Hindawi
Mathematical Problems in Engineering
Volume 2020, Article ID 8858416, 9 pages
https://doi.org/10.1155/2020/8858416

Research Article

Multiobjects Association and Abnormal Behavior Detection for Massive Data Analysis in Multisensor Monitoring Network

Ke Bao and Yourong Ding

Wuxi Institute of Technology, Wuxi 214121, Jiangsu, China

Correspondence should be addressed to Ke Bao; baoke@wxit.edu.cn

Received 14 September 2020; Revised 7 October 2020; Accepted 19 October 2020; Published 3 November 2020

Academic Editor: Yi-Zhang Jiang

With the rapid increase in the number of large-scale distributed cameras and the rapid increase in the monitoring range of the camera network, how to accurately recognize and analyze abnormal behavior is still a challenging problem. In addition, the appearance of moving objects between different cameras without overlapping fields of view undergoes significant changes, making it difficult to obtain accurate association Therefore, multiobjects association and abnormal behavior detection for massive data analysis in multisensor monitoring network are proposed in this paper, which firstly uses belief propagation to associate multiple objects, extracts the object's behavior trajectory characteristics, and then builds a long short-term memory classification network to realize automatic classification of abnormal behaviors. Multiobject association fully considers the timing correlation and object detection probability, as well as the statistical dependence of the measurement on the association matrix. The experimental results show that our proposed method can achieve a high classification accuracy and sensitivity, which meets the requirements of automatic classification of abnormal behavior in complex monitoring network. This further shows that this research has practical application value.

1. Introduction

The key technology of building an intelligent surveillance system is to automatically process the video images captured by the multisensors and understand the semantic information [1]. The development of computer vision technology provides necessary methods and means for intelligent video analysis [2]. However, for multisensor monitoring network, how to intelligently detect abnormal events and recognize the identity of suspicious objects is the core issue of video monitoring [3].

Abnormal events are different from the normal events in the monitoring scene, such as pedestrian chasing, gathering, or fighting. As an effective protection method, abnormal event detection can accurately detect abnormal events in the surveillance video and give timely warnings, which can avoid the further deterioration of the accident [4]. For the suspicious object in the surveillance video, multiobjects association technology can accurately locate the suspicious object's location or trajectory, so as to confirm its identity and behavioral tendency. Therefore, the intelligent analysis based on abnormal event detection and object recognition technology can ensure public safety and promotes the development of safe cities and smart cities [5].

At present, the academia treats the abnormal event detection task as a classification rather than a traditional detection task [6–10]. Through local-feature extraction and behavior-model construction, it can predict the category of the event. With the development of intelligent signal processing technology, deep-learning models represented by convolutional networks have achieved excellent performance in vision applications [7, 9]. The training of deep-learning models generally requires positive and negative samples, but most of the samples in the surveillance video only contain normal behavior information [8]. The frequency of abnormal events is very low, and there are also many types of abnormal events. It is difficult to give an accurate description of abnormal events, which will lead to the fact that most of the

training samples available during model training are positive samples. In other words, there are samples that only contain normal events, and, even in most cases, only positive samples are available. Therefore, the abnormal event detection task in surveillance video can only be regarded as a single-category classification task in most cases. The main solution is to firstly establish a model of normal events, and samples that do not conform to normal events are abnormal events [11]. How to extract the characteristics of normal events and construct an effective normal event model becomes the key to abnormal event detection.

In addition, object association is one of the key technologies in multisensor monitoring network, and its main purpose is to judge whether the feature information of multiple local sensors is from the same object. In a multisensor monitoring system, the role of object association is particularly important. Object missing association will produce redundant objects, and object misassociation will produce wrong judgments due to the fusion of nonsame objects [12]. In particular, wrong association information will reduce the accuracy of identification in the process of identifying abnormal behaviors [13].

The key to abnormal behavior detection lies in long-time object tracking [14]. Through continuous extraction of object behavior characteristics, the abnormal object is finally analyzed. The multiobject tracking algorithm tracks the movement trajectory of an object on the basis of the single-object tracking algorithm [15]. It also needs to distinguish different objects and correctly associate the motion trajectories of different objects in different frames, and it is also a more difficult computer vision task to deal with issues such as mutual occlusion of multiple objects, confusion of similar objects, and mutual motion effects between objects [16].

Object association has a wide range of applications in the field of multisensor information fusion, and its essence is an NP-hard problem that requires an approximate solution method [17]. The existing object association algorithms can be roughly divided into two categories [18–20]. The first category is the object association algorithm based on the nearest neighbor. Singer et al. assumed that the state estimation errors of different sensors for the same object are independent and proposed an effective nearest-neighbor algorithm to improve the accuracy of data association [18]. Considering the influence of common noise, Bar-Shalom et al. proposed an improved weighted statistical distance object association algorithm by optimizing the distance measurement function [19]. Kosaka et al. proposed a nearest-neighbor method combined with fuzzy association [20]. The second category is object association algorithm based on multidimensional allocation, which converts the object association to recursively solve the high-dimensional allocation matrix problem. Bar-Shalom et al. firstly used the Hungarian algorithm to solve the problem of object association between two sensors [21]. Poore et al. used the Lagrangian relaxation operator method to approximately obtain the high-dimensional association matrix in the multisensor object association problem [22]. Since the association algorithm based on the nearest neighbor and the Hungarian method both use hard decision-making association judgment and neither consider the time correlation of object association, when the objects are dense, the hard-decision object association method based on single-frame processing is prone to the mismatch of the association results, which leads to redundant objects and wrong objects [23].

