Pedestrian Detection Based on Incremental Learning

Yu Xia[1], Yongzhen Huang[2], Liang Wang[2], and Xin Geng[1,⋆]

[1] School of Computer Science and Engineering, Southeast University, Nanjing, China
xgeng@seu.edu.cn
[2] National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, China

Abstract. Pedestrian detection is a hot topic in computer vision and pattern recognition. Existing pedestrian detection methods face new challenges in the background of big data, e.g., heavy burdens on computing and memory. To solve these problems, in this paper, we propose a pedestrian detection framework based on incremental learning. Compared with existing pedestrian detection frameworks, it costs much less time and memory. In addition, the performance of our framework is very close to the one which uses all training samples at once. Furthermore, with more new training samples, the performance can be enhanced continually with little time and memory, showing the potential in practical applications.

Keywords: Pedestrian detection, Incremental learning, Converged passive-aggressive, Histograms of oriented gradients.

1 Introduction

Pedestrian detection is an important issue in computer vision and can be applied in many related areas [4,5]. Many effective algorithms are developed. For example, Viola *et al.* [13,14] propose object detection algorithms based on Haar-like features and the cascade structure. However, Haar-like descriptors are not able to describe more complicated objects (such as pedestrians under complex backgrounds). To obtain more robust descriptor, Dalal *et al.* [3] propose HOG (histograms of oriented gradients) for pedestrian detection. HOG describes shape and structure well, and obtains satisfying results in pedestrian detection. However, this approach takes no account of local deformation of objects. To solve this problem, Felzenszwalb *et al.* [6] propose a discriminative part-based approach that models unknown part positions as latent variables. It is one of the most successful approaches for object detection at present.

In the background of big data, it is possible to collect huge pedestrian data sets, which are usually captured continuingly. Existing pedestrian detection methods do not perform well in this situation and face some new challenges.

⋆ Corresponding author.

C. Sun et al. (Eds.): IScIDE 2013, LNCS 8261, pp. 603–610, 2013.

Think about this case: collecting new training sets from various sources (e.g., web, cameras, videos and so on) continuingly after training a detector using former training sets, and updating the detector with new training sets. Traditionally, we need to merge the former and the new training sets, and use the merged training set to train a new detector. In this case, existing methods consume huge time when data sets are very large, e.g., million or billion. What's more, limited memory is not enough to train such a detector. The amount of data is always too large for machines to handle. It is necessary to study incremental learning based pedestrian detection which makes feasible in case of big data, runs fast and maintains satisfying performance.

There is little work on embedding incremental learning into general object detection. Opelt *et al.* [10] introduce incremental learning into visual shape alphabet based multi-class object detection. This method is designed for a small or medium size of training sets but not suitable for big data. Nair *et al.* [7,8,9] combine online learning and object detection in video with simplex and still background. These methods are not effective for general object detection with complex background.

In this paper, we propose an incremental algorithm which is simple but effective. We apply it to pedestrian detection and obtain satisfying results. Our framework has three merits: (1) running very fast; (2) needing only a small size of memory; (3) improving the detection accuracy continually. Our algorithm achieves good performance on the INRIA and the NICTA pedestrian datasets [3,11] in terms of both effectiveness and efficiency, which demonstrates that our algorithm is suitable for pedestrian detection with huge data.

2 Pedestrian Detection Based on CPA

2.1 Framework

The HOG based pedestrian detection algorithm [3] is probably the first pedestrian detection approach which can achieve satisfying performance for real images with complex backgrounds [4,5]. In Fig.1, the diagram inside the dotted line box shows the procedure of HOG based pedestrian detection. Although this approach is less accurate than the part-based model [6], its complexity is far lower and thus it is very suitable to be combined with incremental learning. This is the main reason why we adopt HOG based pedestrian detection as our baseline platform.

Fig.1 shows the framework of incremental learning based pedestrian detection. Intuitively, this framework has two advantages. On one hand, if the data sets are collected continuingly, the system could rapidly and effectively train detectors using new data sets. On the other hand, if a data set is so large that memory can not meet the demand, we could divide it into several parts and use them to train a detector according to the procedure in Fig.1.

