Multiple Mice Tracking and Segmentation through SIFT Flow Analysis

Ario Sadafi[1], Vasiliki-Maria Katsageorgiou[1], Huiping Huang[2], Francesco Papaleo[2], Vittorio Murino[1,3], Diego Sona[1,4]

[1]Pattern Analysis and Computer Vision (PAVIS), Istituto Italiano di Tecnologia, Genova, Italy.
[2]Neuroscience and Brain Technologies (NBT), Istituto Italiano di Tecnologia, Genova, Italy.
[3]Department of Computer Science, University of Verona, Verona, Italy.
[4]Neuroinformatics Lab. (NILab), Fondazione Bruno Kessler, Trento, Italy.
Contact: name.surname@iit.it

***Abstract*—Automatic analysis of high level social interactions between multiple animals in lab environment is of important scientific interest in neurological studies including behavioral effects of pharmaceutical treatments. A robust animal tracking system is therefore an important tool in these studies to make behavior analysis easier to scientists. In this paper, we propose a tracking algorithm which is based on fast morphological operations to segment mice occlusions. Verification and correction of tracking in cases of occlusions is done using SIFT Flow analysis. Like in any other multiple target tracking system, identity preservation and minimization of identity swaps is the main goal. We performed experiments with multiple videos and thanks to the SIFT flow we were able to achieve fewer identity swaps than a state-of-the-art system.**

I. Introduction

The relation between genes and behavior is of important scientific interest. Mice exhibit a great number of social interactions and their low cost allows biologists to create various animal models characterized by specific neurological diseases. Thus, genetically modified mice are widely used nowadays, becoming a unique tool for understanding the links between genes and behavior. Some complex social behaviors span for extended periods and manual investigation of these specific behavioral parameters remains an experimental bottleneck.

Indeed, monitoring and manually scoring video recordings of usually short experiments (5-10 minutes) is very time consuming. Moreover, in order to make statistical analysis more robust and to better understand certain phenomena, long lasting experiments are required, which are, however, infeasible when multiple interacting animals are involved in an experiment. Hence, there is an increasing interest in the development of automated systems for rodents' behavior analysis from videos.

Tracking algorithms are the primary tool for monitoring the behavior in mice. Mice tracking is a uniquely difficult task in computer vision due to the fact that mice are mostly identical and highly deformable objects with erratic motions. In addition to this, an important factor in behavioral studies is the social interaction between individuals making the tracking even more complicated since multiple mice have to be detected along a video recording and occlusions between them are highly possible.

Various systems have been proposed in literature in order to make the monitoring of mice easier to scientists. Most of them are based on 2D data coming from a top view camera and a tracking algorithm, which usually relies on Markov random fields or Bayesian approaches. For example, Branson et al. [10] proposed a multiple deformable object tracker based on particle filtering, combining a multiple blob tracker and a contour tracker. In another, relatively older work, Kervrann et al. [11] used a hierarchical Markov modeling approach to track deformable shapes. Off the shelf software products, such as EthoVision [12], are also helpful to behavioral scientists. The limitation of such software products is that animals must have different fur color. Hence, when animals are identical, the only possible way to distinguish their identity is to use color labeling on the animals, which might affect their behavior. An example of a label-less tracker, was proposed by Giandcardo et al. [9]. In this work, a hybrid approach was presented, where expectation maximization was used to extend watershed into multiple frames and tracking was based on a distance metric. This approach appears to be robust in short partial occlusions, but it is error-prone in longer or more complicated occlusions.

In this paper we propose a novel tracking algorithm based on a pipeline including Kalman filter to cope with noise and SIFT flow analysis [2] to manage mice occlusions and extrapolate their trajectory. In our work, we are not using any kind of labeling, like coloring the mice or implanted RFID chips that would make mice-tracking easier, to prevent any unwanted effect on their behavior. Moreover, we use thermal camera to leave the mice in the dark, which is the condition during which mice produce increased activity.

The rest of the paper is organized as follows: The description of the employed method is given in Section II. Experimental results are given in Section III, followed by conclusion remarks in Section IV.

