Methods of AI for Multimodal Sensing and Action for Complex Situations

Erik Blasch, Robert Cruise, Alexander Aved, Uttam Majumder, Todd Rovito

■ *Artificial intelligence (AI) seeks to emulate human reasoning, but is still far from achieving such results for actionable sensing in complex situations. Instead of emulating human situation understanding, machines can amplify intelligence by accessing large amounts of data, filtering unimportant information, computing relevant context, and prioritizing results (for example, answers to human queries) to provide human–machine shared context. Intelligence support can come from many contextual sources that augment data reasoning through physical, environmental, and social knowledge. We propose a decisions-to-data multimodal sensor and action through contextual agents (human or machine) that seek, combine, and make sense of relevant data. Decisions-to-data combines AI computational capabilities with human reasoning to manage data collections, perform data fusion, and assess complex situations (that is, context reasoning). Five areas of AI developments for context-based AI that cover decisions-to-data include: (1) situation modeling (data at rest), (2) measurement control (data in motion), (3) statistical algorithms (data in collect), (4) software computing (data in transit), and (5) human–machine AI (data in use). A decisions-to-data example is presented of a command-guided swarm requiring contextual data analysis, systems-level design, and user interaction for effective and efficient multimodal sensing and action.*

Artificial intelligence (AI) has sought for the use of machines to solve tasks that humans are not capable of doing (such as large data analytics). The field of AI is exploding based on competitive engineering results, vast amounts of data, access to fast computation, and the promise of autonomy. A prominent example is illustrated by the significant efforts in autonomous cars that expanded from grand challenges (Seetharaman et al. 2006; Urmson et al. 2009). Current interest includes swarms of autonomous coordinated unmanned aerial vehicles (UAVs; Shishika and Paley 2017; Cruise et al. 2018). To enable such robotic systems requires multimodal sensing and action through autonomy. Four types of autonomy (Hintze 2016) include traditional rule-based AI approaches to self-awareness AI (table 1). Self-aware autonomous vehicles interact with humans (Amershi et al. 2014), build conceptual knowledge (Bredeweg et al. 2013), and use context (Adomavicius et al. 2011). A prominent example by Scerri et al. (2015) develops a context-aware situation analysis device by integrating human semantics (for example, social networks), physical sensors (for example, global position system), and models (for example, weather) utilizing mobile displays, ontologies, and multimodal fusion.

  ISSN 0738-4602

Type of AI	Focus	Objective
Type I	Reactive machines	Identify patterns from rules for immediate action
Type II	Limited memory	Estimate response using signal processing
Type III	Theory of mind	Form representations about the world and other agents
Type IV	Self-awareness	Understand self-conscious to interact with prediction

Table 1. Types of AI.

Multimodal sensing and action methods can leverage the recent trends in machine learning (ML) with an emphasis on deep learning (DL) methods. DL operates with large amounts of data to train a statistical model that represents the context from the data collected. A statistical model is one type of contextual analysis from which dynamic information supports cooperative model adaptation for situation awareness, coordinates with first-principles mathematical models representing physical phenomena for situation assessment, and uses social models to provide situation understanding. However, as the power of DL continues to grow, there is a need to consider contextual information as augmented data, contextual constraints selecting relevant data, and contextual prediction as forecasted data. Some DL contemporary research endeavors and future trends (Blasch, Liu et al. 2018) include the following:

Data quality — use more valuable and contextual data before trying to change the model.

Data augmentation — use normal data extension techniques and unsupervised generative models.

Class sampling — model relevant context parameters with equivalent numbers of samples per class.

Ensemble support — train separate networks for classifier combinations to improve accuracy.

Realistic analysis — ensure validation sets and test sets come from the same distribution.

Scalability — design computing methods that expand with more data and model complexity.

Human-level performance metrics — use domain experts and regular users to compare system performances.

The importance of data management for multimodal sensing and action of context-based AI systems supports data at rest — provide structure (that is, translations) between data for integration, analysis, and storage; data in collect — leverage the power of modeling from which data are analyzed for information, delivered as knowledge, and supports prediction of future data needs; data in transit — develop a data as a service architecture that incorporates contextual information, metadata, and information registration to support the systems-of-systems design; data in motion — use feedback control loops to dynamically adapt to changing priorities, timescales, and mission scenarios; and data in use — afford context-based human–machine interactions based on dynamic mission priorities, users, and situations to balance needs, recommendations, and availability (table 2).

One example of these data management methods for analytics occurs in physics-based and human-derived information fusion (information fusion; Blasch et al. 2014) that coordinates data collections through a user-defined operating picture in support of situation analysis (Blasch 2013). Contemporary issues concern situational reasoning, knowledge management, and command and control for AI (where the *A* in AI could extend to automated, augmented, or autonomous).

Intelligence is the ability to recall, reason, and predict. To foster context-based AI requires data models (Blasch, Ravela, and Aved 2018), situational analysis (Snidaro et al. 2016), and systems cooperation (Peterson and Paley 2011). A data model is a computing paradigm that organizes high-dimensional complex information for indexing and recall such as context-aware computing. To predict future events requires a mathematical model based on a set of parameters that is built on repeatable explanations using deductive logic such as context-adaptive control. Finally, reasoning leverages understanding from a conceptual model and coordination with other knowledge for context-enhanced information fusion. Although conceptual models are not well embedded in machines, there are evolutionary, experimental, and intuition-from-inductive-logic methods (for example, nonmonotonic logic) that support knowledge management. To explain situations or events requires various multimodal sensing, scenario modeling, and knowledge reasoning approaches for coordinated action. The power of context-aware, context-adaptive, and context-enhanced multimodal approaches, combined with AI and ML, expands the employment of future systems such as autonomous cars, UAV swarms, and mobile applications.

