

# A Combination of Off-line and On-line Learning to Classifier Grids for Object Detection

Nguyen Dang Binh<sup>1,\*</sup>,

<sup>1</sup>Hue University of Sciences, 77 Nguyen Hue street, Hue city, Vietnam

## Abstract

We propose a new method for object detection by combining off-line and on-line boosting learning to classifier grids based on visual information without human intervention concerned to intelligent surveillance system. It allows for combine information labeled and unlabeled use different contexts to update the system, which is not available at off-line training time. The main goal is to develop an adaptive but robust system and to combine prior knowledge with new information in an unsupervised learning framework that is learning 24 hours a day and 7 days a week. We use co-training strategy by combining off-line and on-line learning to the classifier grids. The proposed method is practically favorable as it meets the requirements of real-time performance, accuracy and robustness. It works well with reasonable amount of training samples and is computational efficiency. Experiments on detection of objects in challenging data sets show the outperforming of our approach.

**Keywords:** context-based learning, classifier grids, object detection, online learning.

Received on 13 April, 2016, accepted on 20 April, 2016, published on 02 May, 2016

Copyright © 2016 *Nguyen Dang Binh*, licensed to EAI. This is an open access article distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/3.0/>), which permits unlimited use, distribution and reproduction in any medium so long as the original work is properly cited.

doi: 10.4108/eai.2-5-2016.151208

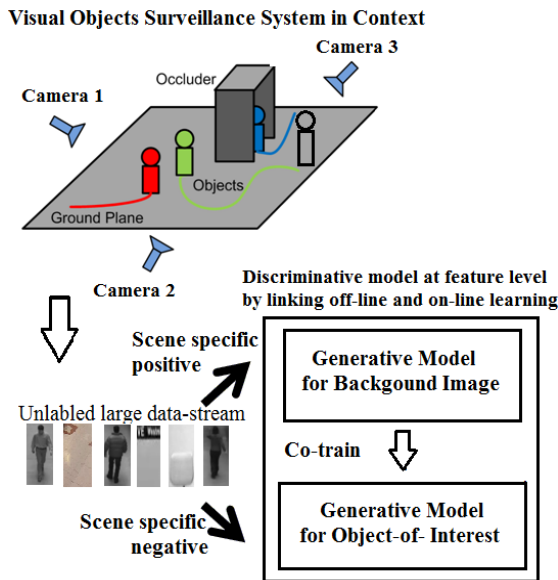
## 1. Introduction

Today the increasing number of surveillance cameras which should have a self-operated monitoring surveillance systems. It is an important thing. Researches focused on detecting the object from the static cameras emphasis on applications in the real world environment in which context information plays an important role. Robust machine learning system to detect objects can be applied in many different fields. For example intelligent surveillance systems, human-computer interaction, information safety, machine vision, data mining. One of the main challenges is how machine learning can combine the information which has not been labeled in different contexts to update the classifications but at the same time preserving the accuracy of the object detection for long periods. This is a very important requirement for the system to detect objects in real time. It may help the selection of information for updating the system better,

limit the changes and reduce the size of the training sample. In many applications, object detection problem is much simpler, such a monitoring system with a still camera 24/7 focusing on one type of object in the same scene. Our approach is a machine learning system having access to large data stream and using different contexts and constantly updated information for specific classifier to detect objects of a class in any context [22]. Moreover, the visual data stream is continuously supplied a large amount of unlabeled data but must be explored to improve detection results as well as the speed of the object detection. The unlabeled large data-stream is analyzed and specific scene of positive and negative samples are collected for continuously updating the classifiers. One simple way to benefit from a static camera is to combine new information unlabeled on the specific situation and data labeled for improving current classifier. However such information only helps reduce the number of false detection [10]. To increase the object detection rate accurately, adaptive learning methods online are

\*Corresponding author. Email: [ndbinh@hueuni.edu.vn](mailto:ndbinh@hueuni.edu.vn)

applied [18]. This method focuses on solving object detection task in a specific scene and has many advantages for data stream to be continuous. In fact, this approach used in the context in training process and post-processing. Thus, online unsupervised learning methods often used to continuously adapt a model. The main problem is how to be able to update the new data which has not been labeled, but to reduce the update failed for the classification.



**Figure 1.** Combination of two generative models in one discriminative model at feature level by linking off-line and on-line learning in context.

In other words, that is to reduce the number in detection objects drifting. Several prominent researchers applied sliding windows technique [6, 7, 8, 16]. The classifier performs a search greed, from top to bottom, from left to right, each area of the image corresponding to the difference in size will be tested. The classifier will estimate the value of the image area corresponding to the object to be detected or not. The corresponding area of the image to match the object will be marked on the image results. Typically, the goal of this approach is that the classification has been chosen for a specific object and be applied to different object detection in different environments [7, 8, 12]. However, if classifier trained a large number of samples to the problem of object detection in general often fail in specific situations. The performance of the detector will have low accuracy. The use of fixed cameras with reasonable constraints for the majority of applications used for a particular situation may help reduce the number of objects to be in detection [10]. In fact, the trained classifier with just a few training data for a specific problem often results in efficient and accurate objects detection. The methods of object detection using traditional models and the sliding window of object search are used a single classification for finding and developing the object on the whole picture

[1, 2, 3, 4, 5]. The training to form the aforementioned classification learning using offline methods will then encounter the following problems: The first, to establish the classification in the offline model of traditional methods (such as support vector machines, neural networks, boosting) the set of training samples must be collected and prepared. Therefore, this dataset can be very large and be up to several thousand samples, depending on the specific application. The preparation of training sample sets is usually made by hand resulting in much time and effort. Also, using classifier on a prepared set sample but when applied on a new context; it should not be effectively promoted. For the effective performance of the system, it must train new additional models to adapt to the new object in the scene. Hence, there should be a new approach in training the classifier. It is online learning. The second, after training completion, object detection must use greedy search methods for all positions and different sizes of objects. Object Detection not only on the current image but also on consecutive frames; so the complexity of the problem of object detection increases. The third, in order to extract features sample learning, the system normally uses a specific method but applied across the sample set. Therefore using a features method of selecting specific extract is only suitable for this sample but not matching other sample. This means not getting the best features for each different sample. Such as the features geometry, textures, colors, shapes. In addition to the system that can be applied in real-time applications, the fast calculation method for selecting features extract is also an example. Thus the research could use a variety of methods for features extraction. How to train each sample to choose the most suitable features for that sample is one of the concerns of our research.

