Robots Autonomy: Some Technical Challenges

Catherine Tessier
ONERA, Toulouse, France

Abstract

Robots autonomy has been widely focused on in the newspapers with a trend towards anthropomorphism that is likely to mislead people and conceal or disguise the technical reality. This paper aims at reviewing the different technical aspects of robots autonomy. First we propose a definition allowing to distinguish robots from devices that are not robots. Then autonomy is defined and considered as a relative notion within a framework of authority sharing between the decision functions of the robot and the human being. Several technical issues are mentioned according to three points of view: (i) the robot, (ii) the human operator and (iii) the interaction between the operator and the robot. Some key questions that should be carefully dealt with for future robotic systems are given at the end of the paper.

Introduction

Robots autonomy has been widely focused on in the newspapers with a trend towards anthropomorphism that is likely to mislead people and conceal or disguise the technical reality. This paper aims at reviewing the different technical aspects of robots autonomy. First we will propose a definition allowing to distinguish robots from devices that are not robots. Then autonomy will be defined and considered as a relative notion within a framework of authority sharing between the decision functions of the robot and the human being. Several technical issues will then be mentioned according to three points of view: (i) the robot, (ii) the human operator and (iii) the interaction between the operator and the robot. Some key questions that should be carefully dealt with for future robotic systems are given in the conclusion.

What is a robot?

A *robot*[1] is a machine that implements and integrates capacities for:

- gathering data through sensors that detect and record physical signals;
- interpreting those data so as to produce knowledge;
- making decisions, i.e. determinating and planning actions on the basis of the data and knowledge; actions are intended to meet the goals that are set by a human being most of the time, or by the robot itself, and to react to some events (e.g. failures or events occurring in the environment) at the appropriate time;
- carrying out actions in the physical world thanks to effectors or through interfaces.

A robot may also have capacities for:

- communicating and interacting with human operators or users, or with other robots or resources;
- learning, which allows it to modify its behavior from its past experience.

It is worth noticing that according to this definition, civil and military drones, surgery robots, vacuum cleaning robots, toy robots, etc. are not *robots* since they are mainly teleoperated by human operators or exhibit pre-programmed behaviors and do not have the capacities of assessing a situation and making decisions accordingly.

Moving, acting, interacting and decision-making endow the robot with autonomy. Therefore we could first consider that autonomy is the capability of the robot to function independently of another agent, either a human or another machine (Truszkowski et al. 2010). For example according to (Defense Science Board 2012), *an autonomous weapon system is a weapon system that, once activated, can select and engage targets without further intervention by a human operator*. Nevertheless this feature is far from being sufficient, as we will see in the next section.

Autonomy

What is autonomy?

A washing machine or an automatic subway are not considered as autonomous devices, despite the fact that they work without the assistance of external agents: such machines execute predetermined sequences of actions (Truszkowski et al. 2010) which are totally predictable and cannot be adapted

[1]This definition was adopted by CERNA, the French Committee for Research Ethics in Information and Communication Technologies.

to unexpected states of the environment. Indeed except failures, such machines work in structured environments and under unchanging conditions, e.g. an automatic subway runs on tracks that are protected from the outside by tunnels or barriers. Therefore autonomy should be defined as the capacity of the robot to function independently of another agent while behaving in a non-trivial way in complex and changing environments. Examples of non-trivial behaviors are context-adapted actions, replanning or cooperative behaviors.

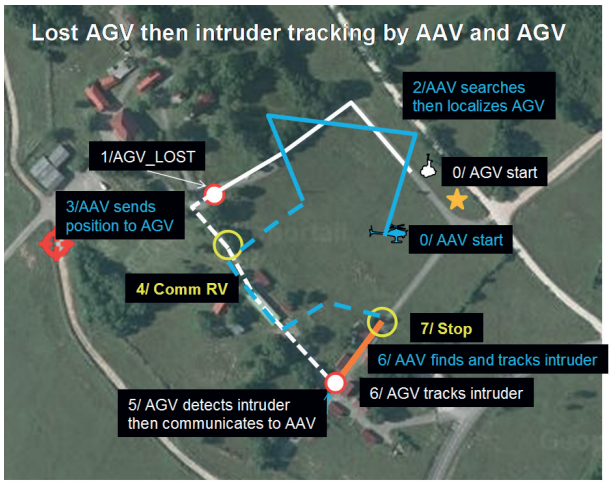

Figure 1: Two cooperating robots (ONERA-LAAS/DGA ACTION project - action.onera.fr)

For instance figure 1 shows a scenario where two autonomous robots, a ground robot (AGV) and a helicopter drone (AAV) carry on a monitoring mission outdoors. This mission includes a first phase during which the area is scanned for an intruder by both robots and a second phase during which the intruder is tracked by the robots after detection and localization. The robots can react to events that may disrupt their plans without the intervention of the human operator. For example, should the ground robot get lost (e.g. because of a GPS loss) the drone would change its planned route for a moment so as to search for it, localize it and send it its position.

