IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON URBAN COMPUTING AND INTELLIGENCE

Received June 25, 2019, accepted July 18, 2019, date of publication July 29, 2019, date of current version August 19, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2931820

Particle Filter-Based Prediction for Anomaly Detection in Automatic Surveillance

XINWEN GAO[1], GUOYAO XU[2], SHUAIQING LI[1], YUFAN WU[2], EDVINS DANCIGS[3], AND JUAN DU[2]

[1]School of Mechatronic Engineering and Automation, Shanghai University, Shanghai 200444, China
[2]SHU-UTS SILC Business School, Shanghai University, Shanghai 201208, China
[3]NYU Shanghai Center for Data Science and Artificial Intelligence, New York University Shanghai, Shanghai 200122, China

Corresponding author: Guoyao Xu (xu78guo6@outlook.sg)

This work was supported in part by the Science and Technology Commission Project, Risk Analysis of Urban Viaduct Traffic Safety under Grant 18DZ1201204, and in part by the Science and Technology Commission Project, Rapid Diagnosis Technology for Infrastructure Structure of Urban Expressway Network under Grant 17DZ1204203.

ABSTRACT Automatic surveillance of abnormal events is a major unsolved problem in city management. By successful implementation of automatic surveillance of abnormal events, a significant amount of human resources in video monitoring can be economized. One solution to this application is computer vision technology. This approach utilizes an image processing algorithm to extract specific features and then uses discriminator algorithms to give an alert. In this paper, we propose to apply a particle filter-based algorithm to feature series extracted from videos in order to give alerts when abnormal events occur. The whole process consists of feature series generation and particle filter tracking. To represent the features of a video, an L2-norm extractor is designed based on the optical flow. Then, the particle filter keeps track of these feature series. The occurrence of abnormal events will cause the shift of feature series and a large error in PF tracking. This, in turn, will allow computers to understand and define the occurrences of anomalies. Experiments on UMN dataset show that our algorithm reaches 90% accuracy in frame-level detection.

INDEX TERMS Event detection, particle filters, video surveillance, signal processing algorithms, optical flow.

I. INTRODUCTION

Pedestrian gathering is a principal cause of many serious accidents such as crushing and trampling accidents. One of the biggest challenges in city management is detection of abnormal events in pedestrian activity to boost accident prevention. For now, many information technology-based methods have been proposed to analyze the flow of people in large-scale gathering places [1], determine the number of people [2], detect the people density in local high-risk areas [3], determine the direction and speed of movement, and their changes [4], [5], and provide certain analytical data. Some researchers focused on abnormal behaviors such as stop-and-go crowds and hedging, to support the management department in conduction of judgments and early warnings regarding the flow of pedestrians [3]. By this, they may gather measures for emergency response and guidance.

In this paper, we propose a particle filter-based video anomaly detection framework to automatically alert the occurrence of abnormal events related to pedestrian behavior. With the help of a large amount of existing monocular cameras, we have an easy access to a large amount of video data. In our framework, we use a particle filter to do prediction of L2-norm of image sequence, which assumes that normal event frame is easy to predict while anomaly is hard to. Here we use particle filter (PF) [6] mainly because it has a strong capacity in tracking and it is also very robust against the environment noise with low computation cost, especially compared to the neural network-based methods (neural network methods need Graphics Processing Unit (GPU) to train, but PF even does not need a training step). The detailed explanation will be given in Sec. III. Before PF is used for tracking, we use Farnebäck optical flow algorithm [7] to acquire the motion information in video and extract the L2-norm series to represent the image series for PF to track. Our feature extractor treats one image as a whole and acquires one L2-norm for each image, so it can detect global anomalies.

In the domain of computer vision, anomaly detection is an unsupervised pattern recognition problem which tends

The associate editor coordinating the review of this manuscript and approving it for publication was Maode Ma.

to automatically separate few frames that contain abnormal events from many normal event frames. It is designed as an unsupervised system because in the real world, abnormal events are very rare, and most surveillance videos are normal. This fact makes it hard to train a well-balanced discriminator by simply tagging training data positive or negative. In some cases, there is only normal training data in the dataset. Consequently, the question of how to build a model that learns the normality in video has become both an academic and a practical problem. Obviously, it can greatly improve the efficiency of real security systems by automatically extracting frames of interest and giving alerts when something abnormal happens.

Lots of efforts have been made towards anomaly detection [8]–[11], [18], [20]–[22]. Among these works, a commonly used strategy was to use reconstruction cost comparison between normal training data and abnormal data [8]–[11]. The final decision regarding the value of the discriminator depends on the reconstruction error.

Different from the reconstruction-based idea, author in [18] used prediction error as the discrimination basis. They built a serial model, which fit the training normal data well, to give a prediction image of the future frame. It was based on an assumption that normal event frame is easy to predict while anomaly is hard to.

Another classical approach for this problem is to use probability-based statistical models**.** This method always requires a probability destiny function which fits normal events' features well, and an input of a test sample. Then we can see the abnormal probability returned by this function and determine its normality. This part will be explained in section II. B.

Further, in terms of features extracted from videos before discrimination**,** almost all existing approaches can be roughly separated into two parts:

1) Hand-crafted feature based approaches [1], [8], [11]. These methods use some specific and interpretable ways to extract features as a representation of motion and appearance information. For example, HOF (Histograms of oriented optical flow) [12], HOG (Histograms of oriented gradients) [13] and trajectory. Based on these stable features, dictionary-learning step or some other algorithms are applied to them to identify the judgment baseline of the discriminator. The better features display the difference between normal and abnormal behavior, the better the discriminator will work.