To overcome these limitations, in this research, we use the feature selection methods such as Haar wavelet, Local binary pattern and Histogram of Oriented. To reduce the complexity of the greedy search for object detection, we use grids classifier model in which each set of classifications that will be used with a few parameters and undertaking the search object in a smaller scope. The classification is done to update and detect simultaneously and allowed to adapt to the current context. Additionally our research uses fast calculation method on the integral image. We propose to use a co-training approach in combination with a novel robust on-line learner. The robust on-line learner keeps two separate models for the positive class as well as two separate models for the negative class. For both, the positive and the negative class, one model is off-line trained and kept fixed during run-time, while the other one is adapted over time. This combination of an off-line model with an on-line adapted model within a single learner allows for incorporating scene specific positive information (i.e., the recall can be increased), while still preserving the accuracy. Our approach to classifier grids in object detection problem from one or multiple cameras in which may combine information unlabeled and labeled use

different contexts to collect samples to update the system from the background image and object classes over time.

The rest of the paper is structured as follows. First, in Section 2, we mention the issues related to recent researches. Next, we consider the idea of combination of off-line and on-line learning for classifier grids in Section 3. We give an empirical evaluation of the approach experimental evaluation and results in Section 4. Finally, we conclude with an outline of research problem in this direction in Section 5.

## 2. Related Works

Recently, to improve the performance of the classifier and reduce the number of samples on the training set, researches using online learning algorithms have withdrawn much interest from researchers worldwide [11, 16, 18]. These algorithms allow to update the information necessary to suit each specific context. This facilitates adapting to each learning sample for the improved system performance of the classifier which is enabling significantly. Also it helps to reduce the number of training samples while assuring the incremental learning. Under this approach, the algorithm often uses self-training [14, 17] and learning combined sample which was labeled and unlabeled in a co-training framework [4, 13] or self-generated sample of learning training process, the method of unsupervised learning [16]. For semi-supervised learning methods, often using a combination of information which was given to exploring the unknown information to form the classification. Self-training method commonly encounters determining updated sample correctly because it is only based on the feedback of the current classification at that time. This is very important because it will affect the performance of the classification, as well as the classification may be drift in the training process. Object detection using classifiers grids overcomes above disadvantages while still assuring effective classification. In contrast to sliding window technique, the main idea here each classification only undertakes an image area size limitations on the image grid. Hence the complexity of each classification reduced and each classification unit designed simpler parameter numbers and number combinations weak classifier. Each classification is only responsible for detecting objects in the specified region by classifying objects from the background. Using online classified system can respond to changing environmental conditions, reduce the complexity of the classification. In general, the approach adapts very effectively to the problem of lost or omitting required information. Such performance of the classifier affected when updating wrong information may pull back by learning to adapt to existing data and update it correctly. To avoid this problem, classifier grids [20] applied to update the fixed strategy using a few positive samples set and negative samples are automatically generated during the training. Each region will have a set image classification and responsible for undertaking

quantified image area to determine whether or not the object appears on the image. This update strategy ensures the stability of the findings in the long run for not being drifted while giving good results. In addition, this approach needs no time to prepare data in advance. In fact, the sampling error during update online learning classifier leads to being degraded that can be restored with a certain number of sample in a short time. An object classifier grids in the same place for a long period of time can be seen as negative examples when major updates lead to the loss of the object detected. In this research, we solve the problem of omitting objects detected in the sequence of continuous image frames by combining current information with the temporarily information of previous images frame and replace most fixed update strategy with a combination adapted between the classification given training to set carefully-selected data as a prior knowledge. The experimental results show that the benefits of the proposed approach. Solving the problem of detecting the object does not move in for a while with better results in terms of both accuracy and performance of the classifier grids.

## 3. Combination of off-line and on-line learning for classifier grids

In the section we describe how to combine off-line boosting for feature selection with on-line boosting for feature selection to allow for combining prior information from labeled data with new information, which is not available at off-line training time. The main purpose of the learning based on the classification grid is to simplify problem detection object, reducing complexity in detecting objects and solving drifting objects. Moreover, classifier grids learning is to incorporate unlabeled information, which classifier adapts to new environment. In this section, we will present the key issues as well as the main idea of our approach to classifier grids learning.

### 3.1. Classifier grids

The main idea of classifier grids follows the division and conquer principle. The image is divided into highly overlapping grid elements (regions), where each grid element corresponds to one simple classifier [9]. Using a separate classifier for each position within the image significantly simplifies the problem. Classifier grids is a set of simple classification. The classification task is handled by the classification of detected objects which is discriminated to the background image. Classifier grids is to reduce complexity of task by training a separate classifier for each position in image and using online learning increase adaptive to classifier. The update of classifier is used fixed rules. We use online boosting learning for feature selection [15] on estimated classifier grids. Using pool of positive samples for updating classifier and the negative samples are directly generated

from image patches corresponding to a classifier grid. The classifier within classifier grids considers a simple problem to discriminate between object of interest and background at one specific location within current image. This is illustrated in Figure 4. In contrast, sliding window technique with which classifier has to evaluate confidence value to different window scales at every position in the image. Classifier in grid has responsibility at one specific region within the image. Hence, this reduces the search space for each classifier in grids and can significantly reduce the number of classifiers.

### 3.2. Linking of off-line and on-line learning

We choose on-line boosting for feature selection as learner for the classifier grids approach. There are many problems in computer vision, where not all training data is available before training. All these problems require on-line learning algorithms. An on-line algorithm showing good results for object detection and object tracking is on-line boosting for feature selection [15]. The feature selection process allows for handling changing situations efficiently by switching between different features and choosing the features that are most suitable for the actual task. In general, on-line classifiers like on-line boosting for features selection are initialized randomly from the set of all possible weak classifiers. However, if the problem is known in advance then it is possible to use a suitable representation describing the actual problem, i.e., features that are suitable for the particular task. Using this prior information, which is often available, can improve the results. Originally, on-line boosting for feature selection initializes the selectors with random features. In order to exploit the often prior available information, we propose to link off-line and on-line boosting for feature selection. Off-line boosting for feature selection allows initializing the classifier with features suitable for the actual task. Therefore, off-line boosting for feature selection needs to be modified. Originally, in each iteration, off-line boosting for feature selection selects one weak classifier, i.e., one particular feature. To allow a subsequent on-line boosting for feature selection, a set of  $J$  weak classifiers has to be selected in each boosting iteration, where  $J$  is the number of weak classifiers within one selector. This admits incorporating prior information about the object class. The huge pool of randomly initialized weak classifiers may contain very similar features. In general, similar features give a similar training error, which is the criterion for selecting the weak classifiers. This does not influence off-line boosting for feature selection, since only the best weak classifier is selected in each iteration. However, to allow a subsequent on-line boosting for feature selection, instead of a single best weak classifier, we select the best  $J$  weak classifiers in each boosting iteration. To avoid having too similar features within one

selector, an additional selection criterion measuring the similarity between features has to be introduced. Too similar features within one selector would hinder the adaptivity of the classifier during on-line updates. Hence, we have to introduce a similarity criterion based on the overlap between the features and the feature types. The overlap criterion considering the spatial position and extend of features within the patch. Features with an overlap larger than a specified threshold are only allowed if they have distinctive feature types, i.e., horizontal vs. vertical or diagonal feature types. The avoidance of too similar features within a selector ensures the adaptivity of the strong classifier, which is required for an on-line adaption. Applying off-line boosting for feature selection to select features appropriate to a specific task allow using less complex classifiers to solve the same problem, since the features within a classifier are well suited for the particular problem. The modified off-line boosting for feature selection algorithm is described in Algorithm 1.

