Information Sciences 366 (2016) 177–187

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Vehicle detection from highway satellite images via transfer learning

Liujuan Cao[a,b,*], Cheng Wang[a,b], Jonathan Li[a,b,c]

[a] Fujian Key Laboratory of Sensing and Computing for Smart City, Xiamen, Fujian, 361005, China
[b] School of Information Science and Engineering, Xiamen University, Xiamen, Fujian, 361005, China
[c] Department of Geography and Environmental Management, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

ARTICLE INFO

Article history:
Received 15 June 2014
Revised 1 October 2015
Accepted 2 January 2016
Available online 20 February 2016

Keywords:
Classification
Satellite images
Transfer learning
Sparse coding

ABSTRACT

Coming with the era of highway satellites, nowadays there is a massive amount of remote sensing images captured. Therefore, it is now feasible to detect vehicles directly from these satellite images, which has attracted extensive research attentions for both academic and industrial applications. However, it is not an easy task at all, mainly due to the difficulty to obtain training data to train vehicle detectors. On the contrary, there has been sufficient amount of labeled information regarding vehicle regions in the domain of aerial images. In this paper, we study the problem of detecting vehicles on highway satellite images, without the time-consuming step of collecting sufficient training data in this domain. Our key idea is to adopt a novel transfer learning technology that transfers vehicle detectors trained in the aerial image domain to the satellite image domain. In doing so, several cutting-edge vehicle detection algorithms can be directly applied. More specifically, our transfer learning scheme is based on a supervised super-resolution algorithm, which learns mutually correlative sparse coefficients between high and low resolution image patches. Then, a linear SVM based detector is trained, the loss function of which is integrated into the sparse coding operation above. Experimental results have shown that the proposed framework can achieve significant improvement over several alternative and state-of-the-art schemes, with high precision and low false alarms.

1. Introduction

Coming with the era of highway satellites and high-resolution imaging techniques, nowadays traffic data can be quickly acquired by using satellite images. Comparing to the traditional method of traffic monitoring that relies on surveillance images or aerial images, traffic monitoring using satellite images merit in its high convenience, low cost, as well as good safety. And the resolution of satellite imagery has been significantly improved, for instance many high-resolution satellites provide 0.5–1 m resolution panchromatic images, such as IKONOS (1m), QuickBird(0.61 m), WordView(0.5 m), etc. Although it is still not comparable to aerial images or surveillance images, the resolution of satellite images is already sufficient for traffic monitoring. Therefore, traffic monitoring using highway commercial satellite images has become a research hot spot in recent years. Among various tasks of traffic monitoring, vehicle detection is one of the core challenges, which has several real-world applications such as battlefield directing and intelligent transportation, etc.

* Corresponding author at: School of Information Science and Engineering, Xiamen University, China. Tel.: +86 15060727861.
E-mail address: caoliujuan@xmu.edu.cn (L. Cao).

http://dx.doi.org/10.1016/j.ins.2016.01.004
0020-0255/

In the literature, vehicle detection from aerial images has attracted extensive research focus in the past years. Different from that of highway satellite images, aerial images are typically with a spatial resolution of 0.35 m or less [4,11,20,25,26], which contains more details on vehicle appearances and backgrounds. In this resolution level, a wealth of robust cutting-edge detectors can be directly used for vehicle detection with high accuracy, meanwhile there are rich training instances and datasets available in the literature. For example, the works in [1,3] focus on detecting moving cars by using airborne optical sensors. The work in [12] focuses on detecting stilling or parking vehicles by supervised sliding window search. In general, the existing approaches for vehicle detection in aerial images can be categorized according to the features used in the methods, i.e.:

- Explicit model [10,12,19], which clusters similar pixels into potential vehicles. Such method usually describes a vehicle as a box, and adopts a top-down matching scheme to find the best-matched candidate in the satellite image. For instance, Hinz [10] proposed a 3D car model based on significant edges. Holt et al. [12] and Lenhard et al. [19] adopt object detection schemes to train car detector.
- Implicit model [15,16], which extracts intensity or texture features surrounding each pixel to implicitly model the vehicle, for instance surrounding contour [16] or histograms of oriented gradients [15]. In the implicit model, a vehicle is usually described by using wire-frame representation. And the detection is performed by checking features surrounding the target region.

To train a robust vehicle detector with high accuracy, the existing works typically needs high-resolution aerial images together with sufficient amount of training samples. The former ensures sufficient details in modeling the appearance of vehicles, while the latter ensures sufficient samples in detector training. However, this is a big challenge when facing low-resolution images captured from highway satellites. One limitation is the resolution, i.e., panchromatic band resolutions of images captured from highway satellites are presently in the range of 0.41–1.0 m. Another limitation is the number of training instance in that resolution. To the best of our knowledge, Correspondingly, limited work has been proposed in the literature to study vehicle detection in low-resolution satellite images [8,14,24,28]. And the above works mainly focus on detecting sparse vehicles on roads.

For a brief review, recently an automatic vehicle detection method was proposed in [6], which can be based for panchromatic images, multi-spectral image of QuickBird satellite, as well as road network data. Leitloff et al. [18] presented an automatic vehicle detection method for satellite images, which adopts an adaptive detector learning scheme by using Haar-like features. He et al. [9] presented a supervised classification and thresholding method to extract traffic information in high resolution satellite images. Mantrawadi et al. [21] proposed an object identification algorithm for high-resolution satellite images based on data mining and knowledge extraction. More recently, Chen et al. [2] proposed a deep learning based vehicle detector.

