Vision and Robotics

Thomas M. Strat, Rama Chellappa, Vishal M. Patel

■ *Vision and robotics has the well-defined goal of meeting or exceeding human-level capabilities in perception, locomotion, and manipulation. Not surprisingly, that is perhaps easier said than done. Beginning in the 1970s, the Defense Advanced Research Projects Agency started the ambitious Imaging Understanding program that would continue for more than 20 years. The Imaging Understanding program began with fundamental research and slowly evolved into a host of more applied efforts with specific systems goals. Robotics programs followed a similar arc as the early research-oriented programs generated capabilities from which practical systems could be built. A culmination of the vision and robotics research was the Defense Advanced Research Projects Agency Grand Challenge, which turned the impossibility of a self-driving car into an imminent reality. This article tells the story of how some of the modern-day technologies we enjoy today trace their evolution from research sponsored by the Defense Advanced Research Projects Agency over the last 40 years.*

The goals of computer vision and robotics are to meet or exceed human-level capabilities in perception, locomotion, and manipulation. Human vision is so effortless and yet computer vision is so difficult. Any ordinary child can discriminate dogs from cats, fire trucks from police cars, and rocks from bushes, but even the most sophisticated object recognition systems cannot match the visual abilities of a young child. Robotic locomotion is similar in that it is easy for animals but challenging for machines. For example, a newborn deer can walk within hours of birth, but robotic locomotion within a forest is still an unsolved challenge. While it is clear that robotic manipulation systems can perform better in tailored environments, paradoxically it is often the case that the easier a physical task appears to be, the harder it is to automate; the same can be said for vision systems. Vision and robotics are artificial intelligence (AI) complete in the sense that solving these pursuits requires general-purpose AI. Vision and robotics are not peripheral fields of study that are ancillary to AI, they are inseparable embodiments of AI.

In 1975, the Defense Advanced Research Projects Agency (DARPA) started what would become the longest running program in its history: Image Understanding (IU). Military applications of computer vision were easy to imagine, and one could readily appreciate the value and potential impact. The original goals of the IU program were clear — explicitly

Vision and Robotics

In September 1977, I joined the hand-eye group, led by Tom Binford, at the Stanford Artificial Intelligence Lab. I had just arrived in the United States as a scruffy 22-year-old from Australia. I was immediately supported in my new research by the ARPA (as it was then) Image Understanding Program. That was at a time when it was possible to personally know every researcher in computer vision in the world, and I soon met most of those in the US through this program.

The key was a six-monthly workshop, rotating among locations next door to the research universities working on computer vision, where we submitted camera-ready copies of papers, which were published and available just a few weeks later at the current ARPA Image Understanding Workshop. And then we met and discussed.

My first workshop was just a few months later in Cambridge, Massachusetts in May 1978, where I presented a coauthored paper. I remember in particular that there I met a very young professor, Takeo Kanade, of Carnegie Mellon University (CMU). National origin made no difference to having either Takeo or me working on important defense problems in the US. I had another paper at the next workshop, in November 1978, and that one was held in Pittsburgh, Pennsylvania, hosted by CMU. And so-on, every six months. I was now really in the vision community, sharing ideas in an extremely timely manner, and making rapid progress, despite what now we would view as having only laughably slow and small computers.

My work was about using high-level knowledge in computer vision, as we know happens in human vision. My ACRONYM model-based vision system was an AI reasoner that used geometric models of the objects it was looking for to drive the low-level vision processes and move up and down the representation stack making inferences in both directions. The ARPA IU program gave me the freedom to pursue radical ideas at the intersection of AI and vision in the interest of new capabilities.

And that friendship with Takeo Kanade? It turned into the two of us cofounding the International Journal of Computer Vision in 1987 — it has a 30+ year history of being one of the leading journals in computer vision.

– Rodney Brooks

envisioned applications included automated cartography, satellite image interpretation, cruise missile guidance, and automatic target recognition (ATR). It is unlikely that program manager David Carlstrom and DARPA management at the time realized the difficulty and complexity of the undertaking that they were embarking on.

In the early days of the IU Program, DARPA realized that they needed first to advance the basic scientific foundations of IU before practical systems could be designed and built. The first steps toward IU research gradually sorted themselves into four schools of thought that can be traced through the community to this day.

Early Vision

Best exemplified by the pioneering works of Azriel Rosenfeld, David Marr, and Tomaso Poggio, this line of research attempted to mimic the processing of the human visual system. This work focused on the lowest levels of visual processing — the first few stages within and after the retina, and gave rise to such important concepts as edge finding, interest points, textures optical flow, stereopsis, 2.5D depth, and the primal sketch.

Physics-Based Vision

In contrast to early vision that took a biologic approach to vision, this school of thought approached the problem from the perspective of a physicist. The idea was to model vision as a system of mathematical equations for the refraction of light by lenses and the reflection of light by surfaces with various material properties. The solution to a computer vision problem could then be found by inverting the mathematics of image formation. Berthold Horn and Thomas Binford are two of the earliest and most preeminent practitioners of this approach.

Statistical Approaches

Some early research focused on modeling the neuron as a computational unit. Minsky and Papert's work on the multilayer Perceptron is widely regarded as the seminal work in this area, which eventually gave rise to the modern-day neural net. Later, methods based on Markov random fields and their variants were proposed for texture and image representation and segmentation. Techniques like simulated annealing were used for image restoration and stereopsis, and Bayesian methods were developed for object recognition.

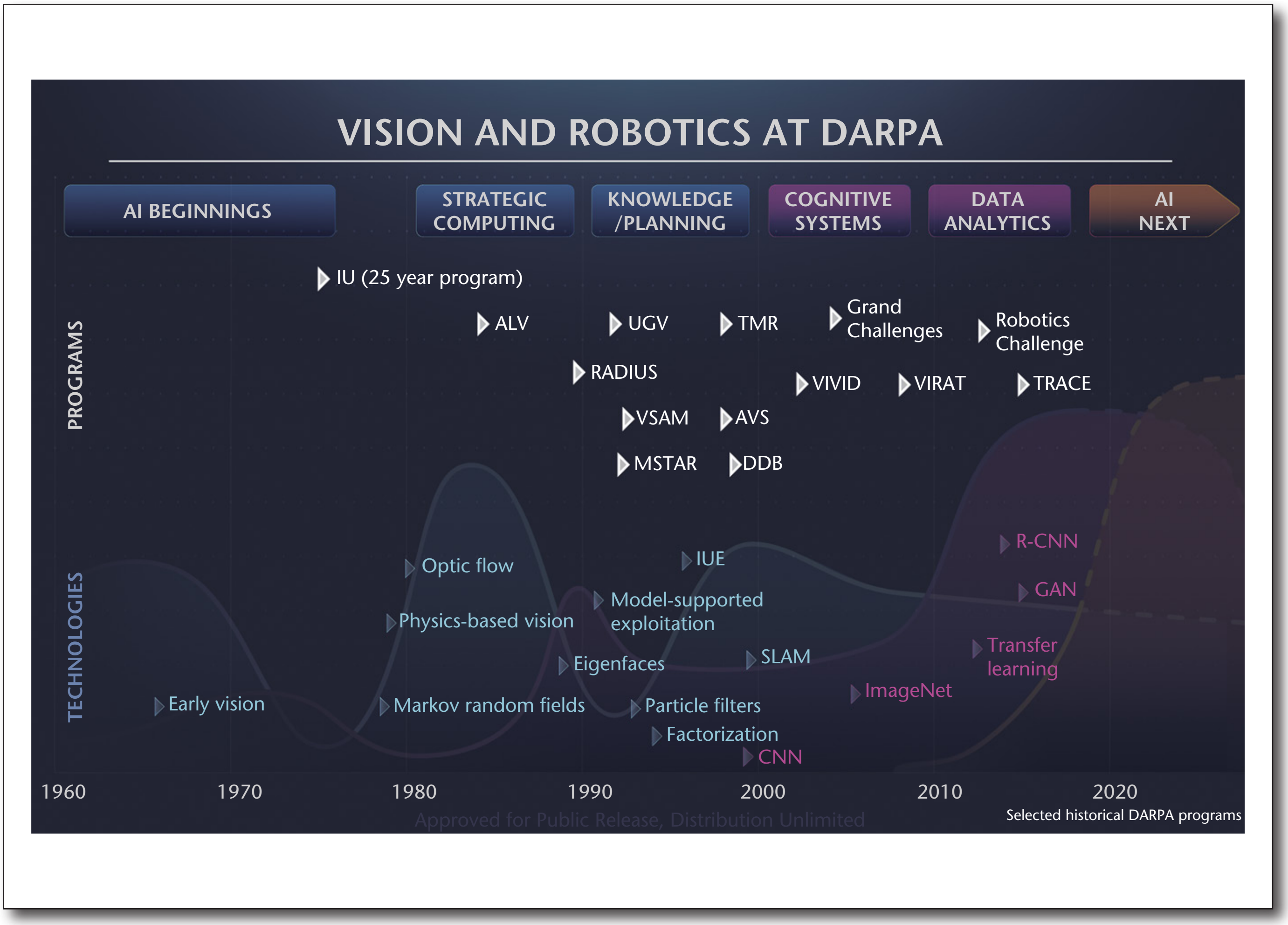

Figure 1. Vision and Robotics at DARPA.

