

Detecting Abnormal Events via Hierarchical Dirichlet Processes

Xian-Xing Zhang¹, Hua Liu², Yang Gao¹, and Derek Hao Hu³

¹ State Key Laboratory for Novel Software Technology,
Nanjing University

flyaway2009@gmail.com, gaoy@nju.edu.cn

² Software Engineering School, Xi'an Jiaotong University
doudoulh@gmail.com

³ Department of Computer Science and Engineering,
Hong Kong University of Science and Technology
derekhh@cse.ust.hk

Abstract. Detecting abnormal event from video sequences is an important problem in computer vision and pattern recognition and a large number of algorithms have been devised to tackle this problem. Previous state-based approaches all suffer from the problem of deciding the appropriate number of states and it is often difficult to do so except using a trial-and-error approach, which may be infeasible in real-world applications. Yet in this paper, we have proposed a more accurate and flexible algorithm for abnormal event detection from video sequences. Our three-phase approach first builds a set of weak classifiers using Hierarchical Dirichlet Process Hidden Markov Model (HDP-HMM), and then proposes an ensemble learning algorithm to filter out abnormal events. In the final phase, we will derive abnormal activity models from the normal activity model to reduce the FP (False Positive) rate in an unsupervised manner. The main advantage of our algorithm over previous ones is to naturally capture the underlying feature in abnormal event detection via HDP-HMM. Experimental results on a real-world video sequence dataset have shown the effectiveness of our algorithm.

1 Introduction

Abnormal event detection is one of the most important problems in analyzing video sequences. Consider the situation that you are facing with a long video sequence, possibly thousands of hours in duration, but all you want is to find critical events among those video sequences. Anomaly detection, in the context of activity recognition, is the process whereby a baseline of normal behavior is established, with deviation from this norm triggering an alert.

Detecting unusual activities is difficult since unusual events by definition rarely occur, they may be unexpected as Zhang et al.[3] suggests, but at the same time they are relevant for the task. This difficulty becomes more significant during training due to the fact that we are in short of such labeled sequences of unusual activities.

Traditional framework for abnormal event detection is mainly addressed by state-space models and differential equation models, such as works in [2][6][7][29].

To address the complexity of time-series data, parametric approaches such as state-space models have to address the problem of model selection. In this context, a methodology for defining an appropriate set of states constitutes a key issue, among those available methods, state decomposition has typically been performed in an *ad hoc* manner [30], this method essentially is a trial-and-error process to choose the number of states. In some particular settings, such as tracking-based abnormal event detection problem, we always can calculate many latent useful features after the tracking procedure, and always wish to select a number of critical features from those features [27], if we choose the Bayesian classifier, we have to decide the number of states for every state-space models[14]; this is impossible if the total number of features are large. Thus, in a word, previous state-space models all suffer the problem of deciding an appropriate number of features, and a non-optimal number of features decided would affect the overall performance a lot.

Here we apply the Hierarchical Dirichlet Process Hidden Markov model (HDP-HMM) [17], which by construction has an infinite number of states and can decide the suitable, or optimal, number of states automatically. Thanks to the effectiveness of HDP-HMM, we can propose a new feature selection algorithm which calculates the confidence vector by an ensemble of HDP-HMM-based weak classifiers. This ensemble of weak classifiers forms a general model, whose threshold is trained mainly to reduce the false positive rate. After we detected abnormal events from this general model, we filter out the misclassified normal activities later on, in order to reduce the false negative rate. Our experimental results show the effectiveness of this framework.

In the remaining of this paper, we will first review relevant background in abnormal event detection in Section 2. In Section 3 we describe our proposed three-phase framework to tackle the problem of abnormal event detection in detail. In Section 4 we present the experimental results to show the effectiveness of our algorithm. Finally we conclude the paper and provide possible directions for future work.

