

Robust Point Cloud Segmentation of Rodents using Close Range Depth Cameras in Controlled Environments

Oscar Fernandez¹, Elsbeth A. van Dam², Lucas P.J.J. Noldus³, Remco Veltkamp⁴
Noldus Information Technology BV (Wageningen, The Netherlands)^{1,2,3}
Utrecht University (The Netherlands)⁴

Abstract

Rodent behavior analysis is an extremely important task for pre-clinical testing of new drugs and neurodegenerative diseases. Some of the underlying mechanisms regulating natural interactions among multiple animals require long term interaction, which implies the usage of automated, objective and systematic video analysis systems. However, current video analysis systems are limited by idiosyncratic features of the 2D video technology, and requiring a great number of parameters. This early stage work focuses on the first step of a depth-based tracking system, offering some methods to robustly perform rodent segmentation in a controlled environment, fully exploiting its geometrical properties, providing a qualitative overview of the results attained so far.

1 Introduction

The use of rodents in controlled environments has proven to be extremely useful for pre-clinical trials, since it allows the systematic repetition of a test environment to assess an initial hypothesis. Fully automated video analysis allows a more objective and quicker behavioral analysis compared to visual inspection [2]. Some systems are commercially available, like the Ethovision XT developed by Noldus Information Technology; and Viewpoint VideoTrack, using top-view video analysis as the underlying base of the system. One of the crucial steps of these types of systems consists of the rodents blob extraction, providing the precise region corresponding to the rodent. Although this might seem a simple task, it is prone to a substantial level of error. Due to the intrinsic features of video cameras, the methods developed upon them are very sensitive to illumination changes, and require a high color contrast between the background and foreground (generally implying the usage of dark cage floors and white

animals, or vice-versa). These limitations are magnified at later stages of the behavioral analysis pipeline, since the positions and pose of the animals in the scene are limited to a 2D horizontal plane. Moreover, given that in common scenarios, rodents often have the same fur color, the correct differentiation of the animals is extremely difficult during close contact, when using solely the overlapping 2D silhouette.

The introduction of affordable indoor consumer depth camera sensors like the Microsoft Kinect 2¹ or the Creative 3DSenZ², have spawned a wide range of novel applications in problem domains that previously were not easily tackled (e.g human pose inference; scene understanding; hand gesture recognition) Their core underlying technique is based on the *Time-of-flight* principle [7], which illuminates the scene with a modulated light source, observing the reflected light thereafter. From the phase shift between illumination and reflection, the distances are calculated for every pixel, and the depth map is extracted. The cameras use a near-infrared LED light ($\approx 850\text{nm}$) which is invisible to human's and rodent's eyes. These cameras have some important properties namely, robustness to illumination changes and the ability to operate in completely dark environments. This makes them very interesting for animal behavior analysis, in particular for rodents, in confined environments. This paper presents the first known contribution, to our best knowledge, that uses this particular type of camera from a top-view perspective, in this stage focusing solely on the segmentation problem, which is one of the core steps of a tracking system.

This work is structured as follows: Section 2 analyzes some of the related work in this area. Section 3 shows the data acquisition and techniques used. Section 4 discusses some of the results obtained. Finally Section 5 explores some of the current limitations and

¹Kinect2: <http://www.microsoft.com/en-us/kinectforwindows/>

²3dSenZ: <http://us.creative.com/p/web-cameras/creative-senz3d>

offers some lines of development.

2 Related Work

Computerized video analysis provides a valuable technique to detect and extract various types behaviors in a much quicker and more systematic fashion than behavioral analysis based on visual inspection. However, common techniques based in standard video analysis present considerable drawbacks inherited by the intrinsic limitations of the data representation attained in those cameras. More specifically, in the rodents case study, when analyzing social behavior among multiple individuals, the most common and difficult problem are occlusions that might occur during those interactions. This clearly hampers the understanding of the animal's pose and the understanding of complex interaction behavior [1, 5, 11].

