2012 9th International Conference on Ubiquitous Intelligence and Computing and 9th International Conference on Autonomic and Trusted Computing

AN INTELLIGENT VIRTUAL FENCE SECURITY SYSTEM FOR THE DETECTION OF PEOPLE INVADING

Jun-Horng Chen, Teng-Hui Tseng, Chin-Lun Lai, and Sheng-Ta Hsieh*

Oriental Institute of Technology
Department of Communication Engineering
New Taipei City, Taiwan
{jhchen,alex,fo001,fo013}@mail.oit.edu.tw

ABSTRACT

This work proposes an intelligent and interactive virtual fence system for security management in a danger- or accident-prone area. The proposed system is designed to be realized in an outdoor and wide-area park. With the proposed system, users can visually and arbitrarily define the polygonic alert area for the sake of security. The Gaussian Mixture Model (GMM) and AdaBoost algorithms are modified and integrated in this work to provide a robust human detection. As soon as the invading people is detected and located in the interior of the alert area, the proposed system will immediately alarm the security guards by MMS of cellular networks or a customized App installed in smart phones. Besides, this work also proposes a sliding-window based alarm mechanism to practically and robustly improve the detection rate and reduce the false alarm rate of detection. The experimental results will demonstrate that the proposed system can be indeed applied in a real field and provide a satisfied performance.

Index Terms— Virtual Fence, Gaussian Mixture Model, Pedestrian Detection, AdaBoost Algorithm

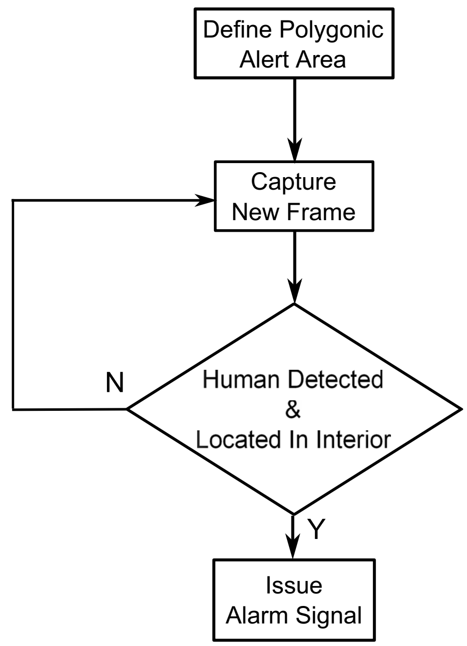


Fig. 1: Operation flowchart of the proposed system.

1. INTRODUCTION

The virtual fence is designed to serve as an enclosure, a barrier, or a boundary, which is relied on other than physical objects. Depending on different concepts of its application, there are various kinds of structures, while “without physical barrier” is their common point [1]. The applications of virtual fence can be in agriculture for grazing animal [2] [3] [1], in fleet or freight management [4], in homeland security [5], in intelligent surveillance systems of public or private space [6] [7] [8] [9], *et cetera*. The main goal of this paper is the applications of intelligent surveillance systems.

As the promising and advanced technology of computer vision, this work will proposed an intelligent and interactive virtual fence system for security management in a danger- or accident-prone area. There is no mandatory facility on site but IP cameras. The landscape is accordingly remained the same as the original. A server will remotely access all the video signals and analyze the content for human detection. With the proposed system, users can visually and arbitrarily point out the polygonic alert area for the sake of security. As soon as the invading people is detected and located in the interior of the alert area, the system will immediately alarm the security guards by *Multimedia Messaging Service* (MMS) of cellular networks or a customized App installed in smart phones. The operation flowchart of the proposed system is shown in Fig. 1.

In this paper, Sec. 2 will introduce some off-the-shelf approaches, which are the key technology adopted in the system. The system architecture and design will be discussed in Sec. 3. Finally, some experimental results will demonstrate

This work was supported by the Far Eastern Resource Development Co. -Tpark, Taiwan, under contract RD1000077.

