Transfer learning for pedestrian detection

Xianbin Cao [a], Zhong Wang [b], Pingkun Yan [c,*], Xuelong Li [c]

[a] BeiHang University, Beijing, 100083, PR China
[b] University of Science and Technology of China, Hefei, 230026, PR China
[c] Center for Optical Imagery Analysis and Learning (OPTIMAL), State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, Shaanxi, PR China

ARTICLE INFO

ABSTRACT

Most of the existing methods for pedestrian detection work well, only when the following assumption is satisfied: the features extracted from the training dataset and the testing dataset have very similar distributions in the feature space. However, in practice, this assumption does not hold because of the scene complexity and variation. In this paper, a new method is proposed for detecting pedestrians in various scenes based on the transfer learning technique. Our proposed method employs the following two strategies for improving the pedestrian detection performance. First, a new sample screening method based on manifold learning is proposed. The basic idea is to choose samples from the training set, which may be similar to the samples from the unseen scene, and then merge the selected samples into the unseen set. Second, a new classification model based on transfer learning is proposed. The advantage of the classification model is that only a small number of samples need to be used from the unseen scenes. Most of the training samples are still obtained from the training scene, which take up to 90% of the entire training samples. Compared to the traditional pedestrian detection methods, the proposed algorithm can adapt to different scenes for detecting pedestrians. Experiments on two pedestrian detection benchmark datasets, DC and NICTA, showed that the method can obtain better performance as compared to other previous methods.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

A major assumption in many traditional pedestrian detection methods is that the training dataset and test dataset should be similar, that is the training images should be taken from similar scenes as the testing images. However, in real pedestrian detection applications, the scene may be changing and become more complex due to varying environment such as light, weather, background and pedestrian clothing. Therefore, the traditional pedestrian detection methods may not work well when the unseen scene is different from the training set. How to find the common part between the different scenes and design pedestrian detection algorithm for different scenes specially is a problem in the current research.

Although the distribution of the training data and unseen data is different, there exists certain correlation between the two dataset because of the same feature space. Specifically, the training set has some samples which are similar to the unseen set. If the samples with high similarity in the training set are transferred to unseen scene, it can expand the scale of the unseen set. Traditional image similarity measurement typically adopts Euclidean distance. However, as the feature dimension of pedestrian image is usually very high, it is difficult to accurately measure the similarity between samples using Euclidean distance. The manifold learning method in the machine learning field has the advantage of feature dimension reduction and data visualization. It can find a low-dimensional manifold from the high dimensional space and computes the corresponding embedding mapping. Therefore, manifold learning theory is used to visualize the samples in the training set and unseen set, and further select the samples of high similarity in the two dataset. This is the most intuitive results which are obtained from the feature level.

In order to further find the shared knowledge among the samples from the two datasets, the transfer learning method is used to solve this problem. According to transfer learning, even though data from the source domain cannot be directly applied, there exists certain portion of the data being useful in learning for the target domain. The related work is to train a transferred classifier using a small number of samples in the target domain. However, transfer learning cannot be directly used for detecting pedestrians due to the complexity of the scenes and the large variation between them.

In this paper, a new method is proposed for pedestrian detection using modified transfer learning. Considering that there

may be only a small number of samples available from the testing dataset, it will be difficult to train a reliable classification model. Although the distribution of the two datasets is different, they are in the same feature space, so there exists certain similarity between the training dataset and the testing dataset. The motivations of our work are as follows. We select some samples in the training set which are similar to the samples in the unseen scene through manifold learning method, and then use samples in the training set as much as possible to assist pedestrian detection in the unseen scene by transfer learning theory.

There are two main contributions of this paper. First, a new sample screening method based on Isomap algorithm is proposed. The method describes the samples in the training scene and the unseen scene visually, and then selects several samples in the training set, which are very similar to the samples in the unseen set. The selected samples are merged into the unseen set, so the method can expand the scale of the unseen scene and provide strong basis for the subsequent classifier. Second, a new classification model based on transfer learning is proposed. This algorithm only uses a small number of samples from the testing scene to assist the construction of an effective classification model. The advantage of the algorithm is that as much as possible to transfer the training scene knowledge to the testing scene for pedestrian detection. Compared to the traditional methods, the method can suit for unseen scenes for pedestrian detection.

Two datasets, DC and NICTA, which are significantly different, are used as the training set and the testing set, respectively. It was shown by our experiments that, by using only a small number of samples from the testing dataset, the proposed method was able to obtain better performance in detecting pedestrians in the NICTA dataset. The following contents of the paper are organized as follows. In Section 2, the related works are discussed. In Section 3, we describe the framework of the proposed pedestrian detection method including sample screening and classification. In Section 4, the experimental results are presented. Finally, we give the concluding remarks in Section 5.