In [8] some of these limitations are tackled by getting a comprehensive physical representation of the animals, using 3D information of the scene instead of the classical 2D information provided by the standard video devices. This is achieved by using depth sensor technology, in this case the MS Kinect, that provides 3D information by measuring distances from the device towards the scene it is capturing, by measuring the deformations that the interest object causes over a projected infrared pattern. However, the device only provides partial 3D information of the captured objects, since it is perspective dependent. So, to get a full object reconstruction of the rats, the authors have used a set of 4 Kinect cameras surrounding the cage, obtaining a full 3D hull representation of the rats. Also, this layout diminishes the possible occlusions on the scene. The complexity of this setup required in order to build a closed hull - the base of the tracking system - makes it difficult to apply in a real application.

Other approaches [6, 12] also considered the usage of 3D information. In the first, the 3D information was achieved by using multiple cameras to achieve voxel reconstruction, an approach with a considerable computational overhead, and with a low level of spacial resolution. In the latter, 3D information was obtained using depth sensor devices, but the degree of information extracted was rather limited, so it could only infer simple behaviors from a single rodent.

Although standard video analysis techniques still play a major role in commercial and academic behavioral analysis systems, it is clear that a new trend, based on depth sensor technology is currently emerging.

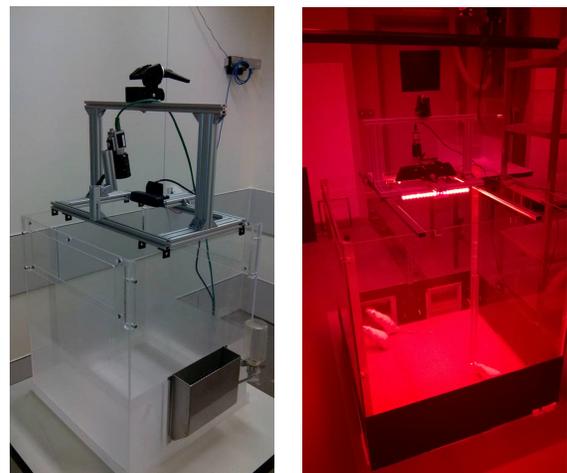
3. Material and Methods

3.1 Overview

Our system has the goal to provide reliable and accurate tracking positional data, and serve as the core base for a richer and robust behavioral analysis system, that can overcome some of current limitations of 2D video technology. We propose a new approach that can take the best from the new depth camera technologies while preserving the top-view hardware setup used by some commercial applications. This way, future developments are not totally disruptive and allow a feasible application to real scenarios it offers the best trade-off between representative quality data where the cage's physical geometric properties can be used to leverage a robust rodent segmentation.

3.2 3D Data Acquisition

In order to develop this new techniques, a novel rodent depth dataset had to be recorded since there is none in the public domain. Therefore, a recording framework had to be designed both in hardware and software in order to collect a representative sample of rats in controlled environments (see Figure 1). For the dataset acquisition the Creative 3D SenZ camera was used, alongside with the Intel Perceptual Computing framework³ and the Point Cloud Library [10] for storage and data analysis.



(a) Top-unit on Noldus Phentyper 4500.

(b) Top-unit on Noldus Phentyper 9000.

Figure 1. Data acquisition setup.

³<https://software.intel.com/en-us/vcsourc/tools/perceptual-computing-sdk>

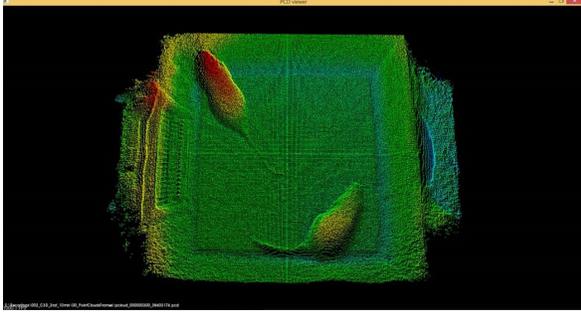


Figure 2. Point cloud sample extracted using Creative 3D Senz depth camera on the PT4500 from a top-view perspective.