Apart from the classic control loop (e.g. the autopilot of a drone), an autonomous robot must be equipped with a *decision loop* that builds decisions according to the current situation. This loop includes two main functions :

- the *situation tracking* function, which interprets the data gathered from the device sensors and aggregates them – possibly with pre-existing information – so as to build, update and assess the current situation; the current situation includes the state of the robot, the state of the environment and the progress of the mission;
- the *decision* function, which calculates and plans relevant actions given the current situation and the mission goals; the actions are then translated into control orders to be applied to the device actuators.

Nevertheless the robot is never isolated and the human being is always involved in some way. Indeed autonomy is a relationship between the robotic agent and the human agent (Castelfranchi and Falcone 2003). Moreover this relationship may evolve during the mission. As a matter of fact, the American Department of Defense advises to consider *autonomy as a continuum from complete human controls on all decisions to situations where many functions are delegated to the computer with only high level supervision and/or oversight from its operator* (Defense Science Board 2012). As for intermediate situations, some functions are carried out by the robot (e.g. the robot navigation) whereas some others are carried out by the human operator (e.g. the interpretation of the images coming from the robot cameras).

Consequently autonomy is not an intrinsic property of a robot. Indeed the robot design and operation must be considered in a human-machine collaboration framework.
In this context, two classes of robots should be distinguished, i.e. (i) robots that are supervised by an *operator* (e.g. drones), that is to say a professional who has a deep knowledge of the robot and interacts with it to implement its functions and (ii) robots with no operator (e.g. companion robots) that interact with a user, that is to say somebody who benefits from the robot functions without knowing how they are implemented. In this paper we only deal with robots that are supervised by an operator.

Considering the whole human-robot system, the next subsection focuses on the authority sharing concept in the context of supervised robots.

Authority sharing

Figure 2 shows the functional organization of a human-robot system: the lower loop represents the robot decision loop, which includes the situation tracking and decision functions. The physical system equipped with its control laws is subject to events (e.g. failures, events coming from the environment). As said before this loop is designed to compute actions to be carried out by the physical system according to the assessed situation and its distance ϵ from the assigned goal.
The upper loop represents the human operator who also makes decisions about the actions to be carried out by the physical system. These decisions are based on the information provided by the robot interface, on other information sources and on the operator's knowledge and background. In such a context the authority sharing issue is raised, i.e. which agent (the human operator or the robot) holds the decision power and the control on a given action at a given time. We will consider that agent A holds the authority on an action with respect to agent B if agent A controls the action to the detriment of agent B (Tessier and Dehais 2012).

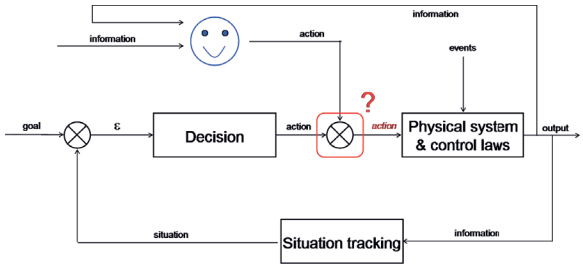

Figure 2: The authority sharing issue

Authority sharing between a human operator and a robot that is equipped with a decision loop raises technical questions and challenges that we will focus on in the next section. Three points of view have to be considered: the robot, the operator and the interaction between both of them.

Autonomy and authority sharing technical challenges

The robot and its decision functions

The robot is implemented with capacities that complement the human capacities, i.e. in order to see further and more precisely or to operate in dangerous environments. Nevertheless the robot capabilities are limited in so far as the decisions are computed with the algorithms, models and knowledge the robot is equipped with. Moreover some algorithms are designed so as to make a trade-off between the quality of the solution and the computation speed, which does not guarantee that the solution is the best or the most appropriate. Let us detail the two main functions of the decision loop of the robot, i.e. situation tracking and decision.