2. Related works

Pedestrian detection method based classification is the mainstream in this field. In practice, the classifier which trained in the training set cannot achieve satisfactory result because of the scene complexity and variation. The previous methods can be divided into three categories: (1) The first kind of methods train a reliable classifier using the training dataset and detect pedestrians in the new scenes directly [1–9]. This kind of methods can achieve successful application when the training set and the testing set are similar. However, when the training set and the test set come from different distributions, the detection performance will decrease for the reason that the detector trained in the training scene may not be able to correctly classify the samples from the unseen scene. (2) The second category of methods are to rebuild a new classification model in the unseen scene for pedestrian detection [25–30]. These methods are either computational expensive or having difficulties in collecting the needed training samples and rebuilding the classification model. Therefore, it is difficult to directly apply these methods in real applications. (3) The last category of methods are to adjust the classifier dynamically for the unseen scene. Although this method achieves greater improvement, it may not meet the effective detection performance for pedestrian detection systems [10].

So far, the existing research on pedestrian detection system focuses on the specific scene, that is the training set and test set

come from the same scene. Most pedestrian detection methods use single classifiers like Support Vector Machine (SVM), Radial Basis Functions (RBF), Neural Networks (NN), or an integrated classifier to detect pedestrians. Multilayer neural networks have also been applied [11,12]. Various feature sets have been used in combination with different classifiers for detection [13,14]. Nonlinear SVM classifier yielded further performance boosts, however, the increase in computational costs and memory requirements restrict its application [15,16]Gavrila [17] trained a RBF network to detect pedestrians. In order to improve the low detection rate, Gavrila et al. [6] adopted simulated annealing to train the RBF based classifier. Szarvas et al. [18] proposed a neural network based classifier to perform detection. Considering that the pedestrians have various postures and motions, it is hardly to distinguish pedestrians from non-pedestrians using a single classifier in a precise and rapid way. Therefore, single classifier based detection method is unsuitable for high performance pedestrian detection. In order to improve the detection performance, integrated classifier was wide concerned over the past decade. Viola et al. [19,20] first proposed cascaded classifiers, which consisted of a series of connected single classifiers. Each single classifier was trained by using the AdaBoost algorithm [21]. The cascaded classifiers received wide acceptance and were adopted by many others [22,23]. Xu et al. [8] proposed a cascade SVM classifier for pedestrian detection. However, the improvement on detection performance is not significant. A cascade of L1-norm minimization based learning classifiers was later proposed by Xu et al. [24] to detect human in images. Although classification based methods have been successfully used in some applications, as pedestrian detection systems often need to handle new scenes, the detection performance will decrease when the training scene and the unseen scene are different.

The second type of methods are rebuilding a new classification model in the unseen scenes. The basic idea is to a small number of samples taken from the testing dataset to assist the construction of a new strong classifier to adapt to the unseen scenes for pedestrian detection. However, there exist only few related works on pedestrian detection systems. Jiang et al. [25] proposed to remove training examples, which may be misleading, from the training set and then build a strong classifier for natural language processing task. Wu et al. [26] integrated the training dataset using an SVM based framework for improving the performance of image classification. However, no quantitative study results were provided. Dai et al. [27] proposed a boosting algorithm to boost the inductive transfer learning, which was then successfully applied for text classification. The key idea is to reduce the weights of the bad samples while increasing the weights of the good samples by dynamically adjusting the weights. Liao et al. [28] proposed to learn the mismatches between the training data and testing data using Migratory–Logistic algorithm. Rosenstein et al. [29] proposed a transfer learning approach using hierarchical simple Bayes and discussed the usage of transfer learning. Wang et al. [30] proposed an approach to align manifolds without correspondence, which can then be used to transfer the knowledge from one source domain to the target domain through the aligned manifolds.

The third type of methods are to adjust the classifiers by making the classifiers dynamically adapt to the unseen scenes. For example, Wang et al. [10] employed the ternary pattern for pedestrian detection. Although this method has achieved certain applications, it may not be able to meet the requirement on the detection performance of pedestrian detection systems.

In this paper, a novel pedestrian detection method based on transfer learning is proposed. Different from the previously reported

methods, the proposed method uses transfer learning to preprocess the training dataset, and then builds a strong classifier only using a small number of samples taken from the testing dataset.

