TÜBİTAK

Research Article

Full occlusion handling for pedestrian tracking via hybrid system

Xiaoyu ZHANG, Shiqiang HU*, Huanlong ZHANG, Xing HU
School of Aeronautics and Astronautics, Shanghai Jiao Tong University, Shanghai, P.R. China

Abstract: Occlusion and lack of visibility even in sparse crowd scenes make it difficult to track individual pedestrians correctly and consistently, particularly in a single view. We present a novel pedestrian tracking approach that connects tracking with reidentification to locate and maintain the identity of certain people who may be occluded for a long time. First, two models are constructed. One model tracks the pedestrian and trains a classifier, while the other model reidentifies the pedestrian of interest from detection results with the trained classifier. Secondly, we design a set of transition rules for model switching. Finally, the two models work alternatively based on the principle of a hybrid system to track the pedestrian. Several typical sets of experiments show that the proposed approach outperforms the state-of-the-art approaches and achieves robust pedestrian tracking in the presence of full occlusion.

Key words: Pedestrian tracking, full occlusion, ellipsoidal gate, hybrid system

1. Introduction

Pedestrian detection and tracking are important in a wide range of applications in our daily lives, such as surveillance systems, airport security, and human–robot interactions, to name just a few. Pedestrian occlusion is a common phenomenon owing to the limitation of camera views. This problem is challenging since it is unknown where the pedestrian of interest will reappear. In addition, it is difficult to ensure the pose of the pedestrian without any change after occlusion.

Over the past years, tremendous efforts in visual tracking [1–3] have been made to overcome the difficulties that often affect tracking performance, such as occlusion, fast motion, and pose variations. To deal with pedestrian tracking in the presence of occlusion, L1 tracking [4] is proposed by casting tracking as a sparse approximation problem in a particle filter framework. During tracking, a target candidate is represented as a linear combination of the template set composed of both target templates and trivial templates. In the case of occlusion, a limited number of trivial coefficients will be activated, but the whole coefficient vector remains sparse. Fast compressive tracking (FCT) [5] is proposed with an appearance model based on nonadaptive random projections that preserve the structure of the original image space. A very sparse measurement matrix is adopted to efficiently compress features from the foreground targets and background ones. The tracking task is formulated as a binary classification problem with online updates in the compressed domain. It can handle occlusion and pose variations well as the adopted scale invariant appearance model is discriminatively learned from the target and background with data-independent measurements, thereby alleviating the influence of background pixels. Overall, these approaches can deal with partial occlusion, even heavy occlusion, with the help of their distinctive target representation and model update mechanisms. They make full use of partial

---

*Correspondence: sqhu@sjtu.edu.cn

features of occlusion targets, but they may lose their ability in the case of full occlusions as their search mechanisms are limited to deterministic and stochastic methods [1]. In other words, they may reinitialize tracking of a pedestrian of interest who reappears in the same place where he disappeared, but they fail to reinitialize tracking of a pedestrian of interest who reappears in a place far away from the place where he disappeared. Tracking-learning-detection (TLD) [6] explicitly decomposes the long-term tracking task into tracking, learning, and detection. The tracker follows the object from frame to frame. The detector localizes all appearances observed so far and corrects the tracker if necessary. The learning process estimates errors of the detector and updates it to avoid these errors in the future. This approach can directly cope with postfailure behavior by reidentifying the object after occlusion. However, Kalal et al. [6] also pointed out that it does not perform well for articulated objects such as pedestrians. Therefore, TLD cannot robustly track pedestrians in the presence of full occlusion. Papadourakis and Argyros [7] proposed multiple-objects tracking in the presence of long-term occlusions. It is capable of handling several challenging situations. However, it is likely to fail when objects to be tracked move with irregular motion patterns. Yang et al. [8] proposed a probabilistic framework for multitarget tracking with mutual occlusion. It can deal with mutual occlusion well. However, a pedestrian occluded by the background may cause tracking failure. Hua et al. [9] proposed an occlusion and motion reasoning for long-term tracking. It can handle occlusion and viewpoint changes well. However, a limited case of this approach is when an object undergoes occlusion and reappears in a viewpoint that has not been seen before the occlusion.