**Algorithm 1:** Off-line Boosting learning for feature selection algorithm.

**Input:** Labeled training data:

$$S = \{(x_1, y_1), \dots, (x_L, y_L)\} | y_i \in \{-1, +1\};$$

**Output:** The final strong classifier :

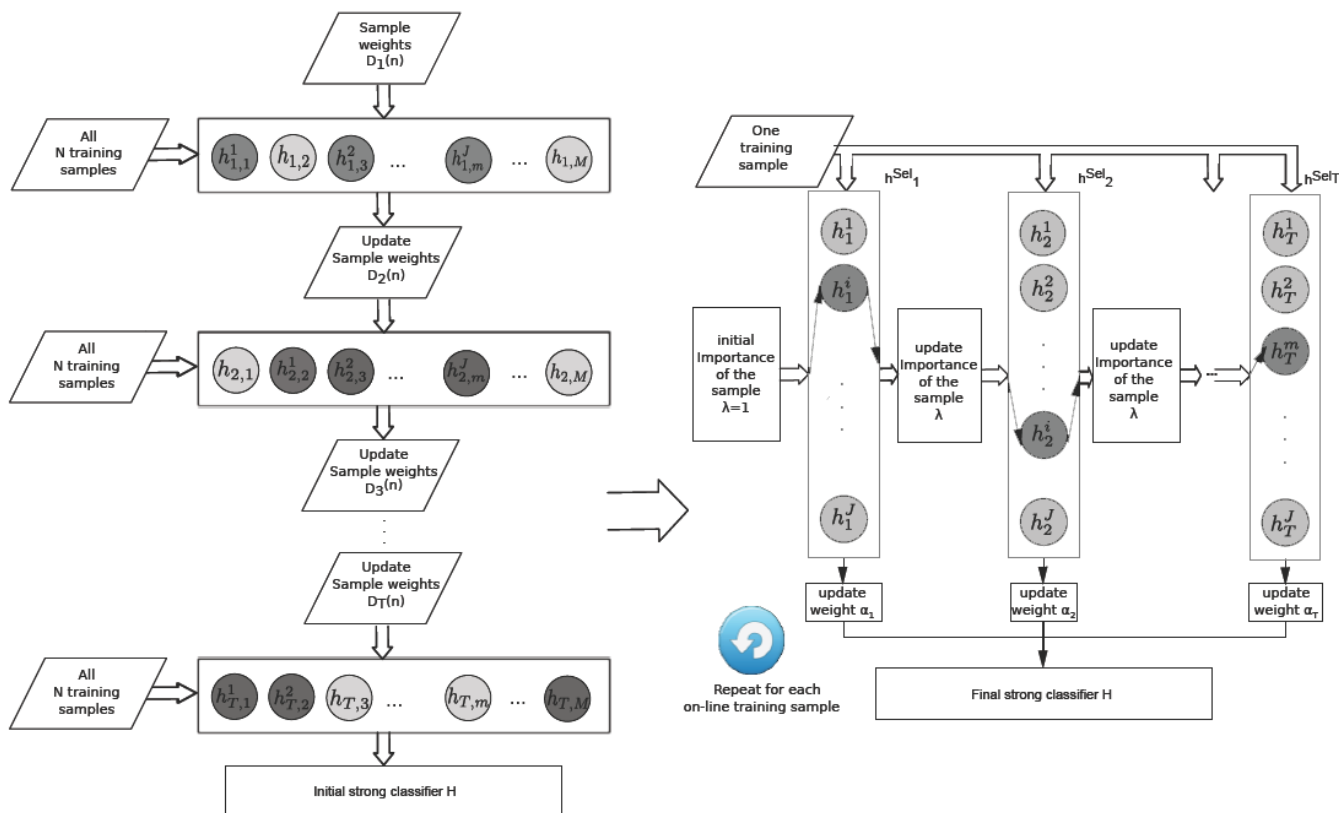
$$H(x) = \text{sign}(\sum_{i=1}^T \alpha_i h_i(x));$$

**Method:**

1. Initialize weights  $D_1(i) = \frac{1}{L}$ ;
2. **for**  $t = 1, 2, \dots, T$  **do**
3. For each feature  $j$  train one weak classifier  $h_j: X \rightarrow Y$  with error with respect to  $D_t$
4.  $e_j^{\text{off-line}} = \sum_{i=1}^L D_t(n) \cdot I(h_j(x_i) \neq y_i)$
5. Select best  $J$  weak classifiers to initialize selector  $\mathbf{t}$  with proper features
6. **end for**
6. Chose  $\alpha_t = \ln \frac{1 - e_t^{\text{off-line}}}{e_t^{\text{off-line}}}$
7. Update weight distribution:
 
$$D_{t+1}(i) = \frac{D_t(i)}{Z_t} \begin{cases} \exp(-\alpha_t) & h(x_i) = y_i \\ \exp(\alpha_t) & h(x_i) \neq y_i \end{cases}$$
8. **end for**

The fixed update strategies are based on a fixed set of hand-labeled samples used for the positive updates to describe the appearance of the object of interest and on samples extracted on the fly from the scene to model the back ground.





**Figure 2.** This is a linking off-line and on-line learning: the off-line boosting for feature selection algorithm has to be adapted to select a set of J weak classifiers within each boosting iteration.