In order to achieve the accuracy of multiobject association and behavior recognition under multiple sensors, this paper proposes a multiobjects association and abnormal behavior detection for massive data analysis in multisensor monitoring network, which first uses belief propagation to associate multiple objects acquired by multiple sensors, extracts the object's behavior trajectory characteristics, and then builds a long short-term memory classification network to realize automatic classification of abnormal behaviors based on object association characteristics. Multiobject association is basis of the Belief Propagation framework, where the multisensor object association problem is transformed into a high-dimensional association matrix allocation problem, and then a high-dimensional matrix probability graph model is established; multiobject association fully considers the timing correlation and object detection probability, as well as the statistical dependence of the measurement on the association matrix; according to the belief propagation criterion of the association matrix, the BP algorithm is finally used to solve the marginal probability of the high-dimensional association matrix. The experimental results based on abnormal behavior database show that our proposed method can significantly shorten the classification time and achieve high classification accuracy and sensitivity, meeting the requirements of real-time and efficient automatic classification of abnormal behavior in complex monitoring network.

Based on the above research, this paper proposes a new multiobject association and abnormal behavior detection method. The work of this paper is summarized as follows:

(1) Connect multiple sensors to use belief propagation. Through the objects acquired by multiple sensors, the behavior trajectory characteristics of the objects are extracted.

(2) Establish a long short-term memory classification network to realize automatic classification of abnormal behaviors based on the associated features of objects.

(3) Multiobject correlation fully considers the temporal correlation and object detection probability and measures the statistical dependence of the correlation matrix. Experimental results show that this method has high classification accuracy and sensitivity and can meet the requirements of automatic classification of abnormal behaviors in complex surveillance networks.

2. Problem Description

In order to analyze our research object intuitively, we show the object association examples in different scenarios in Figure 1, where Figure 1(a) shows the surveillance images

(a) (b)

Figure 1: Object association examples in different scenarios. (a) Multiobject association under different sensors. (b) Multiobject association under the same sensors.

obtained by different sensors at different times and Figure 1(b) shows the surveillance images obtained by the same sensors at different times. It can be seen that the object has passed through the monitoring area of different sensors, and the monitoring results from different sensor-angles are displayed. If the spatiotemporal context characteristics of objects from different sensor networks can be associated, the track of the object can be accurately grasped and the behavioral tendency of the object can be predicted. Therefore, the solved problem in this paper is to estimate the object association matrix A on the basis of the past and current position and attitude of each local sensor and then extract the deep feature of the object's association information, obtain the object behavior feature classification, and finally realize the abnormal behavior detection.

Specifically, the joint probability density function $f(A \mid z)$ of the association matrix is adopted to solve the marginal posterior probability $\left\{f(\alpha^n_{i_1,\ldots,i_s} \mid z)\right\}_{i_1=0,\ldots,i_S=0}^{M^n_1,\ldots,M^n_S}$ of each object association event, where an efficient BP algorithm is used to approximate the marginal posterior probability so as to realize the association of multiple objects acquired by multiple sensors, extract the object's behavior trajectory characteristics, and then build a long short-term memory classification network to realize automatic classification of abnormal behaviors on the basis of the object's associated characteristics. The architecture of the model proposed in this paper is shown in Figure 2. All samples in this experiment are detected by YOLO-V3, and all objects in the image are extracted; belief propagation associates multiple objects acquired by multiple sensors, extracts the object's behavior trajectory characteristics, and then builds a long short-term memory classification network to realize automatic classification of abnormal behaviors based on object association characteristics.

3. Multiobjects Association Model

3.1. Association Matrix of Multiple Objects for Multiple Sensors. It is assumed that there are S sensors in the monitoring area, where $s \in \{1, \cdots, S\}$. Each local sensor detects the object at the same time and reports the object position to the fusion center in real time. The multisensor object association model processes effectively the information of the objects from different sensors so as to determine the same-source object and obtain a unified trajectory chart. The sensor s reports M^n_s object state estimation $z^s_{n,m}$, $m \in [1, \ldots, M^n_s]$, to the fusion center at time n. The estimated error covariance of the sensor object state $z^s_{n,m}$ is p^s_m, and the object detection probability of the local sensor s is p^s_d. It is assumed that all state estimates $z^s_{n,m}$ sent by each sensor to the fusion center have achieved the temporal and spatial registration. Therefore, we can define $z^s_n = [z^s_{n,1}, \ldots, z^s_{n,M^n_s}]^T$, $z_n = [z^{1T}_n, \ldots, z^{sT}_n]^T$, and $z = [z^T_1, \ldots, z^T_n]^T$ (T is transpose operation).

In addition, a binary random matrix $A_n \triangleq \left\{a^n_{i_1,\ldots,i_s}\right\}_{i_1=0,\ldots,i_1=0}^{M^n_1,\ldots,M^n_s}$ with the size of $(M^n_1+1) \times \cdots \times (M^n_S+1)$ is defined at time n. The element $a^n_{i_1\cdots i_s}$ of association matrix is a binary random variable with value of 0 or 1, which represents the object association event. The subscript i_s of $a^n_{i_1,\ldots,i_s}$ indicates the participation of the sensor s in the association event $a^n_{i_1,\ldots,i_s}$, where $i_s = 0$ indicates that all objects of the sensor s do not participate in the association event, and $i_s > 0$ indicates that the i_s − th local object of the sensor s participates in the association event. If $a^n_{i_1,\ldots,i_s} = 1$, the object participating in the association event is the same-source object. According to the definition of association matrix, $a^n_{0,\ldots,0}$ has no concrete meaning. Finally, we can define an association matrix $A \triangleq \{A_1, \ldots, A_n\}$.