2) Deep learning-based approaches. These methods provide end-to-end solutions to most computer vision problems and have shown the most progressive achievements. Instead of hand-crafted features, deep learning methods often use convolutional neural networks to extract features that carry target information, such as motion and appearance. Then these features will be sent to an auto-encoder and will be enforced to reconstruct the normal behavior with small error. However, because of the strong capacity of deep neural networks, the high error might not be guaranteed on abnormal event videos, which is the major loophole of these methods.

We summarize the contributions to our paper as follows:

1) Inspired by related works [14]–[16], we propose a particle filter-based prediction pipeline for anomaly detection, which tries to tackle the abnormal events by comparing their L2-norm with their expectation. Our solution agrees with the concept of anomaly detection that normal events are predictable while abnormal events are unpredictable.

2) For our framework to achieve real-time prediction based on previous information, we design a simplified way to build the L2-norm series for PF to track. This will be discussed in Sec. III. B. After that, with the sequence data generated from the features, we enforce the predicted value of sequence to be as close as possible to the ground truth's sequence value, which will mostly cause larger deviation when abnormal events occur.

Results: we evaluate our framework based on UMN dataset [17]. The results show that our method can distinguish some specific abnormal events from normal ones, as well as detect the start and the end of anomaly in videos.

II. RELATED WORKS

In the domain of computer vision, with the support of high-performance GPU technology, many new deep learning networks have been proposed, which have made some brilliant progress. We summarize some of the related cases below.

A. DEEP LEARNING-BASED ANOMALY DETECTION

In [18], authors used the whole frame generator (U-net and Generative Adversarial Networks (GAN)) as the predictor, which utilized previous sequence information. Their method considered both appearance and motion information and achieved a new baseline for anomaly detection. In [14], authors proposed a hybrid agent approach to detect anomalous behavior in crowded scenarios, and they divided the behavior into individual and group models, proposing a hybrid agent system with static and dynamic agents which effectively allows to distinguish between individual and group behaviors. They proposed using group behavior as a bag-of-words model to determine the abnormal behaviors of groups by integrating static and dynamic proxy information. In [15], authors used a convolutional neural network (CNN) to generate prediction frames for a given input sequence. To deal with the intrinsic fuzzy prediction obtained from the standard mean square error (MSE) loss function, they proposed three characteristic learning strategies: multi-scale structure, confrontation training method and image gradient difference loss function. In [19], authors studied the problem of activity recognition and abnormal behavior detection in patients with Alzheimer's disease, and studied three variants of re-current neural network (RNN): Vanilla RNN (VRNN), long-term short-term RNN (LSTM) and gated recursive unit RNN (GRU). Activity identification was treated as a sequence tag problem, and abnormal behavior was marked by a deviation from the normal mode. In addition, they also proposed a method to expand the sample set size which could reflect certain behaviors in patients. To summarize, most of

these kinds of methods are a combination of CNN-based feature extractor and RNN-based detection model or GAN based frame generator, which have made great progress, but extremely depend on large amount of computation resources (such as high-performance GPU) as well as large amount of training data.

B. HAND CRAFTED-FEATURE-BASED ANOMALY DETECTION

This kind of method generally uses hand-crafted extractor to extract features and then applies some specific rules to them in order to detect the mode shift of samples.

A commonly utilized method to discriminate anomalies from normal patterns is to use statistical models, which provide probability of normality. For example, in [20], authors used optical flow to extract trajectories, and then modeled information from trajectories into chaotic invariant. Then Gaussian mixture model (GMM) is used to describe the probability density function of the normality based on chaotic invariant feature. Similarly, in [21], authors built an optical flow-based Bayesian model to give an abnormal probability of current motion behavior. In another example [22], authors proposed a social force-model to model the motion behavior between different objects. Then the force flow vector from this model is fed into Expectation Maximization (EM) to give its estimation likelihood which can be further separated by a fixed threshold.

Statistic models see anomaly detection as a problem of probability. The model always tries to fit well a certain distribution of the training dataset D and to give a probability of a test sample y under this fitted distribution. A threshold value is then set to discriminate the anomalies, which can be concluded as the following equation [11].

$$f = \begin{cases} normal & p(y \mid D) \geq \theta \\ abnormal & p(y \mid D) \leq \theta \end{cases} \tag{1}$$

where θ is the threshold.

Besides the statistic models, some sparse coding and dictionary learning-based methods like [8], [11] are also frequently used to divide abnormal patterns from normal ones. The fundamental underlying assumption of these methods is that any regular pattern can be linearly represented as a linear combination of basis of a dictionary which encodes normal patterns on a training set. Therefore, a pattern is considered as an anomaly if its reconstruction error is high and vice versa. Authors used matrix decomposition to extract common basis from training normal samples and used it to build a dictionary. Dictionary is then used to reconstruct the test samples and discriminator would rely on the construction error. The abnormal events cannot fit the bases very well and result in high reconstruction loss, which could be anomalies.

III. PARTICLE FILTER-BASED PREDICTION METHOD

Particle filter is often used in object tracking, robot localization. etc., and is robust on various noises. Here we take advantage of its prediction function to estimate the sequence values. In this section, our vision-based method can be roughly separated into two parts: sequence generation and Particle filter prediction. The following flow chart shows how it works.

