



Fusing RFID and Computer Vision for Occlusion-Aware Object Identifying and Tracking

Min Li^{1,2,3}(✉), Yao Chen^{2,3}, Yanfang Zhang^{2,3}, Jian Yang^{2,4},
and Hong Du¹

¹ Beijing Jiaotong University, Beijing, China
{limin, duhong}@iie.ac.cn

² Institute of Information Engineering, Chinese Academy of Sciences,
Beijing, China
{limin, chenyaoyao, zhangyanfang}@iie.ac.cn,
jyang@bit.edu.cn

³ School of Cyber Security, University of Chinese Academy of Sciences,
Beijing, China

⁴ School of Optics and Photonics, Beijing Institute of Technology,
Beijing, China

Abstract. Real-time identifying and tracking monitored objects is an important application in a public safety scenario. Both Radio Frequency Identification (RFID) and computer vision are potential solutions to monitor objects while faced with respective limitations. In this paper, we combine RFID and computer vision to propose a hybrid indoor tracking system, which can efficiently identify and track the monitored object in the scene with people gathering and occlusion. In order to get a high precision and robustness trajectory, we leverage Dempster-Shafer (DS) evidence theory to effectively fuse RFID and computer vision based on the prior probability error distribution. Furthermore, to overcome the drift problem under long-occlusion, we exploit the feedback from the high-confidence tracking results and the RFID signals to correct the false visual tracking. We implement a real-setting tracking prototype system to testify the performance of our proposed scheme with the off-the-shelf IP network camera, as well as the RFID devices. Experimental results show that our solution can achieve 98% identification accuracy and centimeter-level tracking precision, even in long-term occlusion scenarios, which can manipulate various practical object-monitoring scenarios in the public security applications.

Keywords: RFID · Visual tracking · DS evidence theory · Occlusion-aware

1 Introduction

Nowadays, with the increasing number of thefts in museums, important venues, etc., anti-theft systems are gradually integrating with industries such as Closed Circuit Television [1] and intelligent electronic device [2, 3]. However, these anti-theft products can only deal with those cases in which the object is stolen, but fail to identify

and track the thief bringing with the object when multiple people are gathering, which is more and more mentioned as the advanced public security services [4].

Intuitively, both RFID and computer vision are potential solutions for above public security services. RFID has been widely used in many fields because of its light volume, non-line-of-sight propagation, automatic identification, and low cost [5]. A RFID reader can efficiently identify the tag people are wearing within the working range. However, it is very difficult to accurately track moving people based on passive RFID system alone, in which the positioning accuracy is usually at the meter level, such that fine-grained identification and tracking of tagged objects cannot be achieved within this range. On the other hand, computer vision has achieved remarkable accuracy in tracking individual objects and is currently broadly applied in many areas, such as robotics, surveillance, human motion analyses and so on. Nevertheless, the fatal problem faced by computer vision based tracking systems lies in that, such a system can be easily dysfunctional when occlusion occur over the monitored object or the thief [6]. Besides, computer vision cannot distinguish similar items, such as documents, valuable assets, thus it is unfeasible to track these objects solely with computer vision. Furthermore, suppose that there are multiple people gathering, the monitored object is being stolen with a sever occlusion on it. In such a scenario, neither RFID nor computer vision can solely help the safeguard identify who is stealing the object and what track the thief is running out.

To solve such a practical problem, in this paper, we propose a fusing scheme to combine the advantages of RFID and computer vision for online identifying and tracking. More specifically, if the RFID tag is pre-attached to the object in visual surveillance scenarios, we can identify the thief by determining which track is the closest one to the trajectory of the tag. When an item is lost in a multiple-people gathering scenario, the camera detects and tracks all of the individuals in the surveillance area, while the RFID reader detects the tag on the item and tracks the tag. The position and speed of the thief and the tagged stolen item at the same time are often the closest. Therefore, for the detected RFID tag, we try to match it to one of the people that is most likely to be the thief by using a distance and velocity probability matrix in the sliding window. Due to the low accuracy of the RFID trajectory and the fragility of the visual trajectory to occlusion, neither RFID nor visual can solely help a fine-grained and robust tracking. Unfortunately, previous studies only employ the visual trajectory or simple weighting method to obtain the final trajectory [15, 16], which results in a dramatic trajectory error. For this task, we introduce DS evidence theory to fuse the RFID position and the visual position based on a priori probability error distribution to obtain an accurate and robust trajectory. Additionally, visual tracking is prone to fail when the thief is occluded, especially under long-term occlusion. Thereby, the position matrix and peak information of correlation filter are used to determine and correct the false tracking under occlusion.

