Context-Based Classifier Grids Learning for Object Detection in Surveillance Systems

Dang Binh Nguyen(✉)

Hue University of Sciences,
77 Nguyen Hue Street, Hue City, Vietnam
ndbinh@hueuni.edu.vn

Abstract. We propose a new method to adaptive object detector is to incorporate the scene specific information without human intervention to reach the goal of fully autonomous surveillance where the focus is on developing adaptive approaches for object detection from single and multiple stationary cameras that are able to incorporate unlabeled information using different types of context in order to collect scene specific samples from both, the background and the object class over time. The main contributions of this paper tackle the question of how to incorporate prior knowledge or scene specific information in an unsupervised manner. Thus, the goal of this work is to increase the recall of scene-specific classifiers while preserving their accuracy and speed. In particular, we introduce a co-training strategy for classifier grids using a robust on-line learner. The system runs at 24 h per day and 7 days per week with 24 frames per second on consumer hardware. Our evaluation show high accuracy on both synthetic and real test sets. We achieve state of the art in our comparisons with related work and in the experimental results these benefits are demonstrated on different publicly available surveillance benchmark data sets.

Keywords: Context-based learning · Classifier grids · Object detection · Online learning

1 Introduction

Robust learning interactive object detection has applications including surveillance intelligence systems, computer vision, gaming, human-computer interaction, security, and even health-care. One main challenge of incorporating unlabeled information is to preserve the long-term robustness of object detection, which is a major requirement for real-world applications. With the increasing number of surveillance cameras the need for autonomous visual surveillance systems is increasing tremendously. One of the first steps towards autonomous visual surveillance is object detection. The main focus of this research is on object detection from static cameras with specific emphasis on the applicability to real-world environments. To deal with changing environmental conditions which usually occur in real-world environments an adaptive object detector is required. To ensure robust object detection without the need for human intervention we develop different approaches which allow for robustly incorporating scene specific information. Context could help to limit appearance changes and thus scaling down the

P.C. Vinh and V. Alagar (Eds.): ICCASA 2015, LNICST 165, pp. 51–61, 2016.
DOI: 10.1007/978-3-319-29236-6_6

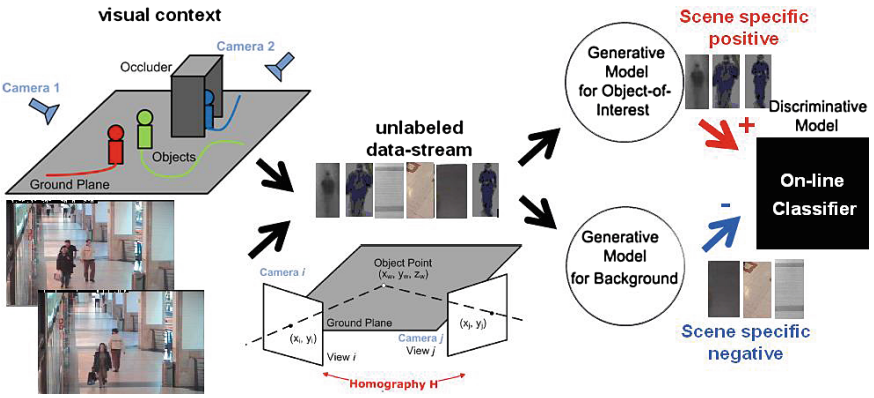


Fig. 1. Proposed approach of context-based classifier grids. 3D Context: A homography, maps a point on the ground plane from one view to another. The unlabeled large data-stream is analyzed and scene specific positive and negative samples are collected for continuously updating the classifiers, i.e. a local grid detector and using various types of context.

training set. It is well known that context plays a very important role, e.g., exactly the same image patch can be interpreted very differently depending on its embedding in the world [21].