This article advocates AI multimodal sensing and action that combines models, measurements, algorithms, and computing. The next section discusses multimodal information fusion (*data at rest*) while the following one highlights contextual reasoning from a variety of ML techniques (*data in collect*). We then discusses the dynamic data-driven applications systems (DDDAS) paradigm for using models (*data in transit*).

Data (Autonomy)	DDDAS	Example
Data at rest	Statistical algorithms	Information fusion (Liu, Z. et al. 2018)
Data in collect	High-dimensional model learning	Road networks (Yang and Blasch 2008)
Data in transit	Systems software computing	Container-based agents (Wu et al. 2016)
Data in motion	Instrumentation and control	Imagery collection (Blasch et al. 2018)
Data in use	Human–machine AI	User-defined operating picture (Blasch 2013)

Table 2. Data Management for Context-Based AI.

The Human–Machine AI section discusses human–machine teaming (*data in use*). The Command Guided Swarms (CGS) section provides a motivation for CGS (data in motion). We then provide an example of context-based AI multimodal image fusion building on the above concepts for CGSs, and then provide conclusions.

Information Fusion

The four types of AI (table 1) build from designing simple devices to complex machines toward the goal of self-awareness (Hintze 2016). Self-awareness typically relates to humans, while self-assessment relates to machines and both are related concepts. Situation analysis (for example, assessment, awareness, and understanding) are commonly researched in the information fusion community leading to human–machine information fusion systems. The data information fusion group model (Blasch et al. 2012), shown in figure 1, leverages AI developments at each processing stage to support assessment (level 0, 1, 2, 3 information fusion) to that of refinement (level 4, 5, 6 information fusion). System management (level 6) provides contextual constraints based on missions, objectives, and goals. The data information fusion group model aligns with AI types (table 3) as: Type 1 — reactive machines with rules support L0 processing; Type II — limited memory signal processing methods are L1 functions; Type III — theory of mind situation representations compose L2/3 goals; and Type IV — self-awareness prediction and interaction result from L4/L5/L6 analyses. The descriptions of the information fusion levels demonstrate the functions aligned to the various levels synergistic with AI opportunities for autonomy. They divide into *sensing* (low level information fusion — assessment, level 1, 2, 3) and *action* (high level information fusion — control, level 4, 5, 6).

Level 0 — Data Assessment: The estimation and prediction of signal/object observable states on the basis of pixel/signal level data association (for example, information systems collections).

Level 1 — Object Assessment: The estimation and prediction of entity states on the basis of data association, continuous state estimation, and discrete state estimation (for example, data processing).

Level 2 — Situation Assessment: The estimation and prediction of relations among entities, to include force structure and force relations, and communications (for example, information processing).

Level 3 — Impact Assessment: The estimation and prediction of effects on situations of planned or estimated actions by the participants; to include interactions between action plans of multiple players (for example, assessing threat or intent actions to planned actions and mission requirements, and performance evaluation.

Level 4 — Process Refinement (this is an element of resource management): The adaptive data acquisition and processing to support sensing objectives (for example, fusion process control and information systems dissemination).

Level 5 — User Refinement (this is an element of knowledge management): The adaptive determination of who queries information and who has access to information (for example, information operations) and adaptive data retrieved and displayed to support cognitive decision-making and actions (for example, human systems integration).

Level 6 — Mission Management (this is an element of platform management): The adaptive determination of spatial–temporal control of assets (for example, airspace operations), route planning, and goal determination to support team decision-making and actions (for example, context operations) under social, economic, and political constraints.

From these AI types and information fusion levels, there is a need to further enhance the systems to adapt and respond in support of multiple users, various situations, and distributed machines, which is developing for automation, augmentation, and autonomy.

A key aspect of the data information fusion group model is the ability to align physics-based and human-derived information fusion over sensed data for action. Scerri et al. (2015) developed a context-aware situation analysis device. Information fusion methods seek a similar goal. Physical data includes processing signals, extracting features, and making decisions, while the human-derived information is analyzed with logic, symbols, and commands as shown in table 4. AI methods have enhanced the ability to process data; and recent results in AI leverage logical and semantic rules.