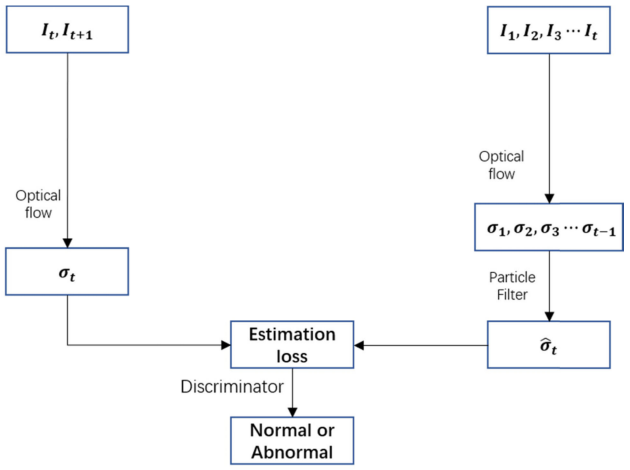


FIGURE 1. The pipeline of our PF-based workflow. "I_1, I_2, I_3, …, I_t" refers to the previous input image sequence. "I_t, I_{t+1}" refers to the two-future input frames. Here we propose an L2-norm generator based on dense optical flow, which will be explained in Sec. III. B. With the input of several frames, particle filter will keep track of their L2-norm and will give estimation value $\hat{\sigma}_t$. This estimation value will then be compared to the real value of next input frame I_{t+1} *to get the estimation loss. Alert would be given by discriminator depend on the estimation loss.* The detail of tracking and discrimination will be explained in Sec. III. C.

Different from some traditional filter algorithms like Kalman Filter [23], PF does not assume any distribution of data. On one hand, it can independently sample distribution parameter and can keep updating it to decrease the residual. On the other hand, some newer methods like PSO-based PF [24] have stronger ability in tracking which is actually not what we want in this application. Because when abnormal events happen, what we want to see is the tracker failure and high residual occurrence. The capacity of some newer filter may enable it to keep tracking but would not give an alert on time.

Our goal in this paper is mainly to test if the particle filter-based prediction is useful for video prediction, as well as to develop an online detection algorithm. So, in our method, to reduce the computational overhead, we simplify each whole frame's motion feature as a specific L2-norm.

A. FEATURES SELECTION

In pedestrian anomalies, movement velocity is a significant factor to indicate the degree of mass. To describe the velocity of objects in video, we think optical flow would be a good choice.

Optical flow is the distribution of apparent velocities which describe movement of brightness patterns in an image. It can give important information about the spatial arrangement of objects and the change rate of this arrangement. In our proposed method, to describe the movement in detail, we use dense optical flow algorithm, which means each pixel in the

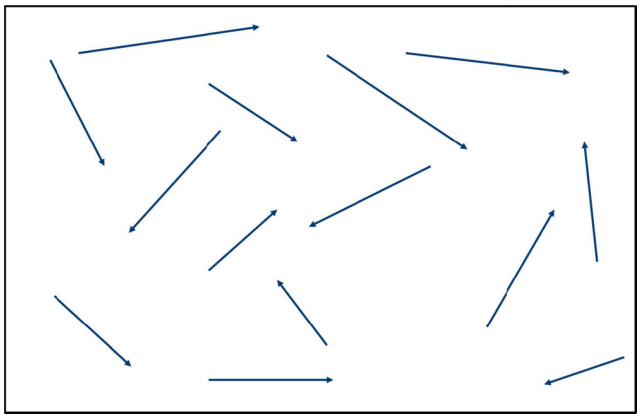


FIGURE 2. Optical flow diagram. Each vector in this diagram refers to a pixel movement. In our proposed method, we use dense optical flow algorithm, which means that each pixel in the image will have a vector as shown in the diagram.

image will be given a certain 2D vector to describe its movement. For two-dimensional image sequences, the optical flow is formulated as the following constraint equation [25]:

$$I_x u + I_y v + I_t = 0 \tag{2}$$

where I_x, I_y and I_t are the derivatives of the image intensity values, along with the x, y and time t dimensions respectively; u, v are the components of the optical flow.

By solving (2), we can get the following result as the final optical flow vector:

$$\begin{bmatrix} u \\ v \end{bmatrix} = -\frac{I_t}{I_x^2 + I_y^2} \begin{bmatrix} I_x \\ I_y \end{bmatrix} \tag{3}$$

The global dense optical flow vectors are displayed as in the following diagram. For each pixel, a vector $\begin{bmatrix} u \\ v \end{bmatrix}$ is given by (3).

B. BUILD THE L2-NORM SEQUENCE FOR PF TO TRACK

For the image-based representation of motion, a common method is to use histograms of optical flow directions weighted with their norm values [26], [22]. Here, we compute global optical flow (OF) using Farneback's method as described in [7]. It calculates optical flow for every pixel in the frame. The OF between two frames is represented in (3).

$$OF\left(I_{t+1}\left(\mathrm{x}, \mathrm{y}\right)\right) = I_t\left(x + u\left(t+1\right), y + v\left(t+1\right)\right) \tag{4}$$

where I_{t+1} refers to one frame of video at instant $t+1$; $u\left(t+1\right), v\left(t+1\right)$ refers to the motion vector between two different instants t;and $t+1$, (x, y) refers to the coordinate of pixels.

Instead of the commonly used HOF (Histograms of Optical Flow) feature, we use the L2-norm as the simplified representation of adjacent frames' motion information, where width and height refer to the width and height of whole image. $\sigma\left(t\right)$ refers to the calculated L2-norm.

$$\sigma\left(t\right) = \sum_{i=1}^{width} \sum_{j=1}^{height} \sqrt{u_{ij}^2\left(t\right) + v_{ij}^2\left(t\right)} \tag{5}$$

By (5), we get one L2-norm for each image. The optical flow extracted from image sequence is converted into a numerical sequence (L2-norm) for PF to track.

Here we summarize this part into the following algorithm.

Algorithm 1 Generation of L2-Norm Before PF Work

Input: Image Sequence $\mathrm{I}_1, \ldots, I_t$
Output: L2-norm $\sigma_1, \ldots, \sigma_t$
Initialize: set sum = 0, OF = 0
1: **For** k = 1 to t
2: OF = Farneback Optical Flow($\mathrm{I_k}$, $\mathrm{I_{k+1}}$)
3: **For** w = 1 to width
4: **For** h = 1 to height
5: sum = sum + sqrt(OF(w,h,u)^2+ OF(w,h,v)^2)
6: **end for**
7: **end for**
8: σ_{k} = sum
9: sum = 0
10: OF = 0
11: **end for**

Additionally, in the line 8 of Algorithm 1, we will compare the calculated σ_t with the estimation value $\hat{\sigma}_{\mathrm{t}}$ by PF and get residual as a discriminator. The estimation $\hat{\sigma}_{\mathrm{t}}$ is calculated by PF, parallelly based on σ_{t-1} in the previous loop. The estimation workflow of PF will be explained in the next section.

