Intelligence Testing for Autonomous Vehicles: A New Approach

Li Li, *Senior Member, IEEE*, Wu-Ling Huang, Yuehu Liu, *Member, IEEE*, Nan-Ning Zheng, *Fellow, IEEE*, and Fei-Yue Wang, *Fellow, IEEE*

***Abstract*—In this paper, we study how to test the intelligence of an autonomous vehicle. Comprehensive testing is crucial to both vehicle manufactories and customers. Existing testing approaches can be categorized into two kinds: scenario-based testing and functionality-based testing. We first discuss the shortcomings of these two kinds of approaches, and then propose a new testing framework to combine the benefits of them. Based on the new semantic diagram definition for the intelligence of autonomous vehicles, we explain how to design a task for autonomous vehicle testing and how to evaluate test results. Experiments show that this new approach provides a quantitative way to test the intelligence of an autonomous vehicle.**

***Index Terms*—Autonomous vehicles, intelligence testing.**

I. Introduction

AUTONOMOUS vehicles usually refers to self-driving vehicles that can fulfill main transportation capabilities of a traditional vehicle [1]–[4]. Such vehicles are viewed as a promising answer to traffic congestion, accident and pollution problems that disturb people around the world; because the movements of vehicles could be controlled in a smoother, safer and economical manner, if the vehicles are built "intelligent" enough.

To reach this goal, some prototypes of autonomous vehicles had been designed and tested during the last few decades. For example, Google and Tesla had demonstrated their autonomous cars can run on road recently [5]–[6]. More companies claimed that they will have their own autonomous vehicles running on road within the next 5 years.

An important question naturally arises as: "*how could we prove an autonomous vehicle is capable to drive in live traffic?*" The accidents that were made by not fully tested Honda cars [7] had demonstrated the disastrous consequences of improper testing. So, unless their reliability and safety can be thoroughly tested and ensured, autonomous vehicles cannot be put into market.

To find an answer, the Defense Advanced Research Projects Agency (DARPA) had sponsored a series of competitions for autonomous vehicles [8]–[9]. The first two "Grand Challenges" had been held in 2004 and 2005 to check whether autonomous vehicles could travel long distances in off-road terrain. The third "Grand Challenges" had been held in 2007 to foster innovation in autonomous driving in busy urban environments [10]–[11]. These tests fired researchers with keenness for autonomous driving.

Similar autonomous vehicle competitions had also been held in Europe and China. National Science Foundation of China had spent over 30 million dollars to support seven "Intelligent Vehicle Future Challenges" that had been held in different cities of China, through 2009 to 2015 [12]. Several prototype vehicles had successfully passed these competition tests.

Notice that competition tests cannot replace real road tests, several countries allowed autonomous vehicles to be tested on ordinary roads since 2010s [13]–[17]. This had triggered a debate on whether it is safe to allow under-testing autonomous vehicles runs into live traffic.

The origin of this debate lies in a long-term hassle to current autonomous vehicle research: *there is no clear definition for "intelligence" of autonomous vehicles*. Generally, intelligence of autonomous vehicles can be viewed as a specialized subfield of artificial intelligence. It can be more generally described as the ability of autonomous vehicles to perceive information, retain knowledge, and adopt adaptive behaviors within an environment [1], [18]. The intelligence of autonomous vehicles has its unique meanings and refers to a set of methodologies dedicated to implement the capabilities of a traditional vehicle.

There were already some initial discussions on autonomous vehicle testing [19]. However, performing tests that produce quantitative, repeatable and comparable results remains challenging for autonomous vehicles, since we do not have a detailed and testable definition of intelligence of autonomous vehicles.

To solve this problem, we provide an exhaustive discussion of intelligence testing for autonomous vehicles in this paper. In the following Section II, we first show that existing testing approaches can be categorized into two kinds: scenario-based testing and functionality-based testing. After discussing the shortcomings of these two kinds of approaches, we propose a new intelligence testing framework to combine the benefits of them in Section III. In Section IV, we discuss how to build adaptive simulation platform for autonomous vehicle testing, using parallel transportation systems. In Section V, a numerical

Manuscript received December 17, 2015; revised March 23, 2016; accepted August 16, 2016. Date of publication September 19, 2016; date of current version December 5, 2016. This work was supported in part by the National Natural Science Foundation of China under Grant 91520301. *(Corresponding author: Wu-Ling Huang.)*

L. Li is with the Department of Automation, Tsinghua University, Beijing 100084, China, and also with Jiangsu Province Collaborative Innovation Center of Modern Urban Traffic Technologies, Nanjing 210096, China (e-mail: forcelee@gmail.com).

W.-L. Huang and F.-Y. Wang are with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100080, China (e-mail: wuling.huang@ia.ac.cn; feiyue.wang@ia.ac.cn).

Y. Liu and N.-N. Zheng are with Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: liuyh@mail.xjtu.edu.cn; nnzheng@mail.xjtu.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIV.2016.2608003

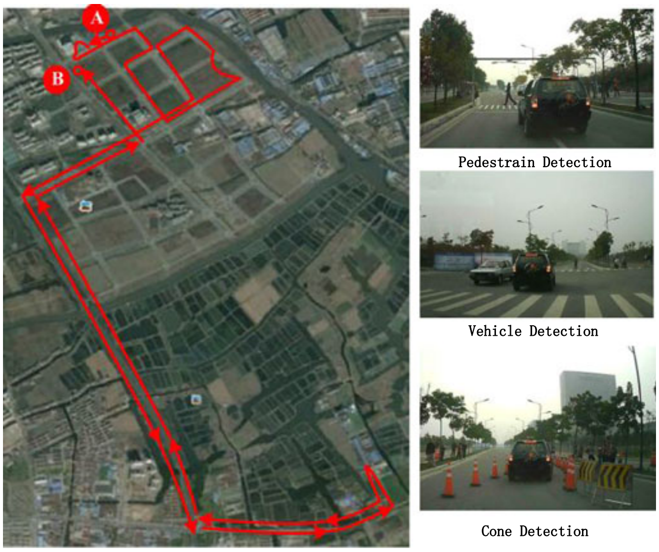


Fig. 1. A scenario-based autonomous vehicle competition which took place in Changshu, China, Nov. 16, 2015. The testing vehicles need to pass several intersections, a tunnel and a working zone, where pedestrians and vehicles may appear.

testing example is provided to show how to design the tasks and how to score the vehicles. Finally, we conclude our findings in Section VI.