It is assumed that, at time n, any trajectory of any local sensor s can only be associated with one object association combination of other sensors at most. Therefore, its mathematical expression can be written as follows:

$$\begin{aligned}
&\sum_{i_1=0}^{M_2} \cdots \sum_{i_s=0}^{M_s} a^n_{i_1,\ldots,i_s} = 1, \quad i_1 = 1, 2, \ldots, M^n_1,\\
&\sum_{i_2=0}^{M_2} \cdots \sum_{i_s=0}^{M_s} a^n_{i_1,\ldots,i_s} = 1, \quad i_2 = 1, 2, \ldots, M^n_2,\\
&\vdots\\
&\sum_{i_1=0}^{M_1} \cdots \sum_{i_{S-1}=0}^{M_{S-1}} a^n_{i_1\cdots i_S} = 1, \quad i_S = 1, 2, \ldots, M^n_S.
\end{aligned} \tag{1}$$

The above constraints are collectively referred to as sum-constraints. Only the data association matrix that satisfies equation (1) is valid, and the set of all the association matrices which satisfies the sum-constraints is A_n. Then, the associated object state estimation set represented by the association matrix element $a^n_{i_1,\ldots,i_s}$ as $z_{i_1,\ldots,i_s} \triangleq \left\{z^1_{i_1}, \ldots, z^N_{i_N}\right\}$ is defined, where $N(1 \le N \le S)$ is the total number of

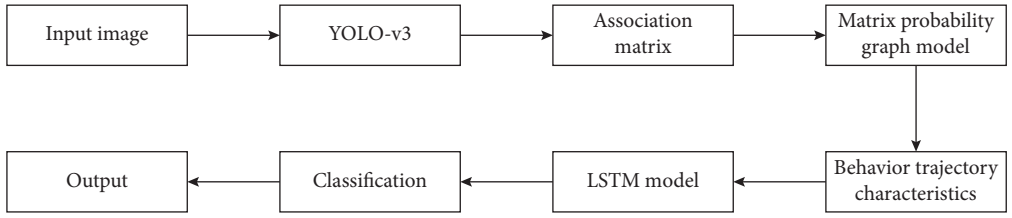

Figure 2: Architecture of our proposed model.

associated objects in the monitoring area. In other words, it is the number greater than 0 in the subscript $a^n_{i_1,\ldots,i_s}$; its corresponding estimation error covariance can be defined as $P_{i_1,\ldots,i_s} \triangleq \{P^1_{i_1}, \ldots, P^N_{i_N}\}$. It is assumed that the local sensor's local estimation error covariance of the same object is independent, and the average distance of the object set $z_{i_1,\ldots,i_s}$ at the time n is defined as follows:

$$\overline{d}\left(z_{i_1,\ldots,i_s}\right) = \begin{cases} \dfrac{2}{N(N-1)} \displaystyle\sum_{j=1}^{N} \sum_{k>j}^{N} d\left(z^j_{i_j}, z^k_{i_k}\right), & \text{if } N > 1, \\ \delta, & \text{if } N = 1, \end{cases} \tag{2}$$

where $d(z^J_{i_j}, z^k_{i_k}) = (z^J_{i_j} - z^k_{i_k})^T (\mathbf{P}^J_{i_j} + \mathbf{P}^k_{i_k})^{-1} (z^J_{i_j} - z^k_{i_k})$ is the estimated Mahalanobis distance between the object positions $z^j_{i_j}$ and $z^k_{i_k}$, and δ is the threshold. When the object association event $a^n_{i_1,\ldots,i_S}$ is true, the probability that the object set $z_{i_1,\ldots,i_S}$ contains the same homologous object is defined as

$$f\left(z_{i_1,\ldots,i_S} \mid a^n_{i_1,\ldots,i_s}\right) = \exp\left(-\frac{\overline{d}\left(z_{i_1,\ldots,i_s}\right)}{2}\right). \tag{3}$$

Therefore, the global likelihood function $f(z_n \mid A_n)$ is used to describe the statistical dependence of the local object z_n on the global association event A_n, which is written as follows:

$$f(z_n \mid A_n) = \prod_{i_1=0}^{M_1^1} \prod_{i_S=0}^{M_1^S} f\left(z_{i_1,\ldots,i_s} \mid a^n_{i_1,\ldots,i_s}\right). \tag{4}$$

