

Identification of Discards in Video Surveillance

Xiaohu Liu, Guojian Cheng

School of Intelligence Science and Information Engineering

Xi'an Peihua University, China

*liuxiaohugood@126.com

Keywords: Bayesian reasoning; Reversible jumping Markov chain Monte Carlo; Multi-target tracking; Posterior probability

Abstract. In this paper, a detection scheme combining Bayesian network modeling and deep neural network feature extraction is used to detect the remaining cabinets in monitoring. The iterative sampling method of RJMCMC is used to track the follow-up multi-targets, and the identification of the legacy box is realized by means of three behavioral modes and tracking states of the RJMCMC process. Experimental results show the effectiveness of the algorithm.

1 Introduction

With the rapid development of the urban economy, the "smart monitoring and analysis system" has developed rapidly[1]. The literature [2] uses the Hidden Markov Model for the detection of legacy parcels and classifies the behavioral characteristics of a series of people such as hand-wrapping, dropping parcels, picking up parcels, etc. This method has a high requirement for the accuracy of the detection and segmentation of moving objects, and it may not be able to accurately separate parcels and people when the package is close to the human body. The literature [3] proposes multi-target tracking based on BPF filtering, uses ADaboot to detect the target, combines particle filtering to achieve tracking, and obtains better results; Monte Carlo method is a method that uses stochastic analog sampling to achieve posterior approximation; The literature [4] adopts the RJMCMC sampling method to realize the approximate reasoning process with variable dimensions, and then realize the tracking of indefinite targets.

The above methods are susceptible to light, deformation, and obscuration. In this paper, based on the deep neural network and RJMCMC, a method for detecting the legacy box is proposed: it will use the deep neural network architecture of Deep Mask [5] to realize the segmentation and detection of objects in the scene, and it is used to calculate the observation probability in a given state to improve the tracking state. At the same time, a variable multi-target tracking algorithm based on RJMCMC is used to realize the modeling of the object state tracking; then according to the detection results and tracking status of Deep Mask, the candidate set of discards is initially determined, and then the "freshman" behavior and tracking status in RJMCMC are combined to determine whether it is a discard.

2 Bayesian Model

Discard detection includes target segmentation identification and multi-target tracking. The overall process adopts dynamic Bayesian network model, which mainly contains two types of variables: the first is a hidden variable used to represent the state space in the tracking process, that is, the configuration of the tracking target, denoted as $X_t = \{X_{i,t}, i \in I_t\}$, i is the target index, and $t = \{1:T\}$ is the time index, which mainly describes the index, category, position and size of the target information; the second is the observation variable, denoted as Z_t , which mainly describes the characteristics of the object in the current frame and other related information.

In the modeling process, the first-order Markov assumption is introduced, combined with Bayesian theory, and formula (1) is used to calculate the posterior probability [6].

$$p(X_t|Z_{1:t}) = \frac{p(Z_t|X_t)p(X_t|Z_{1:t-1})}{p(Z_t|Z_{1:t-1})}. \quad (1)$$

Among then, $p(X_t|Z_{1:t-1})$ available Chapman-Kolmogorov equation is expressed in the form of Equation (2):

$$p(X_t|Z_{1:t-1}) = \int p(X_t|X_{t-1})p(X_{t-1}|Z_{1:t-1})dX_{t-1}. \quad (2)$$

Among then, $p(X_{t-1}|Z_{1:t-1})$ is the posterior probability of the previous moment, $p(X_t|X_{t-1})$ is the state transition probability, also known as the prediction step or the dynamic model. Substituting Equation (2) into Equation (1), and using the constant C to represent $p(Z_t|Z_{1:t-1})$, we can get the formula (3) recursive Bayesian filter equation:

$$p(X_t|Z_{1:t-1}) = C^{-1}p(Z_t|X_t) \int p(X_t|X_{t-1})p(X_{t-1}|Z_{1:t-1})dX_{t-1}. \quad (3)$$

Equation (3) mainly consists of two steps: prediction and updating. A prediction phase in which the state transition probability $p(X_t|X_{t-1})$ between the current time and the previous time and the posterior distribution $p(X_{t-1}|Z_{1:t-1})$ of the previous time are combined, to predict the current state of time $p(X_t|Z_{t-1})$; in the update phase, the observation Z_t of the current time is known, and the likelihood probability $p(Z_t|X_t)$ is used to measure the degree of coincidence between the observation and the predicted state, thereby updating the current date.