Selected historical vision and robotics DARPA programs and technologies. Figure courtesy of DARPA.

Engineering

Many computer vision researchers took an engineering approach devising a wide array of creative solutions in quest of building working systems. Using appropriate combinations of sensors, signal processing, and other mathematical techniques supported by experimentation and systems, IU researchers came up with increasingly well-founded components and innovative solutions to many vision tasks. Some of the most successful early visionaries in this camp include Aggarwal Bajcsy, Brooks, Davis, Fischler, Grimson, Hanson, Huang, Kanade, Mundy, Nayar, Nevatia and Riseman.

In the 1970s and 1980s, achieving the full potential of any of these approaches was impossible due to the lack of available computational power. As a result, experimentation was limited to a relatively small number of images, and technology development was slow. Statistical approaches in particular, which require a very large quantity of exemplars, were slow to find favor.

The IU program continued to make advances for more than 15 years until it gave way in 1992 to a host of more-applied programs focused on system-level results for specific goals. Robotics programs followed a similar arc as the early research-oriented programs generated capabilities from which practical applications followed (see figure 1). The goals of vision and robotics have remained largely unchanged, and have been achieved in limited contexts. This article tells the story of how some of the modern-day technologies we enjoy today originated from research sponsored by DARPA over the last 40 years.

The Golden Years of the DARPA IU Program

From 1975 to 2000, the DARPA IU program that began as a basic research program slowly morphed into an effort supporting specific tasks such as ATR, satellite and aerial image exploitation, visual

Institution	Key Personnel	Areas of Investigation
University of Southern California	Andrews, Nevatia, Price	Aerial image interpretation, object recognition, digital image restoration
SRI International	Barrow, Bolles, Garvey, Tenenbaum	Aerial image interpretation for cartography and intelligence using prior knowledge of the scene
Stanford University	Binford	Stereo photointerpretation using spatial features and spatial relations
University of Rochester	Feldman	Query-driven, top-down approach to aerial image interpretation for ship finding
Purdue University	Huang, Fu	Scene analysis using syntactic approaches, segmentation using texture and gray-level, random field approach to pattern classification, Fourier descriptors of shape
Honeywell, Inc.	Larson	Multiresolution ATR in airborne forward-looking IR imagery
CMU	Reddy	3D-scene understanding, knowledge representation and search, image feature analysis and segmentation, change detection, knowledge acquisition
UMD	Rosenfeld	ATR in forward-looking IR imagery
MIT	Winston, Horn, Marr	Representation as the key issue, reflectance maps for image synthesis and registration, primal sketch for a comprehensive theory of recognition

Table 1. Original Participants of the IU Program.

surveillance, and so forth. During this period of what may be called the golden years, significant strides were made in advancing new theories and computational methods for a variety of computer vision problems. We review the progress and most prominent results made during these years by considering three periods: 1975–1980, 1981–1990, and 1991–2000. The original participants of the IU program are listed in table 1. Due to space constraints, we could not include many other equally important results and efforts.

The first five years of the program (1975–1980) saw breakthroughs in early vision and pioneering research in physics-based vision. The seminal work of Marr and Hildreth in deriving a theory of edge detection laid the foundation for one of the basic problems in computer vision — edge detection. Their pioneering work brought together concepts from signal processing and neuroscience, and showed that the zero-crossings of the Laplacian operator on a Gaussian-smoothed image could serve as edges and generate what is known as the primal sketch. In another pioneering effort, Berthold Horn at the Massachusetts Institute of Technology (MIT) developed the concept of the bidirectional reflectance function and derived an expression that related the 3D depth of a scene or an object to its 2D image through a nonlinear mapping known as the reflectance function. This made it possible to recover 3D structure from a single image by solving a nonlinear partial differential equation using the characteristics strips approach. This period also witnessed the emergence of the seminal computational theory of human stereo vision by Marr and Poggio using the zero-crossing contours extracted from the left and right images of a stereo pair. Grimson extended the Marr-Poggio theory to yield the Marr-Poggio-Grimson algorithm for stereopsis with results on aerial and satellite images.

Throughout the following decade, IU continued to make strides in early vision and physics-based vision. Regularization was the dominant theoretical paradigm that emerged in these years. The key realization was that computer vision tasks such as shape recovery from a single image, a stereo pair, or a video sequence are ill-posed problems; therefore, Poggio and others developed the regularization method that led to a systematic approach for developing computer vision algorithms. In this approach, one would identify a function to optimize with appropriate constraints such as surface or motion smoothness, and derive the Euler-Lagrange equations that could then be solved using discrete optimization methods. Horn used this approach for shape recovery from a single image and computation of optic flow. The regularization approach could be equivalently interpreted as minimizing an energy function or maximizing the posterior probability density function of entities that were being extracted.

This approach could accommodate both deterministic and probabilistic methods to computer vision problems. As an example, the approach was used by Geman and Geman (1984) in their seminal work considering the problem of posterior density maximization when underlying image data are represented using Markov random field models (or equivalently, Gibbs distributions). The DARPA IU program did not directly support this work, but it still had a great impact on many computer vision algorithms such as the maximum posterior marginal methods designed by Marroquin and Poggio. In 1986, Canny published a seminal work on optimal step edge detection in Gaussian noise using a combined metric that is a function of probability of detection, localization error, and false detections. The resulting optimal detector was approximated using the directional derivative of a Gaussian smoothed image. Canny later introduced clever tricks, including hysteresis and noise estimation, to produce more-effective edge-detection results.

The 1980s also saw the emergence of image sequence analysis for estimating the motion and structure of a 3D rigid object. Longuet-Higgins and Prazdny developed a theory relating the structure and motion of a 3D object to optical flow. Horn and his colleagues developed a class of methods known as direct methods for motion and structure estimation from optical flow. Adiv followed these works and demonstrated the recovery of the structure and motion of moving objects by segmentation. While these continuous approaches gained much traction, in parallel, methods based on discrete features such as points, lines, and contours were developed by Huang, Aggarwal, Bolles, Broida, Kanade, Price, Baker, Szeliski, and others. 3D object recognition approaches using 2D images and 3D-range data were also developed, including Grimson's work on interpretation trees, Hummel's work on geometric hashing, and Brooks's work on a scene-labeling system called ACRONYM. Another major development was the notion of purposive vision or active perception proposed by Garvey and elaborated by Bajcsy and Aloimonos (Aloimonos, Weiss, and Bandyopadhyay 1988).

During the 1990s, IU engineering and physics-based approaches prevailed as computational power and the availability of video data created many challenges and opportunities. Due to the availability of high-end computers, the problem of tracking one or more objects from stationary and moving cameras became viable; object tracking methods based on probabilistic data association, Kalman filters, Lucas-Kanade registration algorithm, and particle filters were developed. Particle filters introduced by Isard and Blake in 1996 became appealing due to their ability to handle nonlinear motion and non-Gaussian noise models. Object tracking in video acquired by moving cameras required the stabilization of video sequences before independently moving objects could be detected and tracked. A plethora of real-time and near real-time methods were developed for the problem of video stabilization and mosaic construction. It is worth noting that commercial versions of this algorithm began to be incorporated into hand-held video cameras to eliminate undesired camera motions and jitter.

Progress in detecting, tracking, and classifying moving humans and vehicles was made possible by the DARPA programs Unmanned Ground Vehicles (UGVs) and Visual Surveillance and Monitoring (VSAM). The UGV program demonstrated capabilities such as landmark-based navigation, and reconnaissance, surveillance, and target acquisition. The W4 system developed at the University of Maryland (UMD) and the VSAM testbed developed at CMU are good examples of progress in this area. This period also saw the development of additional work on structure from motion using monocular and binocular sequences. The first data-driven method for recovering the structure and motion of a moving object from point correspondences established over a long sequence under an orthography assumption, known as the factorization theorem, was developed by Tomasi and Kanade. This period also witnessed breakthroughs such as normalized cuts for the problem of texture segmentation by Shi and Malik, anisotropic diffusion methods for edge detection by Perona and Malik, and subspace-based methods for face recognition by Belhumeur, Chellappa, Kriegman, Pentland, and Turk.

By the end of the century, the IU program had matured from one exploring fundamental concepts in computer vision, into a disciplined investigation into many practical applications of IU (figure 2). The field had grown from small research projects at a dozen prominent universities to the point that all major universities in the country had faculty specializing in IU, and computer vision was taught at the undergraduate level of all science and engineering universities. By the year 2000, the research-oriented IU program gave way to a succession of dozens of individual programs focused on specific applications of IU. DARPA was eager to capitalize on the new capabilities afforded by its prior investments in the IU program and the concomitant advancements in compute power and proliferation of low-cost, high-resolution cameras.

Applications of IU and Robotics

In the mid-1980s, DARPA began to explore applications of IU such as autonomous land vehicles, aerial image analysis, and satellite image exploitation in addition to the research-oriented IU program.

