86% | 90.00% |
| Semisitting | 81.15% | 58.76% | 87.25% | 84.78% | 78.69% | 86.57% | 81.02% | 78.12% |
| Turning | 81.40% | 42.07% | 79.35% | 80.19% | 70.98% | 71.68% | 62.17% | 73.78% |
| Jumping | 92.31% | 80.47% | 96.84% | 94.71% | 93.66% | 96.69% | 96.97% | 92.92% |
| Average | 79.56% | 55.84% | 69.59% | 85.50% | 81.93% | 83.32% | 79.01% | 82.05% |

Table 4

Accuracy rates for base learner when classifying a cat's posture, using forward feature selection.

| | Decision tree | Random tree | Random forest | Rule induction* | Support vector machine* | kNN ($k = 4$) | Naïve Bayes | Logistic regression* |
|-------------|---------------|-------------|---------------|-----------------|-------------------------|-----------------|---------------|----------------------|
| Standing | 76.05% | 63.06% | 74.64% | 88.64% | 85.92% | 92.98% | 82.77% | 73.81% |
| Walking | 79.27% | 59.56% | 71.94% | 80.60% | 69.70% | 87.83% | 77.92% | 79.03% |
| Sitting | 91.33% | 87.80% | 90.75% | 94.12% | 83.61% | 94.12% | 96.64% | 88.89% |
| Semisitting | 80.45% | 74.80% | 78.86% | 82.61% | 100.00% | 84.67% | 82.19% | 77.50% |
| Turning | 74.57% | 62.42% | 73.33% | 76.81% | 74.24% | 87.32% | 71.22% | 83.78% |
| Jumping | 93.56% | 92.57% | 97.01% | 96.72% | 96.61% | 97.52% | 96.52% | 89.23% |
| Average | 81.10% | 70.74% | 79.58% | 86.62% | 83.80% | 91.22% | 83.63% | 80.05% |

- Sitting: the cat is sitting on its rear legs, without bending its front legs.
- Turning: the head of the cat and its front part of the body are curving towards one side.
- Semisitting: the cat is sitting on its rear legs but bending towards the front.
- Jumping: the cat jumps on its back legs, head and front paws up, all body extended vertically.

4.1. Supervised classification of body parts and postures

Supervised learning has been used to classify the three different body parts of a cat's contour into head, body and tail. The following features have been used to describe each body part: width and height of the cat's contour, average depth, number of pixels and shape descriptors (second order moments and Hu invariant moments). It should be noticed that the posture has not been considered a feature. In this way, the algorithm can firstly classify a cat's body parts regardless of its posture, and then use that information to properly create the feature vector of a cat's posture as will be explained later in this section. The labeled dataset of cats' body parts contained a total of 4266 feature vectors (1422 feature vectors of each class, i.e. head, body and tail) and the analysis was performed using the RapidMiner data mining tool. Table 2 shows the accuracy of several base classifiers tested using k -fold cross validation (10 validations, stratified sampling) to analyze their performance, except for those marked with *, in which simple split validation was used (70% of the data set used for training, stratified sampling).

Results show rather low average accuracy rates. However, it can be seen that for several classification algorithms such as decision trees and random forest, very promising accuracy rates are obtained when classifying the head and the tail. As we know that for each cat's contour, there is only one head, tail and body, the first two kinds of clusters could be identified using the learned classification model and the latter one, the body cluster, would

be the remaining one. To demonstrate this hypothesis, a combined model has been built using a stacking approach and considering the three best base learning algorithms for this data: rule induction, support vector machine and decision tree. The resulting model, combined using rule induction, has an average accuracy of 83.18% (head = 88.22%, body = 77.75%, tail = 83.97%). As expected, it follows the prediction if the output class is either head or tail, or classifies the cluster as body otherwise. With the head correctly classified, the field of view of the cat can be roughly estimated as an orientation vector from either the center of the body or tail to the center of the head.

Supervised learning has also been used to classify a cat's posture. The same dataset of 1422 manually labeled images of cats in different postures was used. For each posture, the following features were considered: width and height of the cat's contour, clusters basic info (centroid, average depth, and number of pixels), distance between head to body centroids, distance between tail and body centroids, distance between head and tail centroids, angle between the vectors from body to tail and from body to head, depth differences between clusters (head and body, head and tail, body and tail), with a total of 21 features. Base learners performed as shown in Table 3 using k -fold cross-validation (100 validations, stratified sampling) when classifying the aforementioned postures. As can be observed in Table 3, rule induction was shown to be the best performing algorithm considering average accuracy scores. However, some algorithms perform better than others depending on the posture being classified. As an example, classification of jumping postures has slightly better results using a kNN, random forest or Naïve Bayes classifier rather than rule induction.