2.1 The Dynamic Model of the Variable Target

The dynamic model of the variable target corresponds to $p(X_t|X_{t-1})$ in the prediction phase and can be :

$$p(X_t|X_{t-1}) \propto p_V(X_t|X_{t-1})p_0(X_t). \quad (4)$$

Among then, $p_V(X_t|X_{t-1})$ is a multi-objective dynamic model defined as:

$$p_V(X_t|X_{t-1}) = \begin{cases} \prod_{i \in I_t} p(X_{i,t}|X_{t-1}) & \text{if } I_{1:t} \neq \emptyset \\ k & \text{if } I_{1:t} = \emptyset \end{cases}. \quad (5)$$

k is a constant, \emptyset stands for empty set, $p(X_{i,t}|X_{t-1})$ is a single target of the transfer probability, defined as:

$$p(X_{i,t}|X_{t-1}) = \begin{cases} p(X_{i,t}|X_{t-1}) & \text{if } i \in I_{1:t} \\ p(X_{i,t}) & \text{if } i \notin I_{1:t} \end{cases}. \quad (6)$$

$p_0(X_t)$ is the interaction between the targets in the current moment and is used to limit the mutual coverage between the targets., which can be defined as:

$$p_0(X_t) = \prod_{i,j \in C} \phi(X_{i,t}, X_{j,t}). \quad (7)$$

Among then, $\phi(X_{i,t}, X_{j,t}) \propto \exp(-g(X_{i,t}, X_{j,t}))$, that is, the model is modeled based on the relationship between the target immediate neighbors.

2.2 Multi-Targets Joint Likelihood Model

Multi-targets joint likelihood model corresponds to $p(Z_t|X_t)$ in update phase, whose modeling is done by using Deep Mask[5] neural network structure, the overall frame of which shows in Figure.1:

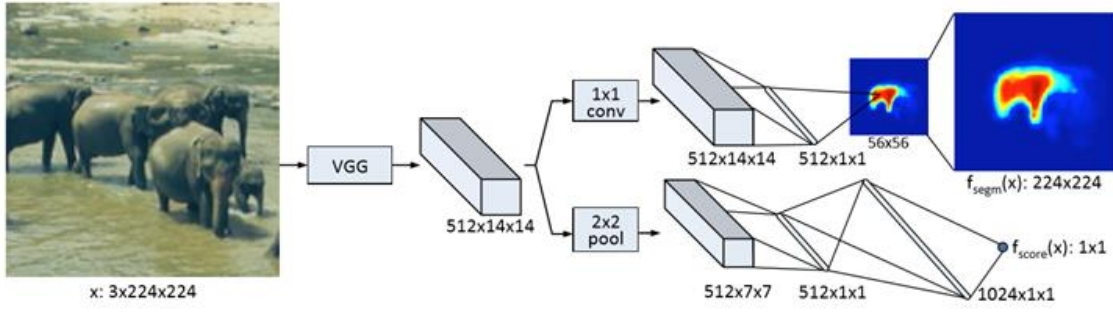


Figure 1. Structure of the Neural Network of Deep Mask

It mainly consists of two parts: one is to obtain the target candidate area and achieve the segmentation of the target; the second is the detection of target areas; the characteristics of the two are implemented in a shared manner in the early period, which makes the network structure more streamlined and efficient, and can reduce the number of candidate areas while ensuring a high recall rate and detection accuracy.

The loss function of this neural network is is:

$$\mathcal{L}(\theta) = \sum_k \left(-\frac{1+y_k}{2\omega^{\circ}h^{\circ}} \sum_{i,j} \log \left(1 + e^{-m_k^{ij} f_{seg}^{i,j}(x_k)} \right) + \lambda \log(1 + e^{-y_k f_{score}(x_k)}) \right). \quad (8)$$

In which, $f_{seg}^{i,j}(x_k)$ is the predicted value to separate parts in the position of (i,j), $f_{score}(x_k)$ is the score value of the predicted target and x_k is the image block of input. In the detection of discards, the predicted output of the training network needs to be a person and a discard object.

Thus, given the current image frame, the joint likelihood function to define multi-targets is:

$$p(Z_t|X_t) = \prod_{i \in I_t} p(Z_{i,t}|X_{i,t}). \quad (9)$$

In which, $p(Z_{i,t}|X_{i,t})$ is the likelihood probability value of the observed quantity in the given current state, which is defined as:

$$p(Z_{i,t}|X_{i,t}) \propto \exp(-\sum_{j,k} m_k^{jk} f_{seg}^{jk}(X_{i,t})). \quad (10)$$

When the model state and the observation are inconsistent, the resulting likelihood probability value is small, so that the state needs to be updated to conform to the observation and then the target is tracked.