II. Existing Intelligence Testing Approaches

A. Scenario-Based Testing

In many tests like Grand Challenge and Urban Challenge, an autonomous vehicle is required to pass a special region safely and legally within a limited time. To reach this goal, a vehicle needs to recognize not only the relatively static environment (e.g. road and static obstacles) but also the dynamic traffic participators (e.g. other vehicles and pedestrians). It should also be able to make the right decision and apply appropriate actions to correctly interact with the environment and traffic participators. Those vehicles which cannot passed the specially formulated traffic scenarios will be taken as NOT "intelligent" enough; see Fig. 1 for an illustration.

The assumption behind such tests is: an autonomous vehicle which had passed the test should be able to successfully run through similar regions.

However, such kind of tests have the following critical shortcomings:

1) The detailed recognition, decision and action behaviors inside a vehicle were neither accurately observed nor fully evaluated. The judgment of driving intelligence is purely based on the outside, Boolean type measurements of the travel. Such indirect measure prevents us from quantitatively assessing the intelligence of an autonomous vehicle.
2) The available traffic scenarios for testing are always limited, although researchers had built some testing ground like M-City to include as many traffic scenarios as possible [20]. Several approaches took years of on-road tests in order to cover all the possible traffic scenarios for autonomous vehicles [5]. This method can ensure the safety of autonomous vehicles but is also time-consuming and economic expensive.
   Since we cannot enumerate all the possible traffic scenarios, we need to set up a theoretical model to analyze the intelligence of an autonomous vehicle from its performance in a limited number of tests. To the best of our knowledge, no reports had discussed this problem.
3) A further problem is how to design a series of tests to evaluate the intelligence of autonomous vehicles in an incremental style, because most prototypes of autonomous vehicles need to be gradually improved.
   However, no studies had set up a theoretical model to study both the complexity of testing scenarios and intelligence of a vehicle simultaneously. As a result, how to design the testing scenarios with respect to the level of driving intelligences also received few discussions in existing studies.
4) A special round of field test is hard to reproduce, if a large number of vehicles are involved. Running tests usually requires expensive hardware and long time. If we want to accurately repeat a special driving test for a certain times, the arrangement costs of the tests would be too high.

Fig. 2. An illustration on a category of vehicle vision based recognition functions: including road boundary detection, lane detection, traffic sign detection, vehicle detection, and etc.

B. Functionality-Based Testing

Analogous to human drivers, functionality-based testing approaches categorized the components of driving intelligence for autonomous vehicles into three parts [21]–[22]: sensing/ recognition functionality, decision functionality according to the recognized information, and action functionality with respect to the decision.

For example, the vision based recognition functionality can be further decomposed into some detailed intellective functions [2], [23]–[24], including road surface recognition, lane marking recognition, vehicle recognition, traffic sign recognition, traffic light recognition, pedestrian recognition, obstacle recognition, etc.; see Fig. 2 for an illustration.

From the viewpoint of functionality-based testing, a fully autonomous vehicle should have all the functionalities that had been retrieved analogous to human drivers.

The most important benefit of functionality-based test is that we could quantitatively evaluate a part of driving intelligence within some specially designed tests. This benefit cannot be easily obtained in traffic scenario-based tests.

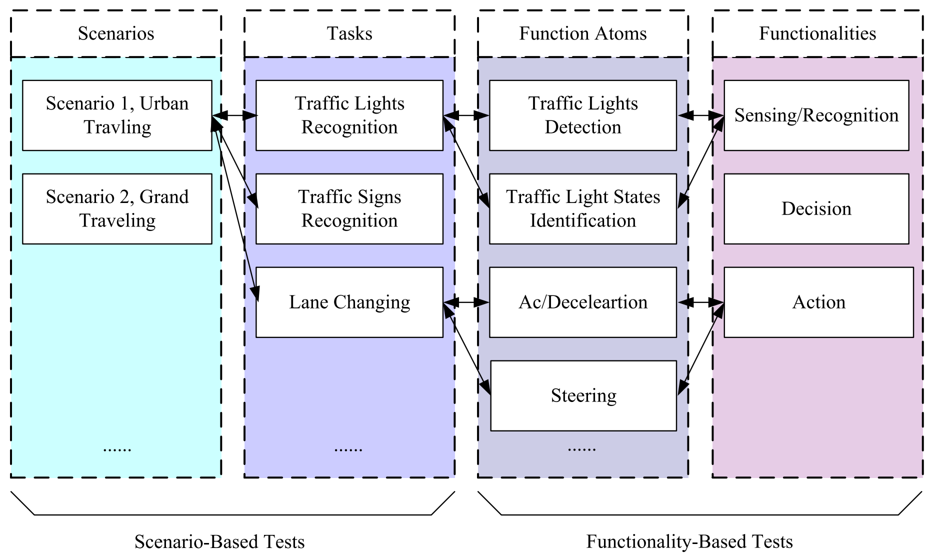


Fig. 3. An illustration of the semantic diagram for driving intelligence of autonomous vehicles.

However, current functionality-based tests also have some critical shortcomings.

1) Many existing functionality-based tests are carried out separately and independently. Only a special functionality was tested in a test. The lack of comprehensive vehicle intelligence tests degrades the reliability of such approaches.
2) We lack standardized benchmark databases that ensuring fair comparisons for functionality-based tests of autonomous vehicles [24]–[25]. The testing scores that claimed in some reports cannot be repeated by others. Though there were some initial attempts to build publicly access benchmark data [26]–[27], the lack of continuous financial support leads to slow progress in building standard test which and hinders the boost of autonomous vehicle research.

III. A New Definition of Driving Intelligence and An Integrated Testing Framework

A. *A Semantic Diagram Definition for Driving Intelligence of Autonomous Vehicles*

In our viewpoint, both scenario-based test and functionality-based test only focus on a part of driving intelligence of autonomous vehicles, respectively. To integrate them, we set up a semantic diagram to define the driving intelligence as shown in Fig. 3.

The hidden links between scenario-based and functionality-based tests are the testing tasks. As shown in the left part of Fig. 3, to pass any a special traffic scenario successfully, a vehicle needs to finish a series of tasks successfully. Here, a task refers to an activity that needs to be accomplished within a limited period of time.

For example, in an urban traveling task, an autonomous vehicle should be able to find a valid path between the origin and the destination. It should also correctly recognize the road markings, traffic signs and traffic lights at certain time/location when it moves along the selected path, so as to move on the right lane without violating traffic laws. It should also be able to recognize the vehicles and obstacles that may interface with its movement and make proper actions to avoid collisions.