Contributions: In summary, this paper makes the following contributions:

- First, we propose an innovative combination of RFID signals and visual signals for online identification and high-precision tracking, which can be applied to find person stealing the tagged object.

- Second, DS evidence theory based on prior error distribution is used to fuse the position information of visual and RFID, which can obtain the fusion trajectory with high robustness and precision.
- Third, a template update strategy and the RFID signal are used to adjust the search box of correlation filter for visual tracking under long-term occlusion.
- Fourth, we implement the system with off-the-shelf camera and RFID products. It's validated that the tracking accuracy is set to the centimeter level and the matching accuracy is 98%. More importantly, the object can be tracked correctly even in long-term occlusion.

The rest of the paper is organized as follows. Section 2 introduces related research. The main design of our scheme is over viewed in Sect. 3. We describe the technical details of our approach to people tracking and identification in Sect. 4. Section 5 presents experimental results of the system. Finally, Sect. 6 concludes the paper.

2 Related Work

Recent studies have shown that people tracking and identification can be accomplished using RFID. The mainstream method in RFID-based people tracking solutions is to leverage Received Signal Strength (RSS) as fingerprints or distance ranging metrics [7]. Mehmood *et al.* [8] used Artificial Neural Networks (ANN) to learn the intrinsic link between RSS fingerprints and position coordinates. Figuera *et al.* [9] added the spectral information of the training data to the Support Vector Machine (SVM). Dwiyasa *et al.* [10] applied Extreme Learning Machine (ELM) to position fingerprint localization algorithm. However, the tracking accuracy of these methods in a complex experimental environment is still low, and it is not possible to accurately locate the tagged object.

On the other hand, recent years have witnessed the rapid advance of computer vision, making it possible for reliable object tracking. Since 2013, Correlation filter-based tracking strategies have been widely studied due to their high efficiency and strong robustness, and its representative method is the Kernelized Correlation Filter (KCF) [11]. Correlation filter predicts the region in the image related to the position of the last frame of the object and then extracts the object region. After a series of transformations, such as Fast Fourier transform, the region with the maximum response in the result is considered as the object's position. Unfortunately, Vision-based tracking is susceptible to illumination, distortion, occlusion, and fast motion, which ultimately leads to a failed tracking [6].

There are some researches focus on the combination of RFID and computer vision for identifying and tracking in different applications. A robot with RFID antennas and camera was used for people tracking in a crowded environment, and this approach was refined by using a fixed camera and RFID antennas [12]. Reference [13] introduce an approximate aggregation for tracking quantiles and range countings in wireless sensor networks. Xuan *et al.* introduce the fusion of electronic and visual signals, but it focuses on the matching algorithm, and the final positioning accuracy is low [14]. Reference [15] performed better in the theory, but the experimental environment was very small and the phase measurement was complicated.

3 Scheme Overview

To address our scheme clearly, we first introduce our scheme overview, as shown in Fig. 1. The whole scheme needs two kinds of input information, RFID information and computer vision information. For RFID part, we need the tag's ID, tracking trajectory, and velocities in all trajectory points. For visual part, we need the detected persons, tracking trajectories and also velocities in all trajectory points. Then, the data processing procedure is composed of three phases, association, fusion and occlusion-aware processing.