978-0-7695-4843-2/12 $26.00 © 2012 IEEE
DOI 10.1109/UIC-ATC.2012.64

the performance of the proposed system in Sec. 4, which is followed by a brief conclusion in Sec. 5.

2. KEY TECHNOLOGY

2.1. GMM Background Modeling

In the application scenario of the proposed system, the alert area pre-defined by the user is supposed to be human-free for security reason. Accordingly, establishing the background model and detecting the foreground object, *i.e.*, the moving object, is one the most intuitive approaches. Stauffer and Grimson [10] proposed the well-known *Gaussian Mixture Model* (GMM) approach to model the multi-modal background image by a mixture of normal distributions. Power and Schoonees [11] provided a good tutorial to understand the GMM algorithm, and showed how to improve it. In [12], the authors modified the GMM approach by integrating the optical flow method for accurately extracting the shapes of the moving objects.

The GMM approach assumes the background is visible more frequently than the moving objects in the foreground, such that the background pixels can be modeled as the mixture of few Gaussian distributions with relative small variance which is due to surface texture, illumination fluctuations, or camera noise. Accordingly, in the pixel-wise GMM approach, each pixel is represented by one state s of a set $s \in \{1, 2, \cdots, S\}$ which is composed of background and foreground surfaces. A process $\mathbf{s}$ generates the state at each frame time t according to the *a priori* probability $Pr(s)$, and $\sum_{s=1}^{S} Pr(s) = 1$. For a given state s, the pixel values are samples of random variable $\mathbf{x}$ with a multivariate distribution of:

$$f_{\mathbf{x}|s}\left(\mathbf{x}|s,\theta_s\right) = \frac{1}{(2\pi)^{\frac{n}{2}} \left|\Sigma_s\right|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{x}-\mu_s)^t \Sigma_s^{-1} (\mathbf{x}-\mu_s)}, \quad (1)$$

where θ_s is the state parameter composed of mean μ_s and variance Σ_s. The dimension n is one for gray-level images and three for color images.

Let Ω be the parameter set of all states including *a priori* probability, *i.e.*, $\Omega = \{\theta_1 \cdots \theta_S\} \cup \{Pr(1), \cdots, Pr(S)\}$, the distribution $\mathbf{x}$ can thus be modelled as the sum of Gaussian mixture

$$f_{\mathbf{x}}\left(\mathbf{x}|\Omega\right) = \sum_{s=1}^{S} Pr(s) f_{\mathbf{x}|s}\left(\mathbf{x}|s,\theta_s\right). \quad (2)$$

In [10], the authors proposed, by the *expectation maximization* (EM) approach [13], the hidden state and parameters of mixture model in Eq. (2) can be iteratively estimated according to the observed data.

Firstly, in the *E-step* of EM algorithm, which of the S surfaces should be determined according to the current sample. The decision should be made by the consideration of which of the S distributions most likely leads to the current observation $\mathbf{x}$. That is, the s which maximizes the likely function

$$\begin{aligned} L(s) &= P\left(s|\mathbf{x},\Omega\right) \\ &= \frac{Pr(s) f_{\mathbf{x}|s}\left(\mathbf{x}|s,\theta_s\right)}{f_{\mathbf{x}}\left(\mathbf{x}|\Omega\right)} \\ &= \frac{Pr(s) f_{\mathbf{x}|s}\left(\mathbf{x}|s,\theta_s\right)}{\sum_{s=1}^{S} Pr(s) f_{\mathbf{x}|s}\left(\mathbf{x}|s,\theta_s\right)} \end{aligned} \quad (3)$$

is the *maximum a posteriori* (MAP) estimate

$$\begin{aligned} \hat{s} &= \arg\max_s L(s) \\ &= \arg\max_s Pr(s) f_{\mathbf{x}|s}\left(\mathbf{x}|s,\theta_s\right) \end{aligned} \quad (4)$$

Obviously, to accomplish the MAP estimate of Eq. (4), all the parameters in Ω should be estimated. This is the major task in the *M-step* of EM algorithm of which detail was give in [10].

