Assessment of Vision-Based Vehicle Tracking for Traffic Monitoring Applications

Dale Joshua R. Del Carmen and Rhandley D. Cajote
Electrical and Electronics Engineering Institute
University of the Philippines Diliman
Quezon City, Philippines
Email: dale.del.carmen@eee.upd.edu.ph and rhandley.cajote@eee.upd.edu.ph

***Abstract*—Multiple-target tracking (MTT) is one of the main components of traffic monitoring systems directly responsible for measuring traffic information. However, assessment and evaluation of these vehicle tracking systems greatly varies and are often incomparable due to different metrics and datasets. This paper focuses on comparing and assessing the viability of online multiple vehicle tracking systems for use in real-time traffic monitoring. Most online vehicle tracking framework uses background subtraction for real-time vehicle detection and various blob-based appearance models for real-time multiple vehicle tracking. The results show that commonly used metrics such as multiple object tracking accuracy (MOTA) and multiple object tracking precision (MOTP) are not necessarily reflective of traffic monitoring performance, particularly in terms of vehicle count accuracy. Furthermore, the track identity switching (IDS) metric is identified to significantly affect the vehicle count accuracy, particularly in terms of count precision, having a correlation coefficient of $r(100) = -0.709$ with P-value $p < 0.001$.**

I. Introduction

The work presented in this paper was motivated in part by the need identified to develop various Intelligent Transport System (ITS) applications to improve the Philippines' traffic management capability by maximizing the utility of road resources in an efficient and effective manner [1]. ITS applications include traffic signal control systems, travel time prediction, incident detection, and traffic counting; all of which require an accurate monitoring of the road infrastructure to extract pertinent data. One of the means to extract data required by these applications is through traffic monitoring. Real-time vision-based traffic monitoring systems, in particular, have progressed significantly due to the advancements in the field of computer vision and computing. Camera sensors are easier to install and maintain, and can record much more complex information compared to other sensors which makes them viable for ITS applications.

A. Vision-based Traffic Monitoring

To have a detailed traffic information of a road network, vehicles must be tracked individually. In the context of traffic monitoring, online tracking is widely used as it uses less resources making it possible to be implemented in real-time on a local processing unit. The typical process [2]–[7] of an online video-based traffic monitoring system is shown in Fig. 1. An image of the current traffic scene is extracted from a video camera in the form of raw frames. A smaller part of the image, called the region of interest (ROI), may then be taken to reduce the resources needed by the system. The ROI may then be further processed to correct perspective distortions either based on the camera configuration or on the traffic flow. After processing the frame, vehicles are detected using motion and/or appearance models. The state of each vehicle detected, such as its position and bounding box, are then estimated based on the model(s) used. This is sequentially done for each frame in the video, with each newly detected vehicle associated with previously detected ones to form vehicle tracks, which are sequences of vehicle states for each unique vehicle. Traffic information can then be extracted from the tracks produced such as vehicle count [7], [8], which ideally corresponds to the number of tracks, and average vehicle speed [2], [7] which can be computed from the tracks. Depending on the appearance model used, each vehicle may also be classified [2], [4], [7] for additional traffic information.

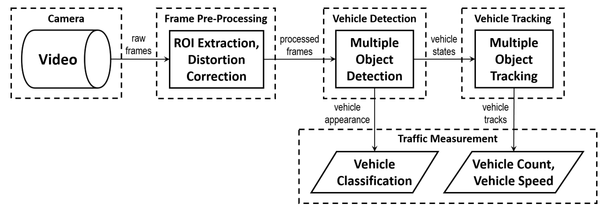


Fig. 1. A typical vision-based traffic monitoring system.

As seen in the typical traffic monitoring framework, vehicle tracking plays a significant part in the measurement of traffic parameters such as vehicle count and vehicle speed, which are predominant metrics needed for many ITS applications. In the context of traffic monitoring, online tracking is widely used as it uses less resources making it possible to be implemented in real-time on a local processing unit. This paper focuses on the performance assessment of online multiple vehicle tracking systems and their reliability in extracting traffic information on different traffic scenes. A comparable assessment of existing state-of-the-art tracking methodologies will aid in identifying appropriate design considerations for multiple vehicle tracking systems used in vision-based traffic monitoring.

II. Related Work

Tracking, in the context of this paper, refers to the tracking-by-detection (TBD) framework. That is, all objects of interest are detected for every frame in the video sequence before being fed to a multiple object tracking algorithm. In particular, online tracking systems, which processes the detections frame-by-frame are explored. These are then put in context of video-based traffic monitoring.

A. Vehicle Detection

Vehicle detection is an important preliminary process for multiple vehicle tracking. It has been shown that the probability of correct detection affects tracking in terms of probability of losing track and probability of switching [9]. This means that a more accurate vehicle detection process is likely to result in a more accurate tracking. As a result, multiple object tracking systems should be assessed jointly with its detections to give a better understanding of its overall performance [10]. The most common vehicle detection technique for daytime static cameras is background subtraction [5]–[8].