Forward feature selection was applied on all base learners in order to discard features that could introduce noise into the classification. Accuracy results for the same cat postures are shown in Table 4. k -Fold cross validation was used to analyze the performance of all classifiers (100 validations, stratified sampling), except for those marked with *, in which simple split validation was

used (70% of the data set used for training, stratified sampling). It can be observed that in few cases the average and individual accuracy rates are significantly improved, such as in the random tree, random forest and kNN classification algorithms. The selected features are different depending on the algorithm being used, hence it seems that there are no noisy features and the best performing subset of features will depend on the algorithm being used. With this approach, the kNN classifier not only has the best average performance score, but also is the best classifier for four of the six postures, and in the remaining two postures it is the second best performing classifier.

As a final validation, a combined model has been trained using a stacking approach, considering the three best performer classifiers (kNN, rule induction and support vector machine) and combining them using rule induction. Forward feature selection has been applied to the combined model. The average accuracy for this model is 88.50%, with the following class accuracy: standing = 87.32%, walking = 80.56%, sitting = 94.00%, semisitting = 86.36%, turning = 89.47%, jumping = 96.72%.

The obtained results are very promising and demonstrate that reliable tracking systems for animals based on depth information can be developed using machine learning algorithms for the classification of the cat's body parts and postures.

4.2. Knowledge-based classification of body parts and postures

The process for supervised learning is time demanding as it requires to label the data in order to construct the training set and build a suitable model to apply to the new data. As this system is envisioned to adapt to cats of different breeds, sizes and physical characteristics, it would be very beneficial if the tasks of labeling, training and deploying the model could be eased. Other works have also considered including knowledge-based models directly encoded in classification algorithms for solving several problems (Li, Goldgof, & Hall, 1993; Winters et al., 2015), when the costs for the preparation of datasets and training required in supervised learning are not feasible. Therefore, an exploratory study on the accuracy of the tracking system using knowledge-based classification has been conducted in order to assess the suitability of such an approach. If feasible, it could allow easy deployment and adaptation of the system to new scenarios and users, i.e. cats of different sizes and breeds, with minimal configuration requirements.

In order to develop such a system, an observational analysis on a sample of the recordings and their corresponding clustered images was conducted. Different cat postures were seen to generate different cat contours in the processed depth frames, and the sizes and depth values of the different clusters also varied from one posture to another. As an example, the cat's depth stream contour when sitting showed a smaller, square-shaped bounding rectangle (see Fig. 3c), while the depth stream when standing or walking revealed a larger and rectangular-shaped bounding rectangle (see Fig. 3a and b). 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. 3c, in which the grey pixels of the head are significantly darker than the pixels of the cat's haunch. In contrast, the cat walking in Fig. 3b has an average depth of its head pixels very similar to the average depth of the haunch and tail pixels. Another aspect to consider is that the k-means clustering algorithm has been fixed to provide three clusters at all times. Hence, the classification algorithm for the different body parts is the one responsible of determining which one would correspond to the lower back of the cat. This cluster, in most cases, corresponds to the tail. However, in some cases in which the tail is not visible such as in Fig. 3c, the detected cluster physically corresponds to the lower back of the animal. These cases usually occur when the cat is in a sitting posture. In this case, the head is clearly differ-

Table 5

Accuracy of the tracking system identifying the cats' postures and body parts.

| | Sitting/semi-sitting | Walking/standing | Jumping | Turning |
|------------|----------------------|------------------|---------|---------|
| Posture | 84% | 49.5% | 69% | 53% |
| Body parts | 99% | 74.5% | 92.5% | 71% |

entiated 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 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. 3b 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.

After the analysis, a decision tree algorithm was developed which considered the following parameters: dimensions of the cat's contour, number of pixels for each cluster and average depth for each cluster. The decisions on the tree are made in terms of the observed average values for each feature on the data used for the analysis, allowing a threshold for the variance of each of these values. Several simplifications have been made. Firstly, the algorithm cannot differentiate between walking and standing positions, nor between sitting and semi-sitting, as considering this basic information the average and threshold values overlap and it is only possible for machine learning algorithms to build such an elaborate classification. Therefore, this basic version of the tracking system is focused on 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. Secondly, with the supervised approach it was possible to isolate the classification of postures from the classification of clusters into cats' body parts. It is not possible to do that in this knowledge-based approach, as for the human observer the identification of each cluster depends on the posture being analyzed, and vice versa. Third, an extra piece of information has been given to the algorithm. If the decision tree cannot find the head of the cat, information from the last previous frame will be used to find the closest cluster which was classified as the head in the last frame. The pseudocode in Algorithm 3 shows the basic behavior of the decision tree algorithm.