As shown in the middle part of Fig. 3, it usually requires an autonomous vehicle to perform several function atoms to fulfil any a single task. For instance, to recognize a traffic light, we must detect the traffic lights from either a single image or a video; and also correctly identify the status of a traffic light.

As indicated in the right part of Fig. 3, the function atoms can be group into three major categories: sensing/recognition functionality, decision functionality and action functionality.

Existing scenario-based tests only addressed the scenario-task relationship on the left side of this semantic diagram; while existing functionality-based tests focused on the task-function relationship on the right side of this semantic diagram.

To give a comprehensive definition, the role of tasks in driving intelligence definition should be emphasized. Actually, some early attempts already defined an intelligent vehicle according to the tasks that it could complete [28]–[29], [2].

On one hand, a vehicle's performance in a task can be accurately measured (see the numerical example added in the revised paper for an example). This solves the problem that the intelligence of a vehicle cannot be quantitatively measured if only scenarios are considered.

On the other hand, a vehicle's performance in a task depends on the capabilities of executing some functions. The scores of the tasks reflect the performance of the functional units. So, we can also appropriately evaluate the interested functional units by designing some special tasks.

In summary, we suggest to use this semantic diagram as a better definition for the intelligence of autonomous vehicles. Moreover, this definition meanwhile delineates how to design and implement tests to evaluate the intelligence of autonomous vehicles.

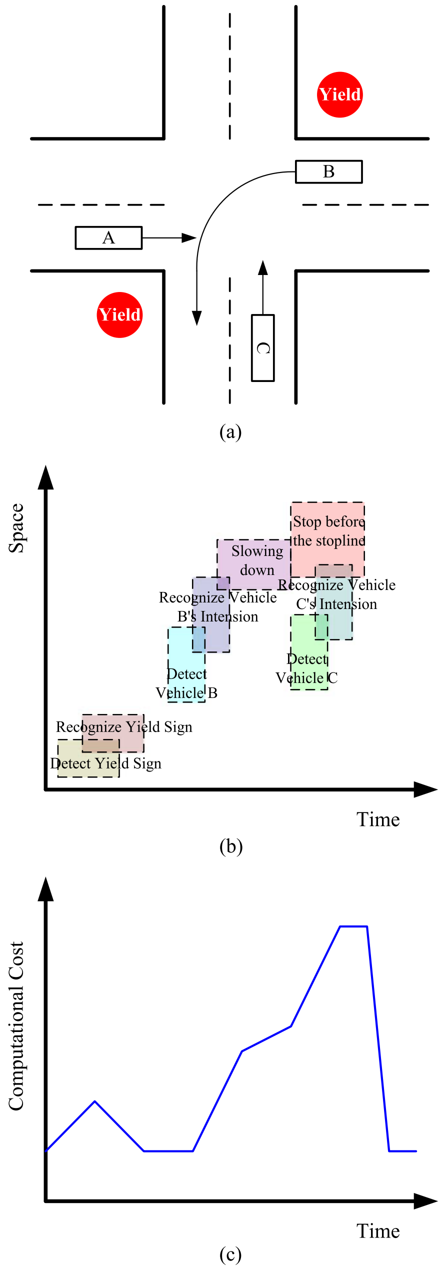


Fig. 4. (a) a typical urban driving scenario, (b) temporal-spatial plot of the assigned tasks, and (c) the corresponding computational cost required along the time.

B. Testing Task Generation for Autonomous Vehicles

Based on the above semantic diagram, we can see that the core of test design for autonomous vehicles is to settle a series of tasks that needs to be finished sequentially within some certain temporal-spatial ranges.

For example, Fig. 4(a) shows a typical urban driving scenario in which three vehicles approach an intersection almost in the same time. *Vehicle B* will arrive a little bit earlier than *Vehicle A* and *Vehicle C* will arrive a little bit later than *Vehicle A*, if no vehicle stops. Both *Vehicle A* and *Vehicle C* aim to pass through the intersection straightly; while *Vehicle B* aims to make a left turn. Suppose *Vehicle A* is the autonomous vehicle under testing. According to the traffic law of many continental countries, Vehicle A should slow down to stop and wait until *Vehicle B* and *Vehicle C* had passed the intersection.

Fig. 4(b) shows the corresponding tasks that autonomous *Vehicle A* needs to fulfill before *Vehicle A* starts to pass through the intersection, in a temporal-spatial plot style. Each task is characterized as a rectangle whose left vertical boundary denotes the minimal possible time that a task is delivered for the vehicle. Its right vertical boundary gives the maximal allowable time that a task must be finished. Similarly, the horizontal boundaries of a rectangle denote the spatial range of a task. As shown in Fig. 4(b), the temporal-spatial range of a task may be overlapped with these of other tasks.

Using this method, we can abstract a test design process into four steps:

1) Select a number of functions that need to be test;
2) Generate a number of tasks that correspond to these functions, respectively.
3) Assign the temporal-spatial ranges of these tasks within a pre-selected temporal-spatial scope, with considerations of the occurring order of some task pairs.
4) Generate the associated traffic scenario according to the temporal-spatial ranges of the assigned tasks. For example, we need to arrange the positions and speeds of *Vehicle B* and *Vehicle C* to trigger the tasks shown in Test of Fig. 4.

In short, the above design process indeed means that we transverse from the right side of the semantic diagram shown in Fig. 3 to the left side of the semantic diagram, in each design process of intelligence tests for autonomous vehicles.

We can see that the complexity of a test within a given temporal-spatial scope is determined by the following factors:

First, the number of tasks in this scope and the difficulties of each task. These two factors control the overall computational costs of the whole testing traffic scenario. In other words, the difficulty of a task is controlled by the difficulties of the functions that need to be performed to finish this task.

Second, the numbers of concurrent tasks at a particular time. This factor characterizes the instantaneous computational cost of the testing traffic scenario at each time. In other words, two testing scenarios may have different levels of difficulties, although they have the same sets of tasks.

So, we can quantitatively analyze and control the complexity of a test by studying the features of temporal-spatial range plot for its tasks. This method provides us a powerful tool to design tests for driving intelligence of autonomous vehicles.

Furthermore, this method also enables us to describe the spaces spanned by all the possible traffic scenarios. As shown in Fig. 5, in the first level, we can categorize the scenario space according to the number and types of tasks in each scenario. Then in the second level, we further divide each category in the first level according to the temporal-spatial ranges of the assigned tasks. In the third level, we can further divide each category in the second level according to the types of instances. Finally, we can link each component in the third level to the instances of traffic scenarios that exist in the scenario space.