The positive updates are correct by definition, since they are taken from a finite set  $X^+$  of hand-labelled positive samples. Thus, for each feature  $f_j \in F$ , where  $F$  is the full feature space, a generative model  $D_j^+$  can be estimated. By drawing from a fixed set of hand-labeled samples  $X^+$  for the positive updates, the positive distributions  $D_j^+$  is not changing over time and can be calculated in an off-line manner, since all information is given in advance. This allows for neglecting these updates during the on-line scene adaptation and results in a fixed distribution for the object class (positive class)  $D_j^+$ . If this step is performed by our modified off-line boosting for feature selection algorithm

(Algorithm 1), we can exploit the advantage of having a classifier which consists only of features suitable for the task of interest, i.e., the features selected during this off-line training stage are well suited to describe the object class. Additionally, the number of updates can be reduced to the half, since positive updates are no longer required because the distribution of the positive information is calculated during off-line training. In order to adapt to changing environmental conditions, the negative

distributions  $D_j^-$  have to be updated all the time. Thus, the input images are directly used to perform updates of corresponding grid elements. We assume that these updates are correct most of the time. Finally, in the particular case of classifier grids the discriminative classifier can be estimated by combining the two generative models  $D_j^+$  (can be calculated off-line) and  $D_j^-$  (has to be calculated on-line) at feature level. This combination can be efficiently realized by using on-line boosting for feature selection. The overall idea is illustrated in Figure 1. Since during the on-line stage of our classifier grids approach only negative updates are performed, the error of the positive samples  $\epsilon_+^{off-line}$  stemming from the off-line training is kept fixed during the on-line stage while the error of the negative samples is adapted all the time. By using solely the error calculated during on-line learning, only the error for negative samples, i.e., the false positive rate can be estimated. However, the fixed distributions of the object class  $D_j^+$  were estimated off-line. Thus, we can use the combined error  $\epsilon = \frac{1}{2}(\epsilon_+^{off-line} + \epsilon_-^{on-line})$  to select the best weak

classifier within the selector. Yet, this linkage of off-line and on-line learning is ideally suited for the classifier grid approach, since prior information can be exploited in the off-line stage while information available at run-time can be considered for on-line adaptation. The linkage of off-line and on-line learning for classifier grids allows for solely performing negative updates for a classifier, whereas the positive representation was trained off-line in advance and kept fix. These update strategies ensure “long-term” stability.

**Algorithm 2:** Linking offline and on-line Boosting learning for feature selection.

**Input:** Labelled training samples  $\langle x, y \rangle, y \in \{-1, +1\}$ ;

Initialize all weights to  $\lambda_{t,j}^{corr} = \lambda_{t,j}^{wrong} = 1$ ;

Sample importance weight  $\lambda = 1$ ;

**Output:** The final strong classifier

$$H(x) = \text{sign}\left(\sum_{t=1}^T \alpha_t h_t(x)\right);$$

**Method:**

1. **for**  $t = 1, 2, \dots, T$  **do**
2.   **for**  $j = 1, 2, \dots, J$  **do**
3.      $h_{t,j} = \text{update}(h_{t,j}, \langle x, y \rangle, \lambda)$
4.     **if**  $h_{t,j}(x) = y$  **then**  $\lambda_{t,j}^{correct} = \lambda_{t,j}^{correct} + \lambda$
5.     **else**  $\lambda_{t,j}^{wrong} = \lambda_{t,j}^{wrong} + \lambda$
6.     **end if**
7.      $e_{t,j}^{on-line} = \frac{\lambda_{t,j}^{wrong}}{\lambda_{t,j}^{correct} + \lambda_{t,j}^{wrong}}; e_{t,j} = \frac{1}{2}(e_{t,j}^{off-line} + e_{t,j}^{on-line})$
8.   **end for**
9.    $j^* = \arg \min_j (e_{t,j}); e_t = e_{t,j^*}; h_t = h_{t,j^*};$
10.    $\alpha_t = \frac{1}{2} \cdot \ln\left(\frac{1 - e_t}{e_t}\right)$
11. Update sample weights
 
$$\lambda = \lambda \times \begin{cases} \frac{1}{2 \cdot (1 - e_t)} & h_t(x_i) = y_i \\ \frac{1}{2 \cdot e_t} & h_t(x_i) \neq y_i \end{cases}$$
12. **end for**

### 3.2. Combination of off-line and on-line learning to classifier grids

**Initialization stage:** Given an input image  $X$  is divided into highly overlapping grid elements, regions  $X_i$  ( $i = 1, \dots, N$ ), where each grid element  $i$  corresponds to one compact classifier  $G_i$  operating in image patch  $X_i$ . We have classifier grid  $G_i$  operating in regions  $X_i$  and one classifier  $C$  operating in sliding window on background subtracted image  $B$ . In our system, classifier grids  $G$ , note  $G_i$  is a classifier  $i$  in grid  $G$ , are trained in a co-training strategy manner. Beginning, classifier grids  $G_i$  and classifier  $C$  are initialized with the same off-line trained classifier. The off-line prior trained classifier has to capture generic knowledge by a few number samples to

update classifiers within a new scene. The classifier grids  $G_i$  is updated using labels generated by a second independent co-trained classifier  $C$  evaluated on the background subtracted image. In otherwise, classifier grids  $G_i$  and classifier  $C$  co-train each other. A confident classification of classifier grids  $G_i$  is used to update classifier  $C$  at position  $i$ . Otherwise, a confident classification of classifier  $C$  at position  $i$  generated an update for classifier grids  $G_i$ . Classifier grids model is a combination of two generative models, one describing the background and one describing the object of interest, which are combined to a discriminative model at feature level by combining off-line and on-line boosting as illustrated in Figure 3. The relations of classifier grids  $G$  and classifier  $C$  on subtracted image updating are summarized in Algorithm 3.

**Algorithm 3:** Classifier Grids Co-training.

**Input:** Classifier grids  $G_i^{t-1}$ ; Classifier  $C^{t-1}$ ;

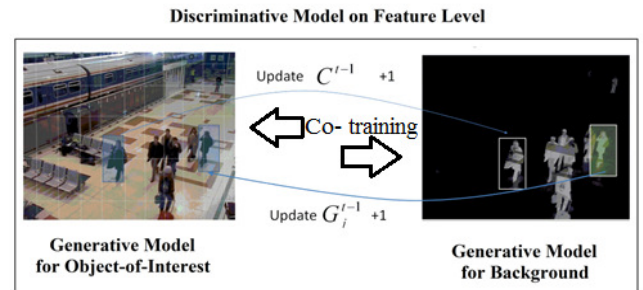
Patch corresponding to grid-element  $X_i$ ;

Background subtracted patch  $B_i$ ;

**Output:** Classifier grids  $G_i^t$  and classifier  $C^t$

**Method:**

1. **if**  $C^{t-1}(B_i) > \theta$  **then**  $\text{update}(G_i^{t-1}, X_i, +1)$ ;
2. **else if**  $C^{t-1}(B_i) < -\theta$  **then**  $\text{update}(G_i^{t-1}, X_i, -1)$ ;
3. **if**  $G_i^{t-1}(X_i) > \theta$  **then**  $\text{update}(C^{t-1}, B_i, +1)$ ;
4. **else if**  $G_i^{t-1}(X_i) < -\theta$  **then**  $\text{update}(C^{t-1}, B_i, -1)$ ;
5. **End**



**Figure 3.** Classifier grids co-training: classifier grids on the left side are co-trained with an independent classifier operating on the background subtracted image on the right side.