2.2. AdaBoost Learning For Human Detection

The AdaBoost algorithm [14] was proposed to solve the difficulties of learning problems in boosting. The authors also gave a short introduction [15] of it. However, the most famous applications of Adaboost algorithm is proposed by Viola and Jones in [16], which used this algorithm to robustly detect the human face in images. Since then, a wide-area of applications used AdaBoost algorithm to solve their learning problem. For example in [17], the authors used the AdaBoost algorithm to detect pedestrian.

It is reported that feature-based detection system has better performance than pixel-based. However, manipulating a large set of features is very time-consuming. The key idea of the AdaBoost algorithm in [16] is to select few features from a very large set of potential features. In [16], the authors combined a series of weak classifies into a strong classifiers. Accordingly, the face-like regions can be quickly processed and background regions can be also quickly discarded. Another contribution of [16] is the concept of "Integral Image" which makes the features be computed very quickly. As [16], this work uses Haar-like rectangle features computed on a pyramid of images, that is, each image is scanned at various scales.

In each iteration of learning process, one feature is selected and the corresponding weak classifier is determined. The so-called weak classifier h can be formulated as:

$$h\left[f(\mathbf{x}), p, \theta\right] = \begin{cases} 1, & \text{if } p \cdot f(\mathbf{x}) < p \cdot \theta \\ 0, & \text{otherwise} \end{cases}, \quad (5)$$

where $\mathbf{x}$ is the input pixel vector of hard examples, $f(\mathbf{x})$ is the corresponding extracted feature, θ is the threshold, and p controls the direction of inequality. In the combined strong classifier, the proportion of each weak classifier is determined

Table 1: The main principles and procedures of AdaBoost algorithm.

• Given a hard sample set $\mathcal{S} = \mathcal{S}^- \cup \mathcal{S}^+ = \{(\mathbf{x}_i, y_i)\}$, where $\mathbf{x}_i$ is pixel vector and $y_i \in \{0, 1\}$ is its corresponding label representing negative and positive sample respectively.

• The initial weight is $w_{1,i} = 1/(2\|\mathcal{S}^-\|)$ and $1/(2\|\mathcal{S}^+\|)$ for each negative and positive sample respectively.

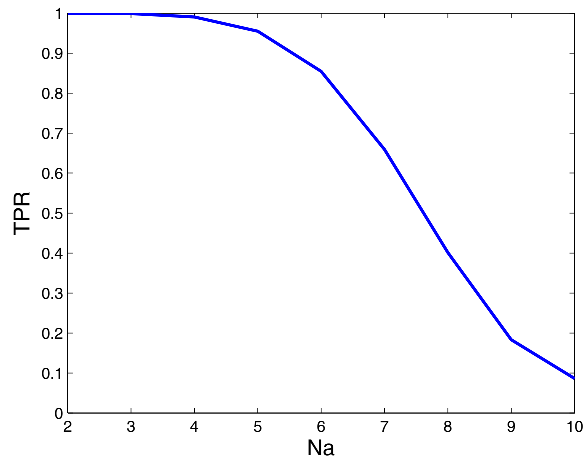
• Let N be the expected number of weak classifiers (*i.e.*, there are totally N iterations in training process), in the n-th iteration,