It is clear that (5) is a simplified way to get the L2-norm. In contrast, by HOF or convolutional neural network, we can get a more complex feature, which can be represented as a long vector. The PF can also be upgraded to a high dimensional version to keep track of this long vector. But this operation would dramatically increase the computation cost and make it harder to achieve online detection result. For an acceptable processing time in practical use, we tend to use (5) as the main conversion method for now.

C. USE STANDARD PARTICLE FILTER TO ESTIMATE THE SEQUENCE'S VALUE

Particle filter, also known as Sequential Monte Carlo method, is a commonly used technique in signal processing [27]–[29].

The core operation of particle filtering can be summarized as using the state distribution of the previous period to predict the state of the next moment. This is an idea based on Hidden Markov Model (HMM). This means that the particle filtering method assumes that the state of the system is correlated in time, which coincides with the assumptions in the anomaly detection. When a normal event occurs, its temporal correlation makes it easy for the particle filter to give an accurate prediction. The occurrence of anomalous events will break this correlation, making the prediction error of particle filtering to increase, which can be used as a basis for discrimination.

Compared with some classical estimators such as Kalman filter, this method has lower computation cost

in implementation and allows complex nonlinear and non-Gaussian estimation problems to be solved efficiently in an online manner, as well as handling the normal events well. However, the particle filter's generalization ability is much weaker than deep neural network in fitting the data, so that it may always cause a large error in estimating the untrained mode, especially for anomaly detection.

In this paper, considering our target is to accurately predict the normal samples' L2-norm and abnormal events' L2-norm inaccuracy, we use standard particle filter algorithm as described in [29], without any optimization steps.

The standard particle filter consists of two main steps: 1) Importance Sampling (IS). 2) Resampling.

In the first step, to compute an expectation $\mu_f = E[f(x)]$ (written as $\hat{\sigma}_t$ in Sec. III), generally we use the following equation:

$$u_f = \int f(x)\, p(x)\, x \tag{6}$$

To implement (6), we use a series of particles' mean to indicate it. The generation of particles yields a certain posterior Probability Destiny Function [6]. Specifically, we assume that these particles respect a posterior distribution $q(x) \sim N(u, \sigma^2)$, so the computation of (6) can be converted into the following equation:

$$u_f = E[w(x) f(x)], w(x) = \frac{p(x)}{q(x)} \tag{7}$$

For IS (Importance sampling) to be accurate (with a limited number of draws m), $q(x)$ is required to be approximately proportional to $p(x)$ for most x [6]. So, in our experiments, we use some normal frames' L2-norm to adjust the distribution $q(x)$, from which particles come from and we enforce it to be close to the distribution of normal frames' L2-norm.

With $w(x)$ given by (7), a sample of independent draws $x^{(1)}, \ldots, x^{(m)}$ from specific Probability Destiny Function can be used to estimate u_f by the following equation, where $f\left(x^{(i)}\right)$ refers to the value of $x^{(i)}$:

$$u_f = \frac{1}{m} \sum_{i=1}^{m} f\left(x^{(i)}\right) w\left(x^{(i)}\right) \tag{8}$$

In the first step, we use (7) as the future-state prediction rule, which gives mean value of particles. However, we still need a method to measure the accuracy of each u_f, so we obtain it by resampling, as show in step 2.

By learning from [6], we use (9) to measure how close u_f is to σ_t (the L2-norm calculated in (3)). The particles with large variance will be deleted and those with small variance will be copied. The variance of u_f can be represented in this way:

$$V\left(u_f\right) = \sqrt{\frac{1}{m} \sum_{i=1}^{m} \left(\frac{w_k^i}{\bar{w}_k} - 1\right)^2}, \bar{w}_k = \frac{1}{m} \sum_{j=1}^{m} w_k^j \tag{9}$$

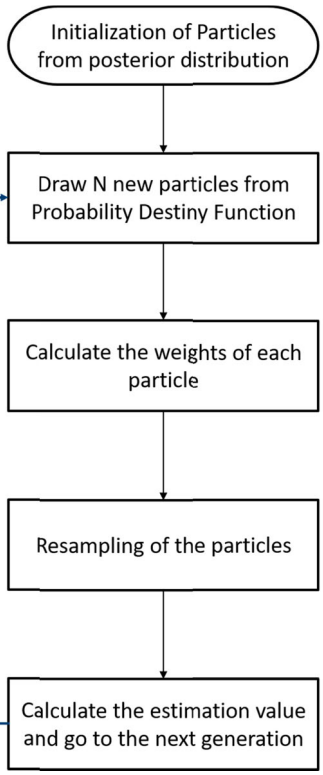


FIGURE 3. Particle filter algorithm workflow. Importance Sampling is composed of the second and third step. The fourth step refers to resampling.

where w_k^j refers to the jth particle's weight in the k round sampling. The value given by (9) is also called weight. Based on the weight value, we will perform resampling step.

The resampling method used in our experiments is called systematic resampling. The key target in resampling method is to determine whether to replicate specific particle $x^{(i)}$ for the next prediction step. Such a decision is made by the following several equations:

$$\mu_s \sim U\left[0, \frac{1}{N}\right) \tag{10}$$

$$\mu^i = \frac{i-1}{N} + \mu_s \tag{11}$$

where $U\left[0, \frac{1}{N}\right)$ refers to a uniform distribution, N refers to the number of particles. In this paper, N = 100.