**Detection stage:** First, the classifier  $C$  is used as an oracle to generate new positive sample as well as negative sample updates of the classifiers within the classifier grids on the background subtracted image. Positive samples updates are spread to all classifiers in the grids whereas negative sample updates are performed for a particular classifiers in grids. In combination with our robust online learning, this oracle can be replaced by the fixed updates rules. In addition, learning framework performs negative sample updates to classifier grids  $G_i$  only if the scene is changing. A confident positive result of classifier  $C$  at position  $i$  in grid generates an update for all classifier grids  $G_i$  ( $i=1..N$ ). In this way, negative sample updates are

performed with the classifier grids  $G_i$  if there is not corresponding detection result at this position for classifier  $C$ . For a new scene specific, positive samples update over the whole classifier grids  $G_i$ . The update algorithm during detection phase for a specific grid element  $i$  is presented in Algorithm 4 and illustrated in Figure 4.

**Algorithm 4:** Classifier Grids Update

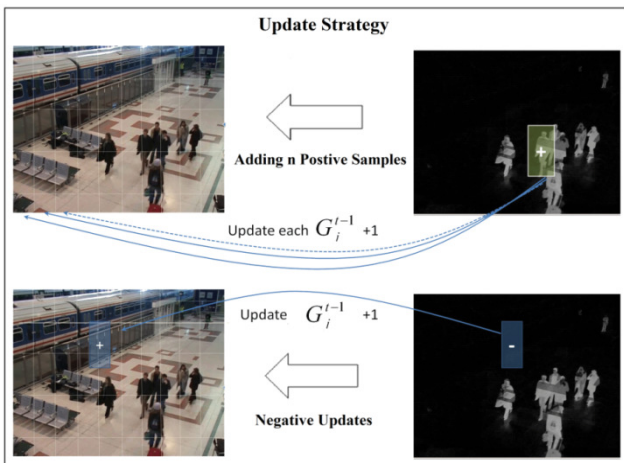
**Input:**

- Classifier grids  $G_i^{t-1}$ ; Classifier  $C^{t-1}$ ;
- Patch corresponding to grid-element  $X_i$ ;
- Background subtracted patch  $B_i$ ;

**Output:** Classifier grids  $G_i^t$

**Method:**

1. **if**  $C(B_i) > \theta$  **then**
2. **For**  $i=1$  **to**  $N$  **do**
3.      $\text{update}(G_i^{t-1}, X_i, +1)$ ;
4. **if**  $C(B_i) < -\theta$  **then**      $\text{update}(G_i^{t-1}, X_i, -1)$ ;
5. **End.**



**Figure 4.** The classifier  $C$  is used as an oracle to perform positive as well as negative updates on the classifiers within the classifier grids. Positive updates are spread to all classifiers in the grids whereas negative updates are performed for a particular classifiers grid.

## 4. Experimental evaluation and results

We evaluate the classifier grids approach on different scenarios to demonstrate the benefits compared to generic state-of-the-art approaches. The aim of our experiments is to demonstrate the efficiency of linked off-line and on-line boosting for classifier grids and the robustness of our framework for object detection problem in surveillance system. In the following, we first give a description of the used data sets, then present the training process, and finally report the performance of our system. To test the proposed object detection system, we used two

challenging data sets. We performed experiments and demonstrated our approach on public different and available datasets for person detection. The first is the public PETS 2006 dataset showing pedestrians at a train station. The second is the CAVIAR dataset. Besides evaluations on two different pedestrian datasets and one car dataset we perform a long-term experiment to demonstrate the stability over time. For the experiment of detecting objects, we start with a random classifier which comprises of 20 weak classifiers and 50 selectors. We use an off-line trained boosted classifier with a size of  $50 \times 20$  weak classifiers which means that we have 50 selectors, each of it containing 20 weak classifiers for the feature selection process. To increase the solidity of the negative samples updates, we collected four background images overlapped activities for different time periods. From these experiments, the benefits of the proposed approach are obvious. We use an overlap between the grid elements of 90%. For the classifier grids approach we first perform updates with each frame within the first 30 frames. After we adapted our classifiers in the initial stage (first 30 frames) we reduce the number of updates by using only every 10-th frame to improve the run-time performance.

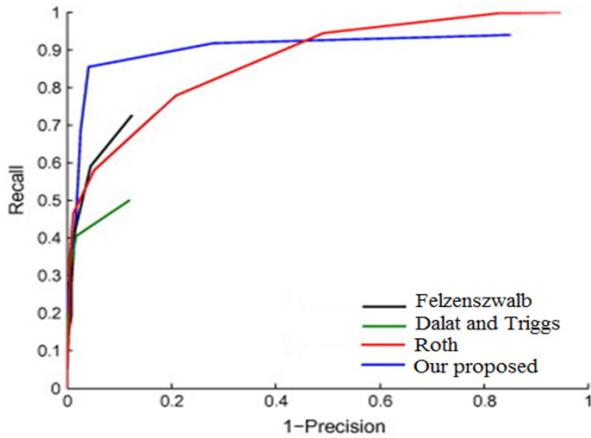
### 4.1. PETS Dataset

In this experiment, we compare the different proposed approaches on a publicly available dataset for pedestrian detection, namely the PETS 2006 dataset. It contains left-luggage scenarios at a train station. We evaluate on Dataset, a scenario where one person leaves a ski equipment. The sequence consists of 308 frames with a resolution of  $720 \times 576$  pixels containing 1265 pedestrians. We compare the classifier grids with other advanced methods, namely the object model deformation Felzenszwalb et al. [7] and the approach chart of Dalal Gradients and Triggs [5]. Both methods use fixed classifier trained offline and are based on the sliding window technique. In addition, we compare our approaches at classifier grids and compare with the approach of Roth et al. [20]. The results of the PET dataset is shown in Figure 5 and the recall and precision for the best F-measure value is shown in Table 1. Illustrated object detection is shown in Figure 6. For all classifier grids approaches we set the overlap between the grid elements to 87%.

