The Role of Situation Awareness in Assuring Safety of Autonomous Vehicles

Andrzej Wardziński

PROKOM Software SA
Podolska 21, 81-321 Gdynia, Poland
wardzinskia@prokom.pl

Abstract. Assuring safety of autonomous vehicles operating in an open environment requires reliable situation awareness, action planning and prediction of actions of other vehicles and objects. Factors that also have to be considered are certainty and completeness of available information and trust in information sources and other entities. The paper discusses the problem of autonomous vehicle safety assurance and proposes dynamic situation assessment to cope with the problem of environment dynamics and incomplete and uncertain situation knowledge. The approach is presented for a simple example of a simulated autonomous vehicle. The situation awareness model and autonomous vehicle control system architecture is presented. The problems of justifying system safety are discussed.

1 Introduction

In 1981 the first man was killed by a robot in a manufacturing plant in Japan [1]. The maintenance worker entered the robot's operating zone to fix its component. Instead of opening the robot's safety gate – which was supposed to cut off its power – he jumped over the barrier fence and accidentally switched the robot back on. The robot sensor was activated then the robot decided that he was an industrial component and crushed him. Other similar accidents have been reported [2]. The main cause of this kind of accidents was that a robot was unaware of a human being present in the area it operated. The fatal accidents as described above were caused not by components failures but weak ability to percept and assess the situation.

The objective of the paper is to investigate the problem of safety assurance for autonomous systems where external events and interaction with the environment and other systems have essential influence on safety. The problem and assumptions are introduced in section 2. Section 3 discusses the concept of situation risk assessment, trust in other agents and the problem of uncertain and incomplete knowledge. An example of an autonomous vehicle is introduced in Section 4. Then a situation awareness model is proposed in Section 5. The way how the situation awareness model is applied for the autonomous vehicle control system is presented in section 6. Hazard analysis of presented architecture is discussed in section 7. A vehicle cooperation process that strengthens safety is proposed in section 8. Experiences from autonomous vehicle simulation experiments are discussed in section 9.

J. Górski (Ed.): SAFECOMP 2006, LNCS 4166, pp. 205 – 218, 2006.

2 The Problem of Autonomous Vehicles Safety

Autonomy relates to an individual or collective ability to make decisions and act without outside control or intervention. Autonomy is quite a new concept for safety-critical systems.

Autonomous vehicle (AV) is a vehicle able to perform action planning and control without human interaction in order to accomplish its long-term mission goals. Autonomous vehicles operate in an open (i.e. non-controlled) environment.

Open environment is defined as an environment in which agents operate and can have different, not consistent missions and strategies. Some regulations can be defined for the environment and agents should follow them however it cannot be guaranteed that every agent would always act in accordance with the regulations. In an open environment an agent cannot assume that all other agents will cooperate and preserve safety. As an *agent* we understand a vehicle or an object.

An *object* can be able to communicate with other agents but not able to move. Examples of objects are traffic lights (which communicate to vehicles if they should stop or drive on) or a command centre (which communicates missions to vehicles).

The essential feature of an autonomous system is its ability to plan actions and achieve some long-term mission goals. Usually AV objective is to:

- accomplish its mission (a long-term goal),
- comply with the regulations if such rules are defined,
- preserve safety (avoid hazardous events).

An example of a hazardous event is a collision – when a vehicle collides with another vehicle or an object. Intuitionally one can say that a collision will happen when two vehicles have crossing courses. There may be many different causes for this event. It can be a sensor or actuator failure, wrong route planning, unexpected events in the environment or manoeuvres of other vehicles.

Accident models are useful for analysis of accidents causes and possible countermeasures. The sequential accident model is the simplest one. More complex models are also available however the sequential model is a good starting point sufficient for the purpose of the presented analysis. The model starts with the safe state followed by a sequence of states (see Fig. 1).