3. Transfer learning for pedestrian detection

3.1. Problem definition

Let $D_a = x_i^a, y_i^a|_{i=1}^n$ be training dataset in the training scenes and it is also an auxiliary dataset. $D_s = x_i^s, y_i^s|_{i=1}^m$ is training dataset in the unseen scenes, which contains only a small amount of labeled samples and they are not sufficient to train a classifier alone. D_s and D_a may come from different distributions. $D_t = x_i^t|_{i=1}^k$ is test dataset in the unseen scenes. n, m and k are the size of the dataset. D_s and D_t come from the same scene and they are in the same feature space, while D_a comes from other scenes. The training dataset $D = D_a \cup D_s$ is divided into two labeled dataset D_a and D_s. The method is to train a classifier for D_t with a small amount of data D_s and many auxiliary data D_a. The total training set $D = D_a \cup D_s = (x_i, y_i)$ is defined as follows:

$$x_i = \begin{cases} x_i^a & i = 1,\ldots,n \\ x_i^s & i = n+1,\ldots,n+m \end{cases}$$

3.2. Sample screening based on manifold learning

Although D_a and D_s come from different scenes, it has certain similarity among the pedestrian samples. The problem is how to find out the common between the two dataset. The dataset D_s only contain 100 samples that they can not construct a reliable classification model, so the goal is to select the useful part of D_a to expand the dataset D_s.

Due to the high dimension of pedestrian samples, the traditional clustering methods use Euclidean distance to measure the sample similarity, so it can not accurately visualize the pedestrian samples in the pedestrian detection scenes. In this paper, manifold learning method was used to visualize the pedestrian samples. Isomap algorithm [31] is used to solve this problem. It may be viewed as an extension to multidimensional scaling (MDS). Isomap algorithm consists of two main steps: (1) Estimate the geodesic distances between points in the input as the shortest paths on the dataset's k-nearest neighbor graph; (2) Find points in low-dimensional Euclidean space using MDS, whose inter-point distances match the distances found in the last step.

Assume that the training data D_s in the unseen scenes are in a low dimensional manifold, the samples in auxiliary training data D_a may be part of the manifold. Our goal is to select this part of training data to expand the scale of the training data in the unseen scenes. The specific sample screening algorithm is as follows.

Step (1) Compute the nearest neighbor points of D_a in D_s using k nearest neighbor method for each sample in the D_s;
Step (2) Construct manifold graph of D_a and D_s respectively in the same graph using Isomap algorithm;
Step (3) Calculate the center point of all the samples, find the path of each sample to the center point in the manifold map of each sample, add the samples of auxiliary dataset in the path to the training set for the unseen scenes.

Fig. 1 gives the two-dimensional visual representation graph using Isomap algorithm in the pedestrian detection scenes. "Red cross" and "green circle" are used to mark the samples in the unseen dataset and training dataset, respectively. As we can see
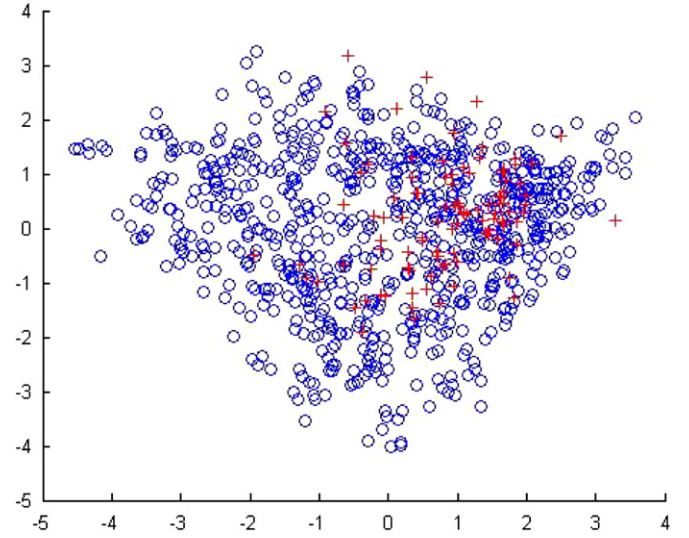

Fig. 1. Two-dimensional visual representation map of same scene.

from the graph, the samples in the training scenes and the unseen scenes basically overlap, which means that they have similar distributions. Thus, the traditional pedestrian detection methods can achieve good performance. Fig. 2 gives the three-dimensional visual representation map when the training scene is different from the testing scene. As we can see from the map, although the distribution of samples in the training dataset and the testing scene is obviously different, part of the samples in the training set and the testing set have overlap with each other. It means that there is certain correlation between them. Owing to the fact that there are few training data in the unseen scenes, Sample screening algorithm was used to select several similar samples from the auxiliary data to make up D_s. Then we design the classification algorithm specifically on D_s.

3.3. Classification based transfer learning

In PDS, although the feature set and labels may be similar, the distributions of data points from the training set and testing set can be significantly different. So it belongs to the transfer learning problem. In addition, in the inductive transfer learning setting, test data and training data are different, no matter if the training and test domains are the same or not. The inductive transfer learning setting can be categorized into two cases: (1) A large number of labeled samples are available from the source domain; (2) No labeled sample available from the source domain. Most of the transfer learning methods in this setting focus on the first case. So the new problem in our paper is an inductive transfer learning problem.

