Proceedings of 10th IEEE International Conference
on CYBER Technology in Automation, Control, and Intelligent Systems
October 10 - 13, 2020, Xi'an, China

Multiple Object Tracking for Similar, Monotonic Targets

Yifei Qian, Hui Shi, Maojin Tian, Runhuai Yang*, Yuping Duan*

***Abstract* - As a vital and extremely hot branch of computer vision, multiple object tracking is widely used in security monitoring and biological behavior research. At present, the majority of the multiple object tracking algorithms are composed of two stages: detection and data association. There is often mutual interference between targets in each frame, which causes great trouble to object detection and correlation. Especially for the microscopic particles, detection and tracking always is a tricky business because of its similar appearance and irregular movement. In this paper, a multiple object tracking algorithm based on two-step detection is proposed for microscopic non-fluorescent labeled object with simple structure, monotonous features, even no obvious features for learning. The specific operation is: first threshold segment each image in video to extract targets from the background, and perform morphological operations on the obtained graph to separate the slightly adhesion targets to avoid incorrect tracking, then estimate object states by Kalman Filter and use Hungarian algorithm for data association, definitively achieve the goal of multiple object tracking. Experiments show that our algorithm is better than the general method, meeting the tracking task of multiple non-fluorescent labeled particles commendably.**

***Index Terms* - multiple object tracking/similar object/simple object/ Kalman filter algorithm/particles tracking**

I. INTRODUCTION

Multiple object tracking is a crucial branch of computer vision and its role in our life is also momentous. Most multiple object tracking algorithms are mainly divided into two parts: object detection and data association. In recent years, with the consecutive development and improvement of deep learning, more and more of the object detection and tracking algorithm have been put forward, [1] proposed a deep affinity network, which inputs a pair of processed video frames to a VGG network for feature extraction, and then performs data association based on the output of the confidence estimator to achieve multiple object tracking. Reference [2] raised a novel detection algorithm that combines deep neural networks with traditional general object detection modes, with area selection module which is responsible for selecting areas where features can be learned, and deep partition learning module which focuses on local features selection and conversion.

Nevertheless, the methods above are aimed at objects with obvious features like texture and color that can be learned in deep learning, just like the texture and contour of human eyes in face detection and vehicle frame and corner point in vehicle tracking. When it comes to some simple objects, such as simple "circles" and bacteria under microscopy, their structure is simple, color and texture are single, and there are no apparent features for learning, deep learning may lose its luster. At these points, if we use deep learning as detection algorithm and then achieve multiple object tracking, there may be some troubles.

In biology, for the tracking of bacteria, viruses and other "particles", the commonly used detection algorithm is spot enhanced filter[3-5], also called Laplace of Gaussian filter(LoG). In [3], a detection algorithm combining edge detection algorithm and limited area sampling is proposed, simultaneously uses the weight of image likelihood as the probability in the Probabilistic Data Association to complete the correlation. Reference [4] suggests a multi-scale spot enhanced filter as detection algorithm, using two-step multi-frame association to achieve multi-target tracking. Reference [5] proposes spot enhanced filter to detect particles, a neural network LSTM to predict the object state, and Hungarian algorithm for data association.

All of the above-mentioned articles track after fluorescent microscope marking which reduce various noise of the image, while for non-fluorescent marking targets, the detection performance of spot enhanced filter is unsatisfactory.

In light of the foregoing, this paper proposes a method based on two-step detection to track that non-fluorescent labeled object with simple structure and no obvious features for learning. First use the primary threshold segmentation as rough detection to extract objects from the background, then employ morphological operations to separate multiple object with slight contact to accomplish refined detection. Then estimate object states through Kalman Filter and the Hungarian algorithm is used to correlate the detections between two frames. The Operational flowchart is shown in Figure 1.