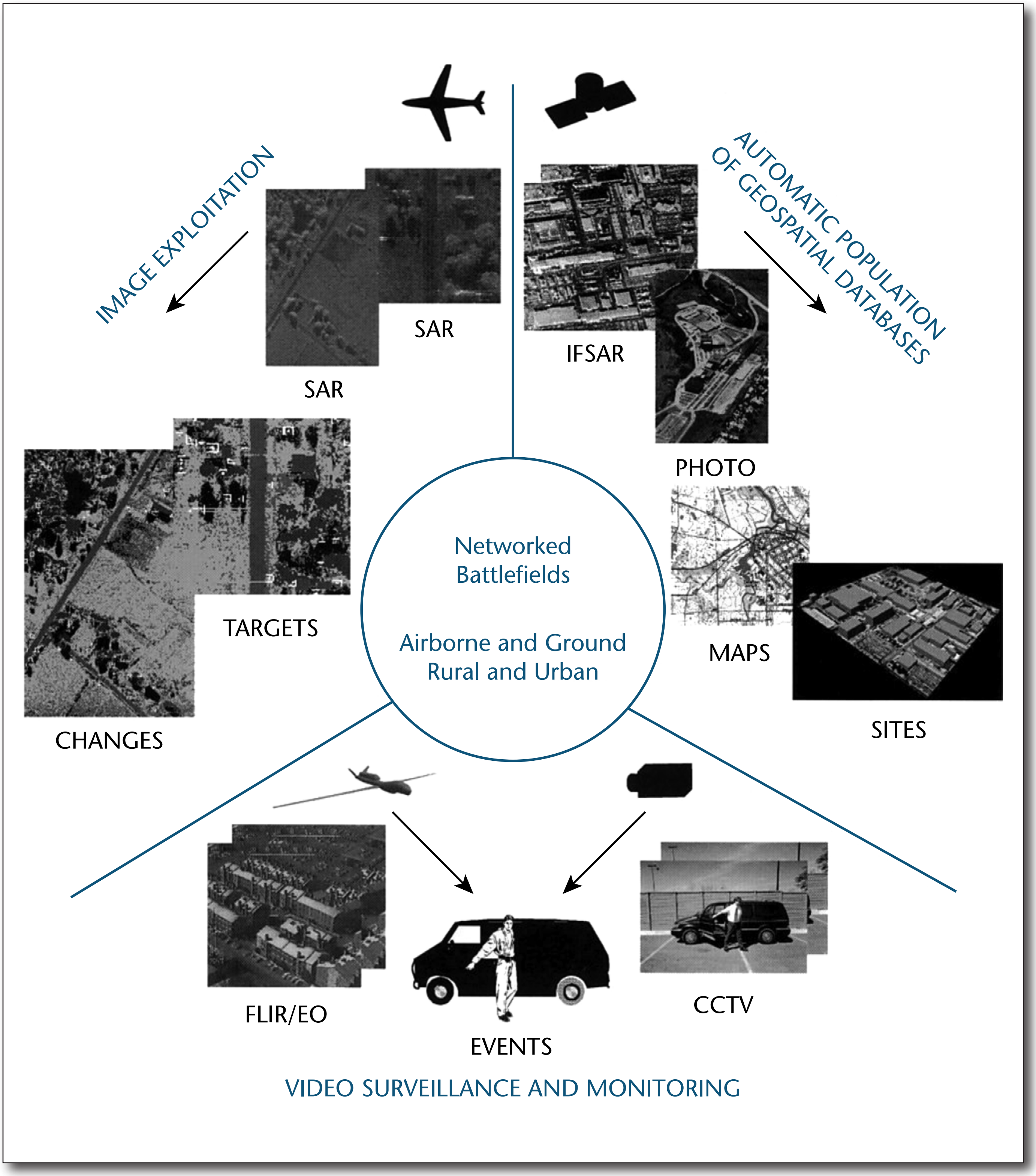
IMAGE EXPLOITATION
AUTOMATIC POPULATION OF GEOSPATIAL DATABASES
SAR
SAR
IFSAR
PHOTO
TARGETS
CHANGES
MAPS
SITES
Networked Battlefields
Airborne and Ground Rural and Urban
FLIR/EO
EVENTS
CCTV
VIDEO SURVEILLANCE AND MONITORING

Figure 2. Applications of IU.

Typical applications of IU that motivated research in the 1990s and spawned military and commercial applications in the years that followed. Adapted from the cover of the *Image Understanding Workshop Proceedings, May 1997*.

Since the early 1990s, increasing emphasis was given to application-driven programs such as UGVs, ATR, and image exploitation (RADIUS). The ATR program encouraged computer vision researchers to look beyond traditional visible spectrum imaging and explore target detection and recognition algorithms using infrared (IR), light detection and ranging (LIDAR), and synthetic aperture radar (SAR) images. An interesting development with ATR using SAR images was the creation of the Moving and Stationary ATR program that incorporated many computer vision concepts such as focus of attention, feature extraction, indexing, prediction of features using sensor physics, and model-based recognition. Given the nature of the problem, Moving and Stationary ATR involved participation from several universities including Ohio State University, UMD, MIT, and companies such as the Environmental Research Institute of Michigan (Ann Arbor, Michigan), ADS Inc., and AlphaTech. IR and LIDAR-based IU algorithms also found applications in the UGV program.

The RADIUS program looked at the problem of model-supported exploitation (figure 3). The approach was to build 3D models of a site using previously collected images, and monitor the site for changes due to new construction or vehicle-related activities. To enable the integration of algorithms that the researchers developed at various institutions (e.g., Lockheed, CMU, SRI International, UMD, USC, and the University of Massachusetts), a software environment known as the RADIUS Common Development Environment was provided. This was one of the first such examples of software integration of computer vision algorithms with the end-goal of providing software systems to the end users.

In 1997, DARPA initiated three application-oriented IU programs: Battlefield Awareness, Automated Population of Geospatial Databases, and Visual Surveillance and Monitoring (VSAM). These efforts covered terrain modeling and exploitation using video, SAR imagery, LIDAR, and Digital Elevation Maps. VSAM was the most prominent of these three programs, due largely to the new availability of video images acquired by unmanned aerial vehicles (UAVs). The VSAM program involved many university research groups and a few companies. In the late 1990s, the Airborne Video Surveillance program with participation from Harris Corporation, Sarnoff Corporation, SRI, and UMD looked at the problem of activity recognition in aerial video sequences. The highlight of the Airborne Video Surveillance program was a live demonstration of real-time video exploitation over the Fort A.P. Hill military reservation.

Programs such as RADIUS and Moving and Stationary ATR gave rise to the Dynamic Database program, which considered the problem of image and signal exploitation using a multitude of data such as electro-optical/IR, SAR, hyperspectral images, and signal intelligence. The trend of analyzing a large collection of aerial videos continued for many more years through programs like VIdeo-based Verification and Identification (or VIVID), Video Image Retrieval and Analysis Tool, and Persistent Stare Exploitation & Analysis System. These programs further developed algorithms and systems for detection and tracking of objects, object verification, and object recognition as well as activity recognition in UAV video and wide-area motion imagery. These problems are still being studied today by the US Department of Defense Maven program, which is making use of research results developed earlier under DARPA funding.

The Autonomous Land Vehicle (ALV) program (figure 4) in its early stages involved many university research groups and companies such as Martin-Marietta for designing the sensors suites and integrating algorithms for object detection, obstacle avoidance, and navigation. The ALV program morphed into the UGV program that built vehicles with sensors such as stereo and IR cameras and LIDAR, enabled by algorithm development at CMU. A high point of the UGV program was the CMU demonstration of a car that drove unassisted for almost 95% of the distance from Pittsburgh to San Diego. The UGV program laid the foundation for the DARPA Grand Challenge that saw many leading groups in the country competing for bragging rights. It is not an exaggeration to say that what we are witnessing today in the area of autonomous cars has its origins in many DARPA programs undertaken since the mid-1980s.

Due to the successes in VSAM, other applications of ground-based video sequences were explored. Of note is human identification at a distance that developed gait and face recognition methods at distances of 10–20 meters. These programs were not continued by DARPA, but other agencies further extended them. For example, the Multi-Disciplinary University Research Initiative on Remote Biometrics in the Maritime Domain by the Office of Naval Research and the Biometrics Exploitation Science & Technology and JANUS programs by the Intelligence Advanced Research Projects Agency (IARPA) organization have looked at developing robust face verification and identification algorithms. The algorithms for action recognition from video sequences originally investigated by the DARPA IU program have inspired ground-based activity recognition in DARPA's Mind's Eye program and served as catalysts for the Advanced Research and Development Activity Video Analysis and Content Extraction (or ARDA-VACE) program as well as IARPA programs including Aladdin and Deep Intermodal Video Analytics.

