# Structured behaviour prediction of on-road vehicles via deep forest

Luntian Mou✉, Shasha Mao, Haitao Xie and Yanyan Chen

Vision-based vehicle behaviour analysis has drawn increasing research efforts as an interesting and challenging issue in recent years. Although a variety of approaches have been taken to characterise on-road behaviour, there still lacks a general model for interpreting the behaviour of vehicles on the road. In this Letter, the authors propose a new method that effectively predicts the vehicle behaviour based on structured deep forest modelling. Inspired by structured learning, the structure information of vehicle behaviour is extracted from the detected vehicle, and then the corresponding structured label is constructed. Especially, the structured label visually expresses the vehicle behaviour as contrast to the discrete numerical label. With the structured label, a structured deep forest model is proposed to predict the vehicle behaviour. Experimental results illustrate that the proposed method successfully obtains the implication of semantic interpretation of vehicle behaviour by the predicted structured labels, and meanwhile it achieves comparable performance with traditional methods.

*Introduction:* With the advancement of imaging technologies, decreasing the camera cost and increasing the computing power, there has been significant research effort dedicated to computer vision based detection and tracking of on-road vehichles in recent years, which aims at enhancing safety and management by monitoring the on-road environment [1–3]. It is not unusual for researchers to report reliable vehicle detection and tracking in real time, over extended periods [4, 5]. Therefore, it is time that higher level of understanding of vehicle behaviour should be pursued. Typically, a bottom-up approach based on vision [6, 7] is taken to achieve the goal of semantic interpretation of the on-road environment. At the lowest level, features such as appearance, motion and size are utilised to detect vehicles in an image. At the medium level, data association, temporal coherence and filtering are exploited for tracking. At the highest level, an aggregate of spatiotemporal features accounts for learning, modelling, classifying and predicting the behaviour of vehicles. Typical approaches to behaviour interpretation include a dynamic visual model to categorise observed vehicle behaviour as normal or abnormal [8], using interacting multiple models to characterise the motion of the oncoming vehicle with a transition probability handling change of states [6], and applying a rotation-invariant version of the longest common subsequence as the similarity metric to classify trajectories [9]. Although a variety of approaches have been taken to characterise on-road behaviour, there still lacks a general model for interpreting and predicting the behaviour of vehicles.

For vehicle behaviour prediction, numerical labels are commonly used in methods of supervised learning. However, while each behaviour is assigned as one numerical label, the important intrinsic properties of the corresponding behaviour are spontaneously and naturally ignored. An example is shown in Fig. 1, including six sequential frames from a surveillance video captured in an intersection. Obviously, when a numerical label is used to express the behaviour of turning left, the white car A in six subfigures will be assigned as the label ’0’ and without the variety among intraclass behaviours. Whereas, from Fig. 1, it is obviously observed that when the white car A turns left with small angles, the black car B can go through, but when it turns left with large angles, the red car C will have to wait. Hence, as opposed to numerical labels, structured labels proposed by structured learning are capable of expressing more structure information of the classes, especially the same class with different angles. Structured learning [10, 11] is a generalisation of the standard paradigms of supervised learning, and it is featured by structured labels. Compared with the scalar discrete or continuous label, the structured label conveys more detailed information of the learned output.

At present, deep learning [12] has been widely applied in computer vision, due to its excellent performance which is generally achieved based on large-scale training datasets. However, for vehicle behaviour recognition, it is difficult to collect a vast amount of samples with labels, especially for real-time data. Hence, an excepted performance may not be achieved by deep learning while the size of the training samples of vehicle behaviour is small. Alternatively, a deep forest approach (multi-grained cascade forest) [13] has been proposed to resolve this issue, and it performs well for small-scale training datasets and is trained without GPUs.

**Fig. 1** *Structured labels expressing turning left with different angles. Top row: the white car A turns left with small angles results in a decision of going through by the black car B. Down row: the white car A turns left with large angles results in a decision of waiting by the red car C*

Based on the above analyses, we propose a new approach of vehicle behaviour recognition by combining deep forest with structured labels in this Letter, generalised as structured deep forest (StruDF), in order to gain more information of vehicle behaviour. Specifically, we construct a dataset of vehicle behaviour with structured labels, where the structure information of vehicle behaviour is extracted from the detected vehicles and as corresponding structured labels to visually express the behaviour of on-road vehicles. Based on the structured labels, behaviour analysis of on-road vehicles is formulated as a problem of structured learning, and then the StruDF method is proposed to recognise the vehicle behaviour. To the best of our knowledge, it is the first time to introduce structured learning into the vehicle behaviour analysis and construct a model of StruDF. Finally, experimental results illustrate that the proposed method can gain the visual behaviour information for detected vehicles while achieving a comparable performance compared with the methods using discrete labels.

