Multi-vehicle Convoy Analysis Based on ANPR Data

A. Homayounfar*, A. T. S. Ho*, N. Zhu*, G. Head†, P. Palmer†

*Department of Computing, School of Engineering and Physical Science, University of Surrey, Guildford, Surrey, UK
†Surrey Police, Mount Brown, Guildford, UK

Keywords: Data Mining, ANPR, Benford's Law, Encryption, Convoy Analysis.

Abstract

This paper focuses on the development and novel application of data mining techniques for convoy analysis of vehicles based on the automatic number plate recognition (ANPR) system. The amount of ANPR data captured daily by traffic cameras in the road networks is very substantial. Data mining techniques are commonly used to extract relevant information and to reduce the amount of data processing and storage. In this paper, we apply data clustering techniques to extract relevant traffic patterns from the ANPR data to detect and identify unusual patterns and irregular behaviour of multi-vehicle convoy activities.

1 Introduction

Convoy Analysis of vehicles is a relatively new research area. There currently exist a few publications in this area because of availability and sensitivity of vehicle data. Gopalan and Narayanaswamy first proposed the use of convoy analysis to determine the routes from origin to destinations of two vehicles [1]. Chardair et al presented some results for a number of instances of the convoy movement problem [2]. Position functions and mathematical modelling were proposed for closely coupled convoys [3, 4]; however these functions and models were similar to work based on GPS tracking of vehicles. Software for convoy intelligence has been developed based on data visualisation and analysis [5]. In this paper, we present a number of data mining and detection techniques to filter and extract the relevant ANPR data for convoy analysis of vehicles. We further investigate different traffic models and scenarios for detecting unusual patterns and irregular behaviour of multi-vehicle convoys.

In Section 2, we present data filtering and detection algorithms for convoy analysis based on ANPR data captured from two or more cameras installed in different road network locations. ANPR cameras are typically set at fixed zoom and angle for number plate capturing The huge amount of ANPR data involved for road network has also a direct impact on limiting the number of ANPR cameras used. By estimating speed limits on the road this paper presents an analytical model that can be used to connect vehicles even if the convoy of vehicles near cameras' locations does not occur. Our proposed techniques for convoy analysis are based on real ANPR data provided by Surrey Police from four different camera locations around the major road network in Surrey; and to the best of our knowledge the error of cameras is negligible. To protect data privacy in this paper, all ANPR data are encrypted using a simple substitution method and the true camera locations on the map are also coarsely referenced without revealing their exact location details. We identify a number of potential convoy scenarios arising unusual behaviours and patterns of vehicles. These scenarios may correspond to movements and activities that may assist police investigations.

2 Model analysis and detection algorithm

Even though the number of traffic cameras installed in each road network is relatively small the amount of ANPR images captured per day is quite substantial. For example, on a typical single lane A* road in the county of Surrey, the number of images could be in the region of 50,000. Dual carriageway and motorways would be handling traffic images in the order of two to three times more data volume. Figure 1 shows a reference satellite map with roadwork graphic overlays for our convoy analysis experiments. Due to the huge data volume, it is paramount that the Police are able to compress and more importantly, to extract the relevant ANPR data that could be submitted for further analysis of unusual patterns and activities related to multi-vehicle convoys.

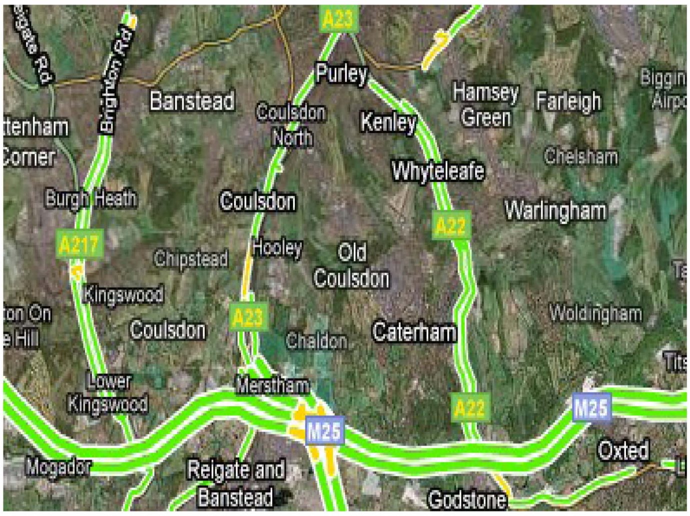

Figure 1: Reference map for analysis in Tables 1-5 of three single lane A* roads and motorway.