In this paper, we propose a robust vehicle detection algorithm for low-resolution highway satellite images, which handles the challenges of both low-resolution and few training samples. Our key innovation is to transfer this problem to the high-resolution aerial image domain, where robust and cutting-edge detectors can be learned. To this end, we proposed a supervised transfer learning scheme which consists of three steps as follows:

- A super resolution framework to transfer the detection task in low-resolution satellite image domain to high-resolution aerial image domain.
- A sparse-coding based reconstruction algorithm that integrates classifier training into the super resolution process, which makes the transferred patches in high-resolution more discriminative to train vehicle detectors.
- A robust vehicle detection via linear SVM based search that supports large-scale parallel.

The output of our framework is a robust vehicle detector run on the low-resolution satellite image domain. Such detector has achieved sufficient accurate that has the potential to be directly used for real-world applications of traffic analysis and road surveillance in military and transportation systems.

The rest of this paper is organized as follows: In Section 2, we introduce the proposed super resolution based transfer learning framework. Then, Section 3demonstrates and analyzes the experimental results. Finally we conclude this paper in Section 4 and discuss our future work.

2. The super-resolution detector transfer framework

2.1. The proposed framework

We first briefly review the overall flowchart of the proposed framework for super-resolution detector transfer.

In preliminary, we assume that vector maps are available to assist the extraction of road area. Nevertheless, other complex algorithms for road detection and segmentation can be deployed without the above assumption. And the step of road detection is not the core contribution of this paper. Without loss of generality, we first align the vector map with the satellite images by their GPS locations. Then only the road areas are extracted, on which the vehicle detectors are run. In such a way, “background” inference such as buildings is removed from the subsequent operations, which largely reduces the false alarm as well as improving the recognition speed.

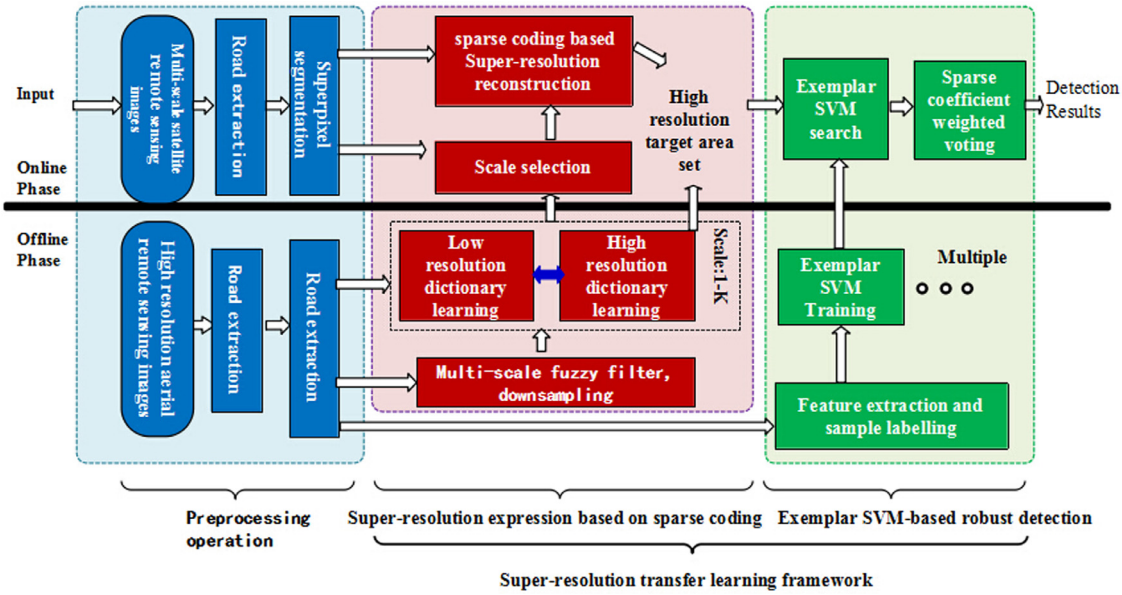

Fig. 1. The proposed transfer learning framework based on super-resolution reconstruction for vehicle detection.

In particular, we adopt the ArcGIS Engine to calibrate vector map and remote sensing image, with the following operations:

- Take the coordinates of remote sensing image as the benchmark, transform the road map according to its coordinates, and align the scales and directions of both images and road maps.
- Load the remote sensing image into the projection coordinate system, which subsequently connects the image with the road information to conduct projection transformation. Therefore, the road data in the database and the images are seamlessly calibrated.
- Compare road element in vector map database with the coordinates of the remote sensing image, and select the road element in line with area, remote the rest areas.

After the above steps, we only carry out vehicle detection on the extracted road areas, with the assumption that such areas should be the only candidates where the vehicles reside. More specifically, we decide the buffer size by the radius of given object space. For a given object A, the buffer size of which can be defined as follows:

$$P = \{x \mid d(x, A) \leq r\} \tag{1}$$

where d is the Euclidean distance, r is the field radius.

The subsequent step is to carry out a K-Means based superpixel segmentation. More specifically, for an image with 4000×6000 pixels, we set the number of clusters as 400, which ensures that the superpixel area is approximately the same as the vehicle size.

After the above preprocessing, we present the super resolution based transfer learning framework, which can be shown in Fig. 1. For a brief review, the proposed super-resolution based transfer learning framework broadly relates to transductive transfer learning [13,22]. One representative work in computer vision comes from [17], which propagates segmentations over ImageNet in a well-controlled manner. For another instance, Rohrbach et al. [23] presented a novel transductive transfer for ImageNet classification with limited training examples. In text, Zhao et al. [27] proposed a crowdsourcing classification scheme for tweets. In this paper, we adopt a new perspective, i.e., a super-resolution based transfer, which handles the challenges (vehicle detection in low-resolution satellite images) into a domain that can be easily handled (vehicle detection in high-resolution aerial images). The details can be described as follows:

In the offline step, we collect large-scale high-resolution remote sensing images and label vehicle rectangles on the road manually. Then, we train vehicle detectors upon such high-resolution labeled instances. We further down-sample the above high-resolution images into their corresponding low-resolution, upon which train sparse coding based super-resolution operator. This operator serves as a "bridge" to link the target-of-interest to be detected in the low-resolution domain and the detectors residing on the high-resolution domain.