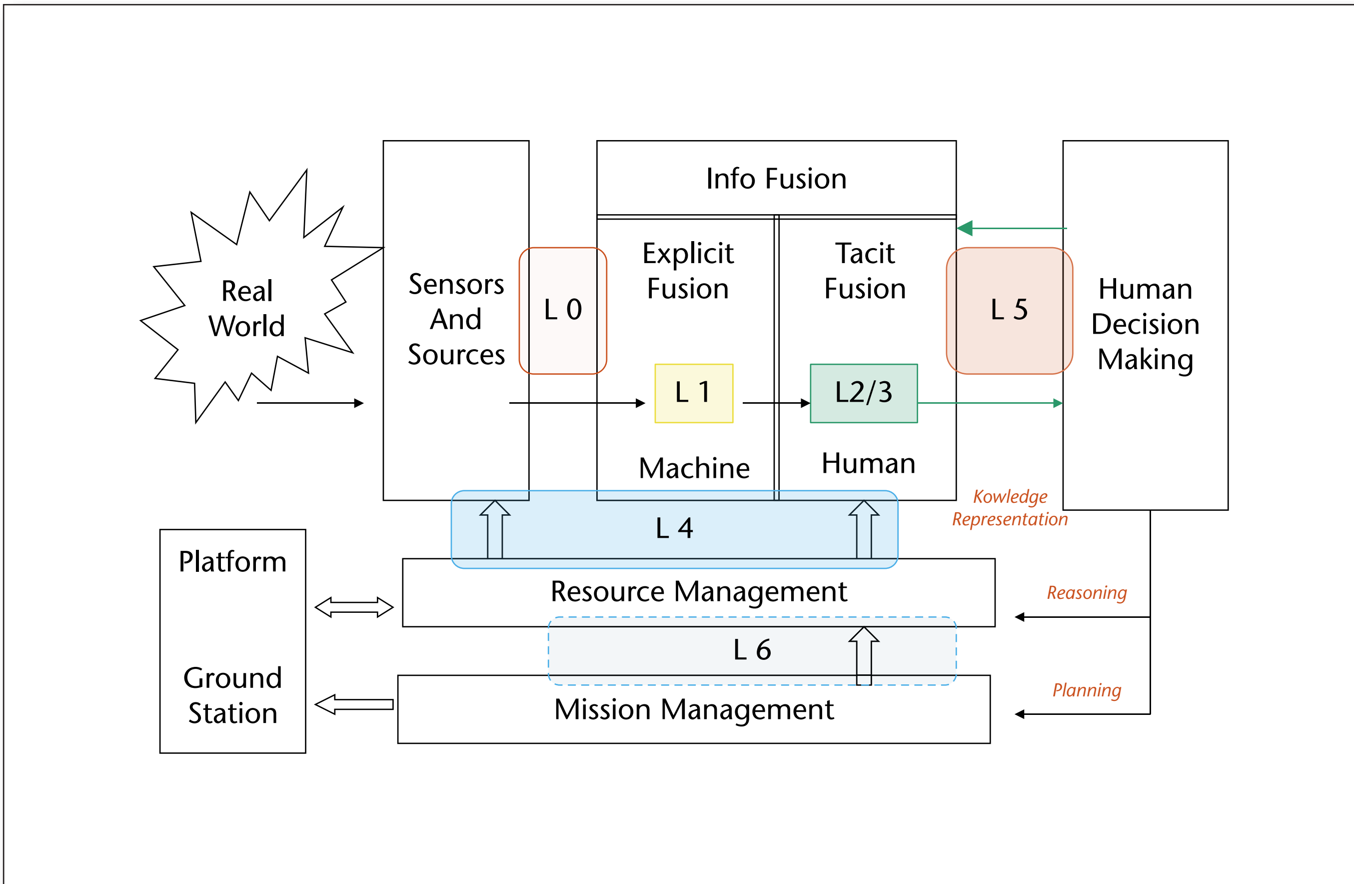

Figure 1. Data Information Fusion Group Model.

However, there is a need to develop AI systems that support automation (that is, knowledge acquisition) to autonomy (that is, to be commanded). To bring together multimodal sensing and action requires the ability to support statistical analysis (for example, probabilistic) with logical analysis (for example, formal axioms).

There are many publications and engineered systems that have designed, developed, and deployed information fusion systems with multiple forms of intelligence. The physical intelligence comes from the sensors, environment, and objects, such as from modeling of high-dimensional contextual information and target tracking (Yang et al. 2012; Dunik et al. 2015). High-performance computing has fostered the ability to take advantage of the various modeling approaches based on first-principles for assessment, estimation, and prediction. The vast amount of data coming from signals, images, and communications provides an influx of sensed data. However, aspects of multimodal and multidimensional spectra, along with spatial and temporal analysis, are needed for machines to process information through statistical or logical means. For example, in surveillance scenarios, data can originate from infrared and visual imagery supported by sensor models, while the context includes road networks and environmental conditions.

Type of AI	Focus	AI with information fusion
Type I	Reactive machines	L0 Data assessment
Type II	Limited memory	L1 Object assessment
Type III	Theory of mind	L2 Situation assessment
		L3 Impact assessment
Type IV	Self-awareness	L4 Process refinement
		L5 User refinement
		L6 Mission refinement

Table 3. AI Types from Table 1 Aligned with information fusion.

Another recent trend in social intelligence is the ability to put the machine in the context of the objectives required for use. Many AI systems must operate within the human, social, cultural, and behavioral modeling and analysis. Although some effort has fostered modeling, it is far from adequate (Blasch et al. 2013). However, the explosion of big data from human information is evident in modern networked society. Hence, intelligence from social information includes text, reports, policies, and laws that seek to

	Processing	Exploitation	Action	Dissemination
Data	Statistical	Features	Decision	Machine engineering (sensor)
Rules	Logic	Symbol	Command	Knowledge engineering (human)

Table 4. Machine-Knowledge Coordination.

provide the community with a set of rules. In a surveillance scenario, the data can come from location and user reports augmented by historical models of object behaviors in different scenarios using latent analysis for categorical assessment (Kashoob et al. 2009; Blasch et al. 2014), while the context includes known cultural norms of the desired area.