1. All weights should be normalized such that all of them can add up to one.
2. Select the best weak classifier $h_n\left[f(\mathbf{x}), p, \theta\right]$ which induces the minimized weighting classification error:
$$\epsilon_n = \min_{f,p,\theta} \sum_i w_{n,i} \left| h\left[f(\mathbf{x}), p, \theta\right] - y_i \right| \ .$$
3. Update the wights for emphasizing the significance of those samples incorrectly classified by $h_n\left[f(\mathbf{x}), p, \theta\right]$, that is
$$w_{n+1,i} = \begin{cases} w_{n,i} \cdot \dfrac{\epsilon_n}{1-\epsilon_n}, & h_n\left[f(\mathbf{x}_i), p, \theta\right] = y_i \\ w_{n,i}, & h_n\left[f(\mathbf{x}_i), p, \theta\right] \neq y_i \end{cases}$$

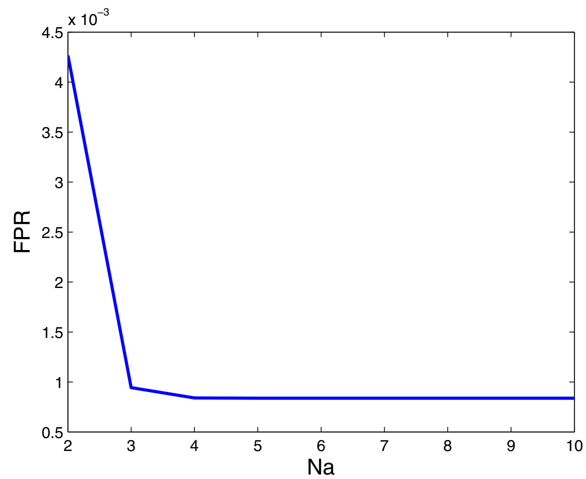
• Because those weak classifiers with lower classification error should be more significantly weighted, the final strong classifier will be

$$H(\mathbf{x}) = \begin{cases} 1, & \sum_{n=1}^{N} \alpha_n h_n\left[f(\mathbf{x}), p, \theta\right] \geq \frac{1}{2} \sum_{n=1}^{N} \alpha_n \\ 0, & \text{otherwise} \end{cases} ,$$

where
$$\alpha_n = \ln \frac{1-\epsilon_n}{\epsilon_n} \ .$$



(a) **TPR**



(b) **FPR**

Fig. 2: The **TPR** and **FPR** for N_b=2.

just when it was selected and related to its classification error. Obviously, those weak classifiers with lower classification error should be more significantly weighted. As for those examples incorrectly classified in the previous weak classifier, they should be re-weighted to emphasize their significance in the next iteration. Table 1 summarizes the main principles and procedures of AdaBoost algorithm.

2.3. Sliding-Window Based Alarm Mechanism

Since the proposed system is mainly applied to detecting people in the alert area, there is no need to alarm the security guard for detecting invading people in only one image frame. Any detection system may have false positive detection. If the alarm is enabled for every invading detection, the security guard will be very tired of those false alarms and thus

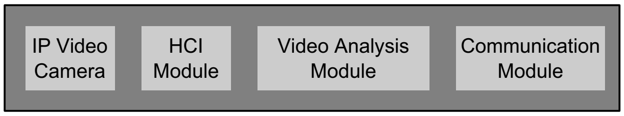


Fig. 3: The function diagram of the proposed system.

involuntarily lose their attention. Furthermore, people is impossibly entering and leaving the alert area in a very short duration. Therefore, this work proposed the *sliding-window based alarm mechanism* which sends the alarm out based on the detection result of consecutive multiple frames instead of a single frame. The so-called sliding window is a fixed length of consecutive frames and shifts with respect to time.