Background subtraction, in general, is the process of differentiating the moving foreground from the static background. In the context of vehicle detection, background subtraction can be used either independently [6], [7], or for foreground localization [5], [8]. Foreground localization approaches uses other image features such as scale invariant feature transform (SIFT) descriptors [8] and fast retina keypoints (FREAK) [5] for improved detection performance and for detection-level occlusion handling. The foreground object regions processed from background subtraction may also be used to classify vehicles according to size [7].

B. Multiple Object Tracking

Vision-based multiple target tracking systems can be categorized using their component models [12]. The appearance model is perhaps the most important model in vision-based tracking systems as it differentiates vision-based tracking systems from general state-based tracking systems. Appearance models describes how an object is represented during detection and/or tracking. Detection-level representation may vary from tracking-level representation in the case of online tracking. Detection-level vehicle appearance models in the literature include 3D models [4], [11], edges [3], and blobs [5], [7]. Edge-based tracking makes use of contour or edge features for tracking vehicles. For blob-based tracking: a vehicle may be represented by its bounding box [3], [7], color histogram [13], [14], or feature points [5], [8] during tracking.

Among the different appearance models used in vehicle tracking, blob-based representation is the most popular. This is due to the fact that online background subtraction algorithms can be implemented for real-time multiple vehicle detection. This makes blob-based tracking a feasible candidate for implementing low-cost real-time multiple vehicle tracking systems.

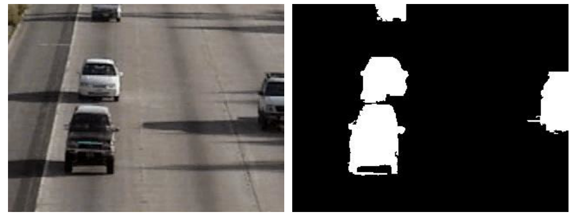


Fig. 2. Extracted vehicle mask for blob-based tracking along with the corresponding original frame. Video source taken from [17].

C. Blob-based Vehicle Tracking Systems

In this vehicle tracking approach, the scene is segmented into vehicular and non-vehicular components. Vehicles are generally represented by blobs or connected pixels with the same values. An example is shown in Fig. 2, where the white blobs represent the detected vehicles. The most common vehicle segmentation scheme is background subtraction.

Among blob-based trackers, state-based tracking [7] is the simplest as each vehicle is represented by the estimated state (e.g. position) of the vehicle's bounding box. The main drawback of this is that nearby vehicles in traffic scenes are likely to have similar states and thus cannot be differentiated under occlusion. Histogram-based tracking improves upon state-based tracking by using particle filters along with additional vehicle information such as color histogram [13], [14]. However, the use of histograms requires proper initialization and object segmentation. On the other hand, feature-based tracking [5] makes use of vehicular feature points and visual descriptors to represent a vehicle. The use of multiple visual descriptors for each vehicle makes it possible to differentiate vehicles during partial occlusion, assuming each vehicle has enough difference in motion.

D. Performance Metrics for Assessment

A traffic monitoring system should be able to measure traffic parameters accurately and as fast as possible to be of benefit to ITS applications. The metrics listed in this section have been used to define the performance of vehicle tracking algorithms in the literature.

1) Execution Time and Track Accuracy: Execution time, as reported, is the average or expected time an algorithm will finish to complete a certain task. This metric is useful, in aggregation to the execution time of other parts of the traffic monitoring system, to determine if a system can be implemented in real-time [16].

The simplest way of reporting the accuracy of a tracking algorithm is by comparing the tracked states with the actual states (e.g. position) of an object of interest [6]. In vision-based tracking, this is usually either the centroid or the bottom center of an object's bounding box. The drawback of this metric is that is only helpful for single target use as it cannot account for target-to-track mismatch or track loss which are both evident in multiple object tracking.

2) Multiple Object Tracking Performance: In the interest of comparing multiple target tracking algorithms, two metrics have been made namely Multiple Object Tracking Accuracy (MOTA) and Multiple Object Tracking Precision (MOTP). The computation of these metrics requires that there be an actual object state to tracker state correspondence for all frames of interest. A threshold-dependent procedure can be done to make a one-to-one correspondence between each object and tracker state. In the context of vision-based tracking, this threshold is related to either the distance between the object and tracker states or the bounding box overlap of the actual object and the estimate of the tracking system. An object with no matching state estimate is considered a *miss* or a *false negative* while a state estimate with no matching object is considered a *false positive*. A distance threshold is also set to determine if the *correspondence* is valid or not. If a correspondence is not valid, the pair will be considered a *mismatch*. The formal definition of each metric is then as follows [18]:

$$MOTA = 1 - \frac{\sum_t (m_t + fp_t + mme_t)}{\sum_t g_t} \quad (1)$$

where m_t is the number of misses or false negatives, fp_t is the number of false positives, mme_t is the number of mismatches, and g_t is the total number of correspondence; all at time t.

$$MOTP = \frac{\sum_{i,t} d_t^i}{\sum_t c_t} \quad (2)$$

where d_t^i is the distance of the *i*th matching valid correspondence at time t, and c_t is the number of valid correspondence at time t. For 2D MOT, d_t^i can be represented as the percent overlap between the object bounding boxes, more commonly known as intersection over union (IOU), or the pixel distance between the objects' point of interest.

