Vehicle Detection in Wide-Area Aerial Imagery: Cross-Association of Detection Schemes with Post-Processings

Xin Gao

Department of Electrical and Computer Engineering
The University of Arizona, Tucson, AZ85721, USA
Email: tzgaoxin2016@hotmail.com

Abstract: Post-processing schemes are crucial for object detection algorithms to improve the performance of detection in wide-area aerial imagery. We select appropriate parameters for three algorithms (variational minimax optimization [2], feature density estimation [3] and Zheng's scheme by morphological filtering [4]) to achieve the highest average F-score on random sample frames, and then follow the same procedure to implement five post-processing schemes on each algorithm. Two low-resolution aerial videos are used as our datasets to compare automatic detection results with the ground truth objects on each frame. The performance analysis of post-processing schemes on each algorithm are presented under two sets of evaluation metrics.

Keywords: Post-processing, object detection, wide-area aerial imagery

Reference to this paper should be made as follows: Gao, X. (xxxx) 'Vehicle detection in wide-area aerial imagery: cross-association of detection schemes with post-processings', *Int. J. Image Mining*, vol. x, no. x. pp. xx-xx, 2018.

Biographical notes: Xin Gao is pursuing Ph.D. degree in Department of Electrical and Computer Engineering with Ph.D. minor in Mathematics at the University of Arizona, Tucson, AZ, USA. Mr. Gao's research interests mainly include communications and coding theory, statistical signal processing, algorithms on digital image analysis and digital image processing, object detection and classification, data (signal, image and video, etc.) compression in intelligent transportation systems, biomedical signal processing and imaging devices, and dynamic quantitative fiscal modelling. He has already published more than 20 English research articles with Google scholar citations of over 240 times during his graduate study by early 2018.

1 Introduction

In wide-area aerial imagery [1], [13]-[15], object detection in a traffic video represents a primary, challenging task which involves a variety of applications such as urban traffic management, remote visual surveillance and emergency rescue. Due to the diversity of appearance model, object occlusion as well as limitations of digital camera resolution, the current state-of-the-art algorithms are still quite difficult to achieve perfect detection performance. We presented a performance comparison of six object detection schemes in [13], while detection algorithms without post-processing bring up with a great number of

false objects; hence, developing reliable and efficient schemes to reduce percentage of wrong classification for each algorithm, generates our research attention.

We studied any possible post-processing schemes that help to improve performance for each detection algorithm: Salem *et al.* presented the binary masking scheme [5] with a median filter for smoothing, followed by a dilation operator to shape the detection on binary images. Samarabandu and Liu proposed the heuristic filtering scheme [6] (with two adjustable thresholds) to discard false detections after region extraction. Sharama *et al.* calculated the shape index (SI) [7] that drops false detections below the lowest SI on each object, while the sieving and opening scheme by Zheng *et al.* [4] sieved out large objects which are impossible to be a vehicle, or too small to be a vehicle comparing to average vehicle size in a wide-area aerial traffic video.

Motivated by the four schemes [4]-[7] as mentioned above, we used a two-stage post-processing scheme [13] to achieve better performance of object detection algorithms. The first stage is discarding any binary object out of a range threshold $A \in [t_{low}, t_{high}]$ on pixel size in every frame, and the second stage is exploiting an unbiased filter to perform morphological closing operation filter to shape detections and reduce errors. Pixel-size based range thresholds are chosen with respect to the size of ground truth vehicles in each dataset, while the size of morphological closing filter is flexible for designing [8] with respect to the metric of average F-score.

