Ann. Data. Sci. (2015) 2(1):21–41
DOI 10.1007/s40745-015-0030-3

Bayesian Nonparametric Approaches to Abnormality Detection in Video Surveillance

Vu Nguyen[1] · Dinh Phung[1] · Duc-Son Pham[2] · Svetha Venkatesh[1]

Received: 11 January 2015 / Revised: 20 February 2015 / Accepted: 13 March 2015 /
Published online: 7 April 2015

Abstract In data science, anomaly detection is the process of identifying the items, events or observations which do not conform to expected patterns in a dataset. As widely acknowledged in the computer vision community and security management, discovering suspicious events is the key issue for abnormal detection in video surveillance. The important steps in identifying such events include stream data segmentation and hidden patterns discovery. However, the crucial challenge in stream data segmentation and hidden patterns discovery are the number of coherent segments in surveillance stream and the number of traffic patterns are unknown and hard to specify. Therefore, in this paper we revisit the abnormality detection problem through the lens of Bayesian nonparametric (BNP) and develop a novel usage of BNP methods for this problem. In particular, we employ the Infinite Hidden Markov Model and Bayesian Nonparametric Factor Analysis for stream data segmentation and pattern discovery. In addition, we introduce an interactive system allowing users to inspect and browse suspicious events.

✉ Vu Nguyen
ntienvu@gmail.com; tvnguye@deakin.edu.au

Dinh Phung
dinh.phung@deakin.edu.au

Duc-Son Pham
dspham@ieee.org

Svetha Venkatesh
svetha.venkatesh@deakin.edu.au

[1] Centre for Pattern Recognition and Data Analytics (PRaDA), Deakin University, Geelong, Australia

[2] Department of Computing, Curtin University, Bentley, Australia

Keywords Abnormal detection · Bayesian nonparametric · User interface · Multilevel data structure · Video segmentation · Spatio-temporal browsing

1 Introduction

In data science, anomaly detection is the process of identifying items, events or observations which do not conform to expected patterns or other items in a dataset. Typically the anomalous items are existing in some kind of specific problem such as bank frauds, medical problems or finding errors in text. There are two major categories of abnormal detection namely *unsupervised abnormal detection* and *supervised abnormal detection*. The former detects anomalies in an unlabeled test data set under the assumption that the majority of the instances in the data set are normal by looking for instances that seem to fit least to the remainder of the data set. The latter requires a data set that has been labeled as 'normal' and 'abnormal' and involves training a classifier (e.g. Support Vector Machine [1], Logistic Regression [2]).

In this paper, we consider specifically the problem of unsupervised abnormality detection in video surveillance. As widely acknowledged in the computer vision community and security management, discovering suspicions and irregularities of events in a video sequence is the key issue for abnormal detection in video surveillance [3–7]. The important steps in identifying such events include stream data segmentation and hidden patterns discovery. However, the crucial challenge in stream data segmentation and hidden patterns discovery are the number of coherent segments in surveillance stream and the number of traffic patterns are unknown and hard to specify.

The theory of Bayesian nonparametric (BNP) [8–13] holds a promise to address these challenges. As such, BNP can automatically identify the suitable number of cluster from the data. Therefore, in this paper we revisit the abnormality detection problem through the lens of BNP and develop a novel usage of BNP methods for this problem. In particular, we employ the infinite hidden Markov model [14] and Bayesian nonparametric factor analysis [15].

The first advantage of our methods is that identifying the unknown number of coherent sections of the video stream would result in better detection performance. Each coherent section of motion (e.g. traffic movements at night time and day time) would contain different types of abnormality. Unlike traditional abnormality detection methods which typically build upon a unified model across data stream. The second benefit of our system is an interface allowing users to interactively examine rare events in an intuitive manner. Because the abnormal events detected by algorithms and what is considered anomalous by users may be inconsistent, the proposed interface would greatly be beneficial.

To this end, we make two major contributions to abnormal detection in video surveillance: (1) proposing to use the Infinite Hidden Markov Model for stream data segmentation, and (2) introducing the Bayesian nonparametric Factor Analysis-based interactive system allowing users to inspect and browse suspiciously abnormal events.

This paper is organized as follows. We present an overview on abnormality detection in video surveillance and the need of segmenting the data and interaction in Sect. 2.

In Sect. 3, we describes our contribution on Bayesian nonparametric data stream segmentation for abnormal detection. Section 4 illustrates our introduced browsing system for abnormal detection. The experiment is provided in Sect. 5. Finally, we present a summary of the paper with some concluding remarks in Sect. 6.

2 Video Surveillance

Ideally, abnormality detection algorithms should report only events that require intervention—however, this is impossible to achieve with the current state-of-art. A large semantic gap exists between what is perceived as abnormal and what are computationally realizable outlier events. An alternative framework in which the algorithm reports a fraction (<1 %) of rarest events to a human operator for potential intervention [4] has been successful commercially (*icetana.com*). By retaining humans in the loop, whilst drastically reducing the footage that needs scrutiny, the framework provides a practical recourse to machine-assisted video surveillance. A typical medium-sized city council has to handle hundreds of cameras simultaneously, and it is imperative that the computational cost is low. This is achieved via an efficient algorithm based on PCA analysis of the video feature data. Motion-based features are computed within a fixed duration video clip (typically 10–30 s). PCA analysis is performed on the training data set to obtain the residual subspace, and the threshold corresponding to a desired false alarm rate. During testing, if the projection of the test vector in the residual subspace exceeds the computed threshold, the event is reported to the operator. Since the algorithm is based on PCA, it is important that the training data is coherent, so as to have most of the energy concentrated within a low dimensional principal subspace. In this case, most normal events remain within the principal space upon projection, and the residual subspace retains the fidelity for detecting subtle but rare events. However, for typical outdoor surveillance, the feature vectors generally exhibit different modes - depending on the time of day, climatic variations etc. If we try to fit all these incoherent modes into a single model, we reduce the sensitivity of detection. If we construct one principle sub-space for a 24 h period, we are likely to miss events at night, because nights have very different motion profiles to that of the daytime.

Thus, it is of paramount importance that video data be separated into coherent sections on which subsequent statistical analysis, for tasks such as anomaly detection, can be performed. One solution to provide this data segmentation into coherent modes is to use Markov models such as the Hidden Markov Model. However, these models requires apriori specification of the number of modes. To circumvent this problem, we model the activity levels as a mixture of Gaussian states for the infinite hidden Markov Model (iHMM) [14] segmentation. We show an application of the model to such stream data and present the collapsed Gibbs inference to achieve automatic data segmentation. To demonstrate the model, we perform experiments with 336 hours of footage obtained from a live surveillance scene. We show how the use of model selection as a preliminary process improves typical downstream processes such as anomaly detection.

The novelty of our contribution is in tacking a novel problem in large-scale stream data—model fitting to find coherent data sections, on which suitable models can be subsequently constructed. The significance of our solution is that the use of iHMM allows incremental use, and thus lends itself to large-scale data analysis. In addition, we introduce a browsing framework assisting users in analyzing and filtering suspicious events to overcome the semantic gap between the returned events by the algorithms and the true events.