An accuracy analysis has been conducted extracting 200 random frames of each posture and processing them offline by the tracking and classification algorithms. Table 5 shows a summary of the results indicating the percentage of frames for each posture in which the algorithm correctly identified the posture and body parts of the cats respectively. Regarding the posture, the results indicate the percentage of cases in which the decision tree assigns the correct label for the analyzed cat's contours. Regarding the body parts, the results indicate the percentage of cases in which the algorithm correctly identifies the head of the cat and its body and/or tail.

As it can be observed, the sitting position is one of the most distinctive ones as it usually comprises a small area in which the head can be clearly identified and hence the body parts are very easily classified and the orientation vector between the head and body/tail can be almost perfectly traced. The algorithm also classifies very well the cat's body parts when jumping, although the classification of this posture has a 31% error rate. This is mostly due to the moments in which the jump is starting or finishing. In these cases, the threshold value established to determine the jump, which is the difference in depth between the highest and the deepest cluster, does not fit well. This accuracy rate could probably be improved by adjusting more carefully this threshold. The algorithm had problems identifying the walking/standing

Algorithm 3 Knowledge-based decision tree algorithm.

```

Input: cat's contour size, clusters' average depth, clusters' number of pixels, clusters'
centroids, head
cluster's centroid from last frame
Output: posture, head cluster, body cluster, tail cluster
BEGIN
1.   if (small bounding rectangle of the cat's contour)
2.     if (difference between highest and deepest cluster > threshold)
3.       posture=jumping
4.       head=highest cluster (cluster with smallest depth value)
5.       body=intermediate cluster
6.       tail=deepest cluster (cluster with biggest depth value)
7.     else if (number of pixels of highest cluster < average tail max. size)
8.       posture=turning
9.       tail=highest cluster (cluster with smallest depth value)
10.      head=intermediate cluster
11.      body=deepest cluster (cluster with biggest depth value)
12.     else if (number of pixels of highest cluster < average head max. size)
13.       posture=sitting
14.       head=highest cluster (cluster with smallest depth value)
15.       body=intermediate cluster
16.       tail=deepest cluster (cluster with biggest depth value)
17.     else
18.       posture=unknown
19.       head=closest cluster to the head cluster detected in last frame
20.       tail=furthest cluster from head
21.       body=remaining cluster
22.   else
23.     if (difference between highest and deepest cluster > threshold)
24.       posture=jumping
25.       head=highest cluster (cluster with smallest depth value)
26.       body=intermediate cluster
27.       tail=deepest cluster (cluster with biggest depth value)
28.     else if (number of pixels of highest cluster < average tail max. size)
29.       tail=highest cluster (cluster with smallest depth value)
30.       head=furthest cluster from tail
31.       body=remaining cluster
32.     if (bounding rectangle of the cat's contour within average turning posture dimensions)
33.       posture=turning
34.     else
35.       posture=walking/standing
36.     else if (number of pixels of second highest cluster < average tail max. size)
37.       tail=second highest cluster
38.       head=furthest cluster from tail
39.       body=remaining cluster
40.     if (bounding rectangle of the cat's contour within average turning posture dimensions)
41.       posture=turning
42.     else
43.       posture=walking/standing
44.     else if (number of pixels of highest cluster > average tail max. size AND
45.       number of pixels of highest cluster < average head max. size AND
46.       depth difference between highest and second highest cluster is very small)
47.       posture=sitting/semisitting
48.       head=highest cluster (cluster with smallest depth value)
49.       body=intermediate cluster
50.       tail=deepest cluster (cluster with biggest depth value)
51.     else
52.       posture=unknown
53.       head=closest cluster to the head cluster detected in last frame
54.       tail=furthest cluster from head
55.       body=remaining cluster
END

```

position as well as the turning one. Nevertheless, in both cases the classification of the body parts offers promising accuracy rates and allows to determine an orientation vector to estimate the animal's field of view.

At this stage the knowledge-based classification algorithm would not be suitable to provide very accurate posture classification in all cases. It is likely that the decisions coded in the algorithm do not represent well the most characteristic features of each posture as it has been a manual process based on observation and codification. This is an issue that supervised classification algorithms are capable of resolve, as it has been demonstrated by the results in [Section 4.1](#). However, the knowledge-based classifica-

tion algorithm presented in this section performs well in terms of body parts classification, allowing to determine the different body parts of the animal correctly with less training time than in a supervised approach.

Therefore, for a system to automatically recognize between different body postures of an animal with very high certainty, supervised classification algorithms are preferred over a basic knowledge-based approach. If more complex systems are to be developed, supervised classification algorithms are the recommended approach. For example, they could provide very reliable information in monitoring systems for animal welfare and behavioral pattern recognition, or remote communication systems

for dogs with occupations based on postures. Nevertheless, a knowledge-based classification technique could provide satisfactory accuracy rates if other types of information derived from the body parts of the animal are required, such as determining its field of view or just detecting a specific part of the animal such as the head. The next section will provide an overview of the different systems in which classification algorithms as the ones presented would be a key component.