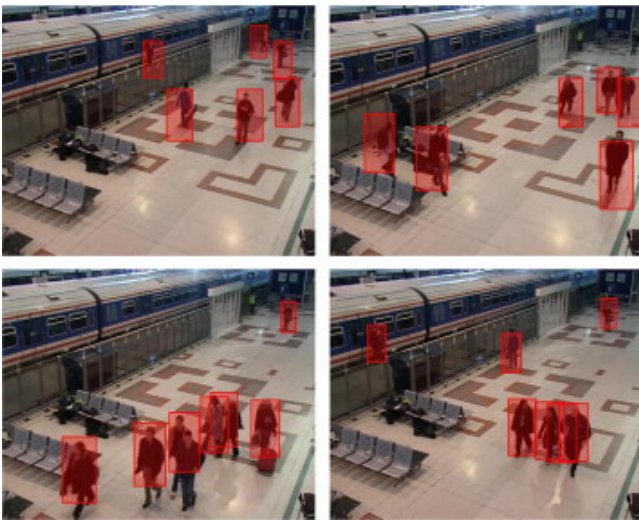
Table 1. The Recall and Precision Comparison

Methods	Comparision		
	Recall	Precision	F-Measure
Felzenszwalb	0.74	0.89	0.79
Dalal and Triggs	0.51	0.88	0.66
Roth	0.80	0.81	0.80
Our proposed	0.82	0.93	0.87





**Figure 5.** Recall-precision curves (RPC) of PET sequence.



**Figure 6.** Some illustrations person detection results of the classifier grids for the PETdataset.

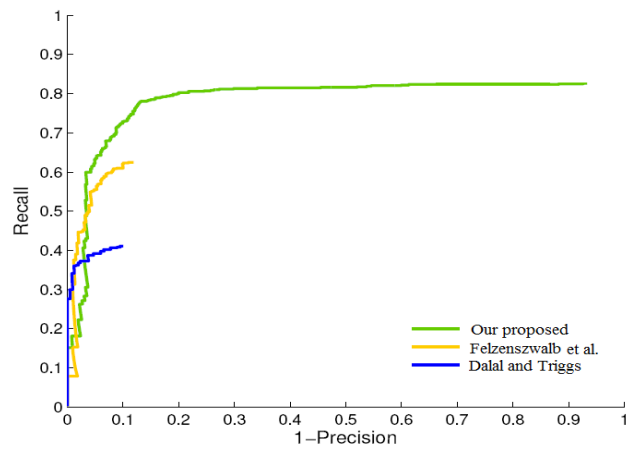
#### 4.2. CAVIAR Dataset

We evaluate the off-line and on-line linking approach on Caviar dataset. The Caviar dataset was provided by the CAVIAR (Context Aware Vision using Image-based Active Recognition) project. This dataset contains a large number of clips. We selected clip ShopAssistant2cor since it contains a large number of pedestrians (i.e., 1265). The sequence consists of 370 frames with an image size of 384 x 288 pixels and shows a Corridor of a public shopping mall. Two persons are entering a shop, crossing the corridor while a number of persons are walking along this corridor or standing at the end of the corridor. To conduct experiments with the approach based on the classification of cells in the data sets Caviar, the following parameters are initialized: the image size: 32 x 64. The selectors used to train online for classifiers is: 10. Number of weak classifier of a selector is 20. The results of the

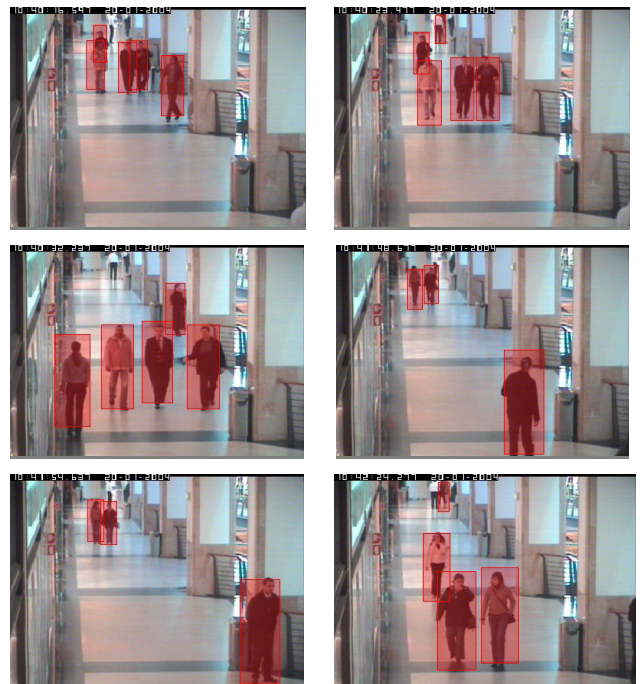
Caviar sequence are shown in Figure 7 and Table 2. Again it can be seen that the classifier grids for object detection (our proposed) better than the generic object detection (Dalat and Triggs and Felzenszwalb et al.), especially Recall. Some illustration person detection results are shown in Figure 8.

**Table 2.** The Recall and Precision Comparison

Methods	Comparison		
	Recall	Precision	F-Measure
Felzenszwalb	0.62	0.90	0.74
Dalat and Triggs	0.62	0.90	0.74
Our proposed	0.78	0.87	0.82



**Figure 7.** Recall-precision curves (RPC) of Caviar dataset



**Figure 8.** Some illustration person detection results of the classifier grids for the Caviar dataset.



### 4.3 Long-term sequences

In the following we demonstrate the long-term sequence of the proposed method. The main goal of classifier grids is to develop an adaptive but still robust system that is learning 24 hours a day and 7 days a week. We captured a sequence of a corridor in our building with 1 fps during 7 days, resulting in 580.000 frames. To show that the performance is unchanged over time, we annotated 2, 500 frames at four different points in time (which corresponds to approx. 45 minutes of video data). The sequences are selected at first day, the third day, the sixth day and the last day. Every single frame was used to perform an update of the classifiers within the classifier grids. The number of pedestrians visible as well as the number of updates performed before the sequence starts are summarized in Table 3 for all four sequences.

Table 3. Sequences of the long-term experiment

	# updates yet performed	# pesons
1 <sup>st</sup> day	3390	475
3 <sup>rd</sup> day	179412	546
6 <sup>th</sup> day	484891	950
7 <sup>th</sup> day	577500	1005

It can be seen that the method is stable over time. The results shown in Figure 9 and Table 4.