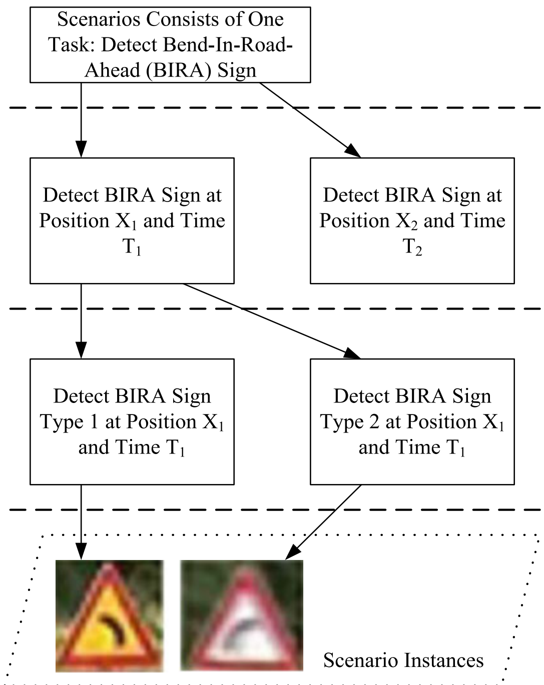


Fig. 5. An illustration of the division for the scenario space.

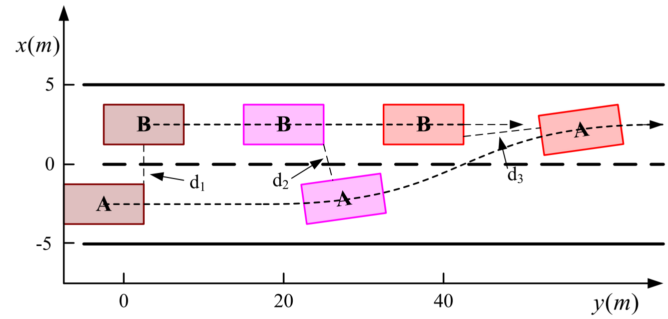


Fig. 6. An illustration of how to calculate the safety score. The shortest distance between two vehicles is d2. This value will be used to calculate the safety score of the test.

On the basis of this category methodology, we could design appropriate strategies to sample the infinite scenario instances existing in the scenario space, so as to keep a good balance between test costs and test coverage. Constrained by the length limit of this paper, we will write a dedicated paper to discuss this issue.

C. Task and Functionality Evaluation

We can set up a series of criteria to evaluate the intelligence of an autonomous vehicle according to its outputs in a task.

The first kind of criteria are Boolean type. They are used to determine whether the vehicle has successfully fulfilled each assigned task within the appropriate temporal-spatial ranges, and without violating any traffic laws.

The second kind of criteria are numerical type. They are used to measure vehicles' output according to a few performance indices, e.g. the 3S scores: smoothness, safety, and smartness.

The smoothness score measures the mechanic output of a vehicle and also indicates the feelings of human passengers. It contains two parts. First, the longitudinal ac/deceleration rates and the jerks (derivative of ac/deceleration rates) of vehicle along time are measured to check whether the vehicle has made too many sudden changes of speed in the task; because inappropriate speed changes increase both oil consumption and pollution, and meanwhile make the travel less comfortable. Second, the rotating rates of vehicle along time are checked for a similar reason.

The safety score is mainly calculated according to the risk/danger level that a vehicle has encountered during the travel in the assigned traffic scenario. For instance, Fig. 6 shows a simple traffic scenario where an autonomous *Vehicle A* needs to make lane change with respect to *Vehicle B*. The safety score in this paper is characterized by the minimum distance between two vehicles during the lane change and the minimum distance to road boundary during the lane change. The closer the two vehicle are, the smaller the safety score is. The closer *Vehicle A* moves toward road boundaries, the smaller the safety score is.

For a more complex traffic scenario, we need to track the relative distances between the testing vehicle and any another component (e.g. vehicle, bicycle, pedestrian, obstacle, stop line) in this traffic scenario and find the most dangerous situation to give a safety score.

The smartness score evaluates the higher level performance of autonomous vehicle. It at least considers three criteria:

First, the time used to fulfil a task. For example, *Vehicle A* detects a traffic sign 2 seconds earlier than *Vehicle B* may receive a higher smartness score. This factor indeed reflects the quality of the algorithms that are designed to implement the function.

Second, the time to pass through a traffic scenario. This factor actually gives a direct, rough and overall summary on a vehicles' comprehensive capability. Due to its simplicity, this criterion is wildly used in autonomous vehicle competitions.

Third, the efficiency of some detailed plans (trajectories) that an autonomous vehicle generates to fulfil a task. For example, *Vehicle A* may execute a very complicated path to adjust its positions and finally park into a pre-selected parking bay; while *Vehicle B* may reach the same goal after a much fewer steps. Then, *Vehicle B* will get a higher smartness score in this task than *Vehicle A*.

It must be pointed out that there is still not a uniform formula to calculate the overall score. Based on the observed vehicle trajectories, researchers had proposed some other performance indices [30]–[32]. Designers may have different viewpoints on the ranks among these scores, too. So, we do not define a uniform way to score the task here.

Based on all these scores, we can transverse from the left side of the semantic diagram shown in Fig. 3 to the right side of the semantic diagram, in order to help quantitatively improve the algorithms that implement the functionalities.

Based on the aforementioned standardized task generation procedures and task evaluation criteria, we could compare