3 iHMM Stream Data Segmentation

For data segmentation using standard HMM, one needs to specify the number of states in advance and use the EM algorithm to estimate the parameters. The iHMM [14] overcomes this restriction, allowing the number of states to grow unboundedly according to the data. In other words, the number of states will be automatically inferred from the data. It was later shown in [16] that this model can be interpreted using the hierarchical Dirichlet process formalism in which the number of groups is dynamically changed according to the state assignments. This interpretation is significant as it provides a deeper understanding and formal framework to work with the iHMM. Interested readers are referred to [14, 16] for details.

3.1 Multi-model Abnormality Detection Framework

We use video footage spanning multiple days for model selection and abnormality detection. A video is divided into a sequence of fixed 20 s clips. Optic flow vectors are computed [17]. For each clip, we first aggregate the total count of optic flow vectors at each pixel location over all the frames, and then spatially bin them into a 10×10 uniform grid. After vectorization, we obtain a 100 dimensional feature vector for each clip (as in [4]). For the model selection phase, we use the total activity level in an hour, computed by summing the feature vectors over an hourly window and then summing across the length of the resultant vector generating a scalar value for the total activity. The activity level is then modeled by a mixture of Gaussian states for the infinite-HMM [14, 16] segmentation. Once we obtain the segmentation of hours based on the activity levels, we run separate anomaly detectors for each model. In the following sections, we present the framework for iHMM followed by a brief description of the core anomaly detection algorithm of [4].

3.1.1 Bayesian Nonparametric

A *Dirichlet Process* [8] DP (γ, H) is a distribution over discrete random probability measure G on $(\Theta, \mathscr{B})$. Sethuraman [18] provides an alternative constructive definition which makes the discreteness property of a draw from a Dirichlet process explicit via the stick-breaking representation: $G = \sum_{k=1}^{\infty} \beta_k \delta_{\phi_k}$ where $\phi_k \overset{\text{iid}}{\sim} H, k = 1, \ldots, \infty$ and $\boldsymbol{\beta} = (\beta_k)_{k=1}^{\infty}$ are the weights constructed through a 'stick-breaking' process. As a convention, we hereafter write $\boldsymbol{\beta} \sim \text{GEM}(\gamma)$. Dirichlet Process has been widely used in Bayesian mixture models as the prior distribution on the mixing measures, resulting in a model known as the *Dirichlet process mixture model* (DPM) [9].

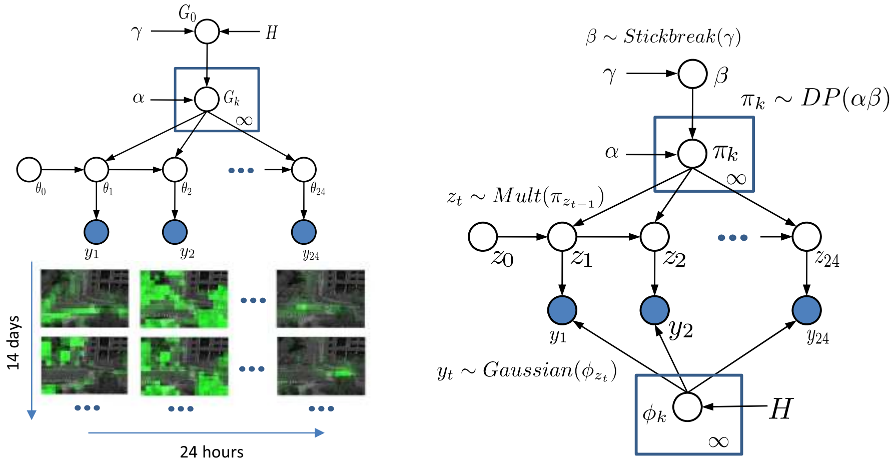


Fig. 1 The infinite Hidden Markov model representation. *Left* Stochastic process of iHMM. Each observation y_t indicates for a traffic movement for an hour t (including 24 hours) collected from 14 days. *Right* Stick-breaking representation of the data

Dirichlet Process can also be constructed hierarchically to provide prior distributions over multiple exchangeable groups. Under this setting, each group is modelled as a DPM and these models are 'linked' together to reflect the dependency among them. One particular attractive approach is the *Hierarchical Dirichlet Processes* (HDP) [16] which posits the dependency among the group-level DPM by another Dirichlet process.

3.1.2 iHMM for Data Segmentation

Under the hierarchical dirichlet process specification [16], the building block property can be adopted to represent the infinite Hidden Markov model (iHMM) [14]. [16] describe the infinite Hidden Markov model, namely a Hierarchical Dirichlet Process Hidden Markov model (HDP-HMM) which provides an alternative method to place a Dirichlet prior over the number of state. Therefore, the (unknown) number of states in HMM is identified in the same way as HDP.

Using HDP [16] as a nonparametric prior for building block, the stochastic process of HDP-HMM is described as:

$$\begin{array}{lll} G_0 \sim \mathrm{DP}\,(\gamma, H \times S) & & \theta_t \overset{\text{iid}}{\sim} G_k \\ G_k \overset{\text{iid}}{\sim} \mathrm{DP}\,(\alpha, G_0) & k = 1, 2, \ldots\infty & y_t \sim F\,(\theta_{t-1}) \quad t = 1, 2, \ldots, T. \end{array}$$

There are T timestamps (e.g. number of hours in a day the data is collected). The stick-breaking of HDP-HMM is illustrated in Fig. 1 in which the parameters have the following distributions:

$$\begin{array}{ll} \boldsymbol{\beta} \sim \mathrm{GEM}\,(\gamma) & \boldsymbol{\pi}_k \sim \mathrm{DP}\,(\alpha, \boldsymbol{\beta}) \\ \phi_k \sim H \qquad k = 1, 2, \ldots\infty & z_t \sim \pi_{z_{t-1}} \qquad t = 1, 2, \ldots, T \\ & y_t \sim F\left(\phi_{z_t}\right). \end{array}$$

Inference for HDP-HMM In this work, we use iHMM at the first stage to segment the data into coherent sections before building the abnormality detection models. The use of Markov model ensures that the temporal dynamics nature of the data is taken into consideration. The number of coherent sections is unknown and will be estimated from the data. Our first goal is perform a rough data segmentation at hourly intervals; thus there are 24 data points for each day using the average motion at each hour as the input. These inputs correspond the observed variables $\{y_t\}$, and $\{z_t\}$ plays the role the latent state variables as in a standard HMM. H is the base measure from which parameters $\{\phi_k\}$ will be sampled from. In our case, we model y_t as a univariate Gaussian and thus each ϕ_k is a tuple of $\left\{\mu_k, \sigma_k^2\right\}$ where both μ_k and σ_k^2 are unknown and treated as random variables. We use H as a conjugate prior, and thus H in our case is a Gaussian-invGamma distribution. A graphical model representation is shown in Fig 1.

We use collapsed Gibbs inference for iHMM as described in [19] in which the latent state z_t and the stick-breaking weight β_k are sequentially sampled by explicitly integrating out parameters $\{\phi_k\}$ for the emission probability and $\{\pi_k\}$ for the transition probability. For example, given $z_{t-1} = i, z_{t+1} = j$ from the previous iteration, the conditional Gibbs distribution to sample z_t has the form:

Here we shortly present the Gibbs sampling for HDP-HMM.