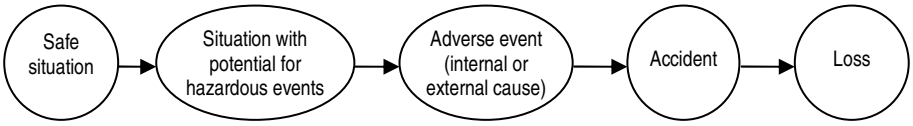

Fig. 1. Sequential accident scenario model

The presented sequential model emphasises that not only adverse events (like components failures) can cause hazards. There are many other factors that have to occur for a hazard to happen. This can be demonstrated on an example of a scenario in which a vehicle starts to overtake another vehicle while being overtaken at the

same moment. This hazardous situation is caused by the vehicles interaction and not by a mechanical component failure. Table 1 shows the increasing risk level for the consecutive accident scenario situations.

Table 1. Examples of accidents scenarios

Model state	Industrial robot	Autonomous vehicle	Risk level
Safe situation	Normal robot operation	No other vehicle in close distance	no risk
Situation with potential for adverse events	Preceding robot is off (no part can be passed on to work on)	Another vehicle in front (see Fig. 2.b)	low risk
	A human in operating area	Decision to overtake the vehicle in front	
Adverse event	The robot switched on	Unexpected manoeuvre of the vehicle in front	high risk
	The human activates a robot sensor		
	Robot arm starts the operation	Braking and course change fails	
Accident	Robot arm hits the human	Vehicles collide	unsafe state
Loss	Fatal injury	Vehicles damaged	

A risk level can be attributed to each situation and can be used to denote how far it is from the unsafe state. The longer is the sequence the better are chances that hazard can be avoided. To assure safety the system should continuously assess the situation and act accordingly when some risk factors are detected.

Similar sequential accident model is used by the European Automobile Manufacturers Association (ACEA) [3] for road traffic accidents. The ACEA model distinguishes "danger" state when an accident can be avoided and "crash unavoidable" state. This depicts the problem that to avoid a hazard the countermeasures should be taken before the risk level is too high.

3 Situation Awareness, Risk and Trust

Situation awareness (SAW) is, generally speaking, the knowledge of what is going around. Situation awareness is an area of research in domains of philosophy, logic, psychology, artificial intelligence, computer science (human-computer interface) and robotics. The research goal in psychology is to examine how a human maintains situation awareness, while in robotics the aim is to create machine SAW. Assessment if a situation is safe or dangerous is one of the situation awareness functions.

The general concept of situation awareness is well known however there is no one agreed definition. Endsley introduces three levels of human situation awareness [4]:

1. perception: basic perception of cues based on direct observation.
2. reasoning: the ability to comprehend or to integrate multiple pieces of information and determine the relevance to the goals the human wants to achieve.
3. predicting: ability to forecast future situation events and dynamics based on the perception and comprehension of the present situation.

Situation awareness is necessary for humans to assess the risk, plan actions and avoid dangerous situations. Reason in [5] describes mechanisms how humans perceive situations and the associated risk, how make errors (mistakes and lapses), and how react in case of unexpected events.

Another area of research on situation awareness is focused on remote controlled robots [6]. Its objective is to provide humans (who are regarded as the only entities with the ability of situation awareness) with the right information to ensure situation awareness and safe robot operation.

Situation risk assessment is one of the key human abilities to preserve safety. When someone drives a car at the speed of 50 km/h it is quite normal to pass a person on the pavement. That person is only two meters away from the passing car and no physical barrier exists between them. The driver trusts that the person will not enter the road in front of the car and therefore assesses the risk as low. On the other hand the person on the pavement trusts that the car will not leave the road. But we do not always trust everybody. For example when we see children we usually slow down – the situation is assessed as more risky.

Trust in an entity is related to certainty that the entity will behave in a predicted way and follow the regulations or some rules. Any entity acting in unpredictable way is not trusted. If an entity is trusted and its behaviour is predictable then the vehicle route can be planed with high probability that it will not cause unsafe events.

Rules are used both for assuring safety (the rule says that the person will not enter the road) and prediction of entity actions (the person will stay on the pavement). It is easier to predict future actions when an agent follows the rules.

Another problem is completeness and certainty of available information. The situation knowledge can be incomplete because of perception limitations, unavailable information sources or external factors. The information can come from unreliable sources and turn out to be false. Some attributes cannot be measured and have to be assessed. Such assessments can be uncertain.

Some improvement in completeness and certainty of the situation knowledge can be a result of communication. Examples of the road traffic communication are the use of turn indicator lights (vehicle-vehicle communication) and traffic lights at crossings (agent-vehicle communication). When a person stands on the edge of road he or she communicates the intention to cross the road. Communication helps to ensure more complete and certain situation knowledge and allows for more accurate prediction.