3.2. Prior Probability of Allocation Matrix. In the process of object association, the object association relationship usually has strong continuity in time sequence. Since the probability distribution of the association matrix at a certain moment often has a great relationship with the probability distribution of the previous moment, it is assumed that the prior probability of the association matrix A_n conforms to the first-order Markov model in timing series. The prior probability of the association matrix is also related to the detection probability of each local sensor. Therefore, the prior probability of the association matrix is described as

$$P(A_n) = \psi(A_n)\phi(A_n) f(A_0) \prod_{i=0}^{n} f(A_i \mid A_{i-1}), \tag{5}$$

where $\psi(A_n)$ is an indicator function that represents the sum-constraints of association matrix and it can be defined as follows:

$$\psi(A_n) = \begin{cases} 0, & A_n \notin \mathbf{\Phi}_n, \\ 1, & A_n \in \mathbf{\Phi}_n, \end{cases} \tag{6}$$

where $\phi(A_n)$ in equation (5) is the confidence function of the detection probability of the local sensor to the prior probability of the association matrix, which is defined as follows:

$$\begin{aligned} \phi(A_n) &= \prod_{i_1=0}^{M_1^1} \cdots \prod_{i_s=0}^{M_1^s} \phi\left(a^n_{i_1,\ldots,i_s}\right) \\ &= \prod_{i_1=0}^{M_1^1} \cdots \prod_{i_s=0}^{M_n^s} \prod_{s=0}^{S} \left(p_d^s\right)^{u(i_s)} \left(1 - p_d^s\right)^{1-u(i_s)}, \end{aligned} \tag{7}$$

where $\mu(x)$ is an indicator function and its definition is denoted as follows:

$$\mu(x) = \begin{cases} 0, & \text{if } x = 0, \\ 1, & \text{otherwise}, \end{cases} \tag{8}$$

where $f(A_i \mid A_{i-1})$ is the state transition matrix of the probability distribution of the association matrix. It is observed that the directly solved marginal probability distribution needs to integrate the association matrix A, and its integration dimension increases with time, the number of objects, and the number of local sensors. Therefore, it is difficult to directly solve the marginal probability distribution for $f(A \mid z)$. In order to solve this problem, Bayesian model is adopted as follows:

$$\begin{aligned} f(A \mid z) \propto f(A) f(z \mid A) &= \prod_{i=1}^{n} f(z_i \mid A_i) f(A_i) \\ &= \prod_{i_1=1}^{M_1^s} \cdots \prod_{i_s=1}^{M_1^s} p\left(a^0_{i_1,\ldots,i_s}\right) \prod_{i=1}^{n} \psi(A_i) \times \prod_{i_1=1}^{M_1^s} \cdots \prod_{i_s=1}^{M_1^s} p\left(z^n_{i_1,\ldots,i_s} \mid a^n_{i_1,\ldots,i_s}\right) \phi\left(a^n_{i_1,\ldots,i_s}\right) f\left(a^n_{i_1,\ldots,i_s} \mid a^{n-1}_{i_1,\ldots,i_s}\right). \end{aligned} \tag{9}$$

It can be seen that the factorization of the above equation can be expressed as a circular factor graph shown in Figure 2. Therefore, the solution of association probability is divided into the following four steps. The first step and the second step are to initialize the confidence of the association variables according to the prior probability of the association matrix and the local object of the sensor. The third step is the iterative propagation of messages between variable nodes and constraint nodes, which plays a key role in data association, namely, how to solve the marginal probability of each associated variable under the conditions of satisfying sum-product constraint. When all the messages in the third step converge to a given threshold or the number of iteration steps reaches the maximum, the fourth step is to extract the highest confidence, which is to say that the marginal posterior probability $f(a_{i_1,\ldots,i_s})$ of $f(A_n \mid z_n)$ can be obtained. So far, we have realized the multiobjective information association and obtained the object association information. It lays a foundation for the next step to adopt the long short-term memory network model for abnormal behavior classification.

4. Abnormal Behavior Detection Based on the Long Short-Term Memory Network Model

4.1. Long Short-Term Memory Network. Once the spatio-temporal association information of the object is obtained, the deep model can be used to obtain the most essential behavior characteristics and perform feature classification [23]. This section adopts that the convolutional neural network automatically learns the association signal features, the input signal is reduced to a small size and fixed length 1-dimensional feature vector, and then the obtained feature vector is input into the long short-term memory (LSTM) network for training and testing and realizes the association signal-based automatic classification of abnormal behavior [24]. CNN can effectively learn data features through convolution operation and its special structure and classify the input information according to the hierarchical structure. In the convolution process, the convolution kernel repeatedly acts on each receptive field of the entire input sample so as to obtain a deep-feature mapping of the input signal. The output feature map will be inputted to the pooling layer for feature selection and information filtering. The pooling layer uses the overall statistical characteristics of the adjacent output at a certain location to replace the output of the network at that location, reducing the amount of network parameters. In this paper, according to the VGGNet network framework combined with the properties of the object trajectory signal, a 1D deep convolutional neural network framework based on the trajectory signal is designed to realize the automatic feature extraction of the trajectory signal. Long short-term memory network is a variant of Recurrent Neural Network (RNN), which can effectively solve the problems of vanishing gradient and exploding gradient in the training process of RNN. It is often used to process time-series signals. There are input gate, forget gate, and output gate in the structure of recurrent neural network, where the input gate is used to learn what information is stored in the memory and to determine the information that needs to be updated; the forget gate is used to understand the length of time through which the information is stored; the output gate is used to determine which information can be used for output [25, 26].