5. Applications within Animal–Computer Interaction

5.1. Behavior recognition, learning behavioral habits and welfare assessment

Non-wearable depth-tracking systems as the one presented in this paper would significantly contribute to the development of ACI research. ACI studies usually rely on the observational clues and expert knowledge provided by pet owners, caretakers, zookeepers or specialists in animal behavior. They can provide insightful interpretations about an animal's body language by analyzing its posture and movements, as understanding and analyzing the animals' body language is one of the potential ways in which ACI researchers can interpret the animals' reactions to a system. Therefore, a tracking system capable of detecting body postures, together with expert knowledge on animal behavior, could have the potential to automatize the interpretation of an animal's responses to a digital system. Wearable systems might not be used in all scenarios; hence, non-wearable technologies would provide great advantages in this regard.

One of the main benefits of not using wearable devices would be the possibility of creating systems capable of constantly monitoring an animal's activity neither disturbing it nor modifying its habits. Complete animal tracking systems such as the one presented in this paper coupled with machine learning algorithms would allow us to learn behaviors and habits of individual animals during their usual daily life. 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.

Zookeepers could make use of these systems in order to assess the wellbeing of animals in zoos by means of detecting abnormal behavioral patterns. Automatically detecting behaviors of animals in shelters could also alert the caretakers in case that abnormal activity patterns are detected on an animal, allowing the caretakers to take rapid action to foster the animal's physical activity, dedicate more time to that animal or prioritizing adoption for cases in which the animal does not respond to those cares. In addition, this automatic behavior detection could allow to create personalized profiles for each animal. When adopting, it is very important that the personality of the animal matches the needs and personality of the human beings who are adopting it. In this way, profiles of the animals' personality could be extracted and presented to the adopters (Alcáidinho et al., 2015). Finally, pet owners could also use this kind of systems in order to detect whether their animals are doing well while left alone at home, for example during working hours, and communicate remotely with them via cameras and microphones if abnormal behaviors are detected.

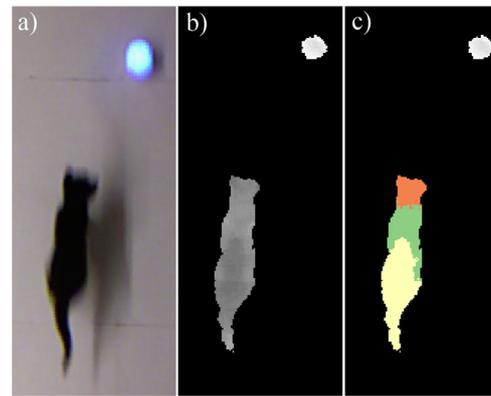


Fig. 4. Kitten chasing a Sphero®.

5.2. 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 playful environments for animals. Current digital games for animals have been usually 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 (Pons et al., 2015b). However, 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.

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 (Pons et al., 2014, 2015b). In order to keep the animals' interest in the game, 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 (Pons & Jaen, 2016b; Pons, Jaen, & Catala, 2017). Hence, being able to interpret an animal's body posture and interactions in a similar manner humans do is very important for providing proper actuations in the playful environment. A tracking system capable of detecting a cat's location, posture and orientation coupled with the detection of digital objects and their movements would allow us to create engaging and realistic games using technological artifacts which adapt to the detected cat's postures and field of view. As an example, we could think of a game in which a cat chases a Sphero® (electronic ball, see Fig. 4) controlled by the system, but the movements of the ball are not random. Instead, the Sphero® could also be tracked by the system and programmed to move depending on the detected context. Detecting the location of the cat inside the play area would allow the game to start whenever the cat approached the electronic ball. The detected cat's field of view could be used as the region on which to deploy and move the technological artifacts to attract and maintain its attention. Posture detection would allow the system to move away the toy if the tracking system detects that, for example, the cat is crouched in a hunting position, waiting for the ball to ap-

proach. The presented tracking system would have other advantages in this regard, as 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 and preferences of an animal during the game in order to adapt it.

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 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.

5.3. Tracking systems for zoo enrichment and open spaces

Depth sensors for animal tracking are a powerful tool in order to help zoos improve their current enrichment practices. Technology is being integrated into enrichment activities at zoos worldwide, as it easily allows creating varied scenarios and cognitive stimulation for the animal (French et al., 2016). In these environments, animals are not allowed to acquire and hold the technology themselves. Therefore, zoo enrichment activities based on technology would benefit from mechanisms that track and augment the animals' interactions with other elements in their environment. In this way, all the technological devices could be placed outside the enclosure while the animal interacts naturally with the augmented elements inside (Pons & Jaen, 2016a). This also leverages the need of a human caretaker being present to provide the technology. For example, current studies include the use of depth sensors to detect the interactions of orangutans with a projection on the floor (Webber et al., 2017). Tangible elements inside the environment have also been augmented using RGB-D sensors. By tracking and detecting non-technological tangible objects, the system could produce responses such as sounds based on the animals' movements with those objects (Pons, Carter, & Jaen, 2016).