In this paper, we use average F-score as a primary evaluation metric, and adapted an image segmentation algorithm using variational minimax optimization (VMO) [2], the feature density estimation (FDE) approach [3] and Zheng's scheme on morphological filtering (MF) [4] by top-hat and bottom-hat transformation for investigation. We present a performance comparison of VMO [2], FDE [3] and MF [4] followed by each of the five post-processing schemes [4]-[7], [13]. We first take each algorithm (in absence of post-processing) for parameter tuning so as to achieve the highest average F-score on random sample frames, then determine the most appropriate parameters for each algorithm with four existing schemes [4]-[7] and the proposed two-stage scheme [14] on post-processing. The selected parameters for each scheme were used to test a sequence of consecutive frames from two aerial videos. Manual vehicles serve as the ground truth to compare with the automatic detection results. We classify a few detection errors on detections, and present the performance analysis of VMO [2], FDE [3] and MF [4] (free of any post-processing) combined with five respective post-processed schemes via two sets of evaluation metrics: i) the basic metrics on information retrieval [10], which comprise precision, recall, F-score and percentage of wrong classification; ii) the CLEAR evaluation metrics [11], [12] including multiple object detection accuracy and multiple object counts. We did not use the metric of multiple object detection precision because the representative detection algorithms [2]-[4] are lacking in spatial overlap information for a fair comparison of detection outputs to the ground truth (GT).

2 Methodology

Two aerial video datasets (spatial resolution: 720×480 pixels per frame, multiple consecutive frames per dataset) obtained from a low-resolution camcorder are used for performance analysis of VMO [2], FDE [3] and MF by Zheng [4] with five post-processing schemes [4]-[7], [13]. We manually crop traffic lanes from the aerial videos (GIS mapping or exploiting an automatic detection method also perform the same task on

road extraction.) In our manual segmentation, there are a total number of 8072 ground truth vehicles (of uniform rectangular shape) in the sample frames of these datasets. The size of vehicle area are distributed from 40 to 150 pixels in the Tucson (T) dataset, and from 20 to 175 pixels in the Phoenix (P) dataset.

We tested VMO [2], FDE [3] and MF [4] free of any post-processing by randomly selected frames from each sample dataset, adjusted parameters of each algorithm to obtain the highest average F-score, then specified the parameters for each algorithm in Table 1. Following the same procedure, we specified the parameters for the five post-processing schemes in Table 2, where each scheme was assigned with a notation (M1: Binary masking [5], M2: Heuristic filtering [6], M3: Filtering by a shape index [7], M4: Sieving and opening [4], M5: The proposed scheme [13], M0: No post-processing; the same notations with Figure 2, Figure 3 and Table 4).

Table 1. General parameter selection for each detection algorithm in two datasets

Algorithm / Year	Dataset	Tucson (T)	Phoenix (P)
	Parameter		
VMO [2]	Iteration times	10	20
	B for Region segmentation (normalized)	M1, M2, M4: $B = 0.7$; M0, M3, M5: $B = 0.6$	M1, M4, M5: $B = 0.7$; M0, M2, M3: $B = 0.6$
FDE [3]	Threshold on Sobel operator	0.30	0.20
	T for Region segmentation	$T = 90$ (grey threshold)	$T = 40$ (gray threshold)
MF [4]	Top-hat, bot-hat transform	Morphological filter, size = 3*3	Morphological filter, size = 3*3

Table 2. Parameter selection for the five post-processing (Post-Proc.) schemes associated with three detection algorithms

Post-Proc.	Parameter	VMO	FDE	MF
M1 [5]	dilation	size = 3	size = 3	size = 3
M2 [6]	[area, width/ height]	[7, 5] (T)	[5, 5] (T)	[5, 4] (T)
M3 [7]	constant $f(.)$ ($\times$ largest SI)	0.17 (T), 0.05 (P)	0.10 (T), 0.05 (P)	0.45 (T), 0.25 (P)
M4 [4]	opening	size = 2 (both)	size = 2(T), 1(P)	size = 1 (both)
M5 [14]	closing	size = 4	size = 3	size = 3
	$[t_{low}, t_{high}]$	[5, 160] (T), [5, 180] (P)		

3 Error Classification

With respect to the procedure of region matching [9], we evaluate each of the detection algorithms by automatically classifying several types of detection errors. The binary output of detections are labelled with 8-connected component which compares the overlap of detection with each object in the ground truth. Each type of detections can be automatically classified as follows [13]:

True positive (TP): correct detection. Only one TP (the one having largest overlap if exists) in each row is counted; even if multiple detections intersect the same ground truth object, or a single detection intersects multiple ground truth objects, TP is still unique in each row;

False negative (FN) or Miss: detection failure, indicated by a ground truth object that does not intersect any detected object;

False positive (FP): incorrect detection, indicated by a detection that fails to intersect any ground truth object;

No true negative (TN) sample exists on each dataset.