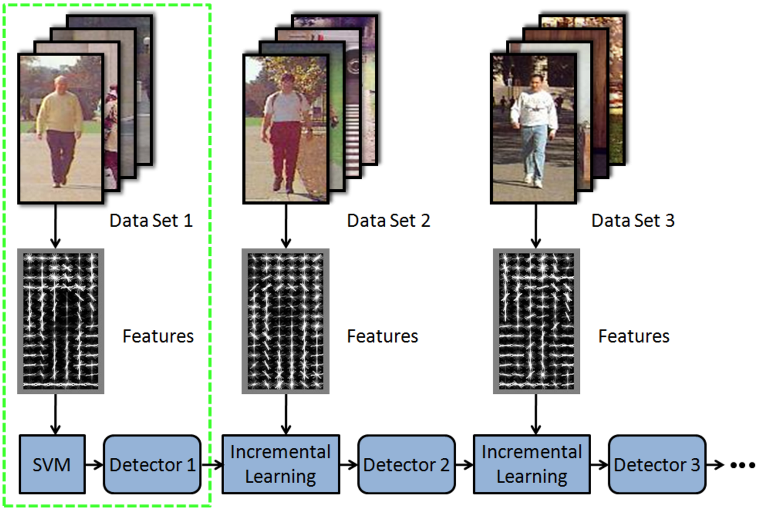


Fig. 1. Framework of incremental learning based pedestrian detection

2.2 Incremental Learning Algorithm

Compared with batch learning algorithms, incremental learning algorithms run faster and cost much less memory. We propose a new incremental learning algorithm called Converged Passive-Aggressive algorithm (CPA) for pedestrian detection.

CPA is based on the Passive-Aggressive algorithm [2]. The objective function of Passive-Aggressive algorithm is,

$$\mathbf{W}_{t+1} = \underset{\mathbf{W}\in \mathrm{R}^n}{\operatorname{argmin}} \frac{1}{n}||\mathbf{W} - \mathbf{W}_t||^2 + C\xi^2 \quad \text{s.t.} \quad \ell(\mathbf{W}; (\mathbf{X}_t, y_t)) \leq \xi. \tag{1}$$

Here $\mathbf{W}_t$ is the vector containing parameters of the detector, $\mathbf{X}_t$ is the sample, and y_t is the label of the sample in round t. ℓ is hingle-loss function, ξ is a non-negative slack variable, and C is the weight of ξ^2. $\mathbf{W}_{t+1}$ is the projection of $\mathbf{W}_t$ onto the half-space of vectors whose hingle-loss is zero for the current sample.

If $\ell(\mathbf{W}; (\mathbf{X}_t, y_t)) > 0$, the Lagrangian of the problem in (1) is

$$L(\mathbf{W}, \xi, \alpha) = \frac{1}{2}||\mathbf{W} - \mathbf{W}_t||^2 + C\xi^2 + \alpha(1 - \xi - y_t(\mathbf{W}^T\mathbf{X}_t)). \tag{2}$$

Here α is a non-negative Lagrange multiplier. Setting the partial derivatives of L with respect to $\mathbf{W}$ and ξ to zero respectively, we obtain

$$\mathbf{W} = \mathbf{W}_t + \alpha y_t \mathbf{X}_t. \tag{3}$$

$$\xi = \frac{\alpha}{2C}. \tag{4}$$

Replacing ξ and $\mathbf{W}$ with (3) and (4), the Lagrangian can be expressed as

$$L(\alpha) = -\frac{\alpha^2}{2}(||\mathbf{X}_t||^2 + \frac{1}{2C}) + \alpha(1 - y_t(\mathbf{W}_t{}^T\mathbf{X}_t)). \quad (5)$$

Setting the derivative of $L(\alpha)$ to zero, we obtain

$$\alpha = \frac{1 - y_t(\mathbf{W}_t{}^T\mathbf{X}_t)}{||\mathbf{X}_t||^2 + \frac{1}{2C}}. \quad (6)$$

It should be noted that each sample is used only once so that the solution may not be convergent. To address this problem, a weight-decay strategy is proposed as shown in Algorithm 1.