In this paper, we propose two models and a set of transition rules based on the principle of a hybrid system for pedestrian tracking in the presence of full occlusion. One model (the tracking model) is capable of tracking the pedestrian of interest while training a robust pedestrian classifier. The other model (reidentification model) uses the trained classifier to reidentify the pedestrian of interest from the detection results. A hybrid system [10,11], which has the benefit of encompassing a class of models within its structure, allowing for model switching based on a set of transition rules, is brought into visual tracking to realize the two models alternatively working. Results show that the proposed approach outperforms the state-of-the-art approaches for pedestrian tracking in the presence of full occlusion.

The rest of the paper is organized as follows. Section 2 gives the details of the proposed approach. The simulation results of the proposed algorithm are provided in Section 3. For illustration Section 3 also discusses the performance of the proposed algorithm as compared with FCT, L1APG, and TLD. Finally, conclusions and future work are presented in Section 4.

2. Robust pedestrian tracking based on a hybrid system

The hybrid system is a dynamical system with both discrete and continuous components. These continuous and discrete dynamics not only coexist but also interact. Its evolution occurs both in response to discrete, instantaneous events and in response to dynamics as described by differential or difference equations in time. The system is widely used in the computer science community, modeling and simulation community, and systems and control community [10,11]. The dynamic characteristic of the hybrid system consists of discrete transitions (from one model to another) and a continuous part evolving in the model invariant. This is in line with our aim of constructing two models working alternatively for pedestrian tracking in the presence of full occlusion. The constitution and evolutions of the two models are described in Sections 2.1 and 2.2, respectively. The discrete transitions are presented in Section 2.3. The representations of some variables are given before we describe the approach. $x = [p_w, p_h, \dot{p}_w, \dot{p}_h, w, h]^T$ is the state of the pedestrian of interest with position (p_w, p_h), velocity

$(\dot{p}_w, \dot{p}_h)$, and size (w, h). $z = [p_w, p_h, w, h]$ is the measurement. $\mathcal{C}$ is a classifier trained during the tracking stage and used to classify the pedestrian of interest from the detection results in the reidentification stage. Collectively, x and $\mathcal{C}$ form the continuous state of the hybrid system $X = [x, \mathcal{C}]^T$. Combined with the discrete state ℓ of the hybrid system, the state of the hybrid system can be represented as $\mathcal{X} = [X, \ell]^T$.

2.1. Tracking model ℓ_0

The function of this model is to track the pedestrian of interest while training a classifier for reidentifying the pedestrian of interest from the detection results in model ℓ_1. The differential rule $X_t = \mathcal{D}(X_{t-1})$ consists of the state evolution and the classifier training.

2.1.1. State evolution

The Kalman filter, which is widely used in visual tracking [12], is utilized for pedestrian tracking. In general, the movement of a pedestrian in the scene always conforms to the constant velocity model. Meanwhile, the measurement is a noisy version of the position and size. The state dynamics model follows a dynamic linear Gaussian model, while the measurement model follows a linear Gaussian, i.e.

$$x_t = Fx_{t-1} + Gu_{t-1}, \quad z_t = Hx_t + v_t, \tag{1}$$

where F is the state transition matrix, G is the input matrix, u_{t-1} is the zero-mean Gaussian white process noise with standard deviation σ_u, H is the measurement matrix, and v_t is the zero-mean Gaussian white observation noise with standard deviation σ_v.

2.1.2. Classifier training

The classifier $\mathcal{C}_t$ trained here is the incremental SVM [13]. It enables classifier training in an incremental way as the samples are given frame by frame. We sample positive samples near the current target location and negative samples far away from the object center in each frame. Those samples are compared with the reference template of the pedestrian of interest to obtain similarity scores. The similarity score should be minimum for negative samples and maximum for positive samples to ensure training a reliable classifier. The salience score proposed by Zhao et al. [14] can effectively handle misalignment caused by a large viewpoint and pose variation. The matched pairs will always hold higher similarity than mismatched pairs with this similarity measurement. Based on the discriminative salience scores, a reliable classifier can be trained.

2.2. Reidentification model ℓ_1

The function of this model is to identify the pedestrian of interest from the detection results. The continuous state of the hybrid system in this model does not have evolution. Hence, there is no differential rule in this model. The reidentification process can be described as follows.