Then selecting particle $x^{(j)}$ for replication, such that:

$$\mu^i \in \left[\sum_{p=1}^{j-1} V^P, \sum_{p=1}^{j} V^P\right) \tag{12}$$

where V is given by (8) for each particle $x^{(j)}$. Every time when μ^i is satisfied (12) for a specific particle $x^{(j)}$, such $x^{(j)}$ will be replicated one time. And after the whole traversal of indexes *i,* the particle will be deleted, without even one time of replication.

To show the workflow of PF, we conclude it into Fig. 3.

FIGURE 4. The upper figure is the normal frame of UMN dataset, the lower figure is the abnormal frame of UMN dataset.

FIGURE 5. Detection results on the first event of scene 1. The detected result starts from 525^{th} to 615^{th} frame. The ground truth starts from 526^{th} to 615^{th}. The first 300 values are input as the training data, which means they are used to extract statistical parameter $\left(u, \sigma^2\right)$. The value series are then normalized using this parameter from the same series. Further, the parameter $\left(u, \sigma^2\right)$ is also used to initialize the distribution of particles. Additionally, such training and initializing can be designed as an automatically conducted step, which means our method can be used to handle real-world surveillance. It is because the model parameter for particle filter can be easily extracted without extra supervised operations.

D. EXPERIMENTS AND DISCUSSION OF PARTICLE FILTER-BASED METHOD

In this part, we evaluate our proposed method on UMN dataset [17] and some real-world scene examples. UMN dataset is used to test if our method can separate the global abnormal events from normal samples and real-world example is used to show its effectiveness in practice. UMN dataset consists of three different scenes of crowded escape events and the total amount of frames is 7739 (1450, 4415 and 2145 for scenes 1∼3, respectively) with a 320∗240 resolution. Its normal events are pedestrians walking randomly filling the whole screen, and the abnormal events are pedestrians swiftly running almost at the same time. There are total of 11 abnormal events in the whole dataset.

In experiments, we use the frame-level detection accuracy as our evaluation metric, which means tagging a 0 or 1 for each frame. In the detection step, once the estimation error is five times larger than all errors that ever occurred in the first several normal frames, the frame would be judged as the start of an abnormal event.

On the contrary, detection of the end of the abnormal event could pose a problem on our method because the strong tracking ability of particle filter may keep track of the consistent motion mode and could fail to give an accurate end time. However, in our experiments, we find that our proposed method can give a very accurate end of events in every scene of UMN dataset. We think this is because the motion mode in UMN dataset has very dramatic change. Pedestrians change their status very swiftly. It makes the optical flow L2-norm change dramatically at the same time, which causes the distribution of L2-norm to have a big difference. Thus, the PF cannot track it very well which could cause a big estimation error in the accuracy of the end time of abnormality.

In terms of the initialization of the PF parameter, we initialize the tracker parameters from the first several frames of each scene, enforcing particles to fit the data well, and leave the others for testing.

As described in (7), we get the target distribution $q(x) \sim N\left(u, \sigma^2\right)$ from normal frames. All frames in the same sequence use the same $\left(u, \sigma^2\right)$ to preprocess the L2-norm extracted by Algorithm 1. The preprocess step uses (13) to enforce input data to be under the same distribution.$\sigma_{\text{test}}^{'}$ refers to the normalized L2-norm:

$$\sigma_{\text{test}}^{'} = \frac{\sigma_{test} - u}{\sigma^2} \tag{13}$$

To be specific, among all the image sequences, the first 300 frames of each sequence are used to extract the statistical parameter $\left(u, \sigma^2\right)$, then all values in this sequence will be normalized by the same $\left(u, \sigma^2\right)$. Further, the PF tracker for this sequence will be initialized by this $\left(u, \sigma^2\right)$.Such a normalization and initialization step can be easily conducted by script, which means our method can be executed in an automated way, without any extra supervised operations.

Based on the steps mentioned above, we implement experiments on the whole UMN dataset and achieve the detection result as shown in Fig. 5, Fig. 6 and Fig. 7. Some additional examples under practical scene are shown as Fig. 8 and Fig. 9. In Fig. 5, the detected event starts from 525^{th} to 615^{th} frame while ground truth is 526^{th} - 615^{th}. In Fig. 6, the detected

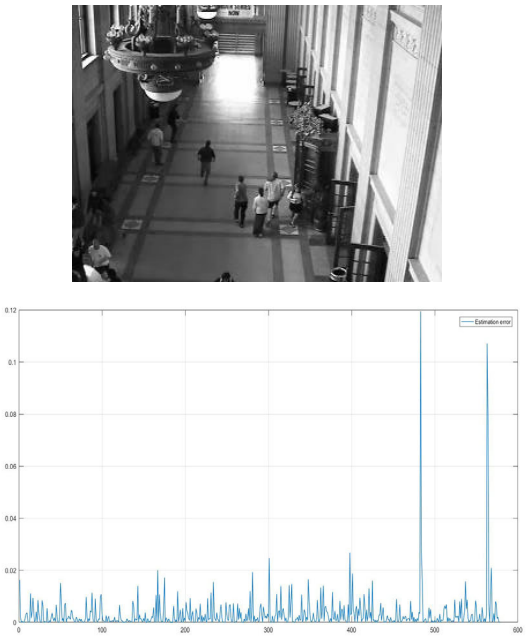


FIGURE 6. Detection results of the first event of scene 2. The detected result starts from 484th to 563th frame. The ground truth starts from 484th to 563th. From our detection result, we also find that the results of PF estimation error somewhat act like impulse response. We think this is a good result because it indicates that the motion mode really has a swift shift at that moment and our proposed method is very sensitive to motion shift. The sudden shift of motion mode can be well captured based on the error of PF estimation. The detected frames' example and the serial results are as shown below.