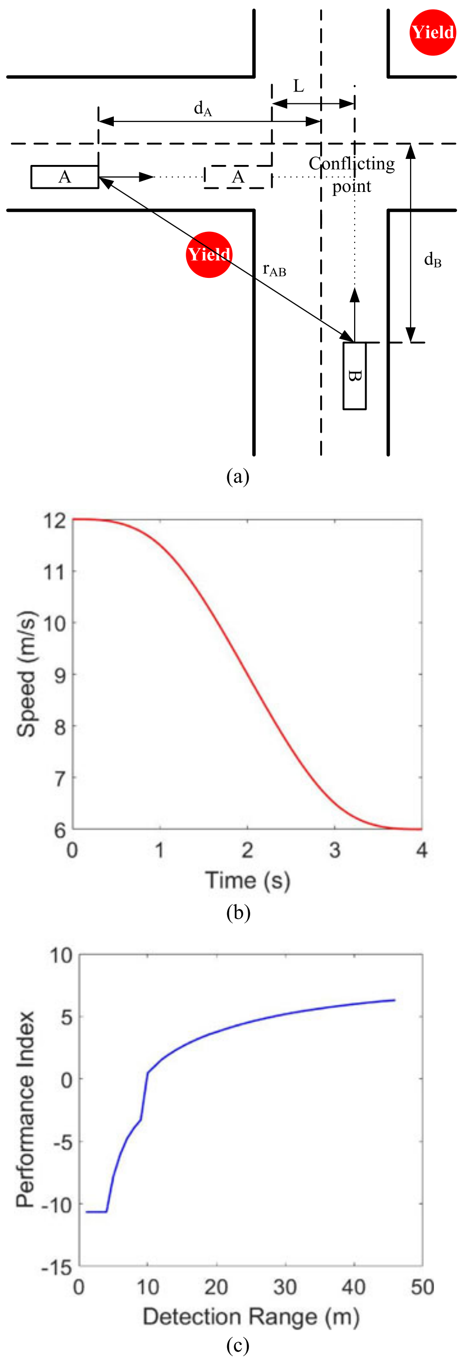


Fig. 7. (a) a typical urban driving scenario, where the initial distance of Vehicle A and Vehicle B are pre-selected. (b) an illustration on the change of the jerk for Vehicle A when it slows down. (c) the corresponding scores of Vehicle A obtained in the simulation test, with the change of detection range of Vehicle A.

the algorithms proposed by different researchers in open and fair mode.

IV. A Numerical Example

In this section, we present an example of simulation test shown in Fig. 7(a), suppose two vehicles are approaching an intersection almost in the same time. Both *Vehicle A* and *Vehicle B* aim to pass through the intersection straightly. Suppose *Vehicle A* is the autonomous vehicle under testing. According to the traffic law of many continental countries, *Vehicle A* should slow down to stop and wait until *Vehicle B* had passed the intersection.

In this test, we focus on testing the vehicle detection range of *Vehicle A* and check the influence of this function to the driving task. As shown in Fig. 7(a), the detection range r_{AB} means how far *Vehicle A* can detect *Vehicle B*. This is a specially designed scenario that contains only one task. So, the score of the task equals to the score of the overall scenario.

Vehicle B keeps at an even speed during the test. *Vehicle A* first runs at an even speed; it then slows down after it detects *Vehicle B*. *Vehicle A* estimates the time that *Vehicle B* passes the conflicting point (say t_B) and determines the time t_A for *Vehicle A* to pass the conflicting point without collision. In this paper, we assume that $t_A = t_B + 1$.

After determining t_A, we control the jerk (the derivative of acceleration according to time) of *Vehicle A* in a piecewise-linear type as

$$J_A(t) = \begin{cases} -Kt, \; 0 \le t < \frac{t_A}{4} \\ Kt - \frac{Kt_A}{2}, \; \frac{t_A}{4} \le t < \frac{3t_A}{4} \\ -Kt + Kt_A, \; \frac{3t_A}{4} \le t < t_A \end{cases} \tag{1}$$

where K describes how much the jerk changes by time. Suppose the distance between *Vehicle A* and the conflicting point is L, when *Vehicle A* first detects *Vehicle B*. K is carefully chosen so that *Vehicle A* completes the slowing down process according to Eq. (1), right after it passes distance L.

As a result, the velocity of *Vehicle A* changes as following, after *Vehicle A* detects *Vehicle B*

$$V_A(t) = \begin{cases} V_{\mathrm{A}0} - \frac{Kt^3}{6}, \; 0 \le t < \frac{t_A}{4} \\ V_{\mathrm{A}0} + \frac{Kt^3}{6} - \frac{Kt_A t^2}{4} + \frac{Kt_A{}^2 t}{16} - \frac{Kt_A^3}{192} \; \frac{t_A}{4} \\ \le t < \frac{3t_A}{4} \\ V_{\mathrm{A}0} - \frac{Kt^3}{6} + \frac{Kt_A t^2}{2} - \frac{Kt_A{}^2 t}{2} + \frac{13Kt_A^3}{96} \; \frac{3t_A}{4} \\ \le t < t_A \end{cases} \tag{2}$$

To ensure the correctness of the simulation, we require the choice of K to guarantee that $V_A(t) \ge 0$.

When both vehicles pass the intersection, the test ends up and we score *Vehicle A* with the following formula

$$\text{score} = k_1 d_{\mathrm{AB}} + k_2 J_A \tag{3}$$

where d_{AB} denotes the minimum distance between *Vehicle A* and *Vehicle B* through the test and J_A denotes the maximum jerk (the derivative of acceleration according to time) of *Vehicle A* during the test. It must be pointed out that the linear weighting mixture given in Eq. (3) may not be the best trade-off formula here. We just choose it for an illustrative example.

These two components of score are chosen to measure the safety and comfortable indices of driving. k_1 is a positive scalar and k_2 is a negative scalar. So, the higher the score is, the better the performance is.

If *Vehicle A* do not have enough time to avoid collision with Vehicle B, we directly set the score equals to 0.

Fig. 7(b) gives the corresponding scores of *Vehicle A* obtained in the simulation test, with the change of detection range of *Vehicle A*. Here, the initial speeds of *Vehicle A* and *Vehicle B* are set as 12 m/s. The initial displacement d_A and d_B of *Vehicle A* and *Vehicle B* to the corresponding centerlines are set as 50 m. k_1 is set as 1.0 and k_2 is set as -0.02.

We can see that the score significantly drops when *Vehicle A*'s detection range is lower than 10 m. This indicates that the detection ability of an autonomous vehicle should at least be larger than 10 m in such conditions. Varying the speeds of *Vehicle A* and *Vehicle B*, we can detect the minimum required detection range of *Vehicle A* in different traffic scenarios.

This example illustrates how we evaluate the function units of an autonomous vehicle according to its performance in specially designed tests. Noticing the unavoidable environment noise disturbance, we can introduce some stochastic factors into the testing parameters and run such tests run for a number of times to determine the reliability of the testing vehicle. Varying the settings of the tests, we can cover the whole scenario space and quantitatively evaluate the intelligence of an autonomous vehicle in a synthetic manner.

V. Simulation Testing Using Parallel Transportation Systems

In Section III, we propose a framework to systematically test autonomous vehicles. In this section, we will discuss how to solve two problems in applying this framework for simulation testing.

