

A Flexible Object Tracking System for Planary Motion

Qinghai Liao, Wencong Zhang, Peng Shi, Ming Liu

Department of Mechanical and Biomedical Engineering

City University of Hong Kong

Email: qinghai.liao@my.cityu.edu.hk [mingliu,pengshi]@cityu.edu.hk

Abstract—Object visual tracking and servo is a challenging research topic in the fields of computer vision, pattern recognition and robotics. Usually, people just utilize PTZ for active surveillance tracking. In this work, we combine visual tracking and visual servo to drive a PTZ to cast laser beam on the target object. The demonstrated tracking system is friendly-using and accurate which can be used for object tracking in plane, such as stage lighting effect, video surveillance, etc. Our proposed tracking system features arbitrarily installation, automatic parameters calibration, high efficiency and low coupling. From the point of mathematical modeling, our model outputs exact solution without any accuracy loss.

I. INTRODUCTION

Visual tracking or monitoring is a challenging research issue that involves computer vision, pattern recognition, and robotics. Visual tracking and monitoring have attracted a large amount of attention in the literature [1] and there are also lots of available algorithms or applications [2], [3]. However, the current tracking system is not able to satisfy our request—not only to track the object in images but to take reaction on the target. In this paper, our proposed tracking system is able to track the object and keep a laser spot on the target. And we give out solution to the more complicated extrinsic calibration since the third-person view instead of traditional camera-in-hand.

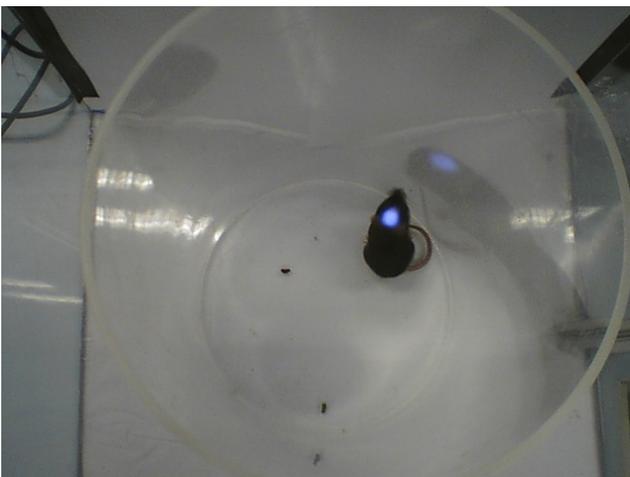


Fig. 1: Effect of tracking system. Illuminate the surgical wound of lab rat head by laser beam

For the large field of view or wide area surveillance, current visual tracking algorithms usually utilize the PTZ with the camera, such as [4]. Guha et al. [5] proposed a method to keep the target always in the middle of the field of view of the camera through adjusting the camera pose via PTZ. They used mean-shift algorithm for visual tracking and derived the error dynamics for a proportional-integral control action. Jain et al. [6] present a stationary-dynamic (or master-slave) camera assemblies to achieve wide-area surveillance and selective focus-of-attention. Their approach features the technique to calibrates all degrees-of-freedom (DOFs) of both stationary and dynamic cameras, using a closed-form solution that is both efficient and accurate. There are many state-of-the-art algorithms or approaches available for surveillance systems indeed, but as Fig.1 presents, our goal is to keep the laser spot continuously on the rat head which demand high precision and dynamic performance. Approaches of Guha and Jain doesn't involve the actuator which is laser generator in our project.

In this work, we combine the visual tracking and visual servo and propose a moving object tracking system to solve this problem. We don't use the PTZ to adjust the camera pose since we only need to guarantee the object within the field of view. On the contrary, we utilize the PTZ as a part of the actuator to help control the laser beam. Due to the difficulty of measuring the extrinsic parameters and efficient deployment, the challenging problem is parameter calibration. In addition, most previous approaches are self-adjusting that the camera is fixed with PTZ. In our tracking system, the actuator is totally irrelevant with camera which makes it different from the camera-PTZ control case.