During the course of the IU Program, the roles played by data availability, hardware improvements, and software developments cannot be ignored. In the early days of IU research, data to support the evaluation of algorithms and systems was not readily available; in particular, ground truth and meta data were hard to obtain. Sensors improved in terms of form factor and performance, making them more accessible and resulting in an increase in data available to researchers. The calibrated imaging laboratory at CMU is a good example of early efforts in collecting

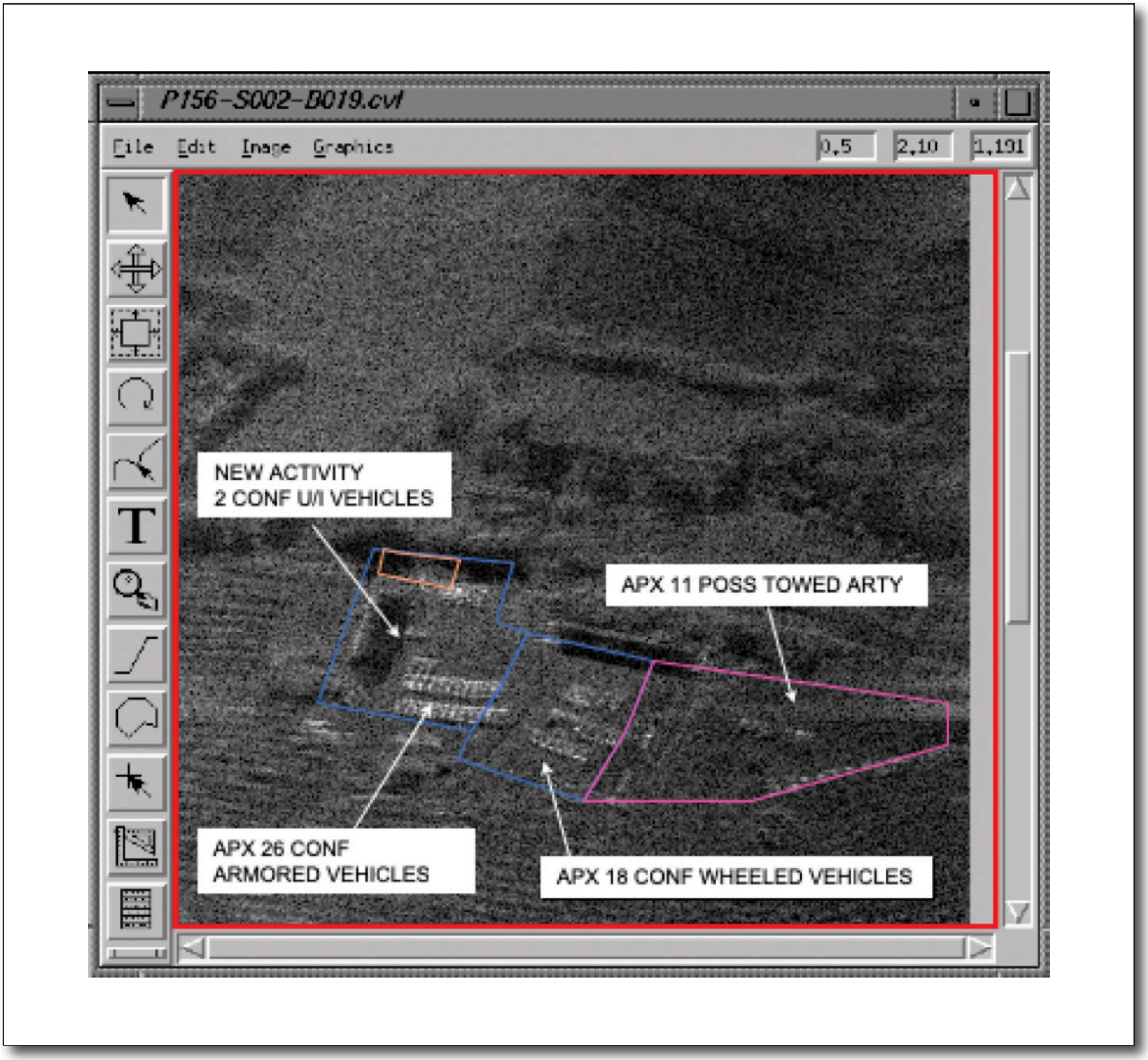

Figure 3. Model-Based Exploitation.

Model-based exploitation as developed in the RADIUS Program used site models to guide IU algorithms for reliable detection and recognition of vehicles in SAR images such as this (DARPATech, 2004).

data for testing many computer vision algorithms. Another good example is the release of high-resolution SAR images in the early 1990s by the MIT Lincoln Laboratory. ALV and UGV program platforms, along with UAV platforms, also enabled the collection of videos from moving sensors used in DARPA research programs and were provided to researchers for use on programs sponsored by other government agencies. The wide availability of multimodal data enabled the quantitative evaluation of algorithms leading to the development of IU systems that can be transitioned from laboratories to the real world. More recently, the availability of large amounts of annotated data has enabled the training and, correspondingly, encouraged the development of deep-learning algorithms.

As mentioned, the RADIUS Common Development Environment, which was developed by Quam at SRI, was one of the earliest software environments that enabled the integration of IU algorithms into a software system. The DARPA IU Environment and

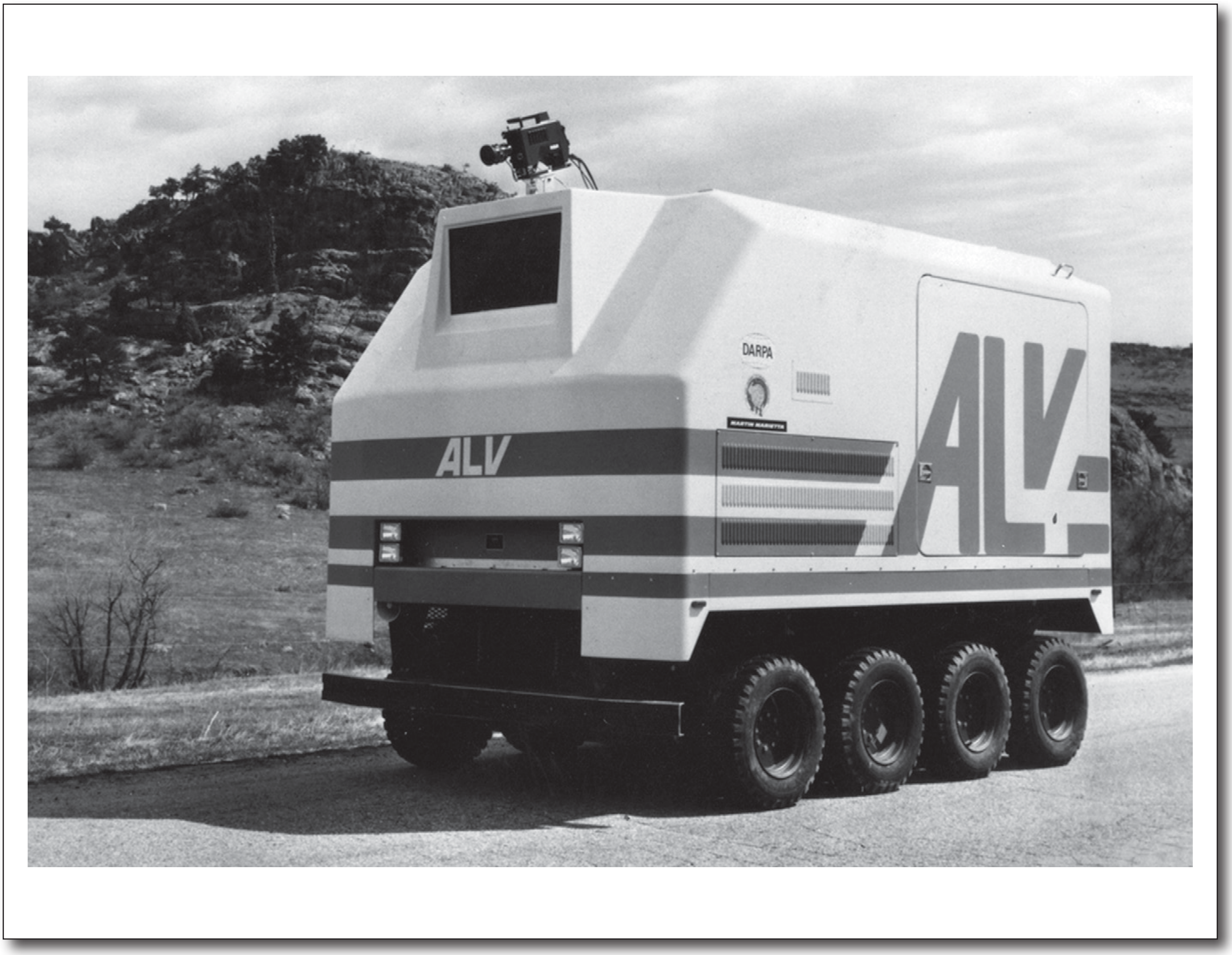

Figure 4. ALV Vehicle.

The DARPA ALV Program was the world's first outdoor vehicle to operate autonomously. The large size of the vehicle was necessitated by the large quantity of computing machinery required to process the sensor data from the cameras and LIDAR. (DARPA, 2018)

its extensions followed soon after. The IU Environment was a community-wide effort led by Mundy and Boult that incorporated the vast majority of IU techniques. The developed software environments were open-source; the availability of the software facilitated collaborations across research groups and accelerated the development of research prototypes.