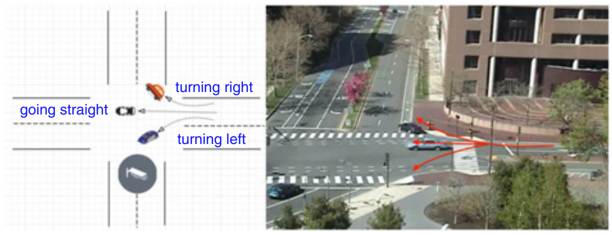


**Fig. 2** *Turning behaviours at one right-to-left way of an intersection. Left: an abstract of the target scene. Right: a real target scene showing going straight, turning right and possibly turning left*

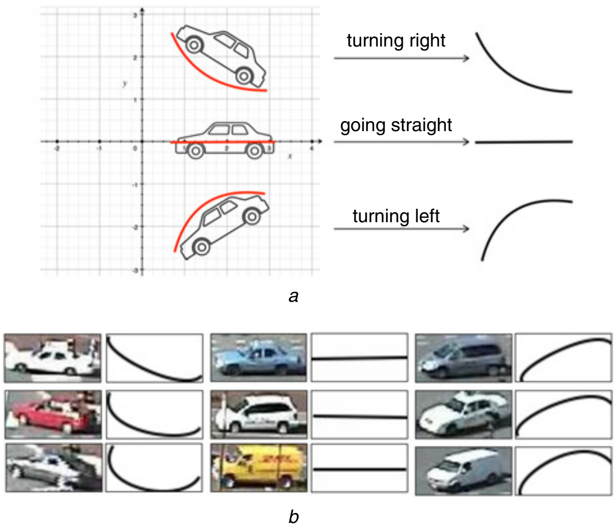


**Fig. 3** *Illustration of structured labels for turning behaviours*

a Drawing curves to represent different turning behaviours
b Nine vehicle examples and corresponding structured labels

*Structured labels of vehicle behaviour:* As discussed above, structured labels can visually illustrate prominent behaviour information of vehicles in typical contexts of on-road environment, thus bear semantic information to some extent. Here, intersections of city roads are taken as an example context, and different turning behaviours are chosen as the targets. As the preliminary exploration, vehicle behaviours, namely, turning left, going straight and turning right, at one right-to-left way of the intersection are analysed and predicted using structured labels (see Fig. 2). From the perspective of human visual system, turning angles are taken as the most prominent feature of turning behaviour.

For constructing the new dataset of vehicle behaviours with structured labels, MIT-CBCL Car Database [14] is employed as the original data. Based on it, first, the locations of vehicles in intersections are manually captured in key frames of an input video, and the images of vehicles are collected. Then, three key points of each detected vehicle, such as approximate centres of headstock, tailstock and middle part, are located in each image of vehicles. Finally, a curve is drawn based on three key points and turning angles. Such curves are regarded as the structured labels to represent corresponding turning behaviours. Fig. 3 shows the example of drawing the structured label and nine samples of vehicle behaviours with structured labels.

*Structured deep forest for vehicle behaviour recognition:* Trajectory is commonly combined with certain probabilistic models to predict vehicle motion [15]. Although using temporal information could help in behaviour analysis, it is computationally expensive, and inefficient in the case that a single frame is enough for behaviour recognition. Therefore, the behaviour prediction is performed on a one-image basis here. Specially, based on the constructed structured labels, Deep forest [13] is extended into the StruDF for vehicle behaviour recognition, and the detail of StruDF is described as follows.