II. DETECTION

Due to the favorable extraction of features, deep learning emerges in image classification and object detection frequently in recent years, with remarkable prosperity. Familiar frameworks include YOLO, SSD, etc., they are all algorithms based on convolutional neural networks with the disadvantages of large computation, long time, and the need for enormous datasets. On the other

[1]Yifei Qian, Postgraduate student, studying in the Department of Biomedical Engineering, Anhui Medical University, Hefei, Anhui, China.
[2]Hui Shi, postgraduate tutor, working in the Department of Biomedical Engineering, Anhui Medical University, Hefei, Anhui, China.
[3]Maojin Tian, working in Hefei National Laboratory for Physical Sciences at the Microscale and Department of Physics, University of Science and Technology of China, Hefei, Anhui 230026, China.
*Corresponding author: Yuping Duan, postgraduate tutor, working in the Department of Biomedical Engineering, Anhui Medical University, Hefei, Anhui, China. (email: duanyuping126@126.com)
*Corresponding author: Runhuai Yang, postgraduate tutor, working in the Department of Biomedical Engineering, Anhui Medical University, Hefei, Anhui, China. (email: yangrunhuai@ahmu.edu.cn)

978-1-7281-9010-5/20/$31.00 © 2020 IEEE

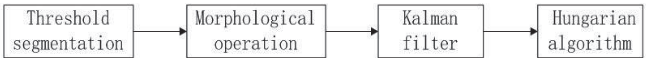
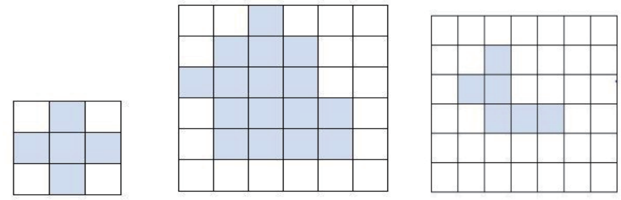


Figure 1. Operational Flowchart



Structural Element B Structural Element A Result of Corrosion

Figure 2. The Schematic Graph of Corrosion

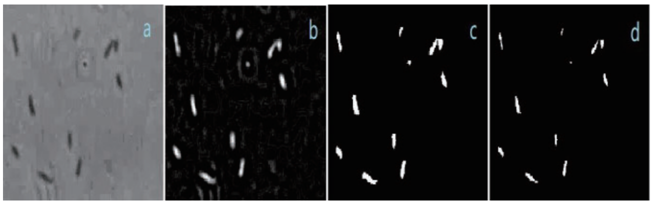


Figure 3. Detection Comparison: a) is the original image; b) is the result of spot enhanced filter(LoG); c) is the outcome of threshold segmentation; d) is the consequence of corrosion of c.

hand, spot enhanced filter underperforms in the presence of various noises without fluorescent labeling. Given that, in this paper, we adopt the simplest threshold segmentation, which takes less computing time, as the first step to separate the target from background to meet the coarse detection requirements.

For (1), f(x) denotes the pixel value after thresholding, g(x) is the pixel value before thresholding, and a signifies the selected threshold.

$$f(x)=\begin{cases}0, g(x)<a\\1, g(x)>a\end{cases} \quad (1)$$

Then, for the object with slight adhesion, we employ morphological operations to separate the targets, and the corrosion formula is shown as (2). The object of this paper is assumed to be on a two-dimensional plane, so occlusion and covering situations are not considered.

$$A \odot B=\{x \mid Bx \subseteq A\} \quad (2)$$

For formula (2), it can be interpreted to that moving structural element B, if all the intersection of B and A is in the valid region of A, the location point is saved, and the set of all the points satisfying the condition is the result of structure A being corroded by structure B. The schematic diagram of corrosion is shown in Figure 2. And detection comparison between spot enhanced filter and our method is drawn in Figure 3.