In the online step, given a low-resolution target-of-interest region obtained from road extraction and superpixel segmentation, our goal is to decide whether this is a vehicle or not. To this end, we first carry out super-resolution reconstruction based on sparse coding, which transfers the detection problem to the high-resolution domain. This step outputs the "augmented" features on which the vehicle detector is run to determine whether the target region is a vehicle.

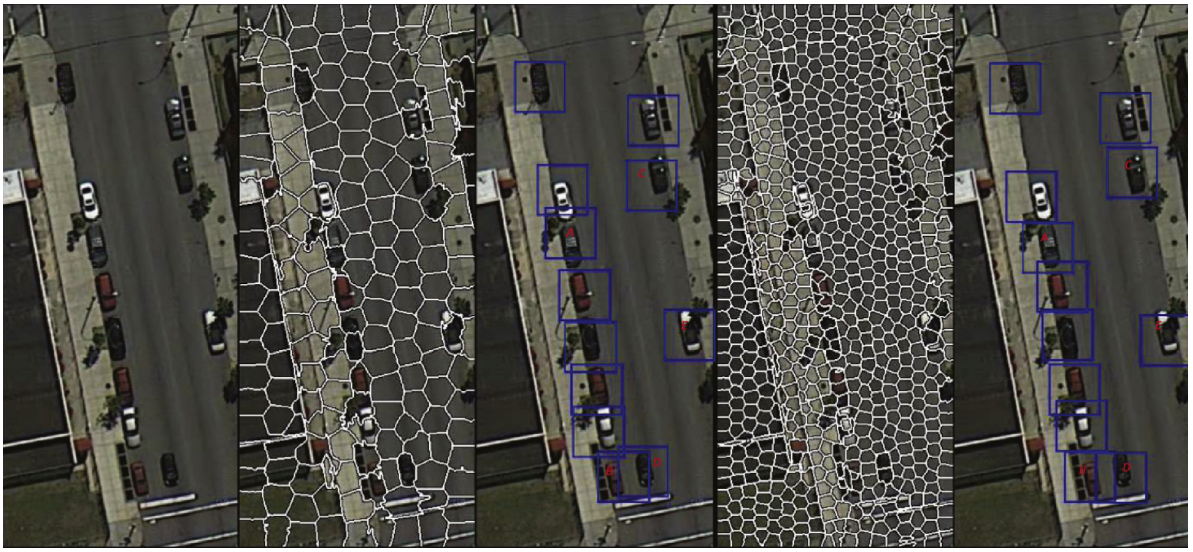

Fig. 2. The first column of this figure is the raw image. In the second column, we show the resulting regions after SLIC based segmentation. Here, we set the number of superpixels n so that the superpixel regions are approximately segmented as a size of 30 × 30. In the third column, the blue rectangles are the detection results that are likely to contain vehicles from regions obtained via SLIC based segmentation. The fourth column is the result after SLIC with n equal to setting the images area as 15 × 15. The fifth column is the 12 highest responded regions after running vehicle detector in this images. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

More specifically, given the kth target detection result of super-resolution reconstruction area $Score_{high}{}^{k}$, we combine $Score_{high}{}^{k}$ with the super-resolution reconstruction coefficients w_k of this region as a weight. We then calculate the overall detection score as follows:

$$Score_{low} = \sum_{k=0}^{K} w_k Score_{high}{}^{k} \tag{2}$$

where K is the number of all non-zero reconstruction coefficients and $Score_{high}{}^{k}$ is the kth voting result of the classifier, i.e., SVM, based outputs [1]. This score is with the range of [0, 1]. w_k is the non-zero reconstruction coefficients of the high-resolution superpixel to the current low-resolution superpixel. And finally, we use a threshold obtained by cross validation to determine whether the target-of-interest is vehicle or not. The calculation of w_k and $Score_{high}{}^{k}$ will be introduced in depth in the following sections.

2.2. SLIC superpixel

In this section, we introduce the details of our superpixel segmentation algorithm run on both low-resolution and high-resolution images. One consideration here is to reduce the computational cost, as we replace the sliding window based scanning to superpixel-level scanning. We assume that, after segmentation, each superpixel should be homogeneous, a.k.a., uniformly contains either vehicle or not. In particular, we adopt the SLIC based segmentation proposed by Achanta et al. to achieve this goal. It can quickly convert a given satellite image into a set of homogeneous regions, which typically with regular shapes and sizes. It is worth to note that the computing speed of running SLIC is extremely fast.

More specifically, we segment a given satellite image into n superpixels. For a fixed-sized image, the larger n means the larger number of the patches and detection boxes, which will increase the detection speed and is closer to the sliding window based approaches. On the other hand, a small n refers to the less number of patches and detection boxes, which will result in the case that some vehicle regions are missed after detection as they are mixed with backgrounds in the large bounding box. We therefore tune n as a parameter to determine the best scale for the superpixel size. We have found that for different image resolutions, the best scale is also different. For instance, as shown in Fig. 2, *A*, *B*, and *C* rectangles are not located very well in column. However, as shown in Column 5, the vehicle regions can be better located. Meanwhile, in Column 5, *D* and *E* rectangles are not located better than they do in Column 3. by combing both scales in SLIC, we can obtain a much better segmentation results for all vehicles. Note that the size of each patch in Fig. 2 is 60 × 60, which is just enough to contain a vehicle at an astonishing ground sampling distance of 0.127 m.