Inspired by structured random forest [11], a two-stage mapping is applied in projecting the structured labels $\mathcal{Y}$ into the discrete ones $\mathcal{C}$. In order to conveniently compute the similarity over $\mathcal{Y}$, an intermediate space $\mathcal{Z}$ is utilised to first define an approximation of the labels $\mathcal{Y}$. Then, $\mathcal{Z}$ is mapped into the discrete space $\mathcal{C}$. As for the first stage projection, $\mathcal{Y}$ is projected into $\mathcal{Z}$ by the transformation $\Phi_1$ to obtain the approximate dissimilarity of $y \in \mathcal{Y}$ by computing Euclidean distance in $\mathcal{Z}$, shown as

$$\Phi_1 : \mathcal{Y} \to \mathcal{Z}. \tag{1}$$

In this projection, $z \in \mathcal{Z}$ is defined as a binary vector that encodes whether every two pixels of $y$ belong to the same or different classification. To reduce the calculation of distance, the strategy that $m$ pairs of pixels are randomly selected is adopted to obtain the binary vector $z$. In particular, $m = 1024$ is experimentally chosen. Similar to [11], principal component analysis (PCA) is employed to reduce the dimensionality of $\mathcal{Z}$. Based on the first projection $\Phi_1$, we achieve the second-stage mapping

$$\Phi_2 : \mathcal{Z} \to \mathcal{C}, \tag{2}$$

which is implemented by using $K$-means to cluster $\boldsymbol{Z}$ into $k$ clusters and $\mathcal{C} \in \{1, \ldots, k\}$. The detailed procedure of the two-stage projection is shown in Algorithm 1.

**Algorithm 1:** The two-stage projection algorithm

1: Input the structured labels set $\mathcal{Y}$, and give the parameter of selecting features $m$, and the number of clustering classes $k$.
2: Obtain the original binary vector $\mathcal{Z}^o$:
  for $i = 1, \ldots, n$
  $z_n^o$: a binary vector encoded based on every two pixels of $y$;
3: Obtain $\mathcal{Z}$ by randomly selecting $m$ elements from $\mathcal{Z}^o$;
4: Reduce the dimensionality of $\mathcal{Z}$ by PCA and obtain the new vector $\mathcal{Z}^r$;
5: Cluster $\mathcal{Z}^r$ into $k$ clusters by using $K$-means and get the discrete;
6: Output the discrete set of labels $\mathcal{C}$: $c \in \{1, \ldots, k\}$.

Based on $\mathcal{C}$ obtained by the two-stage projection, the approximation of the structured labels $\mathcal{Y}$ is involved into the model of gcForest instead of the traditional numerical labels, and StruDF is constructed. In Fig. 4, an example that random forest is given based on the structured labels $\mathcal{Y}$ is shown, where the discrete labels projected by $\mathcal{Y}$ are utilised to calculate the information gain in each node of one tree by

$$I_j = H(\mathcal{X}_j) - \sum_{k \in \{L, R\}} \frac{\left|\mathcal{X}_j^k\right|}{\left|\mathcal{X}_j\right|} H(\mathcal{X}_j^k), \tag{3}$$

where $\mathcal{X}_j$ is the training set with the structured labels $\mathcal{Y}_j$ in the node $j$, and $H(\mathcal{X}_j)$ expresses the Shannon entropy calculated by

$$H(\mathcal{X}_j) = -\sum\nolimits_{c_y} p_{c_y} \log (p_{c_y}), \tag{4}$$

where $p_{c_y}$ is the fraction of elements in $\mathcal{X}_j$ with the approximation label $c$ corresponding to the structured label $y$. For details of gcForest algorithm, refer to [13].

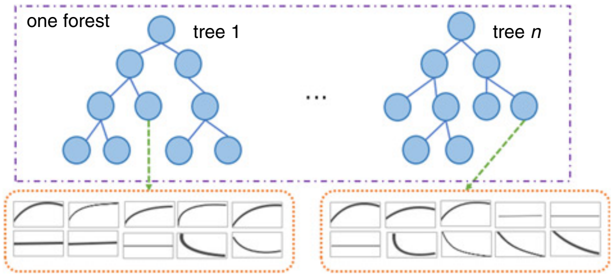


**Fig. 4** *Example of random forest based on structured labels*

*Experiments and analysis:* For collecting the label with structured information, MIT-CBCL Car Database [14] is employed. The MIT traffic dataset is for research on activity analysis and crowded scenes. It holds a traffic video sequence of 90 min long, including vehicle traffic at a four-way signalised intersection. Additionally, 3 h of surveillance videos taken at several intersections in a Chinese city are also used as original datasets. Totally, the new dataset of vehicle behaviour with structured labels includes 900 samples, where the number of samples belonging to turning left, turning right and going-straight is 300, respectively. In the experiment, the samples of each class are divided into two independent parts (80% as the training set and the remaining 20% as the testing set). For gcForest, the same cascade structure as in [13] is used: it consists of 2 complete-random tree forests and 2 random forests in each level, where each forest is composed of 101 trees. Moreover, the parameter of $k$-means is set as 3 or 6 and the dimension reduced by PCA is set as 16.