Assessing the risk is a heavy cognitive task requiring perception skills, knowledge and experience. Not always it is possible to correctly assess the risk. Humans make mistakes, especially when in stress conditions and have too little time. What humans do to cope with that problem is to be aware of limitations of their own assessments and take it into account (for example not to drive fast in the fog). Humans recognise situations with a potential for causing adverse events. In these situations they are more attentive and careful. Generally complete and certain knowledge of the situation is practically impossible due to the broad scope of information needed and short time

for decision. There is always something that we don't know about the situation. Therefore autonomous machines also need the ability to make assessments based on incomplete knowledge and judge on the credibility of its own assessments.

Summarising the section it can be said that:

- situation awareness is necessary to assess the risk of present and predicted future situations,
- the situation risk assessment depends on trust to other agents that they will act in a predictable way,
- situation assessment can be uncertain due to incompleteness of knowledge and uncertain information sources.

The problem how situation awareness can be used for safety assurance will be discussed for an example of a simple simulated autonomous vehicle presented in the next section.

4 An Example of a Simulated Autonomous Vehicle

Simulation was chosen as the first step of the proposed approach verification. The definition of a simulated autonomous vehicle (SAV) is based on the concept of road traffic on a two-lane road. Vehicles can drive along the road in both directions and each of them has been assigned with its own mission. Some objects are also present in the environment: traffic lights and speed limit signs. Some examples of possible situations are presented in Fig. 2.

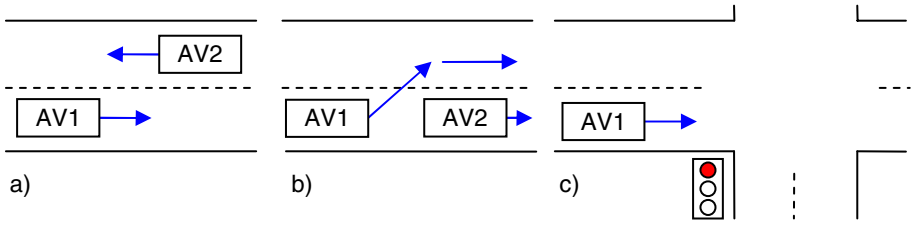

Fig. 2. Examples of simulated autonomous vehicle action scenarios

SAV is equipped with:

- Motors and brakes that enable it to move along the road and change lanes (going off the road is defined as unsafe behaviour and the simulated vehicle is stopped when it leaves the road).
- A set of sensors: speed sensor, clock, position sensor and a set of distance sensors.
- A communication module that allows for communication with other vehicles and objects in a given range.

The objective of SAV is to:

1. accomplish its mission (get to a defined point on the road and then stop),
2. comply with the rules (regulations defined for the environment):
   - drive along the right lane except when overtaking another vehicle,

- do not overtake on crossings,
- stop before the crossing when the traffic light is other then green,
- do not increase the speed when being overtaken,
- keep the speed in the limit according to the traffic signs,
- do not go off the road,

3. avoid unsafe event: a collision with other vehicles or objects.

The simulated AV environment is simplified. The main limitations are:

- Properties of agents and the environment are static. Agents do not have to adapt to new conditions, regulations o changed characteristics of other agents.
- The world is limited to the road. The physics of the simulated world is simplified. Discrete time is used in the simulation (a time tick is defined).
- The image recognition problem is excluded from the analysis (SAV does not have to *see* traffic signs or other agents).
- Objects are considered to be reliable. No object failures are simulated.

The problem to be solved is how to assure safety of presented SAV when situations dynamically change depending on behaviour of other vehicles and objects.

5 Situation Awareness Model for SAV

The situation awareness model is proposed in this section. The model uses the concept of ontology.

Ontology is a description of some domain of real world, which is sometimes called a *universe*, which allows for classification and analysis of some properties [7, 8, 9]. Barry Smith provides the following definition: *Ontology* is the science of what is, of the kinds and structures of objects, properties, events, processes and relations in every area of reality. For an information system, ontology is a representation of some pre-existing domain of reality which:

1. reflects the properties of the objects within its domain in such a way that there obtains a systematic correlation between reality and the representation itself,
2. is intelligible to a domain expert,
3. is formalized in a way that allows it to support automatic information processing.