Our proposed tracking system successfully solves the above mentioned difficulties via independent mathematical model and a low decoupling system framework. We give out accurate mathematical model and derive the parameter calibration solution. Our contributions in this work are listed as follow:

- We propose a rational structure and complete function tracking system. It's easy to deploy and modules decouple from the hardware level.
- We present full parameters mathematical model which is able to give a close-form solution. It means we can greatly reduce the complexity of our tracking system since we don't have to utilize complicated control systems.
- Based on the model we introduce the automatic parameter calibration strategy and it is the foundation of arbitrary

deployment.

II. TRACKING SYSTEM OVERVIEW AND WORKFLOW

Fig.2 shows the deployment of the tracking system in real environment. Camera, laser with Pan-Tilt, and holders compose the tracking system. ③ marked in Fig.2 is the plane where objects, the lab rats, move on. ③ is the IoT control board called ATOM which is small but powerful core, comprised of DUAL high performance processors (STM32 MCU and Linux-in CPU). Our algorithm is mainly implemented on this hardware [7]–[10]. The tracking system includes not only visual tracking but the execution. It features automatic calibration, extensibility, the low coupling of modules and support for the third-part algorithms or approaches. Fig.3 presents the workflow of our tracking system which presents the three working phases.

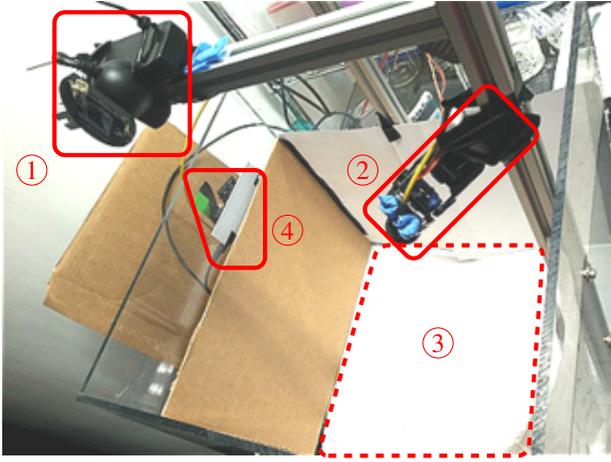


Fig. 2: Framework overview. ①Arbitrarily mounted camera ②Laser with Pan-Tilt ③Target plane ④IoT ATOM board

A. Initialization Phase

In this step the primary work is to install hardware. Like Fig.2 indicates. One purpose of this tracking system is to arbitrarily mount the hardware to improve efficiency without losing accuracy. As result, we can mount the camera and Pan-Tilt with arbitrary pose where the field of view of camera and laser workspace can cover the tracking plane and do not have to take care about the relative position initially.

B. Calibration Phase

To cast the laser beam precisely we have to know the Pan-Tilt intrinsic parameters (Pan-Tilt offsets) and system extrinsic parameters (Pan-Tilt's pose to the plane). We propose an automatic calibration procedure to get these parameters in this phase.

For this situation it's not able to directly measure the required parameters. In section III we model the structure and derive the calibration formulas. In this phase, we drive the Pan-Tilt to zigzag scan the plane, like the first image of Fig.6, and use the camera to capture the laser spot position at the