In Fig. 5, the predictive results of vehicle behaviour obtained by the proposed method (StruDF) are demonstrated, which includes 12 vehicles of turning right, going straight and turning left. It is obviously seen that the proposed method can effectively obtain an accurate behaviour recognition for the detected on-road vehicles. Specifically, the predictive structured labels for vehicle behaviour are highly similar to their corresponding ground truth. Moreover, as demonstrated in Fig. 6, the predicted structured labels can also reflect the relatively small or large angles of turning behaviours. Apparently, the output structured labels are much more meaningful in terms of visual semantic understanding than discrete numerical labels.

**Fig. 5** *Predictive results of vehicle behaviours. The first and fourth columns show the original images of vehicles, the second and fifth columns show the structured labels (ground truth), and the third and sixth columns show the predictive results obtained by the proposed method*

Additionally, in order to validate the performance of the proposed method, the accuracy of vehicle behaviour recognition is evaluated on the testing set. It is worth noting that due to different metrics and datasets used by different approaches, it is difficult to compare the performance of the proposed method with other approaches. Instead, different configurations of structured forest are involved in the evaluation as well as corresponding discrete ones, including deep forest (discrete), random forest (discrete) and structured random forest. Specifically, all methods are implemented assuming the number of classes $c$ to be 3 and 6, respectively. The reason for using $c = 6$ is that it is more elaborate than $c = 3$ for the structured labels, where each class of $c = 3$ will be subdivided into two classes. The experimental results are shown in Table 1.

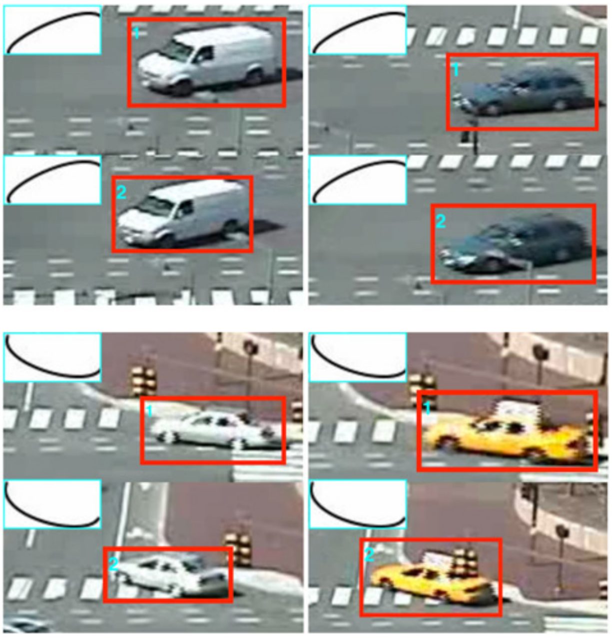

**Fig. 6** *Predictive results of turning behaviours with different angles*

**Table 1:** Results (mean $\pm$ std) of 20 accuracies of behaviour recognition

| Classes | StruDF | Deep Forest | Random Forest | |
|---|---|---|---|---|
| | | | Discrete | Structured |
| $c = 3$ | 81.56 ± 0.09 | 82.45 ± 0.07 | 85.29 ± 0.05 | 84.67 ± 0.09 |
| $c = 6$ | 50.42 ± 0.17 | 62.68 ± 0.22 | 66.27 ± 0.15 | 52.41 ± 0.09 |

In Table 1, the mean $\pm$ std of 20 accuracies obtained by three methods are shown, where ‘StruDF’ represents the proposed method, ‘Deep Forest’ expresses the deep forest with the discrete labels, and ‘Discrete’ and ‘Structured’ are with discrete labels and structured labels, respectively. It can be found that the performance of the proposed method is slightly lower than but comparable to that of Deep Forest with discrete labels, which implies the structured label is feasible for vehicle behaviour recognition besides its advantage of being visually meaningful. Whereas, it is also found that the performance of Random Forest is better than Deep Forest both with discrete and structured labels, which may be caused by small size of training samples. In relation to the number of classes concerned, the performance declines heavily as $c$ changes from 3 to 6 due to increase of errors in the case of finer classification.