2 Related Work

There are many important previous work done in the problem of abnormal activity recognition. However, due to a lack of space, we can only briefly review some related works using state-based models. Another important point in tackling in this problem is the feature selection step, which we will also review some important works.

Most of the previous work proposed in the topic of event detection [2][5][7][14] were centered on the recognition of predefined events in particular situations[13]. Now we consider the problem of abnormal event detection, and according to [7], the approaches for abnormal event detection can be divided into: similarity-based approaches and state model-based approaches. In model-based approaches, HMMs[2][6][7], and other graphical models[5] are most widely adopted.

In this paper, we only care about the trajectories of the objects, which is the results of object movement tracking [12]. Based on the tracking result we can, almost always, get a better result than by just using the original image sequences, because more useful and meaningful features can be computed from these trajectories, for example, semantic primitives can be included [29]. To address the problem of feature selection, which by definition is calculated from the object moving trajectories, [6] proposed a hybrid discriminative / generative approach, where the useful features were extracted by a variation of the AdaBoost algorithm proposed in [27]. At the same time [14] suggested several useful methods for feature selection, including Relief algorithm[15][16].

3 Our Proposed Algorithm

In this section, we present our three-phase framework for abnormal event detection from video surveillance. Intuitively, tracking-based abnormal detection techniques typically compare a profile of all normal activity sequences to objects' moving coordinates, we should calculate candidate features from those coordinates first, then use feature selection techniques to choose critical features to train the model. Any deviation from the model is flagged as a potential alarm.

Here we provide an overview of our proposed framework. First we will use every single calculated features to train an HDP-HMM based classifier using Beam Sampler in a sequential manner, by deciding a suitable model for every feature automatically.

In the second phase, we apply a new AdaBoost-like algorithm to train an ensemble of these weak classifier and get a confidence vector for later classification use, we propose a new algorithm because classical algorithms such as Relief[15] and its variants[16], and ensemble learning framework such as AdaBoost[27] cannot be applied directly to our setting, as explained later in this section. These weak classifiers can identify normal activities with a higher likelihood and assume that everything else is abnormal with a lower likelihood, and when choosing the threshold for the general model, we always intend to reduce the false positive rate, but this somehow involved with increasing the false negative rate. Therefore, abnormal event detection has the potential to detect even the unknown events that have rarely happened[7].

Based on this intuition, our proposed algorithm goes to a third phase process. We adapt the already well-estimated normal event model to a particular unusual event model using detected outlier if low likelihood observed[3], i.e, we start from the usual event model, and move towards an unusual event model in some constrained way, this process is proved to be able to effectively reduce false negative rate. In Section 3.1, we quickly review the Hierarchical Dirichlet Process and Gibbs Sampling methods, and introduce the sequential beam sampling methods. In Section 3.2, we briefly introduce how we calculate features from input coordinate data and how we build an ensemble of weak classifiers. Finally in Section 3.3, we describe how we build suitable model adaptation techniques.

3.1 HDP-HMM and Beam Sampling Method

Hierarchical Dirichlet Process Hidden Markov Model. Consider groups of data, denoted as $\{\{y_{ij}\}_{i=1}^{n_j}\}_{j=1}^J$, where n_j denotes the number of data in group j , J denotes the total number of groups thought to be produced by related, yet unique, generative processes. Each group of data is modeled via a mixture model[17]. A Dirichlet Process(DP) representation may be used separately for each of the data group. In an HDP manner, the base distribution of each of the DPs are drawn from a DP, which is discrete with probability 1, so each of the DPs can share the statistical strength, for instance, this encourages appropriate sharing of information between the data sets. An HDP formulation can decide the right number of states for the hidden Markov model(HMM); from its posterior density function on the appropriate number of mixture components, to some extent, the number of states in HMM can go to infinite if necessary. Besides, it learns the appropriate degree of sharing of data across data sets through the sharing of mixture components.