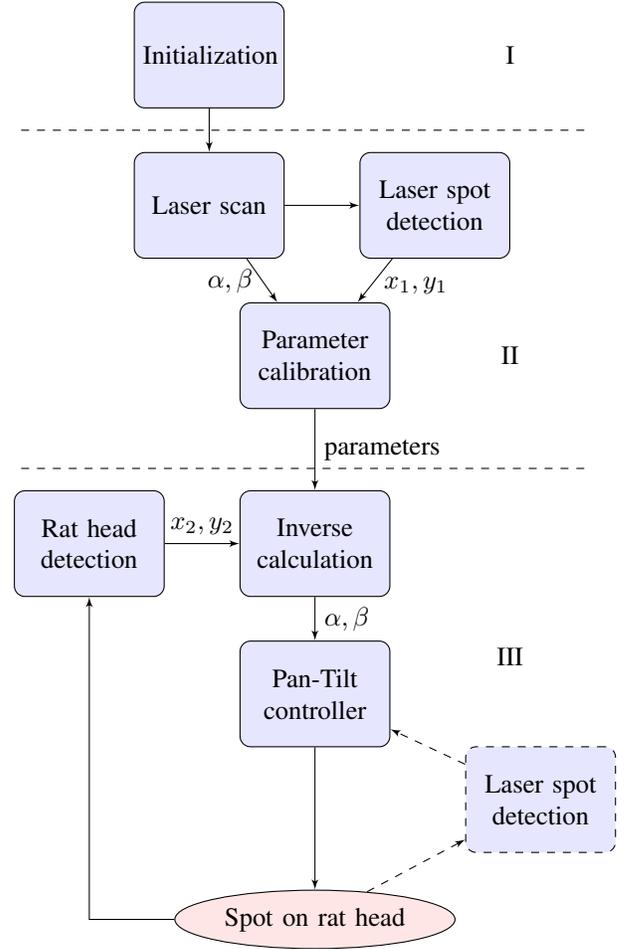


Fig. 3: System work flow. x_1, y_1 are laser spot position, x_2, y_2 are rat head position. α, β are the pan and tilt angle

same time. The sample data should be larger than 7 which is minimum to solve the equation. The Pan-Tilt angle commands and laser spot positions are taken to calibrate the parameters.

C. Tracking Phase

This phase is a common detection-control loop. The camera captures images and we use the method mentioned in section III to locate the rat head, then with the calibrated parameters we can inverse calculate the required α, β which are the expectation of Pan-Tilt controller. In this process, the inverse calculation strategy has been discussed in section II. In Fig.3 there is a dashed border block in phase III which is optional and not implemented in our system but it's easy to integrate into the system.

III. KINEMATIC ANALYSIS

Unlike common PTZ (Pan-Tilt-Zoom) camera based active tracking system, apart from tracking moving object, we have to take action on the target the lab rat in our project. We have a laser which could emit specified frequency laser fixed with the Pan-Tilt. To make sure the laser beam hits the target

accurately, we did kinematic analysis including modeling, parameter calibration and inverse calculation.



Fig. 4: Actuator overview. ①First Servo ②Second Servo ③Laser generator. ④Rotation axis of first servo ⑤Rotation axis of second servo

A. Modeling

Fig.4 shows the detailed structure. We have two servo motors marked ① and ② in the figure. The laser generator, two servo motors and basement are connected in series. The rotation axes of motors marked by white line ④ and ⑤ are perpendicular but not intersecting.

In Fig.5, $O_0x_0y_0z_0$ is the world coordinate where z axis is the same direction as the rotation axis of the first servo motor. We attach the $O_1x_1y_1z_1$ and $O_3x_3y_3z_3$ to the first servo and second servo respectively. Coordinate $O_4x_4y_4z_4$ is attached to the laser whose laser beam is shot in the x axis direction. In addition, α and β are the rotation angle of servos. a_1, a_2, a_3 and a_4 are the offsets to be estimated.

As shown in Fig.5, we compute the coordinate transformations. We present the coordinate of point P or coordinate frame x with respect to coordinate frame y by the notation ${}^y\mathbf{T}_x$. The transformations of coordinate frame in Fig.5 are given as following:

$${}^0\mathbf{T}_1 = \begin{bmatrix} 1 & 0 & 0 & a_1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & a_2 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad {}^1\mathbf{T}_2 = \begin{bmatrix} \cos \alpha & -\sin \alpha & 0 & 0 \\ \sin \alpha & \cos \alpha & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