In Algorithm 1, C_{INIT} is an initial parameter. dr is a scalar which regulates the decreasing rate of C. M is the number of samples. K is the number of iterations in the outer loop. Crammer *et al.* [2] show that setting C to be a small number leads to a slow progress rate and gets better solution when training sets are large. So, C is set to decrease in each round. Generally, the outer loop of CPA needs a small number of (less than 20) iterations.

Algorithm 1. CPA

Input:

Initial Passive-Aggressive parameter, C_{INIT};
Decreasing rate of C_{INIT}, parameter dr;
Initial model $\mathbf{W}_{1,1}$;

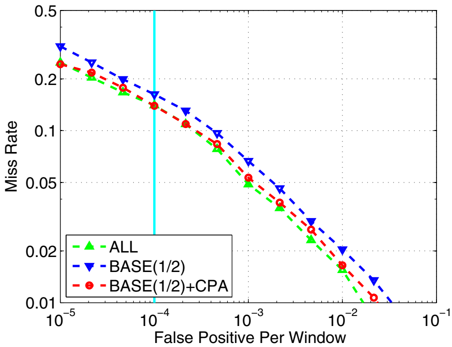
Output:

Updated model $\mathbf{W}_{K,M+1}$.
for $k = 1,2,\cdots,K$ **do**
  $C = \frac{C_{INIT}}{dr \times k}$
  for $t = 1,2,\cdots,M$ **do**
    receive instance: $\mathbf{X}_t \in \mathbf{R}^n$
    receive correct label: $y_t \in \{-1, +1\}$
    compute loss: $\ell_t = \max\{0, 1 - y_t(\mathbf{W}_{k,t}{}^T\mathbf{X}_t)\}$
    compute Lagrange multiplier: $\alpha_{k,t} = \frac{\ell_t}{||\mathbf{X}_t||^2 + \frac{1}{2C}}$
    update: $\mathbf{W}_{k,t+1} = \mathbf{W}_{k,t} + \alpha_{k,t} y_t \mathbf{X}_t$
  end for
  $\mathbf{W}_{k+1,1} = \mathbf{W}_{k,M+1}$
end for

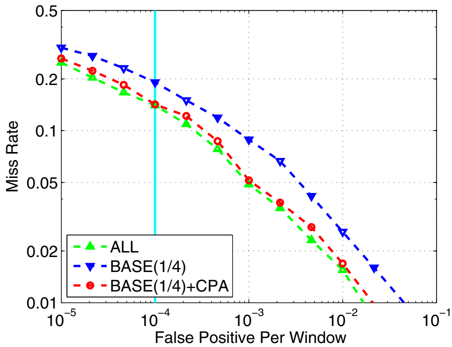
3 Experiments and Results

3.1 Experimental Datasets and Setup

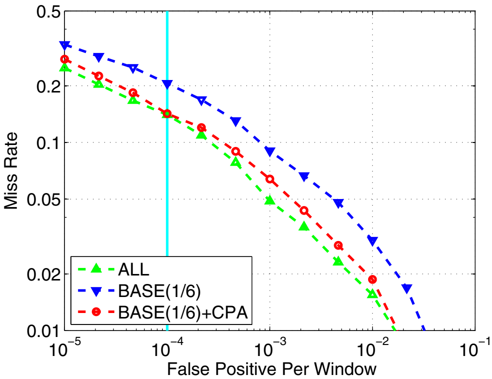
Experiments are conducted on two different pedestrian datasets. The first is the well known INRIA pedestrian dataset [3], containing 3,542 (2,416 for training



(a) BASE is trained with 1/2 training set



(b) BASE is trained with 1/4 training set



(c) BASE is trained with 1/6 training set

Fig. 2. Performance of CPA with different initial detectors

and 1,126 for testing) positive samples and 1,671 (1,218 for training and 453 for testing) background images. The resolution of pedestrian images is 128×64 pixels. We use this database to evaluate the major characteristics of our proposed algorithm. The second is the NICTA pedestrian dataset [11], containing 44,223 (37,344 for training and 6,879 for testing) positive samples and 5,151 (4,147 for training and 1,004 for testing) background images. The resolution of pedestrian images is 80×32 pixels. We use this database to simulate the performance of our framework in the case that training samples are continually obtained.