After all pedestrians in the scene are detected by a detection method such as ViBe [15] or histograms of sparse codes for object detection [16], the salience score $sim(\mathcal{T}\mathfrak{n}, y)$ in [14] is used to measure the similarity between the reference template $\mathcal{T}$ of the pedestrian of interest and the detected pedestrian y. y is the image patch that corresponds with z. Through $sim(\mathcal{T}\mathfrak{n}, y)$ and the trained classifier $f(sim(\mathcal{T}, \cdot))$ in the tracking model, the pedestrian of interest can be identified from the detection results.

2.3. Discrete transitions

Just as described in hybrid systems [10], the model switching is determined by a set of transition rules. Once the pedestrian of interest is fully occluded, the adopted hybrid system should switch to the reidentification model. The reidentification model works until the pedestrian of interest is reidentified. Given that the detection results at time t are $\{z_t^i\}_{i=1}^N$, the details of the three transition rules are described as follows.

2.3.1. Transition rule τ_1

Assume that $l_{t-1} = \ell_0$. If **Rule 1** is not met, $\mathcal{X}_t$ will evolve according to Eq. (2)

$$l_t = \ell_1, X_t = X_{t-1} \tag{2}$$

Rule 1: $\exists! \ z_t^j$ meets with Eq. (3), and then $\mathcal{X}_t$ is evolved according to Eq. (4).

$$\left(z_t^j - Hx_{t-1}\right)^T S^{-1} \left(z_t^j - Hx_{t-1}\right) \leq \gamma \tag{3}$$

$$l_t = \ell_0, X_t = \mathcal{D}\left(X_{t-1}\right) \tag{4}$$

Here, γ is a threshold, and S is obtained by Eq. (5).

$$P_{t|t-1} = Q + FP_{t-1}F^T, \quad S = HP_{t|t-1}H^T + R$$
$$Q = \sigma_u^2 \begin{bmatrix} \frac{\Delta^4}{4}I_2 & \frac{\Delta^3}{2}I_2 & 0_2 \\ \frac{\Delta^3}{2}I_2 & \Delta^2 I_2 & 0_2 \\ 0_2 & 0_2 & 0_2 \end{bmatrix}, R = \sigma_v^2 I_4 \tag{5}$$

Here, I_n and 0_n denote the $n \times n$ identity and zero matrices, respectively. Δ is the sampling period. P_t is the covariance matrix of state x_t at time t.

In order to prove that Rule 1 is effective for model transition, a validation gate [17] is introduced at first. The validation gate is a region where the next measurement is highly probable to appear. Selecting too small of a gate size may lead to missing the target-originated measurement, whereas selecting too large a size may increase the probability of selecting false observations. The ellipsoidal validation gate is optimal for a linear observation model with additive noise shown in Eq. (1). The region, which is delimited by Eq. (3), is like an ellipse. Hence, this gate is called an ellipsoidal gate. The gate threshold γ for measurement dimension M can be computed efficiently since $(z_t - Hx_{t-1})^T S^{-1} (z_t - Hx_{t-1})$ follows a chi-square probability density function. Thus, for $c\%$, which represents the probability that z_t is the valid measurement (z_t is associated with x_{t-1}), γ can be obtained from Eq. (6).

$$c\% = p\left(\frac{M}{2}, \frac{\gamma}{2}\right) \quad p(a, b) = \frac{1}{\Gamma(a)} \int_0^b e^{-t} t^{a-1} dt \tag{6}$$

If γ is calculated by $c\%$, which approaches 1, the measurement z that falls into the ellipse determined by Eq. (3) is the valid measurement with high probability. For example, if c is set as 97.34% and M is set as 4, γ will be 11. Under these circumstances, the only detection result that falls into the ellipse is the pedestrian of interest with high probability. If no detection result falls into that ellipse, the pedestrian of interest is probably occluded or has disappeared from the scene. If more than one detection result falls into the ellipse, we can almost certainly infer that the pedestrian of interest passes by others.