In this section, we propose a novel classification method based on transfer learning, which is an extension of the AdaBoost algorithm. We named it as ITLAdaBoost algorithm which it is an inductive transfer learning algorithm and transfer knowledge from the training scenes to the unseen scenes. Algorithm 1 gives the description of the ITLAdaBoost algorithm. As we can see from the algorithm, ITLAdaBoost uses a small amount of new scene data and large amount of old scene to train a classifier. Due to the difference in distributions between the auxiliary data D_a and new scene data D_s, some of the auxiliary data may be useful in learning for the new scene but some of them may not and could even be harmful. So ITLAdaBoost algorithm attempts to iteratively adjust the weight of auxiliary data for the new scene data.
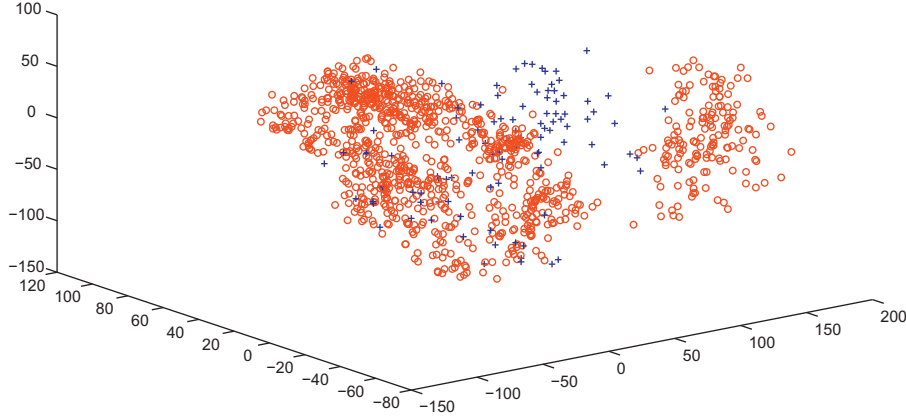
Fig. 2. Three-dimensional visual representation map of different scenes.

Algorithm 1. ITLAda Boost Algorithm

Input: new scene data D_s, old scene data D_a and test dataset D_t

Step 1: Initialize the initial weight vector of each sample

$$w_i^1 = \begin{cases} 1/n & i=1,\ldots,n \\ 1/m & i=n+1,\ldots,n+m \end{cases}$$

Step 2: For $t=1,\ldots,T$

  Step 2.1: Normalize the weights

  $$w_i^t \leftarrow \frac{w_i^t}{\sum_{j=1}^{n+m} w_j^t}$$

  Step 2.2: Train a classifier h_t with the distribution over the training set $D = D_a \cup D_s$ and the test set D_t.

  Step 2.3: Compute the error of h_t on D_a:

  $$\varepsilon_t^a = \sum_{i=1}^{n} \frac{w_i^t |h_t(x_i)-y_i|}{\sum_{i=1}^{n} w_i^t}$$

  Compute the error of h_t on D_s:

  $$\varepsilon_t^s = \sum_{i=n+1}^{n+m} \frac{w_i^t |h_t(x_i)-y_i|}{\sum_{i=n+1}^{n+m} w_i^t}$$

  Step 2.4: Update the weights:

  $$w_i^{t+1} = \begin{cases} w_i^{t+1}(\beta_t^a)^{|h_t(x_i)-y_i|} & i=1,\ldots,n \\ w_i^{t+1}(\beta_t^s)^{-|h_t(x_i)-y_i|} & i=n+1,\ldots,n+m \end{cases}$$

  where $\beta_t^a = \varepsilon_t^a/(1-\varepsilon_t^a)$ and $\beta_t^s = \varepsilon_t^s/(1-\varepsilon_t^s)$.

Step 3: Output the final classifier:

$$C(x) = \begin{cases} 1 & \sum_{t=\lceil T/2 \rceil}^{T} \alpha_t h_t(x) \geq \frac{1}{2}\sum_{t=\lceil T/2 \rceil}^{T} \alpha_t \\ 0 & otherwise \end{cases}$$

where $\alpha_t = \log 1/\beta_t^s$

ITLAdaBoost algorithm uses the same strategy as AdaBoost to update the weight of the incorrectly classified samples. The difference is that ITLAdaBoost adjust the weight of the incorrectly classified samples in the training scenes and unseen scenes using different strategy respectively. AdaBoost algorithm updates the incorrectly classified samples in the same scenes. For each round of iteration, ITLAdaBoost algorithm trains the base classifier on the weighted training and unseen scene data. More specifically, if a sample in the training scene is incorrectly predicted, the sample may likely conflict with the unseen scene data. Then we decrease its weight to reduce its effect through multiplying its weight by

$(\beta_t^a)^{|h_t(x_i)-y_i|}$. If a sample in the unseen scene is mistakenly classified, we increase its weight through multiplying its weight by $(\beta_t^s)^{-|h_t(x_i)-y_i|}$. Thus, in the next round, the misclassified old scene sample will have less effect in the next round and the misclassified new scene sample will have larger weights.