As applications of computer vision became technologically feasible, the need to compute results fast enough to provide real-time solutions for interactive and video processing systems grew. In the 1990s, the need to have faster implementations and real-time video processing algorithms led to the development of specialized computing hardware such as the Connection, Data cube, and Hybercube machines. This trend continued and gave rise to graphics-processing-unit implementations of IU algorithms, and eventually to the development of deep-learning algorithms. The necessity of high-performance hardware and reliable software is especially obvious in real-time applications such as autonomous vehicles.

Perhaps surprisingly, when artificial neural networks (ANNs) made a comeback in the 1980s, largely due to the seminal work of Hopfield, the IU community was not yet totally on board with neural networks. The approach of feeding data into a 3-layer network and receiving object labels as outputs was not appealing as computer vision researchers were more interested in modeling 3D geometry, illumination, and articulation. The 3-layer ANN black-box approach was not seen as explainable or capable of achieving human-level recognition rates. Although the DARPA IU program did not encourage approaches based on ANNs, many computer vision researchers developed algorithms based on neural networks. As examples, we point out Kanade and Poggio's work

DARPA IU and My Research

I came to the United States and joined the CMU – DARPA IU team in 1976. To a young person who had finished his PhD and then worked briefly as a junior faculty member in a Japanese university, where at that time computer vision research was done only on a small scale, everything was different, exciting, and overwhelming. That I could write my programs at any time, even from my apartment, was a paradise to me. The images given to deal with and the tasks to aim for were real. At the DARPA IU Workshop, where all the contractors get together, I found my heart pounding when mingling and talking with legendary researchers such as David Marr, Tom Binford, Marty Fischler, and Azriel Rosenfeld, whom I had known only by name from textbooks.

The mid-1970s to the mid-1980s was a dynamic period for computer vision. Emphasis on use of knowledge was revealing a new aspect of visual understanding, apart from that of pure signal processing. Marr's paradigm was leading the way to relate computer vision with psychology and neuroscience. Many shape-from methods were invented, with eye-opening results. Many challenging applications, such as aerial photo interpretation, industrial, robotic, and medical, were percolating. The DARPA IU Program was the major force in driving forward this emerging field of computer vision.

Our CMU computer vision group has performed in diverse IU projects for more than 35 years. Fond memory was abundant. The theory of the Origami World and shape-recovery from a single image was my IU debut work. The Lucas-Kanade optical flow was originally developed for the image registration task in aerial photo interpretation. The CMU Navlab driverless-car project, which we started in 1985 as part of the DARPA Autonomous Land Vehicle Project, eventually led to our 1995 No Hand Across America demonstration. Carlo Tomasi's factorization method and Shree Nayar's reflection model were among our responses to call for new shape-from methods. For the VSAM project in early 2000, CMU as the system contractor ran big demos yearly involving on- and off-campus networks of surveillance cameras and even a flying airplane for tracking people and cars. Our idea of using a large number of cameras for capturing and modeling scenes, although very common today, was initially regarded strange or even lunatic, but DARPA funded us to develop first a five-camera video-rate stereo machine and then a 51-camera 3D dome, which eventually transformed into the current 480-camera Panoptic Studio; on its way, the technique was used for a movie Matrix-like replay system, EyeVision, in the broadcast of Superbowl XXXV.

Throughout these projects I learned one most important thing. Computer vision must work in the real world, and for that, we must make theories, algorithms, computing, and sensors work together as a system. One may think that it is obvious, but in the early days, computer vision did not work. I was extremely lucky that I learned this lesson by participating in the DARPA IU Program. Researchers in the DARPA IU community were my best comrades and rivals. Successive program managers were tough and supportive. Tasks from military users were challenging and real. All in all, it was the best research environment for me. My whole career would not have existed without DARPA IU.

– Takeo Kanade

on face detection and Dean Pomerleau's demonstration of almost autonomous driving from Pittsburgh to San Diego. In the mid to late 1980s, a program focused on ATR using ANNs was initiated by DARPA. The reemergence of deeper ANNs as deep-learning networks has contributed to the creation of more recent programs such as Explainable AI (or XAI).

Robotics Research

In the robotics arena, beginning in the 1960s, DARPA funded investigations into the fundamental concepts of robots. Efforts included perception, representation, planning, and locomotion. The DARPA-sponsored Shakey robot, developed at SRI, is widely regarded as the first attempt to integrate all of the above-listed aspects into a complete mobile robotic system both situated in the world and responsive to human commands. Since that ambitious beginning, DARPA has pursued many robotic programs in an attempt to accelerate the introduction of robots onto the battlefield. Flying robots proliferated before ground robots, as the airborne environment is simpler than the environment on the ground. UAVs face a far lower obstacle density than UGVs do, and today there are many dozens of different UAVs used by the US military, while only a few UGVs have been adopted operationally.

In 1972, DARPA began developing US military UAVs with the creation of the Remotely Piloted Aerial Observation/Designation System, which led to the production of the Aquila Unmanned Aircraft System, or UAS, in 1983. Since that time, DARPA pioneered the widely used Predator, Global Hawk, and many other UAVs that have entered military service, from micro air vehicles (or MAVs) such as the

6-inch Wasp to the Unmanned Combat Air Vehicle (or UCAV).

In the ground vehicle domain, DARPA's early work on SRI's Shakey and Marc Raibert's one-legged hopping robots paved the way to many other research projects. In 1983, DARPA launched the ALV program (see figure 4) led by Martin-Marietta, and supported by many universities; it was the first attempt to build an unmanned vehicle that could operate outdoors. Progress on ALV led to the creation of follow-on programs such as the Unmanned Ground Vehicle and Perception for Off-road Robotics. The Tactical Mobile Robots Program explored a wide variety of approaches to locomotion, navigation, planning, and control, and gave rise to notable systems including Big Dog, and PackBot. While UGVs have yet to achieve widespread use as combat vehicles, they have proliferated for dealing with improvised explosive devices in Iraq and Afghanistan.

In the manipulation domain, the Autonomous Robotic Manipulation (or ARM) program developed software for intelligent control of manipulators and low-cost rugged and dexterous multifingered hands and arms. In the 1990s, robotic motion planning was an important unsolved problem for planning trajectories under uncertainty while preventing collisions. Jean-Claude Latombe at Stanford University developed many variants of the probabilistic roadmap planner for path planning in high-dimensional configuration space. Oussama Khatib made fundamental advances in dexterous dynamic coordination, virtual linkages to model internal forces in cooperative manipulation, dynamic task decoupling, and human-robot compliant interaction.

Manipulation, locomotion, planning, and control all came together in 2011 when DARPA conducted the DARPA Robotics Challenge (DRC; figure 5). The DRC was a competition motivated by the 2011 nuclear disaster at Fukushima Daiichi in Japan; this competition developed and exhibited human-supervised ground robots capable of executing complex tasks in dangerous, degraded environments using tools and equipment commonly available in human-engineered spaces.

Unmanned naval vessels have also been addressed. Not many details are public, but applications span unattended buoys that sense the ocean environment while harvesting energy from the waves, autonomous surface vessels called wave gliders that travel large distances also by harvesting energy from the waves, and unmanned underwater vessels that probe the ocean at depths unsafe for humans. To date, the pinnacle of unmanned naval vessels is the DARPA Anti-Submarine Warfare Continuous Trail Unmanned Vessel (or ACTUV), which was commissioned in 2016.

Commercial Impact of DARPA Investment in IU and Robotics

DARPA investments in IU and robotics were clearly aimed at accelerating the availability of key technologies for use in military systems by stimulating basic research and by producing early prototypes of previously infeasible components. These investments paid off in many ways that cannot be listed in an open publication, and will continue to do so in the foreseeable future. These same advancements created and enabled by DARPA programs in IU and robotics have given rise to an array of novel commercial products as well, and in some cases have enabled entire industries that are making major contributions to the US economy.

Commercial applications of computer vision began to flourish beginning in the mid-1990s with the simultaneous emergence of three developments: IU algorithms with high reliability, low-cost computer processors and graphics processing units with sufficient processing power, and the proliferation of low-cost high-resolution digital cameras. The most influential commercial applications of IU are ones that we hardly notice. The proliferation of images on the internet created the demand for tools that could rapidly retrieve images based on content, and thus the Google image search capability was born. The emergence of smart phones with cameras gave rise to intrinsic tools such as panoramic mosaics and high-dynamic-range images that had been pioneered in DARPA IU workshops a decade or two earlier. In fact, the entire field of computational photography is now a burgeoning industry with roots that trace to innovators in the IU Program. The rapid progress in face recognition has found its way into hundreds of commercial products and dozens of companies have been created to pursue applications from photo albums to smart-phone authentication to forensics. The introduction of video-understanding algorithms transformed video surveillance from a passive activity to one that actively seeks particular actions and behaviors.