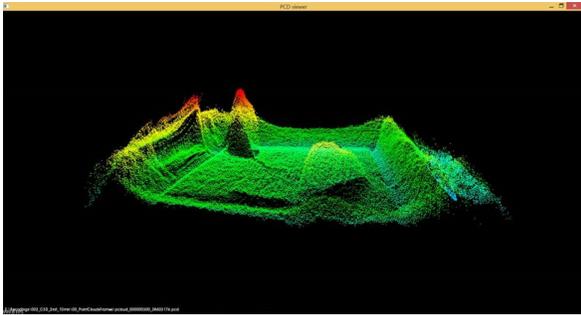


Figure 3. Point cloud sample extracted using Creative 3D Senz depth camera on the PT4500 from a top-view perspective and then rotated to a profile view.

3.3 RANSAC Overview

During the experiment, the rodents are recorded in a controlled environment, more precisely within a Phenotyper cage (PT4500 and PT9000) ⁴ cage, with a known geometry. This prior knowledge can be used to exploit the spatial information provided by the depth camera.

In this cage setup it is known that the scene is composed of a planar floor and walls. This way, knowing the plane model coefficients, all the cloud points that do not fall within the considered plane boundaries are considered outliers to the model. The points non-compliant to the planar geometric primitive are considered belonging to the rodents point cloud data (in our setup, no shelters were considered).

In order to perform this model plane estimation, the algorithm must be robust to a significant proportion of outliers within the unorganized cloud data.

In the literature, traditional regression methods [3, 9]

⁴<http://www.noldus.com/phenotyper> - Noldus

are not robust to the presence of data outliers or fail in the presence of multiple models. These limitations are surpassed by the *RANdom Sample Consensus* (RANSAC) algorithm [4]. This algorithm consists of general purpose parameter estimation, specially designed to cope with a large proportion of outliers. RANSAC is a sampling techniques that generates candidate solutions by randomly selecting a minimum number of observations (data points). Unlike standard sampling techniques using as much of the data as possible to obtain an initial solution and then proceed to prune outliers, RANSAC considers the smallest sampling set possible and then proceeds to enlarge this set with model consistent data points.

3.4 Rodent Blob Extraction

After applying the RANSAC algorithm to the point cloud on the previous step, the plane model coefficients are obtained, and by this the cloud points belonging to floor plane (inliers) can be selected. All the remaining model outliers, represent point data that do not belong to the plane and must belong to the rodents. With this method we get the rodents blob areas by fully exploring the capabilities provided by the depth sensors.

4 Results

In Figures 4 and 5 we demonstrate graphically the results obtained with our method - on the left the full original point cloud; on the right, the resulting rodent segmentation. Initially it is applied an outlier removal filter, followed by a *Pass through* filter, to delimit our 3D region of interest, followed by the RANSAC method that provides us the plane model (red colored) and the extracted rodents blobs (green colored). The obtained result show the quality of the method in dealing with outliers when computing the plane's equation, which is a crucial step to obtain a reliable rodent segmentation.

5 Concluding Remarks

In this paper we present current and ongoing work towards a rodent tracking system for controlled cage environments, capable of fully exploiting the capabilities of affordable depth cameras it performs a data stream analysis that uses the static geometric properties of the environment as a hint for a robust blob detection.

Some limitations were identified, namely the usage of small rodents with black fur reduce the precision of the data acquired by the depth camera, since most of the IR light projected by the camera is absorbed in this

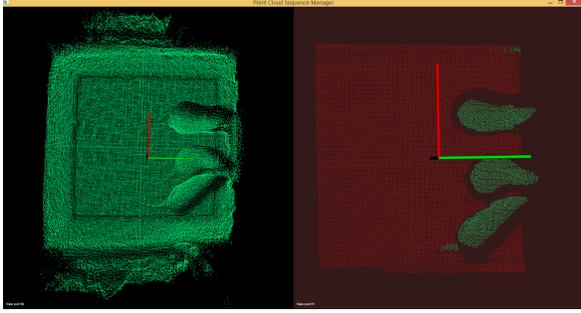


Figure 4. Left: Original point cloud data. Right: Plane(red) and rodents(green) segmentation.