The HDP could be built as follows (Due to space constraint, we will omit the detailed explanation of HDP in this paper, interested readers could refer to [17] for details):

$$G_0(\theta) = \sum_{k=1}^{\infty} \beta_k \delta(\theta - \theta_k) \quad \beta \sim GEM(\gamma) \quad \theta_k \sim H(\lambda) \quad k = 1, 2, \dots \quad (1)$$

$$G_j(\theta) = \sum_{t=1}^{\infty} \tilde{\pi}_{jt} \delta(\theta - \tilde{\theta}_{jt}) \quad \tilde{\pi}_j \sim GEM(\alpha) \quad j = 1, \dots, J, \tilde{\theta}_{jt} \sim G_0 \quad t = 1, 2, \dots \quad (2)$$

$$\bar{\theta}_{ji} \sim G_j \quad y_{ji} \sim F(\bar{\theta}_{ji}) \quad j = 1, \dots, J, i = 1, \dots, N_j. \quad (3)$$

Where $GEM(\cdot)$ stands for the stick-breaking process [17] as follows:

$$\beta'_k \sim Beta(1, \gamma)$$

$$\beta_k = \beta'_k \prod_{l=1}^{k-1} (1 - \beta'_l), \quad k = 1, 2, \dots \quad (4)$$

To better illustrate the construction of HDP-HMM, we introduce another equivalent representation of the generative model using indicator random variables

$$\beta \sim GEM(\gamma) \quad \pi_j \sim DP(\alpha, \beta) \quad z_{ji} \sim Mult(\pi_j) \quad \theta_k \sim H(\lambda) \quad y_{ji} \sim F(\theta_{z_{ji}})$$

Identifying each $G(k)$ as describing both the transition probabilities $\pi_{kk'}$ from state k to k' and the emission distributions parametrized by ϕ_k , we can now formally define the HDP-HMM as follows:

$$\beta \sim GEM(\gamma), \quad \pi_k \sim DP(\alpha, \beta), \quad \phi_k \sim H, \quad (5)$$

$$s_t \sim Mult(\pi_{s_{t-1}}), \quad y_t \sim F(\phi_{s_t}) \quad (6)$$

The Beam Sampler. The beam sampler[18] is a successful combination of slice sampling and dynamic programming, it does not marginalize out neither π nor ϕ . Specifically, the beam sampler iteratively samples the auxiliary variables \mathbf{u} , the trajectory \mathbf{s} , the transition probabilities $\boldsymbol{\pi}$, the shared DP parameters $\boldsymbol{\beta}$ and the hyperparameters conditioned on all other variables. We describe the Beam Sampler as follows:

Sampling \mathbf{u} : $u_t \sim \text{Uniform}(0, \pi_{s_{t-1}s_t})$

Sampling \mathbf{s} : Due to the fact that only the trajectories \mathbf{s} with $\pi_{s_{t-1}s_t} \geq u_t$ for all t will have non-zero probability given \mathbf{u} , and there are only finite number of such trajectories due to the nature of stick-breaking process, as a consequence, we can apply dynamic programming to do the sampling job as classical HMM.[18]

$$p(s_t | y_{1:t}, u_{1:t}) \propto p(s_t, u_t, y_t | y_{1:t-1}, u_{1:t-1}) \tag{7}$$

$$= \sum_{s_{t-1}} p(y_t | s_t) p(u_t | s_t, s_{t-1}) p(s_t | s_{t-1}) p(s_{t-1} | y_{1:t-1}, u_{1:t-1}) \tag{8}$$

$$= p(y_t | s_t) \sum_{s_{t-1}: u_t < \pi_{s_{t-1}, s_t}} p s_{t-1} | y_{1:t-1}, u_{1:t-1}. \tag{9}$$

From (9), we can apply the classic dynamic programming[25] to sample the whole trajectory \mathbf{s} .