The situation awareness model presented below has been developed for SAV described in the previous section. VDM-like notation [10] is used in the definition of the model. The model has been designed to be as simple as possible to be suitable for SAV. More sophisticated situation awareness models should be built for more complex systems.

Universe. For the analysed system the universe state is a set of information that SAV "knows" about itself and the environment. This can be defined as follows:

UniverseState ::

InternalState	: *AttributeId* → (*Value* × *Certainty*)
Events	: *SensorId* → (*Value* × *Certainty*)
Environment	: <AV knowledge about environment: terrain map (road)>
Agents	: *AgentId* → *AgentInfo*

AgentInfo = *AgentAttributeId* → (*Value* × *Certainty*)

UniverseState describes the AV knowledge about itself and the environment for a given moment of time. *UniverseState* is not a real world. It is a SAV's knowledge of the world. Ontology is well defined if the model complies with three conditions defined at the beginning of this section.

InternalState incorporates attributes such as the time, position, speed, mission goal, command for actuators. Examples of agents attributes are agent type, position and speed. Information about environment is static and limited to the road coordinates (that's one of the SAV limitations).

Certainty. Knowledge about the situation can be uncertain and therefore *Certainty* type was introduced to the model. Certainty here does not mean objective probability but a result of SAV assessment on how certain is particular value of an attribute. This was achieved by introducing *BasicProbability* type. The Basic probability is a concept from Dempster-Shafer theory [11, 12]:

$$
\begin{aligned}
&\textit{BasicProbability} :: \\
&\quad \textit{belief} : \mathbf{R} \\
&\quad \textit{disbelief} : \mathbf{R} \\
&\quad \mathsf{invariant}(\textit{belief}, \textit{disbelief}) \equiv \\
&\qquad \textit{belief} \geq 0 \ \wedge \ \textit{disbelief} \geq 0 \ \wedge \ \textit{belief} + \textit{disbelief} \leq 1 \\
&\textit{Certainty} = \textit{BasicProbability}
\end{aligned}
$$

BasicProbability is a tuple of two values: *belief* and *disbelief*. *Belief* is an assessed probability that the value is *true*, while *disbelief* indicates assessed probability of *false*. Basic probability can be used to represent certainty that a situation is safe. Value (1, 0) means that there is 100% assessed probability that the situation is safe. Value (0.5, 0.5) relates to fifty-fifty assessment that the situation is safe or not safe.

When the situation assessment is uncertain, the sum of *belief* and *disbelief* is less then 1. Value (0.2, 0.3) means that there is 20% probability that the situation is safe, 30% that it is not safe however the remaining 50% is uncertain – it's not known if it is safe or not. In other words it can be said that because of incomplete knowledge and uncertain assessment the probability of the safe situation is believed to be somewhere from 20% to 70%. Uncertainty is calculated by function:

$$
\begin{aligned}
&\mathsf{uncertainty}: \textit{BasicProbability} \rightarrow \mathbf{R} \\
&\mathsf{uncertainty}(\textit{bp}) \equiv 1 - \textit{bp.belief} - \textit{bp.disbelief}
\end{aligned}
$$

Situation assessment. Basic probability is used to represent assessments if the situation is safe, the rules (regulations) are followed and whether there is a progress in achieving mission goals:

$$
\begin{aligned}
&\textit{SituationAssessment} :: \\
&\qquad \textit{missionProgress} : \textit{BasicProbability} \\
&\qquad \textit{rulesAccordance} : \textit{BasicProbability} \\
&\qquad \textit{safetyLevel} : \textit{BasicProbability}
\end{aligned}
$$

Trust. Trust to other agents and sensors is also represented by basic probability:

$$TrustAssessment = (AgentId \cup SensorId) \rightarrow BasicProbability$$

Situation awareness. *UniverseState* represents knowledge about a situation at a given moment of time. The Situation awareness is not only the knowledge of the current universe state, but also the past and future states.

$$\begin{aligned} SituationAwareness :: \; & now && : Time \\ & observedSituations && : Time \rightarrow UniverseState \\ & predictedSituations && : Time \rightarrow UniverseState \\ & assessment && : SituationAssessment \\ & trust && : TrustAssessment \end{aligned}$$

$$\text{invariant}(\ \text{mk-}SituationAwareness(\ now, observed, predicted, sa, ta\ )\ ) \equiv$$
$$\forall\ t \in \text{dom}(\ observed\ ) \cdot t \leq now$$

There is a question if the perception, prediction and assessment methods should be included as part of the situation awareness model. This has not been done for the presented example. Intuitively the situation awareness model should comprise this kind of knowledge. This knowledge may also evolve in order to adapt to changing environmental conditions, new agent types or new possible accident scenarios. The problem of adapting and learning from experience is out of the scope of this paper.