- Sampling z_t. Consider the conditional probability of z_t

$$p\left(z_t = k \mid z_{-t}, \boldsymbol{y}, \boldsymbol{\beta}, H\right) \propto \underbrace{p\left(y_t \mid z_t = k, \boldsymbol{z}_{-t}, \boldsymbol{y}_{-t}, H\right)}_{\text{observation likelihood}} \times \underbrace{p\left(z_t = k \mid \boldsymbol{z}_{-t}, \alpha, \boldsymbol{\beta}\right)}_{\text{CRP of transition}}.$$

The first term is the likelihood of the observation y_t given the component ϕ_{z_t}. In other words, this likelihood can be expressed as $\int_{\phi_k} p\left(y_t \mid z_t = k, \phi_k\right) p\left(\phi_k \mid \boldsymbol{y}_{-t}, \boldsymbol{z}_{-t}, H\right) d\phi_k$ which is easily analyzed using the conjugate property. The second probability is simply the Chinese Restaurant Process of transition. Denote n_{ij} as the number of transitions from state i to state j, n_{*j} as the number of all transitions to state j. Similarly, n_{i*} is the number of all transitions departing from state i. The CRP likelihood under Markov property can be analyzed as:

$$p\left(z_t = k \mid \boldsymbol{z}_{-t}, \alpha, \boldsymbol{\beta}\right) \propto \underbrace{p\left(z_t = k \mid z_{t-1}, \alpha, \boldsymbol{\beta}\right)}_{\text{from previous state t}-1\text{ to state t}} \times \underbrace{p\left(z_t = k \mid z_{t+1}, \alpha, \boldsymbol{\beta}\right)}_{\text{from state t to next state t+1}}.$$

We then have four cases to compute this probability:

$$p\left(z_t = k \mid \boldsymbol{z}_{-t}, \alpha, \boldsymbol{\beta}\right) \propto \begin{cases} \left(n_{z_{t-1},k} + \alpha\beta_k\right) \frac{n_{k,z_{t+1}}+\alpha\beta_{z_{t+1}}}{n_{k*}+\alpha} & k \leq K, k \neq z_{t-1} \\ \left(n_{z_{t-1},k} + \alpha\beta_k\right) \frac{n_{k,z_{t+1}}+1+\alpha\beta_{z_{t+1}}}{n_{k*}+1+\alpha} & z_{t-1} = k = z_{t+1} \\ \left(n_{z_{t-1},k} + \alpha\beta_k\right) \frac{n_{k,z_{t+1}}+\alpha\beta_{z_{t+1}}}{n_{k*}+1+\alpha} & z_{t-1} = k \neq z_{t+1} \\ \alpha\beta_{\text{new}}\beta_{z_{t+1}} & k = K + 1. \end{cases}$$

- Sampling stick-breaking $\boldsymbol{\beta}$, and hyperparameters α,γ are exactly the same as for HDP describing in [16].

For robustness we also let the concentration hyper-parameters α and γ to follow Gamma distributions and they will also be re-sampled at each Gibbs iteration.

Abnormality Detection Algorithm Assume that $X \in \mathbb{R}^{d \times n}$ is the data matrix with n centralized feature vectors of d dimensions and C is the covariance matrix with its SVD factorization:

$$C = U \Sigma U^T.$$

We divide the eigenvectors from U in two groups:

$$C = [U_1\ U_2] \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} U^T$$

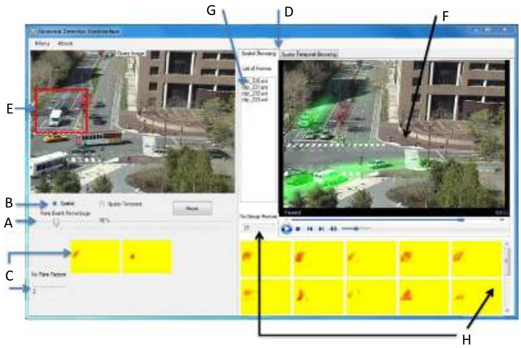
such that $\frac{\mathrm{tr}(\Sigma_1)}{\mathrm{tr}(\Sigma_1)+\mathrm{tr}(\Sigma_2)} = 0.9$, i.e., selecting the most significant eigenvectors such that they cover 90 % of the total energy. U_1 is called the principal subspace and U_2 is called the residual subspace. The abnormality detection algorithm works by projecting the test vectors to the residual subspace U_2 and comparing it to the detection threshold (λ), also called the Q-statistic, and is a function of the non-principle eigenvalues in residual subspace.

4 Interactive System for Browsing Anomalous Events

Security and surveillance systems focus on rare and anomalous events detection. Typically, these events are detected by estimating the statistics from the "normal" data - anything that deviates is termed as *rare*. The problem, however, is that in surveillance data, there is a semantic gap between *statistically* rare events produced by the detection algorithms and what the user would consider as *semantically* rare.

In this section, we raise the question: Is there an alternative to examining these anomalies, at least retrospectively? Consider security officers being given location/time of an incident - they now wish to find the matching footages. We propose a novel interface that permits the operators to specify such queries, and retrieve potential footages of *rare events* that match. This geometric query can be either *spatial* (rare events in region of interest) or *spatial–temporal* (rare events at location A, then B).

Our solution is firstly to find the hidden patterns in the scene. Since the number of latent factors is unknown in advance, we employ recent advances in Bayesian nonparametric factor analysis. The generative process models non-negative count data with a Poisson distribution [20]. The presence or absence of a factor is modeled through a binary matrix. Its nonparametric distribution follows the Indian buffet process [21], and is modeled through a draw from Beta process, which allows infinitely many factors. The extracted factors correspond to patterns of movement in the scene. The rareness of each extracted factor is determined by how much it is used across the whole data set. The factors are then ordered in decreasing rarity, and the user is allowed to choose a proportion of rare factors for consideration. Three top candidates rare factors from MIT dataset are visualized in the right column of Fig. 6 while three other common



A: Adjusting level of rareness.
The indicator is displaying 10% of the total of events.

B: Two browsing schemes: spatial and spatio-temporal.

C: Latent factors detected using % of rare factors chosen.

D: Filtered output.

E: Spatially selected region.

F: Video clip matching filtered result.

G: List of clips containing latent factors.

H: Number of consecutive frames.

Fig. 2 Graphical user interface (GUI) for our browsing system

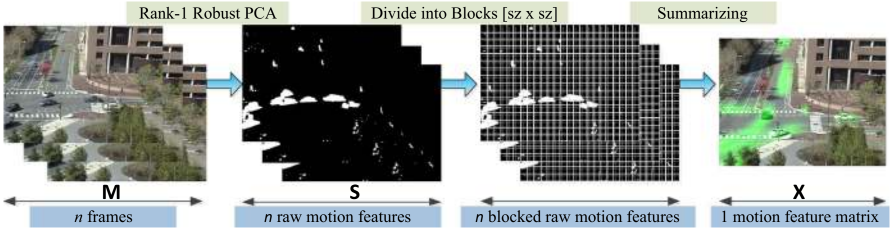


Fig. 3 Foreground extraction and feature computation using rank-1 robust PCA

patterns are on the left hand side. Frames that contain these factors are considered as potential candidates. The solution to a given geometric query is candidate frames that satisfy the specified spatial or spatial–temporal constraints. We demonstrate this browsing paradigm, with spatial and spatio-temporal queries in video surveillance. The user interface of our system is displayed in Fig. 2.