Sampling $\boldsymbol{\beta}$, m_{jk} is as follows[17]:

$$p((\beta_1, \dots, \beta_K, \beta_{\bar{k}}) \mathbf{t}, \mathbf{k}, y_{1:T}, \gamma) \propto \text{Dir}(m_{.1}, \dots, m_{.K}, \gamma). \tag{10}$$

Sampling m_{jk} :

$$p(m_{jk} = m | n_{jk}, \beta, \alpha) = \frac{\Gamma(\alpha\beta_k)}{\Gamma(\alpha\beta_k + n_{jk})} s(n_{jk}, m) (\alpha\beta_k)^m \tag{11}$$

where $s(n, m)$ are unsigned Stirling numbers of the first kind.

The conditional distribution of the observation y_t given an assignment $s_t = k$ and given all other observations y_τ , having marginalized out θ_k , is derived as follows:

$$p(y_t | y_{\setminus t}, s_t = k, s_{\setminus t}, \lambda) \propto \int_{\theta_k} p(y_t | \theta_k) p(\theta_k | \{y_\tau | s_\tau = k, \tau \neq t\}, \lambda) d\theta_k. f \tag{12}$$

Towards each ϕ_k we notice that[17] they are independent of others conditional on \mathbf{s} , \mathbf{y} and their prior distribution H . One of the greatest merits of Beam Sampler is that even if the base distribution H is not conjugate to the data distribution F , an effective and simple Metropolis-Hastings could be developed[18].

3.2 Feature Calculation and Selection

In this section, we mainly introduce the feature calculation and selection method we adopted in our abnormal event detection framework.

Feature Calculation. Our abnormal event detection framework is based on the results of object tracking, so the first step of our method is to calculate useful features from the input coordinate sequences. There are many methods to calculate useful features, we adopt the method proposed in [14]. In [14], features are derived from the target global position provided by a tracker, and they can be divided into three groups:(1) instantaneous measurements, (2) average speed/velocity based features and (3) second order moments/energy related indicators.

Feature Selection. A variation of AdaBoost algorithm for classifier learning is provided by [27] for feature selection, AdaBoost is widely adopted in practice[6], but it’s mainly designed for the discriminant classifiers. Therefore it’s not suitable for our generative HDP-HMM based approach. At the same time, Relief algorithm [15] and its variations [16] are widely accepted in feature selection application, However, due to the fact that the majority of our training data always come from only one class, so neither can we apply Relief algorithm directly. For our HDP-HMM based generative model, we proposed a new AdaBoost-like algorithm as below, which combines merits from those two famous algorithms listed above, namely, our algorithm favors the feature that distinguishes the data best, and this algorithm belongs to an ensemble learning framework. Here we describe our algorithm in the following steps:

1. Given examples of feature sequences $(x_1, y_1) \dots (x_n, y_n)$ where x_i is a matrix of features corresponding to a activity sequence, and $y_i = \{0, 1\}$ for normal and abnormal examples respectively.
2. Initialize weights $u_{i,j} = \frac{1}{2m}, \frac{1}{2l}$, votes $v_{i,j} = 1.1$, and initialize the total number of votes V_i to m, n for every i that $y_i = 0, 1$, corresponds to the normal and abnormal sequences, here m and l are the number of negative and positive sequences respectively, and j stands for one of the J features, where J is the total number of features.
3. Normalize the votes into weights: $u_{i,j} \leftarrow \frac{v_{i,j}}{V_i}$
4. for $t=1:T$, T is the number of iterations
 - for $j=1:J$, J is the total number of features
 - Compute the likelihood $L_{i,j}$ for every normal sequence i by the weak classifier trained by j feature, then multiply the sum of likelihood by its corresponding weight $u_{1,j}$, $L'_j \leftarrow u_{(1,j)} * \sum_{i=1}^m L_j$
 - Transform the weights for every feature back into votes by multiply total positive votes V_1 . $v_{1,j} \leftarrow u_{1,j} * V_1$
 - Choose feature α with the greatest likelihood, increase the $v_{1,\alpha}$ by $voteinc$, and choose the feature β with the lowest likelihood, decrease $v_{1,\beta}$ by $votedec$ if $v_{1,\beta} > 1$, else set it to 0.
5. Do the same procedure for the abnormal sequences
6. then we get two vectors for both the abnormal and normal training sequences($W1$ and $W2$), now we can combine the two vectors into a single weight vector by a linear combination $C = \alpha * W1 + (1 - \alpha) * W2$. We can change the parameter α in order to get a lower false negative rate or a higher false positive rate, here we call the matrix as confidence vector.

