Autonomy Evaluation Methods for the Autonomous Ground Vehicles  
JingLI$^{1, a}$, Zhenmin TANG$^{2, b}$, Yan FANG$^{3, c}$, YefaTAN$^{1, d}$

$^1$College of Field Engineering, PLA University of Science and Technology, Nanjing 210007, China  
$^2$Department of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China  
$^3$Training Division, Military Transportation University, Tianjin 300161, China  
$^a$email: lijingwry@126.com, $^b$email: Tzm.cs@njust.edu.cn, $^c$email:aimee8509@gmail.com,  
$^d$email:tanyefa7651@163.com

Key words: Unmanned system; Autonomous ground vehicle; Autonomy; Evaluation method

Abstract. With the wide application of the autonomous ground vehicles (AGVs) in the military and civilian areas, the autonomy requirements such as autonomous self-awareness, self-planning and self-control are increasingly considered. The autonomy evaluation has been one of the key issues in the research of the AGVs. Firstly, the autonomous definition and significance of the AGVs were discussed. The milestone events in the development of the autonomy evaluation methods were reviewed. Then the autonomy evaluation methods for the AGVs were summarized from both the early methods based on ontology & phenomenology and the modern ones based on cognitive & computational methods, respectively. Finally, an outlook of the autonomy evaluation development for the AGVs was analyzed.

Introduction

With the rapid development of computer science, sensor technology and artificial intelligence theory, the unmanned system has been widely used in space, battlefield, the deepsea, and many other engineering areas. In comparison with the sky and marine environment, the ground is usually considered to be the most intensive place for human activity with complex and changeable conditions. Researchers have focused on the autonomous ground vehicles (AGVs) for a long time, and thus developed their applications in military and civilian areas [1]. For example, in the battlefield environment, the AGVs can complete the tasks of investigation, transportation, and disposal of dangerous goods instead of the soldiers in dangerous situations; in the aerospace area, the AGVs can complete the work such as cruising on mars, mapping or completing scientific experiments instead of human beings; and in the civilian area, the AGVs can complete several tasks such as disaster rescue or mining operations. Moreover, the initiative security system, which is one of the core technologies of the AGVs, has been widely installed in the passenger-vehicles and played an important role in the auxiliary driving.

The working conditions of the AGVs have gradually developed from the simple indoor and structured environment to a variety of dangerous and remote natural environment. Therefore, it shows great importance to enhance the autonomy of AGVs such as self-awareness, self-planning and self-control. The research on autonomy evaluation has become one of the key issues for the AGVs and even the whole unmanned system. Numerous reports on the autonomy evaluation of the unmanned system have been published [2][3].

In this study, the definition and significance of the autonomy of the AGVs were discussed. The milestone events in the development of the autonomy evaluation methods were reviewed. The autonomy evaluation methods for the AGVs were summarized from both the early methods based on ontology & phenomenology and the modern ones based on cognition & calculation, respectively. Additionally, an outlook of the development direction was given as well.
Proposing of the Autonomy Concept

The progress in computer science and communication technology promotes the development of the intelligence and autonomy of the unmanned system. The unmanned system represented by unmanned aerial vehicles (UAVs) began to attract attention in the late 1970s. The early UAVs were manually controlled by skilled operators on the ground or in the manned aerial vehicles. In the 1980s, the manual inputting of signals was replaced by the mode of pre-programmed controlling. In 2000, The U.S. Navy Research Office and the Air Force Research Laboratory (AFRL) proposed the concept of autonomous operation for the first time, and they stated that the key to realize the autonomous operation is to enhance the information processing capability [4][5]. Meanwhile, Pachter et al. [6] claimed that the autonomous control was high level of automatic control in unstructured environment. However, these statements did not consider the environment perception and other important issues. On the basis of these above studies, Boskovic et al. [7] proposed that the autonomy was an optimal way to execute the control strategy when the system was in the uncertain environment without human intervention. The autonomy ensured the unmanned system to accommodate the environment independently, quickly, and efficiently. Therefore, the unmanned system could realize the functions such as scene description, fault detection, strategic planning or obstacle avoidance.

In order to study the autonomous operation more deeply, the AFRL defined the concept of Autonomous Control Level (ACL), and made it a standard to measure the degree of autonomy level for the UAVs. This standard was adopted by the U.S. Defense Ministry to release three versions of the UAV roadmaps (Unmanned Aerial Vehicles Roadmap 2000-2025[8], Unmanned Aerial Vehicles Roadmap 2002-2027[9], and Unmanned Aerial Systems Roadmap 2005-2030). The autonomy level of the representative UAV systems was clearly defined. From then on, the ACL criterion was widely recognized.