[1] Note that we adopt its classification probability, or equivalently its regression score, rather than its binary labels in the above decision.

2.3. Sparse coding based super-resolution

For a given superpixel region in the low-resolution satellite images to be detected, a super-resolution algorithm is first proposed to "transfer" this region into high-resolution, which can "bypass" the detection limitation in low-resolution satellite images. In the most recent literature, the mainstream algorithm for image super-resolution is the learning-based super-resolution construction. Its basic idea is to reconstruct low-resolution superpixels by selecting high-resolution superpixels having the similar appearance, with a set of optimized combination weights learned. However, due to the limitation in selecting similar image patches, the matching performance is typically poor. In order to address the above issue, in this paper we present a collaborative dictionary learning scheme using sparse coding, which reconstructs low-resolution patches using its most similar high-resolution patches with shared coefficients obtained via joint dictionary learning. The main idea is to collaborate the sparse coding process for both high- and low- resolution image patches during dictionary based quantization. Sparse links are then established better high- and low-resolution image patches. In such a way, appearance noise and background clutters can be largely reduced. We introduce the details of our algorithm as follows:

(*1) Cross-correlation sparse dictionary learning:*

The proposed method adopts a learning-based sparse coding scheme, which has stronger adaptive ability and can get more accurate representation. To acquire the training data set, we first obtain the superpixel set $X^h = \{x_1, x_2, \ldots, x_n\}$ from large-scale high-resolution remote sensing image. Correspondingly, we also obtain large-scale high-low resolution training samples via the following Equation 3 via downsampling:

$$Y^l = SHX^h \tag{3}$$

where S denotes the downsampling operation, H denotes the operation of fuzzy filtering, $Y^l = \{y_1, y_2, \ldots, y_n\}$ denotes the obtained low-resolution superpixels. By carrying out the above equation for fuzzy and downsampling for high-resolution aerial images, we obtain a set of high-low resolution superpixel point pairs $P = \{X^h, Y^l\}$.

We then leverage the learned sparse association coefficients from the above sparse coding procedure, based on which we define the training sample pairs as $P = \{X^h, Y^l\}$. P is composed of the extracted superpixels from the high-resolution satellite images as well as their down-sampling results. Each training sample pair $\{x_i{}^h,\ y_i{}^l\}$ is represented by a column vector of image feature block, where $x_i{}^h$ represents the superpixel patches of high-resolution image, and $y_i{}^l$ represents the superpixel patches of low-resolution image.

Sparse coding targets at learning P using the dictionary, by which the high- and low-resolution image patches are aligned. To this end, we adopt the following equation to achieve sparse dictionary learning for high- and low- resolution image patches, respectively:

$$\begin{aligned} V_h &= \arg \underbrace{\min}_{\{V_h, U\}} \|X^h - UV_h\|_F{}^2 + |\lambda| \|U\|_1 \\ V_l &= \arg \underbrace{\min}_{\{V_l, U\}} \|X^l - UV_l\|_F{}^2 + |\lambda| \|U\|_1 \end{aligned} \tag{4}$$

where U denotes the coefficients from sparse representation. V_h and V_l are the learned sparse dictionary for high- and low-resolution patches respectively.

We then combine the learning of both high- and low-resolution dictionaries, as forcing both dictionaries to have the same sparse expression. We design the corresponding learning equation as:

$$\underbrace{\min}_{\{V_h, V_l, U\}} \frac{1}{N}\|X^h - UV_h\|_F{}^2 + \frac{1}{M}\|Y^L - UV_L\|_F{}^2 + \lambda\left(\frac{1}{N} + \frac{1}{M}\right)\|U\|_1 \tag{5}$$

where N and M are the dimensions of image feature vectors for both high- and low-resolution image regions respectively. As coefficients in the above equation, such coefficients should balance the effect of the overall sparse coding framework between low- and high-resolution image regions. Therefore, the above formulation can be further simplified as:

$$\begin{aligned} &\underbrace{\min}_{\{V_h, V_l, U\}} \|X_C - UV_C\|_F{}^2 + \lambda\left(\frac{1}{N} + \frac{1}{M}\right)\|U\|_1 \\ &s.t. \quad X_C = \begin{bmatrix} \frac{1}{\sqrt{N}} FX^h \\ \frac{1}{\sqrt{M}} FY^l \end{bmatrix} \quad V_C = \begin{bmatrix} \frac{1}{\sqrt{N}} V^h \\ \frac{1}{\sqrt{M}} V^l \end{bmatrix} \end{aligned} \tag{6}$$

To learn an optimized coupled dictionary, an iterative optimization scheme is further adopted in this paper. In particular, we carry out such an interactive optimization between the following two steps:

- Fix the sparse coefficients and learn both low- and high-resolution dictionaries.
- Fix both dictionaries and learn the joint (shared) sparse coefficients.

The above process is done iteratively until both are converged. This step outputs two over-complete dictionaries, i.e., V_h and V_l, respectively.

(2) *Super-resolution reconstruction based on sparse coding:*

In the online detection phase, given a low-resolution target superpixel, the following sparse coding model is adopted to construct super-resolution representation:

- First, for the input low-resolution superpixel patch y, we estimate its sparse coding a according to the learned low-resolution dictionary V_l. It can be done as follows:

$$\min \|\alpha\|_0 \quad s.t. \|V_l\alpha - Fy\|_2^2 \ll \omega \tag{7}$$

- Second, we further convert the above L_0 optimization problem to the problem of minimizing L_1 norm, formulated as:

$$\min \|\alpha\|_1 \quad s.t. \|V_l\alpha - Fy\|_2^2 \ll \omega \tag{8}$$

  We then share the obtained low-resolution sparse coefficients a, which are directly used in combination with the high-resolution dictionary V_h via iterative learning. This step outputs the cross-correlation high-resolution superpixel patch set Y^S.
- Third, based on this high-resolution superpixel patch set Y^S, we further combine with the linear SVM learned offline to determine whether the target superpixel region is vehicle or not.