However, due to current limitations of the RGB-D technology, the tracking system described in this paper is only suitable for indoor detection. 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 or wild animals in their habitats. In these scenarios, lighting conditions and required distance for optimal tracking would make the acquisition of accurate information from the depth sensor very difficult. In fact, current works for outdoor animal tracking usually rely on computer vision techniques either from recorded videos from regular video cameras (North, Hall, Roshier, & Mancini, 2015), heat-cameras (van Vonderen, 2015) or drone mounted cameras for survey population (Hodgson, Kelly, & Peel, 2013). The use of drones for posture and behavioral tracking seems promising but still presents many issues that could disturb the animals' natural behaviors: distance from the device to the animal, noise of the device, battery life, transmission rate and range, etc. Although this is an interesting area, it requires deeper discussion and consideration.

6. Conclusions and future work

A tracking system for cats based on depth information has been developed, and two studies comparing its performance using both supervised and knowledge-based approaches for the classification of cat's postures and body parts have been presented. Results have shown to be very promising and therefore tracking systems for cats based on depth information could effectively detect a cat's location and also use classification algorithms to recognize a cat's postures and body parts. Several applications have been envisioned for this kind of tracking systems and its benefits for animal welfare and wellbeing have been outlined.

Our future work would be improving the tracking and classification system, for example by introducing temporal information which could help to better differentiate between several postures such as walking or standing, and allowing to register paths and sequences of movements which are usually performed together. New postures and behaviors could also be identified, and new experiments should be conducted with cats of different sizes and breeds in order to contrast the results. In this regard, improving the knowledge-based classification process could allow to achieve fully adaptation of the system to different cats. For example, the tracking system could automatically detect the size of the cat, in case it is a kitten, and adapt the classifier to it without requiring the researcher or owner of the animal to perform manual labeling of the data to train a new classifier, which is a very time consuming task.

Some improvements could also be done regarding the detection of the cat's field of view. The orientation vector defined as a vector from the body/tail of the cat to its head sometimes gives misleading information when the cat is turning its head around. In these scenarios, the orientation vector might not be accurate when determining the cat's head orientation. Computer vision could then be used to recognize the shape of the cat's head and detect characteristic features, such as nose or ears. This would allow the detection of the head orientation and hence correctly determining where the cat is looking at. Cats have a field of view of 200°; therefore detecting the head orientation would be sufficient to provide a broad area of around 100° centered using the head orientation vector, in which to locate the technological intervention.

Another feature that will greatly benefit the tracking system in terms of flexibility and adaptation is automatic background segmentation. The tracking system presented in this manuscript performed background segmentation assuming a fixed position from the sensor to the floor, and within a clear space with neither obstacles nor furniture. However, from the initial frames of the tracking area without any animals or humans in it, background segmentation could be applied to detect the floor (i.e., deepest area in the image). Once the floor has been extracted from the image, static objects with higher depth than the floor could be detected. This would correspond to furniture, obstacles, etc. Once the area and obstacles have been mapped, any new information with different depth than the one recorded in the initial static frames would correspond to the player/s. This process would allow the tracker to adapt to several scenario configurations without minor interventions from the human, either researcher or owner/caretaker.

It should also be noted that for this study, the depth sensor was a Microsoft Kinect v1. It is known that the newer version of this sensor provides better tracking accuracy. Hence, it remains to be studied whether this new sensor would help to improve the results presented in this paper. Moreover, it could also help to quantify and reduce tracking issues caused by the physical characteristics of the cats, e.g. color, size, density of the coat, etc. In addition, it would be very interesting to evaluate the tracking system with other animal species, not only four legged ones, such as dogs, but also animals with very different skeletal characteristics such as

orangutans. The next step would be the integration of the tracking system and the learned classifiers into an intelligent playful environment, as explained in Section 5.2. In this way, the system could react properly according to the detected behavior, creating engaging playful scenarios for the animal who is playing.

The proposed tracking approach for animals based on depth information as well as the two studies on the performance of different classifiers in this domain are a significant and beneficial contribution for advancing research within and outside the field of Animal Computer Interaction. The outcomes of these two studies will allow to improve the techniques for posture and behavior recognition of animals using non-wearable devices, which will be used in the development of systems to support animal welfare and improve animal wellbeing.