Because the resolution of images in the two datasets is different, we set the size of each cell to 8×8 pixels on the INRIA dataset and 4×4 pixels on the NICTA dataset. To initialize, we use a soft linear SVM trained with libSVM [1].

3.2 Results on INRIA

Experiments are designed to study the learning ability of CPA in various situations. First, dividing the training set into two parts evenly, we use the first part to train a linear SVM and obtain an initial detector denoted as BASE. Then we use the other part to update the detector with CPA and obtain a detector denoted as BASE+CPA. To compare the performance of CPA, all training samples of the INRIA database are also used to train a linear SVM, and get a detector denoted as All. Fig.2(a) shows the DET (Detection Error Tradeoff) curves [3] of BASE, BASE+CPA and ALL.

Table 1. Time(sec) consumed by each detector

	BASE(1/2)	BASE(1/4)	BASE(1/6)	ALL
BASE	56.05	15.60	8.75	145.74
BASE+CPA	56.05+1.57	15.60+1.23	8.75+1.37	-

Then, we do experiments to compare the performance of BASE+CPA with different initial detectors. Dividing the training set into four parts evenly, we use the first part to train a linear SVM and obtain an initial detector. Then we use the rest three parts to update the detector with CPA three times. Each time we use one part. Fig.2(b) shows the DET curves of the detectors. Likewise, dividing the training set into six parts evenly, we train the detector as the previous procedures. This result is shown in Fig.2(c). In addition, the time costed by training each detector is shown in Table 1. From Fig.2 and Table 1, it is easy to obtain the following conclusions:

1. The more training samples, the better the detector performs. For example, the results is gradually enhanced over detectors BASE(1/6), BASE(1/4), BASE(1/2) and ALL.
2. No matter which baseline detector is adopted, our proposed CPA can enhance the baseline effectively and finally performs comparably with the detector ALL.
3. CPA costs very little time compared with SVM, which means that CPA is very efficient to reach satisfying detection accuracy.

We also compare CPA with two incremental learning algorithms (Pegasos [12] and Passive-Aggressive algorithm) on the INRIA database. Pegasos is a popular incremental learning algorithm for linear SVMs. Table 2 shows the miss rate at 10^{-4} FPPW (false positive per window) of different detectors. The results of CPA are the best in all cases. On average, CPA has above 1.5% lower miss rate than Passive-Aggressive and above 0.8% lower miss rate than Pegasos.

3.3 Results on NICTA

To simulate incremental learning based pedestrian detection with training data continually captured, experiments are designed on the NICTA pedestrian database. First, using 1,000 positive and 5,000 negative samples to train a linear SVM, we obtain an initial detector. Then, we use new data (the ratio of positive

Table 2. Miss rate of different detectors at 10^{-4} FPPW

	BASE(1/2)	BASE(1/4)	BASE(1/6)
Passive-Aggressive	15.63%	15.90%	15.36%
Pegasos [12]	14.65%	15.01%	15.19%
CPA	**13.94%**	**14.12%**	**14.21%**

samples to negative samples is 1/5) to train the detector continuingly as the procedures in Fig.1.

Fig.3 shows that the miss rate at 10^{-4} FPPW decreases with incrementally learning from new data. After initializing a detector, we use new training data to update the detector 16 times. Miss rate declines continually, which demonstrates that our framework has good ability for training better detectors.