The association phase aims to match the RFID tag to the visual thief, we calculate the distance and velocity probability matrix in the time window, and select the person corresponding to the tag with the largest probability. The fusion phase is to get an accurate and robust trajectory. We employ DS evidence theory to fuse the coordinate information of the two sensors based on the prior error distribution, and obtain a fusion trajectory superior to the single sensor trajectory. The occlusion-aware processing intends to correct false tracking in case of occlusion when tracking the thief, we use the feedback information from the tracking results and the probability matrix of the two sensors, so that the tracker can still track correctly after long-term occlusion.

Based on the scheme architecture, we then introduce the technic details of the scheme in the next section.

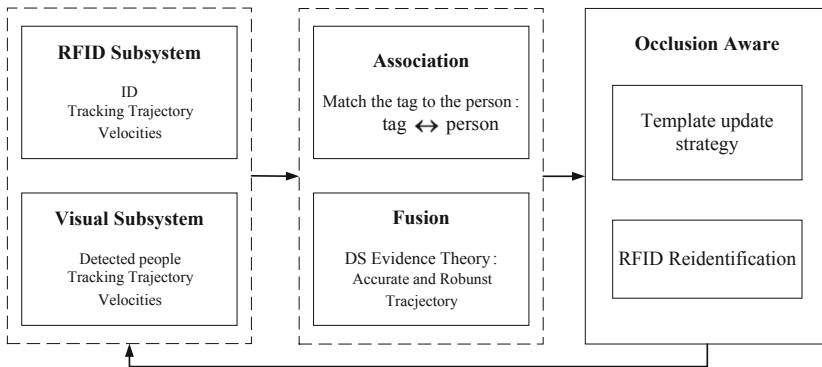


Fig. 1. System architecture

4 Detailed Scheme

4.1 Signal Acquisition

Signal acquisition primarily uses RFID and vision sensors to detect and track objects and obtain their own trajectories. For the RFID subsystem, we need to obtain the tag's ID and the tracking trajectory. We choose ANN [8] to locate RFID tags and get the tracking trajectory for two reasons. First, compared to online learning methods such as Landmarks, ANN only requires reference labels offline for easy deployment. Second,

ANN has a greater impact on nonlinear fitting and positioning accuracy compared to machine learning methods such as SVM [9]. For ANN, in the offline phase, the reference tags are placed 60 cm apart from one another to obtain the training data, which including the RFID signals and the coordinates of the reference tags. In the online phase, the input of the ANN network is the RSS of the four antennas, and the output is the coordinate of the tagged object. Thus, we can get the trajectory of the tagged object.

For the visual subsystem, we need to obtain the detected people and their tracking trajectory. Firstly, to get the detected people, we choose YOLO [16] for pedestrian detection, which has high precision and can run in real time. Secondly, since we need to specifically track the thief, we choose the single object tracking algorithm based on Efficient Convolution Operator (ECO) [17] is used to get the tracking trajectory. ECO is improved on the basis of correlation filter, which is a state-of-the art algorithm. When using the hand-craft features, the speed of ECO can reach 66 fps, which enables real-time tracking. ECO specifies that the template is updated every six frames. But when the model is updated while it is occluded, the template will drift to the occlusion object and lead to false tracking, we will improve the template update strategy in Sect. 4.3.

4.2 Association and Fusion

The two described subsystems provide a set of tags T and a set of visions V , incorporating the identifier and location estimates for every tag and vision, respectively. From these two sets, we want to establish an assignment between individual tags and visions such that a subset T_i is assigned to a particular vision V_j . The described problem can be formulated in a data association context which considers the spatial distance $d_{ij} = \sqrt{x_{ij}^2 + y_{ij}^2}$ between tag T_i and visual V_j . Depending on the localization uncertainty of the RFID and visual system, a zero-mean Gaussian kernel with specific covariance (σ_x, σ_y) is obtained. Then, the spatial distance can be transformed into a probability measure, we have:

$$p_{i,j} = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left(\frac{-x_{i,j}^2}{2\sigma_x^2} + \frac{-y_{i,j}^2}{2\sigma_y^2}\right) \quad (1)$$

On the other hand, the tag and the pedestrian speed are defined as the derivative of the trajectory with respect to time, we have:

$$V_i = \frac{dX_{T_i}(t)}{dt} \quad (2)$$

We can extend them to a function of time, and build an assignment matrix that holds the individual probability measures for each RFID \leftrightarrow visual pair. Tag T_i can then be assigned to the most likely class V_j by finding the maximum value within the sliding window time.