A generic object detector tries Fig. 1. Having access to a large data-stream and using various types of context (e.g., scene knowledge) our approach continuously updates an specific object detector. to solve just the ill-posed problem of detecting the object of a class in any context [22]. Hence, generic detectors often fail in real world scenarios. In many application scenarios the detection problem would be far simpler. For example, in a 24/7 surveillance scenario the camera is often static and focuses always on one and the same scene. Further, there is a continuous data stream providing a huge amount of (unlabeled) data which should be explored for (i) improving detection results as well as (ii) speeding up the detection process. One simple way to benefit from the static camera is to incorporate information about the particular scene (e.g., using a ground plane to limit the size of persons). However, such information usually helps only to reduce the number of false alarms (e.g., [10]). In order to increase also the detection rate, on-line methods adapting to a particular scene have been investigated (e.g., [18]). These methods focus on solving the object detection task in the particular scene and take advantage from the continuous incoming data stream. In fact, these approaches use context (scene knowledge) already in the training process and not just as post-processing. Therefore, on-line unsupervised learning methods are usually used to continuously adapt the model. The main problem, however, is to robustly include the new unlabeled data. If the data is wrongly interpreted, the performance of the detector will be reduced. In other words, the detector might drift and would end in an unreliable state. The most prominent approach is to apply a sliding window technique [6–8, 16]. Each area of the image of a certain image is tested whether it is consistent with a previous estimate model or not, and finally all the images matching the notice results. Typically, the goal of this approach is to develop a general model in which can be applied to all possible scenarios, and the problem of detecting

various objects [7, 8, 12]. However if trained from a very large number of training samples for the detection of common objects ("broad application") often fail in specific situations. Because not all change, especially for coverage negative (e.g., all objects can be the background image), can be obtained results with performance and low accuracy. Assuming a fixed camera, which is a reasonable constraint for most applications, using information specific circumstances may help reduce the number of objects are detected [10]. To further improve the results of the classification of the classification on the specific object can be applied, designed to solve a specific task (for example, detecting objects for a set specific). In fact, the training of classification needs few training data is necessary and for a specific problem which we often better for accuracy and efficiency [13, 18, 19]. The method of detecting objects using traditional models and use the sliding window search object [1–5] then it is usually done coaching a single classification used to detect or identify the object on the whole picture. Therefore, the model of offline learning will encounter the following problems: Firstly, to establish a classification in the offline model of traditional learning (such as SVM, Boosting, neural networks,…) collective training samples must be prepared in advance, can call sample data is big problem depending on the application (several thousand samples). This makes labor expensive and time consuming sample preparation. Additionally due to the sample preparation prior to the application on the new scene to detect objects they can not promote efficiency, want effective it must retrain for new or updated models added adapt the template in this new context. Thus to access online learning. Secondly, after the training is completed the classification exam to detect objects, the classification must perform a search greed from above, from left to right, with all positions and different size search of objects not only on the current image frame in which the entire frame sequence of images. Thus, the complexity of the detection object will increase. Thirdly, usually to extract characterizing select training samples, the system only uses a specific method chosen only deduct certain to form. Therefore, can choose specific extraction method suitable for this kind of data, but may not be the best fit with other data types and in many different problems, but mostly for a particular problem dirty. Characteristics such as geometry, movement of objects, objects change shape, color, texture, and features can quickly calculate matching problem in real time. Therefore, the research to be able to use multiple methods to detect specific selected data in the same form and thus allows the selection of a specific type best characterized of the sort used for system is a matter of concern and this is the approach in this research. In this research, we used 03 extracted choose specific methods wavelet Haar, a local binary pattern, and chart directions simultaneously and choose Gradient method to suit each school model in which the estimated error the smallest model selected. Finally, a problem encountered in the detection methods and update the wrong object is detected errors and omissions objects here's the problem is many researchers focus on finding solutions to improve the ratio object detection system. The approach of this paper is also aimed and basic goals and overcome these drawbacks. Therefore, we develop different approaches that allow incorporating scene specific information for object detection and tracking in static camera setups. This allows adapting to specific scenes, which is beneficial in both single and multiple camera setups.

The rest of the paper is structured as follows. First, in Sect. 2, we mention issues related research recently. Next, we consider the idea of learning classifier grids in

Sect. 3. We give an empirical evaluation of the approach experimental evaluation and results in Sect. 4. Finally, we summarize and conclude the research in Sect. 5.