7. The final classifier is: $Result(i) = \sum_{j=1}^J H(j) * Likelihood(J)$.
8. Sequence(i) is identified as abnormal if $Result(i) < Threshold$.

In our framework, we always choose a threshold with lower false positive rate, to make sure most of the outliers / abnormal activities can be detected and sent to the third phase.

3.3 Model Adaptation

The abnormal event models are derived from a general normal model in an unsupervised manner, and the benefit of such an unsupervised manner is that this framework can address the unbalanced label problem due to the scarcity of training data and the difficulty in pre-defining abnormal events, more specifically, after the second step we may get a high false negative rate, so it's necessary for us to apply a third phase, that is, to adapt models for the abnormal events, and use these abnormal classifiers to reduce the false negative rate. Besides, due to the lack of negative training data, we can not directly build models for abnormal events. However, we can use adaptation techniques to get them during the test time or even in the future use, that is, we can dynamically build the model for the abnormal event after the training phrase. Here we briefly introduce the algorithm's framework[3][7] first. The steps are listed as below:

Prerequisites: A well defined general HDP-HMM with Gaussian observation density trained by all normal training sequences.

- Step 0 : Using the first outlier detected from the former phase - which is considered to be able to represent a particular type of abnormal activities - to train an abnormal event model by adaptation using beam sampler.
- Step 1 : Slice the test sequence into fixed length segments, calculate the likelihood of these segments by the existing models, if the maximum likelihood is given by the general model, we predict this trace to be of a normal activity, then goto Step 4. Else goto Step 2;
- Step 2 : If the maximum likelihood is larger than the threshold, we consider this trace to belong to an existing abnormal model; then we predict this trace to be that kind of abnormal events, goto Step 4, else goto Step 3;
- Step 3 : Using adaptation method to adapt the general model to a new abnormal activity model, then add this adapted abnormal model to the set of models, go to Step 4, here this outlier is regarded to represent one kind of the certain events.
- Step 4 : Goto Step 11 if new outlier comes.

In this procedure, we provide the outlier a second chance to be recognized as a normal events, in this way, we try to identify those normal events that are misclassified due to their unexpectedness or scarcity in training data. Thanks to the effectiveness of beam sampler again, we can do the adaptation effectively without other special design. Suppose now we have the new parameters for the HDP-HMM λ , here the parameters that are adapted are the HDP parameter $\beta, \alpha_0, \gamma, K$ and the π, μ for the HMM.

Here a latent problem for the adaptation is that, we can no longer use traditional adaptation algorithms, such as Maximum Likelihood Linear Regression (MLLR) and Maximum a posterior (MAP), or Kernel Nonlinear Regression (KNLR) because the number of states K will change due to the adaptation, leading to the different dimensions between old and new variables, we adopted a heuristic method to deal with the problem, as will be discussed in the experimental part, and we will discuss it later in the final section.

4 Experiments and Results

In this section, we first describe our dataset and performance measures, then we describe the baseline systems we use to evaluate our results. At last we illustrate the effectiveness of the proposed algorithm framework using the dataset.

4.1 Dataset Description

We have concentrated on low-level, short term events, especially human activities, as those studied in the context of the CAVIAR project [28]. We will explore the availability of a large dataset of video sequences and manually classified activities, in a total of about 16,000 images with ground truth data. The activities (classes) considered can be detected from a relatively short video sequence (a few seconds) and are described in Table 1, we train the general model from a portion of the set of normal sequences, and test the model for the rest of normal sequences and the other sets of abnormal sequences, all of the data is available at [28].