For the cross-association of each post-processing scheme with a detection algorithm, we separately accumulate each type of detections counted on all the sample frames from the two aerial datasets.

4 Error Classification

We use the basic information retrieval (IR) measures [10] to quantify the object detection performance:

Precision (correctness): the ratio of TP to sum of TP and FP;

Recall (completeness): the ratio of TP to sum of TP and FN;

F-score: the harmonic mean of precision and recall; the simplified expression of F-score equals the ratio of $2 \times TP$ to sum of $2 \times TP$, FN and FP;

Percentage of wrong classification (PWC): the ratio of FP and FN to sum of TP, FP, FN and TN. An algorithm with lower PWC indicates better performance on detection.

To assess the system performance of each algorithm with the five post-processing schemes, we also used the CLEAR evaluation metrics [11], [12] for comparison. Let us denote the number of Misses as m_i and the number of FPs as fp_i , the multiple object detection accuracy (MODA) in the i -th frame ($i = 1, 2, \dots, 100$ in each dataset) is computed as [12]:

$$\text{MODA}(i) = 1 - \frac{c_m(m_i) + c_f(fp_i)}{N_G^{(i)}} \quad (1)$$

where c_m and c_f represent the cost functions of the Misses and FPs, respectively; $N_G^{(i)}$ stands for the counts of the ground-truth objects in the i -th frame.

We equally weight $c_m(\cdot) = c_f(\cdot) = 1$ on each algorithm [11] for the performance evaluations; hence, the normalized multiple object detection accuracy (N-MODA) [11], [12] has the same numerical value to the multiple object count (MOC) metric [12] on the count accuracy over all the frames:

$$\text{MOC} = 1 - \frac{\sum_{i=1}^{N_{\text{frames}}} [(m_i) + (fp_i)]}{\sum_{i=1}^{N_{\text{frames}}} N_G^{(i)}} \quad (2)$$

5 Experimental Results

In this section, we present a performance analysis of VMO [2], FDE [3] and MF [4] associated with each of the five post-processing schemes [4]-[7], [13], where the total number of classified detections on each scheme are displayed in Table 3. For comparison of each algorithm without any post-processing, we justify that VMO [2] displays the highest number of TPs and the lowest number of FNs. Comparing the detections of five post-processing schemes on the same algorithm, we justify that the proposed scheme [13] indicates the lowest number of FPs for both VMO [2] and FDE [3], and the second lowest number of FPs for MF [4].

Table 3. Error classification: each detection algorithm free of post-processing and associated with each of the five post-processing schemes

Algorithm	Detection Post-processing	TP	FN	FP
VMO	No Post-Processing	7296	568	18490
	Binary Masking [5]	5495	842	2814
	Heuristic Filtering [6]	6025	1889	3596
	Filtering by SI [7]	6319	1445	10240
	Sieving and Opening [4]	5953	2062	5523
	Proposed Scheme [13]	5678	1169	2476
FDE	No Post-Processing	6480	1233	6706
	Binary Masking [5]	5839	1410	2232
	Heuristic Filtering [6]	6050	1493	2438
	Filtering by SI [7]	5893	2089	3717
	Sieving and Opening [4]	5640	2214	2156
	Proposed Scheme [13]	5508	1598	1400
MF	No Post-Processing	7035	843	47279
	Binary Masking [5]	5721	870	8557
	Heuristic Filtering [6]	5547	2348	7677
	Filtering by SI [7]	5030	2972	24573
	Sieving and Opening [4]	5111	2888	1977
	Proposed Scheme [13]	5715	1721	3107

For all the detection algorithms, a visual comparison of detections without post-processing (with colour-bounded visual vehicles), detections after using a post-processing scheme (colour-boundary on regions) and the ground truth (rectangular regions in white

colour) is depicted in Figure 1. We pick up the 50th frame of both datasets as samples, and select each subimage with the size of 64×64 pixels. From Figure 1, we justify that the proposed scheme [14] combined with VMO [2] and FDE [3], display the best visual performance across each division lanes of traffic among the five schemes. For MF in Zheng's method [4], sieving and opening [4] and our scheme [14] merely exhibits better view than the other three post-processings [5]-[7], which requires follow-up evaluations.