2 Related Works

To improve the strength of classification and continue to reduce the number of training samples of a classification adaptive use online learning algorithms can be applied [11, 16, 18]. Therefore, the system can adapt to the environmental changes (e.g. change of light conditions) and changes without the need to handle by the original model. In fact, in this way the complexity of the problem is reduced and the classification can be more effective training. The adaptive system has a drawback: the new data has not been labeled will be included in a model has been built. This approach typically self trained [14, 17], training and [4, 13], semi-supervised learning [8] or the app itself sample data generated during training [16]. The semi-supervised method, often used by combining the information given and explores new models from available data to form a set of classification. Self-training method or training Frequent synchronization constraints theory of constraints can not be guaranteed in practice or is based on the feedback of the current classification, classification results both unnecessary trust. The classification more effectively avoid the above problems can be trained to use the classification grid [9, 20]. In contrast with sliding window technique, in which a classifier can be quantified with different positions on the image, the main idea of the classification grid is coaching the separate classification for each different location of image. Thus, the complexity of the classification task was handled by a single classification so complexity is significantly reduced. Each classification is only able to distinguish the object to be detected from the background in a particular location in the image. Using the classification system online that can adapt to changing environmental conditions, further reducing the complexity of the classification. Adaptive approach, in general, enjoy problem affecting lost or missing information, for example, due to wrong system update start learning something completely different performance degradation of classification. To avoid this problem in the classification grid [20] have adopted strategies fixed update. Special sampler for updating the classification grid is generated from the corresponding area of the image, while positive samples are trained before and immobilized. This updated strategy to ensure stability in the long run, that is classified is not degraded. In fact, the classification has been chosen the wrong sample labeling update may restore a certain amount of time, this problem we call the short term lost. This could be the case if an object remains in the same place in a longer period of time determined in advance and background information used as samples of audio classes. In this research, we solve the problem of missing object detection in continuous data sequence by combining information temporarily and replace the updated strategy fixed by an adaptive combination between sets classification has been chosen in advance as an initial knowledge of the classification grid adapted, using the classifier is trained beforehand to verify the model before performing the update for each classification unit in net. The experimental results clearly show the benefits of the proposed approach. Especially considering approaches have no moving object can be significantly better handling, increased both in terms of both accuracy and performance of the classifier.

3 Classifier Grids Learning

The main challenge of adaptive object detectors is to incorporate scene specific unlabeled information, which allows for adapting the detector to new environmental conditions in a robust manner without human intervention. In the following, we review the ideas of classifier grids and learning, which build the base for the proposed approach.

(a) (b)

Fig. 2. Classifier Grids. (a) The main idea of classifier grids follows the divide and conquer principle. The image is divided into highly overlapping grid elements (regions), where each grid element. (b) The classifier grids on the left side are updated using labels generated by a second independent co-trained classifier evaluated on the background subtracted image.

3.1 Classifier Grids

The main concept of classifier grids [9] is to sample an input image by using a highly overlapping grid, where each grid element $\boldsymbol{i}$ ($i = 1,..., N$) corresponds to one classifier C_i. This is illustrated in Fig. 2. To reduce the number of classifiers within the classifier grid the ground-plane is pre-estimated. Thus, the classification task that has to be handled by one classifier C_i can be drastically reduced, i.e., discriminating the background of the specific grid element from the object-of-interest. To further reduce the classifiers' complexity and to increase the adaptively, on-line learning methods can be applied, where the updates are generated by fixed rules. For positively updating a grid classifier C_i a fixed pool of positive samples is used; the negative updates are generated directly from the image patches corresponding to a grid element. In general, for estimating the grid classifiers any on-line learning algorithm can be applied, however, on-line boosting has proven to be a considerable trade-off between speed and accuracy [15]. The goal of classifier grids is to further reduce the complexity of the task by training a separate classifier for each position within the image. Using a separate classifier for each position within the image significantly simplifies the problem. In this way, classifier grids follow the, in computer science well-established, divide and conquer paradigm, where the problem is broken down until the sub-problems become simple enough. Afterwards, the solutions to the sub-problems are combined to solve the original problem. Classifier grids divide each input image into a highly overlapping set of grid elements (regions), where each of the grid elements corresponds to one sub-problem of the whole object detection problem which is solved by a separate classifier. This is visualized in Fig. 3(a). The classifiers within the classifier grid can profit from simplifying the problem to discriminate between the object of interest and the background at one specific location within the image. The reduces variability at one specific location within the image