6 SAV Control System

SAV control system has been designed using the layered architecture [13, 14]. The system is decomposed into policy layer, mission layer, tactical layer and control layer. Situation assessment and action planning is located in the tactical layer and this layer is described in this section. The SAV tactical layer is decomposed into two main processes: Situation analysis (SA) and Task planning (TP) presented in Fig. 3.

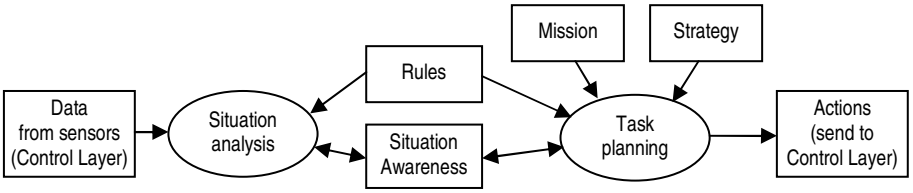

Fig. 3. SAV control system Situation analysis and Task planning processes

The goal of Situation analysis (SA) process is to collect information from available sources to provide situation awareness knowledge. The process is decomposed into following subprocesses:

SA.1. Read sensor values and update *observedSituations*(*now*).*Events* (use sensors *trust* as initial value for data *certainty*)

SA.2. Calculate *observedSituations*(*now*).*Agents* and *InternalState* attributes and check for data consistency (rise or lower trust in appropriate data sources)
SA.3. Update *trust* in other agents depending on their observed behaviour

Task planning process (TP) objective is to predict possible situations and then choose the optimal sequence of actions. The process is decomposed into the following steps:

TP.1. Generate a set of possible scenarios of actions.
TP.2. Assess each scenario for:
  TP.2A. progress towards mission goals,
  TP.2B. compliance with formal rules (regulations),
  TP.2C. situation risk level.
TP.3. Choose the optimal scenario (according to SAV strategy). Update *predicted-Situations* and *assessment*
TP.4. Chosen scenario tasks are communicated to the Control layer.

TP.1 process generates a set of possible action scenarios determined by possible AV actions and predicted behaviour of other agents. Generally the range of possible actions of other agents is very broad. This depends mostly on an agent decision to obey the rules or not. For example for a situation presented in Fig. 2.a an agent AV1 must predict possible set of AV2 actions. If AV2 does not intend to follow the rules then it will be possible that AV2 will suddenly change lane in front of AV1. This could lead to a collision. This would not happen if AV2 followed the regulations. What is needed here is trust. Both agents need to trust each other that they will conform to the regulations (precisely: obey the rule to drive along the right lane).

When the trust in other vehicle is equal to (0.9, 0) it is interpreted as 10% uncertainty that the vehicle will follow regulations. In such case more scenarios of its behaviour would be analysed.

The effectiveness of the situation risk assessment generally depends on two factors. The first is the span, certainty and completeness of the prediction (TP.1). The second factor is the ability of TP.2C function to assess the risk depended on the current situation knowledge. The value of joining risk assessment with prediction algorithm comes from the possibility of examining possible actions for hazard avoidance when unexpected events occur. Even for high risk situations the prediction function will provide for possible scenarios to reach safe state (propose actions that lead to situations with lower risk level).

Two safety assessment functions TP.2B and TP2C have been defined. The first one uses static rules (regulations). The second assessment function examines current and expected future situations to judge the possibility of a hazard. These two assessment functions are both necessary for safe operation of the vehicle. The comparison of these two functions is presented in Table 2.