The U.S. Defense Department unified the unmanned aerial system, unmanned ground vehicle and unmanned maritime system into unmanned systems in 2007, and released two versions of unmanned system roadmap officially (Unmanned Systems Roadmap 2007-2032, Unmanned Systems Integrated Roadmap 2009-2034 [10]). The National Institute of Standards and Technology (NIST) set the autonomy as the performance level of the unmanned systems together with the Department of Commerce, Department of defense, Department of energy, Department of transportation, etc. And they established a working group of Autonomy Levels For Unmanned Systems (ALFUS). Therefore, the autonomy turned into the performance level of the unmanned systems completely.

As the study of autonomy has attracted increasing attention in abroad, researchers in China have also published literature about the autonomous control level [11], autonomous control [12], autonomy [13], and autonomy evaluation [14].

Definition of Autonomy

It can be found from the proposing process of the autonomy concept that it not only emphasizes the high automation degree of the unmanned system in order to reduce human interventions, but also pays attention to the comprehensive ability of perception, understanding, and information processing in the uncertain environment. Numerous studies have focused on the definition of autonomy. Terry et al [15] considered that the autonomous system has the ability to observe, analyze, communicate, plan and act, and it can accomplish the task that human beings have assigned to it. Jefferry et al. [16] decomposed the autonomy into two dimensions, i.e. self-management and self-guidance. And they proposed an adjustable model for the autonomy level. Reed et al. [17] defined the autonomy to be a quantitative value, which was used to distinguish different levels of autonomy. Sean et al. [18] created adjustable autonomous modules in the framework of intelligent control on the basis of the Joint Intention Theory. Considering the factors of standardization and completeness, the definition proposed by the ALFUS is generally adopted:
autonomy of the unmanned systems is the performance level they achieved when they execute the task based on the hardware and software platforms in a dynamic and unstructured environment.

Autonomy is defined based on the complexity of the task and environment, as well as the degree of the human independence. In order to express the definition of autonomy more clearly, the concept of automation and intelligence are listed below.

(1) **Automation**: Automation means that the system can work automatically according to the preset parameters without intervention of the operators. Based on data driven, automatic control shows low tolerance for failure and error [19].

(2) **Intelligence**: Intelligence is smart characteristics of some humanoid or personification. Some static processing ability of the system has been established in the stage of design. Compared with automatic control, the drive modes of intelligent control are mainly based on information such as image, speech or sensor signal, thus it can tolerate a certain degree of error and failure [20].

(3) **Autonomy**: On the basis of the fundamental intelligent functions, autonomy reflects the self-management ability of the system. An autonomous system has the ability to deal with unexpected events that are not pre-defined by the human beings, and it generally presents dynamic capabilities of self-processing and task execution. In comparison with automation and intelligence, autonomy shows knowledge-driven mode according to the results of path planning, task decomposition or task distribution. An autonomous system has the self-repairing ability for failure and error.

**Significance of Autonomy Evaluation for AGVs**

As a typical complex unmanned system, the AGVs not only represent the level of science and technology of a country, but also have great practical value. Autonomy evaluation on environment perception, decision-making and reliability contributes to the design and performance forecasting of a new AGV system. The autonomy evaluation of the AGVs is required by both the scientific research and the technological strategy.

First, autonomy evaluation is an effective method for project-demonstration for the AGVs. Through the evaluation process, the weaknesses of the AGV systems can be correctly analyzed, and thus providing a guarantee for the development direction of the AGVs.

Second, autonomy evaluation provides technical support for the life management of the AGVs. By checking the performance index of the AGV systems and comparing with the theoretical design index, the scheme of the project can be optimized efficiently.

Third, autonomy evaluation for the AGVs can provide reference for the comprehensive performance evaluation of other complex intelligence systems. The AGV systems show several characteristics such as strong real-time, tight coupling, dynamic, complex and intelligent. The autonomy evaluation not only covers the functional properties of the physical body and sensor components, but also highlights the ability of decision-making of the AGV systems. Through the concept of autonomy, the overall performance of other complex intelligent systems can be described directly. Therefore, the autonomy evaluation of the AGVs can provide useful reference for the performance evaluation of other unmanned systems such as unmanned aerial system and unmanned maritime system.