Let N be the length of sliding windows. In the proposed system, the alarm occasion will be the detection of invading people in more than N_a frames or a N_b-frame detection streak. Let **TPR** and **FPR** be the *true positive rate* and *false positive rate* of sliding-window based alarm mechanism, then

$$\begin{aligned}
\mathbf{TPR} &= \sum_{k=N_a}^{N} \binom{N}{k} (\mathbf{tpr})^k (1-\mathbf{tpr})^{N-k} \\
&\quad + \sum_{k=N_b}^{N_a-1} (N-k+1)(\mathbf{tpr})^k (1-\mathbf{tpr})^{N-k} \\
\mathbf{FPR} &= \sum_{k=N_a}^{N} \binom{N}{k} (\mathbf{fpr})^k (1-\mathbf{fpr})^{N-k} \\
&\quad + \sum_{k=N_b}^{N_a-1} (N-k+1)(\mathbf{fpr})^k (1-\mathbf{fpr})^{N-k} \qquad (6)
\end{aligned}$$

where **tpr** and **fpr** are the *true positive rate* and *false positive rate* of the detection in a single frame, respectively. Figure 2 shows the **tpr** and **fpr** with respect to N_a for the case of $N = 10$, $N_b = 2$, **tpr** $= 0.7$, and **fpr** $= 0.01$. Obviously, as long as a proper N_a is chosen, the **TPR** and **FPR** can be significantly increased and decreased compared to **tpr** and **fpr**, respectively.

3. SYSTEM DESIGN OVERVIEW

This section will discuss the design and implementation of each function in the proposed system, as shown in Fig. 3. Initially, the proposed system is designed to applied to a wide-area park in Taiwan. Therefore, to acquire a video signal with an acceptable quality, IP video cameras are equipped near the alert area such that the server can remotely access the video signal via the RTSP streaming protocol. As long as the network bandwidth is sufficient, the video quality is always better than that of analog video cameras of which signal is inevitably attenuated in such a long-distance transmission. Besides, the proposed system uses an infrared camera to make the system can normally operate all day.

The Human-Computer Interaction (HCI) module provides the user to visually and arbitrarily define the polygonic alert area. Once the vertices of the polygon are pointed out, the HCI module will form the corresponding polygon and show it on the screen.

The video analysis module handles the content analysis of the acquired video signal. Based on the aforementioned algorithms, this work proposed the following approaches to detect the invading people:

a. **Background substraction with size screening:** The GMM algorithm is applied in this approach to estimate the background regions. Therefore, the foreground object is easily extracted by subtraction. However, since the camera is used in an outdoor area, many motion events could be mistakenly detected as invading people. For example, strong winds may shake the camera post and blow the water ripples, flying birds may pass through the alert area, and *et cetera*. Accordingly, this approach screens the foreground objects with their size because the size of a human silhouette should be reasonable.

b. **Global human detection by Adaboost:** By the success of Adaboost algorithm in face detection, this approach analogously apply it to detect invading people. However, the temporal information is useless in this approach. Besides, the samples in learning process should be carefully considered because the view angles of cameras are various and dependent on their usage.

c. **Human detection by AdaBoost in foreground region:** This approach aims to combine the GMM and AdaBoost algorithms to detect invading people. The GMM algorithm is firstly applied to extract the foreground moving objects, and followed by applying the AdaBoost to determine whether those moving objects are human or not.

d. **Human detection by AdaBoost in foreground region with size screening:** This approach is the enhanced version of the aforementioned approach. The detected people is further verify by the silhouette size.

The communication module designed to issue the alarm to the security guards. It can be a 3G modem to send an MMS alert out, which may include the screenshot of the invading people. Additionally, it can be a program communicated with the App installed in smart phones.

4. EXPERIMENTAL RESULTS AND DISCUSSION

The core functions the proposed system are realized in a computer equipped with a touchscreen for the implementation of HCI functions, and a GPRS/UMTS/HSDPA modem for the

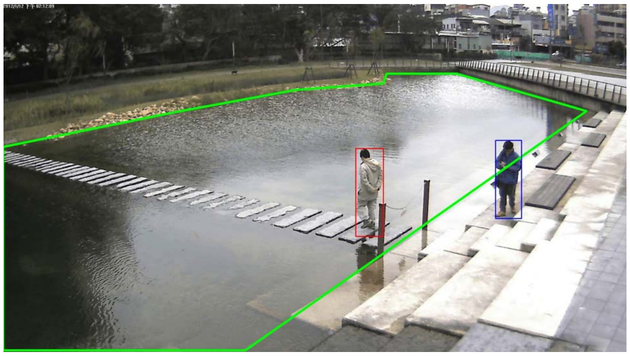


Fig. 4: Demonstration of the proposed system. The polygon bounded in green is the preset alert region. Two people bounded in red and blue are the detected people inside and outsize the alert region, respectively.