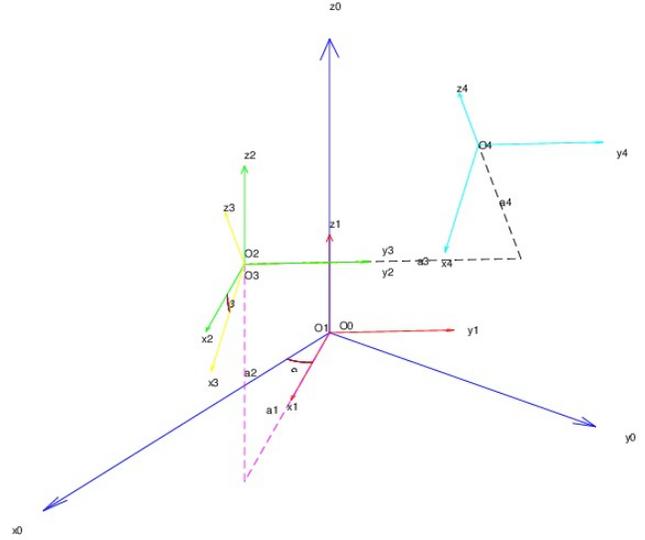


Fig. 5: coordinate overview

$${}^2\mathbf{T}_3 = \begin{bmatrix} \cos \beta & 0 & \sin \beta & 0 \\ 0 & 1 & 0 & 0 \\ -\sin \beta & 0 & \cos \beta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad {}^3\mathbf{T}_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & a_3 \\ 0 & 0 & 1 & a_4 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

We can get the composition coordinate transformation by right multiply the transformations. i.e.

$${}^0\mathbf{T}_4 = {}^0\mathbf{T}_1 {}^1\mathbf{T}_2 {}^2\mathbf{T}_3 {}^3\mathbf{T}_4$$

The result of ${}^0\mathbf{T}_4$ is :

$$\begin{bmatrix} \cos \alpha \sin \beta & -\sin \alpha & \cos \alpha \sin \beta & a_1 + a_3 \sin \alpha + a_4 \cos \alpha \sin \beta \\ \sin \alpha \cos \beta & \cos \alpha & \sin \alpha \cos \beta & a_3 \cos \alpha + a_4 \sin \alpha \cos \beta \\ -\sin \beta & 0 & \cos \beta & a_2 + a_4 \cos \beta \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The first column is the unit base vector of x axis which is chosen as the direction our laser emitting and the fourth column is the origin of coordinate frame $O_4x_4y_4z_4$ which we take as the position of the laser. Thus, we get a laser ray at L with direction \vec{v} .

$$\vec{v} = \begin{bmatrix} \cos \alpha \sin \beta \\ \sin \alpha \cos \beta \\ -\sin \alpha \end{bmatrix} \quad (1)$$

Let $\vec{a} = [a_1 \quad a_2 \quad a_3 \quad a_4]^T$

$$\begin{aligned} \vec{l} &= \begin{bmatrix} a_1 + a_3 \sin \alpha + a_4 \cos \alpha \sin \beta \\ a_3 \cos \alpha + a_4 \sin \alpha \cos \beta \\ a_2 + a_4 \cos \beta \end{bmatrix} \\ &= \begin{bmatrix} \cos \alpha & 0 & -\sin \alpha & \cos \alpha \sin \beta \\ \sin \alpha & 0 & \cos \alpha & \sin \alpha \sin \beta \\ 0 & 1 & 0 & \cos \beta \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix} \\ &= C\vec{a} \end{aligned} \quad (2)$$

Our target is the specified point P on a given plane in the space. Once the laser spot can reach the point we can present the point by the position and direction of the laser.

$$P_i = C_i \vec{a} + r_i \vec{v}_i \quad (3)$$

r_i is the length from the laser (O_4^i) to the point P_i on the plane. Besides, points on the plane can be presented by themselves. We can choose an origin point and two base vector for this plane, such that each point on this plane can be presented as :

$$P_0 P_i = x P_0 P_x + y P_0 P_y \quad (4)$$

$$P_i = x P_0 P_x + y P_0 P_y + P_0 \quad (5)$$

B. Parameters Calibration

Since the rotation axes of Pan-Tilt are not coaxial and the offsets, a_i and r_i , are hard to measure directly and accurately. We propose to estimate the parameters based on sample data.

