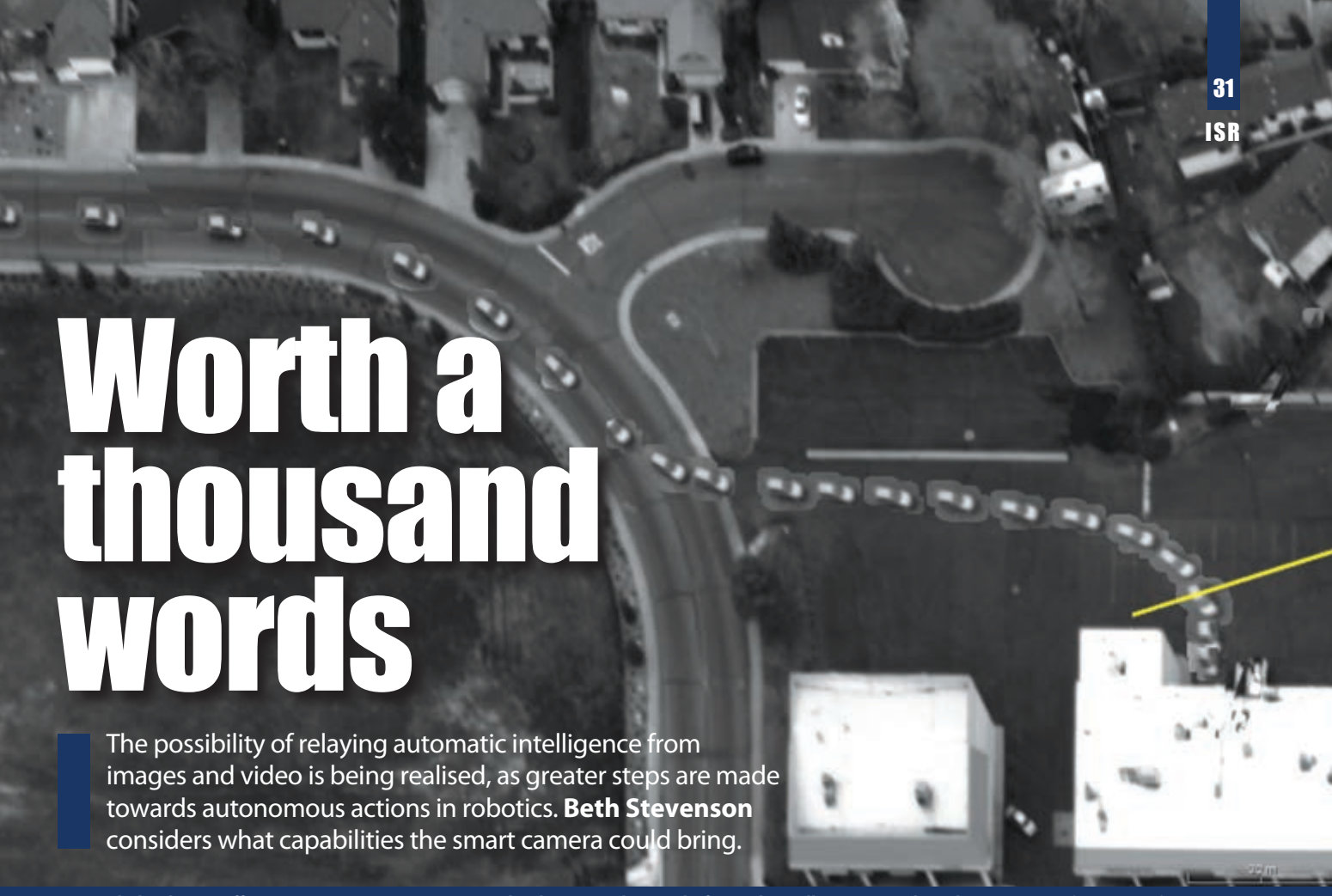
Worth a thousand words

The possibility of relaying automatic intelligence from images and video is being realised, as greater steps are made towards autonomous actions in robotics. **Beth Stevenson** considers what capabilities the smart camera could bring.



ITT Exelis has begun efforts to integrate smart camera technology on airborne platforms that will carry out onboard processing. *(Photo: ITT Exelis)*

New efforts within the computer vision industry are aiming to develop 'smart' camera technology, capable of observing and assessing subjects of interest and disseminating information back to the operator in language form.

The commercial market has experimented with elements of this 'intelligent computing' concept in recent years, with examples including photographic face recognition on Facebook to voice control on Xbox Kinects.

This artificial intelligence (AI) is available in everyday life and is now transcending into the military domain. 'This is based on the same general principles and mathematical techniques, but the particulars differ,' Jeffrey Siskind, associate professor of electrical and computer engineering at Purdue University, told *Digital Battlespace*.

NEXT LEVEL

Hot on the heels of commercial industry, defence companies and agencies are looking to take this intelligence to the next level by passing on even more responsibility to technology and robotics, aiming to produce unmanned ground and airborne platforms able to observe a scene and take on some of the analysis before it reaches the human that requires the information.