allows for using less complex and compact on-line classifiers, which can be evaluated and updated efficiently and further reduces the number of false alarms. Hence, in contrast to standard sliding window approaches which have to evaluate different scales at every position in the image, the use of scale information can significantly reduce the number of classifiers within the classifier grid. The number of classifiers within the classifier grid can be defined by an overlap parameter. There is always a trade-off between run-time and performance of the classifier grid object detector.

3.2 Learning for Classifier Grids

During the initial stage our system is trained in a co-training manner as shown in Fig. 3. Given n grid classifiers $\mathbf{G_j}$ operating on gray level image patches $\mathbf{X_j}$ and one compact classifier $\mathbf{C}$ operating in a sliding window manner on background subtracted images $\mathbf{B}$. To start co-training, the classifiers $\mathbf{G_j}$ as well as the classifier $\boldsymbol{C}$ are initialized with the same off-line trained classifier (see Algorithm 1). The classifiers within the classifier grid $\mathbf{G_j}$ and the classifier $\mathbf{C}$ operating on the background subtracted images co-train each other. A confident classification (no matter if positive or negative) of a classifier $\mathbf{G_j}$ is used to update the classifier C with the background subtracted representation at position $\mathbf{j}$. Vice versa, a confident classification of classifier C at position j generates an update for classifier $\mathbf{G_j}$. The off-line trained prior information already capturing the generic information causes a small number of updates to be sufficient to adapt the classifiers to a new scene. The update procedure during the initialization for a specific grid element $\mathbf{j}$ is summarized in Algorithm 1.

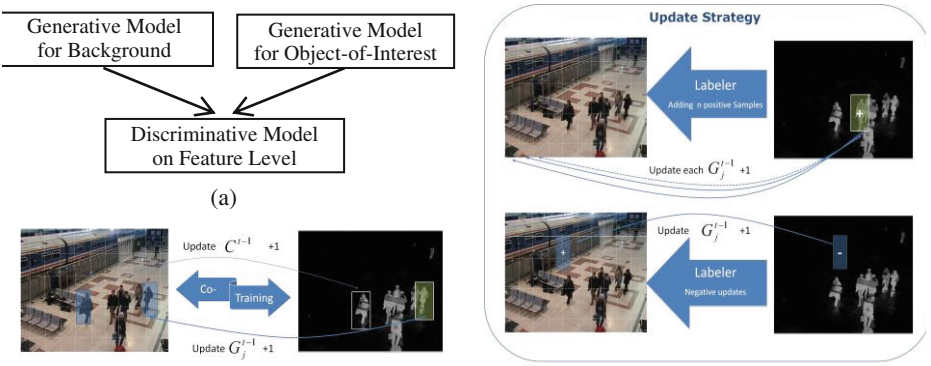


Fig. 3. (a) The grid-based classifiers can be interpreted as a combination of two generative models, one describing the back ground and one describing the object of interest, which are combined to a discriminative model at feature level by linking off-line and on-line boosting. (b) Co-grid initialization stage: the grid classifiers on the left side are co-trained with an independent classifier operating on the background subtracted image on the right side. (c) Co-grid detection stage: the classifier C is used as an oracle to perform positive as well as negative updates of the classifiers within the classifier grid. Positive updates are spread to all classifiers in the grid whereas negative updates are performed for a particular classifiers in grid.