The significance of this paradigm is that it allows an operator to browse rare events, either spatially or spatial–temporally, at different "scales" of rarity. The use of non-parametric factor analysis models allows the framework to gracefully adapt to the data, without the need for *a priori* intervention. The framework can also easily be extended to accommodate multiple cameras. To our knowledge, there is no such existing system in the literature. Our main contributions in this interface include:

- The anomaly detection frame work based on part-based matrix decomposition that utilizes our recently introduced rank-1 robust background subtraction for motion video from static camera and nonparametric pattern analysis.
- The new browsing scheme allowing users not only to control the rareness degree but also to query spatial or spatial–temporal searching to overcome the difficulty due to the semantic gap.

4.1 Proposed Browsing Framework

A schematic illustration of the proposed system is shown in Fig. 3. The first step is to perform background subtraction followed by the feature extraction step detailed in Sect. 4.2. Once the features are extracted, latent factors are learned as detailed in Sect. 4.3. We use non-parametric factor analysis to recover the decomposition of factors (motion patterns) and constituent factor weights. For each latent factor, a rareness score is derived based on their overall contribution to the scene, and sorted in an decreasing order of rareness level. Since we follow a part-based decomposition approach for scene understanding, each latent factor is a *sparse* image having the same dimension of the original video frame. Therefore, a query for rare events at a spatial location can directly 'interact' with latent factors. The user is then able to select a proportion of rare factors for consideration. Based on the rareness degree of each latent factors, the interface returns to the user the corresponding footages. We shall now describe these steps in detail.

4.2 Robust Foreground Extraction and Data Representation

Since our framework focuses on scene understanding and therefore, features are extracted directly from the foreground information. To do so, we require a robust foreground extraction algorithm which can operate incrementally and in real-time. To this end, we utilize a recently proposed robust PCA approach [22] which is a special case of the robust PCA theory [23,24] developed specifically for static surveillance camera. Given a short window time size of n and $\boldsymbol{M} = [M_1, M_2, \ldots, M_n]$ being the data matrix consisting of n consecutive frames, the goal of robust PCA theory is to decompose

$$\boldsymbol{M} = \boldsymbol{L} + \boldsymbol{S},$$

where L is a low-rank matrix and S is a sparse matrix. A standard algorithm to perform robust PCA is principal component pursuit (PCP) [23] which involves SVD decomposition at each optimization iteration step. However, it can be very costly to compute. Static cameras, on the other hand, pose a strong rank-1 characteristic wherein the background remains unchanged within a short duration. Given this assumption, an algorithm for rank-1 robust PCA can be efficiently developed which is shown to be a robust version of the temporal median filter [22]. This makes the foreground extraction, contained in $\boldsymbol{S}$, becomes extremely efficient [1] since it can avoid the costly SVD computation in the original formulation of [23]. Moreover, it can be operated incrementally in real-time.

Next, using the sparse matrix $\boldsymbol{S}$, a fixed $sz \times sz$ block is super-imposed and the foreground counts in each cell is accumulated to form a feature vector X summarizing the data matrix $\boldsymbol{M}$ over a short window time of size n. An illustration of this step is shown in Fig. 3.

4.3 Learning of Latent Factors

Recall that a foreground feature X_t is collected for each short window t. Let $\boldsymbol{X} = [X_1 X_2 \ldots X_T]$ be the feature matrix over such T collections. Our next goal is to learn latent factors from $\boldsymbol{X}$, each of which represents a 'part' or basis unit that constitutes

[1] In practice, it is noted to be 10–20 times faster than a standard optical flow implementation.

our scene. Using a part-based decomposition approach, a straightforward approach is to use nonnegative matrix factorization (NNMF) of [25] which factorizes $\boldsymbol{X}$ into

$$\boldsymbol{X} \approx \boldsymbol{W}\boldsymbol{H}, \tag{1}$$

where $\boldsymbol{W}$ and $\boldsymbol{H}$ are nonnegative matrices. The columns of $\boldsymbol{W}$ contains K latent factors and $\boldsymbol{H}$ contains the corresponding coefficients of each factor contribution to the original data in $\boldsymbol{X}$. Due to the nonnegativity of $\boldsymbol{H}$, a part-based or additive decomposition is achieved and each columns of $\boldsymbol{X}$ is represented by $X_j = \sum_{k=1}^{K} W_k H_{kj}$. However, a limitation of NNMF for our framework is that it requires a manual specification of the number of latent factors K in advance. This can severely limit the applicability of the proposed framework since such knowledge on K is very difficult to obtain.

To address this issue, we employ recent advances in Bayesian nonparametric factor analysis for this task which can automatically infer the number of latent factors from the data [15,26]. In particular, we use a recent work [20] that models count data using Poisson distribution. For the sake of completeness, we shall briefly describe it here. A nonparametric Bayesian factor analysis can be written as follows:

$$\boldsymbol{X} = \boldsymbol{W}\left(\boldsymbol{Z} \odot \boldsymbol{F}\right) + \boldsymbol{E}, \tag{2}$$

wherein $\odot$ denotes as the Hadamard product, $\boldsymbol{Z}$ is a newly added binary matrix whose nonparametric prior distribution follows an Indian Buffet Process (IBP) [21]. Its binary values indicates the presence or absence of a factor (i.e. a column of matrix $\boldsymbol{W}$) and the matrix $\boldsymbol{F}$ contains the coefficients when working with matrix $\boldsymbol{Z}$. Formally, $Z_{kn} = 1$ implies that the k-th factor is used while reconstructing the n-th data vector, i.e. n-th column of the matrix $\boldsymbol{X}$. In this nonparametric model, $\boldsymbol{Z}$ is modeled through a draw from Beta process which allows infinitely many factors. Given the data, the number of active factors [2] are automatically discovered using the inference procedure.

The distributions on the parameters $\boldsymbol{W}$, $\boldsymbol{F}$ of the above nonparametric model is as below

$$W_{mk} \sim \text{Gamma}\left(a_w, b_w\right), \quad \boldsymbol{F}_i \sim \Pi_{k=1}^{K}\text{Gamma}\left(a_F, b_F\right), \tag{3}$$

where a_w, b_w, a_F and b_F are the shape and scale parameters. Similarly, given the parameters, the data is modeled using a Poisson distribution in the following manner

$$X_i \mid \boldsymbol{X}, Z_i, F_i \sim \text{Poisson}\left(\boldsymbol{X}\left(Z_i \odot F_i\right) + \lambda\right), \tag{4}$$

where λ is a parameter which expresses modeling error $\boldsymbol{E}$ such that $E_{mn} \sim \text{Poisson}\left(\lambda\right)$.