As we all know, the signal features extracted by the convolutional neural network have strong representational capacity and Identification ability [26]. LSTM can highlight the timing of trajectory signals and remember the internal connections of signals. The depth of the neural network model is the key to ensuring feature abstraction and model generalization. However, if the network model is too deep, model degradation problems will occur [27]. Therefore, a 15-layer 1D convolutional neural network feature learning model is designed to extract trajectory signal features. In proposed deep model, the Rectified Linear Unit (ReLU) is selected as the activation function of the convolutional neural network, and batch normalization (BN) layer is added after each convolution layer to prevent overfitting of the model, which can obtain signal features with strong representational capacity and distinguishable ability [28]. The final output is a one-dimensional association signal feature. These features are used as the input of the long short-term memory network classifier, and the fully connected layer and the Softmax function are used to obtain the probability values of different categories, so as to realize the automatic classification of abnormal behaviors based on the association signals.

In the CNN feature extraction module, a convolution kernel with a size of (5, 1) and a sliding step of 1 is used to automatically extract deep features. The numbers of convolution kernels are 16, 32 and 64, respectively; the number of convolution kernels is finally passed. The results of all branches are recombined by convolution layer with 1 convolution kernel to form a feature mapping [29]. In addition, batch normalization is used to increase the internal covariance shift of feature vectors and prevent overfitting. The pooling layer selects the maximum pooling operation with a size of 2 and a sliding step of 2 to compress the feature vector size. Meanwhile, all intermediate layers use ReLU as the activation function. After the entire convolution operators are finished, the length of the association signal is compressed from 300×1 to 38×1 feature vector. The detailed structure parameters of each layer of the long short-term memory network model are shown in Table 1.

The signal spatial features extracted in the convolution structure are decomposed into sequential components and sent to the 32-unit LSTM network for analysis. Its special gate structure is used to improve the problems of gradient disappearance and gradient explosion, and the deep learning model can mine the deep-level features between trajectory signals and perform learning and mapping to help the model abstract the temporal feature in these feature vectors [30, 31]. Finally, the fully connected layer and the Softmax function are used to calculate the predicted probability according to the output of the LSTM module to complete the abnormal behavior classification.

4.2. Model Training and Classification. The initial values of network weight W and bias b are set to random numbers with approximate 0, which are continuously adjusted by the

Table 1: Detailed structure parameters of each layer in LSTM.

Layer	Module	Size of convolution kernel	Number of convolution kernels	Activation function	Step	Parameter	Output
0	Input	—	—	—	—		300×1
1	1d-Conv	5×1	16	ReLU	1	96	300×16
2	BN	—	—	—	—	128	300×16
3	1d-Conv	5×1	16	ReLU	1	1424	300×16
4	BN	—	—	—	—	1456	300×16
5	Max pool	2	16	—	2	32	150×16
6	1d-Conv	3×1	32	ReLU	1	3024	150×32
7	BN	—	—	—	—	3088	150×32
8	1d-Conv	3×1	32	ReLU	1	6192	150×32
9	BN	—	—	—	—	6256	150×32
10	Max pool	2	32	—	2	64	75×32
11	1d-Conv	5×1	64	ReLU	1	16560	75×64
12	BN	—	—	—	—	16688	75×64
13	1d-Conv	5×1	1	ReLU	1	28800	75×1
14	BN	—	—	—	—	28928	75×1
15	Max pool	2	1	—	2	2	38×1

network during the training process to obtain the meaningful spatial information in the multisensor monitoring data. The adopted features are extracted from convolution layer and pooling layer, and then these features are sent to the long-term and short-term memory network units for analyzing the abnormal behavior [32]. The adaptive moment estimation (AME) algorithm based on stochastic gradient descent [28] is used to train the model. After bias correction, the learning rate of each iteration has a certain range, which makes the parameters more stable. The weight is updated by the stochastic gradient descent (SGD) method, and the cross entropy loss (CEL) function is used to calculate the loss rate [33]. The learning rate was fixed at 0.0001, the momentum was 0.9, and the batch size was 20:

$$L(\theta) = -\frac{1}{N}\sum_{k=1}^{N}\sum_{j=1}^{C} Y_{k,j}\left(\log_2\left(P_{k,j}\right)\right) + \lambda\|\theta\|^2, \tag{10}$$

where λ is the regularization term of L2 norm, $Y_{k,j}$ is the category of abnormal behavior, c is the number of abnormal categories, and its number is 5; N is the number of training samples, α is the learning rate, θ is the hyperparameter of model and its updating formula is $\theta \leftarrow \theta + \partial L(\theta)/\partial\theta$. In classification module, only the final output of the LSTM network is transmitted to the full connection layer. The trajectory signals are divided into five categories by using Softmax activation function. The six nodes in Softmax layer represent walking, running, waving, jogging, boxing, and handclapping, respectively [34]. Finally, the probability of five categories can be calculated, and its calculation formula is as follows:

$$\begin{aligned} p_i(y \mid x) &= \text{soft}\max\left(w^x h + b^x\right), \\ \text{soft}\max(i) &= \frac{e^i}{\sum_j e^j}, \end{aligned} \tag{11}$$

where w^x, b^x are the weight and bias of the Softmax layer and p_i represents the probability of being divided into the i-th class.