II. Methods

The proposed algorithm consists of five main units. Background subtraction for the identification of the possible mice blobs, tracking based on Kalman filter, segmentation of mice occlusions and identity preservation. Finally, mice position coordinates (including head, tail and the centroid of the body) are returned for every frame to be used for behavior analysis.

A. Background Subtraction & Blob Detection

Initial step in tracking applications is the separation of the object of interest from the background pixels. In our case, background includes the arena, bedding, possible eliminated excretory products and thermal reflections of the mice on the arena panels. Since the temperature of the mice differs significantly from the one of the arena and all other confounds, background subtraction based on thresholding is performed. To find the best threshold, some sample frames randomly distributed over the video are selected. Starting from the lowest temperature, threshold is gradually increased until the temperature that is high enough to correctly separate the mice from the arena is reached. Since the number of expected mice is already known, by performing blob analysis on the frame a thresholding temperature that is high enough to avoid any noise or reflections while it still preserves pixels of mice bodies well is found.This threshold is then fixed for all the frames of the experiment. After thresholding, pixels intensity is normalized in $[0,1]$ for a higher dynamic range:

$$P_{i,j} = \begin{cases} \frac{P_{i,j}-\tau}{max(P_{i,j})-\tau} & \text{if} \quad P_{i,j} > \tau \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

where τ is the computed threshold value.

Separation of blobs when there are no occlusions is easy. Morphological operations, like image opening and image closing with different structure element sizes, are performed to obtain smoother and better shaped blobs. In particular, 3×3 and 15×15 pixel rectangular elements are used for opening and closing operations, respectively. In order to make sure that the blobs do not contain any hole, image filling method [7] is used.

B. Kalman-based Tracking

The assumption that the trajectory of a moving object is constrained in a known mathematical model can facilitate its tracking. Kalman filter [14] is a smooth, dynamic model widely used in such applications. In our case, in order to have a smoother path and a good estimation of trajectories, mice are assumed to be moving with constant acceleration. The trajectory of each mouse is predicted by a Kalman filter with constant acceleration method using the previous mouse locations.

As a next step, the predicted positions of each mouse are used to compute a cost matrix on the basis of the distance between the projected position of each mouse and each detected blob.

$$\forall m \in M, \forall d \in D \quad cost(m,d) = \sqrt{||d-m||_2} \tag{2}$$

Where M is the set of all mice (tracks) and D is the set of the detections in current frame. $Cost$ is a matrix holding the euclidean distance of all detections to all tracks. Munkres assignment algorithm [8] is then used to assign the new detections to the tracks and the Kalman filter is corrected according to the detection. Finally, Kalman parameters are fine tuned to fit to the mice normal movements, which are usually observed in the experiment.

In the proposed tracking algorithm, mice identities in case of occlusion do not rely on the assignment provided by Kalman, but correct tracking of mice identities is verified by SIFT flow, as will be discussed in the next section.

C. Occlusion Segmentation

Accurate mice position and their body description is vital for behavior analysis. When mice are moving individually in the arena, detection is flawless and enough blobs exist, each representing a mouse body. When the mice are in contact instead, thresholding based background subtraction cannot separate them properly and a large occluded blob is returned. In such a case, segmentation methods are helpful to separate the occluded mice bodies, such that Kalman filter can be updated accordingly.

The proposed segmentation method is based on mathematical morphology on grayscale images. The occluded image is first eroded such that mice blobs are shrank. Remaining pixels are used as the mask for morphological grayscale image reconstruction, which returns the union of connected components of the original image [1]. Next, the reconstructed image and its dilation are complemented to be used for another operation of image reconstruction. This reconstruction prevents a lot of noise in the segmentation results. Finally, the blobs are extracted from the obtained image. In this way, the occluded blob is segmented and newly generated detections can be used by Kalman filter.