In the offline process, given a set of high-resolution aerial images as training samples, the proposed algorithm first carry out an unsupervised learning to learn both high- and low-resolution dictionaries respectively. In the online process, for a given low-resolution satellite image, we first determine its scale, a.k.a., resolution, upon which we carry out the sparse coding on its corresponding scale.

2.4. Training linear SVM detector in high-resolution

Based on the super-resolution reconstruction results, we can obtain a set of a high-resolution image patches that are most relevant to the low-resolution aerial image patches, as well as the reconstruction coefficients using the coupled dictionaries and sparse coding algorithm. In this paper, we do not investigate the details of using complex detectors to detect vehicles in the high-resolution aerial image domain. Instead, we resort to the most commonly-used detector/classifier, i.e., linear SVM, to train our vehicle detector [2]. Note that in such a case, the classifier should be written as a linear model to suit in the detector loss as introduced in the super-solution based transfer learning.

More specifically, given a set of labeled vehicle rectangles, we treat them as positive examples to train a set of linear SVMs by using the libsvm toolbox. We then carry out a calibration step over this set of linear SVMs using the validation set, which results in a set of weights reflecting the confidence of the detectors. For the given test aerial image, once detection scores of different superpixel regions are obtained, a non maxima suppression is further adopted to produce a non redundant detection results without spatial overlapping. The overlapping area between the detection result (rectangle) and the ground truth rectangle is calculated to evaluate the detection accuracy, which follows the PASCAL VOC operation mode. Test results overlapping with real standard of more than 0.5 will be classified as a positive example, and test results overlapping with real standards below 0.2 will be classified as a negative example.

It is worth to note that such scores are fitted with a logistic function to unify their distribution. The proposed calibration step can be interpreted as a simple adjusting and moving operation for decision boundary. In other words, through moving the decision boundary towards the direction of the sample, the SVMs worse performed will be suppressed, and by moving the decision boundary away from the direction of the sample, the SVMs better performed will be promoted.

Although the resulting decision boundary is not optimal for local SVM, the comparison precision will be greatly improved between different samples. Formally speaking, given a test example x and a set of learned sigmoid parameters (α_E, β_E), the calibration test score of sample E is:

$$f(x|w_E, \alpha_E, \beta_E) = \frac{1}{1 + e^{-\alpha_E(w_E{}^T x - \beta_E)}} \tag{9}$$

In online testing, we adopt a threshold value of the original SVM score -1 (negative border) to adjust the Sigmoid parameter for each classifier. Given the super-resolution reconstruction results, we can obtain the high resolution superpixels (multiple) that are used to reconstruct the target low-resolution area, with which we can conduct binary detection.

3. Experiments

In this section, we present in depth our quantitative evaluation. All experiments are done on Visual C++ 10.0 developing environment with OpenCV.

3.1. Road map extraction

For the step of road map extraction, a MapInfo/Tab format 2D-vector map and ArcGIS Engine is used to test the performance of our scheme. Fig. 3 shows some exemplar results of the calibration operation between the satellite/aerial images

[2] Without loss of generality, other complicated detectors can be also deployed, which is orthogonal to the contribution of this paper.

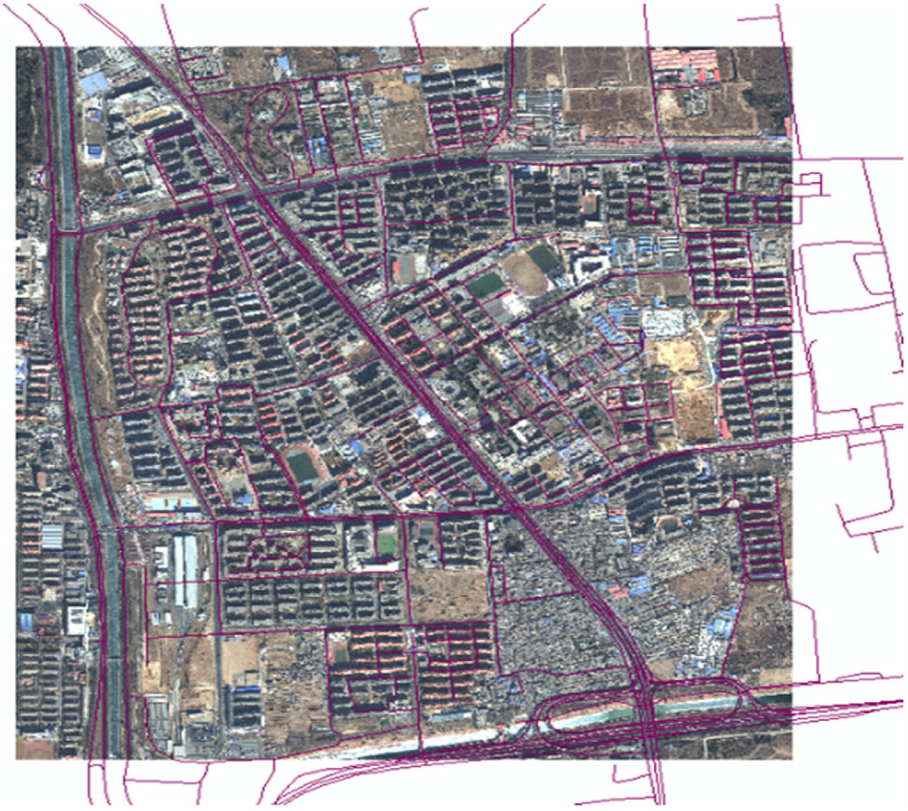

Fig. 3. Registration result of remote sensing image and road vector information.

and the vector maps. In addition, Figs. 3(a) and 4(b) show some registration result of remote sensing images and road vectors.