Existing systems and algorithms only focus on visual trajectories, or use simple weighting methods. However, visual tracking is easy to fail due to problems such as occlusion. Therefore, we use DS evidence theory to combine the trajectories of RFID

and visual object based on the priori error distribution to obtain robust and accurate results. DS evidence theory fuses basic probability assignment functions in the identification framework through combination rules and finally makes decisions, which is introduced as follows [18].

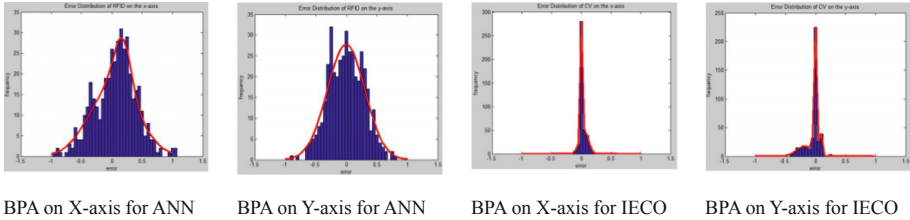


Fig. 2. The basic probability distribution function (BPA) for ANN and IECO without occlusion.

Identification Framework: We define the identification framework $\Theta = \{H_1, H_2, \dots, H_M\}$ as a set of discrete locations within the object space region, which is the mean error distance from the real position.

Basic Probability Assignment Functions (BPA): The probability values of RFID positioning error and visual positioning error in the identification framework are calculated separately. Thus, we get the BPA m_1, m_2 of ANN and IECO under different areas based on the Gaussian distribution. Figure 2 shows the BPA without occlusion, while the BPA under occlusion is discussed in Sect. 4.3.

Combination Rules: The combination rule based on evidence credibility can be defined as follows:

$$\begin{cases} m(H) = p(H) + K \cdot \varepsilon \cdot q(H), \\ m(\emptyset) = 0, \\ m(\Theta) = p(\Theta) + K \cdot \varepsilon \cdot q(\Theta) + K \cdot (1 - \varepsilon) \end{cases} \quad (3)$$

for $H \neq \emptyset, \Theta$, where $p(H)$ represents the combination mass without normalization, and $q(H)$ represents the average support of proposition H . ε is the evidence credibility. We have:

$$p(H) = \sum_{\cap_{i=1}^N H_i=H} m_1(H_1)m_2(H_2) \cdots m_n(H_N) \quad (4)$$

$$q(H) = \frac{1}{N} \sum_{i=1}^N m_i(H) \quad (5)$$

By taking consideration the local conflict degrees, the evidence credibility ε can be calculated. Thus, the introduction of the local conflict concept is the innovation of this modified combination method. We therefore have

$$\varepsilon = e^{-K} \quad (6)$$

$$\bar{K} = \frac{1}{N(N-1)/2} \sum_{i < j < N} K_{ij} \quad (7)$$

$$K_{ij} = \sum_{H_i \cap H_j = \emptyset} m_i(H_i) m_j(H_j) \quad (8)$$

where K_{ij} is the local conflict of evidence m_i, m_j . While \bar{K} is the average conflict degree of all local conflict K_{ij} , ($i, j = 1, 2, \dots, N$), which represents the global conflict of system.

4.3 Occlusion-Aware

Most existing trackers are prone to long-term or short-term occlusion because of obstacles, pedestrians, etc. However, these phenomena will lead to drifting problems and missed the thief. In the proposed method, we use template update strategies and RFID signals to solve occlusion problems in visual tracking.