Table 2: The **tpr** and **fpr** performance of the approaches proposed in Sec. 3

	a	b	c	d
tpr	0.50	0.52	0.79	0.78
fpr	0.001	0.001	0.04	0.003

implementation of communication functions. The users can thus define the alert area on demand. As shown in Fig. 4, the polygon bounded in green defined by the users is the preset alert region. Two people bounded in red and blue are the detected people inside and outsize the alert region, respectively.

Table 2 shows the performance comparison of those approaches proposed in Sec. 3. All the **tpr** and **fpr** shown in table. 2 is measured based on the single frame detection. The well-known fact in detection systems is the **tpr** may be increased by adjusting the parameters, nevertheless the **fpr** will accordingly be increased also. To avoid unacceptable **fpr**, the approach of Sec. 3a can only have the performance of **tpr**=0.5, which is the worst performance among those four approaches. This is not so surprising because many motion events could be mistakenly detected as invading people in outdoor field. The approach of Sec. 3b does not perform significantly better than the approach of Sec. 3a. A reasonable explanation is the existing camera in this case is too far from the target field such that the captured images are with low-resolution, which can not provide enough and effective pixel information for AdaBoost algorithm.

The approaches of Sec. 3c and 3d combine the GMM and AdaBoost algorithms to increase the **tpr** and decrease the **fpr**. Furthermore, the approach of Sec. 3d filters out the objects with unreasonable size to decrease the **fpr**, and thus performs the best outcome among those approaches. Figure 5 shows

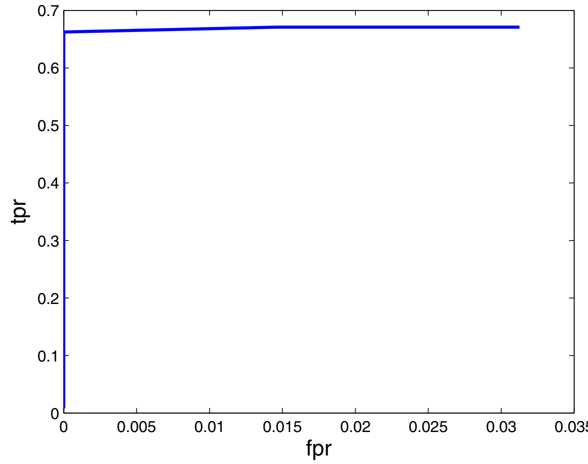


Fig. 5: The Receiver Operating Characteristic (ROC) curve of the approach Sec. 3d

the Receiver Operating Characteristic (ROC) curve of the approach of Sec. 3d.

To examine the performance of the *sliding-window based alarm mechanism* in Sec. 2.3, this work set $N = 5$, $N_b = 2$, and $N_a = 2$ or 3 and then measured the **TPR** and **FPR**. As expected, the **FPR** is vanish, and the **TPR** is 0.93 and 0.83 for N_a=2 and 3 respectively. This result is almost agreed with the Eq. (6).

5. CONCLUSIONS

This work proposed a practical virtual fence system which can be realized in a wide-area park. To improve the detection performance, this work utilized and modified the off-the-shelf technology, GMM and AdaBoost algorithms in the detection of invading people. Furthermore, this work proposed the Sliding-Window Based Alarm Mechanism, which can effectively increase the true positive rate and decrease the false positive rate of detection. Besides, the proposed system provide a user-friendly HCI module which makes the users can visually and arbitrarily define the polygonic alert area for the sake of security. The whole proposed system has already been installed in a wide-area park of Taiwan, and provides a satisfied performance.