Fig. 5 shows several groups of the superpixel segmentation results. From Fig. 5, we can see that the edge information of the vehicle can be well preserved, which means in this case vehicles can be recognized much easier.

Fig. 5 shows one original vector map and 3 watermarked vector maps. By comparing the same part of each map (zoomed out), we can see that the watermarked vector maps (with iteratively embedding) have more smooth appearance, which can induce lower shape distortions and improve the detection accuracy of the proposed scheme. And with the increasing of the iterative embedding times, the capacity also increase accordingly, thus the proposed scheme has a better performance on perception invisibility. In addition, Fig. 5 shows the watermark images that iteratively extracted twice from Fig. 5(b), five times from Fig. 5(c) and ten times from Fig. 5(d), respectively.

3.2. Data collection and ground truth labeling

We test our algorithm on both satellite and aerial image datasets. To build the satellite image dataset, we collect 80 satellite images from Google Earth (©2009 Google https://earth.google.com/), in which each image is with 979 × 1348 resolution, covering the road maps in New York City. Correspondingly, we further collect 80 corresponding aerial images covering the same road map of New York City by zooming in the Google Earth into the finest resolution. We ask a group of volunteers to manually labeled vehicle regions with both low- and high-resolution images collected above, resulting in in total 1482 vehicle annotations. Fig. 6 shows several groups of vehicle rectangles, which are quite in visual appearance, imaging conditions, and camera viewing angles.

3.3. Baselines and evaluation protocols

From the aerial images, we adopt an 1:5 leave-one-out (training:cross-validation) scheme for parameter tuning. Note that labels from the low-resolution satellite images are used for validation purpose only. The accuracy of the proposed scheme is tested by using precision-recall curves and mean classification error.

We compare the proposed scheme to a serial of baselines and state-of-the-art approaches, including: (1) ***k*NN-source**: It adopts *k*NN to search most similar superpixels and assign labels by majority voting, which is operated on the source

Fig. 4. Road information extraction.

Fig. 5. The result of superpixel segmentation.

(low-resolution) domain. (2) **Linear SVM-source**: It adopts SVM with linear kernel for detection and operates on the source domain. (3) ***k*NN-target**: It adopts *k*NN to search most similar superpixels and assign labels by majority voting, which is operated on the target (high-resolution) domain. (4) **Linear SVM-target**: It adopts SVM with linear kernel for detection and operates on the target domain. It is the final method adopted in this paper.

For all above approaches, HoG based descriptors [5] are adopted to extract features[3].

3.4. Preliminary settings

In preliminary, both satellite and aerial images are processed by applying a low-pass filter to remove noises. We then extract road maps from both satellite and aerial images. More specifically, a MapInfo/SHP format 2D-vector map and ArcGIS Engine(https://www.arcgis.com/) is used to align vector maps and satellite (and aerial) images. Then, the road maps are extracted from images regions that coincide on roads of vector maps. Specially, active shape model is adopted to extract precise boundaries.

[3] It is quite clear that, even with the current "high-resolution" vehicle regions, the superpixels are still quite small, which by nature hesitates complex and premature detection models such as Deformable Part-based Model [7] to be deployed. Further, more complex features can be further adopted, which is orthogonal to the contribution of this paper.

Fig. 6. Examples of vehicles extracted from low (left) and high (right) resolution imageries.

Subsequently, superpixel segmentation is adopted to segment the roads into superpixels, on which the vehicle detector is run. Fig. 5 shows the superpixel segmentation results. We tune a best segmentation scale to ensure that the size of superpixel is approximately the size of vehicles.

3.5. Parameter tuning

We have identified and tuned the following parameters that affect the performance of the proposed super-resolution transfer algorithm:

- *Dictionary size*: The proposed coupled dictionary learning is affected by the dictionary size. As shown in Fig. 7, we tune to seek the best size by using the validation set.

3.6. Quantitative analysis

As in Fig. 8, the Precision-Recall curves have demonstrated that our approach achieves consistent and promising performance comparing to the five baselines as introduced above. Especially, there is a significant performance boosting by transferring from the source (low-resolution) domain to the target (high-resolution) domain. And another performance gain can be clearly observed by replacing *k*NN and Linear SVM based classifiers with the proposed E-SVMs. The proposed learning-to-rank based classifier calibration further push the Precision-Recall curves to the best one as evaluated in our entire experiment.

We parallelize individual classifiers of each positive examples using 10 servers each with Intel(R) i5 3.20 GHz cpu, which has achieved approximately 7–8 times of speedup overall. It requires on average 2.02 h. in learning dictionary, 28 m. in training E-SVMs, and 167 s. in calibration. It is not a linear speedup, as data communications and ranking score aggregations are required in a sequential manner.

4. Conclusion

In this paper, we study the problem of vehicle detection in low-resolution satellite imagery. Two critical challenges are there, i.e., limited training examples, as well as the difficulty to train robust detectors directly on the satellite domain. To this end, we contribute in the following aspects: First, we propose to transfer the detection problem into high-resolution aerial imagery, which is based on a supervised super-solution algorithm with coupled dictionary learning. Second, in the targeted domain, a simple detection scheme based on linear SVMs is proposed, which is very efficient while moderately accurate to cope with the domain transfer variations. We have tested the proposed scheme extensively with comparisons to a serial of existing and alternative approaches. Significant performance gains have been reported.