4. Experimental results

4.1. Datasets

The DC and NICTA datasets were used for validation of the proposed method. Since these two datasets are taken from different scenes, the former was used as the training set and the latter was taken as the scenes for testing. The datasets were manually labeled. Pedestrians are marked by rectangular in videos. The dataset was equally split into five subsets. Three of them were used for training and the rest two were used for testing. Each subset consists of 4800 pedestrian samples, which were scaled to the size of 18×36. Some pedestrian samples from the two datasets are shown in Fig. 3. In our work, the experiments were performed on a computer with Intel 1.6 GHz Dual-core CPU and 1 GB RAM.

4.2. Sample screening

The sample screening algorithm based Isomap algorithm was used to preprocess the two dataset. Specially, 100 pedestrian samples in NICTA dataset as the unseen data and 1000 pedestrian samples in the DC dataset were selected. The proportion of size of the two dataset is 1:10. Harr feature was used to represent the pedestrian samples. Due to the large scale of harr features, different feature numbers were selected to test. In this paper, the number of neighbor k in the Isomap algorithm was also test. The specific values of k are (5, 8, 10). We also test different feature numbers (500, 1000, 2000, 5000) for each different k.
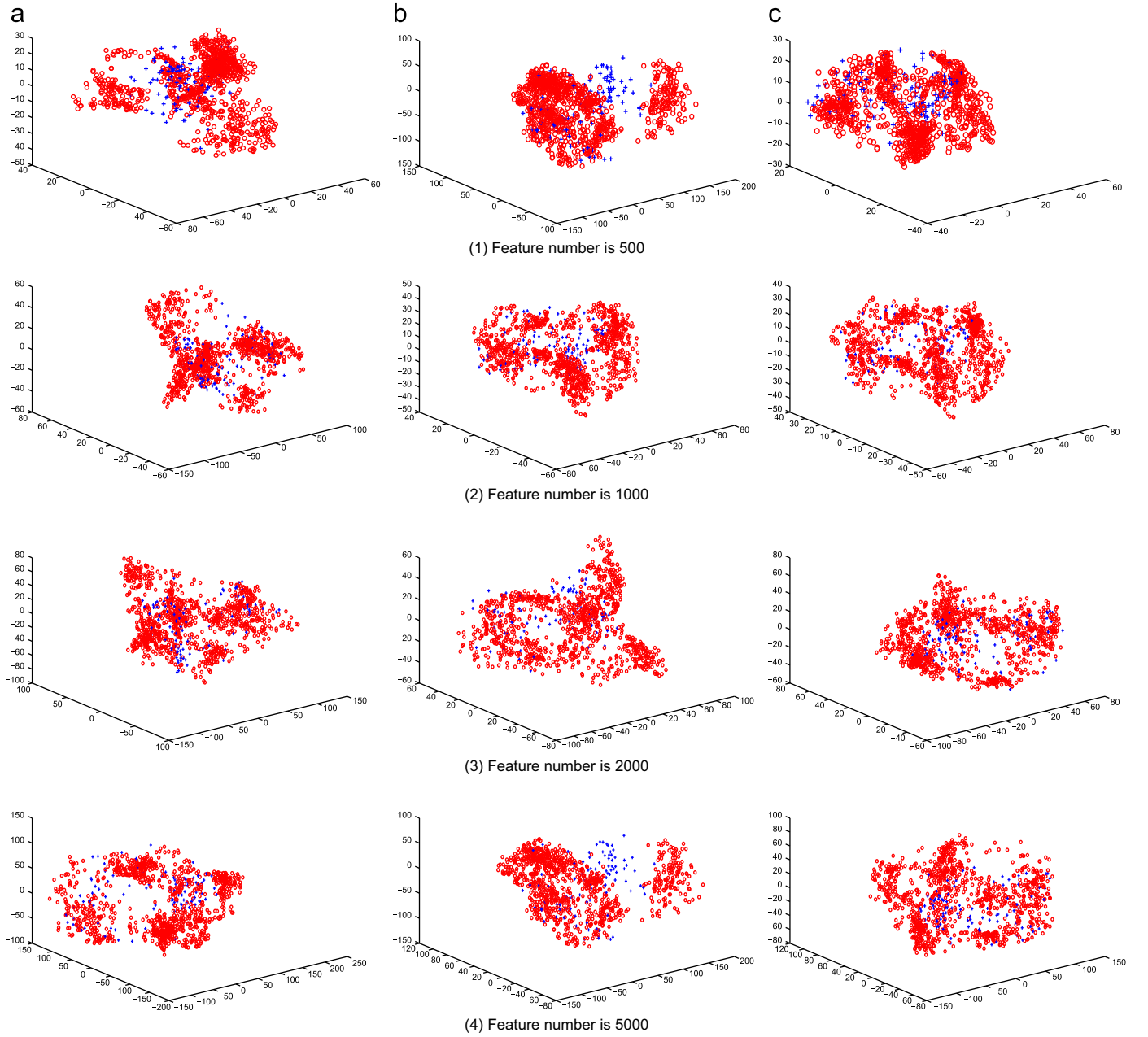
Fig. 4 gives the manifold graph with different feature numbers and neighbor numbers in the DC dataset and the NICTA dataset.