D. Identity Preservation

The most important part of multiple-animal tracking is preserving their correct identity throughout the whole experiment. The segmentation method presented in the previous section can handle occlusions in almost all cases but the distance based assignment of the tracks to detections is error-prone to mice swap and is highly probable that the mice identities are lost. Tracks can easily swap if detections are allocated based on euclidean distance alone when mice are too close. In the proposed pipeline, when an occlusion is detected, an identity preservation module starts storing sequence of the frames until the occlusion is over. Using this occlusion subsequence frames, correct allocation of mouse identities is verified ensuring that there will be no identity swap due to poorly segmented frames. Identity preservation is performed according to the following steps:

1) Occlusion Bounding Box: A bounding box is defined such that it holds the occluded part completely for all of the subsequence frames. The resulting bounding box is simply the smallest holding box that contains the union of all the bounding boxes detected for every frame in the occluded part. For better results with SIFT analysis, the resulting bounding

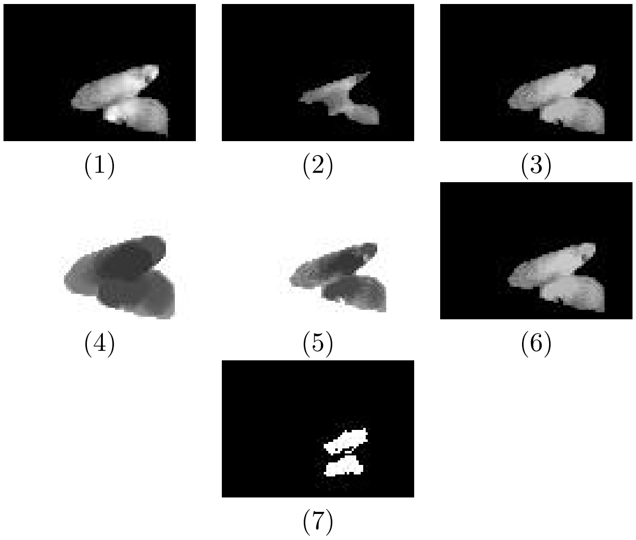

Fig. 1. Snapshots demonstrating the steps of segmentation. From left to right, top to bottom: 1. Original Image 2. Eroded image 3. Reconstruction of the image using the eroded image as mask 4. Image complement of the dilation of the original image 5. Image complement of previous reconstruction 6. Complement of the image obtained after image reconstruction using the complements of dilation and first reconstruction 7. Obtained blobs

box is enlarged by 15 pixels on each side to have safe margins around the occlusions.

2) SIFT Flow Analysis: SIFT flow is inspired from optical flow algorithm that produces pixel to pixel dense correspondences between two images by matching the SIFT descriptors extracted from each pixel by the "coarse to fine" algorithm [2].

Using all frames of the occlusion subsequence and calculating sift flow for each two of them would be too expensive computationally. We observed, on the contrary, that first and middle frames were the most informative frames of every occlusion subsequence. The two frames are, therefore, cropped using the occlusion bounding box calculated as described above. The SIFT descriptor is computed on the occlusion in these two frames and is passed to the "coarse to fine" algorithm for the SIFT flow analysis to obtain the flow matrix *SF*. To analyze the occlusion we compute the flow for each mouse m identified in the last frame before the occlusion begin as follows:

$$\forall m \in Mice,\ \forall i,j \qquad D^m_{i,j} = \begin{cases} 1 & \text{if pixels } i,j \in Mice_m \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

$$\forall m \in Mice, \quad V^x_m = |SF^x \circ D^m| \qquad V^y_m = |SF^y \circ D^m| \tag{4}$$

where V is the vector of average flow for each mouse. Then, projection of the position of centroids is computed according to:

$$\forall m \in Mice : P_m = Centroid_m + \kappa \times V_m \tag{5}$$

A coefficient κ is defined to increase the impact of flow vector as it normally has small values. In our experiments we empirically set κ to 15 for the best results.