6. REFERENCES

[1] Christina Umstatter, "The evolution of virtual fences: A review," *Computers and Electronics in Agriculture*, vol. 75, pp. 10–22, 2011.

[2] Zack Butler, Peter Corket, Ron Peterson, and Daniela Rust, "Virtual fences for controlling cows," in *Proceed-*

ings of the IEEE International Conference on Robotics and Automation, 2004, pp. 4429–4436.

[3] M. O. Monod, P. Faure, L. Moiroux, and P. Rameau, “A virtual fence for animals management in rangelands,” in *Proceedings of the 14th IEEE Mediterranean Electrotechnical Conference (MELECON 2008)*, 2008, pp. 337–342.

[4] Fabrice RECLUS and Kristen DROUARD, “Geofencing for fleet & freight management,” in *Proceedings of the 9th IEEE Intelligent Transport Systems Telecommunications Conference (ITST 2009)*, 2009, pp. 353–356.

[5] Robert N. Charette, “The virtual fence’s long goodbye,” *IEEE Spectrum*, vol. 48, pp. 8–8, 2011.

[6] J.L. Castro, M. Delgado, J. Medina, and M.D. Ruiz-Lozano, “Intelligent surveillance system with integration of heterogeneous information for intrusion detection,” *Expert Systems with Applications*, vol. 38, pp. 1182–1192, 2011.

[7] Ruiquan Ge, Zhenfang Shan, and Hao Kou, “An intelligent surveillance system based on motion detection,” in *Proceedings of the 4th IEEE International Conference on Broadband Network and Multimedia Technology (IC-BNMT)*, 2011, pp. 306–309.

[8] Durgesh Patil, Sachin Joshi, Milind Bhagat, and Swapnil Aundhkar, “Survey on wireless intelligent video surveillance system using moving object recognition technology,” *Computer Engineering and Intelligent Systems*, vol. 2, pp. 25–30, 2011.

[9] Frederick Tung, John S. Zelek, and David A. Clausi, “Goal-based trajectory analysis for unusual behaviour detection in intelligentsurveillance,” *Image and Vision Computing*, vol. 29, pp. 230–240, 2011.

[10] Chris Stauffer and W. E. L. Grimson, “Learning patterns of activity using real-time tracking,” *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, pp. 747–757, 2000.

[11] P. Wayne Power and Johann A. Schoonees, “Understanding background mixture models for foreground segmentation,” in *Proceedings of the Image and Vision Computing Conference, New Zealand*, 2002, pp. 267–271.

[12] Dongxiang Zhou and Hong Zhang, “Modified gmm background modeling and optical flow for detection of moving objects,” in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, 2005, pp. 2224–2229.

[13] A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum likelihood from incomplete data via the em algorithm,” *JOURNAL OF THE ROYAL STATISTICAL SOCIETY, SERIES B*, vol. 39, no. 1, pp. 1–38, 1977.

[14] Yoav Freund and Robert E. Schapire, “A decision-theoretic generalization of on-line learning and an application to boosting,” in *Proceedings of the Second European Conference on Computational Learning Theory*, 1995, pp. 23–37.

[15] Y. Freund and R. Schapire, “A short introduction to boosting,” *Japonese Society for Artificial Intelligence*, vol. 14, pp. 771–780, 1999.

[16] Paul Viola and Michael Jones, “Robust real-time face detection,” *International Journal of Computer Vision*, vol. 57, pp. 137–154, 2004.

[17] D. Gerónimo, A. Sappa, A. López, and D. Ponsa, “Pedestrian detection using adaboost learning of features and vehicle pitch estimation,” in *Proceedings of the International Conference on on Visualization, Imaging, and Image Processing*, 2006, pp. 400–405.