Figure 1. Visual comparison on vehicle detections: VMO, FDE and MF with no post-processing, and those with each of the five post-processing schemes

The evaluation metrics [10]-[12] for each algorithm respectively combining each of the five post-processing schemes are presented in Figures. 2-3 and Table 4, where a t -test with sample mean and 95% confidence interval (CI) was performed on all the sample frames in datasets. By the basic IR metric [10] in Figure 2, we justify that among the five schemes, the proposed scheme (M5) displays the highest precision for VMO and FDE, the second highest precision (next to M4: sieving and opening) for MF; binary masking (M1) displays the highest recalls on all three algorithms, while our scheme (M5) exhibits the second best recalls for VMO and MF, a median recall for FDE; meanwhile, our scheme (M5) displays the highest F-score, and the lowest PWC score (Table 4) for each

Title

algorithm using the five schemes. In Figure 3, our scheme (M5) also displays the best performance judging from the CLEAR metric [11], [12], (where the paired MODA and MOC in each diagram closely coincide in the quantitative values), among the five post-processing schemes for VMO, FDE and MF. Our statistical data is supported by the tight CIs for each of these evaluation metrics.

(a) VMO

(b) FDE

(c) MF

Figure 2. Comparison on basic IR metrics [10] of no post-processing and five post-processing schemes associated with: (a) VMO, (b) FDE, and (c) MF.

Author

(a) VMO

(b) FDE

(c) MF

Figure 3. Comparison on CLEAR metrics [11], [12] of no post-processing and five post-processing schemes: associated with: (a) VMO, (b) FDE, and (c) MF.

Table 4. Comparison on PWC: VMO, FDE and MF (all free of any post-processing), or each combined with any of the five post-processing schemes (mean and 95% CI)

Metric / Scheme Algorithm	PWC %					
	M0	M1 [5]	M2 [6]	M3 [7]	M4 [4]	M5 [14]
VMO	71.08 ± 0.98	39.22 ± 1.08	47.44 ± 1.32	64.40 ± 0.76	54.82 ± 1.11	38.62 ± 1.09
FDE	54.70 ± 0.89	35.80 ± 2.21	37.87 ± 1.88	45.54 ± 2.27	43.12 ± 1.11	33.83 ± 1.87
MF	82.34 ± 1.44	59.16 ± 1.87	59.69 ± 2.48	81.06 ± 1.41	47.08 ± 1.88	45.07 ± 1.93

6 Conclusions

In sum, we have presented a performance evaluation on five post-processing schemes [4]-[7], [13], where each scheme are crossly associated with three detection algorithms named as VMO [2], FDE [3] and MF [4]. Visual comparison and quantitative analysis from two sets of evaluation metrics [10]-[12] of our experimental design, indicate better validity and efficiency of the proposed post-processing scheme [13] by an area-thresholding sieving process and morphological closing transformation in comparison to the other four schemes. Regarding to the effectiveness of cross association on selecting two schemes among the five post-processing schemes for three detection algorithms, the proposed two-stage scheme on sieving and closing, and the binary masking scheme, represent the best two methods for VMO [2] and FDE [3], while our scheme as well as sieving and opening, indicate the best two methods for MF (which was adapted from Zheng's scheme in [4]).