Three different assessment functions are used in the TP.2 process and a question arises what strategy should be used to choose the so-called optimal scenario (TP.3 process). Many different strategies are possible. Three of them are:

1. *Care for yourself* strategy – if there are some risk factors (*safetyLevel.disbelief* > 0) then choose the safest scenario or otherwise select the scenario with the best combination of assessment values. In case of any risk factor the vehicle will ignore all rules and will perform any action to achieve no-risk state.

2. *Stick to the rules* strategy – always choose the scenario with the highest TP.2B assessment, then the second priority is the safety level and mission progress.
3. *Mission first* strategy – always choose scenario with the best mission progress assessment (this strategy can be used to implement hostile vehicles).

Table 2. Rule-based and risk-based assessment function comparison

Rules Compliance (TP.2B)	Situation Risk Assessment (TP.2C)
Based on static rules	Based on dynamic risk assessment
Easier for design and verification	More difficult for design and verification
Supports external judgment if a vehicle obeys the rules	Supports decision making when planning individual and group actions
Can be used to assure safety in normal situations. May not assure safety if rules are violated in the environment	Can be used to assure hazard avoidance in any situations
Will not assure safety if rules are incomplete for a context of a given environment	Will not assure safety if the knowledge is incomplete or uncertain in context of a given environment
When function fails: breaking the rule will sometimes (but not always) cause a hazard depending on the situation	When function fails: choosing unsafe scenario of actions will lead to hazard if no other countermeasures are taken
Is required in order to assure compliance with the regulations	Is required in order to assure hazard avoidance in open environment

A question comes to mind what strategy is the right one. Even when more sophisticated strategies are defined, the question remains when an autonomous agent can violate regulations in order to avoid hazard. Never? Whenever it perceives any risk factor? Only when the risk is high? How high? All rules can be violated or only some of them? And what if the agent risk assessment is wrong? These are hard questions to answer.

7 The Problem of Autonomous System Hazard Analysis

The autonomous vehicle is considered to be a safety-critical system therefore demonstrating that the solution assures safety is essential. This can be done by developing a safety case for the system. The system architecture is more complex then for most embedded systems and to develop such a safety case can be a nontrivial task.

The first problem encountered when constructing the safety case was the decision how the hazards should be defined. The first attempt was to define accepted probability of hazard occurrence. This led to probabilistic claims and assumptions on external events (like behaviour of other vehicles). Quantitative arguments were left off for the moment and the analysis focus was on what situations and combinations of events can lead to hazard (qualitative arguments).

The argument for the claim that the hazard will be avoided (e.g. SAV will not collide) intuitively leads to safety requirements like: the Task Planning process will not give unsafe action plan as an output. A fault tree-like analysis can be used to track down the causes for such event and then define requirements for processes and their input data. Such analysis on the components level has been carried out. The result of this analysis was a set of safety requirements. Some of them relate to situation awareness model and assessment functions.

Justification for claims like completeness of situation awareness model, correctness of prediction process TP.1 or correctness of risk assessment function TP.2C requires providing evidence based on the environment and AV mechanical engineering models and accident models. This is an area of future research.

8 Extending SAV with a Collaboration Process

The vehicle presented in Section 4 is fully autonomous. It does not cooperate with other agents and does not exchange any information with them. Cooperation can offer possibility for:

- verification of situation awareness knowledge by comparison to information from other agents,
- more reliable prediction of other agents behaviour,
- using recommendation mechanism for more accurate assessments of trust in other agents.

These properties strengthen justification for some safety requirements identified in the safety case.

Extending AV with a communication process is a big change of autonomous vehicle characteristics. A set of non-communicating autonomous vehicles is now transformed into a set of vehicles which cooperate to assure safety although each of them has its own mission.

The proposed solution is to extend Situation analysis process (described in section 6) with additional process decomposed into four steps:

1. Communicate with other agents to exchange a subset of situation awareness information. The scope of the information exchange depends on the AV strategy. For the simulated AV the data exchange scope is the vehicle identification data, position, speed, planned actions and also trust levels to other vehicles.
2. Analyse consistency of the recommendations and own trust assessments and accordingly update trust in a recommended agent, the recommender and own trust assessment function data sources.
3. Consistency check – for each situation awareness attribute find related data received from other agents, if found then check consistency and adjust the data, its certainty and the data sources trust levels according to the detected data consistency or discrepancy.
4. Situation awareness extention – add information on other agents planned actions.