Simulation based tests are widely used for vehicle function tests, since the complexity of testing scenarios can be easily controlled and a particular test can be accurately repeated for a number of times without spending too much financial/time costs [33]–[40].

However, existing simulated based vehicle testing has two problems:

First, the behaviors of all the agents in simulation systems are manually coded and rigid. For example, the vehicles only use a certain (usually simplified) car-following model to guide their movements [41]–[43]. This restriction often makes the system fail to reproduce some complex behaviors of traffic participators observed in real lives.

Second, the vision data are purely generated by 3D virtual reality engines. Although the capability of cutting-edge 3D VR engines is powerful and is being continuously improved, the generated view data sometimes do not appear enough real for some vision function tests.

To solve these two problems, we design a new simulation test platform that incorporates parallel traffic systems [44]–[46]; see Fig. 8(a). In brief, we setup a mirror of the real testing ground in the virtual spaces, which is called parallel traffic systems. This system will simulate the testing scenario and test how vehicles interact with the scenario.

Because of the flexibility provided by the parallel systems, we can record and learn some complex behaviors of real traffic participators and reproduce such behaviors in virtual reality for testing vehicles. As pointed out in [44]–[46], existing parallel

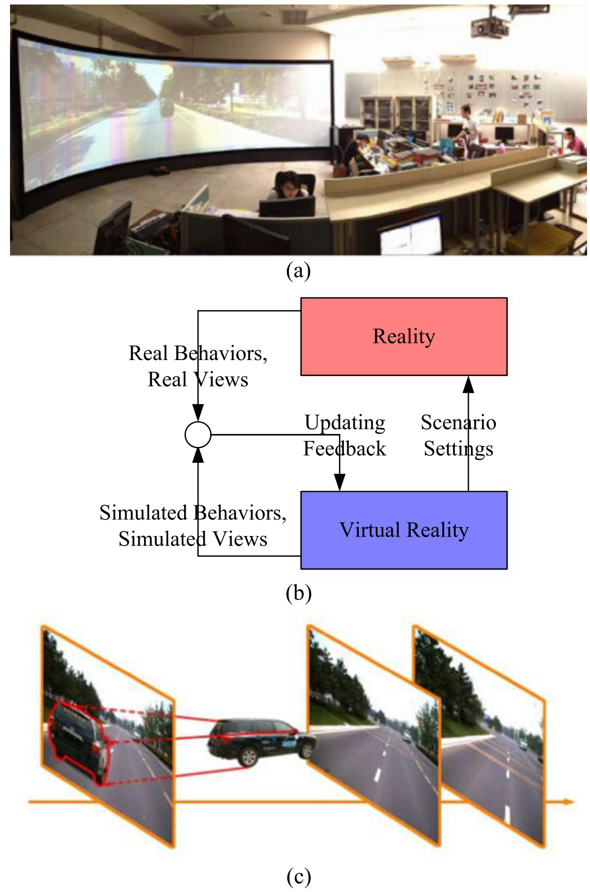


Fig. 8. (a) A simulation test platform designed by Xi'an Jiaotong University, Institute of Automation CAS, and Tsinghua University in China, which incorporates parallel traffic systems; (b) the cyclic updating mechanism of the co-evolution between real testing ground and parallel virtual testing ground; (c) an illustration on mapping the real view data to virtual reality space for flexible testing.

traffic systems had been proven to be able to compare the real behaviors observed in practice and the simulated behaviors generated in virtual space so as to continuously upgrade the parallel system; see Fig. 8(b).

Moreover, we can also map the view data collected by video cameras, radars and laser scanners into the virtual spaces, so that the testing vehicles can received vision data as real as possible for testing. This data-driven method can significantly boost the facticity of the tests [47].

VI. Conclusions

In this paper, we propose a new semantic diagram definition of driving intelligence which explains the relationship between testing scenarios, tasks, and vehicle functions. We show that the testing scenario/tasks generation process and the functions evaluation process can be viewed as two transverses (with opposite directions) in the proposed semantic diagram.

Based on this definition, we discuss how to synthetically test driving intelligence of an autonomous vehicle. The new task generation method enables us to design appropriate strategies to sample scenario instances, so as to keep a good balance between test cost and test case coverage. We can also design a series of tests to evaluate the intelligence of autonomous vehicles in an incremental style.

The new function evaluation method allows us to openly, quantitatively and fairly measure sensing/recognition, decision and action behaviors inside a vehicle. To verify the proposed method, a numerical example is provided to intuitively explain how the proposed quantitatively method measures the performance of vehicles and automatically detects which scenarios a vehicle may fail to pass.

In addition, we discuss how to design simulation tests for autonomous vehicles. Such simulation testing data can be easily converted to standard benchmarks for both beginners and veterans in autonomous vehicle filed. Those researchers who cannot afford a fully autonomous vehicle or those researchers who are only interested in developing a special function of autonomous vehicles can conveniently use such simulation testing benchmarks in their studies.

The new testing theory and methods are approved to be used to guide the coming "Intelligent Vehicle Future Challenge" that will be held in China in 2016 and 2017. We believe this set of new testing methods will help improve the developing of autonomous vehicles in the future.

References

[1] L. Li, J. Song, F.-Y. Wang, W. Niehsen, and N.-N. Zheng, "IVS 05: New developments and research trends for intelligent vehicles," *IEEE Intell. Syst.*, vol. 20, no. 4, pp. 10–14, Jul./Aug. 2005.

[2] L. Li and F.-Y. Wang, *Advanced Motion Control and Sensing for Intelligent Vehicles*. Berlin, Germany: Springer, 2007.

[3] A. Eskandarian, ed., *Handbook of Intelligent Vehicles*. Berlin, Germany: Springer, 2012.

[4] T. J. Gordon and M. Lidberg, "Automated driving and autonomous functions on road vehicles," *Veh. Syst. Dyn.*, vol. 53, no. 7, pp. 958–994, (2015).

[5] Google Self-Driving Car Project, [Online]. Available: https://www.google.com/selfdrivingcar/

[6] Tesla's Musk sees fully autonomous car ready in 5 years, (2014). [Online]. Available: https://my.teslamotors.com/fr_CH/forum/forums/tesla%E2%80%99s-musk-sees-fully-autonomous-car-ready-5-years

[7] A. Zubinskiy, The Quality of Embedded Software, or the Mess has Happened. (2013). [Online]. Available: http://www.viva64.com/en/a/0083/

[8] S. Thrun *et al.*, "Stanley, the robot that won the DARPA Grand Challenge," *J. Field Robot.*, vol. 23, no. 9, pp. 661–692, 2006.