3) Verification of centroids: After updating the centroid positions and obtaining projections, a deviation matrix, which is the cost matrix based on euclidean distance defined by equation 2, is computed between the projected centroids and tracked centroid positions. This matrix is used as the allocation cost matrix for Munkres assignment algorithm, which allocates each mouse centroid to its closest projected centroid. If any swap is detected the mice representations are swapped to correct the error. After the correction, stored subsequence is released.

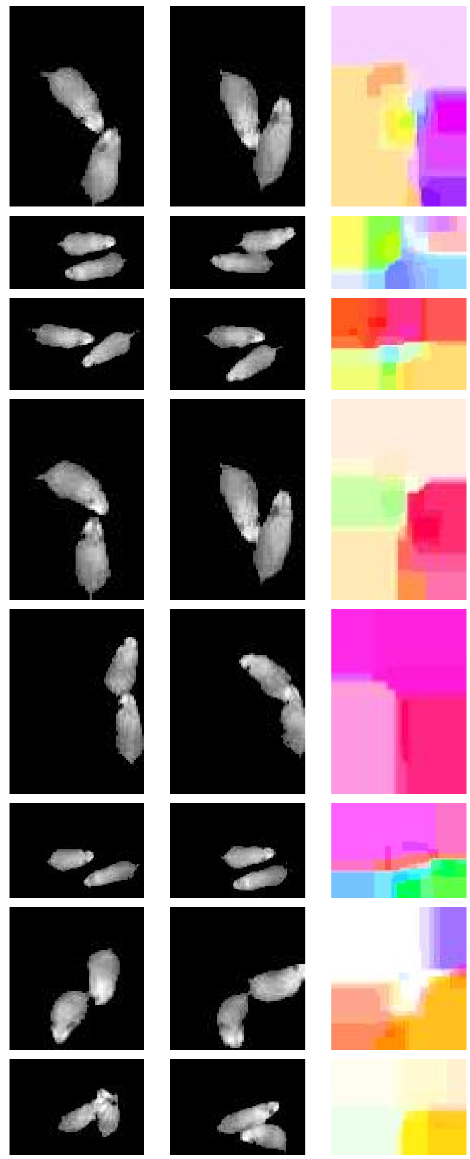

Fig. 2. Some of the occlusions and their corresponding detected flow. The left and middle column are the frames used in sift flow analysis. The third column is the color demonstration of the computed flow vectors.

Some examples of occlusions are demonstrated in Fig 2 where We used color coding introduced in [13] which is originally used in [2] to show obtained flow vectors. When both mice are moving in the same direction (as in the fifth example from top in Fig. 2) the flow for both of them is towards the same direction as well. Although two mice receive SIFT flow vectors not different much in terms of direction and magnitude and their centroids are projected by almost equal vectors, their relative distance is preserved so that it would not lead to any problem.

III. Experiments & Results

To evaluate the performance of the proposed tracker, we used six different video recordings (each lasting about one hour) with two mice interacting in an arena. The tracking results of the tracker proposed in [3] manually corrected by behavioral scientists, were used as ground truth to evaluate the performance of our tracker.

A. Experimental Setup

Experiments were carried out in an $50cm \times 50cm$ open arena with a thermal camera mounted about $1.5m$ above it. Mice were free to move and interact in the arena while their interactions were recorded with a FLIR A315 thermal camera having a resolution of 320×240 at the rate of 30 frames per second. The recordings were carried out in almost complete darkness since this is less stressful for the animals and they are in their most active phase. Animals were recorded without any tagging to prevent any alteration to their behavior and interactions.

B. Results

We compared our method to the one previously developed by Giancardo et. al.[3]. We used both trackers on the same datasets and we evaluated their results using the ground truth, counting the number of mice-inversions observed in each tracking.

TABLE I
Tracking Results and comparison with the tracker of Giancardo et al. [3]. Total number of identity swaps is reported for each experiment and each tracker.

	Video Length	Giancardo et al	Sift Flow
Exp. 1	1:00:54	27	17
Exp. 2	1:09:07	45	23
Exp. 3	1:11:45	54	29
Exp. 4	1:15:32	49	21
Exp. 5	1:09:11	75	28
Exp. 6	1:05:39	14	8

The algorithm was implemented in Matlab® 2016 running almost in real time on an Intel® Xeon® core at 3.20 GHz.