The proposed cooperation process makes stronger arguments for justification of claims like completeness of situation awareness model, correctness of prediction process TP.1 or correctness of risk assessment function TP.2C.

9 Experiments Results

A simulation tool has been designed and developed using Java development environment. Some number of simulation experiments has been carried out to verify the approach. The main limitation of the simulation tool is that AV processes are run as sequential tasks for each simulated time tick. No concurrency has been implemented and no time requirements have been analysed.

Analysis of simulation results is difficult except for simple situations. For a given scenario the risk and trust evolve in time but have no directly measurable values in the real world. The problem is that humans also have different perception of how risky are some situations. Therefore justification for a particular risk value for a given situation is usually questionable. Risk assessment function results depend on many parameters. Changing slightly some parameters can sometimes cause big change in the risk assessment value. Designing the risk assessment algorithms is a non-trivial task even for the simple simulated AV. The conclusion of the experiments is that the critical issue is to start with explicit definition of safe and unsafe situations, which are denoted as the extreme *BasicProbability* values (1,0) and (0,1).

The safe level assessment value (1,0) was defined as lack of any risk factors and full certainty of the situation knowledge. The risk factors are derived from the accident model analysis. The safety level value (0,1) has been defined as accident (vehicles collide). It is not required to justify the exact risk assessment values however risk assessment consistency should be justified (the same risk value should be assigned to the situations of the same risk level).

The SAV accident scenario analysis was made manually and tool support for this task is needed. The plan is to use accident scenario risk profiles. A risk profile is a chart showing change of safety level in time together with labels for relevant events. The objective of using risk profiles is to ensure that the assessment is consistent with the concept of the increasing risk level for accident scenarios (compare to Table 1) and to stretch the period of time when the risk is perceived.

Another issue analysed in the experiments was what initial trust level should be assigned to other vehicles. Three approaches were tested: full trust, uncertain or no trust. Also the ways how humans assess trust have been analysed. This led to the first impression mechanism. Humans usually judge any person within the first three seconds when they meet. This strategy was chosen as the best one. Initial trust level for any new vehicle should be uncertain, denoted as value (0,0). For some short period of time the trust assessment function should be more sensitive to make wide range of trust adjustment possible. When the trust level is stabilized, the function should become less sensitive. That leads to the problem which factors should be considered as relevant for initial trust or distrust. Some strategies for building trust (e.g. regulations conformance, identifying itself to other vehicles, cooperation) have been analysed. The mechanism of the first impression needs further research.

Some number of scenarios have been simulated and analysed.

Scenario 1 (see Fig 2.a in section 4) relates to a situation when two vehicles pass each other on separate lanes. The situation risk level depends on the distance, speed and direction of the vehicles and available space (limited to the road by the rule). When AV does not fully trust other vehicle then it tries to keep safe distance. That may lead to speed decrease and turning to the road side. The safe passing scenario (Fig. 2.a) is possible when vehicles follow the rules and trust each other.

Scenario 2 (see Fig. 2.b) relates to a situation when one vehicle is followed by another one and finally can be overtaken. The first issue is the safe distance between the two vehicles. When vehicles cooperate and trust each other the distance can be smaller. When there is no trust relation then the safe distance is longer. The second issue is the decision when to overtake safely. Some problems were caused by limitations of the prediction function as the predicted scenario was shorter then the overtake manuevre. Longer scenario would ease the risk assessment. Some work was needed to elaborate a risk assessment function that would preserve safety when for example another vehicle is approaching on the left lane or there is a crossing not far away. Effect of safe overtaking was achieved after some experimental trials however the systematic process for accident model analysis is needed.

Scenario 3 (see Fig. 2.c) was introduced to investigate how cooperating vehicles behave in dynamically changing environment. The first problem encountered was the influence of action plan changes on trust. In the tested scenario two vehicles were driving one following another. The first vehicle stopped when the traffic lights changed from green to yellow. When the second vehicle noticed the change in planned actions it lowered the trust level in AV1. This approach is too simplified. The justification for other vehicle plan change should be assessed before the trust level is altered. Analysis how this can be achieved is a possible future work.