**Early Autonomy Evaluation Methods for AGVs Based on Ontology & Phenomenology**

In the early development stages of the AGVs, there is no clear description for the autonomy concept. At that time, the judging criterion of self-awareness, self-planning and self-control for the AGVs were mainly based on the comparative analysis of the running data obtained in the control system. So they belonged to evaluation methods based on ontology and phenomenology. From the several famous robot research projects below, these evaluation methods were analyzed and summarized as follows.

(1) **Evaluation in strategic computing program**

The American Defense Advanced Research Projects Agency (DARPA) started the “Strategic
Computing Program” in 1980s, and thus attracting many universities and companies to participate init [21]. The sample car of Martin Marietta Company named “ALVin” realized the functions of path tracking and automatic obstacle avoidance, which was driven on the unstructured road for 4.2 kilometers at the average speed of 10 kilometers per hour. Carnegie Mellon University successfully developed the “Navlab” autonomous platform with the functions of cross-country investigation, highway automatic running, cross-country obstacle avoidance, and driver assistance in the crowded urban environment. Among the platform series, the NavLab5 completed a 2840-miles travel across the North American continent in 1995, in which the autonomous driving completed 2797 miles [22]. The artificial control behavior in the test only included the brake and the throttle. Moreover, the artificial interference rate is less than 2%. In 1998, Navlab11 also completed the travel across the American continent from Pittsburgh to California. The testing and evaluation methods in this period are mainly based on the system debug, and the testing of physical performance and sensor functions were the key factors of the autonomous platform. According to the quantitative degree of artificial intervention, the autonomy level could be obtained intuitively.

(2) Evaluation in perception for off-road robotics program

The Perception for Off-Road Robotics (PerceptOR) program was hosted by American DARPA, which mainly focused on the sensor system testing and navigation ability evaluation of the AGVs. The former platforms were generally based on different platforms and mainly used for testing and showing, but the PerceptOR program adopted the unified Honda Foreman Rubicon all-terrain autonomous platform. All research institutions were requested to develop their system on this unified platform. The project sponsor would give objective testing results and autonomy evaluation to all the systems at last.

In the PerceptOR program, the AGVs were tested by the way of manual driving, remote control and autonomous driving in the forest, grass, desert and mountainous regions. Through the comparison and analysis of the running data from the vehicle control system, the overall evaluation of the system performance was accomplished. Meanwhile, the sensor system was improved according to the testing results, including the increase of the resolution of the laser radar imaging instrument and the stereo FLIR camera, thus the evaluation information feedback could be realized to the system performance.

(3) National qualification event (NEQ) tests of the DARPA challenge

The unmanned highly-mobile platforms are the base of the AGVs. According to this characteristic, DARPA organized two cross-country ground vehicle challenges (Grand Challenge) in the field environment and one urban vehicle challenge (DARPA urban challenge, DUC).

The Grand Challenge organized by DARPA learned from the evaluation methods of the PerceptOR program. During the first two Grand Challenges in 2004 and 2005, the evaluation target was to investigate the autonomous navigation, path planning and safe driving ability of the vehicles in the off-road environment. The National Qualification Event (NQE) tests during the preliminaries mainly focused on the basic abilities of the vehicles such as obstacle avoidance, intersection of two cars, travelling on heaving roadways and the roads of ups and downs, or running through the mountains and sharp corners. The final stage was arranged with a remote and large span of the autonomous navigation task. In order to realize the test and evaluation target of the vehicles, they were tested and evaluated by a following vehicle in the moving process.

In 2007, the competition was mainly for the urban traffic environment, and thus the testing of urban traffic capacity was introduced. All the AGVs were requested to travel on a 96-kilometers long road named “model city blocks (no traffic violations and accidents)”, where they must avoid the obstacles, obey the traffic signals, incorporate into the general traffic jam, and avoid other vehicles autonomously. The testing apparatus and the simulation of traffic flow were positioned in the final field, and the evaluation method based on the penalty system was established [23]. The evaluation results not only depended on the completion time of the task, but also the safety of the vehicles and the efficiency of the path selection [24].
Autonomy Evaluation Methods for AGVs Based on Cognition & Calculation

It can be found from the research program and development process of the AGVs that the autonomy evaluation always develops with the improvement of the prototype system. With the rapid development of the intelligence technology and robot cognition theory, a series of evaluation models and methods based on cognition & calculation were proposed.

(1) Heuristic calculation method

In the AGV systems, the interaction between human and the AGVs reflects their autonomy level to some degree. Sheridan et al. [25] proposed an evaluation model for the human interaction level, and decomposed the task execution into four steps, i.e., information acquisition, information analysis, decision choice and act realization. The evaluation criteria included human performance, system reliability and costs of the action sequences. Through this heuristic calculation model, the problem caused by calculation of human interaction level can be solved.