3 Approximate Posterior Estimate

In the real scene, because of the interaction among targets, there exist varied corresponding relations between observed quantity and the state. This article uses the way of Reversible Jump Markov Chain Monte Carlo (RJCMCMC) [7] to approximately solve the posterior probability of state.

In RJCMCMC, any transformation that causes a dimensional change must be reversible, and an invertible transformation defines a transfer function h_v from the current state X to another proposed state X^* , and h_v is differentiable and isomorphic. In order to configure the dimensional change between states, it is usually necessary to introduce an additional variable U [7], then the acceptance rate of the state X transferred to the state X^* can be defined as:

$$\alpha(X_t, X_t^*) = \min \left\{ 1, \frac{p(X_t^*|Z_{1:t})}{p(X_t|Z_{1:t})} \times \frac{p(v^R)}{p(v)} \times \frac{q_v^R(X_t, U|X_t^*, U^*)}{q_v(X_t^*, U^*|X_t, U)} \times \left| \frac{\partial h_v(X_t, U)}{\partial (X_t, U)} \right| \right\}. \quad (11)$$

In which, $p(X_t|Z_{1:t})$ is the likelihood probability of model, $p(v)$ is the probability of conversion type, $q_v(X_t^*, U^*|X_t, U)$ is the transition probability, and $\left| \frac{\partial h_v(X_t, U)}{\partial (X_t, U)} \right|$ is the Jacobian matrix when the

dimension changes. Thus, in order to make the Jacobian matrix 1, the following three conversion methods are designed::

1) Newborn: it means a new target $X_{i^*}^*$ is added in the case that the current state is not changed. It can be defined as:

$$X_{i,t}^* = \begin{cases} X_{i,t}, & i \neq i^* \\ U, & i = i^* \end{cases} \quad (12)$$

Here, the proposal distribution is:

$$q_{birth}(X_t^*|X_t, U) = \sum_{i \in D_t \cup \{i^*\}} q_{birth}(i) q_{birth}(X_t^*|X_t, U, i). \quad (13)$$

In which, $q_{birth}(i)$ is the probability of the new-added target, i^* is the optional target index, D_t is the current disappearing target, and the proposal distribution of the new-added target is defined as:

$$q_{birth}(X_t^*|X_t, U, i) = \begin{cases} \frac{1}{c(X_t)} \frac{1}{N} \sum_{n=1}^N p(X_{i,t}^*|X_{t-1}^{(n)}) \prod_{j \in I_t} p(X_{j,t}|X_{t-1}^{(n)}) \delta(X_{j,t}^* - X_{j,t}) & \text{if } i = i^* \\ 0 & \text{otherwise} \end{cases} \quad (14)$$

In which, $p(X_{j,t}|X_{t-1}^{(n)})$ can be obtained by sampling in the prediction phase by formula (5). $p(X_{i,t}^*|X_{t-1}^{(n)})$ is the probability of dimension matching variable U . When $i = i^*$, $p(X_{i,t}^*|X_{t-1}^{(n)}) = p(X_{i,t}^*)$.

Here, the acceptance rate of sample is:

$$\alpha_{birth} = \min\left\{1, \frac{p(Z_t|X_t^*)}{p(Z_t|X_t)} \times \frac{p(v=death)}{p(v=death)} \times \frac{q_{death}(i^*)}{q_{birth}(i)}\right\} \quad (15)$$

In which, $p(Z_t|X_t)$ is the likelihood probability, $p(v)$ is the probability of conversion type, q_{birth} is the newborn probability.

2) Disappear: it means one target is removed from the current state in order to return to the previous tracking state, which is reversible with newborn conversion, that is, $h_{birth}^R = h_{death}$, $h_{death}^R = h_{birth}$, which is defined as:

$$\begin{cases} X_{i,t}^* = X_{i,t} & i \neq i^* \\ U^* = X_{i,t} & i = i^* \end{cases} \quad (16)$$

Here, the proposal distribution is:

$$q_{death}(X_t^*, U^*|X_t) = \sum_{i \in I_t} q_{death}(i) q_{death}(X_t^*, U^*|X_t, i). \quad (17)$$