From the definition, it can be seen that MOTA is penalized by untracked objects, false state estimates, and low individual object tracking accuracy. On the other hand, MOTP is only dependent on the matching pairs. Thus, both metrics are reported hand in hand to get a complete evaluation of a multiple object tracking system.

3) Track Quality: Aside from metrics regarding tracking accuracy, track quality metrics were used to compare the relative performance of different algorithms in multiple object tracking benchmarks [10], [19], as used previously for pedestrian tracking [20], to quantify the typical errors observed in multiple target tracking.

The first set of track quality metrics assesses how accurately each individual object is tracked. Tracks are classified as: mostly tracked (MT) if the track accuracy is greater than 80%; mostly lost (ML) if the track accuracy is less than 20%; and, partially tracked (PT) otherwise. Reporting of these metrics may be normalized by the total number of tracks available for comparison across different datasets. An ideal multiple object tracker should be able to maximize the amount of MT tracks while minimizing the amount of ML tracks.

The other two track quality metrics quantifies two common errors observed in multiple target tracking: identity switching (IDS) and track fragmentation (Fgmt). Both are reported as a cumulative measure for the whole data set and may be normalized by the Recall rate of the detections [10]. The IDS metric is reported as the sum of the total number of times a target object is incorrectly matched to a different object track while the Fgmt metric is reported as the sum of the total number of times each object is unmatched throughout its lifespan. The IDS also corresponds to the number of mismatches during MOTA computation. An ideal multiple target tracker should be able to minimize these two metrics.

4) Traffic Information Accuracy: Traffic information that can be extracted by multiple vehicle tracking systems include cumulative vehicle count and average vehicle speed. The cumulative vehicle count is incremented by the arrival and exit of vehicles; that is, the number of tracks created for a certain span should amount to the cumulative vehicle count in the same span. On the other hand, the individual average vehicle speed is estimated using the state of each vehicle during entry and exit and the frame rate of the test sequence.

The speed of a vehicle is computed using the formula:

$$s \approx kr \frac{||p_{exit} - p_{entry}||}{T_{exit} - T_{entry}} \quad (3)$$

where s is the average speed in meters per second, p is the vehicle point of interest coordinates in pixels, r is the frame rate in Hz, k is the video scale factor in meters per pixel, and T is the frame index of each p.

For traffic monitoring applications vehicle count accuracy [7], [8] and vehicle speed accuracy [2], [7], [16] have been used to determine the capability of the vehicle tracking system to extract information on the road network. The accuracy of these traffic information metrics is usually expressed using relative measurement accuracy, defined as:

$$Accuracy = \left(1 - \frac{|x - \hat{x}|}{x}\right) \times 100\% \quad (4)$$

where $\hat{x}$ is the measurement of the actual value x.

5) Comparison of Tracking Algorithms: Video-based tracking has a wide range of applications. Thus, it is expected to have several comparisons in the literature. Object tracking algorithms, in particular, have been surveyed with great detail [21]. Particularly, different tracking frameworks were compared in terms of their capability to handle object entry, object exit, occlusion, and multiple objects. Another comparison included the evaluation of certain algorithms to handle tracking under translation, rotation, and scaling. However, such comparisons were qualitative and lack the comparison of actual tracking implementations.

Vision-based multiple object detection and tracking benchmarks have been made in recent years in response to the lack of comparison among existing state-of-the-art algorithms [10], [19]. However, the tracking systems evaluated in these benchmarks are mostly pedestrian-oriented or detection-independent offline trackers which are not feasible for real-time implementation.