We use Gibbs sampling to infer $\boldsymbol{W}$ and $\boldsymbol{F}$. Since the condition posteriors are intractable, auxiliary variables are introduced to make the inference become tractable. For example, the Gibbs update equation for i-th row of $\boldsymbol{W}$, denoted by $\boldsymbol{W}_{(i)}$, is given as:

[2] e.g. k-th factor is an active factor, if k-th row of the matrix Z has at least one non-zero entry.

$$p(\boldsymbol{W}_{(i)}\big|\boldsymbol{Z}, \boldsymbol{F}, \boldsymbol{X}, \lambda, \mathbf{s}) \propto \Pi_{k=1}^{K} (W_{ik})^{a+\sum_{j=1}^{T} s_j^{ik}-1} \times \exp\left\{-\left(b+\sum_{j=1}^{T} H_{kj}\right) W_{ik}\right\}, \tag{5}$$

where the auxiliary variables $\mathbf{s} = \left\{s_j^{ik}\right\}_{k=1}^{K+1}$ can be sampled from a Multinomial distribution for each $j \in \{1, \ldots, T\}$ satisfying $\sum_{k=1}^{K+1} s_j^{ik} = 1$:

$$p\left(s_j^{i1}, \ldots s_j^{iK}, s_j^{i(K+1)} \mid \cdot\right) \propto \frac{X_{ij}!}{\prod_{k=1}^{K+1} s_j^{ik}!} \Pi_{k=1}^{K} \left(W_{ik} H_{kj}\right)^{s_j^{ik}} \lambda^{s_j^{i(K+1)}}. \tag{6}$$

The matrix $\boldsymbol{F}$ and $\boldsymbol{Z}$ can also be sampled in a similar manner proposed in [26].

4.4 Browsing Functionalities

Using the latent factors learning in the previous steps, we propose the following functionalities for our system.

4.4.1 Discovering Rare Factors and Footages

For each factor W_k within K factors discovered in the previous step, we define a score to measure its rareness based on its overall contribution to the scene. Since $X_j = \sum_k W_k H_{kj}$, it is clear that H_{kj} is the contribution of factor W_k to reconstruct X_j. Hence, we have the term of $\sum_j H_{kj}$ is the overall contribution of factor k to $\boldsymbol{X}$. We define the rareness score of a factor as a function reciprocal to this quantity:

$$\text{r-score}\,(W_k) = -\log\left(\sum_j H_{kj}\right). \tag{7}$$

In our system, we rank the scores for those factors learned in Sect. 4.3 using Eq. 7 and allows the user to interactively choose the percentage α of rare factors to be displayed and interacted with (cf. Figs. 8, 2a). The list of footages associated with this factor is also returned to the user (cf. Fig. 2g). Denote $S\,(W_k)$ as the corresponding index set, then:

$$S\,(W_k) = \left\{j \mid H_{kj} > \epsilon, j = 1, \ldots, T\right\}, \tag{8}$$

where ϵ is a small threshold, mainly used for the stability of the algorithm. Further, let K_α be the collection of all rare factors, then the index set of all detected footages is:

$$\mathscr{F}_\alpha = \bigcup_{W \in K_\alpha} S(W). \tag{9}$$

4.4.2 Spatial Searching

Given a spatial region of interest R being input to the system, spatial filtering on rare events can now be efficiently carried out by analyzing the intersecting region between the spatial region R and the set of rare factors W. First we extend R to R' to have the full size of the video frame by zero padding and mask it with each rare factor W which will be selected if the resultant matrix is non-zero. Let $\mathrm{SP}_\alpha(R)$ be the set of output indices returned, then formally we have:

$$\mathrm{SP}_\alpha(R) = \bigcup_{W \in \mathrm{SPF}(K_\alpha, R)} S(W) \quad \text{where}$$
$$\mathrm{SPF}(K_\alpha, R) = \left\{ W \mid W \in K_\alpha, \left|\left| W \odot R' \right|\right|_0 > 0 \right\}.$$

Here, α is a percentage of rareness degree in as described in Sect. 4.4.1 and $\odot$ is element-wise multiplication, $||A||_0$ is the l_0-norm which counts the number of non-zero elements in the matrix A. The demonstration of this browsing capacity is shown in Fig. 8 which reveals that the security officer can scrutinize the red rectangle region in the left window to inspect any unusual things happened in the right panel such as an event that one person is crossing the street.

4.4.3 Spatial–temporal Searching

More significantly, the spatio-temporal criteria searching is included in our model in Fig. 8. The semantic can be understood as "show me the events here (red rectangle) followed by the events there (blue rectangle)" that is set temporally as within Δt seconds. Once again our filters extracted the frames data into the potential candidates for rare frames. Initially, an user indicates a queue region of interest. For this purpose, we illustrate them into two regions, say red and blue rectangle. Spatial scanning in previous section will be applied into both rectangles. Those output patterns are considered as the necessary input for this process. In accordance with the mathematical formula in Eq. 10, the typical illustration of this searching category can be found in Fig. 8.

$$\mathrm{ST}_\alpha(R_1, R_2) = \{(i, j) \mid i \in \mathrm{SP}_\alpha(R_1), \quad j \in \mathrm{SP}_\alpha(R_2), \quad |i - j| < \Delta t\}. \tag{10}$$

5 Experiment

In this experiment, we first demonstrate quantitatively the abnormality detection performance, then present the user interface system.

5.1 Quantitative Experiment

We use a 14 day long video footage from an existing surveillance camera mounted at a street junction overlooking a busy thoroughfare. For each hour, we have 14 separate observations from each of the 14 days—this is used as the input matrix for the iHMM inference. The total number of Gibbs iterations performed for the inference is 1500, with 500 burnings. An example of the discovered segmentation is shown in Fig 4. We discover two segments including 8.00 a.m.–8.59 p.m., and 9.00 p.m.–7.59 a.m.. The total running time is 10.58 s on the X5690 based server.

We next show why such data segmentation improves downstream processes like anomaly detection. We divide the data into two parts. The first 7 days are used for training, i.e. computing residual subspace and the detection threshold set. The detection threshold (λ) is set at 0.1 %. The remaining 7 days of video are used for testing, i.e. projecting each feature vector onto the residual space and declaring an anomaly if the projected energy in the residual space exceeds λ. We run two anomaly detectors:

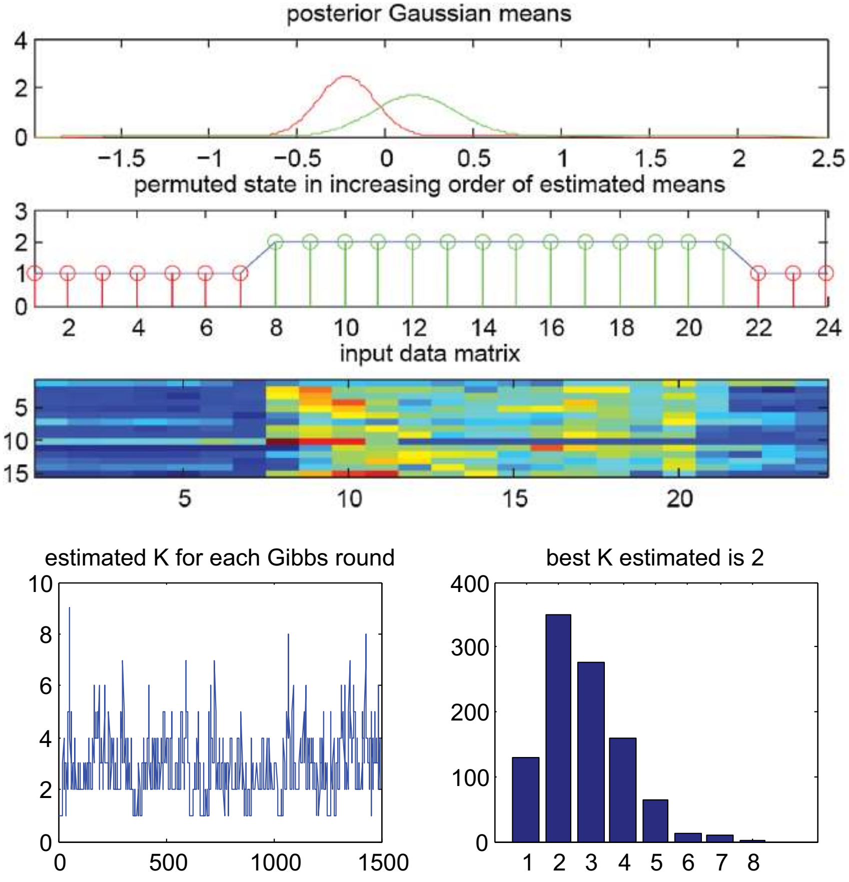


Fig. 4 Example of of iHMM segmentation for 1-day data

(1) The uni-model, that runs on the whole data, and (2) The multi-model, catering to multiple modes for the segmented hours as obtained by iHMM, with separate anomaly detectors for each mode.

The energy distribution of the test vectors in the residual subspace for the two settings are shown in Fig. 5a. The energy distribution for the multi-model decays more sharply, and thus an application of detection threshold will not ‘leak’ normal events as anomalous ones. Fig. 5b shows the energy signal for a chain of anomalous events—a street fight followed by police intervention. It shows that whilst the overall projection energy is higher for the uni-model, the detection threshold is also much higher, resulting in missed events (between frames 40 and 45 of Fig. 5a, for example). For the multi-model, the detection threshold is low, and the energy for this entire period remains above the detection threshold.

This effect is illustrated quantitatively in Table 1 which shows the number of events detected by both set-ups. The multi-model is more effective than the uni-model—detecting more loitering events (all of which occur at night, and thus are missed by the uni-model) and the full sequence of events in the street fight period. Incidentally, both models declared one (different) event as anomalous, which we consider a false positive. For both models, the training and testing of the total 14 days of video were achieved in less than 0.5 s.

5.2 User Interface Demonstration

Next, we demonstrate the proposed system using the MIT dataset [27]. In this public dataset, the traffic scene are recorded by the static camera, especially the traffic flows such as truck, car, pedestrian, bicycle, and other noisy motions such as leaves flickering due to wind etc. These objects generate various motion patterns in the intersection area of the traffic scene. The image dimension of the traffic scene is 480 × 720 pixel per frame (cf. Fig. 6). As mentioned earlier in Sect. 4.2, static cameras own the rank-1 property which is the necessary condition for our background subtraction task.

For the motion feature extraction stage, we choose the block size of 20 × 20 and a sequential footage of $n = 200$. In order to deal with matrix factorization problems when we do not know the number of latent factors beforehand, one possible solution is to do model selection by varying the number of latent factor K. The visualization of the model selection step is depicted in Fig. 1, in which we restrain the parameter scope from 20 to 56. Using our nonparametric model, however, the parameter K is automatically identified as 40. From 40 learned patterns (cf. Fig. 7), we sort all in an increasing order of rareness amount that is explained in Sect. 4.4.1. For example, three candidates for common factors and three rare factors are shown in Fig. 6.

We establish the browsing paradigm by assisting users to restrict their searching region by spatial and spatial–temporal criteria. One typical example is presented in Fig. 8. An user draws two regions: red and blue rectangles to investigate which patterns will followed by others in those windows. Initially, the system will automatically detects suitable candidate patterns in those regions with regard to the proportion of rareness level that user are querying. Through the candidate factors, we will reverse

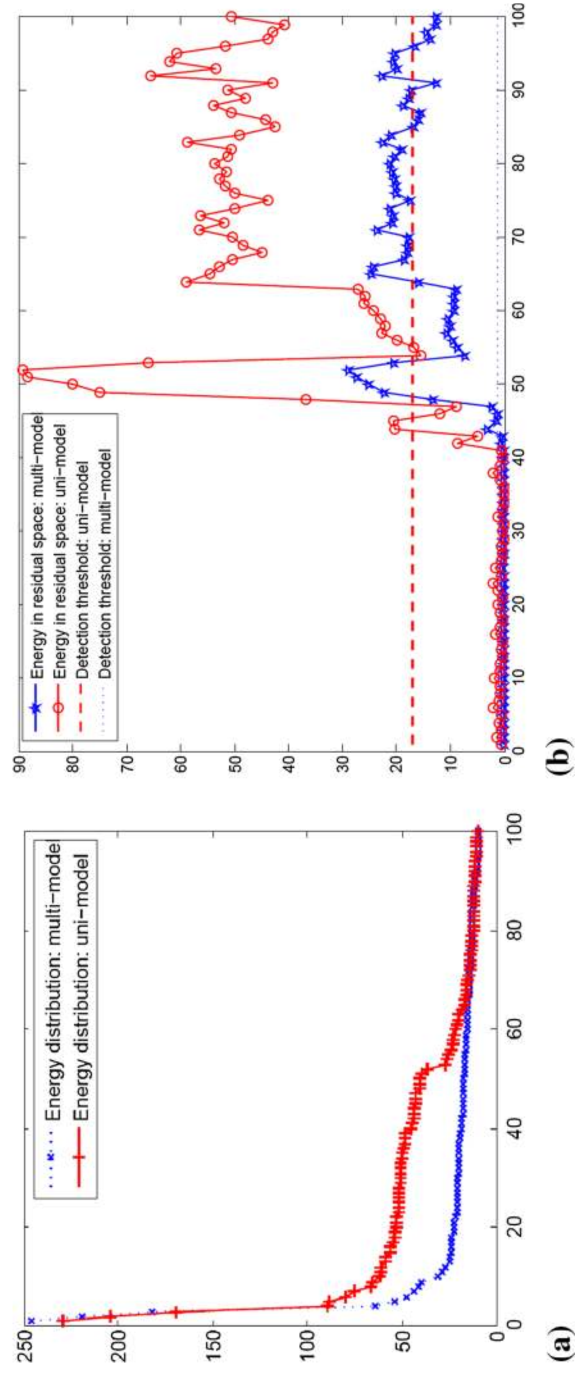


Fig. 5 Comparative signal in residual space. **a** Energy distribution in the residual subspace. **b** Residual signals

Table 1 Description of anomalous events

Event type	# detected (uni-model)	# detected (multi-model)
Street fight	57	63
Loitering	1	7
Truck-unusual stopping	4	4
Big truck blocking camera	2	2
No apparent reason	1	1