2.3.2. Transition rule τ_2

Assume that $l_{t-1} = \ell_1$. If $\left\{ y_t^i \middle| f\left(sim(\mathcal{T}, y_t^i)\right) < 0,\ i = 1, 2, \cdots, N \right\}$, l_t and X_t evolve according to Eq. (7).

$$l_t = \ell_1, X_t = X_{t-1} \tag{7}$$

2.3.3. Transition rule τ_3

Assume that $l_{t-1} = \ell_1$. If $f\left(sim(\mathcal{T}, y_t^j)\right) > 0$, l_t and X_t evolve according to Eq. (8).

$$l_t = \ell_0, x_t = [z_t^j(1), z_t^j(2), \dot{p}_w, \dot{p}_h, z_t^j(3), z_t^j(4)]^T, X_t = [x_t, \mathcal{C}_{t-1}]^T \tag{8}$$

Eq. (8) means that the pedestrian of interest is reinitialized and the state of x_t needs to be reinitialized.

The overall framework of the hybrid system in visual tracking is shown in Figure 1. Given that the detection results at time t are $\{z_t^i\}_{i=1}^N$, the main steps of the proposed approach can be expressed as follows:
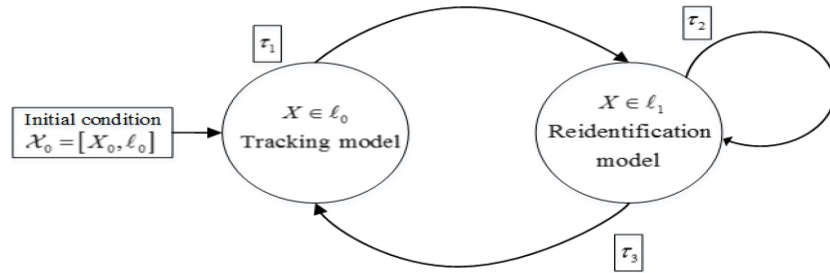

Figure 1. Hybrid system in visual tracking.

1. Assume that the hybrid system works in tracking model $\mathcal{X}_{t-1} = [X_{t-1}, \ell_0]$ at time $t - 1$.

   - If transition rule τ_1 is met, the hybrid system will switch to reidentification model ℓ_1 and the state of the hybrid system will stay in X_{t-1}.

   - If transition rule τ_1 is not met, then $\mathcal{X}_t$ is evolved according to the tracking model ℓ_0 described in Section 2.1.

2. Assume that the hybrid system works in reidentification model $\mathcal{X}_{t-1} = [X_{t-1}, \ell_1]$ at time $t - 1$.

   - If transition rule τ_2 is met, the hybrid system will still work in reidentification model ℓ_1 and the state of the hybrid system will stay in X_{t-1}.

   - If transition rule τ_3 is met, the hybrid system will switch to tracking model ℓ_0 and the continue state of the hybrid system will be reinitialized according to the new reidentified result.

3. Experiments

We evaluate our method with 3 state-of-the-art approaches on ten sequences with full occlusion. Two tests of them contain other moving objects (bicycle or vehicle) as well. Those sequences cover three main challenges in object tracking. The three challenges are interference caused by other similar pedestrians, large pedestrian pose variations, and uncertain direction of motion before and after occlusion. The 3 trackers that we compare with are TLD [6], FCT [5], and L1APG [18]. These methods all have the ability of handling occlusion.

3.1. Experimental setup

The proposed approach is implemented in MATLAB on a Windows PC with an Intel 3.20 GHz Dual Core CPU and 4 GB memory. Assuming that the camera is static, the detection results in each frame are obtained by ViBe [15]. After initiating the pedestrian of interest with a bounding box, our aim is to track this pedestrian robustly in the presence of full occlusion. The measurement dimension M is 4. The search radius for drawing positive samples is set to a quarter of the width of the pedestrian of interest. The inner and outer radii for generating negative samples are set to a half of the pedestrian's width and the pedestrian's width, respectively. The number of positive samples and negative samples in each frame is 25 and 35, respectively. After c is set as 97.34% and M is set as 4, the gate threshold γ calculated by Eq. (6) is 11. This value of γ ensures that the measurement that falls into the gate is the pedestrian of interest with high probability. The standard deviation σ_u of the process noise is set to 3 to ensure that the adopted model can cope with the model's inaccuracy. The standard deviation σ_v of the measurement noise is set to 5 to ensure that the adopted model can cope with the difference between the measurement and the actual position and size of the pedestrian of interest. The sampling period Δ is set according to the frame rate of each test sequence.