Another aspect is cooperative intelligence, which is the ability of multiple systems to coordinate actions. Many AI systems must develop joint training methods to determine if coordinated action enhances performance. Recently, efforts have been focused on cooperative deep neural networks such as multimodal information fusion (Ngiam et al. 2011), signals fusion (Shen et al, 2018), and image fusion (Zheng et al. 2018). The multimodal analysis can serve not only with assessment, but also in developing models for coordinated action. In a surveillance scenario, cooperative sensing can position sensors collecting data at the correct locations using contextual information to facilitate relevant data collection for object detection as in swarm behavior (Cruise et al. 2018).

Elements of AI Systems for Contextual Reasoning

Physical, social, and cooperative intelligence includes similar aspects of model building, entity extraction, relationship linking, and event assessment. Emerging concepts include graphical models, Markov logic networks, statistical relational learning, and DL networks. Together these approaches support contemporary efforts in AI/ML toward contextual reasoning for machines to be self-aware in explaining choices for actions.

The three waves of AI include the first phase (1960–1980) for handcrafted knowledge and rules (Fogg 2017; Cruise et al. 2018), as shown in figure 2. The second phase (starting in 1990) includes popular methods in ML using statistical analysis such as natural language processing and computer vision. The third wave (current) seeks to develop explainable methods for scenarios and situations. However, there is still a gap in machine–human teaming, machines that think, and machines that rival humans with common sense. To address the issues of human–machine teaming, it is important to understand the different types of AI and ML methods as well as combining data such as deep multimodal image fusion (Liu, S. et al. 2018).

Learning Methods

ML attempts to build models in support of AI goals, for which a variety of methods exist (Domingo 2015; Cruise et al. 2018). The types of ML approaches foster multiple opportunities for analysis, many of which are common in information fusion methods and typically designed for the specific scenarios, data, and reasoning desired (table 5).

The various ML approaches stem from the types of data that are collected (figure 3), either *unlabeled* or *labeled*. If a world model is known with prior information (for example, two-class decisions), then probabilistic Bayesian methods can be used. However, a Bayesian world model is hard to validate for completeness (as in moving from a two-class problem to an unknown set of classes). Hence, most model labeling is incomplete. Labeled data for training is termed *supervised*. Supervised learning, which is often used in DL, affords methods of categorical classification with analogistic, connectionist, and symbolic reasoning. When unlabeled data are processed, there is a need to use alternative methods such as possibilistic and evolutionary approaches; however, these methods are an ad hoc approach toward understanding and repeatability. The future of ML is most likely a combination of these approaches, such as semi-supervised learning that includes a small amount of labeled data among a large set of unlabeled data.

The development of DL extends these ML approaches that can be organized from the methods in figure 3. An initial question is whether contextual a priori, or historical, information can be used; is available; and represents the real world (or constrained subset of the real world as a specific situation of a given scenario). With prior statistics as contextual information, probabilistic methods provide appropriate, sufficient, and reasonable results. While Bayesian decision theory is based on complete world knowledge, if a subset of the world knowledge is assumed to be well known (for example, modeled), then the maximum a priori analysis is used; however, if the world is not known, a comparative maximum likelihood estimator can be used. Which method to use is based on data availability and model completeness (which is generally incomplete to cover all circumstances). If the data are not available, then physical first-principles or syntactic grammar methods are used to support symbolic analysis. The symbolic approaches can also be appropriate if there are few decision boundaries that lead to general parametric decision rules. Recent trends

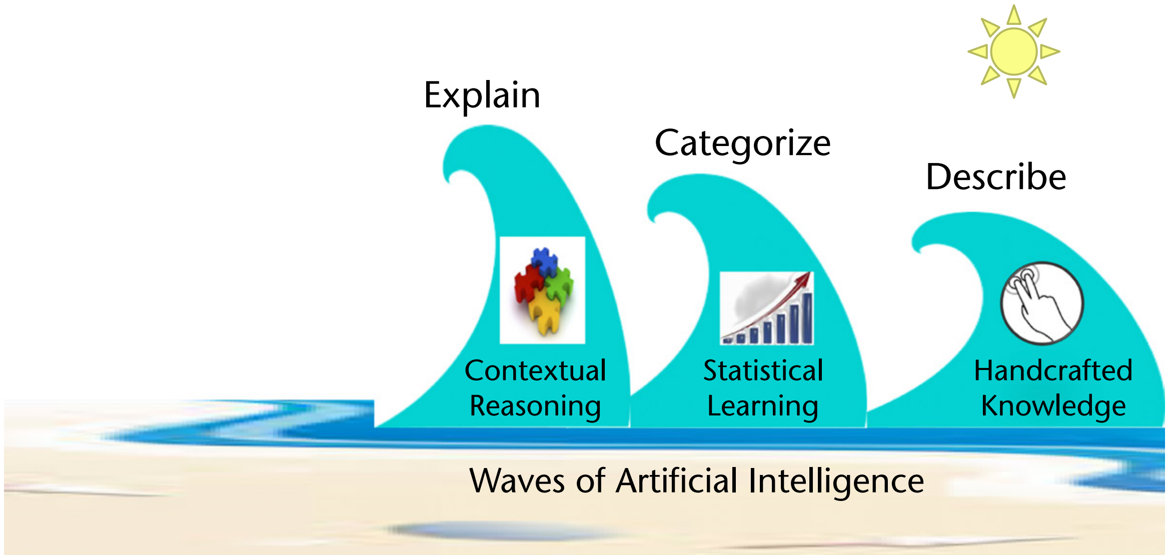

Figure 2. Three Waves of AI.