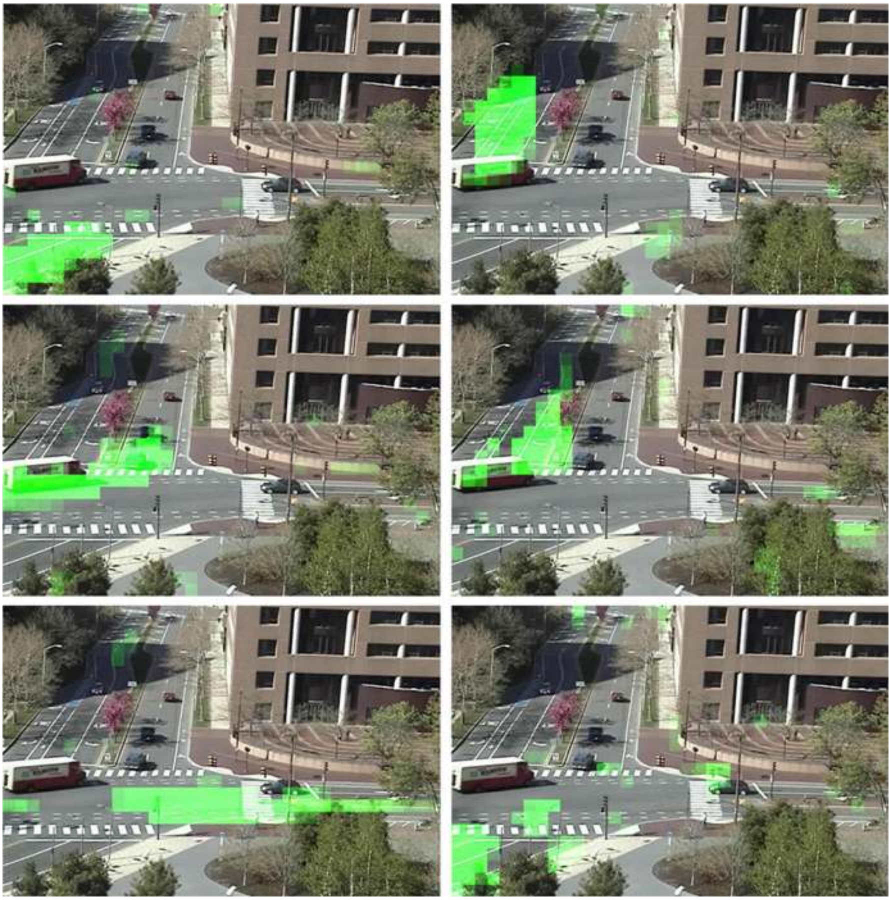


Fig. 6 Illustration of our learned factors overlayed with data the from MIT dataset. The *left column* presents three common patterns. Three rare factors are displayed in the *right column*

to all the consecutive frames and clips associated with the selected factors. Then, the most appropriate event will be discovered following Eq. 10. In Fig. 8, people who cross the zebra-crossing (red rectangle) and turn right (blue rectangle) are caught by our system.

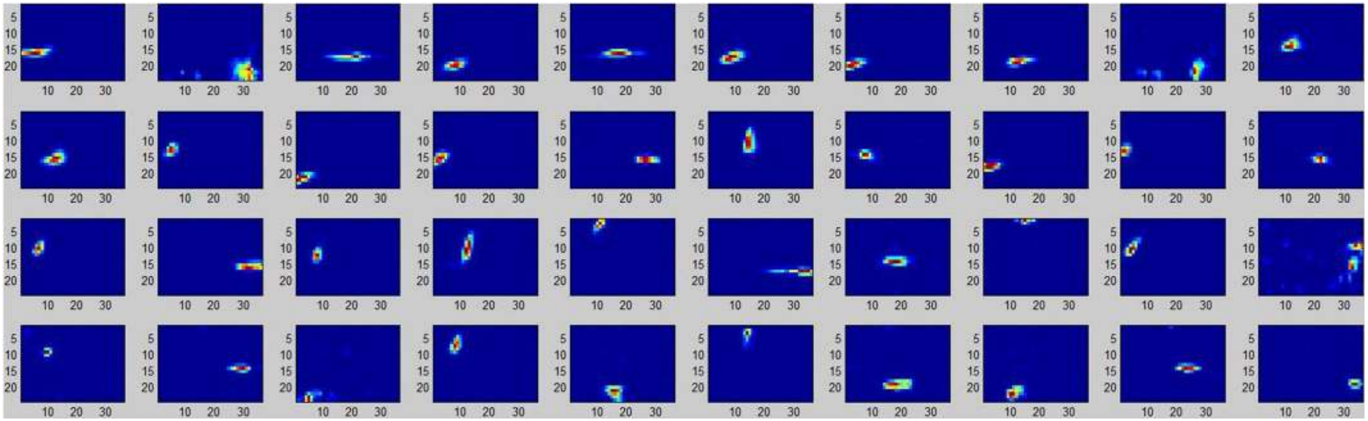


Fig. 7 40 factors learned from MIT dataset, shown from *top-left* to *right-bottom* in the increasing order of their ‘rareness’

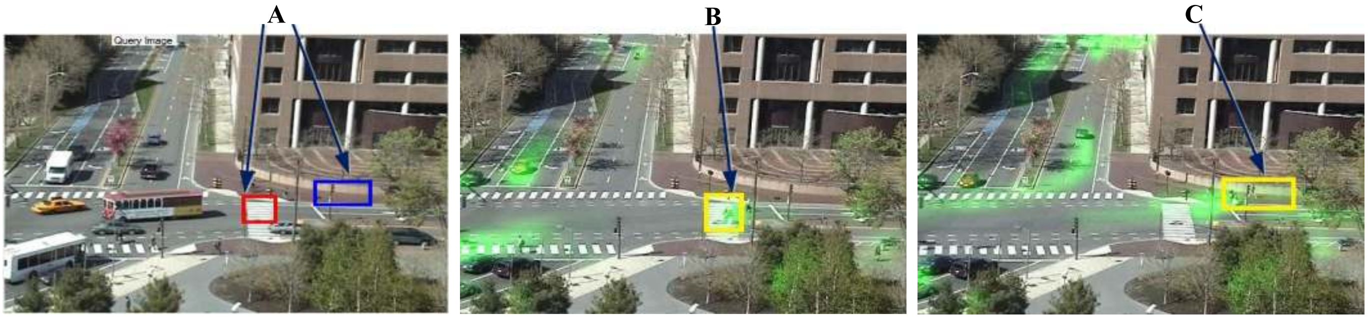


Fig. 8 Example of spatial–temporal browsing. User draw two *rectangles*: *red* and *blue* to find the abnormal incident that turn up at *blue area* followed the one in *red* section. The system caught the pedestrian which is compatible with his motion direction in this zones. **a** user drags two *rectangles* to search which events in *red* one followed by the *blue* one. **b, c** Detected frame corresponding in *red/blue* region. (Color figure online)

One false positive is also recorded. Because of the big traffic flow in the period of $n = 200$ serial frames in the selected rectangle, the system will treat it as an abnormal episode. When a user draws a spatial interrogation in this area, the machine will give back this flow as a possible candida for abnormality. However, the user can control, fortunately, the rareness level and alter it following the true abnormality semantically in the scene. Concerning with the input rareness rate, multiple patterns and clips are discovered so that the user can decide which one is a real affair. Thus, our proposed framework surmounts successfully the semantic gap between the statistical perspective and human perception.