5. Experimental Results and Analysis

5.1. Dataset. The model proposed in this paper is mainly used in multisensor monitoring network to achieve multiobject detection and association and to analyze the behavior trend of the object. In order to evaluate the performance of our proposed abnormal behavior detection algorithm, the training set uses the international common behavior recognition data: UCSD, KTH , UCF101, and HMDB5 [31–35]. The KTH database includes six types of behaviors with 25 different pedestrians in four scenarios, namely, walking, running, waving, jogging, boxing, and handclapping, where the camera of these samples is relatively fixed and the background is simple; UCF101 and HMDB5 are complex datasets containing a large number of behavior categories, most of the data comes from video clips, pedestrian movement is complex, the angle of view changes greatly, and there are a lot of multiperson interaction.

ped1 and ped2 datasets in UCSD are databases that researchers at the University of California have collected for abnormal behavior detection [29]. The ped1 dataset contains 34 training video clips and 36 testing video clips, while the ped2 dataset has 16 training video clips and 12 test video clips negative. In these datasets, the training samples only contain normal behavior, while the testing samples not only have normal behavior but also mix many abnormal behaviors. In these two datasets, the main background is public pedestrian roads, so riding bicycles, roller skating, and the presence of vehicles on nearby pedestrian roads are regarded as abnormal events. The dataset distribution is shown in Table 2.

5.2. Qualitative and Quantitative Analysis. To achieve the accuracy of multiobject association and behavior recognition under multiple sensors, this paper proposes a multiobjects association and abnormal behavior detection for massive data analysis in multisensor monitoring network, which first uses belief propagation to associate multiple objects acquired by multiple sensors, extracts the object's behavior trajectory characteristics, and then builds a long

Table 2: Dataset distribution.

	Ped1	Ped2	KTH	UCF101	HMDB5
Training	13000	12200	15000	15700	12000
Validation	1200	1500	1500	1500	2000
Testing	5500	4500	4620	3500	5000
In total	19700	18200	21120	20700	19000

short-term memory classification network to realize automatic classification of abnormal behaviors based on object association characteristics. Therefore, in order to analyze the effectiveness of our proposed algorithm, this experiment carries out qualitative and quantitative analysis from two experimental perspectives, namely, multiobject association and abnormal behavior detection.

5.2.1. Comparison Analysis for Multiobject Association. It is worth noting that all samples in this experiment are detected by YOLO-V3, and all objects in the image are extracted [30]. Multiobject association is basis of the belief propagation framework, where the multisensor object association problem is transformed into a high-dimensional association matrix allocation problem, and then a high-dimensional matrix probability graph model is established; multiobject association fully considers the timing correlation and object detection probability, as well as the statistical dependence of the measurement on the association matrix; according to the belief propagation criterion of the association matrix, the BP algorithm is finally used to solve the marginal probability of the high-dimensional association matrix.

In order to verify the effectiveness of the proposed multiobject association algorithm, this experiment considers the monitoring scenario with three-sensor joint monitoring system when the six objects are relatively parallel and cross-moving and selects Temp_Ass and Sift_Ass as comparison algorithms, which can be found in [12]. The iterative termination threshold of the belief algorithm is 10–6, and the maximum number of iterations is 1000. When the object is associated for three consecutive frames, the association starts, and if the object is not associated for four consecutive frames after the start of the association, the association ends.

The evaluation index of the experiment adopts the CLEAR MOT metrics commonly used in the field of multiobject association. This criterion defines the accuracy of multiobject association (MAA) , which measures the accuracy of the location of the associated results. In addition, we also adopted precision to comprehensively measure the accuracy of the number of associated objects (MAN), Number of False Positives (FP), Number of False Negatives (FN), and Number of Identity Switches (ID-Sw). It is assumed that the tracking result set of the k-th frame of the video sequence is T_k, and the ground-truth result set is GT_k; then the solved problem to measure the performance is the matching problem of the two sets of T_k and GT_k. Suppose the matching mapping of the two sets of T_k and GT_k is $\mathrm{Map}_k = \{o_i, h_j\}$, where o_i represents the ith real object and h_j represents the corresponding object obtained from the association result. The experimental results are shown in Table 3.

Table 3: Performance comparison of object association.

Models	MAA%	MAN%	FP	FN	ID-Ww
Temp_Ass	68.5	64.8	852	1251	75
Sift_Ass	72.8	69.1	753	1058	39
Proposed	77.3	72.8	718	860	27

Table 3 shows the quantitative evaluation results of the three algorithms. As shown in the table, compared with other algorithms, our proposed algorithm achieves the best association results in this video. This is because the video has many moving objects, but the objects are relatively sparse and the background is relatively simple. Our algorithm can accurately and completely detect the moving objects, and the number of false alarms and missing detections is less, so it has obvious advantages. Due to the limitation of the number of training samples and the accuracy of classifier, Temp_Ass and Sift_Ass have many false alarms and missed detection. Therefore, our proposed algorithm also obtains good test results in number of identity switches, which shows that the proposed algorithm can ensure the accuracy of multiobject association.

5.2.2. Comparison Analysis for Abnormal Behavior Detection. The long short-term memory network is implemented under the PyTorch deep learning framework. In order to improve the efficiency of feature optimization and sample learning, Adam optimization algorithm is used to update the network weight and parameter. The parameters are set as follows: alpha = 0.001, beta1 = 0.900, beta2 = 0.999, and epsilon = 10–8. For the behavior characteristics' extraction module, the learning rate is set to 10^{-6}, and the training iterations are 200K times.