Name	Device	Approach to Data Analysis
Symbolic	Logical statements Expert systems	First order logic, truth tables
Probabilistic	Graphical models Subjective logic	Bayesian statistics, conditional probabilities
Connectionist	Artificial neural networks Multilayer artificial neural networks (DL)	Computational model of linked statistical error gradient minimization
Analogistic	Support vector machines Kernel methods	Pattern recognition via distance computations in feature hyperspace
Evolutionary	Genetic algorithms Genetic programming	Competitive random variations for discovery of survival adaptations
Possibilistic	Fuzzy inference systems Evidential reasoning	Expansion of classic logic to accommodate ambiguous partial truths

Table 5. Types of ML.

explore more-complex scenarios to include large volumes of data from which it is difficult to completely model the entire world, and hence, statistical methods are used, such as DL.

The challenge for most complex scenarios is that the a priori information is unknown, the data collected is unknown, and the objective is unknown. Although many techniques are exploratory, the discovery of some attributes infers relationships. For example, evolutionary approaches mimic an understanding of genetic diversity that allows for mutations and adaptations. If a reasonable solution is found from clustering, then some unknown classes are revealed, leaving a set of data still unclassified. In other cases, partial learning results explain some details. Another approach is more propositional, leveraging the symbolic and probabilistic combinations such as a Gaussian mixture model based on known classes.

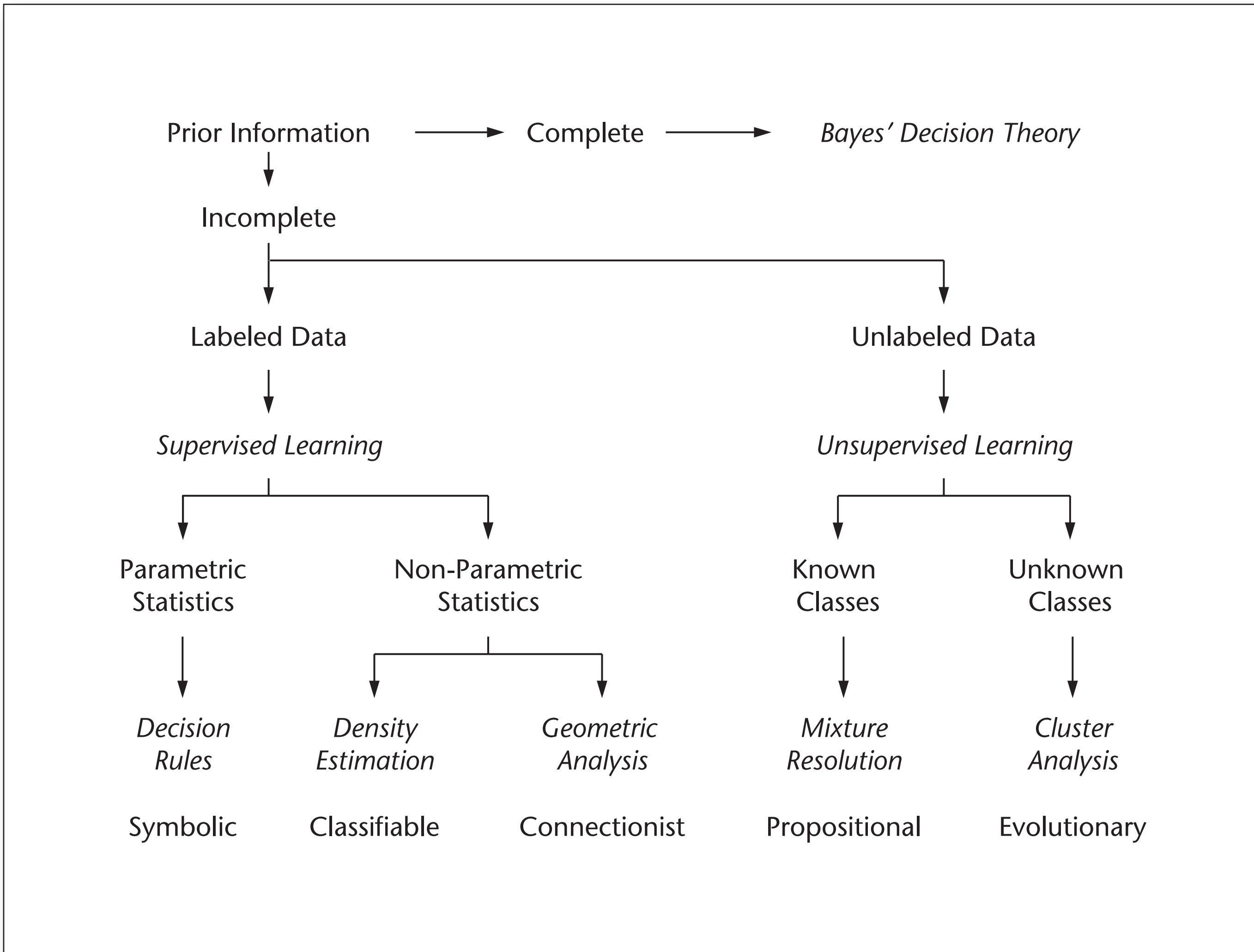

Figure 3. ML Approaches Based on Available Data.

Another set of techniques focuses on nonparametric statistics from labeled data to include classification and connectionist approaches. Classification methods seek to organize the data into categories by determining a decision boundary between the categories (for example, support vector machines). These methods have grounded mathematical theory for repeated results, but provide limited robust operations in complex scenarios.