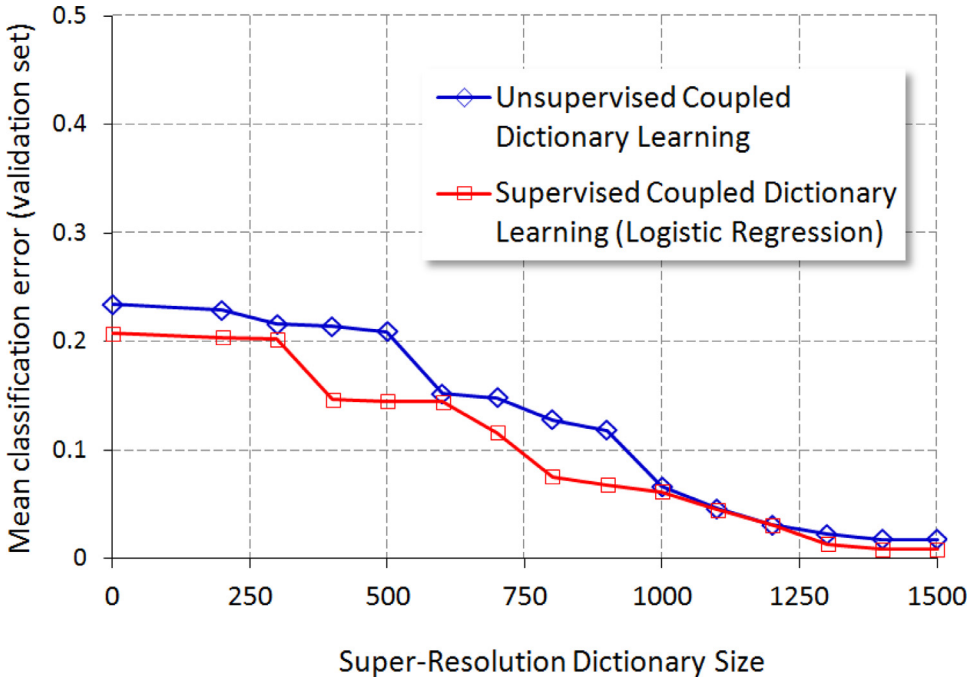

Fig. 7. Parameter tuning on the dictionary size.

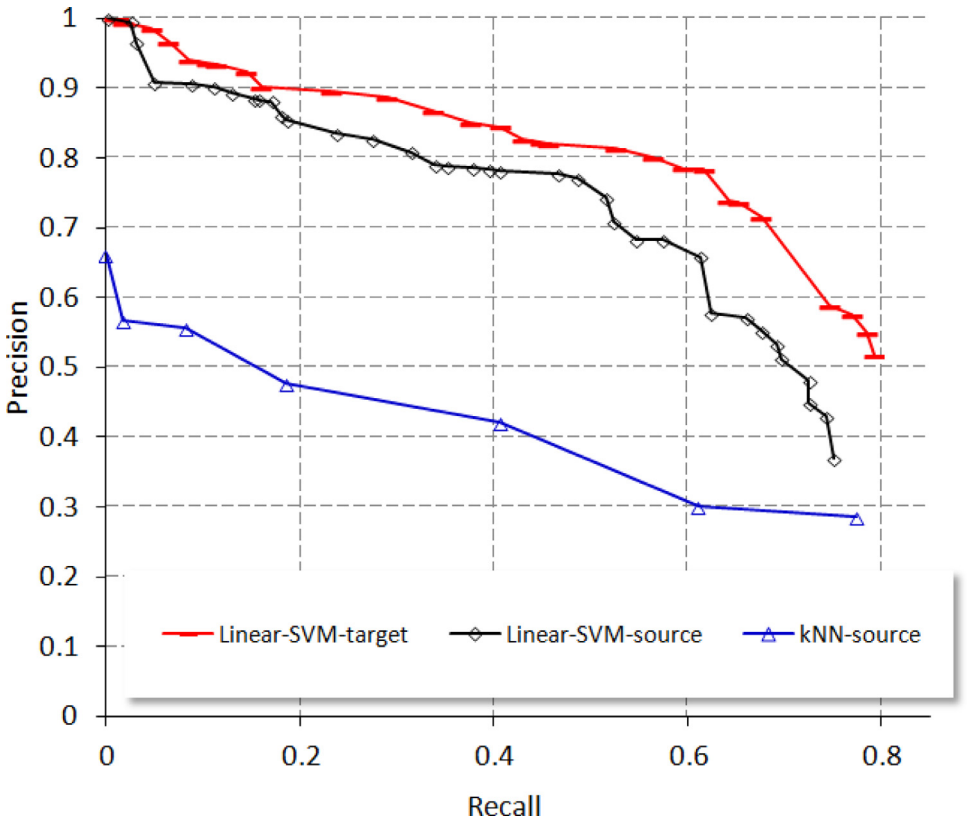

Fig. 8. Quantitative comparisons of PR curves between the proposed algorithm and alternatives.

In our future work, we will pay further attention on handling the multi-scale issues, i.e., the detected windows can be further extended into a pyramid matching setting to ensure finding vehicles using satellites with various spatial resolutions. In addition, it would be beneficial to break the restriction that the vehicle has to be on the road map, which is, in our current scenario, a practical setting which has significantly reduce our computational complexity. And finally, it would be beneficial to incorporate labels from satellite domain (while current very limited), if any, to assist the training of detectors and super-resolution based transfer.