It can be seen from the results in Table 4 that our proposed model has higher classification and recognition accuracy. SRM is a nondeep learning model that uses SVM classifiers to achieve final behavior recognition, and the recognition accuracy of simple background images in a fixed field of view is relatively high, while the result for video sequences with complex scenes is only 63%. BMP is a commonly used spatiotemporal flow recognition mode, and MSR is to combine the information of all channels to obtain the final feature description. Our proposed algorithm is based on the comprehensive improvement of each model, using the association information of the object, extracting rich global context information, and enhancing the accuracy of recognition. Therefore, our model has achieved the best result of 90.55% on the UCF101 dataset.

Since there are many behavior categories in different datasets, in order to accurately analyze the behavior recognition performance under different categories, this article uses the classification confusion matrix of the ped2 data recognition results to quantitatively describe the results as shown in Figure 3. It can be seen from Figure 3 that behaviors have false positives, but the proportion is small, and most behaviors have an 80% accuracy. In addition, the difference in sequence between walking and jogging in the dataset is small, but the accuracy of the detection result is

Table 4: Comparison for Recognition accuracy.

Models	Ped1%	Ped2%	HMDB5%	UCF101%	KTH %	Avg %
SRM	76.12	77.58	73.65	81.51	98.57	88.93
BMP	82.92	78.25	82.14	86.99	98.62	88.58
DSS	85.71	79.39	85.95	86.93	98.47	91.58
MSR	84.58	86.08	85.27	88.0	98.28	88.85
DCL	89.21	84.21	83.15	84.76	98.1	87.67
Proposed	89.95	85.01	87.26	90.55	98.72	90.71

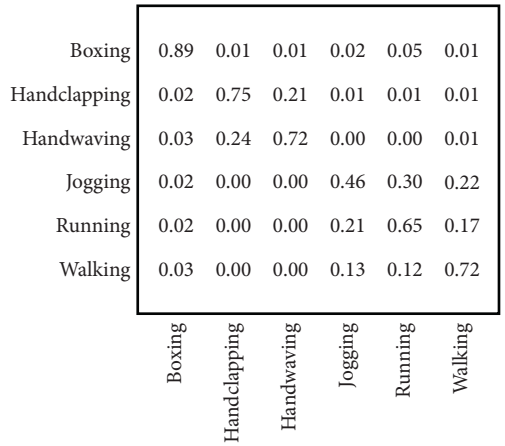

Figure 3: Classification confusion matrix of the Ped2 dataset.

high, which fully shows that our proposed can obtain a more ideal abnormal behavior classification effect.

6. Conclusion

With the rapid increase in the number of large-scale distributed cameras and the rapid increase in the monitoring range of the camera network, how to accurately recognize and analyze abnormal behavior is still a challenging problem. In addition, the appearance of moving objects between different cameras without overlapping fields of view undergoes significant changes, making it difficult to obtain accurate association Therefore, multiobjects association and abnormal behavior detection for massive data analysis in multisensor monitoring network is proposed in this paper, which firstly uses belief propagation to associate multiple objects, extracts the object's behavior trajectory characteristics, and then builds a long short-term memory classification network to realize automatic classification of abnormal behaviors. Multiobject association fully considers the timing correlation and object detection probability, as well as the statistical dependence of the measurement on the association matrix. The experimental results show that our proposed method can achieve high classification accuracy and sensitivity, which meets the requirements of automatic classification of abnormal behavior in complex monitoring network. However, as the number of cameras increases, the performance of the algorithm proposed in this article will gradually decrease. This is exactly where this research needs to be improved in the future.

Data Availability

The labeled datasets used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This work was supported by the 13th Five-Year Plan of Educational Science in Jiangsu Province (Grant no. B-b/2020/03/29).

References

[1] X. Tian, H. Li, and H. Deng, "An improved two-steps saliency detection algorithm based on binarized normed gradients and nuclear norm model in video sequences," *Journal of Information Hiding and Multimedia Signal Processing*, vol. 9, no. 4, pp. 841–852, 2018.

[2] D. A. Forsyth, *Computer Vision: A Modern Approach*, Prentice-Hall, Englewood Cliffs, NJ, USA, 2011.

[3] Z. Cai, L. Wang, X. Peng et al., "Multi-view super vector for action recognition," in *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 596–603, IEEE Computer Society, Washington, DC, USA, 2014.

[4] H. Wang and C. Schmid, "Action recognition with improved trajectories," in *Proceedings of the 2013 IEEE International Conference on Computer V Ision*, pp. 3551–3558, IEEE Computer Society, Washington, DC, USA, 2014.

[5] X. Peng, L. Wang, X. Wang, and Y. Qiao, "Bag of visual words and fusion methods for action recognition: comprehensive study and good practice," *Computer Vision and Image Understanding*, vol. 150, pp. 109–125, 2016.

[6] L. Wang, Y. Qiao, and X. Tang, "MoFAP: a multi-level representation for action recognition," *International Journal of Computer Vision*, vol. 119, no. 3, pp. 254–271, 2016.

[7] A. Karpathy, G. Toderici, S. Shetty et al., "Large-scale video classification with convolutional neural networks," in *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1725–1732, IEEE Computer Society, Washington, DC, USA, 2014.

[8] D. Tran, L. Bourdev, R. Fergus et al., "Learning spatiotemporal features with 3d convolutional networks," in *Proceedings of the 2014 IEEE International Conference on Computer Vision*, pp. 4489–4497, IEEE Computer Society, Washington, DC, USA, 2015.

[9] G. Varol, I. Laptev, and C. Schmid, "Long-term temporal convolutions for action recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 6, pp. 1510–1517, 2018.