In this paper, data sorting, filtering and detection algorithms have been proposed and developed for convoy analysis.

Initial results are encouraging, confirming the functionality and accuracy of the proposed algorithms. The scalability of algorithms is an important consideration during the design and development stage as the algorithm must be able to support more traffic cameras leading to a significant further increase in data volume as well as support for any specific range of time. To achieve a scalable algorithm, two threshold values are introduced to detect single occurrence in different camera locations and two-vector distance (or time difference) among multi-vehicle convoys as illustrated as flow diagram in Figure 2.

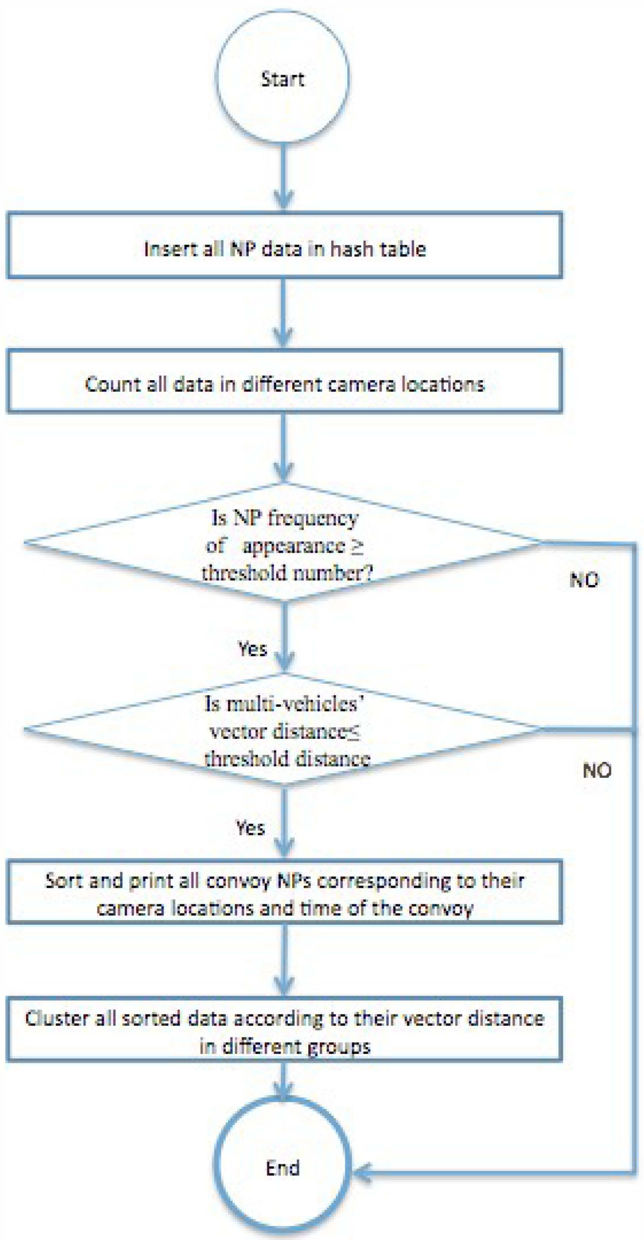

Figure 2: Flow chart of the algorithm. The vector distance among multi-convoys is used for sorting and clustering data in different groups.