Based on the four steps (Observe, Orient, Decide, Act, OODA) during the decision making process of the human beings, Ryan et al. [26] proposed an evaluation method using task evaluation and re-planning capability of the system according to the modular function.

Meanwhile, the network communication capability was also considered as a suitable measurement for autonomy and thus attracted some attention [27]. Dorais et al. [28] used the time delay of the network communication to evaluate the network status, and adjusted the resource configuration of each intelligent body in the multi-agent system dynamically. Zhou et al. [29] studied the relationship between the autonomous ability and the quality of the network communication. They found that the autonomy of the system decreased as the time delay of the network communication decreased.

Environmental perception is the basis of autonomous planning and control for the AGVs. In order to study the navigation and environment perception ability of the AGVs in unstructured natural terrain condition, Raia et al. [30] divided the terrain into several types such as the highly reliable area, the low reliable area, the high reliable barrier and the low reliable barrier, etc. They completed the terrain classification by the training classifier, and adopted the accuracy of classification results as the evaluation basis of the visual perception ability. Michael et al. [31] built the terrain model library and evaluated the learning ability of AGVs by the method of artificial marking and checking the terrain identification error after multi-pass the same terrain. Shirkhodaie et al. [32] proposed the Terrain Traversability Assessment Measure (TTAM). By counting the terrain deviation between the testing and actual values, the road detection capability was obtained, and the trafficability of the road tested was evaluated.

In addition, researchers proposed several other evaluation methods based on costs analysis, cost and benefit analysis [33], trajectory fragment analysis [34], and intelligent behavior analysis [35], respectively. These different methods above have their specific application scope and emphasis, which contributed to the development of autonomy evaluation for the unmanned systems. But it’s a pity that they didn’t form the systematic evaluation methods.

(2) Axis analysis method

The axis analysis method focuses on each individual aspect of the evaluated system. The evaluation results are determined from the weighted level of the overall aspects of the system. This method covers the multiple factors that affect the autonomy level, thus it plays an important role in autonomy evaluation.

One of the most representative axis analysis methods is the ALFUS framework proposed and established by the NIST in 2003, which contributed to testing and evaluating all kinds of unmanned systems including the AGVs comprehensively [36]. It stressed that the autonomy level of the unmanned system was the unity of three factors (Mission Complexity, Environment Complexity, and Human Interference) [37]. Each of the three factors represented an axis in the three-dimensional coordinate system, so that the autonomy level could be shown obviously. However, the ALFUS framework also showed several limitations. For example, the three factors above were not always completely independent from each other in actual applications, so it required complex
evaluation model to unify them into the uniform standards.

Another representative axis analysis method is the cobweb evaluation model [38]. This model has an original point with several radiating axes, and each axis indicates an aspect of the autonomy evaluation. Each aspect has nine levels, and these levels represent the degree of technology maturity. For each AGV system, there’s a corresponding mapping to a level of each evaluation aspect, and the corresponding level for each axis constitutes the latitude of the cobweb. By this means, the autonomy level of the AGVs can be evaluated.

(3) Formulation method

The formulation method is usually used to calculate the autonomy level of the unmanned system with a multivariable function directly. The multivariable function usually has two categories, i.e., multiplicative type and weighted type. It is high-demanding in the function design and variables choices. The formulation method is generally based on plenty of tests or data, and only for a specific system. In this paper, the formulation of autonomy, reliability, and efficiency of the system are shown as follows:

(i) Calculation of autonomy

In order to evaluate the collision-avoidance performance in the multiple unmanned system, Curtin et al [39] proposed a formulation for calculating the autonomy in Eq. (1):

\[ A = C_n \left( \frac{S_C}{E} \right)^i \left( \frac{S_T}{T} \right)^j \]

Where \( A \) is autonomy of the unmanned system, \( C_n \) represents complexity index of the task, \( S_C \) and \( S_T \) represents the controlled variable and controlled time, respectively, \( E \) represents the gross information content, \( T \) is the total task time, and \( i \) and \( j \) are the preset parameters.