Table 1. Examples of Normal and Abnormal Events

Normal events	Abnormal events
Walker: Walking	...
Browser: Browsing	Fighter: Running
Fighter: Fighting	Fighter: Walking
...	Walker: Running
...	...

4.2 Performance Measures

One of the most challenging problems involved with abnormal detection is the different misclassification costs corresponding to different types of error [19]. According to [7], cost-sensitive learning are useful for learning problems where the classes are extremely unbalanced. And in most of the time it's more important to recognize the negative class. We usually apply cost-sensitive learning under such situation, for example, Receiver Operating Characteristic (ROC) curves are recommended instead of accuracy [20]. In [21], the authors show that class-imbalance often affects the performance of cost-sensitive classifiers: when the

misclassification costs are not seriously unequal, cost-sensitive classifiers generally favor natural class distribution although it might be imbalanced; when misclassification costs are seriously unequal, a balanced class distribution is more favorable.

As we mentioned above, the problem of abnormal event detection can be recognized as a two-class classification problem, where the positive class(normal events) takes dominating majority in the whole training data, and so the traditional accuracy-based performance measure is no longer suitable for this setting. Here we evaluate the performance of the abnormal event detection algorithm framework in terms of two error rates: detection rate and false alarm rate. The detection rate is the same as true-positive rate while the false alarm rate is the same as false positive rate.

4.3 Results on CAVIAR Sequences

In order to evaluate the performance of our proposed algorithm, experiments were carried out on a real data set from [28]. For comparison, two other approaches were used as baselines. The first one is composed by HDP-HMM + AL, the second one is composed by HDP-HMM LITE + Adaptation, and our proposed algorithm is referred to as HDP-HMM + AL + Adaptation, here AL is referred to our proposed AdaBoost-like feature selection algorithm, and LITE is referred the feature selection algorithm proposed in[14], which is proven to be effective to the problem of activity recognition also based on data set[28].

We randomly select a 50% subset of normal event sequences(50 sequences) for training, these sequences are of 13KB size in average. The other 50 normal traces and all the 20 abnormal traces were randomly mixed together for testing. Fig. 1 shows the ROC curve with respect to the detection rate and the false alarm rate.

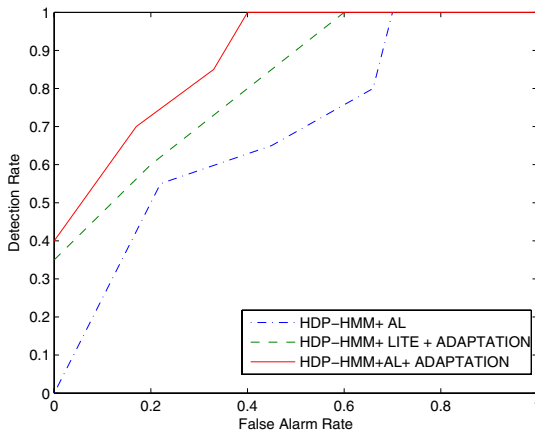


Fig. 1. Comparison of the detection rate and the false alarm rate

We can see from Figure 1 that, HDP-HMM + AL gives the poorest detection result because it achieves a high detection rate at the cost of incurring a high false alarm rate. From Figure 1, it is also noteworthy to mention that our AL method for feature selection performs much better than the LITE algorithm, which will achieve a perfect detection rate of 1 at the cost of 0.4 false alarm rate, while LITE algorithm requires a 0.6 False Alarm Rate when the detection rate is 1.