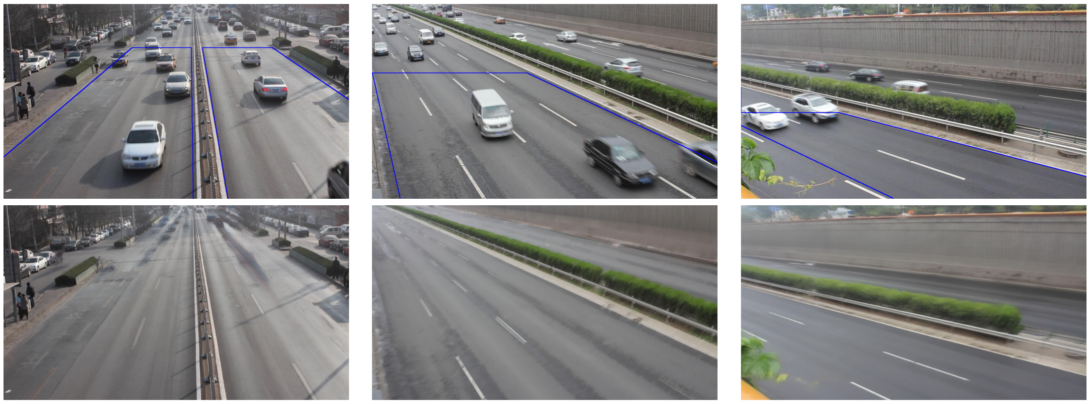


Fig. 3. Example of stable daytime traffic scenes from the UA-DETRAC dataset and their generated background model [10]. The blue lines correspond to the boundary of the region of interest.

III. METHODOLOGY

A comparative testbed is adapted from a typical background subtraction-based traffic monitoring framework and shown in Fig. 4. The test traffic sequences are processed, frame-by-frame. First, a mask containing the region of interest is applied on the current frame to be processed. After which, the initial background image is compared with the current region of interest (ROI) using background subtraction [22] to detect the vehicular blobs in the ROI. The vehicle bounding boxes are then determined from the detected vehicular blobs; and, depending on the system, the vehicle will be represented by a certain track-level appearance model. The state of the vehicle (e.g. bounding box) is then predicted, based on the previous state, using the tracking system's probabilistic inference model. The next frame is then processed. After which, the resulting detected vehicles are associated with the previously detected vehicles, based on the appearance model. The series of states, formed from associated vehicle detections, then form several vehicle tracks which are all compared to the annotated bounding boxes to determine the multiple object tracking performance, track quality, and traffic monitoring performance of the tracking system evaluated.

A. Dataset

A total of 34 stable daytime sequences are chosen from the UA-DETRAC training dataset [10]. An initial background mask is generated from the first 100 frames (4 seconds) of each sequence using a standard temporal median filter to avoid ghosting artifacts due to the use of background subtraction. A user-defined region of interest (ROI) containing the traffic flow is defined in each sequence resulting to a total of 140,644 bounding boxes from 2,213 vehicles in a total of 43,241 frames for assessment. Example images of the sequences used along with the user-defined ROI and generated background model is shown in Figure 3.

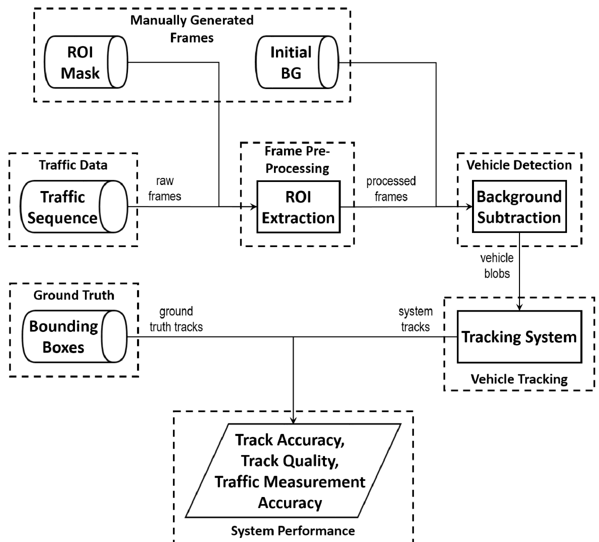


Fig. 4. Comparative framework for the assessment of multiple vehicle tracking systems.

B. Tracking Implementation

Three baseline tracking systems are implemented using the Hungarian optimization data association model through the Munkres algorithm. This data association model should be sufficient for the various traffic scenes in the dataset which are high frame rate video sequences in which vehicle movement is minimal in consecutive frames. Another common assumption used for all baseline tracking systems is that a tracked object can only be lost or misdetected for at most 5 frames. Otherwise, the tracked object is terminated and not registered as a vehicle track. The specific implementation of the three blob-based tracking systems are summarized below.

1) State-based Tracking: This system uses a Kalman filter to predict the vehicle bounding box on the next frame [7]. The Munkres association cost is then set to the amount of bounding box overlap, commonly known as intersection over union (IOU), instead of pixel distance. The (minimum) overlap threshold is set to $T_{overlap} = 0.5$. A simple initialization and termination procedure is implemented and adapted from UrbanTracker [5].

2) Feature-based Tracking: The UrbanTracker [5] is adapted for evaluating feature-based tracking. This tracker implements a state-machine for track initialization and termination, and merge-and-split procedures for occlusion handling. The FREAK keypoint detector is set to a threshold value of 25 and octave number of 3. The default values of the FAST feature extractor in OpenCV is used.

3) Histogram-based Tracking: The Kalman filter in the state-based tracking system is replaced with a standard particle filter. All other multiple object tracking models are identical with the state-based tracker. The number of particles used is 200 and a total of 24 histogram bins are used for storing the color histogram, 8 for each of the RGB channels.