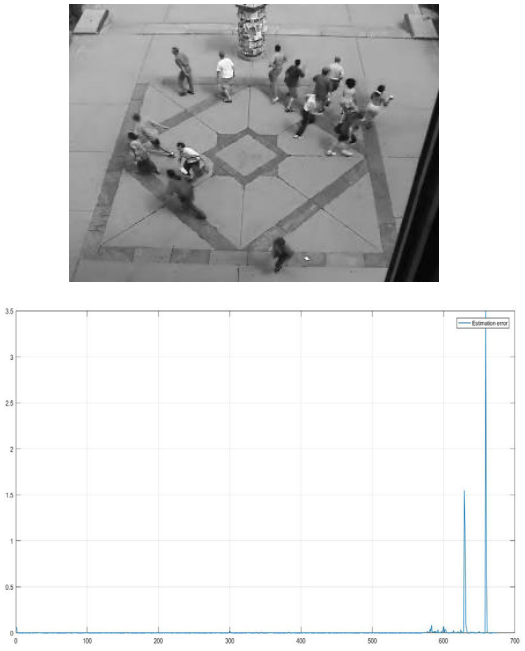


FIGURE 7. Detection results on the first event of scene 1. The detected result starts from 629th to 659th frame. The ground truth starts from 628th to 659th frames.

event starts from 483th to 563th frame while ground truth is 484th to 563th. In Fig. 7, the detected event starts from 628th to 659th frame while ground truth is 630th to 659th.

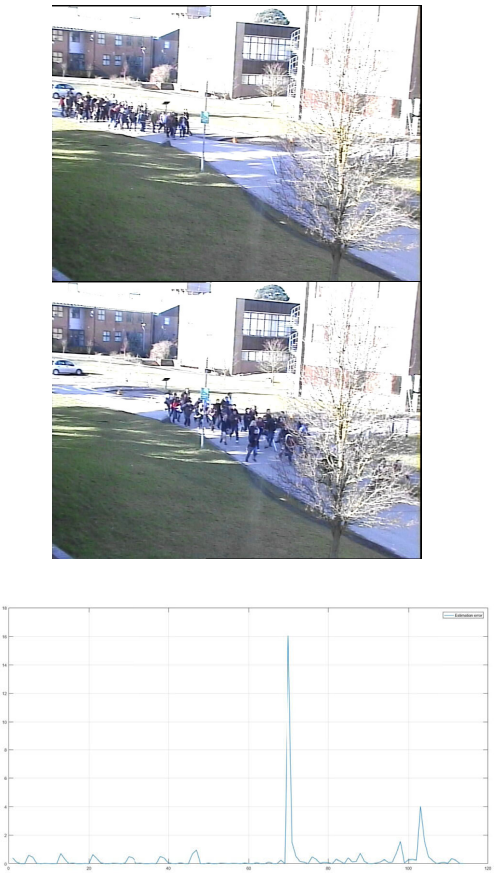


FIGURE 8. Detection result of a practical escaping event. The detected result starts from 71th to 110th. The ground truth starts from 70th to 113th.

To give a comparison with other state-of-art methods, we collect statistics from the detection result and list them in Table 1. The proposed method is compared with other state-of-art hand-crafted feature based methods and our previous work [30], including BM [21], CI [27], SF [22], SRC [11], PF [31]. The method stated in this paper is an improved version of our previous work. All these methods were previously tested on UMN dataset, and we used the provided data in these papers as a comparison. Results are shown in Table 1 below.

Table 1 shows the accuracy comparison of five methods for three different scenes of UMN dataset. Overall, our proposed method achieves the second-best accuracy with an average accuracy rate of 89.89%, which is higher than the accuracy of CI (87.91%), SF (85.09%), SRC (84.70%), but lower than BM (96.40%). We observe that the proposed method performs well on Scene 1 and Scene 3 but poorly on Scene 2. We carefully watch the video and find that this is because scene 1 and scene 3 are outdoor scenes but scene 2 is an indoor scene. The indoor activity is always limited by the space between furniture, so that pedestrians cannot move very quickly and the separation between optical flow from different motion mode becomes much more complex. Therefore, all methods in table 1 got worst accuracy on scene 2.

FIGURE 9. Another example is a real-world traffic scene, which shows our approaches' practical effectiveness. There are no anomalies in this video. The first 500 frames are training data. In the whole sequence, all error is smaller than that which occurred in training data, so it is judged as a normal video.

Nevertheless, even though BM got the best accuracy, it has a defect in its environment adaptability. Because the Bayesian model that was used for modeling of location, magnitude and direction of foreground objects, the trained model can only be used in the same environment as the training data (the change of background may cause the distribution of data to be different). This is because the Bayesian model is very sensitive to data distribution variation. Any difference of training data and test data in distribution will lead to a bias problem. From the same perspective, our proposed method does not have such a concern because it just extracts some simple statistic features from training data like mean and variance. Such an extraction is also easy to be automatically implemented with script, which means our proposed method can somewhat adapt the environment change while BM cannot (training of Bayesian model needs carefully tuning of parameters and is difficult to automate).

TABLE 1. Accuracy (%) comparison of the proposed method with BM, CI, SF, SRC in the UMN dataset.

	Scene 1	Scene 2	Scene 3	Overall Accuracy
Proposed Method	99.00	82.74	**98.11**	89.89
BM	**99.03**	**95.36**	96.63	**96.40**
CI	90.62	85.06	91.58	87.91
SF	84.41	82.35	90.83	85.09
SRC	90.52	78.48	92.70	84.70
PF	96.77	70.32	92.12	86.40

In practical use, we also consider the probable detection result at nighttime. Since the feature series used for detection are built based on optical flow, we can conclude that the accuracy of optical flow determines the final effectiveness. We found that recent advances in optical flow can assure its high accuracy, so we think the effectiveness of our proposed method at nighttime is as good as it is in daytime.

