(English below)

Que verront les robots de demain ?

De plus en plus au contact d'êtres humains, les robots autonomes ont besoin non seulement d'interpréter le monde qui les entoure mais aussi de s'y repérer. Les caméras peuvent répondre à ce double défi.

Par Jacques Henno

Publié le 2 déc. 2025 à 10:00

Le petit robot mesure à peine 20 centimètres de long sur 15 de large, mais il est capable de suivre, à la vitesse de 1,5 mètre par seconde un chemin après l'avoir mémorisé en une seule fois. A l'arrivée, il n'a dévié que de quelques centimètres par rapport à la route prévue. Antcar n'a pas de GPS, ne coûte que 150 euros, fonctionne avec un nano-ordinateur Raspberry Pi 4, une caméra à 360° en mode dégradé qui ne capte que les grandes masses autour d'elle et un réseau de neurones inspiré de celui des insectes.

« En reproduisant la vision et le cerveau des insectes, Antcar n'a besoin que de 20 kilo-octets d'information pour effectuer un trajet de 50 mètres, soit l'équivalent d'un courriel sans photo », s'enthousiasme Julien Serres, membre de l'équipe Systèmes Bio-Inspirés qui a développé Antcar, à l'Institut des Sciences du Mouvement Etienne-Jules Marey (Aix-Marseille Université-CNRS). L'intérêt ? « Notre système pourrait servir de roue de secours à un robot autonome : en cas de perte du signal GPS en intérieur ou dans un 'canyon urbain', il serait capable de retourner sur ses pas pour retrouver une connexion ou de rentrer au dépôt », ajoute Julien Serres.

« En robotique, le marché le plus prometteur actuellement est celui des robots autonomes : robots-taxis, robots collaboratifs en usine, robots humanoïdes à domicile…, constate Fabien Moutarde, directeur du Centre robotique de Mines Paris. Ces machines seront en permanence au contact d'êtres humains. Elles doivent donc parfaitement percevoir cet environnement, très instable, pour s'y adapter. Tout en consommant le moins d'énergie possible. »

Or les caméras sont des capteurs peu chers et très polyvalents. « La vision permet tout d'abord à un robot de reconnaître les objets qui l'entourent, de les identifier et parfois de les manipuler, détaille Franck Ruffier, spécialiste de la robotique bio-inspirée au laboratoire Lab-STICC (ENSTA Brest) et codéveloppeur d'Antcar. Mais une vision avec moins de pixels permet aussi de se localiser dans l'espace, par rapport à des contrastes créés par d'imposants

objets, comme des immeubles, ou même avec la polarisation de la lumière, dont l'évolution, après calculs, 'indique' le nord géographique. »

Reste à choisir le modèle de caméra le plus adapté au robot. Puis à traiter de façon les données transmises. Tout en gardant un oeil sur les questions éthiques que posent des robots désormais capables, grâce à leur caméra, de deviner nos modes de raisonnement (voir encadré ci-dessous).

Dans certains cas, les robots demeurent aveugles et doivent être guidés de l'extérieur. « Il existe des caméras de 1 millimètre sur 1 millimètre que l'on peut intégrer dans des endoscopes de 2 millimètres de diamètre, explique Brahim Tamadazte, spécialiste de la microrobotique à l'ISIR (Institut des Systèmes Intelligents et de Robotique - Sorbonne Université-CNRS-Inserm). Mais pour les robots utilisés pour réparer des vaisseaux à l'intérieur de l'oeil ou sélectionner des spermatozoïdes lors d'une FIV, on doit suivre leurs déplacements grâce à un capteur dit 'extéroceptif' : un microscope optique ou électronique à balayage pour les nanorobots. »

Dans les autres cas, la panoplie est large : radar, caméras perspective (elles équipent les smartphones), caméras fisheye avec une vision à 120°, omnidirectionnelle à 360° et les petites dernières, les caméras événements, très prometteuses dans des environnements mouvants. « Les caméras événements sont constituées de pixels indépendants qui envoient un signal dès qu'ils détectent un changement de luminosité, décrit Davide Scaramuzza, directeur du groupe Robotique et Perception à l'université de Zurich. Comme elles émettent très peu de données, le traitement de ces informations est très rapide et donc la réactivité du robot est très élevée. » Ce très faible temps de latence - de l'ordre de 0,2 milliseconde - est très utile pour les voitures autonomes, les drones, les robots dans l'espace ou le contrôle qualité.

Une fois la caméra choisie, il faut parfois redresser ses images. « Plus le champ de vision est grand - 120°, 360° -, plus les bords de l'image sont distordus », résume Guillaume Caron, codirecteur du JRL (Joint Robotics Laboratory), commun au CNRS et à l'AIST (National Institute of Advanced Industrial Science and Technology - ministère de l'Economie japonais), à Tsukuba. « On risque donc de perdre les informations que ces bords contiennent, quand l'image est projetée sur un plan pour pouvoir être traitée par un ordinateur. » Un problème vieux comme les cartes qui sont une représentation aplatie d'une sphère… Avec Antoine André, chercheur au JRL, Guillaume Caron a mis au point une méthode de calcul qui permet à l'ordinateur de reconstituer la sphère d'origine tout en conservant des temps de traitement acceptables. « Cela va permettre à

un robot de service de reconnaître la pièce dans laquelle il se déplace », avance Antoine André.