The overall 6 hours 52 minutes 8 seconds of video was tested on both trackers. The state of the art tracker had a total of 264 swaps compared to 126 swaps for the proposed one, which is about 50% less. On average, using the state of the art tracker we observe a swap once in every 94 seconds, while with the proposed tracker only once in every 196 seconds.

IV. Conclusion

Robust mice tracking is a challenging problem considering that all mice look the same in shape and color. They tend to touch each other or even pass above others' body Reliable automated mice tracking enables scientists to have experiments that can last for days in order to have better insights about the hidden unknown behavior patterns of mice that linger days to exhibit and may not be observable in short term experiments. Thermal cameras are easily available in the market and are getting more affordable which is expected to be widely used in different experiments. This algorithm works with thermal data of the camera but as the algorithm is modular it can be easily extended to be used with different cameras with small changes in image acquisition section or for different animals with small changes in detection part.

References

[1] Vincent L. Morphological grayscale reconstruction in image analysis: applications and efficient algorithms. IEEE transactions on image processing. 1993 Apr;2(2):176-201.

[2] Liu C, Yuen J, Torralba A, Sivic J, Freeman WT. Sift flow: Dense correspondence across different scenes. InEuropean conference on computer vision 2008 Oct 12 (pp. 28-42). Springer Berlin Heidelberg.

[3] Giancardo L, Sona D, Huang H, Sannino S, Managò F, Scheggia D, Papaleo F, Murino V. Automatic visual tracking and social behaviour analysis with multiple mice. PloS one. 2013 Sep 16;8(9):e74557.

[4] Peters SM, Pinter IJ, Pothuizen HH, de Heer RC, van der Harst JE, Spruijt BM. Novel approach to automatically classify rat social behavior using a video tracking system. Journal of neuroscience methods. 2016 Aug 1;268:163-70.

[5] Hong W, Kennedy A, Burgos-Artizzu XP, Zelikowsky M, Navonne SG, Perona P, Anderson DJ. Automated measurement of mouse social behaviors using depth sensing, video tracking, and machine learning. Proceedings of the National Academy of Sciences. 2015 Sep 22;112(38):E5351-60.

[6] Goto T, Tomonaga S, Okayama T, Toyoda A. Murine depression model and its potential applications for discovering foods and farm products with antidepressant-like effects. Frontiers in neuroscience. 2016;10.

[7] Soille P. Morphological image analysis: principles and applications. Springer Science & Business Media; 2013 Mar 14.

[8] Munkres J. Algorithms for the assignment and transportation problems. Journal of the society for industrial and applied mathematics. 1957 Mar;5(1):32-8.

[9] Giancardo L, Sona D, Scheggia D, Papaleo F, Murino V. Segmentation and tracking of multiple interacting mice by temperature and shape information. InPattern Recognition (ICPR), 2012 21st International Conference on 2012 Nov 11 (pp. 2520-2523). IEEE.

[10] Branson K, Belongie S. Tracking multiple mouse contours (without too many samples). In2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) 2005 Jun 20 (Vol. 1, pp. 1039-1046). IEEE.

[11] Kervrann C, Heitz F. A hierarchical Markov modeling approach for the segmentation and tracking of deformable shapes. Graphical Models and Image Processing. 1998 May 31;60(3):173-95.

[12] Noldus LP, Spink AJ, Tegelenbosch RA. EthoVision: a versatile video tracking system for automation of behavioral experiments. Behavior Research Methods, Instruments, & Computers. 2001 Aug 1;33(3):398-414.

[13] Baker S, Scharstein D, Lewis JP, Roth S, Black MJ, Szeliski R. A database and evaluation methodology for optical flow. International Journal of Computer Vision. 2011 Mar 1;92(1):1-31.

[14] Kalman RE. A new approach to linear filtering and prediction problems. Journal of basic Engineering. 1960 Mar 1;82(1):35-45.