[9] M. Buehler, K. Iagnemma, and S. Singh, Eds., *The DARPA Urban Challenge*. Berlin, Germany: Springer, 2009.

[10] M. Campbell, M. Egerstedt, J. P. How, and R. M. Murray, "Autonomous driving in urban environments: Approaches, lessons and challenges," *Philosoph. Trans. A*, vol. 368, no. 1928, pp. 4649–4672, 2010.

[11] A. Broggi, *et al.*, "TerraMax vision at the urban challenge 2007," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 1, pp. 194–205, Mar. 2010.

[12] K. Redmill, U. Ozguner, S. Biddlestone, A. Hsieh, and J. Martin, "Ohio State University Experiences at the DARPA Challenges," *SAE Int. J. Commer. Vehicles*, vol. 1, no. 1, pp. 527–533, 2009.

[13] National Science Foundation of China, Intelligent vehicle future challenges: 2009-2015, Nat. Sci. Found. China, Beijing, China, Tech. Rep., 2015.

[14] UK to Lead Development of Driverless Car Technology, (2015). [Online]. Available: https://www.gov.uk/government/news/uk-to-lead-development-of-driverless-car-technology

[15] Virginia Tech Transportation Institute and Partners Unveil Virginia Automated Corridors, (2015). [Online]. Available: http://www.vtti.vt.edu/featured/?p=260

[16] K. Connolly, Germany Creates Laws for Driverless Cars, (2015). [Online]. Available: http://www.theguardian.com/technology/2015/feb/01/germany-laws-driverless-cars-autobahns-google-industry

[17] D. J. Fagnant and Kara Kockelman, "Preparing a nation for autonomous vehicles: Opportunities, barriers and policy recommendations," *Transp. Res. Part A: Policy Pract.*, vol. 77, pp. 167–181, 2015.

[18] N. Nilsson, *The Quest for Artificial Intelligence: A History of Ideas and Achievements*. New York, NY, USA: Cambridge Univ. Press, 2009.

[19] A. Broggi *et al.*, "Extensive tests of autonomous driving technologies," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 3, pp. 1403–1415, Sep. 2013.

[20] Mcity Test Facility. (2015). [Online]. Available: http://www.mtc.umich.edu/test-facility

[21] W. Huang, D. Wen, J. Geng, and N.-N. Zheng, "Task-Specific performance evaluation of UGVs: Case studies at the IVFC," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 5, pp. 1969–1979, Oct. 2014.

[22] L. Li, D. Wen, N.-N. Zheng, and L.-C. Shen, "Cognitive cars: A new frontier for ADAS research," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 1, pp. 395–407, Mar. 2012.

[23] M. M. Trivedi, T. Gandhi, and J. McCall, "Looking-in and looking-out of a vehicle: Computer-vision-based enhanced vehicle safety," *IEEE Trans. Intell. Transp. Syst.*, vol. 8, no. 1, pp. 108–120, Mar. 2007.

[24] Z. Li, L. Li, and Y. Zhang, "IVS 09: Future research in vehicle vision systems," *IEEE Intell. Syst.*, vol. 24, no. 6, pp. 62–65, Nov. 2009.

[25] J. Gomez *et al.*, "Where are we after five editions? Robot vision challenge, a competition that evaluates solutions for the visual place classification problem," *IEEE Robot. Autom. Mag.*, vol. 22, no. 4, pp. 147–156, Dec. 2015.

[26] The German Traffic Sign Recognition Benchmark and the German Traffic Sign Detection Benchmark (GTSDB), (2014). [Online]. Available: http://benchmark.ini.rub.de/?section=home&subsection=news

[27] LISA Traffic Sign Dataset, (2016). [Online]. Available: http://cvrr.ucsd.edu/LISA/lisa-traffic-sign-dataset.html

[28] C. Little, The intelligent vehicle initiative: Advancing 'human-centered' smart vehicles, *Fed. Highw. Adm., Public Roads, U.S. Dep. Transp.*, [Online]. Available: http://www.fhwa.dot.gov/publications/research/.../pavements/.../publicroads/97septoct/p97sept18.cfm

[29] F.-Y. Wang, X. Wang, L. Li, and P. Mirchandani, "Creating a digital-vehicle proving ground," *IEEE Intell. Syst.*, vol. 18, no. 2, pp. 12–15, Mar./Apr. 2003.

[30] S. Glaser, B. Vanholme, S. Mammar, D. Gruyer, and L. Nouvelière, "Maneuver-based trajectory planning for highly autonomous vehicles on real road with traffic and driver interaction," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 3, pp. 589–606, Sep. 2010.

[31] Y. Kuwata *et al.*, "Real-time motion planning with applications to autonomous urban driving," *IEEE Trans. Control Syst. Technol.*, vol. 17, no. 5, pp. 1105–1118, Sep. 2009.

[32] L. Ma, J. Xue, K. Kawabata, J. Zhu, C. Ma, and N.-N. Zheng, "Efficient sampling based motion planning for on-road autonomous driving," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 4, pp. 1961–1976, Aug. 2015.

[33] C. Kasnakoglu and U. Ozguner, "Recent advances on the OSU virtual environment system test-bed and its applications," in *Proc. IEEE Intell. Veh. Symp.*, pp. 537–542, 2003.

[34] J. J. Olstam, J. Lundgren, M. Adlers, and P. Mastoms, "A framework for simulation of surrounding vehicles in driving simulators," *ACM Trans. Model. Comput. Simul.*, vol. 18, no. 3, 2008, Art. no. 9.

[35] V. Punzo and B. Ciuffo, "Integration of driving and traffic simulation: Issues and first solutions," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 2, pp. 354–363, Jun. 2011.

[36] J. Olstam, S. Espie, S. Maardh, J. Jansson, and J. Lundgren, "An algorithm for combining autonomous vehicles and controlled events in driving simulator experiments," *Transp. Res. Part C: Emerging Technol.*, vol. 19, no. 6, pp. 1185–1201, 2011.

[37] F. Han, W. Deng, S. Zhang, J. Wu, Y. Wang, B. Liu, B. Shang, and S. Qiu, "A real-time virtual simulation environment for Advanced Driver Assistance System development," SAE Tech. Paper 2014-01-0194, 2014.

[38] C. Zhang, Y. Liu, and D. Zhao, "RoadView: A traffic scene simulator for autonomous vehicle simulation testing," in *Proc. IEEE Int. Conf. Intell. Transp. Syst.*, Qingdao, China, 2014, pp. 1160–1165.