Une fois obtenue une représentation mathématique de l'image de bonne qualité et compréhensible par l'ordinateur du robot, celui-ci doit en tirer des informations qui fassent « sens ». Par exemple, dans le cas d'une voiture autonome, « à droite, il y a un piéton qui risque de traverser ». « C'est la sémantisation des images, pointe Fabien Moutarde. Depuis trois ans sont apparus en intelligence artificielle des VLM, des modèles vision-langage capables de faire le lien entre l'apparence visuelle et la sémantique. » En mode conduite assistée, une voiture pourra ainsi dire au conducteur « si tu tournes à droite, fais attention au vélo qui arrive par-derrière ». De nouveaux modèles, dits « VLAM », pour vision-langage-action, devraient même permettre à un robot d'apprendre à reproduire une action en observant un humain.

Problème, « l'IA peut halluciner », rappelle Ezio Malis, responsable de l'équipe Acentauri (Intelligence artificielle et algorithmes efficaces pour la robotique autonome - Inria). Il travaille donc sur des méthodes dites « hybrides », mélange de règles (un visage a forcément deux yeux, un nez…) et d'apprentissage automatique (la machine apprend à reconnaître un visage après avoir vu des milliers d'images) capables non seulement d'interpréter une image, mais aussi de limiter les extrapolations hasardeuses.

Quand les caméras interpréteront nos pensées

Dans le film « 2001, l'Odyssée de l'espace », de Stanley Kubrick, l'ordinateur HAL lit sur les lèvres des astronautes. La réalité a rattrapé la fiction : les robots seront bientôt capables de deviner nos stratégies de raisonnement. « Chez la souris, les micromouvements des oreilles, du museau ou des abajoues ne sont pas aléatoires mais stéréotypés : en les observant avec une caméra, on peut inférer ce que ces rongeurs pensent », révèle Fanny Cazettes, chercheuse à l'INT (Institut de Neurosciences de la Timone - Aix-Marseille Université/CNRS), qui avec des collègues français et portugais vient de publier un article scientifique en ce sens. La méthode est-elle applicable aux humains ? « Oui, pour deux raisons : on pourra entraîner l'algorithme de prédiction avec des données encore plus fiables en observant un être humain et en lui demandant à quoi il pense ; et le visage humain qui comporte 42 muscles est plus expressif que chez la souris qui n'en possède qu'une vingtaine », avertit Fanny Cazettes, qui souligne non seulement les bienfaits (détection d'enfants souffrants de troubles psychologique), mais aussi les dangers (manipulation de la part du robot…) d'une telle approche.

_____

Les chiffres clés

Une dizaine de milliards de dollars
La taille du marché mondial des robots humanoïdes d'ici à 2030.

130 à 80 images par seconde
La fréquence à laquelle travaillent les caméras qui équipent robots ou drones. Il faut que les microprocesseurs, derrière, puissent suivre.

3 mètres
La distance parcourue par une voiture autonome, roulant à 108 km/heure en 0,1 seconde, soit l'écart entre le moment où une caméra « voit » un événement et où l'ordinateur réagit.

1 millimètre sur 1 millimètre
La taille des caméras les plus petites actuellement disponibles sur le marché.

Jacques Henno

-> ENGLISH

What will the robots of tomorrow see?

Increasingly in contact with humans, autonomous robots need not only to interpret the world around them but also to orient themselves within it. Cameras can meet this dual challenge.

By Jacques Henno

Published on Dec 2, 2025 at 10:00 AM

The tiny robot measures barely 20 centimeters long by 15 centimeters wide, but it can follow a path at a speed of 1.5 meters per second after memorizing it in one go. Upon arrival, it has deviated by only a few centimeters from the planned route. Antcar has no GPS, costs only 150 euros, runs on a Raspberry Pi 4 nano-computer, a 360° camera in degraded mode that only captures large objects around it, and a neural network inspired by that of insects.

“By replicating the vision and brain of insects, Antcar only needs 20 kilobytes of information to travel 50 meters, the equivalent of an email without a photo,” enthuses Julien Serres, a member of the Bio-Inspired Systems team that developed Antcar, at the Etienne-Jules Marey Institute of Movement Sciences (Aix-Marseille University-CNRS). The advantage? “Our system could serve as a backup for an autonomous robot: if it loses its GPS signal indoors or in an ‘urban canyon,’ it would be able to retrace its steps to regain a connection or return to the depot,” adds Julien Serres.

“In robotics, the most promising market right now is that of autonomous robots: robot taxis, collaborative robots in factories, humanoid robots in the home…,” observes Fabien Moutarde, director of the Robotics Center at Mines Paris. “These machines will be in constant contact with human beings. They must therefore perfectly perceive this very unstable environment in order to adapt to it, all while consuming as little energy as possible.”