DARPA is pioneering one such effort, having decided a few years ago to push AI forward with the launch of a programme looking to develop camera technology that could report what it sees in plain English text, specifically from UGVs.

Originally a five-year project, the Mind's Eye programme was launched in 2010, under which 12 research teams were tasked with developing machine-based visual intelligence, alongside three companies developing integration concepts.

The 12 research teams contracted to develop the camera technology were: Carnegie Mellon University; Co57 Systems; Colorado State University; Jet Propulsion Laboratory/Caltech; Massachusetts Institute of Technology; Purdue University; SRI International; State University of New York; TNO (Netherlands); University of Arizona; University of California Berkeley; and the University of Southern California.

Meanwhile, the three companies exploring integration onto the UGVs were General Dynamics Robotic Systems, iRobot and Toyon Research.

DARPA claims that a 'truly "smart" camera' would be able to describe with words everything it sees and reason about what it could not, just like manned operators working from traditional observation posts (OPs).

LEARN AND INTERPRET

The agency claimed that the programme sought to develop the capability for visual intelligence by automating the ability to learn representations of actions between objects in a scene directly from visual inputs, and then interpret those learned representations.

'A key distinction between this research and the "state of the art" in machine vision is that the latter has made continual progress in recognising a wide range of objects and their properties – what might be thought of as nouns in the description of a scene,' explained a DARPA spokesman.

'The focus of Mind's Eye is to add the perceptual and cognitive underpinnings for recognising and reasoning about the verbs in those scenes, enabling a more complete narrative of action in the visual experience.' ➤

Data can become intelligence faster with image-to-language technology. *(Photos: Purdue University)*

Essentially, the agency is transforming commercial technology that is able to identify the 'what' to defence technology that is able to communicate the 'why'.

However, some 18 months after the programme began, it was effectively cancelled, although funding was stretched to continue for some 24 months more, leading to an eventual termination of the programme scheduled in December 2013, according to Siskind, a member of one of the contracted visual intelligence teams.

Nevertheless, three years' worth of research has been developed to date and the teams are still currently working on Mind's Eye as part of an overall research objective that each institute has in this area. The programme aims to support the 'robotic scout' on surveillance operations – a type of system mounted on a UGV that would be sent over enemy lines. Siskind explained: 'Normally, the scout role is today played by humans and is a very dangerous role.'

SMART SCOUTS

The manned OP is typically operated for long periods of time, with personnel having to not only conduct surveillance, but at the same time provide their own security. Again, Siskind described how such a 'smart' robotic scout would save on manpower, while increase the overall safety of those in OPs, as well as saving on bandwidth.

'The mission objective… was to observe activity and report the activity back, not as video, but as natural language text, the idea being that natural language text has a much lower bandwidth than video,' he noted.

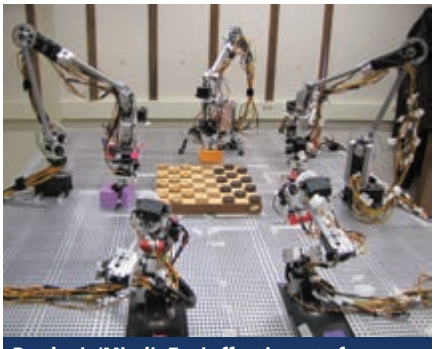
'If you had many of these out in the field and had limited bandwidth, you could have more information by text than video. If there was then a particular thing you needed to see as a result of the video, you could tune into specific videos to get video feed if you need it.'

As a result, the analysis and dissemination of data is reduced, and a smaller number of analysts need to monitor feeds.

Referring to the system Purdue is currently developing, Siskind noted: 'We have a real-time system that runs on a desktop or portable computer, and takes live camera inputs [and] watches people interact with objects, like briefcases, toy rifles and jerry cans.'

The system recognises people and their position in the FoV, as well as objects and their position and orientation, tracking them in real time between 2 and 4fps.

'It will track two people and three to four objects, and it recognises the activities that the people are performing with the objects,' he continued. 'It has a vocabulary of around a dozen verbs, things like walk, carry, pick up… and it will produce a real-time spoken description in audio of what's happening in the video.'



Purdue's 'Mind's Eye' effort is part of a broader intelligent robotics development programme at the university.

WESTERN INFLUENCE

Siskind said that the department also has a collection of ten full-length films, mainly 'Westerns', that the system can search through in response to search criteria. The genre has a preponderance of objects such as horses and people, making it ideal for the system to search through.

'We can take a potential query like "the person rode the horse leftward towards another horse" and search the entire video corpus to find short video clips that depict that sentence,' he explained. 'It has a vocabulary of around 20 words that you can string together into a coherent, semantically meaningful English sentence, and type that sentence in just like with a Google search, and you would get video hits that depict the sentence that you typed in.'

Before Mind's Eye comes to an end, Purdue is looking to make the system more robust, as well as decreasing the number of errors occurring, thus increasing the reliability of the object detectors, trackers and event recognisers.