The commercial adoption of robotics has been even more striking. When the Roomba was introduced by iRobot as a household vacuum cleaner in 1999, consumer robots were still the subject of science fiction. At the time, iRobot was a DARPA-sponsored company that created the PackBot, a robot designed for elimination of explosives on the battlefield. Today more than a dozen companies market household robots, and the industry is valued at more than $2B in the US alone.

Finally, it is hard to overestimate the size of the impact that the DARPA Grand Challenge has had on the automotive industry. From a humble but ambitious beginning in the Mojave Desert in 2004, the Grand Challenge has redefined the prospects of self-driving cars. From "that will never happen in my lifetime" to self-driving components that are being offered today by nearly every automobile manufacturer (e.g., intelligent cruise control, parking assist, lane following, blind-spot warning, collision avoidance), DARPA has helped create a $100 billion industry.

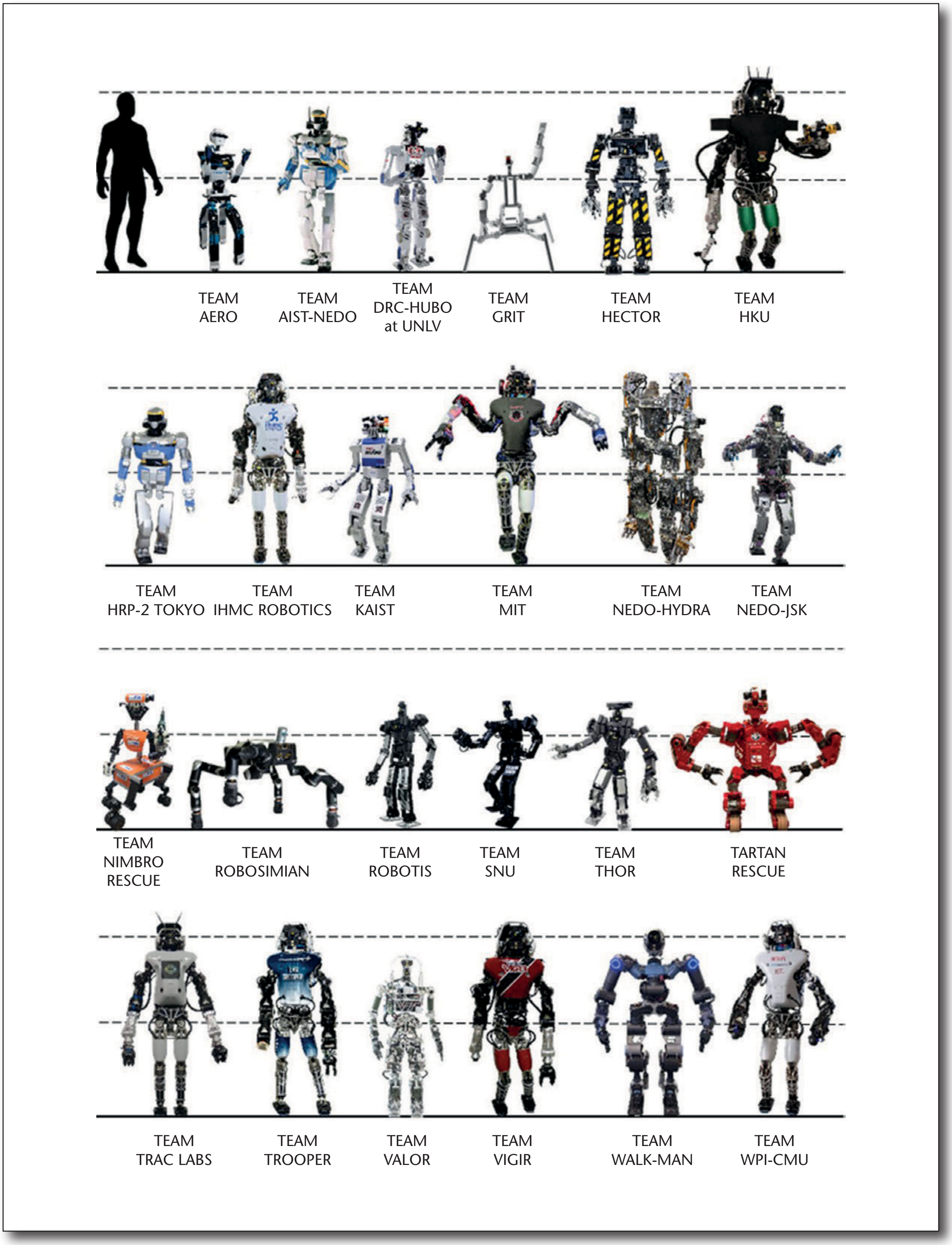

Figure 5. The 2011 DRC.

The 2011 DRC saw a diverse range of robots. The 24 robots pictured above participated in the DRC Finals. Dashed lines show heights of 1m and 2m, respectively. ©2016 Wiley Periodicals, Inc. Reproduced with permission from E. Krotkov, D. Hackett, L. Jackel, M. Perschbacher, J. Pippine, J. Strauss, G. Pratt, and C. Orlowski, 2017, "The DARPA Robotics Challenge," *Journal of Field Robotics*, 34(2)2: 229–40.

The DARPA Grand Challenge: A Case Study

The DARPA Grand Challenge represents a premier example of DARPA accelerating progress on a challenging research area of critical importance to the US Department of Defense. In 2002, DARPA had multiple programs underway with goals to advance the state of the art in UGVs. DARPA had pursued these technologies beginning with the ALV program as part of the Strategic Computing Initiative begun in the late 1980s. The research encompassed all aspects of autonomous driving including sensors, computation, obstacle detection, road-following, subsystem control, and higher-level planning and reasoning. These programs sponsored research and development projects at many defense contractors and research universities that featured a steady procession of increasingly sophisticated demonstrations. Much progress was made, but the goal of UGVs to support military operations remained elusive, and frustration was growing with unmet expectations.

In parallel with these technical developments, DARPA was innovating on another front. While DARPA is widely recognized for its many technical innovations, it also has an extensive, although less well-known, track record of innovation in the acquisition of research and development. With exceptional foresight, DARPA requested, and was granted by Congress, the authority to issue cash prizes in reward for technological achievement. This obscure provision was a historical first for any agency in the federal government. In this case, the limit was set at $1 million, but the specific technical area was left to DARPA's discretion. Early in 2002, DARPA Director Tony Tether issued a call for ideas and eventually selected a race for autonomous ground vehicles as the subject for what is now known as the first DARPA Grand Challenge for Autonomous Ground Vehicles. Dr. Tether had two primary challenge criteria: The topic should be accessible by many — anyone with a car and a laptop computer could attempt to assemble an entry; and the winner should be clear to all — the vehicle that completes the course in the shortest amount of time wins the $1 million prize.

The concept was simple. DARPA specified the route that must be followed — a desert course over rough terrain from the outskirts of Los Angeles to the fringes of Las Vegas — but kept the 140-mile course secret until 2 hours before the start of the event. Colonel Jose Negron was the Director of the Grand Challenge held in April 2004.

The response was tremendous. More than 100 robot teams eventually formed and completed the application process to participate in the Grand Challenge. Interest came from all corners. Many top universities, with CMU the early favorite, and defense contractors, were represented. There were even quite a few individual entries comprised of grease monkeys, mechanics, computer geeks, veterans of the TV show Junk Yard Wars, and other colorful characters who contributed novel approaches to sensing, control, and suspension systems, not to mention marketing. There was even a high school team — Palos Verdes High School — that convinced Acura to donate a brand-new SUV. The team had 63 kids, only two of whom had a driver's license!

The vehicles were as varied as the entrants themselves. From SUVs and pickup trucks, to dune buggies and custom-tracked vehicles, pit row was a veritable hodgepodge of every vehicle imaginable. With no limits set by DARPA, the vehicles ranged in size from the 32,000-pound behemoth military transport entered by the OshKosh Truck Corporation to the autonomous motorcycle entered by Anthony Levandowski of the University of California at Berkeley.

The results of the Grand Challenge of 2004 are well ensconced in history. After an extensive selection process, 15 vehicles started the race on the morning of March 13, 2004, but only four traveled more than 4 miles, with Red Whittaker's Sandstorm from CMU taking top honors with a distance of 7.4 miles. While that fell far short of DARPA's lofty goal of 142 miles, it was still farther than any autonomous vehicle had traveled previously.

While the press tended to put a negative spin on the results as a disappointment, DARPA pointed out that the Wright Brothers' first flight at Kitty Hawk lasted less than 30 seconds, and like the Wright Brothers, DARPA was not deterred. DARPA announced a second Grand Challenge for Autonomous Ground Vehicles to be held 18 months later on October 4, 2005. The results of that second Grand Challenge were much different, with five vehicles, including that 16-ton truck from OshKosh, completing the entire course. Stanford University (see figure 6) led the pack with the fastest time while CMU vehicles took second and third place. The age of self-driving vehicles had begun.

What changed? Why the sudden progress after decades of frustration? In a word, everything, including the way DARPA had gone about procuring technology:

DARPA hosted the race, but it didn't contract with any of the teams that entered. Instead, each team was left to its own resources. In academia, many universities had departments or clubs that were pursuing mobile robotics; DARPA simply gave them a common cause on which to focus their resources. The academic spirit to solve the challenge better than their peers took over and drove the universities through a friendly competition.