7 References

- [1] R. Porter, A. M. Fraser, and D. Hush, "Wide-area motion imagery," *IEEE Signal Process. Mag.*, vol. 27, no. 5, pp. 56-65, Sept. 2010.
- [2] B. N. Saha and N. Ray, "Image thresholding by variational minimax optimization," *Pattern Recognit.*, vol. 42, no. 5, pp. 843-856, May 2009.
- [3] J. Gleason, A. V. Nefian, X. Bouysounousse, T. Fong, and G. Bebis, "Vehicle detection from aerial imagery," in *2011 IEEE Int'l Conf. Robotics Automat. (ICRA' 2011)*, May 9-13, 2011, Shanghai, China, pp. 2065-2070.
- [4] Z.-Z. Zheng, G.-Q. Zhou, Y. Wang, Y.-L. Liu, X.-W. Li, X.-T. Wang, and L. Jiang, "A novel vehicle detection method with high resolution highway aerial image," *IEEE J. Sel. Top. Applied Earth Observ. Remote Sens.*, vol. 6, no. 6, pp. 2338-2343, Dec. 2013.
- [5] M. A. Salem, N. Ghamry, and B. Meffert, "Daubechies versus biorthogonal wavelets for moving object detection in traffic monitoring systems," *Informatik-Berichte 229*, ISSN: 0863-095X, pp. 8-9, 2009.

- [6] J. Samarabandu and X.-Q. Liu, "An edge-based text region extraction algorithm for indoor mobile robot navigation," *Int'l J. Signal Process.*, vol. 3, no. 4, pp. 273-280, 2007.
- [7] B. Sharma, V. K. Katiyar, A. K. Gupta, and A. Singh, "The automated vehicle detection of highway traffic images by differential morphological profile," *J. Transp. Technol.*, vol. 4, pp. 150-156, 2014.
- [8] S. Li, G.-Q. Zhou, Z.-Z. Zheng, Y.-L. Liu, X.-W. Li, Y. Zhang, and T. Yue, "The relation between accuracy and size of structure element for vehicle detection with high resolution highway aerial images," in *2013 Int'l Geosci. Remote Sens. Symp. (IGARSS' 2013)*, July 21-26, 2013, Melbourne, Australia, pp. 2645-2648.
- [9] J. C. Nascimento, and J. S. Marques, "Performance evaluation of object detection algorithms for video surveillance," *IEEE Trans. Multimedia*, vol. 8, no. 4, pp. 761-774, Aug. 2006.
- [10] S. H. Shaikh, K. Saeed, and N. Chaki. *Moving Object Detection Using Background Subtraction*, Springer 2014, pp. 30-31.
- [11] R. Kasturi, D. Goldgof, P. Soundararajan, V. Manohar, J. Garofolo, R. Bowers, and J. Zhang, "Framework for performance evaluation of face, text, and vehicle detection and tracking in video: data, metrics, and protocol," *IEEE Trans. Pattern Anal. Mach. Intell. (PAMI)*, vol. 31, no. 2, pp. 319-336, Feb. 2009.
- [12] R. Kasturi, D. Goldgof, P. Soundararajan, V. Manohar, M. Boonstra, and V. Korzhova, "Performance evaluation protocol for face, person and vehicle detection & tracking in video analysis and content extraction (VACE-II)," *Comput. Sci. Eng. University of South Florida, Tampa, FL*, pp. 17-18, Jan. 2006.
- [13] X. Gao, S. Ram, and J. J. Rodríguez, "A performance comparison of automatic detection schemes in wide-area aerial imagery," in *2016 IEEE Southwest Symp. Image Anal. Interpret. (SSIAI'2016)*, March 6-8, 2016, Santa Fe, NM, pp. 125-128.
- [14] S. Workman, R. Souvenir, and N. Jacobs, "Wide-area image geolocalization with aerial reference imagery," in *2015 IEEE Int'l Conf. Comput. Vis. (ICCV'2015)*, December 13-16, 2015, Santiago, Chile, pp. 3961-3969.
- [15] A. Elliethy and G. Sharma, "Automatic registration of wide area motion imagery to vector road maps by exploiting vehicle detections," *IEEE Trans. Image Process.*, vol. 25, no. 11, pp. 5304-5315, Nov. 2016.