C. Assessment and Testing

The following performance metrics will be used to evaluate each candidate tracking methodology: (1) multiple object tracking performance, using MOTA and MOTP, as well as the False Positives (FP) and False Negative (FN) counts used in MOTA computation; (2) track quality, through IDS, Fgmt, MT, and ML; (3) traffic count estimation performance, through count precision and recall, and count measurement accuracy; and (4) traffic speed estimation performance, using speed measurement accuracy.

The multiple object tracking performance and track quality metrics are computed using a general MOT toolkit [19] which compares the ground truth annotations with the tracking system output tracks. Similarly, the traffic count and speed measurement performance are computed by comparing the system vehicle count and system average vehicle speed output with the actual value based on the dataset ground truth annotations.

Measurement of the true positive count for large datasets is impractical since it would need to be manually counted. Instead, the vehicle count true positive can be analytically determined by counting how many ground truth tracks uniquely correspond to each system output track. The ground truth vehicle count in a sequence can be easily determined by from the number of tracks in the ground truth annotations; analytically, this corresponds to the sum of the vehicle count true positive and false negative. Meanwhile, the vehicle count measured by a tracking system analytically corresponds to the sum of the vehicle count true positive and false positive. The vehicle count binary metrics can thus be approximated by implementing an association threshold on the ground truth and system output tracks to determine the upper bound for vehicle count true positives.

TABLE I
SUMMARY OF METRICS FOR PERFORMANCE ASSESSMENT

Metric	Ideal	Description
MOTA	100%	Multiple Object Tracking Accuracy
MOTP	100%	Multiple Object Tracking Precision
FP	0	False Positives
FN	0	False Negatives or Misses
IDS	0	Identity Switches or Mismatches
Fgmt	0	Fragmentation
MT	100%	Mostly Tracked targets (>80% tracked)
ML	0%	Mostly Lost targets (<20% tracked)
CnPrn	100%	Count Precision
CnRcl	100%	Count Recall
CnAcc	100%	Count Accuracy
SpAcc	100%	Speed Accuracy

The computation for vehicle count precision and count recall is then as follows:

$$CnPrn = \frac{TP_{cn}}{TP_{cn} + FP_{cn}} \times 100\% \quad (5)$$

$$CnRcl = \frac{TP_{cn}}{TP_{cn} + FN_{cn}} \times 100\% \quad (6)$$

where:

TP_{cn} is the number of correctly counted vehicles, determined analytically;

FP_{cn} is the number of incorrect count increment; and

FN_{cn} is the number of vehicles not counted correctly.

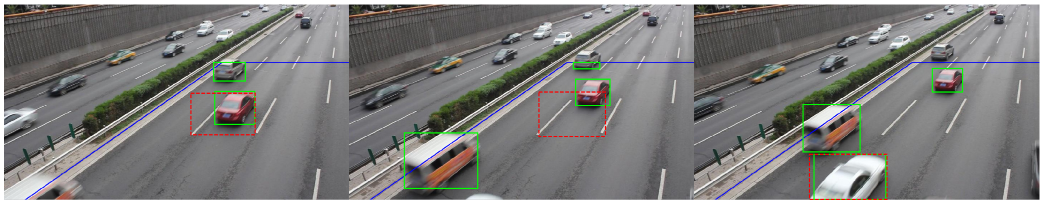
Lastly, the vehicle speed measurement is done by comparing the average vehicle speeds computed from the associated ground truth and system output tracks. Specifically, it is taken as one (1) minus the average relative error, times one-hundred percent (100%). This makes the two traffic measurement accuracy metrics independent of one another. Thus, for a tracking system to have a good traffic monitoring performance, it must have a high count and speed accuracy; as well as a high count precision and recall, to validate the reliability of the measured count accuracy. A summary of the different assessment metrics to be used are shown in Table I.

D. Overlap Threshold

The performance metrics used in multiple object tracking algorithms are dependent on how each annotated (ground truth) bounding box are matched to a tracking system's output bounding box. For 2D vision-based tracking, this is done by employing an overlap threshold, t_{iou}, between the actual bounding box and the output bounding box. This overlap threshold is usually set to $t_{iou} = 0.5$ [19] for general cases but has also been set to $t_{iou} = 0.7$ [10] for strict multiple vehicle tracking assessment. For background subtraction based systems, detection bounding boxes may significantly be affected by cast shadows. Having a strict association threshold thus penalizes detections with shadows which is inherently present in the systems to be compared. Thus, we shall use the conservative threshold $t_{iou} = 0.5$ for reporting the evaluation metrics.