In which, $q_{death}(i)$ is the target probability of option, i^* is the target index, which will be added into the disappearing target set D_t . The proposal distribution of disappearing target is defined as:

$$q_{death}(X_t^*, U^*|X_t, i) = \begin{cases} \prod_{j \in I_t, j \neq i^*} \delta(X_{j,t}^* - X_{j,t}) & \text{if } i = i^* \\ 0 & \text{otherwise} \end{cases} \quad (18)$$

Here, the acceptance rate of sample is:

$$\alpha_{\text{death}} = \min\left\{1, \frac{p(Z_t|X_t^*)}{p(Z_t|X_t)} \times \frac{p(v=\text{birth})}{p(v=\text{death})} \times \frac{q_{\text{birth}}(i^*)}{q_{\text{death}}(i)}\right\}. \quad (19)$$

3) Update: it means in the case that the states of other targets remain not changed, the states of the current selected targets are updated and itself is reversible. It can be defined as:

$$X_{i,t}^* = \begin{cases} X_{i,t}, & i \neq i^* \\ U, & i = i^* \end{cases}$$

Here, the proposal distribution is:

$$q_{\text{update}}(X_t^*, U^* | X_t, U) = \sum_{i \in I} q_{\text{update}}(i) q_{\text{update}}(X_t^*, U^* | X_t, U, i). \quad (20)$$

In which, $q_{\text{update}}(i)$ is the probability of the selected updating target and the proposal distribution of the selected target is:

$$q_{\text{update}}(X_t^*, U^* | X_t, U, i) = \frac{1}{N} \sum_{n=1}^N p(X_{i,t}^* | X_{t-1}^{(n)}) \prod_{j \neq i} p(X_{j,t}^* | X_{t-1}^{(n)}) \delta(X_{j,t}^* - X_{j,t}). \quad (21)$$

In which, $p(X_{i,t}^* | X_{t-1}^{(n)})$ is the probability of dimension matching variable.

Here, the acceptance rate of sample is:

$$\alpha_{\text{update}} = \min\left\{1, \frac{p(Z_t|X_t^*)}{p(Z_t|X_t)}\right\}. \quad (22)$$

4 Discard Detection

After completing the modeling of variable multi-objective detection tracking, the detection of discards and personnel in the current frame, as well as the overall target quantity, index, category, position and range information can be obtained. Furthermore, whether the target is discarded can be judged according to the moving speed of the target and the interaction between the target and the target, which mainly includes the judgment of the target movement and the judgment of the relationship between the target and the target.

A moving judgment part of a box, are judged by differential frames, be defined as:

$$v_{i,t} = ||\text{blob}_{i,t} - \text{blob}_{i,t-t_d}||. \quad (23)$$

Among them, t_d is ahead of the current d frame, $\text{blob}_{i,t}$ as the center of the target, when the speed is zero, is that the discard is not moving.

When the detection target movement speed is zero, it needs to be combined with the target activity cycle to further determine whether it is a discard. Combining the "new students" in the conversion types in the three parts, it is determined: that is, when the objects are discarded, the personnel and the objects should be integrated: the same state in the tracking process; when the object is discarded, the object and the person should be split into two objects, which appear as a new person or object during the tracking process.

The whole algorithm is as follows:

Step1: Initialization algorithm.

Take a sample i from all targets of the time of $t - 1$, use dynamic models $\prod_{i \in I_t} p(X_{t,i} | X_{t-1,i}^{(n)})$ to renew each target and record the sample $n = 0$.

Step2: RJMCMC sampling

1) According to $v \sim p(v)$ to select a conversion type in which the type includes "new", "disappeared" and "updated";

2) According to $q_v(i)$ to select the target i^* from the selected conversion type;

3) According to the proposed probability, take sample state X_t^* of the target i^* from the selected

conversion type, As follows:

- a) *new*: Sample the new target $X_{i,t}^*$ according to formula (13)
- b) *disappeared*: Remove the target $X_{i,t}^*$ from the current state according to formula (17)
- c) *updated*: Update the status of the target $X_{i,t}^*$ according to formula (20)

4) Calculate acceptance rate α_v based on selection conversion type("new" adopts formula (15), "disappeared" adopts formula (19); "updated" adopts formula (22));

5) Judge whether accept Adopted sample X_t^* . When $\alpha_v \geq 1$, accept the sample and add it to the Markov chain, Otherwise, the previous sample is added to the Markov chain;

6) Combining the likelihood probability of the formula (10) and the sampling sample, the posterior estimation of the state is computed;

7)In combination with formula (22) and the “newborn” transition between corresponding personnel and targets, determine whether the target is a discard.