The advantage of using the vector distance method compared with the time difference method is that the vector distance always identifies the coupled convoys independently of their time difference detections, which is related to the speed of each vehicle. Hence, unlike the time difference detection, the vector distance is always an efficient way of identifying coupled convoys even during a traffic jam, when it is too complicated to estimate speeds of vehicles. Table 1 shows convoy scenarios for two cars or couple pair travelling in close proximity of each other. The first two characters of the number plates of the couple pair are encrypted as shown in the Table. The specific detection threshold values used in Table 1 for the convoy analysis experiments are three for the frequency of appearance of the NP, and 10 for the multi-vehicle vector distance for all camera locations. In sections 3-5 to detect more challenging scenarios we exceed the number of multi-vehicle blocks to 100 for M25 and 50 for A* lanes, whilst keeping the threshold value for frequency of appearance at three. A proximity time of approximately 30 seconds or less (except in rush hours) is considered to be a reasonable proximity for cars during a typical journey. The two cars identified as a convoy pair in Table 1 have similar number plates, may potentially belong to the same organization, for example, two tour coaches or delivery lorries. We will present more detailed analysis and scenarios of unusual patterns with irregular time of journey and dependency based on static and dynamic analysis of ANPR data in Sections 3 to 5.

NP	Camera Time	Location	Direction
ZA*9*KM	18/**/20**-10:07	A22	South
ZA*8*HN	18/**/20**-10:07	A22	South
ZA*9*KM	18/**/20**-10:19	M25	East
ZA*8*HN	18/**/20**-10:19	M25	East
ZA*9*KM	18/**/20**-11:32	M25	West
ZA*8*HN	18/**/20**-11:32	M25	West
ZA*9*KM	18/**/20**-11:45	A22	North
ZA*8*HN	18/**/20**-11:45	A22	North

Table 1: Couple convoys travelling very close to one another and according to their car number plates may belong to the same organization.

To perform real-time convoy analysis, it is important to ensure a good performance trade-off between detection accuracy and processing speed, given the sheer amount of ANPR data involved on a daily basis. Data mining and extraction techniques of unusual traffic patterns can be efficiently achieved via supervised classification. The hidden patterns among tens of thousands of ANPR data need to be fully understood and classified according to their irregular or unusual behaviours as compared to "ordinary or normal behaviour" vehicles. Once classified, irregular patterns of ANPR data should be stored and continuously updated in a database for further analysis in future as stored multi-vehicle convoy data may help to predict some linkage and hierarchy among those detected cars. This will be discussed in more detail in Section 5.

In this paper, the first data mining method proposed involves two main processes: firstly, the ANPR data are first filtered to remove regular behaviour or activities between

By measuring the distance (D) of the camera locations from the M25 to the A22, $t_{min} \approx 8$ minutes. Based on available online traffic information data, the daily delay during rush hour is estimated at 10-20 minutes [10]. This gives an average time value, t_{AVG}, from one camera location to the next one. t_{AVG} does not exceed more than twice of t_{min}; hence t_{max} at the worst case (for example during traffic jam) for the above four specific locations are $3t_{min} < t_{max}$. Therefore, we can only take into account of vehicles that have appeared as convoy more than the maximum time (unusual patterns). It is evident that a vehicle cannot exceed t_{max}, unless it breaks its journey but when two vehicles both exceed t_{max} and detected by traffic cameras several times in close proximity to each other, then it would clearly highlight a strong linkage between these vehicles.

3.1 Convoy analysis of unusual journey times and locations

As discussed in previous sections, we attempted to analyse only the most probable cases of irregular convoys. This also protects against invasion of privacy for normal vehicles with regular patterns and behaviours. For example, during a daily routine it might not be considered unusual for people who might work in the same organization who happened accidentally to travel close together in the same path around 8-10 am and 5-6 pm. However, a suspected scenario is shown in Table 2, where there were two vehicles detected at three different time periods by traffic cameras outside the normal peak periods by the convoy analysis algorithm. In the following examples by considering the distance of each camera location to the next one, it is estimated that even during rush hours the time of travel does not exceed by three times of t_{AVG}. Therefore, all scenarios mentioned in the following sections, which exceeded this threshold value, are considered as irregular behaviours.