Acknowledgments

This work is funded by the European Development Regional Fund (EDRF-FEDER) and supported by Spanish MINECO with Project TIN2014-60077-R. It also received support from a post-doctoral fellowship within the VALi+d Program of the Conselleria d'Educació, Cultura i Esport (Generalitat Valenciana) awarded to Alejandro Catalá (APOSTD/2013/013). The work of Patricia Pons is supported by a national grant from the Spanish MEC (FPU13/03831). Special thanks to our cat participants and their owners, and many thanks to our feline caretakers and therapists, Olga, Asier and Julia, for their valuable collaboration and their dedication to animal wellbeing.

References

- Alcaldinho, J., Valentin, G., Tai, S., Nguyen, B., Sanders, K., Jackson, M. M., et al. (2015). Leveraging mobile technology to increase the permanent adoption of shelter dogs. In *Proceedings of the 17th international conference on human-computer interaction with mobile devices and services, MobileHCI'15* (pp. 463–469). New York, NY, USA: ACM Press.
- Baskin, S., Anavi-goffer, S., & Zamansky, A. (2015). Serious games: Is your user playing or hunting? In *Proceedings of the 14th international conference on entertainment computing (ICEC 2015)* (pp. 475–481).
- Bednarik, J., & Herman, D. (2015). Human gesture recognition using top view depth data obtained from Kinect sensor. In *Proceedings of the Excel@FIT – Student conference on innovation, technology and science in IT* (pp. 1–8).
- Benko, H., Wilson, A. D., Zannier, F., & Benko, H. (2014). Dyadic projected spatial augmented reality. In *Proceedings of the 27th annual ACM symposium on user interface software and technology, UIST'14* (pp. 645–655).
- Bozkurt, A., Roberts, D. L., Sherman, B. L., Brugarolas, R., Mealin, S., Majikes, J., et al. (2014). Toward cyber-enhanced working dogs for search and rescue. *IEEE Intelligent Systems*, 29(6), 32–39. <https://doi.org/10.1109/MIS.2014.77>.
- Brugarolas, R., Loftin, R. T., Yang, P., Roberts, D. L., Sherman, B. L., & Bozkurt, A. (2013). Behavior recognition based on machine learning algorithms for a wireless canine machine interface. In *Proceedings of 2013 IEEE international conference on body sensor networks* (pp. 1–5). IEEE.
- Carter, M., Webber, S., & Sherwen, S. (2015). Naturalism and ACI: Augmenting zoo enclosures with digital technology. In *Proceedings of the 2nd international conference on animal computer interaction: Proceedings of the 2015 workshops on advances in computer entertainment conference, ACE'15 workshops*.
- French, F., Kingston-Jones, M., Schaller, D. T., Webber, S. E., Väättäjä, H., & Campbell, M. (2016). Don't cut to the chase: Hunting experiences for zoo animals and visitors. In *Proceedings of the 3rd international conference on animal-computer interaction, ACI'16* (pp. 1–6). New York, NY, USA: ACM Press.
- French, F., Mancini, C., Sharp, H., & Smith, N. (2014). Designing smart toys for the cognitive enrichment of elephants. In *Proceedings of intelligent systems for animal welfare*.
- Hirskjy-Douglas, I., Luo, H., & Read, J. C. (2014). Is My Dog Watching TV? In *Proceedings of workshop on animal-computer interaction (ACI): Pushing boundaries beyond "human," NordiCHI'14* <https://doi.org/10.13140/2.1.2608.6401>.
- Hodgson, A., Kelly, N., & Peel, D. (2013). Unmanned aerial vehicles (UAVs) for surveying marine fauna: A dugong case study. *PLoS One*, 8(11), 1–15. <https://doi.org/10.1371/journal.pone.0079556>.
- Hu, G., Reilly, D., Alnusayri, M., Swinden, B., & Gao, Q. (2014). DT-DT: Top-down human activity analysis for interactive surface applications. In *Proceedings of the 9th ACM international conference on interactive tabletops and surfaces, ITS'14* (pp. 167–176). New York, NY, USA: ACM Press.
- Jackson, M. M., Kshirsagar, Y., Starner, T., Zeagler, C., Valentin, G., Martin, A., et al. (2013). FIDO – facilitating interactions for dogs with occupations: Wearable dog – activated interfaces. In *Proceedings of the 17th annual international symposium on International symposium on wearable computers, ISWC'13* (p. 81). New York, NY, USA: ACM Press.
- Jackson, M. M., Valentin, G., Freil, L., Burkeen, L., Zeagler, C., Gilliland, S., et al. (2015). FIDO – Facilitating interactions for dogs with occupations: Wearable communication interfaces for working dogs. *Personal and Ubiquitous Computing*, 19(1), 155–173. <https://doi.org/10.1007/s00779-014-0817-9>.
- Jones, B., Benko, H., Ofek, E., & Wilson, A. D. (2013). IllumiRoom: Peripheral projected illusions for interactive experiences. In *Proceedings of the SIGCHI conference on human factors in computing systems, CHI'13* (pp. 869–878).
- Jones, B., Shapira, L., Sodhi, R., Murdock, M., Mehra, R., Benko, H., et al. (2014). RoomAlive: Magical experiences enabled by scalable, adaptive projector-camera units. In *Proceedings of the 27th annual ACM symposium on user interface software and technology, UIST'14* (pp. 637–644).