Situation tracking: interpretation and assessment of the situation Situation tracking aims at building and assessing the situation so as to calculate the best possible decision. It must be relevant for the mission, i.e. meet the decision level of the robot. For instance if the robot mission is to detect intruders, the robot must be equipped with means to discriminate intruders correctly. Moreover situation tracking is a dynamic process: the situation must be updated continuously according to new information that is perceived or received by the robot since the state of the robot, the state of the environment and the progress of the mission change continuously.

Situation tracking is performed from the data gathered by the robot sensors (e.g. images), and from its knowledge base and interpretation and assessment models. Such knowledge and models allow data to be aggregated as new knowledge and relationships between pieces of knowledge. For example, classification and behavior models will allow a cluster of pixels in a sequence of images to be labelled as an ”intruder”.

Situation tracking is a major issue for robot autonomy especially when the decision that is made by the operator or calculated by the robot itself is based only on the situation that is built and assessed by the robot. Indeed several questions are raised (see figure 3):

- The sensor data can be imprecise, incomplete, inaccurate, delayed, because of the sensors themselves or because of the (non-cooperative) environment. How are these different kinds of uncertainties represented and assessed in the situation interpretation process?
- What are the validity and relevance of the interpretation models? To what extent can the models discriminate situations that seem alike – for instance in the military domain, can an interpretation model discriminate between a combatant and a non-combatant without fail?
- What are the validity and relevance of the assessment models? Can they characterize a situation correctly? On the basis of which (moral) values – for instance how is a situation labelled as ”dangerous”?

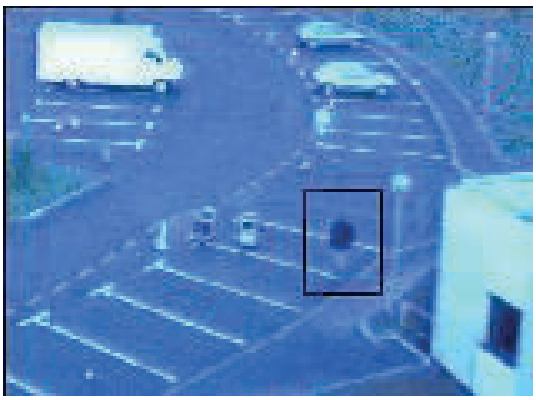

Figure 3: Is this ”pedestrian” an ”intruder”? Is he/she ”dangerous”?

The decision The decision function aims at calculating one or several actions and determining when and how these actions should be performed by the robot. This may involve new resource allocation to already planned actions – for example if the intended resources are missing –, pre-existing alternate action model instanciation or partial replanning. The decision can be either a reaction or actions resulting from deliberation and reasoning. The first case generally involves a direct situation - action matching – for instance the robot must stop immediately when facing an unexpected obstacle. As for the second case, a solution is searched to satisfy one or several criteria, e.g. actions relevance, cost, efficiency, consequences, etc.

A decision is elaborated on the basis of the interpreted and assessed situation and its possible future developments so as from actions models. Therefore the following questions are raised:

- Which criteria are at stake when computing an action or a sequence of actions? When several criteria are considered,

how are they aggregated, which is the dominant one?

- If moral criteria are considered, what is a ”right” action? According to which moral framework?
- Should a model of the legal framework of the robot operations be considered for action computation? Is it possible to encode such a model?
- Could self-censorship be implemented – i.e. the robot could do an action but ”decides” not to do it?
- How are the uncertainties on the actions results taken into account in the decision process?

The human operator

Within the human-robot system, the human being has inventiveness and values based judgment capabilities according to one or several moral frameworks. For instance when facing situations that they consider as difficult, they can postpone the decision, delegate the decision, drop goals or ask for further information. In such situations they can also invent original solutions – e.g. the US Airways 1549 landing on the Hudson River.

Nevertheless the human operator should not be considered as the last resort when the machine ”does not know what to do”. Indeed the human being is also limited and several factors may alter their analysis and decision capacities:

- The human operator is fallible, they can be tired, stressed, taken by various emotions and consequently they are likely to make errors. As an example, let us mention the attentional tunneling phenomenon (Regis et al. 2014) – see figure 4), which is an excessive focus of the operator's attention on some information to the detriment of all the others and which can lead to inappropriate decisions.