Cameras are inexpensive and highly versatile sensors. "Vision allows a robot to recognize the objects around it, identify them, and sometimes manipulate them," explains Franck Ruffier, a specialist in bio-inspired robotics at the Lab-STICC laboratory (ENSTA Brest) and co-developer of Antcar. "But vision with fewer pixels also allows it to locate itself in space, relative to contrasts created by large objects, such as buildings, or even with the polarization of light, the evolution of which, after calculation, 'indicates' true north."

The next step is to choose the most suitable camera model for the robot. Then, the transmitted data needs to be processed. All while keeping an eye on the ethical questions raised by robots now capable, thanks to their cameras, of guessing our reasoning processes (see box below).

In some cases, the robots remain blind and must be guided externally. “There are cameras measuring 1 millimeter by 1 millimeter that can be integrated into endoscopes with a diameter of 2 millimeters,” explains Brahim Tamadazte, a microrobotics specialist at ISIR (Institute of Intelligent Systems and Robotics - Sorbonne University-CNRS-Inserm). “But for robots used to repair blood vessels inside the eye or to select sperm during IVF, their movements must be tracked using an ‘exteroceptive’ sensor: an optical microscope or a scanning electron microscope for nanorobots.”

In other cases, the range of options is vast: radar, perspective cameras (like those found in smartphones), fisheye cameras with a 120° field of view, 360° omnidirectional cameras, and the latest addition, event cameras, which show

great promise in moving environments. “Event cameras are made up of independent pixels that send a signal as soon as they detect a change in brightness,” explains Davide Scaramuzza, head of the Robotics and Perception group at the University of Zurich. “Because they emit very little data, the processing of this information is very fast, and therefore the robot's responsiveness is very high.” This very low latency—on the order of 0.2 milliseconds—is very useful for autonomous vehicles, drones, robots in space, and quality control.

Once the camera is chosen, its images sometimes need to be straightened. "The wider the field of view—120°, 360°—the more the edges of the image are distorted," summarizes Guillaume Caron, co-director of the JRL (Joint Robotics Laboratory), a joint research unit of the CNRS (French National Centre for Scientific Research) and the AIST (National Institute of Advanced Industrial Science and Technology—a branch of the Japanese Ministry of Economy), in Tsukuba. "Therefore, we risk losing the information contained in these edges when the image is projected onto a plane for processing by a computer." This is a problem as old as maps, which are a flattened representation of a sphere. With Antoine André, a researcher at the JRL, Guillaume Caron has developed a calculation method that allows the computer to reconstruct the original sphere while maintaining acceptable processing times. "This will allow a service robot to recognize the room it is moving through," explains Antoine André.

Once a high-quality mathematical representation of the image, understandable by the robot's computer, is obtained, the robot must extract meaningful information from it. For example, in the case of a self-driving car, "on the right, there's a pedestrian who might cross." "This is the semanticization of images," explains Fabien Moutarde. "For the past three years, VLMs (Vision-Language Models) have emerged in artificial intelligence, capable of linking visual appearance to semantics." In assisted driving mode, a car could thus tell the driver, "If you turn right, watch out for the bicycle coming from behind." New models, called "VLAMs" (Vision-Language-Action Models), should even allow a robot to learn to reproduce an action by observing a human.

The problem is, "AI can hallucinate," notes Ezio Malis, head of the Acentauri team (Artificial Intelligence and Efficient Algorithms for Autonomous Robotics - Inria). He is therefore working on so-called "hybrid" methods, a mix of rules (a face necessarily has two eyes, a nose, etc.) and machine learning (the machine learns to recognize a face after seeing thousands of images) capable not only of interpreting an image, but also of limiting risky extrapolations.

When cameras interpret our thoughts

In Stanley Kubrick's film "2001: A Space Odyssey," the computer HAL reads the astronauts' lips. Reality has caught up with fiction: robots will soon be able to guess our reasoning strategies. "In mice, the micromovements of the ears, snout, or cheek pouches are not random but stereotyped: by observing them with a camera, we can infer what these rodents are thinking," reveals Fanny Cazettes, a researcher at the INT (Timone Institute of Neuroscience - Aix-Marseille University/CNRS), who, along with French and Portuguese colleagues, has just published a scientific article on this subject. Is the method applicable to humans? "Yes, for two reasons: we can train the prediction algorithm with even more reliable data by observing a human being and asking them what they are thinking; "and the human face, which has 42 muscles, is more expressive than that of the mouse, which has only about twenty," warns Fanny Cazettes, who emphasizes not only the benefits (detection of children suffering from psychological disorders), but also the dangers (manipulation by the robot…) of such an approach.

_____

Key figures

Around ten billion dollars
The size of the global humanoid robot market by 2030.

130 to 80 frames per second
The operating frequency of the cameras mounted on robots or drones. The microprocessors behind them must be able to keep up.

10 foot
The distance traveled by an autonomous car, traveling at 108 km/h in 0.1 seconds, is the difference between the moment a camera "sees" an event and the computer reacts.

1 millimeter by 1 millimeter
The size of the smallest cameras currently available on the market.

Jacques Henno