In correlation filter, the ideal response map should have only one sharp peak and the waveform is smooth in other areas when the detected object is highly matched to the correct object. Otherwise, when the object is occluded or has undergone violent deformation, the entire response map will fluctuate intensely. Therefore, the peak value and the fluctuation of the response map can reveal the confidence degree of the tracking results to a certain extent. Hence, we explore a high confidence feedback mechanism with two criteria. The first criterion is the maximum response score. The F_{\max} of the response map $F(s, y; w)$ can be defined as follows:

$$F_{\max} = \max F(s, y; w) \quad (9)$$

The second criterion is a novel one to measure the Fluctuation of the Waveform named F_w measure. We have

$$F_w = \frac{|F_{\max} - F_{\min}|}{\text{mean}(\sum_{w,h} (F_{w,h} - F_{\min})^2)} \quad (10)$$

where F_{\max} , F_{\min} and $F_{w,h}$ denote the maximum, minimum and the w -th row h -th column elements of $F(s, y; w)$. When the two criteria F_{\max} and F_w of the current frame are lower than their respective historical average values with certain ratios β_1 and β_2 , the correlation filter will not be updated, which can reduce the impact of short term occlusion.

Furthermore, in addition to the above two indicators, when both the distance probability matrix and the velocity probability matrix in the sliding time window are larger than the experimentally set threshold β_3 and β_4 , we can conclude that the object tracking fails due to long-term occlusion. In the sliding time window, we mainly focus

on the BPA of RFID, while the visual BPA is negligible due to occlusion. To correct the occlusion, the search box of the correlation filter is adjusted to the position of the RFID, thus the tracking is performed again.

5 Experimental Results

In this section, we conduct performance evaluation of the hybrid scheme in our lab environment.

5.1 Evaluation Methodology

We conduct indoor experiments in a laboratory (10 m × 7 m) by using an off-the-shelf camera and four RFID antennas with different population sizes, RFID antennas are placed at the center of each side. In the scene, there are multiple individuals, with one of them (the thief) having an attached RFID tag (the tagged object). The surveillance camera is calibrated with known intrinsic parameters, including focal length, lens distortion, and relative rotation and translation with respect to the scene, so that image coordinates can be transferred to world coordinates. Ground truth was annotated by clicking on individuals' head in each image and using calibration information to reconstruct their coordinates. The hybrid system runs on a computer with CPU, and its running speed is 66 fps, which enables real-time tracking. We adopt the error distance, defined as the Euclidean distance between the result and the ground truth, as our base metric.

5.2 Performance of Matching

For the thief identification, we test our ability to identify the thief with multiple people moving in the experimental area. The matching accuracy rate R denotes the success rate of matching the tagged object to the thief, which can be defined as follows:

$$R = \frac{\text{of successful matches}}{\text{of experiments in total}} \times 100\% \quad (11)$$

The experiments in Table 1 show that system has perfect performance, which demonstrates the effectiveness of our association algorithm. Besides, with the number of people increasing, the performance gets worse because people movement cause signal reflection and extra visual noise.

Table 1. Matching accuracy with different numbers of people

# of tracks	1	2	3	4
Matching accuracy	100%	99%	99%	98%
Localization accuracy	10 cm	21 cm	29 cm	39 cm

When matching the thief, the length of the sliding window, i.e., the number of location points, usually affects the success rate of matching. We set the number of locations from 10 to 80 at an interval of 10, and for each case, we test on 2, 3 and 4 trackers respectively. Figure 3 describes how the matching accuracy changes along with the sampling number. When the number of location points increases, the accuracy also tends to increase, i.e., from 59% when the location number is 5 to 98% when the location number reaches 80. In practical experiments, we set the number of points to 60, which not only obtains relatively higher matching accuracy but also reduces the complexity of the experiment.

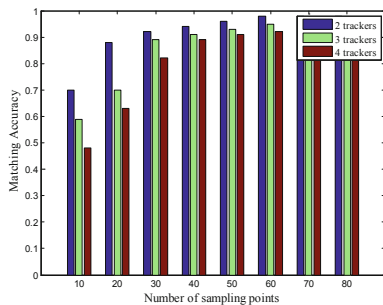


Fig. 3. Matching accuracy ratio vs position number

5.3 Performance of Fusing Algorithm

We then evaluate the thief tracking accuracy of our scheme and compare the performance with other well-known systems. As shown in Fig. 4(a), the path of RFID is dispersed to a large extent, while the video and the fusion paths are closer to the real path. Besides, the Cumulative Distribution Functions (CDF) of the error distance for different methods are shown in Fig. 4(b). For RFID-only method, most of the errors are within 100 cm. While 90% of the errors are less than 30 cm with the fusion method and video tracking.