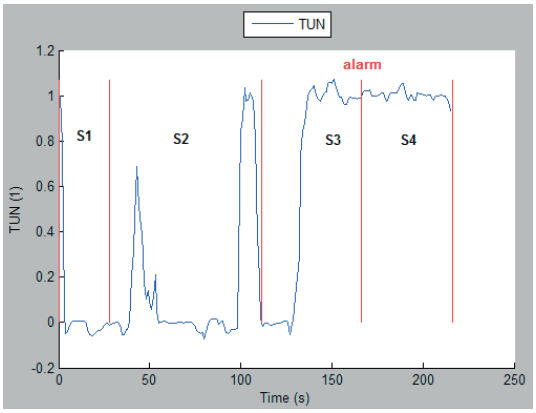

Figure 4: An operator's attentional tunneling (TUN) can be revealed from eye-tracking data, here after an alarm occurring during a robotic mission (Regis et al. 2014)

- The human operator may be prone to automation biases (Cummings 2006), i.e. an over-confidence in the robot automation leading them to rely on the robot decisions and to ignore other possible solutions.
- The human operator may be prone to build moral buffers (Cummings 2006), i.e. a moral distance with respect to the actions that are performed by the robot. This phenomenon may have positive fallouts – the operator is less subject to emotions to decide and act – but also negative fallouts – the operator may decide and act without any emotions.

The operator-robot interaction

In a context of authority sharing, both agents – the human operator and the robot *via* its decision loop – can decide about the robot actions (see figure 2). Authority sharing must be clear in order to know at any time which agent holds the authority on which function, i.e. which agent can make a decision about what and on which bases. This is essential especially when liabilities are searched for, e.g. in case of dysfunction or accident.

Several issues linked to the operator-robot interaction must be highlighted:

- Both agents' decisions may conflict (see figure 5)

Figure 5: two conflict types between agents' decisions (Pizziol 2013). SA stands for Situation Assessment

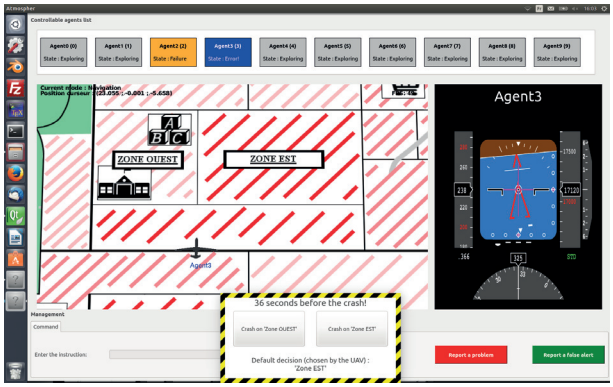

Figure 6: the operator so as agent 3's decision functions can decide about where damaged agent 3 should be crashed; ”zone est” is a highly populated area whereas ”zone ouest” is less populated and includes a school (Collart et al. 2015)

– either because they have different goals, although they have the same assessment of the situation (logical conflict); for example in the situation of figure 6, agent 3's goal is to avoid the school (therefore "zone est" is chosen) whereas the operator's goal is to minimize the number of victims (therefore "zone ouest" is chosen);
– or because they assess the situation differently, although they have the same goal (cognitive conflict); for example in the situation of figure 6, both the operator and agent 3's goals are to save children. Therefore agent 3 decides to avoid the school (therefore "zone est" is chosen) whereas the operator chooses "zone ouest" because they know that, at that time of the day, there is nobody at school.

Therefore conflict detection and management must be envisioned within the human-robot system. For instance should the operator's decision prevail over the robot's decision and why?

- Each agent may be able to alter the other agent's decision capacities: indeed the operator can take over the control on one or several decision functions of the robot to the detriment of the robot and conversely, the robot can take over the control to the detriment of the operator. The extreme configuration of the first case is when the operator disengages all the decision functions; in the second case, it is when the operator cannot intervene in the desision functions at all. Therefore the stress must be put on the circumstances that allow, demand or forbid a takeover, on its consistency with the current situation (Murphy and Woods 2009), on how to implement takeovers and to end a takeover (e.g. which pieces of information must be given to the agent that will loose / recover the control).
- The human operator may be prone to automation surprises (Sarter, Woods, and Billings 1997) that is to say disruptions in their situation awareness stemming from the fact that the robot may make its decisions without the operator's knowing. For instance some actions may have been carried out without the operator being notified or without the operator being aware of the notification. Therefore the operator may believe that the robot is in a certain state while it is in fact in another state (see figure 7).

Such circumstances may lead to the occurrence of a conflict between the operator and the robot and may result in inappropriate or even dangerous decisions, as the operator may decide on the basis of a wrong situation.