Siskind added: 'We're going to try to, if possible, transfer this technology to the US military. There are a number of technology demonstrations that they're trying to schedule to show… military officers to see if there's any interest in this, and they're also interested in what they call "forensic video surveillance".'

This, he explained, refers to the proposed ability to search through large amounts of data in the wake of an incident to find targeted activity, triggered by recent events such as the Boston Marathon bombings.

FROM DAY TO DAY

Purdue's participation in the Mind's Eye programme is part of a long-term research effort into the potential benefits of human-robotic interaction in day-to-day activities.

'My interests lie in trying to understand and replicate human intelligence on the part of computers and particular robots,' added Siskind. 'My interests lie within the grounding of semantics and language in both visual perception and action.

'I want to tie computers and robots to a language, so that robots can describe what they see and perform actions that you tell them. Ideally, a robot that can see, do and talk can interact productively in a cooperative fashion with a human.'

Purdue is currently attempting to teach robots the game of draughts, as well as making them capable of interacting with humans, asking questions when something is not understood. This idea has been mastered with noughts and crosses, with chess being the next, longer-term objective. ➤

(Photo: Purdue University)

'The idea is to learn the rules by a combination of English instruction and demonstration – both vision and language and robotic interaction with people, and have the robot not only able to be a student, but once it learns, to also be a teacher,' said Siskind.

He suggested that military applications for this will likely be logistical, so that personnel can concentrate on other tasks. 'Robotic maintenance crews that fix vehicles and aeroplanes… that understand how the machines work and how to fix and repair them and figure out what's wrong,' he continued. 'That is in the real long-term future. You can imagine robots that collaborate with soldiers in the battlefield, that don't have to be told everything that needs to be done at every step along the way.

'We're looking at increasing the autonomy of things like autonomous ground vehicles, UAVs, autonomous underwater vehicles. There are all sorts of autonomous vehicles, but they aren't really autonomous today. Every one of these has a full-time, dedicated human controlling them by remote control, and we're working on trying to increase the level of autonomy.'

IN THE AIR

This intelligent computing concept is not suited to ground applications alone, it is also being integrated into airborne sensor suites.

'The concept of having ground-based sensors that are learning and taking video data or spectral data and turning it into information, words and detections – that is a concept that has been going on for a long time,' Bernard Brower, a chief scientist at ITT Exelis Geospatial Systems, told *DB*. 'We do something similar on the airborne spectral side.'

The company has developed an OP technology that it has been working on for two years, incorporating it into its Jagwire data management software. It is now looking to transfer this technology to an airborne system.

He explained that hyperspectral sensing data processing can be done on board an aircraft, which enables operators on the ground to only receive the information that they specify and often a representative image to go alongside it.

'In general, not everybody trusts the fact that we tell you this is the location, they want to see it for themselves,' explained Brower. 'This is common, but we don't have to do that because we've detected and told them where it is, but we often send a context image with it.'

'We're expecting to replace ten to 15 analysts per sensor with watchboxes and having the computers watch locations.'

The hyperspectral onboard processing enables the systems to collect data, specify the different materials being surveyed and identify and transfer the information down to the operator.

'We have operational systems of this type,' he continued. 'But in more of the video context we're focused on how to… decrease manpower through using computers.

'Instead of having ten people looking at ten locations, you put ten watchboxes and say "just tell me when something happens there". You can then either go over and look at it or do an action based on what it is told to do.'

PATTERN OF LIFE

This concept translates into the wide area persistence domain, benefitting the operator who would otherwise be flooded with information. Warfighters will refer to the generation of a 'pattern of life'.

'We are part of the Gorgon Stare team, and what we're working on now is having a camera… and the concept is that it is persistently watching a location,' noted Brower. 'And so from that it can learn the location and start building up history and information, and as they are tracking and identifying people and motions… a critical part of that is that you have to have the persistent information.

'You have to have it watching all of the time characterising information, and then when you identify something that is threatening you see that pattern again and can convert that into words and text to say what is happening.'

Brower distinguished between wide-area airborne and ground-based data collection of this type, as the latter is collecting limited volumes of data so a large amount of processing is not required.

'But as we start moving into the airborne environment, and doing the persistent surveillance type applications, that becomes a much more difficult task,' he said. 'Even if you are an HD video camera, you're taking 2MP of data, but on these airborne wide-area persistent surveillance cameras, you're taking hundreds of megapixels to gigapixels of data, and you're trying to process that and convert it into intelligence into the words.'

MALE INTEGRATION

ITT has been working on intelligent computing for airborne applications for some three months, and hopes to begin testing some time next year. The first integration will be on board a medium-altitude long-endurance (MALE)-type UAV, such as the MQ-9 Reaper, although systems for smaller UAVs are also in the pipeline.

'I truthfully think that at some level this [technology] is unlimited, but in the near term over the next couple of years we're expecting to replace ten to 15 analysts per sensor with watchboxes and having the computers and intelligence watch locations,' commented Brower.

'But ultimately, as computers increase the amount that you can do, the more intelligence you have… you can start looking at patterns of life, and applications and events that way.' DB