In defining the Grand Challenge, DARPA specified the task, but did not specify the solution. Innovators were free to approach the challenge in practically any way they chose.

The military-focused programs all insisted upon retaining a human-in-the-loop. The Grand Challenge was the first time that a major program focused on fully autonomous operations. It turns out that it

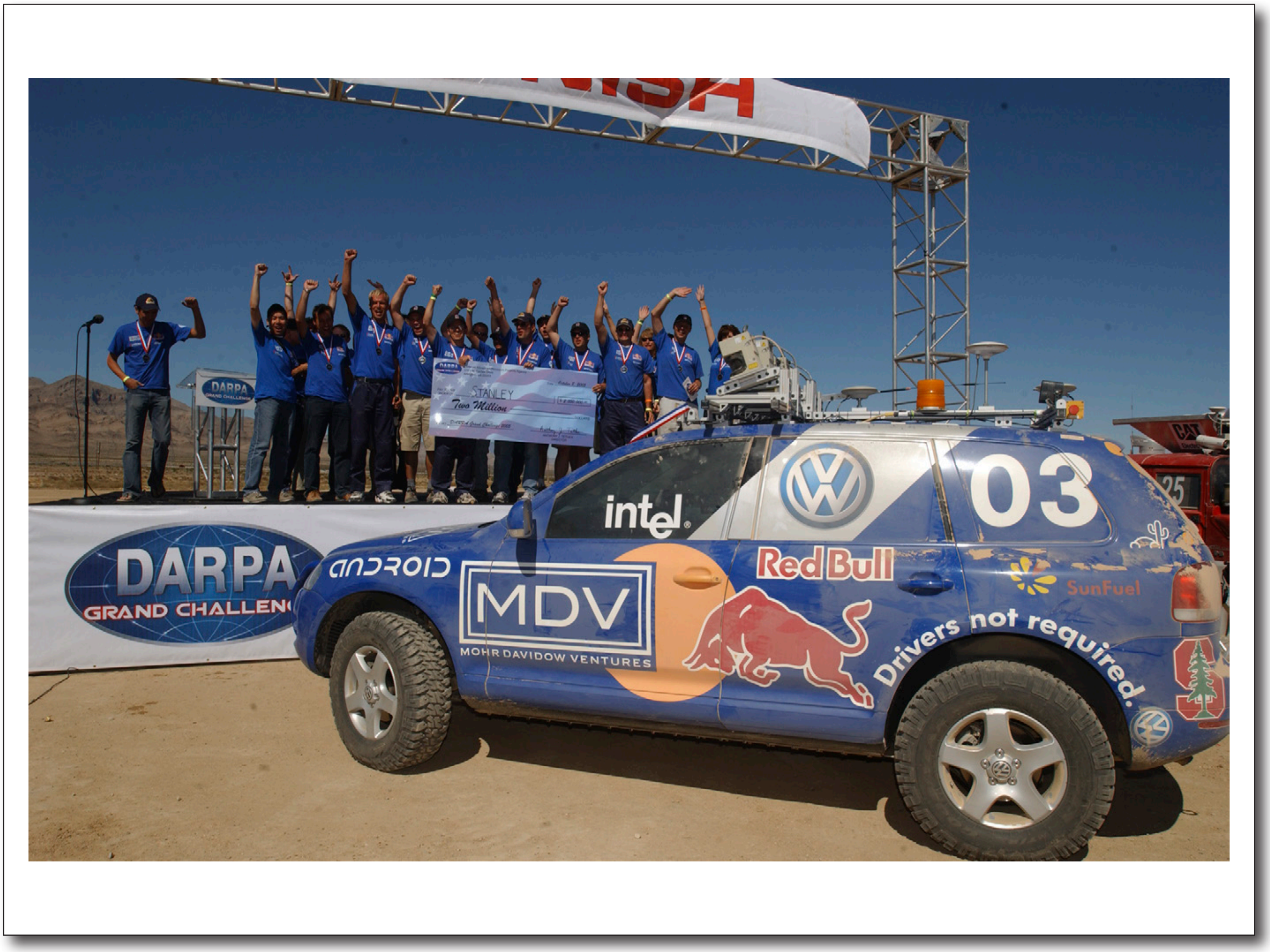

Figure 6. Stanley.

The winner of DARPA's 2005 Grand Challenge — Stanley — and the team that built the autonomous vehicle. DARPA's Assured Autonomy program leverages some of the research pioneered more than a decade ago in DARPA's 2005 Grand Challenge. (DARPA, 2018)

is harder to interface with human operators than it is to provide complete automation.

By 2004, the prior research sponsored by DARPA and others had progressed to the point where solving the Grand Challenge was possible. What remained was the impetus to integrate the components. The Grand Challenge became that impetus.

Deep-learning technology was just maturing to the point where it could be put to practical use for road detection and following. The Stanford team did this exceedingly well, and rode it all the way to the finish line.

Interestingly, the major automobile manufacturers were notably absent from the Grand Challenge teams. Many car manufacturers' executives were present as observers and some had donated vehicles to the teams in the race, but mostly they were unwilling to risk the reputations of their corporations on a technology as preposterous as a self-driving car. After his victory in 2005, Sebastian Thrun, the leader of the Stanford Team in the 2005 Grand Challenge, went on to create the self-driving car program at Google. Many of the other leading contenders at the Grand Challenge went on to prominent positions at self-driving car programs at other major corporations, and the automobile industry as a whole began to invest large sums of money in autonomous vehicle technologies. Today these capabilities are available as components of cars and trucks from most major manufacturers, and fully self-driving cars appear to be just around the corner.

Autonomous vehicle capabilities are proving to be a tremendous safety innovation as evidenced by the rapid endorsement of the auto insurance industry. Many military vehicles are equipped with self-driving vehicle capabilities as well, both for safety and for the enhanced military capabilities it affords. In the case of self-driving vehicle technologies, DARPA

Challenge Name	The Task	Why It Is Hard	Why It Is Important
Describe The Scene	Given an image (or short video clip), provide a narrative description of the scene, its contents, and what is happening in the scene	This would build on prior work in object recognition and activity recognition and face recognition, but would require putting these pieces together to reason about what is happening in the scene. It would presumably require elements of common-sense reasoning, 3D modeling, recognition of specific people, places, and events, and combine it with the ability to generate natural language descriptions	The visual world is a rich source of information for humans, but our mechanisms for providing information about that world in an accessible form is extremely limited. A solution to this general perception problem is essential for robots to operate in the world and interact in natural ways. Computer systems ingest information from pictures
Recognize Novel Objects	Train an object recognition system to learn a new class of objects from a single exemplar	Given that humans learn about objects from very few examples, robust methods for object/face/action detection and recognition from very few samples are needed. Another open challenge that existing deep-learning methods cannot handle is recognition or detection of new objects that were not present in the training data. Novel methods for zero-shot detection and recognition of objects are needed	While deep-learning methods have produced impressive performance gains for many computer vision problems such as object/face detection and recognition, they need large amounts of annotated training data to train. That makes system development costly, as annotated training data are often expensive to obtain. In other cases, large numbers of training data simply are not available, such as ATR in search of a newly introduced military vehicle
Integrate Autonomous Robots	Design mobile, manipulating, autonomous agents to interact with people in day-to-day lives	Today's robots don't interact smoothly with people. Robots can't go where people go; robots can't manipulate objects and materials as easily as people; robots are frustrating to talk to; robots have no common sense	To realize the full potential of robotic assistants, they need to assimilate into our everyday lives. Examples of robots we want but don't have: robotic maids; robotic gardeners; robotic chefs; robotic shoppers
Vision-Enabled Virtual Reality/ Augmented Reality Systems	Design the next generation of Virtual Reality/ Augmented Reality systems that integrate outputs of computer vision algorithms	Real-time generation of computer vision outputs is not there yet	Virtual Reality/Augmented Reality systems see the world, but do not know much about it

Table 2. Representative Problems.

has clearly achieved what DARPA was created to do — to accelerate the development of defense-related technologies while maintaining US technological superiority.

The impact of the Grand Challenge was felt far beyond the autonomous vehicle technology itself. Universities used the Grand Challenge as a design exercise for robotics classes and laboratories, regardless of their participation in the Challenge. And that high school team that entered the race? Their teacher/advisor started a robotics class that fall and had 350 high school students sign up. It became known as the DARPA class and became the most popular class on campus. That robotics class continues to be offered today at Palos Verdes High School.

Open Research

Tremendous progress has been made in advancing the art and practice of vision and robotics over the last decade, but classes of problems remain for which no solution is in sight. As mentioned at the outset, perception is an AI-complete problem, meaning that it is unlikely that general-purpose computer vision or robotics systems will be created anytime soon. In table 2 we present some representative problems, the solution of which would represent fundamental advancement in the foundations of vision and robotics research. These problems have been selected specifically because they do not submit readily to the machinery of deep-learning neural nets or other known techniques. The solutions to these problems would represent fundamental advancement in the foundations of vision and robotics research. Solutions to these fundamental problems will no doubt lead to new applications of great importance to both military and commercial enterprises.