As we can see from above graph, the distribution of DC and NICTA is different obviously, but there exist part of correlation, this is because there are some similar pedestrian samples in the two dataset. "Green cross" was used to mark the samples in the unseen dataset and "red circle" was used to mark the samples in the training dataset. As we can see from the figure in the last group, when $k=8$, the manifold graph can accurately reflect the distribution of the two dataset, samples in the training dataset can separate from the samples in the

Fig. 3. Pedestrian samples from (a) DC dataset and (b) NICTA dataset.

(1) Feature number is 500

(2) Feature number is 1000

(3) Feature number is 2000

(4) Feature number is 5000

Fig. 4. Manifold graph of DC and NICTA dataset with different feature number and *k*. (a) *k*=5, (b) *k*=8 and (c) *k*=10.

unseen dataset and there are several samples overlapping, this is because these samples in the two dataset are basically similar. When *k*=5 and 10, it cannot achieve acceptable result.

Different feature number leads to different result, as we can from the graph, when feature number is 5000, we can achieve the best results.

4.3. Performance comparison

The detection performance was evaluated by detection rate and false positives rate. The performance of three other algorithms has been included for comparison. The first method trains a classifier using AdaBoost with training scenes data D_a. The second method trains a classifier with D_a and D_s. The third method trains a classifier using ITLAdaBoost with training data D_a and D_s.

In this paper, we first propose sample screening algorithm to preprocess the training data, and then construct a high-quality classifier with ITLAdaBoost algorithm. Table 1 gives the detection performance of the three methods and our proposed method. As it can be seen from the table, classifier trained using only the training scenes was not suitable for detecting pedestrians in unseen scenes. That is, AdaBoost algorithm does not work well when training set and test set are in the different distribution. Although the detection rate of method 2 has increased, the auxiliary dataset are much larger than the unseen dataset, so the corresponding classification model is not suitable for pedestrian detection in the unseen scenes. ITLAdaBoost algorithm designs a new strategy for samples in the auxiliary dataset specially and retains the samples, which are similar to the unseen scenes, so the performance has certain improvement. Our method adopt sample screening algorithm based manifold learning to select samples in the auxiliary which suitable for the unseen scenes, and then design ITLAdaBoost algorithm to train a new classification. The experiments show that our method can adapt to pedestrian detection in changing scenes. Fig. 5 gives the ROC curve under different methods and it verifies the effectiveness of the proposed method.

ITLAdaBoost algorithm is an extension of AdaBoost algorithm and is also an iteration algorithm. So the convergence is an important feature of ITLAdaBoost algorithm. Fig. 6 gives the detection rate results of the number of iterations. As we can see from the curve, ITLAdaBoost algorithm converges well but not
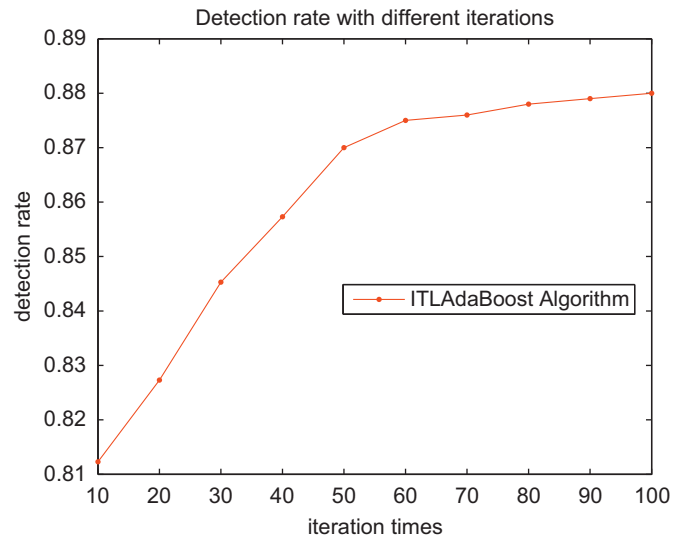
Table 1
The comparison of the detection performance.