Detection Stage: After the initial stage, as described above, the classifier ***C*** operating on the background subtracted images is no longer updated and is applied as an oracle to generate new positive and negative samples as illustrated in Fig. 3. In combination with our robust learning algorithm this oracle can now be to replace the fixed update rules. Moreover, we perform negative updates for the classifiers $\mathbf{G_j}$ only if they are necessary, i.e., if the scene is changing. Even if the oracle classifier **C** has a low recall, the precision is very high. Thus, only very valuable patches are used to update the classifier $\mathbf{G_j}$, which leads to an increasing performance of the classifiers within the classifier grid. In particular, a confident positive classification result of classifier **C** at position **j** generates an update for all classifier $\mathbf{G_j}$, j = 1,..., n in the classifier grid. In this way new scene specific positive samples are disseminated over the whole classifier grid. Negative updates are performed for classifiers $\mathbf{G_j}$ if there is no corresponding detection reported at this position for classifier C. The update procedure during the detection phase for a specific grid element **j** is summarized in Algorithm 2.

Algorithm 1: Co-Grid Initialization
Input:
grid-classifier G_j^{t-1}
co-trained classifier C^{t-1}
patch corresponding to grid-element X_j
background subtracted patch B_j
Output: grid-classifier G_j^t and classifier C^t
Method:
1. **if** $C^{t-1}(B_j) > \theta$ **then** update(G_j^{t-1}, X_j, +1)
2. **else if** $C^{t-1}(B_j) < -\theta$ **then** update(G_j^{t-1}, X_j, -1)
3. **end if**
4. **if** $G_j^{t-1}(X_j) > \theta$ **then** update(C^{t-1}, B_j, +1)
5. **else if** $G_j^{t-1}(X_j) < -\theta$ **then** update(C^{t-1}, B_j, -1)
6. **end if**

Algorithm 2: Co-Grid Update
Input:
grid-classifier G_j^{t-1}
co-trained classifier C
patch corresponding to grid-element X_j
background subtracted patch B_j
Output: grid-classifier G_j^t
Method:
1. **if** $C(B_j) > \theta$ **then**
2. $\forall_j$: update(G_j^{t-1}, X_j, +1)
3. **end if**
4. **if** $C(B_j) < -\theta$ **then**
5. update(G_j^{t-1}, X_j, -1)
6. **end if**

4 Experimental Evaluation and Results

In the following, we demonstrate our approach on different publicly available datasets for multi-camera person detection. We first describe our experimental setup and evaluation methods used and then evaluate our approach on two different datasets. To demonstrate the benefits of the proposed approach, we conducted two experiments (person object). We selected some of the data is publicly available for research to quantify the results to conduct experiments. From these experiments the benefits of the proposed approach is obvious. For the experiment of detecting pedestrians, we use the classification of 20 of selectors, each selector has 10 weak classifiers. For detecting person experiment we used the classification of 50 of selectors. each selector has 30 weak classifiers. When the classification we use simple decision tree than on the response characteristics Haar- like. To increase the solidity of the negative samples updates, we collected four background images overlap activities four different time periods.

4.1 PETS Dataset

In this experiment, we used a series of publicized PETS 2006 data includes 308 frames (720 × 576 pixels), including 1,714 pedestrians. We compare our approach with other advanced methods, namely the object model deformation Felzenszwalb et al. in 2008 (FS) [7] and the approach chart of Dalal Gradients and Triggs 2005 (DT) [5]. Both methods use fixed classifier was trained offline and is based on the sliding window technique. In addition, we compare our approached at classifier grids and compare with the approach of Roth et al. in 2009 (CG) [20]. The results of the Pet dataset is shown in Fig. 4 and Table 1. Illustrated object detection is shown in Fig. 6(a).

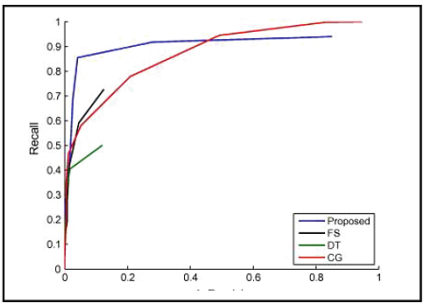


Fig. 4. Recall-precision of Pet dataset.