The mean error distance of our fusion approach is compared with methods using only RFID or camera for people tracking (Table 2). The ANN model has better performance than SVM [8] while the ultimate positioning accuracy is at the decimeter level. In terms of visual methods, the improved ECO (IECO) has higher position accuracy than the original ECO [24]. The fusing approach has better results than the RFID-only and visual-only approach, and its positioning error is 0.097 m, which demonstrates that our fusion method is effective.

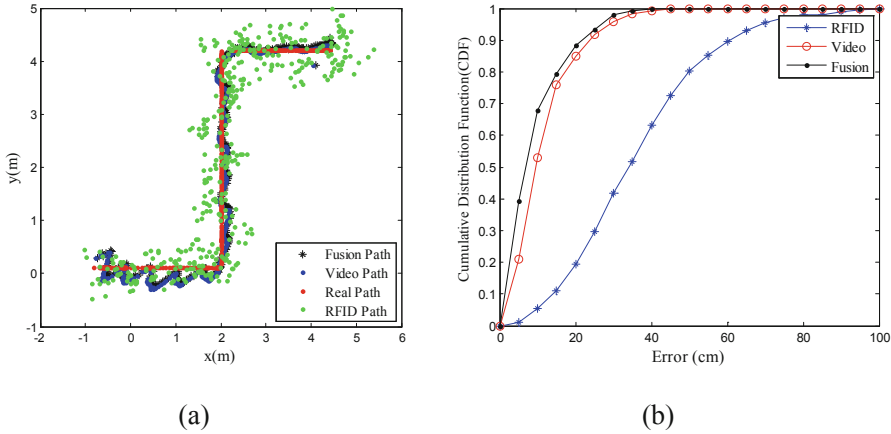


Fig. 4. Comparison of tracking performance when there is no occlusion. (a) Shows the paths corresponding to three different methods and the ground truth. (b) Shows the cumulative distribution function (CDF) corresponding to different methods.

Table 2. Comparison of the mean error distance for different tracking systems

System	Sensor type	Error (m)	Error in x-axis (m)	Error in y-axis (m)
SVM	Passive RFID	0.530	0.310	0.263
ANN	Passive RFID	0.400	0.279	0.234
ECO	Camera	0.124	0.086	0.089
IECO	Camera	0.118	0.063	0.081
Fusion	RFID and camera	0.097	0.039	0.070

5.4 Performance of Occlusion-Aware

The experiments on long-term occluded scenes are also validated. As shown in the first row of Fig. 6, the IECO algorithm drifts after the object is occluded for a long time and then tracks another object. For the fusion algorithm, the object is completely occluded in frame 516 and lost for a certain time. However, after performing the correction using RFID signal, the object is correctly tracked in the 886th frame.

The path of the occlusion experiments is shown in Fig. 5(a), the trajectory of the visual-only (IECO) tracking method is completely erroneous after being occluded, while the RFID and fusion methods do not deviate from the true trajectory. As shown in Fig. 5(b) and Table 3, the mean error distance of the RFID-only method is 0.52 m, and 90% of the errors is within 0.9 m. Although the video path is concentrated, the error is increased due to the tracking error caused by the occlusion. However, the fusion error is 0.2 m, which has better results. Furthermore, 90% of the errors are within 50 cm, which suggest enhanced performance compared to the single sensor. Therefore, our fusion method can still track the thief correctly under long-term occlusion, and has higher tracking accuracy than the single sensor.