*Conclusion:* In this Letter, we propose a structured learning approach for behaviour understanding of on-road vehicles via combining deep forest with structured labels, which exploit the prominent visual information of vehicle behaviours. Experimental results illustrate that the proposed method can obtain comparable performance of recognition accuracy with traditional methods, which validates the effectiveness of structured labels in vehicle behaviour recognition. In the future, we will further explore the structured labels in learning more complex and various vehicle behaviours and understanding better the driving environments. Moreover, we will build a large-scale benchmark dataset and corresponding metrics for evaluating vehicle behaviour analysis methods and systems as well.

*Acknowledgments:* This work was supported by the National Natural Science Foundation of China (grant nos. 61672068 and 61806154).


Submitted: *05 February 2019* E-first: *21 March 2019*
doi: 10.1049/el.2019.0472
One or more of the Figures in this Letter are available in colour online.

Luntian Mou, Haitao Xie and Yanyan Chen (*Beijing Key Laboratory of Traffic Engineering, Beijing University of Technology, Beijing, People's Republic of China*)

✉ E-mail: ltmou@bjut.edu.cn

Shasha Mao (*School of Artificial Intelligence, Xidian University, Xi'an, People's Republic of China*)

## References

1 Shirazi, M.S., and Morris, B.T.: ‘Looking at intersections: A survey of intersection monitoring, behavior and safety analysis of recent studies’, *Trans. Intell. Transp. Syst.*, 2017, **18**, (1), pp. 4–24
2 Sivaraman, S., and Trivedi, M.M.: ‘Looking at vehicles on the road: a survey of vision-based vehicle detection, tracking, and behavior analysis’, *Trans. Intell. Transport. Syst.*, 2013, **14**, (4), pp. 1773–1795
3 Buch, N., Velastin, S.A., and Orwell, J.: ‘A review of computer vision techniques for the analysis of urban traffic’, *Trans. Intell. Transp. Syst.*, 2011, **12**, (3), pp. 920–939
4 Wen, X.Z., Shao, L., Fang, W., *et al.*: ‘Efficient feature selection and classification for vehicle detection’, *Trans. Circuits Syst. Video Technol.*, 2015, **25**, (3), pp. 508–517
5 Yan, G., Yu, M., Yu, Y., *et al.*: ‘Real-time vehicle detection using histograms of oriented gradients and adaBoost classification’, *Optik-Int. J. Light Electron Opt.*, 2016, **127**, (19), pp. 7941–7951
6 Barth, A., and Franke, U.: ‘Tracking oncoming and turning vehicles at intersections’. Proc. 13th Int. IEEE Conf. on Intelligent Transportation Systems (ITSC), Funchal, Portugal, September 2010, pp. 861–868
7 Sivaraman, S., Morris, B.T., and Trivedi, M.M.: ‘Learning multi-lane trajectories using vehicle-based vision’. Proc. IEEE Int. Conf. Comput. Vision Workshop, Barcelona, Spain, November 2011, pp. 2070–2076
8 Cherng, S., Fang, C.-Y., Chen, C.-P., *et al.*: ‘Critical motion detection of nearby moving vehicles in a vision-based driver-assistance system’, *Trans. Intell. Transp. Syst.*, 2009, **10**, (1), pp. 70–82
9 Hermes, C., Wohler, C., Schenk, K., *et al.*: ‘Long-term vehicle motion prediction’. Proc. IEEE Intelligent Vehicles Symp., Xi'an, China, June 2009, pp. 652–657
10 Nowozin, S., and Lampert, C.H.: ‘Structured learning and prediction in computer vision’, *Found. Trends Comput. Graphics Vis.*, 2011, **6**, pp. 185–365
11 Dollar, P., and Lawrence Zitnick, C.: ‘Fast edge detection using structured forests’, *Trans. Pattern Anal. Mach. Intell.*, 2015, **37**, pp. 1558–1570
12 Yann, L., Bengio, Y., and Hinton, G.: ‘Deep learning’, *Nature*, 2015, **521**, (7553), p. 436
13 Zhou, Z.H., and Feng, J.‘Deep forest: Towards an alternative to deep neural networks’, 2017, arXiv preprint arXiv:1702.08835
14 Shirazi, M.S., and Morris, B.: ‘Contextual combination of appearance and motion for intersection videos with vehicles and pedestrians’, International Symposium on Visual Computing (ISVC), Las Vegas, NV, USA, December 2014, pp. 708–717
15 Wiest, J., Hoffken, M., Kresel, U., *et al.*: ‘Probabilistic trajectory prediction with Gaussian mixture models’. Proc. IEEE IV Symp., Alcala de Henares, Spain, June 2012, pp. 141–146