The current set of AI techniques follow from DL rooted in connectionist neural networks. Minsky and Papert (1969) highlighted that the original perceptron (one layer) could not perform the exclusive OR (XOR) function. The challenge was quickly solved by adding another layer in the network (two layers) such that the middle layer learned the XOR function, while the final layer did the classification. An example is finding the XOR bounding box (inside versus outside the classification boundary box). Adding the middle layer for the four sides of the box (top, left, bottom, right) supported XOR learning. Extensions of these ideas have resulted in the convolutional neural network (CNN). The CNN includes choosing a kernel (for example, 5 × 5 bounding box in an image), learning the classification, and typically using backpropagation with stochastic gradient descent to learn one layer at time. By connecting all of the partial layer results, reasonable classifications are obtained — subject to the appropriate choice of the number of layers. More exemplar data increase the likelihood the system would be able to perform well with increasingly complex situations. The advantage of the CNN is that the kernels do not need the full transformation equations, and with an autoencoder as an artificial neural network, neighborhood functions can be used to filter the results for parametric reduction. The number of layers in deep NN is part of ongoing work; for example, hyperparametric learning has replaced the sigmoid function in traditional NNs to include the rectified linear unit for gradient analysis.

Multimodal DL

Multimodal DL (Ngiam et al. 2011) has focused on leveraging the different learning approaches over the type of data and application (Zuo et al. 2017) that extends methods such as image fusion clustering methods (Kim et al. 2016). For example, while the CNN is showing promise for images, the recurrent neural network (RNN) has shown better results for language data. The CNN works for physics-based data that has spatial relationships, while the RNN supports human-derived data for feature analysis, entity extraction, and relationships modeling. Combining the RNN with the CNN for different types of data supports a more robust multimodal sensing, action, and event analysis (table 6). Likewise, it is important to note that implementation on computers is different in that the CNN supports graphical processing units while the RNN is appropriate for neuromorphic computations.

The challenges of using AI and ML techniques relate to the data available and the performance metrics desired. Vyas et al. (2012) highlight data fusion performance variations of data variability, relevance, and reliability; these do not always correspond to model accuracy. Not all data are available so there is a need for contextual information to support incomplete data. Contextual information is available from dynamic statistical analysis (data at rest); historical data from models (data in collect), real-time data sensing (data in motion); data flows such as partitioned data in the cloud or at the edge over distributed communications (data in transit); and finally, relevancy — that for which the user is most interested in (data in use). Sensing relevant data supports human–machine AI reasoning, while modeling helps in collection needs and context information supports explainable results. Using models to seed relevant data for complex situations when data are not available will be required for robust DL applications. The augmentation of real-time data collects with simulated data from models is a thrust of the DDDAS paradigm.

Vision	Language
CNN	RNN
Corpus of images	Corpus of books
Relation: Spatial neighborhood	Relation: Probability word order

Table 6. Multimodal ML Methods.

DDDAS AI

The DDDAS paradigm (Blasch et al. 2018) brings together modeling, measurements, statistics, and simulation-based computational frameworks. Examples of theoretical analysis from DDDAS that are used in AI theory include uncertainty quantification, model reduction, and objective satisfaction. Current DDDAS methods span UAV systems analysis to the Internet of Things (Li and Darema 2017). The use of a model augments data collections where there is incomplete, partial, or sparse data. For example, the model can support data extraction, feature classification, and knowledge generation for robust DL. The data can come from the physical system via the sensor reconfiguration loop or from a simulated system through the data assimilation loop. Simulated data can extend AI and ML methods by harnessing the ability to get access to labeled data with model predictions that are based on the first principles from mathematical or formal models. Additional methods develop categorical models from human-derived data with a human (sensor) reconfiguration loop.

A simple outdoor surveillance example presents a use case where a day and night luminance model can provide a general understanding of the situation when streaming data are not completely available (Li et al. 2017). Changing lighting conditions determine whether an infrared or a visual image should be used, and in the case of dawn or dusk, whether both should be used (Zheng et al. 2018). Engineered systems, such as a UAV, could use Bayesian filtering for flight control, symbolic logic for image analysis, and ontology categories for object labeling during different times of the day; but the complex situation would benefit from additional model understanding to provide relevant data to a multimodal DL. Multimodal DL, along with context, would support streaming data analysis (data in motion) from the images available from onboard sensors and text classifications from the ground human observers (data at rest). The notional example is based on the ability to simulate an enormous number of situations for unknown scenarios, from which DL provides classification decisions for these complex events.

Building from the DDDAS control model, there is the sensor reconfiguration loop and the data assimilation loop. Data are available from the UAV for flight control, while luminance information is available from the environment. The augmented data comes from the simulated model to capture the high-dimensional information when only a small portion of images are available for certain luminance values; that is, the changing sun patterns, sensor-to-object profiles, and interpretability options. Likewise, there is a low-dimensional model of the UAV dynamics for path determination. The evaluations of (fused) information from the lighting conditions and the UAV dynamics are combined for real-time control and object analysis. Advances in AI support the computational analysis that comes from model learning, parameter assessment, and model reduction; for example, a high-dimensional manifold reduced to an appropriate dimension to coordinate multimodal fusion with a low-dimensional UAV model. For final decision-making, human–machine AI coordinates the display for the UAV operator to provide general commands (for example, where to image) for automatic target recognition as for data in use. During real-time dynamic operations, the AI system provides situation understanding by