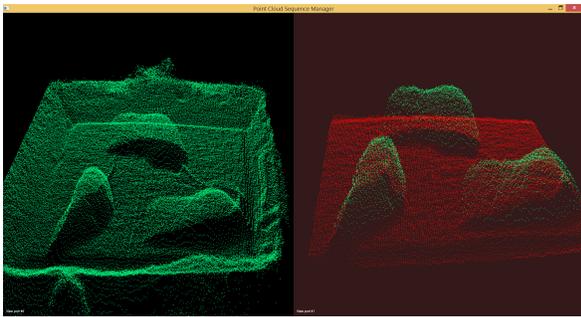


Figure 5. Rotated point cloud.

case. Moreover, a better data precision was achieved using white (non-reflective) floor in the cage, which renders this setup unusable in conjunction with common video analysis systems, since they rely on high contrast between the animal and the floor color.

In future work we will focus on the most difficult scenarios such as animal overlapping in complex social interaction. This case is remarkably difficult in 2D video analysis setups since both animals are extremely similar both in terms of texture and color. Considering our point cloud data, this task can be simplified by searching the discontinuities created along the body overlaps, both in terms of depth and surface normals. This would be an intermediate step for a more advanced multiple rodent tracking system.

Nevertheless, this affordable technology enables the system to harness the potential provided by three dimensional scene data. This aspect combined with the well known physical layout constraints, provides invaluable prior knowledge that could and should be used to the system's advantage.

Acknowledgments

This work was possible partially due to the support of *Delta Phenomics* that allowed the data recording on their labs. In particular, I would like to thank the time and support provided by Jan van de Kieft and Elisavet Kiryakou throughout the sessions.

This research is financed by the European Commission with the Marie Curie EID project PhenoRat (No. 317259).

References

- [1] K. Branson and S. Belongie. Tracking multiple mouse contours (without too many samples). In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 1039–1046 vol. 1, June 2005.
- [2] F. de Chaumont, R. D.-S. Coura, P. Serreau, A. Crescent, J. Chabout, S. Granon, and J.-C. Olivo-Marín. Computerized video analysis of social interactions in mice. *Nature Methods*, 9:410–417, 2012.
- [3] P. de Groen. An introduction to total least squares. *Nieuw Archief voor Wiskunde*, 1:237–253 vol. 14, 1996.
- [4] M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395, June 1981.
- [5] L. Giancardo, D. Sona, H. Huang, S. Sannino, F. Manag, and et al. Automatic visual tracking and social behaviour analysis with multiple mice. *PLoS ONE* 8(9): e74557, 2013.
- [6] P.-L. Lai, D. M. Basso, L. C. Fisher, and A. Sheets. 3d tracking of mouse locomotion using shape-from-silhouette techniques.
- [7] L. Li. Time-of-flight camera - an introduction. *Texas Instruments - Technical White Paper*, 2014.
- [8] J. Matsumoto, S. Urakawa, Y. Takamura, R. Malcher-Lopes, E. Hori, C. Tomaz, T. Ono, and H. Nishijo. A 3d-video-based computerized analysis of social and sexual interactions in rats. *PLoS One*, 8(10):e78460, 2013.
- [9] P. J. Rousseeuw. Least median of squares regression. *Journal of the American Statistical Association*, 1984.
- [10] R. Rusu and S. Cousins. 3d is here: Point cloud library (pcl). In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 1–4, May 2011.
- [11] A. Weissbrod, A. Shapiro, G. Vasserman, L. Edry, M. Dayan, A. Yitzhaky, L. Hertzberg, O. Feinerman, and T. Kimchi. Automated long-term tracking and social behavioural phenotyping of animal colonies within a semi-natural environment. *Nature communications* 4, 2013.
- [12] T.-H. O. Yang, M.-L. Tsai, C.-T. Yen, and T.-T. Lin. An infrared range camera-based approach for three-dimensional locomotion tracking and pose reconstruction in a rodent. *J Neurosci Methods*, 2011.