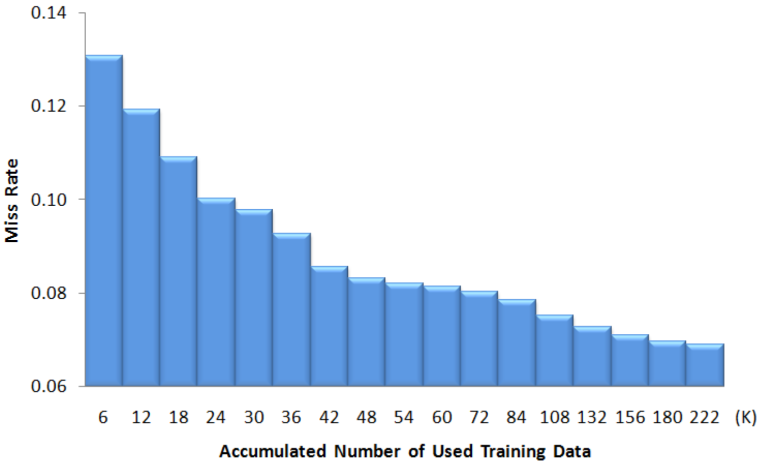


Fig. 3. The performance of the detector with increasing new training samples

4 Conclusions

In this paper, we have proposed a pedestrian detection framework with incremental learning. The detection accuracy of our method is comparable to traditional pedestrian detection which uses all training samples at once, but consumes much less time and memory. Furthermore, we demonstrate that the detector can be continually enhanced with more new training samples, which is very valuable for practical applications.

Finally, we conclude this paper with two main contributions. First, we have proposed a new incremental learning algorithm CPA to enhance online Passive-Aggressive algorithm. Second, we have embedded the enhanced incremental learning algorithm into pedestrian detection to effectively and efficiently solve problems of existing pedestrian detection methods in the background of big data.

Acknowledgment. This research was supported by the National Science Foundation of China (61273300, 61232007, 61203252, 61175003), the Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry, the Excellent Young Teachers Program of Southeast University, the Key Lab of Computer Network and Information Integration of Ministry of Education of China.

References

1. Chang, C.C., Lin, C.J.: Libsvm: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology 2(27), 1–27 (2011)
2. Crammer, K., Dekel, O., Keshet, J., Shalev-Shwartz, S., Singer, Y.: Online passive-aggressive algorithms. The Journal of Machine Learning Research 7, 551–585 (2006)
3. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Computer Vision and Pattern Recognition, vol. 1, pp. 886–893 (2005)
4. Dollár, P., Wojek, C., Schiele, B., Perona, P.: Pedestrian detection: An evaluation of the state of the art. IEEE Transactions on Pattern Analysis and Machine Intelligence 34(4), 743–761 (2012)
5. Enzweiler, M., Gavrila, D.M.: Monocular pedestrian detection: Survey and experiments. IEEE Transactions on Pattern Analysis and Machine Intelligence 31(12), 2179–2195 (2009)
6. Felzenszwalb, P.F., Girshick, R.B., McAllester, D., Ramanan, D.: Object detection with discriminatively trained part-based models. IEEE Transactions on Pattern Analysis and Machine Intelligence 32(9), 1627–1645 (2010)
7. Grabner, H., Bischof, H.: On-line boosting and vision. In: Computer Vision and Pattern Recognition, vol. 1, pp. 260–267 (2006)
8. Javed, O., Ali, S., Shah, M.: Online detection and classification of moving objects using progressively improving detectors. In: Computer Vision and Pattern Recognition, vol. 1, pp. 696–701 (2005)
9. Nair, V., Clark, J.J.: An unsupervised, online learning framework for moving object detection. In: Computer Vision and Pattern Recognition, vol. 2, pp. 310–317 (2004)
10. Opelt, A., Pinz, A., Zisserman, A.: Incremental learning of object detectors using a visual shape alphabet. In: Computer Vision and Pattern Recognition, vol. 1, pp. 3–10 (2006)
11. Overett, G., Petersson, L., Brewer, N., Andersson, L., Pettersson, N.: A new pedestrian dataset for supervised learning. In: IEEE Intelligent Vehicles Symposium, pp. 373–378 (2008)
12. Shalev-Shwartz, S., Singer, Y., Srebro, N.: Pegasos: Primal estimated sub-gradient solver for svm. In: International Conference on Machine Learning, pp. 807–814 (2007)
13. Viola, P., Jones, M.J.: Robust real-time face detection. International Journal of Computer Vision 57(2), 137–154 (2004)
14. Viola, P., Jones, M.J., Snow, D.: Detecting pedestrians using patterns of motion and appearance. International Journal of Computer Vision 63(2), 153–161 (2005)