3.2. Quantitative evaluation

Center location errors (CLEs) and success rates [5], which are commonly used for quantitative analysis, are shown in Tables 1 and 2. A CLE is defined as the Euclidian distance between the central location of the tracked object and the ground truth. Overlap rate is defined as the area of intersection between the tracking bounding box and the ground truth bounding box, divided by the area of the union of the two boxes. The tracking result in one frame is considered as a success when this score is above 0.5. The success rate is defined as the number of success frames divided by the total number of frames in a video sequence. The lower the CLE is, the better the tracker. The higher the success rate is, the better the tracker. In Table 1, compared with other trackers (FCT, L1APG, and TLD), the CLE of the proposed method is much lower than the best one of the others. For average success rates, we can see that our method obtains the highest value among the four trackers. From Tables 1 and 2, we can see that our method performs best in terms of both average CLE and success rates. To evaluate the performance of our approach comprehensively, we follow the protocol of a benchmark study [1] for quantitative analysis as well. We report the results in one-pass evaluation (OPE) and spatial robustness evaluation (SRE) using the distance precision and overlap success rate in Figure 2. OPE is used to run trackers through a test sequence with initialization from the ground truth position in the first frame and report the average precision or success rate. SRE is used to test the spatial robustness of trackers to an initial bounding box. The larger the area under the curve for each precision (success) plot is, the more accurate the tracker. Figure 2 shows that our approach always performs best among the four trackers in OPE and SRE for pedestrian tracking in the presence of full occlusion. FCT and L1APG can track pedestrians before they are fully occluded. However, they cannot reinitialize tracking of the pedestrian after the pedestrian of interest reappears at a place far away from the place where he disappeared. TLD has the ability of reinitiating. However, due to the pose variations of the pedestrian of interest during movement, both the tracker and detector in TLD often fail to obtain a reliable tracking or detection result. Our method has the ability of training a reliable classifier with discriminative samples in the tracking stage. After the pedestrian of interest is fully occluded, the transition rule in our method switches the tracking model to the reidentification model. Together with the discriminative salience score, the trained classifier can reidentify the pedestrian of interest from detection results. The reliable reinitiating ability makes our method for reinitializing and tracking the pedestrian after full occlusion have a low CLE and high success rate. However, our method may cause tracking failure if only a few frames (such as 6

frames) are given before the pedestrian of interest is fully occluded. Under these circumstances, the incremental SVM cannot obtain a good generalization for reidentifying the pedestrian from the detection results.

Table 1. The average CLE of the evaluated tracks on ten test sequences.

	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6	Test 7	Test 8	Test 9	Test 10
FCT	58	142	3	30	28	51	81	22	44	13
TLD	102	58	17	22	46	61	32	17	22	13
L1APG	28	100	16	32	21	56	51	9	9	15
Ours	15	27	2	8	1	3	2	4	4	10

Table 2. The average success rates of the evaluated trackers on ten test sequences.

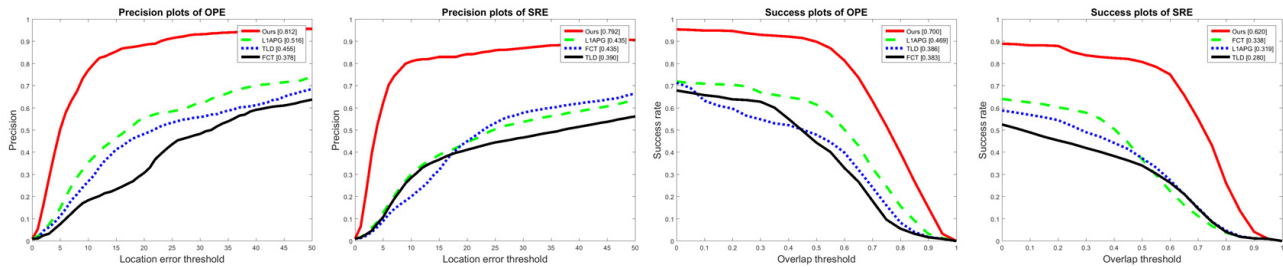|	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6	Test 7	Test 8	Test 9	Test 10
FCT	0.16	0.12	0.95	0.12	0.63	0.60	0.01	0.57	0.43	0.82
TLD	0.44	0.37	0.32	0.65	0.36	0.52	0.70	0.74	0.07	0.60
L1APG	0.21	0.19	0.53	0.54	0.65	0.61	0.53	0.92	0.99	0.97
Ours	0.69	0.65	0.98	0.88	0.99	0.99	0.95	0.99	0.98	0.86

Figure 2. Plots of OPE and SRE. The performance score for each tracker is shown in the legend.