Table 1. The Recall and Precision Comparison

Methods	Comparison		
	Recall	Precision	F-Measure
Felzenszwalb and el.	0.74	0.89	0.79
Dalal and Triggs	0.51	0.88	0.66
Roth and el.	0.80	0.81	0.80
Our proposed	0.88	0.99	0.92

4.2 CAVIAR Dataset

Datasets Caviar show a corridor in a shopping center from two different angles. First corner side of the corridor, the second corner of the face directly. Because we are interested in the process of discovering who the percentage change should we focus on first dataset. Data may or JPEG and MPEG resolution is 384 × 288. For our experiment, we choose a fairly complex data set to assess the ShopAssistant2cor because it contains a large number of pedestrians (e.g. 1265). There are 370 frames with size 384 × 128 image. 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 × 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 dataset is shown in Fig. 5 and Table 2. Again it can be seen that the detection adaptive grid (CG-OOL) better than the generic object detection (HOG-DT and DPM-FS), especially Recall. Results illustrated object detection is shown in Fig. 6(b).

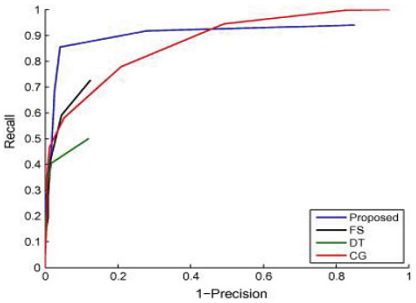
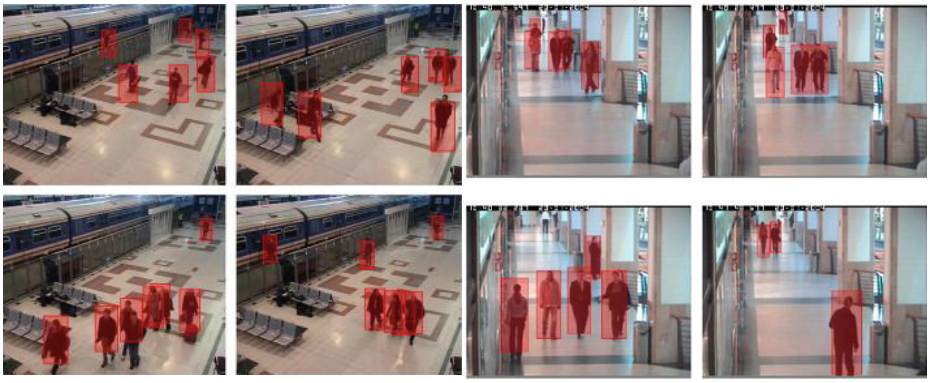


Fig. 5. Recall-precision of Caviar dataset

Table 2. The Recall and Precision Comparison

Methods	Comparison		
	Recall	Precision	F-Measure
Felzenszwalb and el.	0.62	0.90	0.74
Dalal and Triggs	0.41	0.91	0.57
Roth and el.	0.78	0.87	0.82
Our proposed	0.92	0.93	0.92



(a) (b)

Fig. 6. (a) Illustrative detection results of person detector for the Pet Sequence. (b) Illustrative detection results of person detector for the Caviar Sequence.

5 Conclusion

We developed a approaches to incorporate this information based on the idea of classifier grids. Classifier grids divide the input image into highly overlapping grid elements, where each grid element contains its own classifier. Based on the idea of classifier grids learning strategies. To allow for incorporating both, scene specific object information as well as scene specific background information we propose a co-training related approach for the classifier grids, i.e., classifier co-grid. In combination with our robust learning algorithm this allows for incorporating unlabeled information from the scene but still preserving the reliable labeled information. This approached aim to preserve long-term stability, which is given by either using specific update strategies or by using our robust learning algorithm. We demonstrate the long-term stability of the proposed approaches empirically by evaluating them on a real-world surveillance scenario, where a corridor in a public building is monitored over one week and object detection is performed. Even though the whole approach is

updated without supervision, the robustness is preserved. The experimental results, demonstrating against the problem with different objects, clearly shows that very good results detected with high precision, while ensuring that it can perform online, adapting with many environmental and drifting problem detection system for short-term performance improvements explicit.