- Karlsson, J., Ren, K., & Li, H. (2010). Tracking and identification of animals for a digital zoo. In *Proceedings of 2010 IEEE/ACM international conference on green computing and communications & international conference on cyber, physical and social computing* (pp. 510–515). IEEE.
- Ladha, C., Hammerla, N., Hughes, E., Olivier, P., & Ploetz, T. (2013). Dog's life: Wearable activity recognition for dogs. In *Proceedings of the 2013 ACM international joint conference on pervasive and ubiquitous computing, UbiComp'13* (p. 415). New York, NY, USA: ACM Press.
- Lee, S. P., Cheok, A. D., James, T. K. S., Debra, G. P. L., Jie, C. W., Chuang, W., et al. (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. <https://doi.org/10.1007/s00779-005-0051-6>.
- Li, C., Goldgof, D. B., & Hall, L. O. (1993). Knowledge-based classification and tissue labeling of MR images of human brain. *IEEE Transactions on Medical Imaging*, 12(4), 740–750. <https://doi.org/10.1109/42.251125>.
- Lin, S., Liu, A., Hsu, T., & Fu, L. (2015). Representative body points on top-view depth sequences for daily activity recognition. In *Proceedings of 2015 IEEE international conference on systems, man, and cybernetics* (pp. 2968–2973).
- Mancini, C. (2011). Animal-computer interaction: A manifesto. *Magazine Interactions*, 18(4), 69–73. <https://doi.org/10.1145/1978822.1978836>.
- Mancini, C. (2013). Animal-computer interaction (ACI): Changing perspective on HCI, participation and sustainability. In *Proceedings of extended abstracts on human factors in computing systems, CHI'13* (pp. 2227–2236). New York, NY, USA: ACM Press.
- Mancini, C., van der Linden, J., Bryan, J., & Stuart, A. (2012). Exploring interspecies sensemaking: Dog tracking semiotics and multispecies ethnography. In *Proceedings of the 2012 ACM conference on ubiquitous computing, UbiComp'12* (pp. 143–152). New York, NY, USA: ACM Press.
- Moreno, A., Van Delden, R., Poppe, R., Reidsma, D., & Heylen, D. (2015). Augmenting traditional playground games to enhance game experience. In *Proceedings of the 7th international conference on intelligent technologies for interactive entertainment, INTETAIN: 17* (pp. 140–149).
- North, S., Hall, C., Roshier, A., & Mancini, C. (2015). HABIT: Horse automated behaviour identification tool – A position paper. In *Proceedings of the British Human computer interaction conference – Animal computer interaction workshop* (pp. 1–4).
- Noz, F., & An, J. (2011). Cat cat revolution: An interspecies gaming experience. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 2661–2664).
- Paldanius, M., Kärkkäinen, T., Väänänen-Vainio-Mattila, K., Juhlin, O., & Häkkinen, J. (2011). Communication technology for human-dog interaction: Exploration of dog owners' experiences and expectations. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 2641–2650). New York, NY, USA: ACM Press.
- Pons, P., Carter, M., & Jaen, J. (2016). Sound to your objects: A novel design approach to evaluate orangutans' interest in sound-based stimuli. In *Proceedings of the 3rd international conference on animal computer interaction* (p. 7:1–7:5).
- Pons, P., & Jaen, J. (2016a). Tangible user interfaces for zoo enrichment. In *CHI 2016 workshops: HCI goes to the zoo* (pp. 1–4).
- Pons, P., & Jaen, J. (2016b). Towards the creation of interspecies digital games: An observational study on cats' interest in interactive technologies. In *Proceedings of the 2016 CHI conference extended abstracts on human factors in computing systems, CHI EA'16* (pp. 1737–1743). New York, NY, USA: ACM Press.
- Pons, P., Jaen, J., & Catala, A. (2014). Animal ludens: Building intelligent playful environments for animals. In *Proceedings of the 2014 workshops on advances in computer entertainment conference, ACE'14 workshops* (pp. 1–6). New York, NY, USA: ACM Press.
- Pons, P., Jaen, J., & Catala, A. (2015a). Developing a depth-based tracking system for interactive playful environments with animals. In *Proceedings of 2nd international conference on animal computer interaction: Proceedings of the 2015 workshops on advances in computer entertainment conference, ACE'15 workshops*.
- Pons, P., Jaen, J., & Catala, A. (2015b). Envisioning future playful interactive environments for animals. In A. Nijholt (Ed.), *More playful user interfaces* (pp. 121–150). Springer.
- Pons, P., Jaen, J., & Catala, A. (2017). Towards Future Interactive Intelligent Systems for Animals: Study and Recognition of Embodied Interactions. In *Proceedings of the 22nd international conference on intelligent user interfaces, IUI'17* (pp. 389–400). New York, NY, USA: ACM Press.
- Ribeiro, C., Ferworn, A., Denko, M., & Tran, J. (2009). Canine pose estimation: A computing for public safety solution. In *Proceedings of 2009 Canadian conference on computer and robot vision* (pp. 37–44). IEEE.
- Trindade, R., Sousa, M., Hart, C., Vieira, N., Rodrigues, R., & França, J. (2015). Purrfect crime. In *Proceedings of the 33rd annual ACM conference extended abstracts on human factors in computing systems, CHI EA'15* (pp. 93–96).
- Väättäjä, H., & Pesonen, E. (2013). Ethical issues and guidelines when conducting