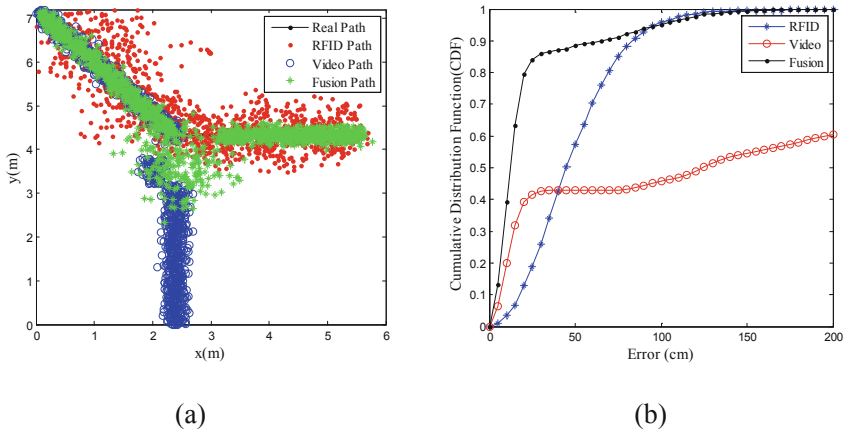


Fig. 5. Comparison of tracking performance under occlusion. (a) Shows the paths corresponding to different methods and (b) shows the cumulative distribution function (CDF) corresponding to different methods.

Table 3. Performance comparison of different tracking systems

System	Sensor type	Error (m)	Error in x-axis (m)	Error in y-axis (m)
SVM	Passive RFID	0.656	0.410	0.346
ANN	Passive RFID	0.52	0.32	0.34
ECO	Camera	2.224	1.886	1.689
IECO	Camera	1.95	1.69	0.94
Fusion	RFID and camera	0.21	0.12	0.15

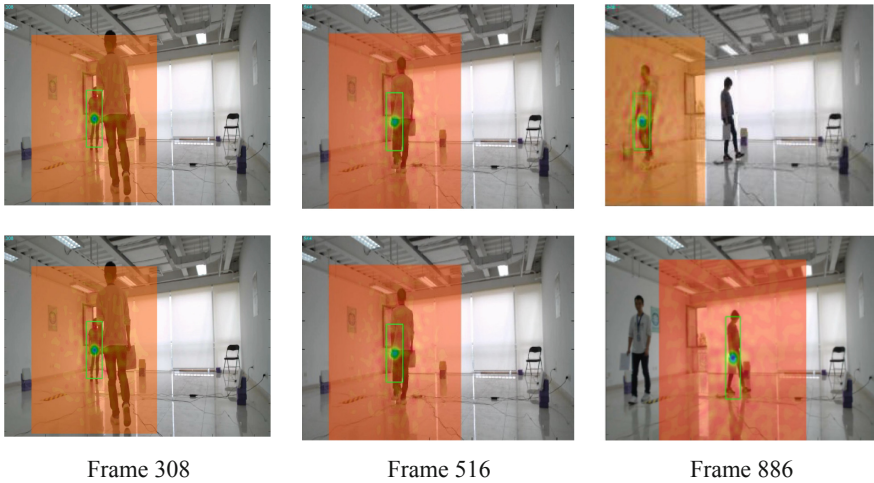


Fig. 6. Comparison of tracking performance under occlusion. The top row is the visual tracking method (IECO), and the second row is the fusion tracking method.

6 Conclusion

In this study, we propose a hybrid system that combines the identities of RFID and high-precision tracking of computer vision for online tracking and identifying. The distance and velocity probability matrix in the time window are used to match the RFID tag and the visual thief. In order to get a robust trajectory, the coordinate information of the two sensors is fused by DS evidence theory based on the prior error distribution. In addition, we use the characteristics of the correlation filter and the signal strength of the RFID to address the occlusion issue. The system performance under different number of people is evaluated, which depicts that the system can achieve fairly good precision and strong robustness, objects can still be tracked correctly under long-term occlusion. Due to the poor positioning performance of RFID in dense crowds, we will focus on improving the positioning performance of RFID in the future.

Acknowledgement. This work was supported by Collaborative Fund of China Electronics Technology Group Corporation (Project No. 6141B08080401).