Another problem encountered in the experiments was that the risk level definition does not take in account the hazard consequences. The conclusion was to extend the situation risk definition from *BasicProbability* to a set of tuples containing accident consequences (the loss) and probability assessment. Quite interesting is that such model could be used for risk assessment compliant with Three Laws of Robotics defined by Asimov [15]. An example of such situation risk assessment can be a set { (*human-harm*, (0, 0.9)), (*robot-damage*, (1,0)) }. Other possibilities for the situation awareness model extensions are also analysed.

10 Summary

The main conclusion of the paper is that the situation awareness is the key factor in autonomous vehicles safety assurance. Autonomous vehicles need to be able to perceive current situation, assess it and predict future situations. Situation awareness model should be built on a sound ontology which describes the vehicle and its environment. The model should provide means to cope with the problems of trust in other agents and uncertain and incomplete knowledge. The proposed solution is based on simple patterns of human situation awareness.

Three situation assessment functions were distinguished: situation risk level assessment, regulations compliance assessment and mission progress assessment. Hazard avoidance is dependent on the perception and assessment of the situation risk.

Communication and cooperation has been proposed to strengthen safety in a situation of dynamic agent interactions.

The proposed approach was demonstrated on a small simulated example of an autonomous vehicle. For most scenarios the simulated vehicle could perceive unsafe situations and avoid them. Limitations of the method and experiences from the experiments have been discussed in Section 8.

Systematic process for accident model analysis and building safety argument is an area of planned research. Plans for future work include also application of presented approach to a laboratory autonomous vehicle and extending the situation awareness model. A challenge that is foreseen is to enable AV to adapt to changing environment characteristics. Autonomy is a novel concept for safety-critical systems and will require a lot of research work to provide sound arguments for safety justification.

References

1. Dennet, D. C.: Did HAL Commit Murder?, Cogprints Cognitive Science E-print Archive, http://cogprints.org/430/00/didhal.htm (1998)
2. Fatal Accident Summary Report: Die Cast Operator Pinned by Robot, FACE Report 8420, National Institute for Occupational Safety and Health, http://www.cdc.gov/niosh/face/In-house/full8420.html (2005)
3. Huang Y., Ljung M., Sandin J., Hollnagel E.: Accident models for modern road traffic: changing times creates new demands, IEEE International Conference on Systems, Man and Cybernetics (2004)
4. Endsley, M. R.: Direct Measurement of Situation Awareness: Validity and Use of SAGAT in (Eds.) Mica R. Endsley and Daniel J. Garland, Situation Awarenss Analysis and Measurement (2000)
5. Reason J.: Human Error, Cambridge University Press, (2000)
6. French H. T., Hutchinson A.: Measurement of Situation Awareness in a C4ISR Experiment, In: The 7th International Command and Control Research and Technology Symposium, Quebec City (2002)
7. Goczyła K., Zawadzki M.: Processing and inferring from knowledge of different trust level, In: Kozielski S., Małysiak S., Kasprowski P., Mrozek P. (eds.): Data Bases – Models, Technologies, Tools, WKŁ (2005) 207-213 (in Polish)
8. Matheus C. J., Kokar M. M., Baclawski K.: A Core Ontology for Situation Awareness, International Conference of Information Fusion, IEEE, Cairns, Australia (2003) 545-552
9. Smith B.: Ontology, in Floridi (ed.) Blackwell Guide to the Philosophy of Computing and Information, Oxford: Blackwell (2003) 155–166
10. Jones, C.: Systematic Software Development using VDM, Prentice Hall International (1990)
11. Shafer, G.: Mathematical theory of evidence, Princetown University Press (1976)
12. Górski J., Zagórski M.: Using Dempster-Shafer approach to support reasoning about trust in IT infrastructures, In: Warsaw International Seminar on Intelligent Systems (WISIS 2004), Warsaw (2004) 39-57
13. Kelly, A. J.: Predictive Control Approach to the High-Speed Cross-Country Autonomous Navigation Problem, Ph.D. thesis, Carnegie Mellon University, Pittsburg (1995)
14. Broten G., Monckton S.: Unmanned Ground Vehicle Electronic Hardware Architecture – A Flexible and Scalable Architecture for Developing Unmanned Ground Vehicles, Defence R&D Canada – Suffield TM 2004-122 (2004)
15. Asimov I., I, Robot (1950)