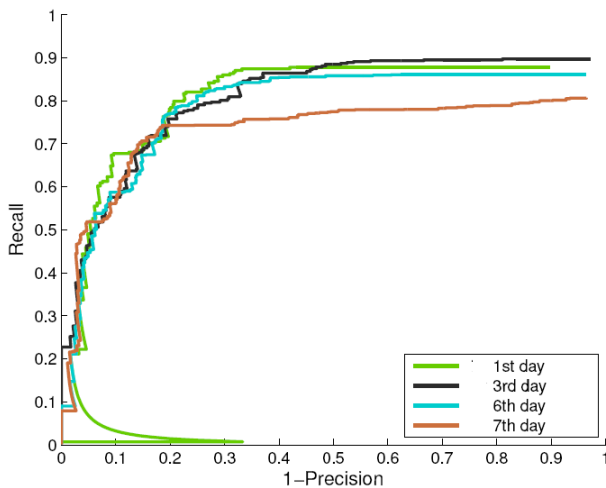


Figure 9. RPC for the long-term data experiment

Table 4. The Recall and Precision Comparison

Day	Comparision		
	Recall	Precision	F-Measure
1 <sup>st</sup> day	0.80	0.79	0.80
3 <sup>rd</sup> day	0.76	0.80	0.78
6 <sup>th</sup> day	0.78	0.79	0.79
7 <sup>th</sup> day	0.74	0.81	0.78

The slight variations in the curves can be explained by the different levels of complexity for the four sequences (i.e., number of persons, density of persons, etc.) and the F-measure is unchanged over time. Finally, we illustrate the significant changing conditions (i.e., natural light, artificial lighting, inadequate lighting, etc.) in Figure 10. We had to deal with during these 7 days. These drastical changing conditions arise the need for adaptive approaches, like the classifier grids approach.

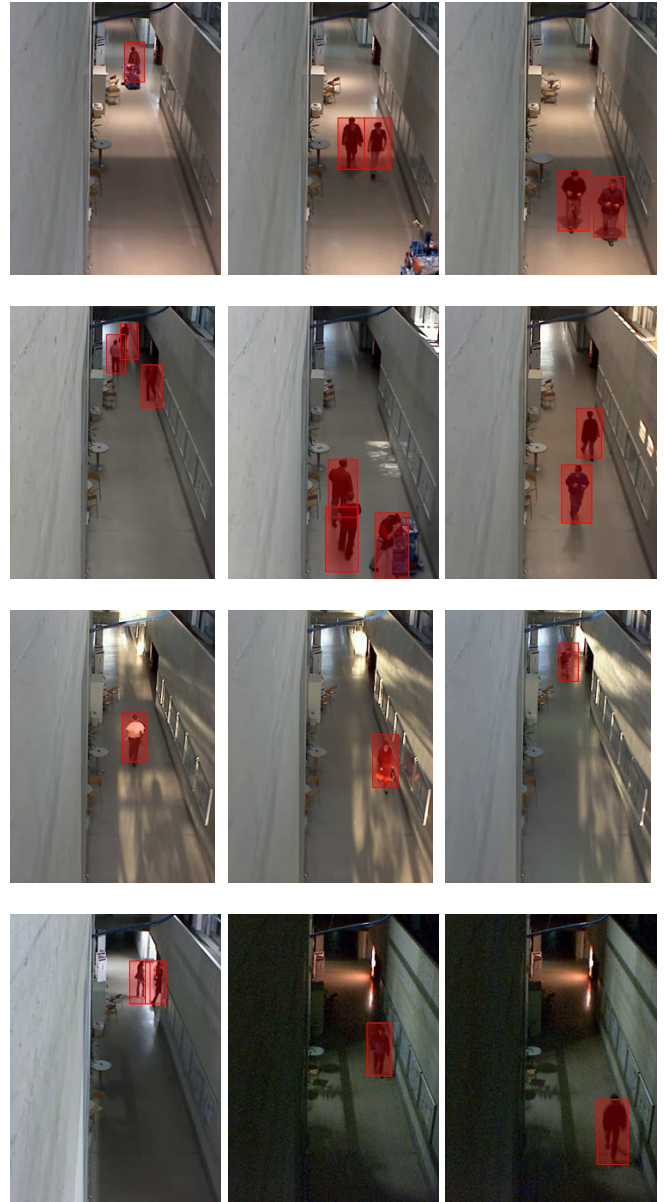


Figure 10. Illustrative detection results of the classifier grids obtained during long-term experiment. Each row corresponds to one time of day, morning, noon, afternoon, and night respectively.

### 4.4 Car detection

We evaluate the off-line on-line linkage approach on a sequence showing one direction of a public highway. This dataset as “highway dataset”. The whole scene consists of 1000 frames and contains a total number of 1952 cars from the rear view. The overlap between the grid elements was set to 92%. We compare our method to existing established methods, namely the Implicit Shape Models (ISM) of Leibe et al. [23] and the deformable part of model detector (DPM-FS). Illustrative detection results obtained by the proposed approach are shown in Figure 12 The detection characteristics are summarized in Table 5.

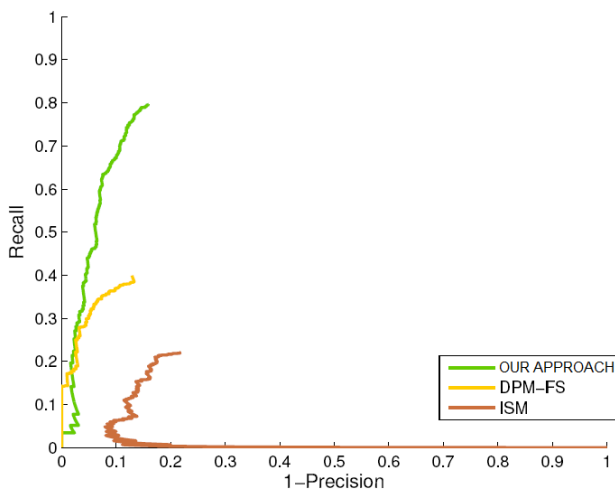


Figure 11. RPC for the Highway sequence.

Detection characteristics of the Highway Sequence for different methods sorted by the F-measure.

Table 5. The Recall and Precision Comparison

Methods	Comparison		
	Recall	Precision	F-Measure
Our approach	0.79	0.87	0.82
DPM-FS	0.40	0.87	0.55
ISM	0.22	0.78	0.35

From the results shown in Figure 12 it can be seen that the proposed method clearly outperforms the generic car detectors (ISM and DPM-FS).