Combine equation (3) and (5):

$$C_i A + r_i V_i = x P_0 P_x + y P_0 P_y + P_0 \quad (6)$$

By just taking $P_0 P_1$ and $P_0 P_2$ as P_x, P_y :

$$P_0 P_i = P_i - P_0 = (C_i - C_0) A + r_i V_i - r_0 V_0 \quad (7)$$

$$\begin{aligned} C_i \vec{a} + r_i \vec{v}_i &= x_i P_0 P_1 + y_i P_0 P_2 + P_0 \\ &= x_i (C_1 \vec{a} + r_1 \vec{v}_1 - r_0 \vec{v}_0) \\ &\quad + y_i (C_2 \vec{a} + r_2 \vec{v}_2 - r_0 \vec{v}_0) + C_0 \vec{a} + r_0 \vec{v}_0 \end{aligned}$$

$$\begin{aligned} [x_i(C_1 - C_0) + y_i(C_2 - C_0) - (C_i - C_0)] \vec{a} \\ + (1 - x_i - y_i) r_0 \vec{v}_0 + x_i r_1 \vec{v}_1 + y_i r_2 \vec{v}_2 - r_i \vec{v}_i = 0 \quad (8) \end{aligned}$$

Reorganize the equation (8) and we can get the homogeneous system of linear equations in shape of $B \vec{x} = 0$ and $\vec{x} = [a_1 \ a_2 \ a_3 \ a_4 \ r_1 \ \dots \ r_n]^T$. Since there is no scale, we can not get the true parameters, but it does not matter. As given the input $[\alpha_i \ \beta_i \ x_i \ y_i]$ ($i \geq 7$), using the SVD to decompose the coefficient matrix to get the exact solution. The 2nd column of coefficient matrix is constant zero vector which means a_2 is not feasible to compute but it make no influence on the result.

C. Inverse Kinematic

We get the sample data based on the forward kinematic which calculate the coordinate of P_i and r_i via given $[\alpha \ \beta]$, A and plane. However, inverse kinematic is to figure out the $[\alpha \ \beta]$ by given $[x \ y]$ and calibrated parameters.

For P_i :

$$\begin{aligned} O_4^i P_i &= O P_0 + P_0 P_i - O O_4^i \\ &= C_0 \vec{a} + r_0 \vec{v}_0 + x_i P_0 P_2 + y_i P_0 P_3 - C_i \vec{a} \end{aligned}$$

Let set $\vec{v} = C_0 \vec{a} + r_0 \vec{v}_0 + x_i P_0 P_2 + y_i P_0 P_3$, a vector with 3 elements. Since $O_4^i P_i$ should be parallel to \vec{v}_i thus we can get the equations:

$$\vec{v}_i \times (\vec{v} - C_i \vec{a}) = \mathbf{0} \quad (9)$$

The expanded and simplified form of the equation(9) is:

$$\begin{cases} a_1 \sin \alpha_i \sin \beta_i + a_3 \cos \alpha_i \sin \beta_i + a_4 \sin \alpha_i \\ \quad - v(2) \sin \beta_i - v(3) \sin \alpha_i \cos \beta_i = 0 \quad (10) \\ a_1 \cos \alpha_i \sin \beta_i - a_3 \sin \alpha_i \sin \beta_i + a_4 \cos \alpha_i \\ \quad - v(1) \sin \beta_i - v(3) \cos \alpha_i \cos \beta_i = 0 \quad (11) \end{cases}$$

By solving the above equation array, we can get exact solutions $[\alpha_i \ \beta_i]$ for a given P_i , which are the commands sent to the Pan-Tilt.