[39] Y. Li, Y. Liu, C. Zhang, and D. Zhao, "The 'floor-wall' traffic scenes construction for unmanned vehicle simulation," in *Proc. IEEE Int. Conf. Intell. Transp. Syst.*, Qingdao, China, 2014, pp. 1726–1731.

[40] D. Zhao, Y. Liu, C. Zhang, and Y. Li, "Autonomous driving simulation for unmanned vehicles," in *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, 2015, pp. 185–190.

[41] S. C. Yang, M. Li, T. Q. Tang, and Y. Lin, "An electric vehicle's battery life model under car-following model," *Measurement*, vol. 46, no. 10, pp. 4226–4231, 2013.

[42] B. Ciuffo and V. Punzo, 'No Free Lunch' theorems applied to the calibration of traffic simulation models," *IEEE Trans. Intell. Transp. Syst.*. vol. 15, no. 2, pp. 553–562, Apr. 2014.

[43] L. Pariota, G. N. Bifulco, F. Galante, A. Montella, and M. Brackstone, "Longitudinal control behaviour: Analysis and modelling based on experimental surveys in Italy and the UK," *Accident Anal. Prevention*, vol. 89, pp. 74–87, 2016.

[44] Zhang, F.-Y. Wang, F. Zhu, D. Zhao, and S. Tang, "DynaCAS: Computational experiments and decision support for ITS," *IEEE Intell. Syst.*, vol. 23, no. 6, pp. 19–23, Nov./Dec. 2008.

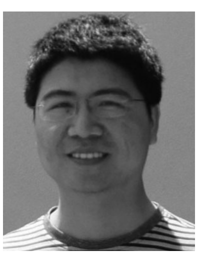
[45] F.-Y. Wang, "Parallel control and management for intelligent transportation systems: Concepts, architectures, and applications," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 3, pp. 630–638, Sep. 2010.

[46] L. Li and D. Wen, "Parallel systems for traffic control: A rethinking," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no 4, pp. 1179–1182, Apr. 2016.

[47] J. Zhang, F.-Y. Wang, K. Wang, W.-H. Lin, X. Xu, and C. Chen, "Data-driven intelligent transportation systems: A survey," *IEEE Trans, Intell. Transp. Syst.*, vol. 12, no. 4, pp. 1624–1639, Dec. 2011.

Li Li (S'05–M'06–SM'10) is currently an Associate Professor with the Department of Automation, Tsinghua University, Beijing, China, working in the fields of complex and networked systems, intelligent control and sensing, intelligent transportation systems, and intelligent vehicles.

Dr. Li had published more than 50 SCI indexed international journal papers and more than 50 international conference papers as a first/corresponding author. He serves as an Associate Editor of the IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS.



Wu-Ling Huang received the Ph.D. degree in control theory and control engineering from the Institute of Automation, Chinese Academy of Sciences, Beijing, China. He is currently an Associate Professor with State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, working in the fields of intelligent transportation systems and intelligent vehicles.

Yuehu Liu received the B.E. and M.E. degrees from the Xi'an Jiaotong University, Xi'an, China, in 1984 and 1989, respectively, and the Ph.D. degree in electrical engineering from Keio University, Tokyo, Japan, in 2000. He is currently a professor with Xi'an Jiaotong University. His research interests are focused on computer vision, computer graphics, and simulation testing for autonomous vehicle. He is a member of the IEICE.

Nan-Ning Zheng (SM'93–F'06) graduated from the Department of Electrical Engineering, Xi'an Jiaotong University, Xi'an, China, in 1975, and received the M.S. degree in information and control engineering from Xi'an Jiaotong University in 1981, and the Ph.D. degree in electrical engineering from Keio University, Yokohama, Japan, in 1985.

He joined Xi'an Jiaotong University in 1975, and he is currently a Professor and the Director of the Institute of Artificial Intelligence and Robotics. His research interests include computer vision, pattern recognition and image processing, and hardware implementation of intelligent systems.

Prof. Zheng became a member of the Chinese Academy of Engineering in 1999, and he is the Chinese Representative on the Governing Board of the International Association for Pattern Recognition. He also serves as the President of Chinese Association of Automation.

Fei-Yue Wang (S'87–M'89–SM'94–F'03) received the Ph.D. in computer and systems engineering from Rensselaer Polytechnic Institute, Troy, New York, NY, USA, in 1990. He joined the University of Arizona in 1990 and became a Professor and Director of the Robotics and Automation Lab and Program in Advanced Research for Complex Systems. His current research focuses on methods and applications for parallel systems, social computing, and knowledge automation. In 1999, he founded the Intelligent Control and Systems Engineering Center at the Institute of Automation, Chinese Academy of Sciences (CAS), Beijing, China, under the support of the Outstanding Oversea Chinese Talents Program from the State Planning Council and "100 Talent Program" from CAS, and in 2002, was appointed as the Director of the Key Lab of Complex Systems and Intelligence Science, CAS. In 2011, he became the State Specially Appointed Expert and the Director of the State Key Laboratory of Management and Control for Complex Systems.

Prof. Wang was the Founding Editor-in-Chief (EiC) of the *International Journal of Intelligent Control and Systems* from 1995 to 2000, the Founding EiC of IEEE INTELLIGENT TRANSPORTATION SYSTEMS MAGAZINE from 2006 to 2007, the EiC of IEEE INTELLIGENT SYSTEMS from 2009 to 2012, and the EiC of the IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS from 2009 to 2016. He is currently the EiC of the IEEE/CAA JOURNAL OF AUTOMATICA SINICA, and China's *Journal of Command and Control*. Since 1997, he has served as the General or Program Chair of more than 20 IEEE, INFORMS, ACM, and ASME conferences. He was the President of IEEE ITS Society from 2005 to 2007, Chinese Association for Science and Technology in 2005, the American Zhu Kezhen Education Foundation from 2007 to 2008, and the Vice President of the ACM China Council from 2010 to 2011. Since 2008, he has been the Vice President and Secretary General of Chinese Association of Automation. He is an elected Fellow of INCOSE, IFAC, ASME, and AAAS. In 2007, he received the 2nd Class National Prize in Natural Sciences of China and the Outstanding Scientist by ACM for his work in intelligent control and social computing. He received the IEEE ITS Outstanding Application and Research Awards in 2009 and 2011, respectively, and the IEEE SMC Society Norbert Wiener Award in 2014.