Our focus is on browsing interactively the abnormal activities locally in a scene. There is no such existing interactive system available for comparison. Moreover, the difficult thing in evaluating our experimental results for interactivity is that there is no suitable ground truth which can satisfy all of user spatial and temporal queries. Because an user can examine in different locations: top left, right bottom, or middle region, and with different window sizes and time interval. For that reason, the quantitative evaluation of our abnormality detection approach can be referred to Sect. 5.1.

This system is programmed in C# and Matlab. The experiment is running on a PC Intel Core i7 3.4 GHz, with 8GB RAM. A query system takes approximately less than 0.2 s, as the motion feature extraction step was preprocessed. As mentioned, the rare patterns are understood as human perception, so we select roughly $p = 10\,\%$ for the number of rare events that the user can slide the bar to alter the number of rare events following their interests.

6 Conclusion

Identifying meaningfully anomalous events in video surveillance is essential to security management. In this paper, we address the problem of abnormality detection in video surveillance data using Bayesian nonparametric methods. We propose a framework for nonparametric data segmentation and multi-modal abnormality detection. By building multiple abnormality detection models on different coherent sections of the stream data, our proposed framework is more robust for abnormality detection in large-scale video data. Especially, when the video cameras are monitored across many days and exhibit strong variations in the data. Our experiments on a collection of video data over 14 days have demonstrated the superior performance of the proposed multi-modal anomaly detector compared to uni-model detectors.

In addition, we have addressed the problem of interactive monitoring in video surveillance, allowing users to examine rare events. They are detected in an unsupervised manner and can be filtered out interactively. We establish the browsing paradigm with spatial and temporal–spatial treatments to overcome the limitation of pure computational processing.

The main contributions in this paper are (1) proposing to use the Infinite Hidden Markov Model for stream data segmentation, and (2) introducing a user interface, using Rank-1 Robust PCA for feature extraction and Bayesian Nonparametric Factor Analysis for pattern discovery, allowing users to inspect and browse suspiciously abnormal events.

References

1. Cortes C, Vapnik V (1995) Support-vector networks. Mach Learn 20(3):273–297
2. Bishop CM et al (2006) Pattern recognition and machine learning, vol 1. springer, New York
3. Nanri T, Otsu N (2005) Unsupervised abnormality detection in video surveillance. In MVA, p 574–577
4. Budhaditya S, Pham DS, Lazarescu M, Venkatesh S (2009) Effective anomaly detection in sensor networks data streams. In: Proceedings of 9th IEEE international conference on data mining, p 722–727. IEEE, 2009
5. Nguyen TV, Phung D, Rana S, Pham DS, Venkatesh S (2012) Multi-modal abnormality detection in video with unknown data segmentation. In: Pattern recognition (ICPR), 2012 21st international conference on, pp 1322–1325, Tsukuba, November 2012, IEEE
6. Nguyen TV, Phung D, Sunil G, Venkatesh S (2013) Interactive browsing system for anomaly video surveillance. In: Proceedings of IEEE 8th international conference on intelligent sensors, sensor networks and information processing (ISSNIP), Melbourne, pp 384–389, April 2013
7. Kratz Louis, Nishino Ko (2009) Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models. In Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, pp 1446–1453. IEEE, 2009
8. Ferguson TS (1973) A Bayesian analysis of some nonparametric problems. Ann Stat 1(2):209–230
9. Antoniak CE (1974) Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. Ann Stat 2(6):1152–1174
10. Teh YW, Jordan MI (2009) Hierarchical Bayesian nonparametric models with applications. In: Hjort N, Holmes C, Müller P, Walker S (eds) Bayesian nonparametrics: principles and practice. Cambridge University Press, Cambridge, p 158
11. Jordan MI (2010) Bayesian nonparametric learning: expressive priors for intelligent systems. In: Geffner H, Dechter R, Halpern J (eds) Heuristics, probability and causality: a tribute to Judea Pearl. College Publications, London
12. Phung D, Nguyen X, Bui H, Nguyen TV, Venkatesh S (2012) Conditionally dependent Dirichlet processes for modelling naturally correlated data sources. Technical report, pattern recognition and data analytics, Deakin University
13. Nguyen V, Phung D, Venkatesh XL, Nguyen S, Bui H (2014) Bayesian nonparametric multilevel clustering with group-level contexts. In: Proceedings of international conference on machine learning (ICML), Beijing, China, pp 288–296
14. Beal MJ, Ghahramani Z, Rasmussen CE (2002) The infinite hidden Markov model. In: Advances in neural information processing systems (NIPS), MIT, vol 1, pp 577–584
15. Paisley J, Carin L (2009) Nonparametric factor analysis with Beta process priors. In: Proceedings of the international conference on machine learning (ICML), pp 777–784. ACM
16. Teh YW, Jordan MI, Beal MJ, Blei DM (2006) Hierarchical dirichlet processes. J Am Stat Assoc 101(476):1566–1581
17. Horn BKP, Schunck BG (1981) Determining optical flow. Artif Intell 17(1–3):185–203
18. Sethuraman J (1994) A constructive definition of Dirichlet priors. Stat Sin 4(2):639–650
19. Van Gael J, Saatci Y, Teh YW, Ghahramani Z (2008) Beam sampling for the infinite hidden Markov model. In: Proceedings of international conference on machine learning (ICML), ACM pp. 1088–1095
20. Gupta S, Phung D, Venkatesh S (2012) A nonparametric Bayesian Poisson Gamma model for count data. In: Proceedings of international conference on pattern recognition (ICPR), pp. 1815–1818
21. Griffiths T, Ghahramani Z (2006) Infinite latent feature models and the Indian buffet process. Adv Neural Inf Process Syst 18:475
22. Pham DS, Rana S, Phung D, Venkatesh S (2011) Generalized median filtering—a robust matrix decomposition perspective. Preprint
23. Candes Emmanuel J, Li Xiaodong, Ma Yi, Wright John (2011) Robust principal component analysis? J ACM (JACM) 58(3):11
24. Eriksson A, van den Hengel A (2010) Efficient computation of robust low-rank matrix approximations in the presence of missing data using the l1 norm. In: Proceedings of IEEE international conference on computer vision and pattern recognition (CVPR)
25. Lee DD, Seung HS (2001) Algorithms for non-negative matrix factorization. Adv Neural Inform Process Syst 13:556–562

26. Teh YW, Gorur D, Ghahramani Z (2007) Stick-breaking construction for the Indian buffet process. In: Proceeding of the international conference on artificial intelligence and statistics (AISTAT), vol 11
27. Wang X, Ma X, Grimson WEL (2008) Unsupervised activity perception in crowded and complicated scenes using hierarchical Bayesian models. IEEE Trans on Pattern Anal Mach Intell (PAMI) 31(3):539–555