IV. VISION BASED DETECTION

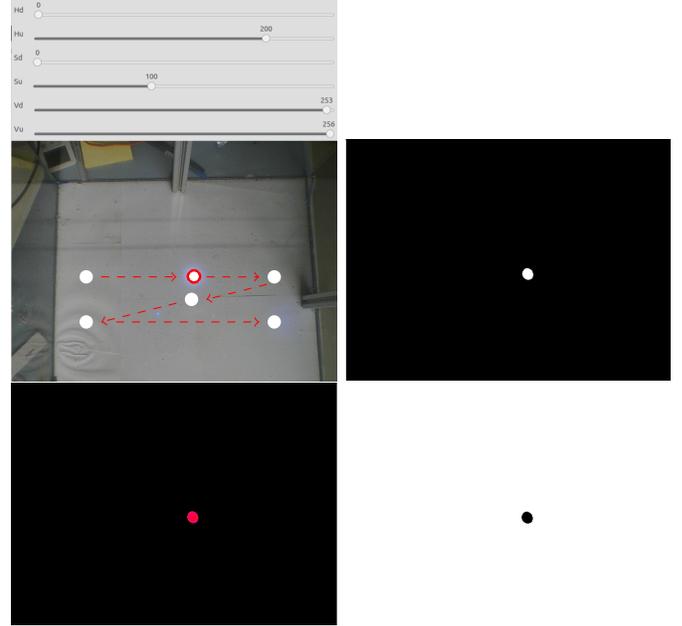


Fig. 6: Laser spot detection

Computer vision community has proposed lots of available approaches for detection of the interested region or objects, which serve well for the automatic tracking system. For complicated objects and environment, we have several algorithms to track the moving objects, the noticeable methods are Kalman filter-based algorithms [11] and CONDENSATION algorithm [12]. For regular objects or specified environment, blob detection [13], [14] and contour detection can help a lot. These algorithms work very fast which usually means real-time performance and more robust.

In our project, vision-based tracking approach serves in both parameter calibration and tracking process. For parameter calibration, we utilize the camera to capture the image of the target plane and get the laser spot location which is the input for calibration. And for tracking process we surveil the rat on the plane and get the rat head location where the laser spot should be. In Fig.6 the laser spot is a typical blob and in Fig.7 we can easily find three spot areas which are ears and wound after the operation. In consideration of the characteristic of tracking the target, laser spot and rat head, we utilize the blob detection to locate the target.

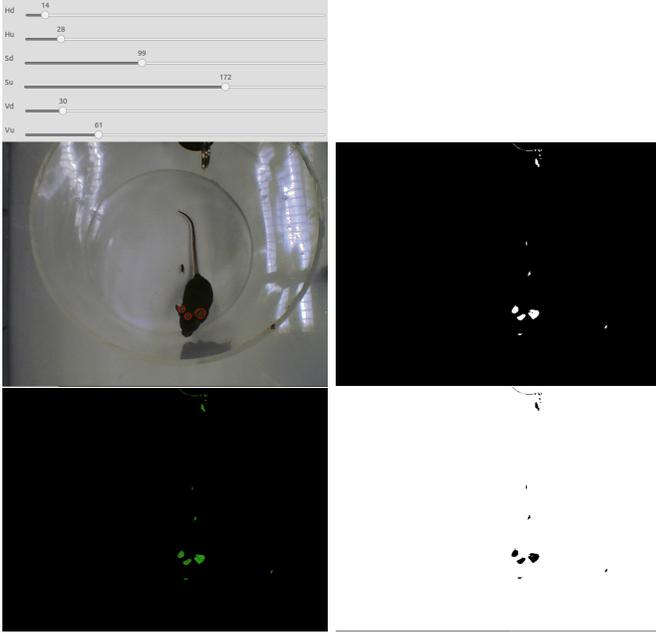


Fig. 7: Rat head detection

We use the blob detector from OpenCV [15] which provides various and out-of-the-box algorithm implementations. Shown in Fig.6 and Fig.7, we convert the image to HSV and get the threshold which the key parameter for blob detector. For two different blobs we manually set the filter parameters of the detector. As shown in Fig.7 we take the average of the blobs location as the rat head location.