Acknowledgements

This work is supported by the Special Fund for Earthquake Research in the Public Interest No. 201508025, the Nature Science Foundation of China (No. 61402388, No. 61422210 and No. 61373076), the Fundamental Research Funds for the Central Universities (No. 20720150080 and No. 2013121026), the CCF-Tencent Open Research Fund, and the Open Projects Program of National Laboratory of Pattern Recognition.

References

[1] X. Cao, R. Lin, P. Yan, X. Li, Visual attention accelerated vehicle detection in low-altitude airborne video of urban environment, IEEE Trans. Circuits Syst. Video Technol. 22 (3) (2012) 366–378.
[2] X. Chen, S. Xiang, C.-L. Liu, C.-H. Pan, Vehicle detection in satellite images by parallel deep convolutional neural networks, in: Proceedings of the IAPR Asian Conference on Pattern Recognition, 2013, pp. 181–185.
[3] H.-Y. Cheng, C. Weng, Y.-Y. Chen, Vehicle detection in aerial surveillance using dynamic Bayesian networks, IEEE Trans. Image Process. 21 (4) (2012) 2152–2159.
[4] J.-Y. Choi, Y.-K. Yang, Vehicle detection from aerial image using local shape information, Adv. Image Video Technol. 5414 (2009) 227–236.
[5] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005, 1, IEEE, 2005.
[6] L. Eikvil, L. Aurdal, H. Koren, Classification-based vehicle detection in high resolution satellite images, ISPRS J. Photogramm. Remote Sens. 64 (2009) 65–72.
[7] P.F. Felzenszwalb, et al., Object detection with discriminatively trained part-based models, in: Proceedings of the IEEE Transactions on 32.9 Pattern Analysis and Machine Intelligence,, 2010, pp. 1627–1645.
[8] A. Gerhardinger, D. Ehrlich, M. Pesaresi, Vehicles detection from very high resolution satellite imagery, Int. Arch. Photogramm. Remote Sens. 36 (3) (2005) 2795–2806.
[9] X.F. He, L.Q. Zhou, J. Li, Extraction of traffic information in high resolution satellite images, Urb. Geotech. Investig. Surv. 3 (2011) 49–51.
[10] S. Hinz, Detection of vehicles and vehicle queues in high resolution aerial images, Photogramm. Fernerkund. Geoinf. 3 (4) (2004) 201–213.
[11] S. Hinz, A. Baumgartner, Vehicle detection in aerial images using generic features, grouping, and context, in: Proceedings of the DAGEM Symposium, 2191, 2001, pp. 45–52.
[12] A.C. Holt, E.Y.W. Seto, T. Rivard, G. Peng, Object-based detection and classification of vehicles from high-resolution aerial photography, Photogramm. Eng. Remote Sens. 75 (7) (2009) 871–880.
[13] S. Jethro, S. Coupland, Fuzzy transfer learning: Methodology and application, Inf. Sci. 293 (2015) 59–79.
[14] X. Jin, H.D. Curt, Vehicle detection from high-resolution satellite imagery using morphological shared-weight neural networks, Image Vis. Comput. 25 (2007) 1422–1431.
[15] A. Kembhavi, D. Harwood, L. Davis, Vehicle detection using partial least squares, IEEE Trans. Pattern Anal. Mach. Intell. 33 (6) (2011) 1250–1265.
[16] K. Kozempel, R. Reulke, in: U. Stilla, F. Rottensteiner, N. Paparoditis (Eds.), Fast vehicle detection and tracking in aerial image bursts, 38, 2009, pp. 175–180.
[17] D. Kuettel, M. Guillaumin, V. Ferrari, Segmentation propagation in imagenet, in: Proceedings of the European Conference on Computer Vision, 2012, pp. 459–473.
[18] J. Leitloff, S. Hinz, U. Stilla, Vehicle detection in very high resolution satellite images of city areas, IEEE Trans. Geosci. Remote Sens. 48 (2010) 2795–2806.
[19] D. Lenhart, S. Hinz, J. Leitloff, U. Stilla, Automatic traffic monitoring based on aerial image sequences, Pattern Recognit. Image Anal. 18 (3) (2008) 400–405.
[20] R. Lin, X. Cao, Y. Xu, C. Wu, H. Qiao, Airborne moving vehicle detection for video surveillance of urban traffic, IET Comput.Vis. 2 (1) (2008) 1–12.
[21] N. Mantrawadi, M. Nijim, Y. Lee, Object identification and classification in a high resolution satellite data using data mining techniques for knowledge extraction, in: Proceedings of the IEEE International Systems Conference, 2013, pp. 750–755.
[22] S. Pan, Q. Yang, A survey on transfer learning, IEEE Trans. Knowl. Data Eng. 22 (10) (2010) 1345–1359.
[23] M. Rohrbach, S. Ebert, B. Schiele, Transfer learning in a transductive setting, in: Advances in Neural Information Processing Systems, 2013, pp. 46–54.
[24] G. Sharma, Ohio State University, 2002 Master of Science Thesis.
[25] M. Shen, D.-R. Liu, S.-H. Shann, Outlier detection from vehicle trajectories to discover roaming events, Inf. Sci. 294 (2015) 242–254.
[26] X. Wen, L. Shao, Y. Xue, W. Fang, A rapid learning algorithm for vehicle classification, Inf. Sci. 295 (2015) 395–406.
[27] Z. Zhao, D. Yan, W. Ng, S. Gao, A transfer learning based framework of crowd-selection on twitter, in: Proceedings of the ACM International Conference on Knowledge Discovery and Data Mining, 2013, pp. 1514–1517.
[28] H. Zheng, Y. Yu, A morphological neural network approach for vehicle detection from high resolution satellite imagery, J. Neural Inf. Process. (2006) 90–106.