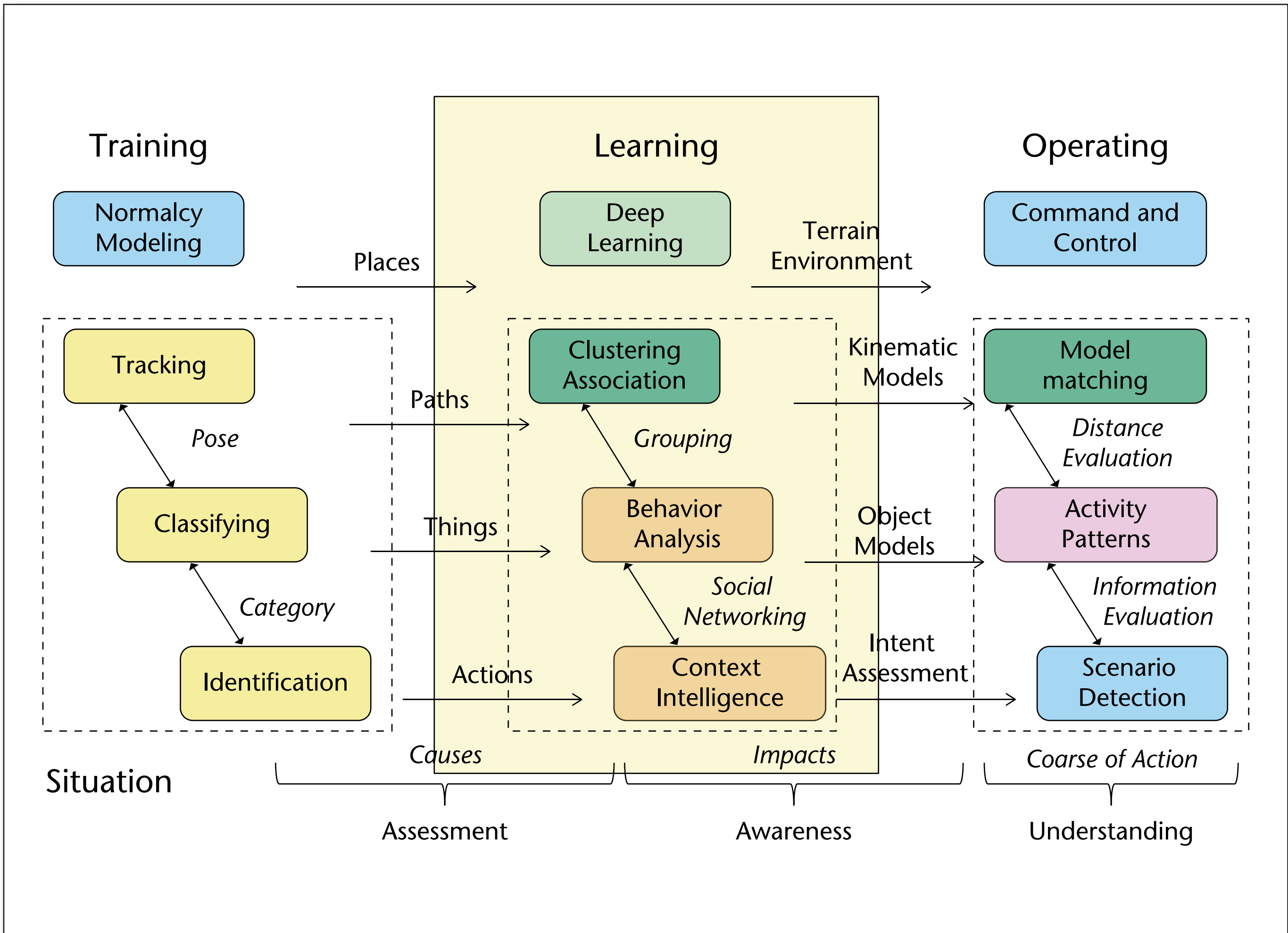

Figure 4. Methods of Situation Understanding.

Given the advances in ML techniques, human–machine AI teaming is required to gain an advantage in complex situations and increase the efficiency and effectiveness of sensing and action.

leveraging situation assessment and situation awareness through contextual intelligence.

Figure 4 describes the conceptual shift from context-based situation assessment, to awareness, to understanding. The typical AI approaches serve advances in machine-level situation assessment, while presentation of the information to the user is an element of situation awareness. Future AI systems for command and control would be mission-driven or goal-driven to incorporate situation understanding (that is, as aligned with the context). Using AI models and analyses, predictions of future data collections that are optimal would enhance AI capabilities. One example is scenario detection in which the trained AI data analytics, combined with augmented contextual intelligence, determine which set of sensors and fusion methods would optimize the command and control of situation analyses. As a common occurrence of context intelligence, interdependence among system components may require dramatic changes upon training and operations to achieve a desired understanding (that is, social situations among humans produce different cluster associations and analyses of behavior that result in different identification interpretations and scenario detections).