From the impulse response of the test results, we also found that this coincides with our assumptions about how particle filters work. The occurrence of anomalous events destroys the HMM hypothesis of particle filtering, which makes the tracker unable to give accurate tracking results at a certain moment, thus providing our discriminator with a basis for determining abnormal events.

In terms of the running time, our framework is implemented with MATLAB 2016b on CPU i5-7300HQ. We analyze the computation cost of our framework see that the computation of global optical flow cost the most time. Since all of the methods discussed above used optical flow as a motion information extractor, the difference in computation cost is only contributed by the post-processes of extractor and the discriminator algorithm. For post-processes of extractor, our computation complexity of L2-norm process is O(w∗h) (width and height of an image). For the discriminator algorithm, complexity of standard particle filter is $O(n^2)$ (n here refers to the number of particles, in our experiments n=100) [6]. By combining these algorithms, we get a little decrease in accuracy, but we achieve a much better average running time of 20 fps, while in BM the average running time is only 1 fps. This shows that our framework runs faster than the state-of-art method (BM).

IV. CONCLUSION

Based on an assumption that normal event frame is easy to predict while anomaly is hard to, we propose a particle filter-based method to automatically detect the abnormal events in videos. We characterize crowd motion by optical flow and construct the L2-norm series by extracting optical flow of each image. Then PF is used to track the value series.

Once the tracking error is 5 times larger than for the training frames, the error occurrence time will be seen as a start or end of an abnormal event. Usually, the first and last large error time means the start and end of an event. The experiments on UMN dataset and real-world video show that our proposed method is effective.

In anomaly detection domain, prediction-based method is a new approach for solving such a problem. Compared to the reconstruction-based method, this kind of approach has a much more reasonable assumption that normal events are easier to predict than the abnormal ones. Inspired by such a new idea, we allocate the particle estimator to this problem. Our assumption was that, as the predictor takes advantage of both history information and the newest observation, it might work well in predicting the normal events' L2-norm but might not work well on the sudden occurrences of abnormal events' L2-norm**.** We also considered that because of the strong capacity of particle filter in tracking, even though the unexperienced mode occurs, the estimator may still work after a short period of inaccurate estimations. Consequently, we utilize such a short time of inaccuracy as the judging criteria for detection of abnormal events. In our future work, we plan to add more spatial feature extractors at the front end of our pipeline, which can enable our method to give the location of anomalies in a single frame.

ANNOUNCEMENTS

This paper is an addition to our previous work "Particle filter-based prediction for anomaly detection" [31], which has been published as a conference paper of ATCI2018(2018 International conference of Applications and Techniques in Cyber Intelligence). Compared to the previous paper, we rewrite the whole paper by adjusting the application of PF algorithm, conducting necessary comparison experiments for practical use and adding more detailed discussion.

REFERENCES

[1] F. Tung, J. S. Zelek, and D. A. Clausi, "Goal-based trajectory analysis for unusual behaviour detection in intelligent surveillance," *Image Vis. Comput.*, vol. 29, pp. 230–240, Mar. 2011.

[2] W. Song, J. Luo, Y. Zhou, J. Lin, and Z. Shu, "Method of real-time traffic statistics using mobile network signaling," *Appl. Res. Comput.*, vol. 31, no. 3, pp. 776–779, 2014.

[3] S.-W. Chen, Y.-D. Bian, F. Hu, and C. Wang, "Population surveillance and trend alert analysis of mass gatherings," *Chin. J. Netword Inf. Secur.*, vol. 12, pp. 45–53, Dec. 2017.

[4] J. Yick, B. Mukherjee, and D. Ghosal, "Analysis of a prediction-based mobility adaptive tracking algorithm," in *Proc. 2nd Int. Conf. Broadband Netw.*, vol. 1, Oct. 2005, pp. 753–760.

[5] S. C. Mukhopadhyay, "Wearable sensors for human activity monitoring: A review," *IEEE Sensors J.*, vol. 15, no. 3, pp. 1321–1330, Mar. 2015.

[6] F. Wang, M. Lu, Q. Zhao, and Z. Yuan, "Particle filtering algorithm," *Chin. J. Comput.*, vol. 37, no. 8, pp. 1679–1694, 2014.

[7] G. Farnebäck, "Two-frame motion estimation based on polynomial expansion," in *Proc. Scand. Conf. Image Anal.*, vol. 2749. Springer-Verlag, 2003, pp. 363–370.

[8] C. Lu, J. Shi, and J. Jia, "Abnormal event detection at 150 FPS in MATLAB," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 2720–2727.

[9] B. Zhao, L. Fei-Fei, and E. P. Xing, "Online detection of unusual events in videos via dynamic sparse coding," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2011, pp. 3313–3320.

[10] M. Hasan, J. Choi, J. Neumann, A. K. Roy-Chowdhury, and L. S. Davis, "Learning temporal regularity in video sequences," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 733–742.

[11] C. Yang, J. Yuan, and J. Liu, "Abnormal event detection in crowded scenes using sparse representation," *Pattern Recognit.*, vol. 46, no. 7, pp. 1851–1864, 2013.

[12] R. Chaudhry, A. Ravichandran, G. Hager, and R. Vidal, "Histograms of oriented optical flow and Binet-Cauchy kernels on nonlinear dynamical systems for the recognition of human actions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 1932–1939.

[13] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Jun. 2005, pp. 886–893.

[14] S.-H. Cho and H.-B. Kang, "Abnormal behavior detection using hybrid agents in crowded scenes," *Pattern Recognit. Lett.*, vol. 44, pp. 64–70, Jul. 2014.