Methods	Our method	Method 1	Method 2	Method 3
DR (%)	88	68	75	84
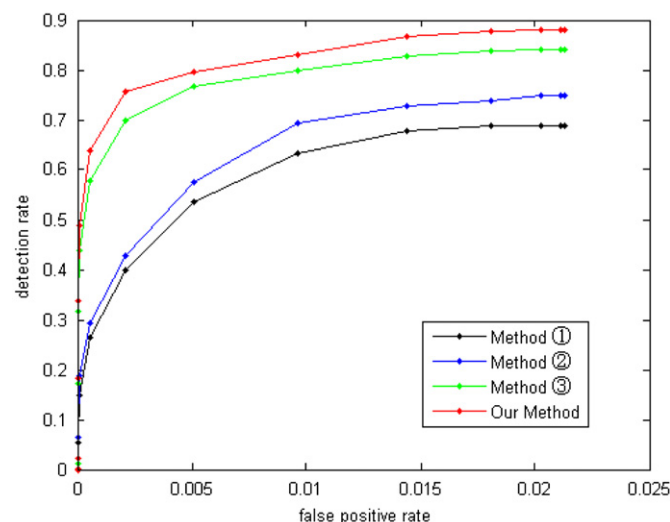
Fig. 5. ROC curves of different methods.

Fig. 6. Detection rate with different iterations.

fast. When ITLAdaBoost algorithm performs 50 iterations, the corresponding detection rate tends to smooth.

5. Conclusions

This paper presented a transfer learning based method for detecting pedestrians in changing scenes. The contributions of this paper include: (1) A new framework for detecting pedestrians with sample screening algorithm based manifold learning and classification based transfer learning. (2) Sample screening algorithm use Isomap algorithm to select useful samples in the training scenes for extending large training set. (3) ITLAdaBoost was employed to solve the problem of training set and testing set coming with different distributions. Our experimental results showed that the proposed method can achieve better performance compared with the traditional methods. In our future work, we will investigate the influence of the amount of the data from the unseen scenes on the performance of detection. Other methods of incorporating the knowledge from changing scenes will also be studied.

References

[1] M. Enzweiler, D.M. Gavrila, Monocular Pedestrian Detection: Survey and Experiments,, IEEE Trans. Pattern Anal. Mach. Intell. 31 (12) (2009) 2179–2195.
[2] G. Overett, L. Petersson, N. Brewer, L. Andersson, N. Pettersson, A New Pedestrian Dataset for Supervised Learning, in: Proceedings of the IEEE Intelligent Vehicles Symposium, 2008, pp. 373–378.
[3] X.B. Cao, H. Qiao, J. Keane, A low-cost pedestrian detection system with a single optical camera,, IEEE Trans. Intell. Trans. Syst. 9 (1) (2008) 58–67.
[4] Y.P. Huang, S. Fu, C. Thompson, Stereovision-based object segmentation for automotive applications,, EURASIP J. Appl. Signal Processing 2005 (14) (2005) 2322–2329.
[5] M. Bertozzi, A. Broggi, A. Lasagni, M. Del Rose, Infrared stereo vision-based pedestrian detection, in: Proceedings of the IEEE International Conference on Intelligent Vehicles Symposium, 2005, pp. 24–29.
[6] D.M. Gavrila, J. Geibel, Shape-based pedestrian detection and tracking, in: Proceedings of the IEEE Intelligent Vehicles Symposium, 2003, pp. 8–14.

[7] G.M.A. Sessler, T. Martoyo, F.K. Jondral, RBF based multiuser detectors for UTRA-TDD, in: Proceedings of the IEEE Vehicular Technology Conference, 2001, pp. 484–486.

[8] F.L. Xu, X. Liu, K. Fujimura, Pedestrian Detection and Tracking with Night Vision, IEEE Transactions on Intelligent Transportation System 6 (2005) 63–71.

[9] S. Maji, A. Berg, J. Malik, Classification using intersection kernels Support Vector Machines is efficient, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2008, pp. 1–8.

[10] Z. Wang, X.B. Cao, Rapid classification based pedestrian detection in changing scenes, in: Proceedings of the IEEE Conference on SMC, 2010, pp. 1591–1596.

[11] D.M. Gavrila, S. Munder, Multi-Cue Pedestrian Detection and Tracking from a Moving Vehicle, Int. J. Comput. Vision 73 (1) (2007) 41–59.

[12] S. Munder, D.M. Gavrila, An Experimental Study on Pedestrian Classification, IEEE Trans. Pattern Anal. Mach. Intell. 28 (11) (2006) 1863–1868.

[13] V.D. Shet, J. Neumann, V. Ramesh, L.S. Davis, Bilattice-Based Logical Reasoning for Human Detection, in: Proceedings of the IEEE Int'l Conf. Computer Vision and Pattern Recognition, 2007, pp. 1–8.

[14] L. Zhang, B. Wu, R. Nevatia, Detection and Tracking of Multiple Humans with Extensive Pose Articulation, in: Proceedings of the Int. Conf. Computer Vision, 2007, pp. 1–8.

[15] C. Papageorgiou, T. Poggio, A Trainable System for Object Detection, Intl. J. Comput. Vision 38 (2000) 15–33.

[16] M. Szarvas, A. Yoshizawa, M. Yamamoto, J. Ogata, Pedestrian Detection with Convolutional Neural Networks, in: Proceedings of the IEEE Intelligent Vehicles Symp., 2005, pp. 223–228.