V. EXPERIMENTS

For the model, we generate samples via $\vec{a} = [1 \ 2 \ 3 \ 4]^T$ and plane $x - y + z + 4 = 0$ and then use our approach to calculate the parameters.

TABLE I: Sample data

i^{th}	α_i ($^\circ$)	β_i ($^\circ$)	x_i	y_i
1	0	0	0	0
2	25.0000	0	1.0000	0
3	0	10.0000	0	1.0000
4	25.0000	10.0000	1.6037	2.1394
5	22.0000	3.0769	0.8744	0.4043
6	6.0000	6.1538	0.1680	0.6003
7	3.0000	3.8462	0.0752	0.3389

TABLE I presents the sample data which are used to calibrate the parameters and TABLE II is the result. Since there is no scale, we choose the a_1 as the scale factor to scale the result (2nd column) and get the 3rd column. As we mentioned before the scale has no effects on the inverse kinematic. TABLE II presents clearly that our approach gets the exact solution.

Fig.8 records the field test results. Although errors still exist in the figures, the laser spot is not the ideal point but a quite large dot which can compensate the position error to some degree and taking the scale and output resolution of the Pan-Tilt into consideration, we think the system meets

TABLE II: Parameters

	Ground truth	result	rescaled result
a_1	1.0000	-0.0282	1.0000
a_2	2	-	-
a_3	3.0000	-0.0846	3.0000
a_4	4.0000	-0.1128	4.0000
r_1	-8.0000	0.2256	-8.0000
r_2	-13.4320	0.3789	-13.4320
r_3	-10.6438	0.3002	-10.6438
r_4	-22.3729	0.6310	-22.3729
r_5	-13.5716	0.3828	-13.5716
r_6	-10.2247	0.2884	-10.2247
r_7	-9.1648	0.2585	-9.1648

our requirement. And owing to the simple architectures and accurate model the tracking system is able to work efficiently. However, due to the current limitation of current system and measurement, we don't do contrast experiment which we have to update later.

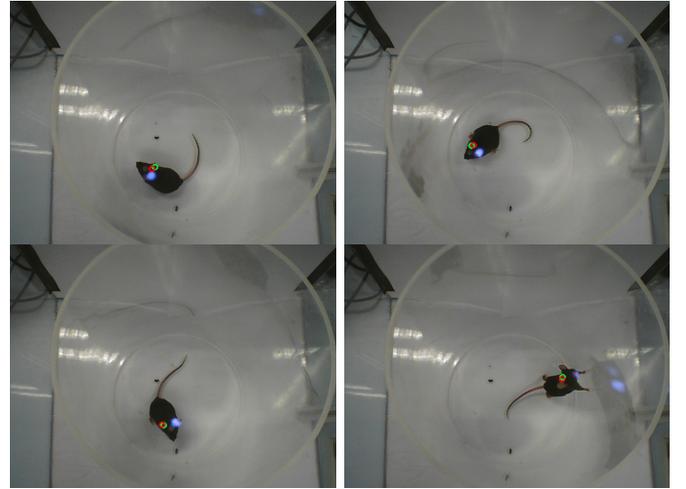


Fig. 8: Field test. Circles are the position detected by camera. purple spot is laser spot

VI. DISCUSSION AND CONCLUSION

In this work, we presented a visual tracking system featuring easy deployment, efficiency, and low coupling. Our mathematical model offers no error output in theory. In addition, we designed an automatic parameters calibration strategy and make it able to arbitrarily install hardware without accuracy loss. Owing the precise mathematical model, we can reach our goal at low cost instead of relying on the complex control system.

Our mathematical model is designed for tracking objects on plane. However, objects usually are three-dimensional, imaging the rat stands in our projects (the last one in Fig.8), and our current model is doomed to failure or errors. Thus, involving another dimension and a more comprehensive model is an interesting direction in the future. Moreover, integrating and testing better visual object detection algorithms are worth trying.