NP	Camera Time	Location	Direction
S*2*MO	02/**/20** 11:24	M25	West
H*71N*W	02/**/20** 11:25	M25	West
S*2*MO	02/**/20** 11:35	A23	North
H*71N*W	02/**/20** 11:36	A23	North
S*2*MO	02/**/20** 14:58	A22	South
H*71N*W	02/**/20** 14:58	A22	South

Table 2: Convoy time of two vehicles measured at three camera locations.

The journey time between traffic cameras A22 and A23 is approximately three and a half hours for both S*2*MO and H*71N*W with $t_{AVG} \approx$ 10-15 minutes (time of travel >14 t_{AVG}). Their close proximity appearing at around the same time periods and locations would certainly indicate a potential relationship between them.

This scenario is not so easy to analyse for a short distance and $t \approx t_{AVG}$ to determine whether or not the two vehicles are convoyed. However, when exceeding time is more than the threshold value t_{max}, and if there is also an unusual journey path that the two vehicles are seen at certain locations repeatedly, even with a couple of minutes delay, then this may be considered for further analysis as it might be more serious than a normal convoy.

Table 3 highlights a convoy scenario of two vehicles that could not have happened accidentally for three camera locations, which are not in the same path (for camera locations, the path of the A22 and A23 are parallel roads and the M25 path intersects these two). The maximum time distance between the A23 and A22, t_{max}, does not exceed 30 minutes, which is less than seven times for the journey of B*51*VJ and J*71*ZY. Hence for J*71*ZY not following B*51*VJ is therefore considered highly improbable.

NP	Camera Time	Location	Direction
B*51*VJ	04/**/20** 11:28	A23	North
J*71*ZY	04/**/20** 11:30	A23	North
B*51*VJ	04/**/20** 14:03	A22	South
J*71*ZY	04/**/20** 14:04	A22	South
B*51*VJ	04/**/20** 14:14	M25	East
J*71*ZY	04/**/20** 14:18	M25	East

Table 3: Example of a car following another at three different times and locations and exceeding t_{AVG} between camera locations at A23 and A22.

4 Dynamic analysis in addition of camera bases

As given in Tables 1-3, the time of analysis is based on static locations of cameras, where the convoy is based on the exact time detected by the ANPR cameras. It is possible to extend the static analysis to dynamic analysis by an approximate time measured for the cars when the convoy is not exactly at the cameras' location.

Table 4.1 contains convoy results similar to results presented in Tables 1-3, which give the time of the two vehicles as V*28*AJ and E*82*ZU passing three traffic cameras on 04/**/20**. From this information it is possible to find out the approximate time and location of the two vehicles when they are a convoy pair within a few miles range of the ANPR cameras. The main difference between this scenario and previous scenarios presented in Sections 2 and 3 is that this convoy also occurs in other locations where there is no existence of any cameras.

NP	Camera Time	Location	Direction
V*28*AJ	10:10	A23	North
E*82*ZU	10:17	A23	North
V*28*AJ	10:20	A217	North
E*82*ZU	10:27	A217	North
V*28*AJ	10:32	A217	South
E*82*ZU	16:46	A22	North
V*28*AJ	16:47	A22	North
V*28*AJ	16:59	A22	South

Table 4.1: Convoy pair in three locations

By knowing the speed limit on the A217 (40 mph) and A22 (30 mph), it is possible to make an approximation of the convoy pair vehicles within a few miles in the proximity of the traffic cameras.

Convoy Time	Convoy Location
t ≈ 10:29 - rendezvous	d<2 miles north of A217 camera
16:46<t<16:47-convoy	A22

Table 4.2:. In addition to the convoy, this provides a potential rendezvous scenario with a domain up to two miles north of the camera location on the A217 road.