3.3. Qualitative evaluation

In the experiments, the proposed approach could obtain superior tracking results in many sequences. For simplicity, we make qualitative evaluations for the following cases (Figures 3a–3j).

3.3.1. Interference caused by other similar pedestrians

For View-001 of S2L1 [19] shown in Figure 3a, the pedestrian of interest is occluded twice. FCT, L1APG, and our method track the pedestrian before the pillar obscures him. TLD often tracks other pedestrians with similar appearances. During the occlusion from frame #24, all approaches drift. After the pedestrian of interest reappears in frame #39, the proposed approach reinitializes tracking of the pedestrian, while other approaches are still drifting. The second severe occlusion happens from frame #51. After the pedestrian reappears in frame #64, only FCT loses track. L1APG reinitializes the pedestrian of interest because he reappears in its search area.

In order to make it more intuitive for the reader to understand the behavior of the proposed method, we also give the changes of the state in each frame for this test in Figure 4.

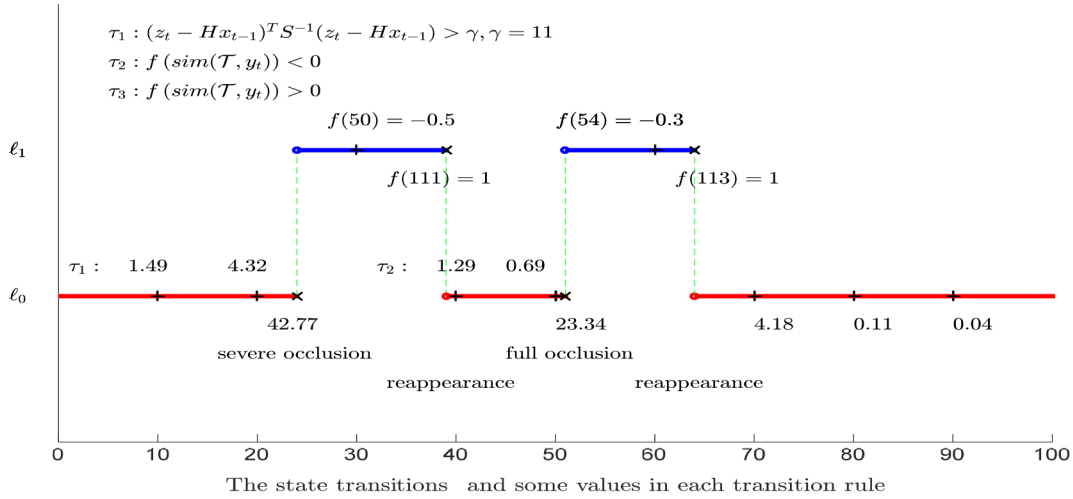**Figure 3.** Qualitative evaluation of four methods on eight image sequences with full occlusion.

Figure 4. The state transitions and some values in each transition rule for test 1.

3.3.2. Large pedestrian pose variation before and after occlusion

For View-007 of S2L1 [19] shown in Figure 3b, the pedestrian of interest goes into the crowd with his back facing the camera and out of the crowd with his front facing the camera. FCT, L1APG, and our method track the pedestrian until frame #37. From frame #38 to #48, the pedestrian goes across two other pedestrians. The pedestrian of interest gets away from them after frame #48 and is reinitialized by the proposed approach and TLD. With the limitation of searching and update schemes in L1APG and FCT, they both fail to reinitialize the pedestrian of interest. When the pedestrian gets out of the crowd with a front view after frame #98, our approach reinitializes him. L1APG seems to reiniate tracking of the pedestrian of interest as well. However, the tracking results of the pedestrian from frame #103 show that the object representation in L1APG has already stored the error information about the pedestrian of interest.