[15] M. Mathieu, C. Couprie, and Y. LeCun, "Deep multi-scale video prediction beyond mean square error," *CoRR*, vol. abs/1511.05440, pp. 1–14, Nov. 2015.

[16] W. Sultani, C. Chen, and M. Shah, "Real-world anomaly detection in surveillance videos," *CoRR*, vol. abs/1801.04264, pp. 1–10, Jan. 2018.

[17] *Unusual Crowd Activity Dataset Made Available by the University of Minnesota*. [Online]. Available: http://mha.cs.umn.edu/Movies/Crowd-Activity-All.avi

[18] W. Liu, W. Luo, D. Lian, and S. Gao, "Future frame prediction for anomaly detection—A new baseline," *CoRR*, vol. abs/1712.09867, pp. 1–10, Dec. 2017.

[19] D. Arifoglu and A. Bouchachia, "Activity recognition and abnormal behaviour detection with recurrent neural networks," in *Proc. FNC/MobiSPC*, 2017, pp. 86–93.

[20] S. Wu, B. E. Moore, and M. Shah, "Chaotic invariants of Lagrangian particle trajectories for anomaly detection in crowded scenes," in *Proc. IEEE Computer Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 2054–2060.

[21] S. Wu, H.-S. Wong, and Z. Yu, "A Bayesian model for crowd escape behavior detection," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 1, pp. 85–98, Jan. 2014.

[22] R. Mehran, A. Oyama, and M. Shah, "Abnormal crowd behavior detection using social force model," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 935–942.

[23] E. A. Wan and R. V. Merwe, "The unscented Kalman filter for nonlinear estimation," in *Proc. IEEE Adapt. Syst. Signal Process., Commun., Control Symp.*, Oct. 2000, pp. 153–158.

[24] W. Jing, H. Zhao, C. Song, and D. Liu, "A optimized particle filter based on PSO algorithm," in *Proc. Int. Conf. Future BioMed. Inf. Eng. (FBIE)*, Dec. 2009, pp. 122–125.

[25] T. Wang and S. Hichem, "Histograms of optical flow orientation for abnormal events detection," in *Proc. IEEE Int. Workshop Perform. Eval. Tracking Surveill. (PETS)*, Jan. 2013, pp. 45–52.

[26] M. Baccouche, F. Mamalet, C. Wolf, C. Garcia, and A. Baskurt, "Sequential deep learning for human action recognition," in *Proc. Int. Workshop Hum. Behav. Understand.*, 2011, pp. 29–39.

[27] T. Zhang, S. Liu, C. Xu, B. Liu, and M.-H. Yang, "Correlation particle filter for visual tracking," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 2676–2687, Jun. 2018.

[28] F. Wang, B. Lin, and X. Li, "An ant particle filter for visual tracking," in *Proc. IEEE/ACIS 16th Int. Conf. Comput. Inf. Sci. (ICIS)*, May 2017, pp. 417–422.

[29] J. Vermaak, C. Andrieu, A. Doucet, and S. J. Godsill, "Particle methods for Bayesian modeling and enhancement of speech signals," *IEEE Trans. Speech Audio Process.*, vol. 10, no. 3, pp. 173–185, Mar. 2002.

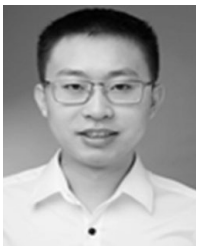
[30] A. Dosovitskiy, P. Fischer, E. Ilg, P. Hausser, C. Hazirbas, V. Golkov, P. van der Smagt, D. Cremers, and T. Brox, "FlowNet: Learning optical flow with convolutional networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 2758–2766.

[31] X. W. Gao, G. Y. Xu, and Y. F. Wu, "Particle-filter-based prediction for anomaly detection," in *Proc. Int. Conf. Appl. Techn. Cyber Secur. Intell. (ATCI)*, 2018, pp. 546–553.

XINWEN GAO received the M.E. degree in computer application from Guizhou University, Guizhou, China, and the Ph.D. degree in control theory and control engineering from Shanghai University, Shanghai, China, where he is currently a Teacher with the School of Mechatronic Engineering and Automation. His research interests include artificial intelligence and robot control.

GUOYAO XU received the B.S. degree in electronic and information engineering from Shanghai University, where he is currently pursuing the M.S. degree in information management of urban public facilities. His research interests include computer vision and image processing. Recently, he has been focusing on solutions for computer vision applications in automatic surveillance.

SHUAIQING LI received the B.S. degree in mechatronic engineering from the Shandong University of Technology, China, in 2016. He is currently pursuing the M.S. degree in mechanical engineering with Shanghai University, China. His research interests include computer vision and its applications in facility management.



YUFAN WU was born in Xining, Qinghai, China, in 1991. He received the B.S. degree in civil engineering from the Huazhong University of Science and Technology, Wuhan, in 2014. He is currently pursuing the M.S. degree in urban public facilities information management with Shanghai University, China.

From 2017 to 2019, his research has been focused on how to use machine learning to solve real problems of management of urban public facilities. His research interests include computer vision with deep learning and automobile OBD data mining.

EDVINS DANCIGS was born in Latvia. He is currently pursuing the B.S. degree in computer science with New York University Shanghai, China. He speaks Latvian, Russian, English, and Chinese. His previous research has been focused on how to apply blockchain technology in the field of intellectual property. His research interests include artificial intelligence and blockchain technology.

JUAN DU received the M.Sc. degree in information system and management from Warwick University and the Ph.D. degree in management sciences and engineering from the Shanghai University of Finance and Economy. She is currently a Lecturer with the SHU-UTS SILC Business School, Shanghai University, where she is also a Researcher with the Shanghai Urban Construction Industry Research Center. Her current research interests include building information modeling and ontology-based information integration.