[17] D.M. Gavrila, Pedestrian detection from a moving vehicle, in: Proceedings of the European conference on Computer Vision, 2000, pp. 37–49.

[18] M. Szarvas, A. Yoshizawa, M. Yamamoto, J. Ogata, Pedestrian detection with convolutional neural networks, in: Proceedings of the IEEE Intelligent Vehicles Symposium, 2005, pp. 224–229.

[19] P. Viola, M. Jones, D. Snow, Detecting pedestrians using patterns of motion and appearance, in: Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2003.

[20] P. Viola, M. Jones, Robust real-time object detection, IJCV 57 (2) (2003) 137–154.

[21] P. Viola, M. Jones, Rapid Object Detection Using a Boosted Cascade of Simple Features, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, 2001, pp. 511–518.

[22] A. Shashua, Y. Gdalyahu, G. Hayun, Pedestrian Detection for Driving Assistance Systems—Single-frame Classification and System Level Performance, in: Proceedings of the IEEE International Conference on Intelligent Vehicles Symposium, 2004, pp.1–6.

[23] P. Sabzmeydani, G. Mori, Detecting Pedestrians by Learning Shapelet Features, in: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, 2007, pp. 1–8.

[24] R. Xu, B. Zhang, Q. Ye, J. Jiao, Cascaded L1-norm Minimization Learning (CLML) Classifier for Human Detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2010, pp.3566–3569.

[25] J. Jiang, C. Zhai, Instance weighting for domain adaptation in NLP, in: Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, 2007, pp. 264–271.

[26] P. Wu, T.G. Dietterich, Improving SVM accuracy by training on auxiliary data sources, in: Proceedings of the 21st International Conference on Machine Learning, 2004, pp. 871–878.

[27] W. Dai, Q. Yang, G. Xue, Y. Yu, Boosting for transfer learning, in: Proceedings of the 24th International Conference on Machine Learning, Corvalis, 2007, pp. 193–200.

[28] X. Liao, Y. Xue, L. Carin, Logistic regression with an auxiliary data source, in: Proceedings of the 21st International Conference on Machine Learning, 2005, pp. 505–512.

[29] M.T. Rosenstein, Z. Marx, L.P. Kaelbling, T.G. Dietterich, To transfer or not to transfer, in: Proceedings of NIPS 2005 Workshop on Inductive Transfer: 10 Years Later, 2005.

[30] C. Wang, S. Mahadevan, Manifold alignment using procrustes analysis, in: Proceedings of the 25th International Conference on Machine learning, 2008, pp. 1120–1127.

[31] J.B. Tenenbaum, V. de Silva, J.C. Langford, A global geometric framework for nonlinear dimensionality reduction,, Science 290 (5500) (2000) 2319–2323.
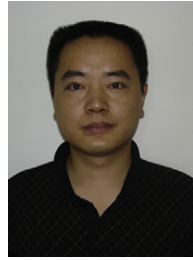
Xianbin Cao received the B.S. degree in computer science and the M.S. degree in information and system from Anhui University, Hefei, China, in 1990 and 1993, respectively, and the Ph.D. degree in intelligent information processing from the University of Science and Technology of China (USTC), Hefei, in 1996. He has been with the USTC since 1996 and became an Associate Professor with the Department of Computer Science and Technology from 1999 to 2005, where he is currently a Professor and the Administrative Director of the Anhui Province Key Laboratory of Software in Computing and Communication, Hefei since 2005. Since 2009 he has also been a professor in the school of electronic and information engineering, Beihang University PR China and is the director of the lab of intelligent transportation system. His current research interests include intelligent transportation systems, airspace transportation management and intelligent computation. He has been publishing more than 100 books, book chapters, and papers in these areas since 1993.

Zhong Wang is a PhD candidate in the Department of Computer Science and Technology, the University of Science and Technology. He received the BS degrees in computer science from AnQing Teachers College, China, in 2006. His research interests include machine learning, computer vision, and pedestrian detection. He is a student member of IEEE.

Pingkun Yan received the B.Eng. degree in electronics engineering and information science from the University of Science and Technology of China, Hefei, China and the Ph.D. degree in electrical and computer engineering from the National University of Singapore, Singapore. He is a full professor with the Center for OPTical IMagery Analysis and Learning (OPTIMAL), State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, Shaanxi, PR China. He was a Senior Member of the Research Staff of Philips Research North America, Briarcliff Manor, NY. Before that, he worked as a Research Associate with the Computer Vision Laboratory, University of Central Florida, Orlando, FL. His research interests include computer vision, pattern recognition, machine learning, and their applications. He is a senior member of the IEEE.

Xuelong Li is a full professor with the Center for OPTical IMagery Analysis and Learning (OPTIMAL), State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, Shaanxi, PR China.