The proposed approach is robust in reidentifying the pedestrian with dramatic pose variations for the following reasons:

- When τ_1 is not violated, the reliable classifier is trained with distinguishing positive and negative samples obtained by the salience score.

- When ℓ_1 works, the salience score between $\mathcal{T}$ and the pedestrian of interest with pose variation can still be classified as a positive sample. This is due to the distinguishing ability of the salience score, which can keep the absolutely higher value with matched pairs than with unmatched pairs.

3.3.3. Uncertain direction of motion before and after occlusion

Figure 3c shows that the pedestrian of interest walks upstairs and is fully occluded for a long time, then reappears at the place where he disappeared. FCT, L1APG, and our method track the pedestrian before he is fully occluded in frame #90. Another pedestrian is used to test the distinguishing capability of the reidentification model in the proposed approach after frame #90. Our approach succeeds in avoiding tracking the other pedestrian. After the pedestrian of interest reappears after frame #311, FCT, L1APG, and our method reinitialize the pedestrian of interest. As this pedestrian walks downstairs, only FCT and our approach can continue tracking him. An example shown here is frame #386. Figure 3d shows that the pedestrian of interest walks past a pillar. The tracking results show that only our approach can successfully track the pedestrian as he reappears after

frame #161. In addition, we can see that the reidentification model in our approach does not classify other pedestrians as the pedestrian of interest in frame #119. It shows the distinguishing capacity of the trained classifier in our method. The proposed method can handle pedestrian tracking with long-term full occlusion and reappearance with uncertain direction of motion for the following reasons:

- Before the pedestrian of interest is full occluded, τ_1 ensures that ℓ_0 works.

- After the pedestrian of interest is full occluded, τ_1 ensures that ℓ_1 works. τ_2 ensures that ℓ_1 keeps working.

- After the pedestrian of interest reappears, the classifier in ℓ_1 reinitializes him based on image feature no matter where he reappears. Then τ_3 reinitializes the state of ℓ_0 for subsequent tracking.

In order to show the robustness of the proposed method, tracking results of another six sequences with full occlusion are also given. From the tracking result of the fourth image in Figure 3j, we can see that our method fails to track the pedestrian of interest as the moving vehicle and the tracked pedestrian are treated as one detection result by ViBe [15]. However, our method can reidentify and track the pedestrian after there is no overlap between the two moving objects (the fifth image in Figure 3j).

3.4. Computational cost

All compared methods are implemented in the Windows platform with an Intel 3.20 GHz Dual Core CPU and 4 GB memory. All compared methods are implemented on a single thread with no GPU acceleration. The image size is 800 × 600 pixels. The average processing speeds for all ten tests and the running environment are shown in Table 3.

Table 3. The average processing speed and running environment.

	Environment	Average speed (frames per second)
FCT	MATLAB & C/C++	8.00
TLD	MATLAB & C/C++	3.35
L1APG	MATLAB & C/C++	3.67
Ours	MATLAB	0.10

As a SIFT feature, patch matching, and many positive and negative samples are used in our method, together with our unoptimized MATLAB code, the average processing speed of our method is very slow. Though the salience score ensures a distinguishing similarity measurement for the matched and unmatched pairs, the computation cost of this score is high. In order to optimize the processing speed of the proposed method, this score can be written in C++ and compiled into a MEX file like other methods, such as TLD. In addition, the methods described at http://www.mathworks.com/matlabcentral/fileexchange/5685 can be used to improve the processing speed of our method.

4. Conclusion and future work

This paper brings a hybrid system into the visual tracking area and realizes a robust pedestrian tracking by model switching in the presence of full occlusion. In detail, we design two models: a tracking model and a reidentification model. The tracking model tracks the pedestrian of interest and trains a reliable pedestrian classifier. The reidentification model uses the trained classifier to reidentify the pedestrian of interest from the

detection results. When the pedestrian is reidentified from the detection results, the state of tracking model is reinitialized. The transitions between the two models are guaranteed by a hybrid system with a set of valid rules. Experimental results and comparisons with other state-of-the-art trackers on a pedestrian walking across a crowd with interference from other pedestrians with similar appearance, large pedestrian pose variations, and uncertain direction of motion before and after full occlusion show the superiority of our approach.