Further research into problems like these is necessary to advance the field and move ever closer to general-purpose vision and robotics systems. DARPA can lead the way as it has done for the last 40 years with investigations into both fundamental research as well as problem-oriented applications. DARPA's continued focus on perception and robotics will accelerate the innovations and advances in these important areas of AI. These innovations hold enormous importance to maintaining technological superiority as well as providing robust economic stimulus in the form of new companies, new products, and new capabilities to lead the nation.

While great advances have been made at the level of fundamental science as well as practical applications, the goals of computer vision and robotics remain largely unchanged from the early days — to match or exceed human-level capabilities in perception, locomotion, and manipulation. This goal has already been achieved in some narrow contexts, but general-purpose solutions that exhibit the full-range of performance of human visual systems and human arms, legs, hands, and feet will require additional investigation, funding and devotion — the type of commitment to which DARPA has proven itself to be uniquely suited.

Acknowledgments

We would like to acknowledge the tremendous accomplishments of the DARPA vision and robotics research community. While we have mentioned a few of their achievements in these pages, we have barely scratched the surface — their contributions will live on in the form of commercial products, defense systems, and foundational research ideas for generations to come. We would also like to thank Larry Davis, Eric Krotkov, Bill Mark, Kristen Fuller, Lois Hollan, and Zachary Lapin, who generously reviewed earlier drafts of this article.

The Reference list provides a sampling of the many thousands of papers, books, reports and systems that were produced by the DARPA research community.

References

Ali, K.; Langley, P.; Maloof, M.; Sage, S.; and Binford, T. 1998. Improving rooftop detection with interactive visual learning. In *Proceedings of the Image Understanding Workshop*, 479–92. San Francisco: Morgan Kaufmann.

Aloimonos, J.; Weiss, I.; and Bandyopadhyay, A. 1988. Active vision. *International Journal of Computer Vision* 1(4): 333–56. doi.org/10.1007/BF00133571.

Ballard, D. H., and Brown, C. M. 1982. Computer Vision. Englewood Cliffs, NJ: Prentice-Hall.

Binford, T. 2000. *Context and Quasi-Invariants in Automatic Target Recognition (ATR) with Synthetic Aperture Radar (SAR) Imagery: Final Report for Period 03 June 1997–June 03, 2000*. Technical Report AFRL-SN-WP-TR-2001-1101. Wright-Patterson Air Force Base, OH: Air Force Research Laboratory, Sensors Directorate.

Brainard, D. H., and Wandell, B. A. 1992. In *Analysis of Retinex Theory of Color Vision*. Healey, G. E.; Shafer, S. A.; and Wolff, L. B., editors. Boston, MA: Jones and Bartlett.

Canny, J. 1986. A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis Machine Intelligence* 8(6): 679–98. doi.org/10.1109/TPAMI.1986.4767851.

Defense Advanced Research Projects Agency 60 Years: 1958-2018, Faircount Media Group, Sep. 2018, URL: https://www.darpa.mil/attachments/DARAPA60_publication-no-ads.pdf

Defense Advanced Research Projects Agency DARPATech 2004 Symposium, Anaheim, CA, March 2004.

Firschein, O., and Strat, T. 1997a. *RADIUS: IU for Imagery Intelligence*. San Francisco: Morgan Kaufmann.

Firschein, O., and Strat, T. M. 1997b. *Reconnaissance, Surveillance and Target Acquisition for Unmanned Ground Vehicle: Providing Surveillance Eyes for an Autonomous Vehicle*. San Francisco: Morgan Kaufmann.

Fischler, M. A., and Firschein, O. 1987a. *Intelligence: The Eye, the Brain, and the Computer*. Reading, MA: Addison-Wesley Longman Publishing.

Fischler, M. A., and Firschein, O. 1987b. *Readings in Computer Vision: Issues, Problems, Principles and Paradigms*. San Francisco: Morgan Kaufmann.

Geman, S., and Geman, D. 1984. Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 6, 99.721-741, Nov. 1984.

Grimson, W. E. L. 1981a. Computer Implementation of a Theory of Human Stereo Vision. Philosophical Transactions of the Royal Society of London. *Series B, Biological Sciences* 292: 217–53.

Grimson, W. E. L. 1981b. *From Images to Surfaces: A Computational Study of the Human Visual System*. Cambridge, MA: The MIT Press.

Hanson, A. H., and Riseman, E. M. 1978. *Computer Vision Systems*. Boston, MA: Academic Press.

Horn, B. K. P. 1986. *Robot Vision*. Cambridge, MA: The MIT Press.

Horn, B. K. P., and Brooks, M. J. 1986. The Variational Approach to Shape from Shading. *Computer Vision Graphics and Image Processing* 33(2): 174–208. doi.org/10.1016/0734-189X(86)90114-3

Horn, B. K. P., and Schunck, B. G. 1981. Determining Optical Flow. *AI Journal* 17: 185–203.

Huang, T. S. (Editor). 1981. *Image Sequence Analysis*. Berlin, Germany: Springer doi.org/10.1007/978-3-642-87037-8

Kanade, T. 1987. *Three-Dimensional Machine Vision*. Amsterdam, the Netherlands: Kluwer doi.org/10.1007/978-1-4613-1981-8.

Krotkov, E.; Hackett, D.; Jackel, L.; Perschbacher, M.; Pippine, J.; Strauss, J; Pratt, G.; and Orlowski, C. 2017. The DARPA Robotics Challenge Finals: Results and Perspectives. *Journal of Field Robotics*. 34(2): 229–40. doi.org/10.1002/rob.21683.

Lucas, B. D., and Kanade, T. 1981. An Iterative Image Registration Technique With An Application To Stereo Vision. In *Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI'81),* Vol. 2, 674–79. San Francisco: Morgan Kaufmann.

Marr, D. 1982. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco: W. H. Freeman and Company.

Marr, D., and Hildreth, E. 1997. Theory of Edge Detection. *Proceedings of the Royal Society of London. Series B, Biological Sciences* 207(1167): 187–217. doi.org/10.1098/rspb.1980.0020.

Marr, D., and Poggio, T. 1997. A Computational Theory of Human Stereo Vision. *Proceedings of the Royal Society of London. Series B, Biological Sciences* 204(1156): 301–28. doi.org/10.1098/rspb.1979.0029.

Nevatia, R. 1982. *Machine Perception*. Englewood Cliffs, NJ: Prentice-Hall.

Rosenfeld, A. 1969. Picture Processing by Computer. *ACM Computing Surveys* 1(3): 147–76. 10.1145/356551.356554.

Rosenfeld, A., and Kak, A. C. 1982. *Digital Picture Processing,* Vol. 2, 2nd Ed. Boston, MA: Academic Press.

Strat, T., and Hollan, L. 2004. *VIVID: Automated Video Processing for Unmanned Aircraft*. Arlington, VA: DARPA.

Tomasi, C., and Kanade, T. 1992. Shape and Motion from Image Streams under Orthography: A Factorization Method. *International Journal of Computer Vision* 9(2): 137–54. doi.org/10.1007/BF00129684.

Ullman, S. 1979. *The Interpretation of Visual Motion*. Cambridge, MA: The MIT Press doi.org/10.7551/mitpress/3877.001.0001.

Thomas M. Strat was program manager at DARPA for 8 years between 1995 and 2006. He managed the IU Program, created a half-dozen other programs involving vision, robotics, and autonomy, and served as the deputy director of the first two Grand Challenges for Autonomous Ground Vehicles. Strat received the BS and MS in computer science from the MIT AI Laboratory and received his PhD in computer science from Stanford University. Currently, he is the Chief Executive Officer of DZYNE Technologies, a company that creates intelligent, autonomous aircraft.

Rama Chellappa is a Distinguished University Professor and a Minta Martin Professor of Engineering in the Department of Electrical and Computer engineering at UMD and the UMD Institute Advanced Computer Studies. His current research interests are computer vision, pattern recognition, and machine intelligence. He has received numerous research, teaching, service, and mentoring awards from the University of Southern California, UMD, IBM, the Institute of Electrical and Electronics Engineers Computer Society, the Institute of Electrical and Electronics Engineers Signal Processing Society, the Institute of Electrical and Electronics Engineers Biometrics Council, and the International Association of Pattern Recognition. He is a Fellow of the Institute of Electrical and Electronics Engineers, the International Association for Pattern Recognition, the Optical Society of America, the American Association for the Advancement of Science, the Association for Computing Machinery, the Association for the Advancement of Artificial Intelligence, and holds six patents.

Vishal M. Patel is an assistant professor in the Department of Electrical and Computer Engineering at Johns Hopkins University. Prior to joining Johns Hopkins, he was an A. Walter Tyson Assistant Professor in the Department of Electrical and Computer Engineering at Rutgers University and a member of the research faculty at the UMD Institute for Advanced Computer Studies. His current research interests include signal processing, computer vision, and pattern recognition with applications in biometrics and imaging.