III. Object Tracking

Kalman filter is a state estimation algorithm for linear system, and frequently appear in the state estimation of object tracking. There are two main stages in data prediction: one is the estimation of object state originated from prediction of the model based on the state of previous

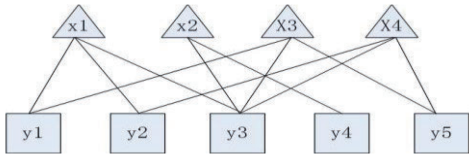


Figure 4. Schematic Diagram of Hungarian Algorithm

frame, and the other is the measurement of object trajectory obtained from detection and used for updating. The prediction model of Kalman filter[3, 6, 7] is:

$$x_k = a x_{k-1} \quad (3)$$

$$p_k = a p_{k-1} a^T + q \quad (4)$$

Where x_k is the estimation state of the current frame of the object, x_{k-1} is the state of the previous frame, a represents the state transition matrix, a^T is the transpose of a, q is the process noise, and p_k is the covariance matrix of the previous frame.

To make the state accurately after estimating each frame, the parameters of the model must to be updated and the renewal formulas are shown as (5-8), where c is the covariance weighting matrix, w is the observation equation, b is the variance matrix of the observation noise, g_k is the gain matrix, and z_k is the measurement value of object.

$$c = w p_k w^T + b \quad (5)$$

$$g_k = p_k w^T c^{-1} \quad (6)$$

$$\hat{x}_k = x_k + g_k (z_k - x_k) \quad (7)$$

$$\hat{p}_k = p_k - g_k c g_k \quad (8)$$

IV. Data Association

In addition to the precision of detection, multiple object tracking generally accompanied by another trouble because it is not like a single object tracking whose detection result and trajectory is one-to-one in each frame, the detection results and trajectories in multiple object tracking are many-to-many. Hence, how to pair detection results with trajectories is the problem we have to work out. In [8], the nearest neighbor matching algorithm is proposed, that is, each existing trajectory is paired with the detection result closest. When targets are intensively packed, the nearest result of detection is not necessarily the target of the trajectory. The probability data association (PDA) has also been mentioned[3], but it also has disadvantages when facing the difficult scenarios.

In this paper, the Hungarian algorithm, an optimization algorithm to find the minimum of maximum matching loss sum, is proposed to match the detection results with the existing trajectories. The schematic diagram of Hungarian algorithm is demonstrated in Figure 4. In order to make one-to-one pairing between x and y, it is known that x1-x4 and y1-y5 can be matched into at most four pairs, but the matching paths are various, for example, y1 can be paired