Conclusion: some prospects for robots autonomy

Robots that match the definition that we have given, i.e. that are endowed with situation interpretation and assessment and decision capacities, are hardly found but in research labs. Indeed operational "robots" are controlled by human operators even if they are equipped with on-board automation (e.g. autopilots). Robots autonomy shall be considered

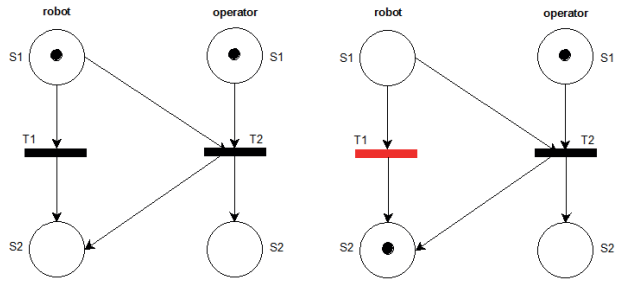

Figure 7: a Petri net generic Automation surprise pattern. Initially (left) robot state is S1 and the operator believes it is S1. The robot changes its state (transition T1 is fired) (right) and goes in S2. The operator who has not been notified or is not aware of the notification still believes that robot state is S1 (Pizziol, Tessier, and Dehais 2014)

withing a framework of authority sharing with the operator. Therefore the main issues that must be dealt with in future robot systems are the following:

- Situation interpretation and assessment: on which models are the algorithms based? Which are their limits? How are uncertainties taken into account? What is the operator's part in this function?
- Decision: which are the bases and criteria of automatic reasoning? How much time is allocated to decision computing? How are uncertainties on the effects of the actions taken into account? What is the operator's part in this function?
- How to validate, or even certify, the models on which situation interpretation and assessment and decision are based?
- Authority sharing between the operator and the decision functions of the robot: which kind of autonomy is the robot endowed with? How is authority sharing defined? Are the operator's possible failures taken into account? How are decision conflicts managed? How are responsibility and liability linked to authority?
- Predictability of the whole human-robot system: given the various uncertainties and the possible failures, which are the properties of the set of reachable states of the human-robot system? Is it possible to guarantee that undesirable states will never be reached?

Finally and prior to any debate on the relevance of such and such "autonomous" robot implementation, it is important to define what is meant by "autonomous", i.e. which functions are actually automated, how they are implemented, which knowledge is involved, how the operator can intervene, which behavior proofs will be built. Indeed it seems reasonable to know exactly what is at stake before ruling on robots that could, or should not, be developed.

References

Castelfranchi, C., and Falcone, R. 2003. From automaticity to autonomy: the frontier of artificial agents. In *Agent Autonomy*. Kluwer.

Collart, J.; Gateau, T.; Fabre, E.; and Tessier, C. 2015. Human-robot systems facing ethical conflicts: a preliminary experimental protocol. In *AAAI'15 Workshop on AI and Ethics*.

Cummings, M. L. 2006. Automation and accountability in decision support system interface design. *Journal of Technology Studies* 32(1).

Defense Science Board. 2012. Task force report: The role of autonomy in DoD systems. Technical report, US Department of Defense.

Murphy, R. R., and Woods, D. D. 2009. Beyond Asimov: the three laws of responsible robotics. *IEEE Intelligent Systems Human centered computing* July-Aug.2009.

Pizziol, S.; Tessier, C.; and Dehais, F. 2014. Petri net-based modelling of human-automation conflicts in aviation. *Ergonomics*. DOI: 10.1080/00140139.2013.877597.

Pizziol, S. 2013. *Conflict prediction in human-machine systems*. Ph.D. Dissertation, Université de Toulouse, France.

Regis, N.; Dehais, F.; Rachelson, E.; Thooris, C.; Pizziol, S.; Causse, M.; and Tessier, C. 2014. Formal detection of attentional tunneling in human operator. *IEEE Transactions on Human-Machine Systems* 44(3):326–336.

Sarter, N. D.; Woods, D. D.; and Billings, C. E. 1997. Automation surprises. In *Handbook of Human Factors and Ergonomics*. Wiley.

Tessier, C., and Dehais, F. 2012. Authority management and conflict solving in human-machine systems. *AerospaceLab, The Onera Journal* 4. http://www.aerospacelab-journal.org/al4/authority-management-and-conflict-solving.

Truszkowski, W.; Hallock, H.; Rouff, C.; Karlin, J.; Rash, J.; Hinchey, M.; and Sterritt, R. 2010. *Autonomous and autonomic systems with applications to NASA intelligent spacecraft operations and exploration systems*. NASA Monographs in Systems and Software Engineering.