[10] K. Simonyan and A. Zisserman, "Two-stream convolutional networks for action recognition in videos," in *Proceedings of the 2014 Conference on Neural Information Processing Systems*, pp. 568–576, NJ Curran Associates, New York, NY, USA, 2014.

[11] Y. H. Ng, M. Hausknecht, S. Vijayanarasimhan et al., "Beyond short snippets: deep networks for video classification," in *Proceedings of the 2015 IEEE Conference on Computer Vision*

and Pattern Recognition, pp. 4694–4702, IEEE Computer Society, Washington, DC, USA, 2015.

[12] L. M. Wang, Y. J. Xiong, Z. Wang et al., "Temporal segment networks: towards good practices for deep action recognition," in *Proceedings of the 2016 European Conference on Computer V Ision*, pp. 22–36, Springer, Berlin, Germany, 2016.

[13] L. Wang, Y. Qiao, and X. Tang, "Action recognition with trajectory-pooled deep-convolutional descriptors," in *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4305–4314, IEEE Computer Society, Washington, DC, USA, 2015.

[14] C. Szegedy, V. Vanhoucke, S. Ioffe et al., "Rethinking the inception architecture for computer vision," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2818–2826, IEEE Computer Society, Washington, DC, USA, 2016.

[15] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*, p. 22, MIT Press, Cambridge, UK, 2012.

[16] B. K. P. Horn and B. G. Schunck, "Determining optical flow," *Artificial Intelligence*, vol. 17, no. 1-3, pp. 185–203, 1981.

[17] Y. G. Jiang, J. Liu, A. Zamir et al., "Competition track evaluation setup," in *Proceedings of the First International Workshop on Action Recognition With a Large Number of Classes*, vol. 12, p. 23, Sydney, Australia, 2018.

[18] R. A. Singer and A. J. Kanyuck, "Computer control of multiple site track correlation," *Automatica*, vol. 7, no. 4, pp. 455–463, 1971.

[19] Y. Bar-Shalom and H. Chen, "Track-to-Track association using attributes," *Journal of Advances in Information Fusion*, vol. 2, no. 1, pp. 49–59, 2007.

[20] M. Kosaka, S. Miyamoto, and H. Ihara, "A track correlation algorithm for multi-sensor integration," *Journal of Guidance Control & Dynamics*, vol. 10, no. 2, pp. 160–171, 2015.

[21] A. B. Poore and A. J. Robertson III, "A new Lagrangian relaxation based algorithm for a class of multidimensional assignment problems," *Computational Optimization and Applications*, vol. 8, no. 2, pp. 129–150, 1997.

[22] J. M. Wolfe and T. S. Horowitz, "What attributes guide the deployment of visual attention and how do they do it?", *Nature Reviews Neuroscience*, vol. 5, no. 6, pp. 495–501, 2004.

[23] R. Desimone and J. Duncan, "Neural mechanisms of selective visual attention," *Annual Review of Neuroscience*, vol. 18, no. 1, pp. 193–222, 1995.

[24] S. K. Mannan, C. Kennard, and M. Husain, "The role of visual salience in directing eye movements in visual object agnosia," *Current Biology*, vol. 19, no. 6, pp. 247-248, 2009.

[25] C. Rother, V. Kolmogorov, and A. Blake, "GrabCut," *ACM Transactions on Graphics*, vol. 23, no. 3, pp. 309–314, 2004.

[26] X. Hou and L. Zhang, "Saliency detection: a spectral residual approach," in *Proceedings of the CVPR*, pp. 1–8, Salt Lake City, UT, USA, 2007.

[27] Z. Tang and S. Wang, "Object tracking based on sparse appearance model of local structure in DCT," *Video Engineering*, vol. 41, no. 7, pp. 140–146, 2017.

[28] J. L. Zhang and Z. Tang, "Bayesian framework with non-local and low-rank constraint for image reconstruction," *Journal of Physics Conference Series*, vol. 787, 2017.

[29] X. Tian, H. Li, and H. Deng, "Object tracking algorithm based on improved context model in combination with detection mechanism for suspected objects," *Multimedia Tools and Applications*, vol. 78, no. 12, pp. 16907–16922, 2019.

[30] X. Tian, H. Li, and H. Deng, "Object tracking algorithm based on improved siamese convolutional networks combined with deep contour extraction and object detection under airborne platform," *Journal of Imaging Ence and Technology*, vol. 64, 2020.

[31] Y. Hu, "Design and implementation of abnormal behavior detection based on deep intelligent analysis algorithms in massive video surveillance," *Journal of Grid Computing*, vol. 18, pp. 1–11, 2020.

[32] X. Hu, "A weakly supervised framework for abnormal behavior detection and localization in crowded scenes," *Neurocomputing*, vol. 383, pp. 270–281, 2020.

[33] L. Xia and Z. Li, "A new method of abnormal behavior detection using LSTM network with temporal attention mechanism," *The Journal of Supercomputing*, vol. 10, 2020.

[34] C. Guo, "Crowd abnormal event detection based on sparse coding," *International Journal of Humanoid Robotics*, vol. 16, no. 4, pp. 1220–1223, 2019.

[35] J. Cai, X. Zhang, and S. Xie, "Video crowd detection and abnormal behavior model detection based on machine learning method," *Neural Computing and Applications*, vol. 31, 2019.