- HCI studies with animals. In *Proceedings of extended abstracts on human factors in computing systems, CHI'13* (pp. 2159–2168).
- Valentin, G. (2014). Gestural activity recognition for canine-human communication. In *Proceedings of the 2014 ACM international symposium on wearable computers adjunct program, ISWC'14 adjunct* (pp. 145–149). New York, NY, USA: ACM Press.
- Valentin, G., Howard, A., & Jackson, M. M. (2015). Towards a canine-human communication system based on head gestures. In *Proceedings of 2nd international conference on animal computer interaction: Proceedings of the 2015 workshops on advances in computer entertainment conference, ACE'15 workshops*.
- van Vonderen, J.. *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-quit/6256558>.
- Watanabe, S., Izawa, M., Kato, A., Ropert-Coudert, Y., & Naito, Y. (2005). A new technique for monitoring the detailed behaviour of terrestrial animals: A case study with the domestic cat. *Applied Animal Behaviour Science*, 94(1–2), 117–131. <https://doi.org/10.1016/j.applanim.2005.01.010>.
- Webber, S., Carter, M., Sherwen, S., Smith, W., Joukhadar, Z., & Vetere, F. (2017). Kinecting with orangutans: Zoo visitors' empathetic responses to animals' use of interactive technology. In *Proceedings of the 2017 CHI conference on human factors in computing systems, CHI'17* (pp. 6075–6088). New York, NY, USA: ACM Press.
- Weilenmann, A., & Juhlin, O. (2011). Understanding people and animals: The use of a positioning system in ordinary human-canine interaction. In *Proceedings of the 2011 annual conference on Human factors in computing systems, CHI'11* (pp. 2631–2640). New York, NY, USA: ACM Press.
- Weiss, G. M., Nathan, A., Kropp, J. B., & Lockhart, J. W. (2013). WagTag: A dog collar accessory for monitoring canine activity levels. In *Proceedings of the 2013 ACM conference on pervasive and ubiquitous computing adjunct publication* (pp. 405–414).
- Westerlaken, M., & Gualeni, S. (2014). Felino: The philosophical practice of making an interspecies videogame. In *Proceedings of the philosophy of computer games conference* (pp. 1–12). Retrieved from.
- Wingrave, C. A., Rose, J., Langston, T., & LaViola, J. J. J. (2010). Early explorations of CAT: Canine amusement and training. In *Proceedings of extended abstracts on human factors in computing systems, CHI'10* (pp. 2661–2669).
- Winters, M., Mealin, S., Brugarolas, R., Yuschak, S., Majikes, J., Sherman, B. L., et al. (2015). Knowledge engineering for unsupervised canine posture detection from IMU Data. In *Proceedings of the 2nd international conference on animal computer interaction: Proceedings of the 2015 workshops on advances in computer entertainment conference, ACE'15 workshops*.
- Yonezawa, K., Miyaki, T., & Rekimoto, J. (2009). Cat@Log: Sensing device attachable to pet cats for supporting human–pet interaction. In *Proceedings of the international conference on advances in computer entertainment technology, ACE'09* (pp. 149–156). New York, NY, USA: ACM Press.
- Zeagler, C., Byrne, C., Valentin, G., Freil, L., Kidder, E., Crouch, J., et al. (2016). Search and rescue: Dog and handler collaboration through wearable and mobile interfaces. In *Proceedings of the third international conference on animal–computer interaction, ACI'16* (pp. 1–9). New York, NY, USA: ACM Press.