In the future, a simple and effective similarity measurement can be used to replace the salience score in our proposed method. Hence, faster pedestrian tracking can be achieved in the presence of full occlusion.

Acknowledgment

References

[1] Wu Y, Lim J, Yang M. Online object tracking: a benchmark. In: IEEE 2013 Conference on Computer Vision and Pattern Recognition; 25–27 June 2013; Portland, OR, USA. New York, NY, USA: IEEE. pp. 2411-2418.

[2] Yilmaz A, Javed O, Shah M. Object tracking: a survey. ACM Comput Surv 2006; 38: 1-45.

[3] Ekinci M, Gedikli E. Silhouette based human motion detection and analysis for real-time automated video surveillance. Turk J Elec Eng & Comp Sci 2005; 13: 199-229.

[4] Mei X, Ling H. Robust visual tracking and vehicle classification via sparse representation. IEEE T Pattern Anal 2011; 33: 2259-2272.

[5] Zhang K, Zhang L, Yang M. Fast compressive tracking. IEEE T Pattern Anal 2014; 36: 2002-2015.

[6] Kalal Z, Mikolajczyk K, Matas J. Tracking-learning-detection. IEEE T Pattern Anal 2012; 34: 1409-1422.

[7] Papadourakis V, Argyros A. Multiple objects tracking in the presence of long-term occlusions. Comput Vis Image Und 2010; 114: 835-846.

[8] Yang M, Liu Y, Wen L, Zhisheng Y, Li SZ. A probabilistic framework for multitarget tracking with mutual occlusions. In: IEEE 2014 Conference on Computer Vision and Pattern Recognition; 23–28 June 2014; Columbus, OH, USA. New York, NY, USA: IEEE. pp. 1298-1305.

[9] Hua Y, Alahari K, Schmid C. Occlusion and motion reasoning for long-term tracking. In: Springer 2014 European Conference on Computer Vision; 6–12 September 2014; Zurich, Switzerland. Berlin, Germany: Springer. pp. 172-187.

[10] Schaft A, Schumacher H. An Introduction to Hybrid Dynamical Systems. 1st ed. London, UK: Springer, 2000.

[11] Sankaranarayanan S, Sipma H, Manna Z. Constructing invariants for hybrid systems. In: Alur R, Pappas GJ, editors. Hybrid Systems: Computation and Control. Philadelphia, PA, USA: Springer Press, 2004. pp. 539-554.

[12] Chen S. Kalman filter for robot vision: a survey. IEEE T Ind Electron 2012; 59: 4409-4420.

[13] Diehl CP, Cauwenberghs G. SVM incremental learning, adaptation and optimization. In: IEEE 2003 Proceedings of the International Joint Conference on Neural Networks; 20–24 July 2003; Portland, OR, USA. New York, NY, USA: IEEE. pp. 2685-2690.

[14] Zhao R, Ouyang W, Wang X. Unsupervised salience learning for person re-identification. In: IEEE 2013 Conference on Computer Vision and Pattern Recognition; 25–27 June 2013; Portland, OR, USA. New York, NY, USA: IEEE. pp. 3586-3593.

[15] Barnich O, Van Droogenbroeck M. ViBe: A universal background subtraction algorithm for video sequences. IEEE T Image Process 2011; 20: 1709-1724.

[16] Ren X, Ramanan D. Histograms of sparse codes for object detection. In: IEEE 2013 Conference on Computer Vision and Pattern Recognition; 23–28 June 2013; Portland, OR, USA. New York, NY, USA: IEEE. pp. 3246-3253.

[17] Blackman SS, Popoli R. Design and Analysis of Modern Tracking Systems. Boston, MA, USA: Artech House, 1999.

[18] Bao C, Wu Y, Ling H, Ji H. Real time robust L1 tracker using accelerated proximal gradient approach. In: IEEE 2012 Computer Vision and Pattern Recognition Conference; 16–21 June 2012; Rhode, Island, USA. New York, NY, USA: IEEE. pp. 1830-1837.

[19] Bo Y, Nevatia R. Multi-target tracking by online learning of non-linear motion patterns and robust appearance models. In: IEEE 2012 Computer Vision and Pattern Recognition Conference; 16–21 June 2012; Rhode, Island, USA. New York, NY, USA: IEEE. pp. 1918-1925.