References

1. Agarwal, S., Awan, A., Roth, D.: Learning to detect objects in images via a sparse, part-based representation. IEEE Trans. Pattern Anal. Mach. Intell. **26**(11), 1475–1490 (2004)
2. Andrews, S., Tsochantaridis, I., Hofmann, T.: Support vector machines for multiple-instance learning. In: Advances in NIPS, pp. 561–568 (2003)
3. Babenko, B., Yang, M.-H., Belongie, S.: Visual tracking with online multiple instance learning. In: IEEE Conference on CVPR (2009)
4. Blum, Mitchell, T.: Combining labeled and unlabeled data with co-training. In: Proceedings of Conference on Computational Learning Theory, pp. 92–100 (1998)
5. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: IEEE Conference on CVPR, vol. I. pp. 886–893 (2005)
6. Viola, P., Platt, J.C., Zhang, C.: Multiple instance boosting for object detection. In: Advances in Neural Information Processing Systems, pp. 1417–1426 (2005)
7. Felzenszwalb, P., McAllester, D., Ramanan, D.: A discriminatively trained, multiscale, deformable part model. In: IEEE Conference on CVPR (2008)
8. Goldberg, B., Li, M., Zhu, X.: Online manifold regularization: a new learning setting and empirical study. In: Proceeding on European Conference on Machine Learning and Knowledge Discovery in Databases, vol. I, pp. 393–407 (2008)
9. Grabner, H., Roth, P.M., Bischof, H.: Is pedestrian detection really a hard task?. In: Proceeding of IEEE Workshop on Performance Evaluation of Tracking and Surveillance (2007)
10. Hoiem, D., Efros, A.A., Hebert, M.: Putting objects in perspective. In: Proceeding of IEEE Conference on Computer Vision and Pattern Recognition, vol. II, pp. 2137–2144 (2006)
11. Javed, O., Ali, S., Shah, M.: Online detection and classification of moving objects using progressively improving detectors. In: Proceeding of IEEE Conference on CVPR, vol. I, pp. 696–701 (2005)
12. Leibe, B., Leonardis, A., Schiele, B.: Robust object detection with interleaved categorization and segmentation. Int. J. Comput. Vis. **77**(1–3), 259–289 (2008)
13. Levin, P., Viola, Freund, Y.: Unsupervised improvement of visual detectors using co-training. In: Proceedings of ICCV, vol. I, pp. 626–633 (2003)
14. Li, L.J., Wang, G., Fei-Fei, L.: Optimol: automatic online picture collection via incremental model learning. In: Proceeding of IEEE Conference on CVPR, pp. 1–8 (2007)
15. Grabner, H., Bischof, H.: On-line boosting and vision. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, vol. I, pp. 260–267 (2006)
16. Nair, V., Clark, J.J.: An unsupervised, online learning framework for moving object detection. In: Proceedings of IEEE Conference on CVPR, vol. II, pp. 317–324 (2004)
17. Rosenberg, C., Hebert, M., Schneiderman, H.: Semi-supervised self-training of object detection models. In: IEEE Workshop on Applications of Computer Vision, pp. 29–36 (2005)

18. Wu, B., Nevatia, R.: Improving part based object detection by unsupervised, online boosting. In: Proceedings of IEEE Conference on CVPR, pp. 1–8 (2007)
19. Roth, P.M., Grabner, H., Skočaj, D., Bischof, H., Leonardis, A.: On-line conservative learning for person detection. In: Proceedings of IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance, pp. 223–230 (2005)
20. Roth, P.M., Sternig, S., Grabner, H., Bischof, H.: Classifier grids for robust adaptive object detection. In: Procceding of IEEE Conference on CVPR (2009)
21. Torralba, A.: Contextual priming for object detection. IJCV **53**(2), 169–191 (2003)
22. Koller, D., Heitz, G.: Learning spatial context: using stuff to find things. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part I. LNCS, vol. 5302, pp. 30–43. Springer, Heidelberg (2008)