(ii) Calculation of reliability

When the AGVs are influenced by the outer environments and inner conditions, something will happen such as breakdown, error, or invalidation [40]. Therefore, it shows significant importance to evaluate the performance of the AGVs based on reliability index. In general, the reliability of the system mainly includes the reliability degree, “mean time to failure” (MTTF) and the failure rate [41]. The reliability degree can be expressed as \( R(t) \) in Eq. (2):

\[ R(t) = 1 - F(t) \]

Where, MTTF represents the mean working time of the AGVs before failure occurs, thus it is also called “mean working time”. The relationship between the MTTF and the reliability can be expressed in Eq. (3):

\[ MTTF = \int_0^{\infty} R(t)dt \]

The ratio of total repair time and total number of failures is called “Mean Time to Repair” (MTTR), which represents the mean time of a maintenance activity. The mean time between the two adjacent failures is called “mean time between failures” (MTBF), which consists of MTTF and MTTR shown in Eq. (4):

\[ MTBF = MTTF + MTTR \]

Based on the above equations, the reliability degree of AGVs can also be expressed as Eq. (5):

\[ R(t) = \frac{MTTF}{MTBF} \times 100\% \]

The reliability index method focuses on the weak link of the system. By calculating the reliability degree of the weak parts or subsystems in the complex system, the overall failure condition of the system can be obtained.

(iii) Calculation of effectiveness

The effectiveness analysis is mainly for the products or equipment with specific performance, and generally used to evaluate the task completion ability of the system with specialized, mathematical or linguistic description [42]. Different calculation methods will be obtained from different effectiveness models. For example, in the AN model, as shown in Eq. (6), the effectiveness of the system shows positive relationships with the performance, effectiveness ratio, and utilization factor of the system.
\[ P_{SE} = P \cdot A \cdot V \] (6)

Where \( P_{SE} \) is the comprehensive effectiveness of the system, \( P \) represents the performance of the system, \( A \) is the effectiveness ratio, and \( V \) is the utilization factor [43].

Comparatively, the system effectiveness of the ARING model, as shown in Eq. (7), focused on the index of preparation conditions, task reliability and design appropriateness.

\[ P_{SE} = P_{OR} \cdot P_{MR} \cdot P_{OA} \] (7)

Where, \( P_{OR} \) represents the preparation conditions, \( P_{MR} \) is the task reliability, and \( P_{OA} \) represents the design appropriateness [44].

Both the AN model and ARING model reflect the reliability degree of the unmanned system, and can be obtained by numerical method. However, it is still difficult to obtain accurate index-values of the performance, such as the effectiveness ratio and design appropriateness of the system. Compared with the AN model and ARING model, the WSEIAC model for evaluating the performance of the unmanned system is more clear and accurate, and thus becomes a generally accepted model for evaluating the effectiveness of the system. The WSEIAC model can be expressed as Eq. (8):

\[ E = A \cdot D \cdot C \] (8)

Where, \( A \) is the availability vector, which represents the measure of the system status at the beginning stage of the task execution; \( D \) is the reliability matrix, which indicates the measure of the system status during the task execution process; and \( C \) is the ability matrix, representing the measure of the ability that the system owns to complete the specified task.

Since the WSEIAC model has three measuring index such as \( A \), \( D \) and \( C \), so it is also called the ADC effectiveness analysis model. It can be considered that the system effectiveness in the ADC model shows function relationship with the system status at the beginning stage of the task execution, the measure of the system status during the task execution process, and the measure of the ability that the system owns to complete the specified task. In practical applications, the availability and reliability of the system are generally simplified to the solving of MTTF and MTTR, and the setting and predicting of the ability matrix are closely related to the task.

**Conclusion**

In the past 30 years, the AGVs have attracted increasing attention and thus showed great potential in military and civilian areas. It shows great importance to enhance the autonomy of AGVs such as self-awareness, self-planning and self-control. In this study, on the one hand, the definition of autonomy and the significance of autonomy evaluation for AGVs were discussed, and the milestone events in the development of the autonomy evaluation methods were reviewed. On the other hand, the autonomy evaluation methods for the AGVs were summarized from both the early methods based on ontology & phenomenology and the modern ones based on cognitive & computational methods, respectively.

Development of the autonomy evaluation methods has become one of the key issues for the AGVs and even for the whole unmanned system. However, it still has a long way to go. For prospect, several aspects can be drawn for the future development of autonomy evaluation methods as follows:

1. Due to the technological gaps, cultural differences, technology confidentiality or other reasons, the proposing of effective universal evaluation methods is an important challenge. Therefore, further research is still required on universal evaluation methods.
2. Further work is required to focus on the quantitative index system and the corresponding evaluation methods.
3. Research on the cognition and understanding of the environment complexity should be further carried out, so as to form a novel classification standard of environment complexity for different intelligent agents.
References


[19] Correll N. Parameter estimation and optimal control of swarm-robotic systems: A case study in