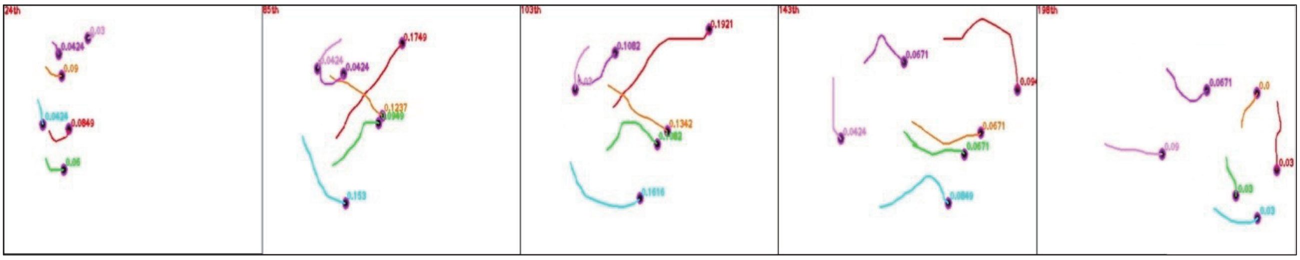


Figure 5. Simulation Video Results

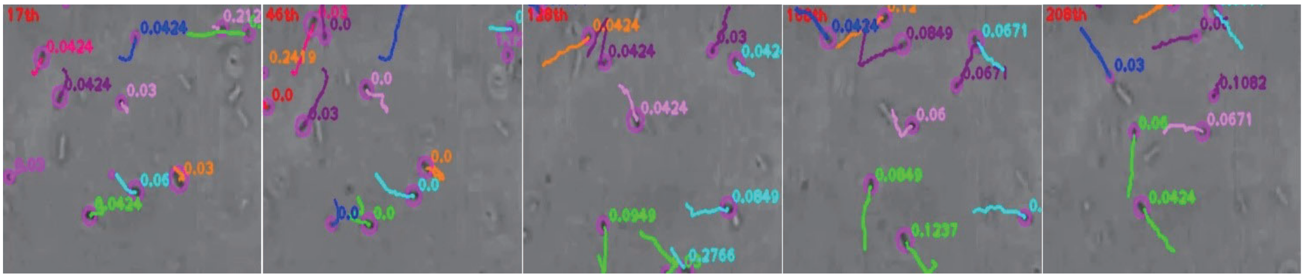


Figure 6. Bacteria 1 (Scattered Distribution) Tracking Results

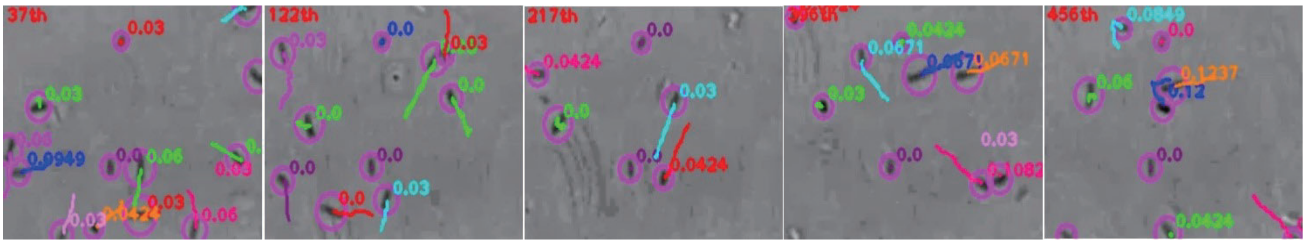


Figure 7. Bacteria 2 (Micro-contact) Tracking Results

with x1 and x3, y2 can be paired with x1 and x4...Every pair of x and y is accompanied by a corresponding loss, the loss can be the distance between x and y, the Hungarian algorithm is to find the minimum value of the sum of these losses while ensuring the maximum match.

V. EXPERIMENTAL RESULTS

The experiment is based on python 3.6.5 programming environment, after programming the aforementioned algorithms, we get the experimental results.

In order to prove the feasibility of the theory, we synthesize a video of multiple "circles" in a two-dimensional environment whose background is neat with little noise. Since it is two-dimensional, there is no object occlusion in the video, and only collisions, that is, micro-contact between object, occur. The experimental results are shown in Figure 5. There are six small "circles" in each frame, two of which collide on frame 85. The trajectories of the object are drawn and the motion velocities are calculated in each frame. Frames 24, 85, 103, 143 and 198 from the tracking results are respectively captured. The running results are good, which proves the feasibility of our theory to some extent.

Then, for reasons of universality, we collect two dynamic videos of multiple bacteria in nature that are characteristically monotonous. The bacteria in one of the videos, called Bacteria 1, are scattered, while in the other named Bacteria 2, they are relatively dense and some of them are micro-contact. After running the code, frames 17, 46, 128, 166 and 208 in the video of Bacteria 1 and frames 37, 122, 217, 356 and 456 in the video of Bacteria 2 are respectively captured. We keep trajectories of the last thirty frames and print them out. Lines of different colors represent the trajectories of different particles. We can view that in these two scenarios, the backgrounds of bacteria are complex, and there are aggregation and micro-contact between particles, but our tracking performances are still good. The experimental results are shown in Figure 6 and Figure 7.

As seen in Figure 5, 6 and 7, our method solves the hardship of micro-contact which can not be realized by the approach based on spot enhanced filter, just like depicted in Figure 8 that a few defect tracking occurred.