TABLE II
PERFORMANCE OF SELECTED VEHICLE TRACKING SYSTEMS

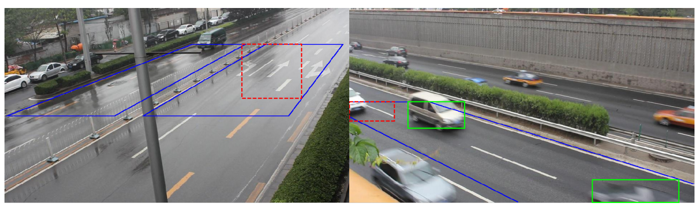
System	MOTA	MOTP	FP	FN	IDS	Fgmt	MT	ML	CnPrn	CnRcl	CnAcc	SpAcc
State	37.4%	70.9%	33155	53726	1147	2067	44.9%	17.3%	57.8%	88.5%	47.0%	83.3%
Histogram	26.6%	70.0%	31072	71075	1054	2012	20.8%	29.4%	57.1%	78.5%	62.6%	75.7%
Feature	28.3%	68.3%	41413	59056	421	2588	41.5%	21.4%	78.7%	86.4%	90.3%	83.6%



(a) frame n (b) frame $n+5$ (c) frame $n+10$

Fig. 5. Errors due to track loss.

The blue lines correspond to the boundary of the ROI. The green solid bounding boxes correspond to ground truth while the red dashed bounding boxes corresponds to tracked vehicles. Notice the tracked vehicle in (a). The tracker fails to update its hypothesis in (b), then the vehicle is lost in (c).



(a) tracked tree reflection (b) tracked lane marking

Fig. 6. Errors caused by misdetection.

The blue lines correspond to the boundary of the region of interest. The green solid bounding boxes correspond to ground truth while the red dashed bounding boxes corresponds to tracked vehicles. The reflection of a tree is tracked in (a) while a lane marking is tracked in (b).

IV. RESULTS AND ANALYSIS

The performance of each blob-based tracking system is evaluated in terms of multiple vehicle tracking performance, namely, tracking accuracy and track quality; and traffic monitoring performance, namely, count and speed accuracy. The cumulative performance metrics of each system for all 34 test sequences is shown in Table II.

The results show that in terms of multiple vehicle tracking, the state-based tracker is the most accurate in terms of matching bounding boxes to the ground truth annotation. The state-based system has the best MOTA, MOTP, MT, and ML metrics, followed by the feature-based system, with the histogram-based system having the lowest performance. On the other hand, the feature-based tracking system is the most accurate in terms of traffic monitoring performance. The feature-based system has the highest vehicle count and vehicle speed measurement accuracy. Meanwhile, the state-based and histogram-based systems have similar traffic monitoring performance which are both lower than that of the feature-based system; the feature-based system having a higher vehicle count accuracy, and the state-based system having a higher vehicle speed accuracy.

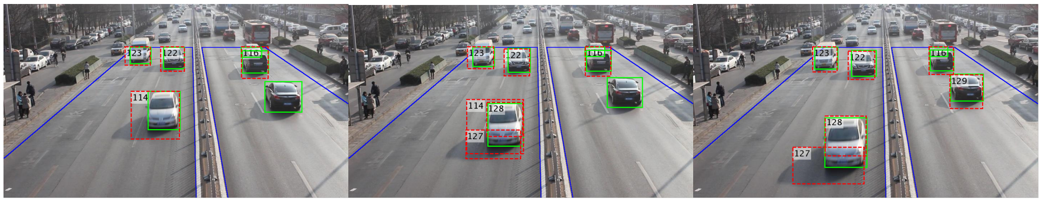
A. Analysis of Tracking Accuracy Metrics

We first analyze the result in terms of multiple vehicle tracking performance. As seen in Table II, the state-based system has a lower number of false positives and false negatives compared to the other two systems. Specifically, the histogram-based system has a higher amount of false negatives due to track loss which is shown in Fig. 5. This tendency to lose tracks also leads to the histogram-based system having the least amount of false positives since losing wrong tracks will result in a lower number of false positives. This can be attributed to the performance of the standard particle filter at the tested number of particles.

Meanwhile, the significantly higher number of false positives of the feature-based system compared to the other two systems can be attributed to tracking of non-vehicular blobs as shown in Fig. 6. This is likely due to differences in low-level tracking logic with the other two systems.

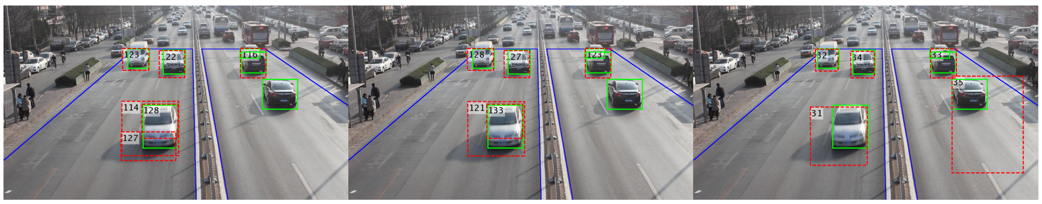
B. Analysis of Track Quality Metrics

Next, we analyze the results in terms of track quality. The other error metric accounted for in the MOTA is the number of identity switches (IDS), which is also treated separately to



(a) frame n (b) frame $n+5$ (c) frame $n+10$

Fig. 7. Count false positives due to vehicle fragmentation.
The blue lines correspond to the boundary of the region of interest. The green solid bounding boxes correspond to ground truth while the red dashed bounding boxes corresponds to tracked vehicles. Notice the vehicle on the lower half of the left lane. The vehicle hypothesis ID changes from (a) 114, to (b) either 114 or 128, to (c) 128.