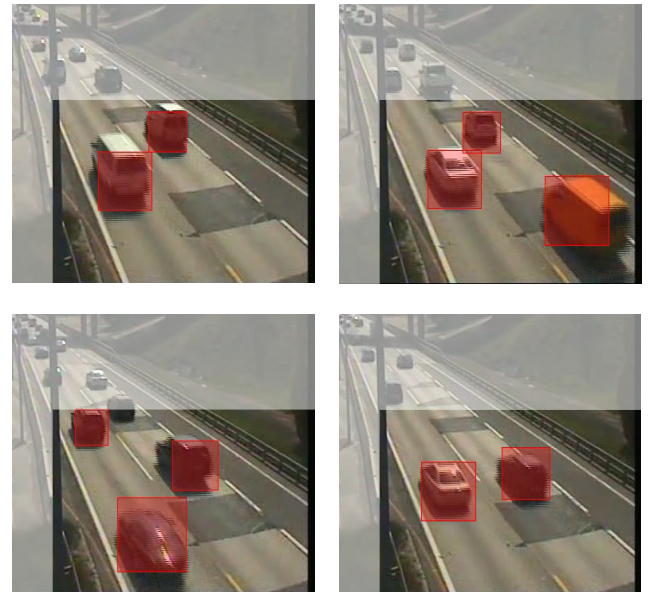


Figure 12. illustrative detection results of the generic car detector for the Highway sequence.

### 5. Conclusion

We have shown the success of object detection in surveillance systems based on idea of linking off-line and on-line learning for classifier grids. We have demonstrated the efficiency of classifier grids detection using the proposed framework. Also, the robustness of the detector against vast changes of pedestrians is clearly mentioned. The advance of our method is capability in dealing with unlabeled information incorporation for object detection from stationary cameras which is to preserve robustness classifier grids. We use semi-supervised learning with co-training approach for the classifier grids. The idea of classifier grids could be used to combine both scene specific information and background information. This combination allows to incorporate unlabelled large information-streams but still in preservation of the reliable labelled information. Empirical experiments demonstrated the performance of our system for object detection on several challenging data sets on real-world surveillance scenario, where a corridor in public building is monitored over 24 hours per day and 7 days per week and object detection performed. Our approach is updated without supervision, the classifier grids is to preserve long-term stability and robustness. By using update strategies and combination of off-line and on-line learning algorithm, these classifier grids are modified over time to ensure scene adaptation of our object detection framework. Even though the updates are completely unsupervised, empirical results on long-term evaluations demonstrated that this approach is robust over time.

## References

- [1] Agarwal, S., Awan, A., Roth, D. (2004) Learning to detect objects in images via a sparse, part-based representation, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 26 (11), pp. 1475–1490.
- [2] Andrews, S., Tsochantaris, I., Hofmann, T. (2003) Support vector machines for multiple-instance learning, *In: Advances in NIPS*. pp. 561–568.
- [3] Babenko, B., Yang, M.-H., Belongie, S. (2009) Visual tracking with online multiple instance learning, *In IEEE Conference on CVPR*.
- [4] Blum, Mitchell, T. (1998) Combining labeled and unlabeled data with co-training”. *In: Proc. Conf. on Computational Learning Theory*. pp. 92–100.
- [5] Dalal, N., Triggs. (2005) Histograms of oriented gradients for human detection. *In IEEE Conference on CVPR*, Vol. I. pp. 886–893.
- [6] Viola, P., Platt, J. C., Zhang, C. (2005) Multiple instance boosting for object detection. *In: Advances in Neural Information Processing Systems*. pp. 1417–1426.
- [7] Felzenszwalb, P., McAllester, D., Ramanan, D. (2008) A discriminatively trained, multiscale, deformable part model”. *In IEEE Conference on CVPR*.
- [8] Goldberg, B., Li, M., Zhu, X. (2008) Online manifold regularization: A new learning setting and empirical study”. *In Proceeding on European Conf. on Machine Learning and Knowledge Discovery in Databases*, Vol. I. pp. 393–407.
- [9] Grabner, H., Roth, P. M., Bischof, H. (2007) Is pedestrian detection really a hard task?”. *In Proceeding IEEE Workshop on Performance Evaluation of Tracking and Surveillance*.
- [10] Hoiem, D., Efros, A.A., Hebert, M. (2006) Putting objects in perspective. *In: Proceeding IEEE Conf. on Computer Vision and Pattern Recognition*, Vol. II. pp. 2137–2144
- [11] Javed, O., Ali, S., Shah, M. (2005) Online detection and classification of moving objects using progressively improving detectors”. *In Proceeding IEEE Conference on CVPR*, Vol. I. pp. 696–701.
- [12] Leibe, B., Leonardis, A., Schiele, B. (2008) Robust object detection with interleaved categorization and segmentation, *International Journal of Computer Vision*, 77 (1–3), pp. 259–289.
- [13] Levin, P., Viola, P., Freund, Y. (2003) Unsupervised improvement of visual detectors using co-training”, *In: Proc. ICCV*, Vol. I. pp. 626–633.
- [14] Li, L.J., Wang, G., Fei-Fei, L. (2007) Optimol: automatic online picture collection via incremental model learning”. *In Proceeding IEEE Conference on CVPR*, pp. 1–8.
- [15] Grabner, H., and Bischof, H. “On-line boosting and vision”. *In Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, vol. I, pp. 260–267 (2006)
- [16] Nair, V., Clark, J. J. (2004) An unsupervised, online learning framework for moving object detection”, *In: Proc. IEEE Conference on CVPR*, Vol. II. pp. 317–324
- [17] Rosenberg, C., Hebert, M., Schneiderman, H. (2005) Semi-supervised self-training of object detection models”, *In IEEE Workshop on Applications of Computer Vision*, pp. 29–36
- [18] Wu, B., Nevatia, R. (2007) Improving part based object detection by unsupervised online boosting. *In: Proc. IEEE Conference on CVPR*, pp. 1–8,
- [19] Roth, P.M., Grabner, H., Skočaj, D., Bischof, H., Leonardis, A. (2005) On-line conservative learning for person detection. *In Proc. IEEE Int’l Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance*, pp. 223–230.
- [20] Roth, P. M., Sternig, S., Grabner, H., Bischof, H. (2009) Classifier grids for robust adaptive object detection. *In Proceeding IEEE Conference on CVPR*.
- [21] Torralba, A. (2003) Contextual priming for object detection. *IJCV*, 53(2):169–191.
- [22] Heitz, G., and Koller, D. (2008) Learning spatial context: Using stuff to find things. *In Proceeding ECCV*.
- [23] Bastian L., Alesš L., and Bernt S. (2008) Robust object detection with interleaved categorization and segmentation. *Int’l Journal of Computer Vision*, 77(1-3): 259–289.