References

1. Alshammari, A., Rawat, D.B.: Intelligent multi-camera video surveillance system for smart city applications. In: 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, pp. 0317–0323, IEEE (2019)
2. Liang, Y., Cai, Z., Yu, J., Han, Q., Li, Y.: Deep learning based inference of private information using embedded sensors in smart devices. *IEEE Netw.* **32**(4), 8–14 (2018)
3. Zheng, X., Cai, Z., Yu, J., Wang, C., Li, Y.: Follow but no track: privacy preserved profile publishing in cyber-physical social systems. *IEEE Internet Things J.* **4**(6), 1868–1878 (2017)
4. Han, Y., Chen, Z., Guo, T.: Design of equipment anti-theft tracker based on wireless sensor network. In: 2017 First International Conference on Electronics Instrumentation and Information Systems (EIIS), Harbin, pp. 1–5. IEEE (2017)
5. Subedi, S., Pauls, E., Zhang, Y.D.: Accurate localization and tracking of a passive RFID reader based on RSSI measurements. *IEEE J. Radio Freq. Identif.* **1**(2), 144–154 (2017)
6. Dong, X., Shen, J., Yu, D., Wang, W., Liu, J., Huang, H.: Occlusion-aware real-time object tracking. *IEEE Trans. Multimed.* **19**(4), 763–771 (2017)
7. He, S., Chan, S.H.G.: Wi-Fi fingerprint-based indoor positioning: recent advances and comparisons. *IEEE Commun. Surv. Tutor.* **18**(1), 466–490 (2016)
8. Mehmood, H., Tripathi, N.K.: Optimizing artificial neural network-based indoor positioning system using genetic algorithm. *Int. J. Digit. Earth* **6**(2), 158–184 (2013)
9. Figuera, C., Rojo-Álvarez, J.L., Wilby, M., et al.: Advanced support vector machines for 802.11 indoor location. *Signal Process.* **92**(9), 2126–2136 (2012)
10. Dwiyyasa, F., Lim, M.H.: Extreme learning machine for active RFID location classification. In: Handa, H., Ishibuchi, H., Ong, Y.S., Tan, K.C. (eds.) Proceedings of the 18th Asia Pacific Symposium on Intelligent and Evolutionary Systems. PALO, vol. 2, pp. 657–670. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-13356-0_52
11. Henriques, J.F., Caseiro, R., Martins, P., Batista, J.: High-speed tracking with kernelized correlation filters. *IEEE Trans. Pattern Anal. Mach. Intell.* **37**(3), 583–596 (2015)

12. Yuan, J., Chen, H., Sun, F., Huang, Y.: Multisensor information fusion for people tracking with a mobile robot: a particle filtering approach. *IEEE Trans. Instrum. Meas.* **64**(9), 2427–2442 (2015)
13. He, Z., Cai, Z., Cheng, S., Wang, X.: Approximate aggregation for tracking quantiles and range countings in wireless sensor networks. *Theor. Comput. Sci.* **607**(3), 381–390 (2015)
14. Teng, J., Zhang, B., Zhu, J., Li, X., Xuan, D., Zheng, Y.F.: EV-Loc: integrating electronic and visual signals for accurate localization. *IEEE ACM Trans. Network.* **22**(4), 1285–1296 (2014)
15. Duan, C., Rao, X., Yang, L., Liu, Y.: Fusing RFID and computer vision for fine-grained object tracking. In: *IEEE INFOCOM 2017 - IEEE Conference on Computer Communications*, Atlanta, pp. 1–9. IEEE (2017)
16. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: unified, real-time object detection. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, pp. 779–788 (2016)
17. Danelljan, M., Bhat, G., Shahbaz Khan, F., Felsberg, M.: ECO: efficient convolution operators for tracking. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6638–6646 (2017)
18. Wang, J., Liu, F.: Temporal evidence combination method for multi-sensor object recognition based on DS theory and IFS. *J. Syst. Eng. Electron.* **28**(6), 1114–1125 (2017)