(a) State-based system (b) Histogram-based system (c) Feature-based system

Fig. 8. Count false positives due to vehicle fragmentation.
The blue lines correspond to the boundary of the region of interest. The green solid bounding boxes correspond to ground truth while the red dashed bounding boxes corresponds to tracked vehicles. Notice the vehicle on the lower half of the left lane. The state-based and histogram-based tracking systems tracks the vehicle into separate parts likely due to blob separation. Meanwhile, the feature-based tracking system can merge the blobs into one vehicle.

measure track quality. The feature-based tracker has the least amount of identity switches compared to the other two tracking systems. A common case of identity switching is shown in Fig. 7.

C. *Analysis of Traffic Measurement Accuracy Metrics*

Finally, we analyze the results based on the traffic monitoring performance. The feature-based system has the highest count and speed accuracy even though it does not have the best multiple vehicle tracking performance. The state-based system, which has better multiple vehicle tracking performance, is affected mainly by its low count precision, which means that it tends to over-count vehicles, resulting to a low vehicle count measurement accuracy. The same over-counting problem is encountered by the histogram-based system. An example of how false positive counts occur is shown in Fig. 8. Meanwhile, the state-based and feature-based systems have a higher speed measurement accuracy than histogram-based systems due to having less false negatives which cause late track initializations.

From the analysis of results, both track identity switching and vehicle count false positives are mainly caused by vehicle fragmentation due to errors during background subtraction. Thus, the track identity switching metric, in the case of blob-based tracking systems, affects the vehicle count measurement accuracy in terms of vehicle count precision. This can be further validated by plotting the normalized IDS vs CnPrn for

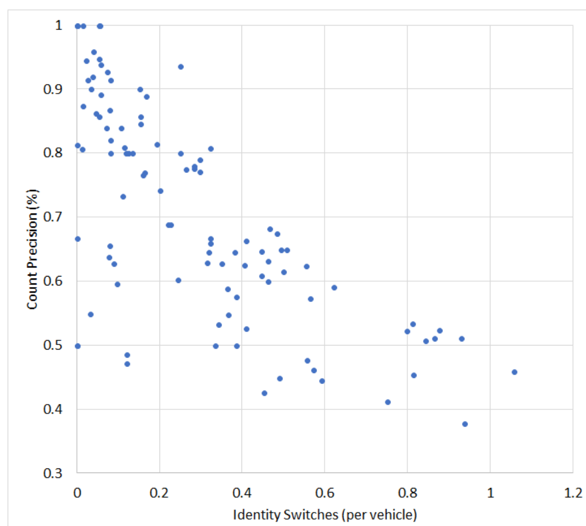


Fig. 9. Plot of normalized identity switches (IDS) against vehicle count precision (CnPrn).
The normalized IDS is the number of identity switches over the total number of vehicles in a sequence.

all possible sequence-to-system combinations, resulting to a total of 102 data points as shown in Fig. 9. Statistical analysis between the two normalized metrics results to a correlation coefficient of $r(100) = -0.709$ with a P-value of $p < 0.001$. For reference, the MOTA and CnPrn metrics have a correlation coefficient of $r(100) = -0.059$ and a P-value of $p = 0.554$.

V. Conclusion

This paper presents a standardized assessment and analysis of background subtraction-based vehicle tracking systems. The use of labeled bounding box annotations to determine the upper bound of object count true positives is introduced. The resulting count precision and count recall from the said true positive upper bound makes it easier to automate the count measurement accuracy of tracking systems reliably, albeit on labeled datasets, by considering the possibility of having undercount and over-count cancellations.

The feature-based tracking system has been found to be the best tracker in terms of traffic monitoring performance. Utilizing a complex appearance model decreases the amount of identity switching with the drawback of less accurate tracking overall (in terms of MOTA). Analysis of results show that track quality (in terms of IDS) affects traffic monitoring performance (in terms of CnPrn) much more significantly than the tracking performance (in terms of MOTA).

References

[1] "Roadmap for Transport Infrastructure Development for Metro Manila and Its Surrounding Areas (Region III and Region IV-A) in the Republic of the Philippines Final Report: Summary," tech. rep., Japanese International Cooperation Agency, Mar. 2014.