Furthermore, after we adopted the Adaptation technique, both HDP-HMM + LITE + Adaptation and HDP-HMM + AL + Adaptation received better results. This is because the adapted abnormal activity models contribute to distinguishing normal and abnormal activities, which can significantly reduce the false alarm rate, in this sense the formerly misclassified normal events get a second chance to be differentiated from the abnormal events. The essential advantage of AL, compared to LITE is that it not only gains the information of how the feature contributes to classification.

5 Conclusions

In this paper, we propose a novel approach for abnormal event detection. To deal with the scarcity of training data for abnormal events, we proposed a three-phase abnormality detection algorithm. In the first phase, we train a set of weak classifiers based on HDP-HMM, in particular, we have chosen beam sampling for inference, which combines slice sampling with dynamic programming. In the second phase, we developed an ensemble learning algorithm to filter out abnormal events, and the suspicious traces are then passed on to a collection of abnormal events models adapted from the general model for further detection. The major advantage of our approach is that HDP-HMM can naturally capture the underlying features and the hierarchical detection phase can achieve a better tradeoff between detection rate and false alarm rate. We demonstrate the effectiveness of our approach using real data collected by CAVIAR PROJECT.

One limitation of our approach is the adaptation technique we adopted in this framework. We have chosen HDP-HMM because it does not select a fixed number of states, and a full posterior density function is inferred for all model parameters, this introduces one problem: since the number of parameters may change after the adaptation, as a result, we can not employ classic adaptation techniques such as MLLR[6]. Our solution is just based on the sequential nature of beam sampler but not consider the mismatch between an initial model and the adaptation data. Therefore we need to be able to develop an adaptation technique to address this problem. Another latent problem involved with the adaptation step is that there is a risk of generating a large number of abnormal models when abnormal activities suddenly become the norm, as suggested in [7], we need to detect when abnormal activities turns normal and vice versa.

Acknowledgement

We would like to acknowledge support for this project from the National Science Foundation of China(NSFC grant No.60775046 and No.60721002) and the

National Grand Fundamental Research 973 Program of China (grant No.2009CB320700).

References

1. Yao, Y., Wang, F., Wang, J., Zeng, D.D.: Rule+Exception Strategies for Security Information Analysis. *IEEE Intelligent Systems* 20(5), 52–57 (2005)
2. Wang, J., Xu, C., Chng, E.S., Tian, Q.: Sports highlight detection from keyword sequences using HMM. In: *Proc. IEEE ICME, Taiwan* (June 2004)
3. Zhang, D., Gatica-Perez, D., Bengio, S., McCowan, I.: Semi- Supervised Adapted HMMs for Unusual Event Detection. In: *CVPR*, pp. 611–618 (2005)
4. MacKay, D.J.C.: Ensemble learning for hidden Markov models. *Tech. Rep., Cavendish Lab, Univ. Cambridge, Cambridge, U.K* (1997)
5. Gong, S., Xiang, T.: Recognition of group activities using a dynamic probabilistic network. In: *Proc. IEEE ICCV, Nice* (October 2003)
6. Lester, J., Choudhury, T., Kern, N., Borriello, G., Hannaford, B.: A Hybrid Discriminative/Generative Approach for Modeling Human Activities. In: *Proc. 19th Int'l Joint Conf. Artificial Intelligence (IJCAI 2005), July-August 2005*, pp. 766–772 (2005)
7. Yin, J., Yang, Q., Pan, J.J.: Sensor-based Abnormal Human-Activity Detection *IEEE Transactions on Knowledge and Data Engineering (IEEE TKDE)*. IEEE Computer Society Digital Library. IEEE Computer Society, Los Alamitos (2007)
8. Pruteanu-Malinici, I., Carin, L.: Infinite Hidden Markov Models for Unusual-Event Detection in Video. *IEEE Transactions On Image Processing* 17(5) (May 2008)
9. Bradley, A.P.: Shift-invariance in the discrete wavelet transform. *Digit. Image Comput. Tech. Appl.*, pp. 29–38 (December 2003)
10. Hurri, J., Hyvarinen, A., Karhunen, J., Oja, E.: Image feature extraction using independent component analysis. In: *The IEEE Nordic Conf. Signal Processing* (1996)
11. Kohonen, T.: Emergence of invariant-feature detectors in the adaptive- subspace self-organizing map. *Biol. Cybern.* 75(4), 281–291 (1996)
12. Stauffer, C., Grimson, W.E.L.: Learning patterns of activity using real-time tracking. *IEEE Trans. Pattern Anal. Mach. Intell.* 22(8), 747–757 (2000)
13. Viola, P., Jones, M., Snow, D.: Detecting pedestrians using patterns of motion and appearance. *IJCV* 63(2), 153–161 (2005)
14. Ribeiro, P.C., Santos-victor, J.: Human activity recognition from video: modeling, feature selection and classification architecture. In: *International Workshop on Human Activity Recognition and Modeling (HAREM)* (2005)
15. Kira, K., Rendell, L.: A practical approach to feature selection. In: *Proc. 9th Int. Workshop on Machine Learning*, pp. 249–256 (1992)
16. Robnik-Sikonja, M., Kononenko, I.: Theoretical and Empirical Analysis of ReliefF and RReliefF. *Machine Learning Journal* 53, 23–69 (2003)
17. Teh, Y.W., Jordan, M.I., Beal, M.J., Blei, D.M.: Hierarchical dirichlet processes. *J. Amer. Statist. Assoc.* 101, 1566–1581 (2006)
18. Gael, J.V., Saatci, Y., Teh, Y.W., Ghahramani, Z.: Beam Sampling for the Infinite Hidden Markov Model. In: *ICML* (2008)
19. Elkan, C.: The Foundations of Cost-Sensitive Learning. In: *Proc. 17th Int'l Joint Conf. Artificial Intelligence (IJCAI 2001)*, pp. 973–978 (August 2001)