5 Experimental Results and Analysis

The PETS2006 data source was used to verify the performance of the algorithm. Because of the randomness of the RJMCMC, each series of data is modeled seven times and the results are averaged as a result of the calculation.

Test video sequence S1, S2, S3, S4, S5, S6, and S7 for the discard detection, and the results are shown in Figure.2.

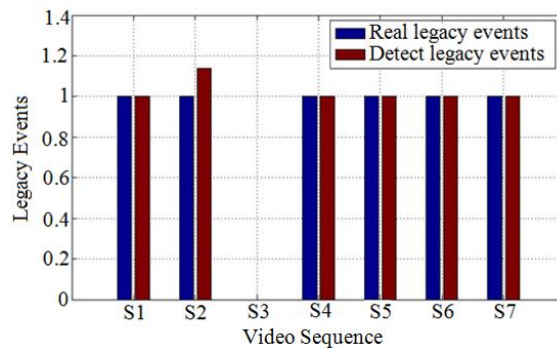


Figure 2. Discard Detection

Among them, gray represents the number of incidents of real discards and yellow is the number of discarded events detected. It can be seen from the above that: In addition to S4, the model was successfully detected as a discard event and there was a false alarm in sequence S2. In the event that S3 did not have an event, it could also be processed. For the sequence S4, the owner of the discard in this sequence sits on the discard, while another person appears close to the discard, and the owner of the discard leaves again. In this case, the model erroneously detects that the tank and personnel are not separated.

Figure 3 shows the discard detection results in S5. The detected targets are marked with different outlines. When discards are detected, the discards are highlighted.



Figure 3. Discard Detection Results in S5

The performance of the proposed algorithm is analyzed by the method based on the tracking rate and detection rate and compared with the BPF algorithm, as shown in Figure 4. BPH algorithm uses particle filter algorithm and GMM detection method, Adaboost algorithm combined with the method of testing, which discards the identification part of the detection object size and regional characteristics to achieve.

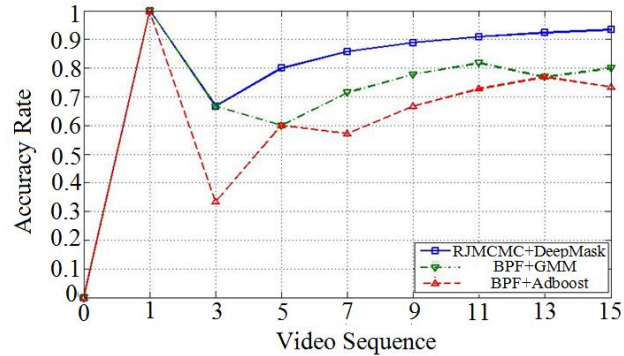


Figure 4. The Recall Rate

6 Conclusions

This paper proposes a discard detection method based on deep neural network and RJMCMC.

This method integrates tracking, segmentation and detection into the Bayesian inference framework. The segmentation and detection parts are implemented using deep neural network Deep mask and used to calculate the likelihood of the model; The model solver uses RJMCMC sampling to estimate the posterior probability, and combines the moving speed of discards and the interaction between people and discards to determine whether there is a discard event. Experimental results show that the algorithm has good performance.

The RJMCMC solution part is more time-consuming in the experiment. The next step is to use the neural network to approximate the posterior probability of the state to improve the detection speed.

References

- [1] Guohua Liao. 2008. The Key Technologies in Video Monitoring System, Communication & Information Technology, 4:58-61.
- [2] ZHANG Chun-jie, LI Ji-yun. 2009. Unattended Packages Recognition Based on HMM, Computer Technology and Development, 19(3):28-31.
- [3] Kreucher C, Morelande M, Kastella K. 2005. Particle Filtering for Multitarget Detection and Tracking, IEEE Transactions on Aerospace and Electronic Systems, 41:1396-1414.
- [4] Zhao Tao, Nevatia R. 2004. Tracing multiple humans in complex situations, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2196-2210.
- [5] Pinheiro P O, Collobert R, Dollar P. 2015. Learning to segment object candidates, neural information processing systems, 1990-1998,
- [6] S. Maskell. 2002. A Tutorial on Particle Filters for On-Line Nonlinear/Non-Gaussian Bayesian Tracking. IEEE Transactions on Signal Processing, 50(2):174-188.
- [7] P. Green. 1995. Reversible Jump MCMC Computation and Bayesian Model Determination. Biometrika, 82:711-732.