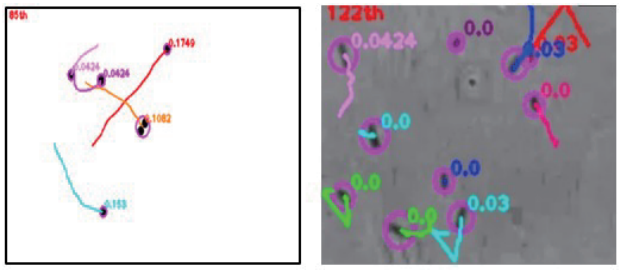


Figure 8. Results Based on Spot Enhanced Filter

Videos	Tracking Accuracy(Detected by LoG)	Tracking Accuracy(Detected by Our Approach)
Simulation Video	86.40%	99.863%
Bacteria 1	85.54%	94.041%
Bacteria 2	88.12%	91.397%

Figure 9. Tracking Accuracy Comparisons

By analyzing the statistics of correct and incorrect tracking in each frame, we quantitatively calculate the tracking accuracy $\mathrm{P}_{\mathrm{track}}$ in each video:

$$\mathrm{P}_{track} = \frac{n_{correct}}{n_{overall}} \tag{9}$$

Where $\mathrm{n}_{correct}$ represents the number of correct tracking and $\mathrm{n}_{\mathrm{overall}}$ is the count of overall tracking object. Run the proposed method and the spot enhanced filter on the same videos respectively, and the results are shown in Figure 9. We can see that the tracking accuracies of our approach are all above 90%, higher than the spot enhanced filter, which attest the feasibility of the algorithms.

VI. CONCLUSION

In this paper, we propose an approach that uses threshold segmentation and morphological operations as the two-step detection algorithm, Kalman filter algorithm as the tracking algorithm, and the Hungarian algorithm as the data association algorithm. The method is respectively applied to simulation and real video with similar, monotonous targets successfully. Certified by the experiments, the tracking accuracy are more than 90%, which can be applied for multiple non-fluorescent labeled particles tracking.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (Nos. 61973003, 61603002), The Basic and Clinical Collaborative Research Improvement Project of Anhui Medical University (2019xkjT017)

REFERENCES

[1] Sun, S., et al., "Deep Affinity Network for Multiple Object Tracking," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019: p. 1-1.

[2] Xu, H., et al., "Deep Regionlets: Blended Representation and Deep Learning for Generic Object Detection," IEEE Trans Pattern Anal Mach Intell, 2019.

[3] Godinez, W.J. and K. Rohr, "Tracking Multiple Particles in Fluorescence Time-Lapse Microscopy Images via Probabilistic Data Association", IEEE Transactions on Medical Imaging, 2015. 34(2): p. 415-432.

[4] Jaiswal, A., et al., "Tracking Virus Particles in Fluorescence Microscopy Images Using Multi-Scale Detection and Multi-Frame Association", IEEE Transactions on Image Processing, 2015. 24(11): p. 4122-4136.

[5] Roman Spilger B, T.W.Y.Q., B.M.U.O. Ji Young Lee and A.K. Rohr, "Deep Particle Tracker: Automatic Tracking of Particles in Fluorescence Microscopy Images Using Deep Learning", Deep Learning in Medical Image Analysis, 2018. 11045: p. 128 – 136.

[6] Faragher, R., "Understanding the Basis of the Kalman Filter Via a Simple and Intuitive Derivation", IEEE Signal Processing Magazine, 2012. 29(5): p. 128-132.

[7] Zhao, X., S. Yan and Q. Gao," An Algorithm for Tracking Multiple Fish Based on Biological Water Quality Monitoring", IEEE Access, 2019. 7: p. 15018-15026.

[8] Edwards, M.R., et al., "Swimming characterization of Serratia marcescens for bio-hybrid micro-robotics", Journal of Micro-Bio Robotics, 2014. 9(3-4): p. 47-60.