20. Ling, C.X., Huang, J., Zhang, H.: AUC: A Statistically Consistent and More Discriminating Measure than Accuracy. In: Proc. 18th Int'l Joint Conf. Artificial Intelligence (IJCAI 2003), pp. 519–526 (August 2003)
21. Liu, X.-Y., Zhou, Z.-H.: The influence of class imbalance on cost-sensitive learning: An empirical study. In: Perner, P. (ed.) ICDM 2006. LNCS, vol. 4065, pp. 970–974. Springer, Heidelberg (2006)
22. Yilmaz, A.: Object Tracking by Asymmetric Kernel Mean Shift with Automatic Scale and Orientation Selection. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2007) (2007)
23. MacKay, D.J.C.: Ensemble learning for hidden Markov models. Tech. Rep., Cavendish Lab, Univ. Cambridge, Cambridge, U.K (1997)
24. Hongeng, S., Nevatia, R., Bremond, F.: Video-based event recognition: activity representation and probabilistic recognition methods. *Computer Vision and Image Understanding* 96, 129–162 (2004)
25. Rabiner, L.R.: A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proc. IEEE* 77(2), 257–286 (1989)
26. Beal, M.J., Ghahramani, Z., Rasmussen, C.E.: The infinite hidden Markov model. In: *Advances in Neural Information Processing Systems*, pp. 577–584. MIT Press, Cambridge (2002)
27. Viola, P., Jones, M.: Rapid Object Detection using a Boosted Cascade of Simple Features. *Computer Vision and Pattern Recognition* (2001)
28. CAVIAR PROJECT, <http://homepages.inf.ed.ac.uk/rbf/CAVIAR/>
29. Chan, M.T., Hoogs, A., Schmiederer, J., Perterson, M.: Detecting rare events in video using semantic primitives with HMM. In: Proc. ICPR (August 2004)
30. Runkle, P., Bharadwaj, P., Carin, L.: Hidden Markov model for multi-aspect target classification. *IEEE Trans. Signal Process.* 47(7), 2035–2040 (1999)