[2] D. Beymer, P. McLauchlan, B. Coifman, and J. Malik, "A Real-time Computer Vision System for Measuring Traffic Parameters," in *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference*, pp. 495–501, IEEE, 1997.

[3] R. Rad and M. Jamzad, "Real Time Classification and Tracking of Multiple Vehicles in Highways," *Pattern Recognition Letters*, vol. 26, pp. 1597–1607, July 2005.

[4] Z. Zhang, T. Tan, K. Huang, and Y. Wang, "Three-DimensionalDeformable-Model-Based Localization and Recognition of Road Vehicles," *IEEE Transactions on Image Processing*, vol. 21, pp. 1–13, Jan. 2012.

[5] J.-P. Jodoin, G.-A. Bilodeau, and N. Saunier, "Urban Tracker: Multiple Object Tracking in Urban Mixed Traffic," in *Applications of Computer Vision (WACV), 2014 IEEE Winter Conference on*, pp. 885–892, IEEE, 2014.

[6] E. Nateghinia and H. Moradi, "Video-based Multiple Vehicle Tracking at Intersections," in *Robotics and Mechatronics (ICRoM), 2014 Second RSI/ISM International Conference on*, pp. 215–220, IEEE, 2014.

[7] H. Yu, R. Bin, K. Yang, S. Li, and J. Zhang, "Video-based Traffic Data Collection System for Multiple Vehicle Types," *IET Intelligent Transport Systems*, vol. 8, pp. 164–174, Mar. 2014.

[8] K. McBride, "Vehicle Tracking in Occlusion and Clutter." 2008.

[9] M. de Feo, A. Graziano, R. Miglioli, and A. Farina, "IMMJPDA versus MHT and Kalman Filter with NN Correlation: Performance Comparison," in *IEE Proceedings Radar, Sonar and Navigation*, vol. 144, pp. 49–56, IEEE, Apr. 1997.

[10] L. Wen, D. Du, Z. Cai, Z. Lei, M. Chang, H. Qi, J. Lim, M. Yang, and S. Lyu, "DETRAC: A New Benchmark and Protocol for Multi-Object Tracking," 2015. http://arxiv.org/abs/1511.04136.

[11] K.-H. Lee, J.-N. Hwang, and S.-I. Chen, "Model-Based Vehicle Localization Based on 3-D Constrained Multiple-Kernel Tracking," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, pp. 38–50, Jan. 2015.

[12] W. Luo, X. Zhao, and T. Kim, "Multiple Object Tracking: A Review," 2014. http://arxiv.org/abs/1409.7618.

[13] Y. Hao, Y. Yin, and J. Lan, "Vehicle Tracking Algorithm Based on Observation Feedback and Block Symmetry Particle Filter," *JECE*, vol. 2014, pp. 4:4–4:4, Jan. 2014.

[14] M. Yildirim, I. Ince, Y. Salman, J. Song, J. Park, and B. Yoon, "Direction-based Modified Particle Filter for Vehicle Tracking," *ETRI Journal*, vol. 38, pp. 356–365, 2016.

[15] B. Coifman, D. Beymer, P. McLauchlan, and J. Malik, "A Real-time Computer Vision System for Vehicle Tracking and Traffic Surveillance," *Transportation Research Part C: Emerging Technologies*, vol. 6, no. 4, pp. 271–288, 1998.

[16] J.-C. Tai, S.-T. Tseng, C.-P. Lin, and K.-T. Song, "Real-time Image Tracking for Automatic Traffic Monitoring and Enforcement Applications," *Image and Vision Computing*, vol. 22, pp. 485–501, June 2004.

[17] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati, "Detecting Objects, Shadows and Ghosts in Video Streams by Exploiting Color and Motion Information," in *Image Analysis and Processing, 2001. Proceedings. 11th International Conference on*, pp. 360–365, IEEE, 2001.

[18] K. Bernardin and R. Stiefelhagen, "Evaluating Multiple Object Tracking Performance: The CLEAR MOT Metrics," *EURASIP Journal on Image and Video Processing*, vol. 2008, pp. 1–10, 2008.

[19] A. Milan, L. Leal-Taixé, I. D. Reid, S. Roth, and K. Schindler, "MOT16: A benchmark for multi-object tracking," *CoRR*, vol. abs/1603.00831, 2016.

[20] B. Wu and R. Nevatia, "Tracking of Multiple, Partially Occluded Humans based on Static Body Part Detection," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, vol. 1, pp. 951–958, June 2006.

[21] A. Yilmaz, O. Javed, and M. Shah, "Object Tracking: A Survey," *ACM Computing Surveys*, vol. 38, pp. 13–es, Dec. 2006.

[22] O. Barnich and M. V. Droogenbroeck, "Vibe: A UniversalBackground Subtraction Algorithm for Video Sequences," *IEEE Transactions on Image Processing*, vol. 20, no. 6, pp. 1709–1724, 2011.