ACKNOWLEDGMENT

This work is supported by the Research Grant Council of Hong Kong SAR Government, China, under project No. 16206014 and No. 16212815; National Natural Science Foundation of China No. 6140021318, awarded to Prof. Ming Liu.

REFERENCES

- [1] R. T. Collins, A. J. Lipton, T. Kanade, H. Fujiyoshi, D. Duggins, Y. Tsin, D. Tolliver, N. Enomoto, O. Hasegawa, P. Burt, and L. Wixson, *A System for Video Surveillance and Monitoring*, 2000, vol. 823, no. 8. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=868676>
- [2] C.-S. Yang, "PTZ camera based position tracking in IP-surveillance system," *2008 3rd Int. Conf. Sens. Technol.*, pp. 142–146, 2008. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4757089>
- [3] R. Cucchiara, A. Prati, and R. Vezzani, "Advanced Video Surveillance with Pan Tilt Zoom Cameras 2 . People Tracking from Active Camera."
- [4] M. Lalonde, S. Foucher, L. Gagnon, E. Pronovost, M. Derenne, A. Janelle, and S. S. West, "A system to automatically track humans and vehicles with a PTZ camera," *Processing*, vol. 6575, 2007. [Online]. Available: http://www.crim.ca/perso/langis.gagnon/articles/6575-1_final.pdf
- [5] P. Guha, D. Palai, D. Goswami, and A. Mukerjee, "DynaTracker: Target tracking in active video surveillance systems," *ICAR '05. Proceedings., 12th Int. Conf. Adv. Robot. 2005.*, pp. 621–627, 2005. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1507473>
- [6] A. Jain, D. Kopell, K. Kakligian, and Y. F. Wang, "Using stationary-dynamic camera assemblies for wide-area video surveillance and selective attention," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 1, pp. 537–544, 2006.
- [7] L. Wang, M. Liu, M. Q. H. Meng, and R. Siegwart, "Towards real-time multi-sensor information retrieval in cloud robotic system," in *Multisensor Fusion and Integration for Intelligent Systems (MFI), 2012 IEEE Conference on*, Sept 2012, pp. 21–26.
- [8] L. Wang, M. Liu, and M. Q. H. Meng, "Towards cloud robotic system: A case study of online co-localization for fair resource competence," in *Robotics and Biomimetics (ROBIO), 2012 IEEE International Conference on*, Dec 2012, pp. 2132–2137.
- [9] M. Tan, L. Wang, D. Tardioli, and M. Liu, "A resource allocation strategy in a robotic ad-hoc network," in *Autonomous Robot Systems and Competitions (ICARSC), 2014 IEEE International Conference on*, May 2014, pp. 122–127.
- [10] L. Wang, M. Liu, and M. Q. H. Meng, "Real-time multisensor data retrieval for cloud robotic systems," *IEEE Transactions on Automation Science and Engineering*, vol. 12, no. 2, pp. 507–518, April 2015.
- [11] J. Ferryman, S. Maybank, and A. Worrall, "Visual surveillance for moving vehicles," in *Visual Surveillance, 1998. Proceedings., 1998 IEEE Workshop on*, Jan 1998, pp. 73–80.
- [12] M. Isard and A. Blake, "Condensation - conditional density propagation for visual tracking," *Int. J. Comput. Vis.*, vol. 29, no. 1, pp. 5–28, 1998. [Online]. Available: <http://link.springer.com/article/10.1023/A:1008078328650>
- [13] M. Isard and J. MacCormick, "Bramble: a bayesian multiple-blob tracker," in *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, vol. 2, 2001, pp. 34–41 vol.2.
- [14] T. B. Nguyen and S. T. Chung, "An improved real-time blob detection for visual surveillance," in *Image and Signal Processing, 2009. CISP '09. 2nd International Congress on*, Oct 2009, pp. 1–5.
- [15] G. Bradski, "OpenCV," *Dr. Dobb's Journal of Software Tools*, 2000.