5 Hierarchies and clustering among ANRPs

In addition to convoy analysis, the proposed algorithm can also detect some irregular linkage and relationship between the vehicles; this can be considered as a kind of clustering approach to classify ANPR data into different relationship groups. From the experiments and analysis of 200,000 daily ANPR data provided by Surrey Police, only a few convoys fall in this category.

To allow easy search of ANPR data records, a SQL database has been created with each vehicle number plate assigned an ID field that can be used as a primary key to access the data records efficiently in a hierarchical manner. This offers a significant advantage in performing clustering and data mining of multi-vehicle convoys and in predicting unusual or irregular behaviours of suspect vehicles

ID	Encoded NP	Encoded Date -Time	Direction
A	J*27*KA	2/**/20**-21:04	M25-East
B	N*73*SW	2/**/20**-21:28	M25-West
C	U*20*FI	2/**/20**-21:36	M25-East
A	J*27*KA	3/**/20**-03:38	M25-West
C	U*20*FI	3/**/20**-03:40	M25-West
B	N*73*SW	3/**/20**-03:41	M25-East
D	G*79*FH	3/**/20**-03:49	M25-West
A	J*27*KA	3/**/20**-03:50	A22-South
C	U*20*FI	3/**/20**-03:53	A22-South
D	G*79*FH	3/**/20**-05:52	M25-East
B	N*73*SW	3/**/20**-05:53	M25-West

Table 5: Connections among four vehicles in three groups in two continuous days.

From Table 5, it is possible to identify a connection between different vehicles that appeared together as convoy pairs in different days to associate unusual activities via a hierarchy relationship.

6 Conclusions and Future work

This paper presented initial research and results of convoy analysis of ANPR data collected from traffic cameras installed at various road network across Surrey. The work was in collaboration with the Surrey Police who kindly provided the data for the research. The huge amount of ANPR data was first reduced via filtering followed by sorting and identifying relevant vehicle data into convoy sets. A number of suspected convoy scenarios were highlighted that could be considered for further analysis by the Police.

In future work, we will investigate advanced pattern recognition techniques for convoy analysis, as well as continuing the initial work based on the Benford's law for data forensics of accurate detection of multi-vehicle convoys.

References

[1] R. Gopalan and C. D. Narayanaswamy, "Analysis of algorithm for an online version of the convoy movement problem", Journal of the Operational Research Society, **60**, 1230-1236, (2009).

[2] P. Chardair, G. P. McKeown, S. A. Verity-Harrison, and S. B. Richardson, "solving a time-space network formulation for the convoy movement problem", Operations Research, **53 (2)**, 219-230, (2005).

[3] W. Travis and D. M. Bevly, "Trajectory duplication using relative position information for automated ground vehicle convoys," in Proceedings of IEEE/ION Position, Location, and Navigation Symposium, 1022-1032, (May 2008).

[4] W. Travis and D. M. Bevly, "Performance Analysis of a Closely Coupled GPS/INS Relative Positioning Architecture for Automated Ground Vehicle Convoys," Proceedings of The national Technical Meeting -Institute of navigation, VOL 2 999-1008 (2009)

[5] Visual Analysis. Internet: http://www.visualanalysis.com, July 2011

[6] Map Container Objects, MATLAB R2011a Documentation [On-line]. Available: http://www.mathworks.com/help/techdoc/ref/containers_map.html

[7] F. Benford, "The law of anomalous numbers," Proc. Amer. Phil. Soc., **78**, 551–572, (1938).

[8] M. Nigrini. Forensic Analytics: Methods and Techniques for Forensic Accounting Investigations: Wiley Corporate F&A, 2011, 43-44 and 85-98.

[9] X. Zhao, A.T.S. Ho, Y. Q. Shi, "Image Forensics using Generalized Benford's Law for Accurate Detection of Unknown JPEG Compression in Watermarked Images," 16th International Conference on Digital Signal Processing, 5-7 July 2009, Santorini, Greece

[10] Traffic Information: Internet: http://www.highways.gov.uk/traffic/