Human–Machine AI

The user would like a trustable result with high reliability, confidence, and credibility, which also render low uncertainty. The problem with many AI and ML techniques that seek a classification boundary is that they do not perform uncertainty quantification. Probabilistic approaches (for example, Bayesian filter) support data in motion for which the immediate real-time control needs to satisfy reducing the uncertainty by minimizing the error. If the data are from multiple sources, perspectives, and situations, then the uncertainty can be further reduced. A control paradigm for data in use is a sense–act–learn paradigm

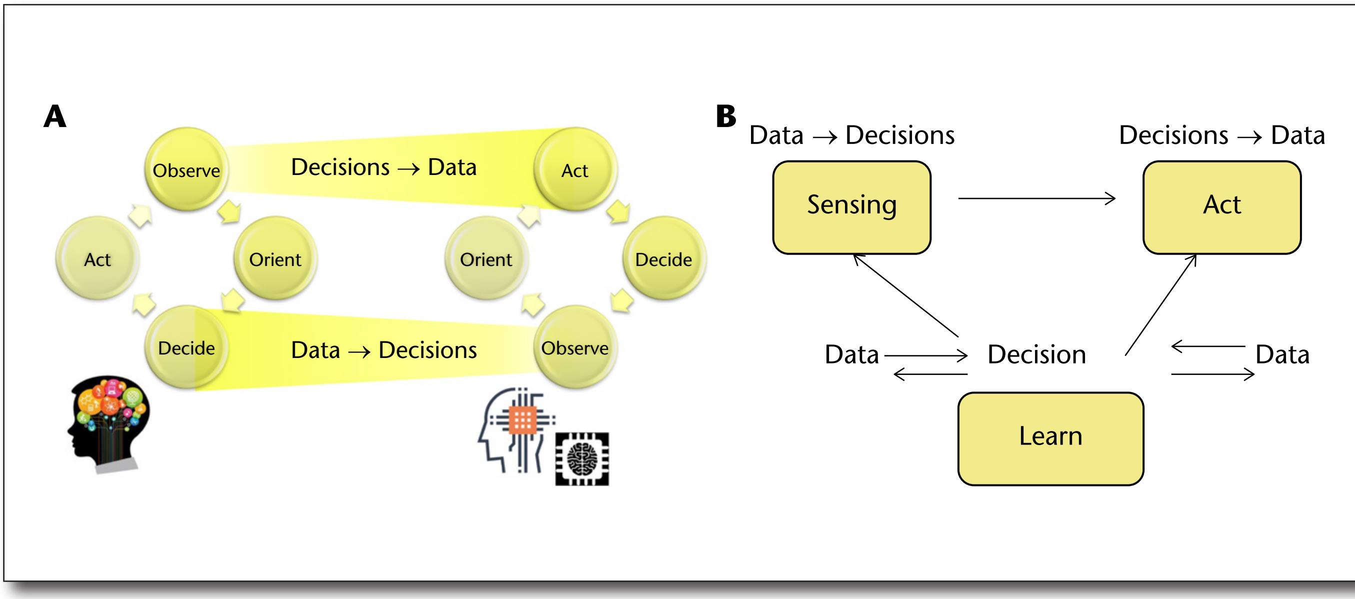

Figure 5. Human–Machine AI Teaming.

(A) Observe-orient-decide-act loops. (B) The sense–learn–act loop.

that builds from the observe-orient–decide–act loop (Blasch, Bosse, and Lambert 2012). Although the orient–decide–act loop is well known (figure 5a), the goal is to have the human and machine team in the learning process. The human seeks results from the machine (observe) to decide and act, which is currently known as data-to-decisions. From the human observation of the situation, they would rather select data in concert with the machine decision boundaries — decisions-to-data. Hence, the decision and data selection should be a combination between machines and humans in sensing, acting, and learning (figure 5b). Data-to-decisions (classification learning) from the machine should be coordinated with the human that requires decisions-to-data (command and control data collection). Such an example is a user query for object recognition and the machine to output relevant classifications. Together, human–machine AI would help to enhance the usability of current AI systems.

The design of the human–machine AI system should take advantage of contextual data that a machine can process from its large collection of historical data (for example, terrain, weather, infrastructure) that is compiled into available first-principle (for example, for sensor data), categorical (for example, for human-derived data), or statistical (for example, for unknown data collection) models. The use of contextual data provides methods for DL. However, the question remains, what is learning in the sense of human cognition, machine computation, and explainable analysis? Learning is the ability to understand the surroundings through experience. Humans learn through situational awareness; although the machine is not necessarily aware, it can be equipped with related capabilities to provide an assessment.

The use of the physical models supports physics intelligence, which is a model of cause and effect in the world. For example, if the challenge is identifying parts of a vehicle, does the vehicle have four wheels or two wheels based on partial observations, planar images, or few pixels? If the context is known, such as a bike path, the human can determine it has two wheels from its trajectory, but a computer cannot. Another model is social intelligence, such as the networking of interactions, which requires human assessment. When the UAV sees a vehicle in an image, the machine cannot tell if the vehicle is preparing to turn left or right. Hence, DL cannot learn the intent of objects, only the spatial temporal relations of an object's parts. The future of AI is cooperative intelligence that leverages the contextual knowledge between humans and machines as context awareness (figure 6). Future explorations require a combination of AI approaches as shown in table 7.

AI systems must also coordinate multiple distributed user and machines such as a CGS.

Command Guided System

An example of multimodal sensing and action is a CGS. Key challenges of CGSs in current development (Rajkumar et al. 2017; Cruise et al. 2018) are: (1) Coordinate multiple disciplines — merge expertise in software, hardware, sensing, and modeling; (2) Characterize time — record discrete time multiresolution, latencies, and out-of-order time stamps; (3) Guarantee clock synchronization — align sensors and platforms, clock drift, and distributed entities; (4) Establish real-time sensing — verify constraints, software/physical interactions, and scheduling; (5) Determine component interactions — design for asynchronous

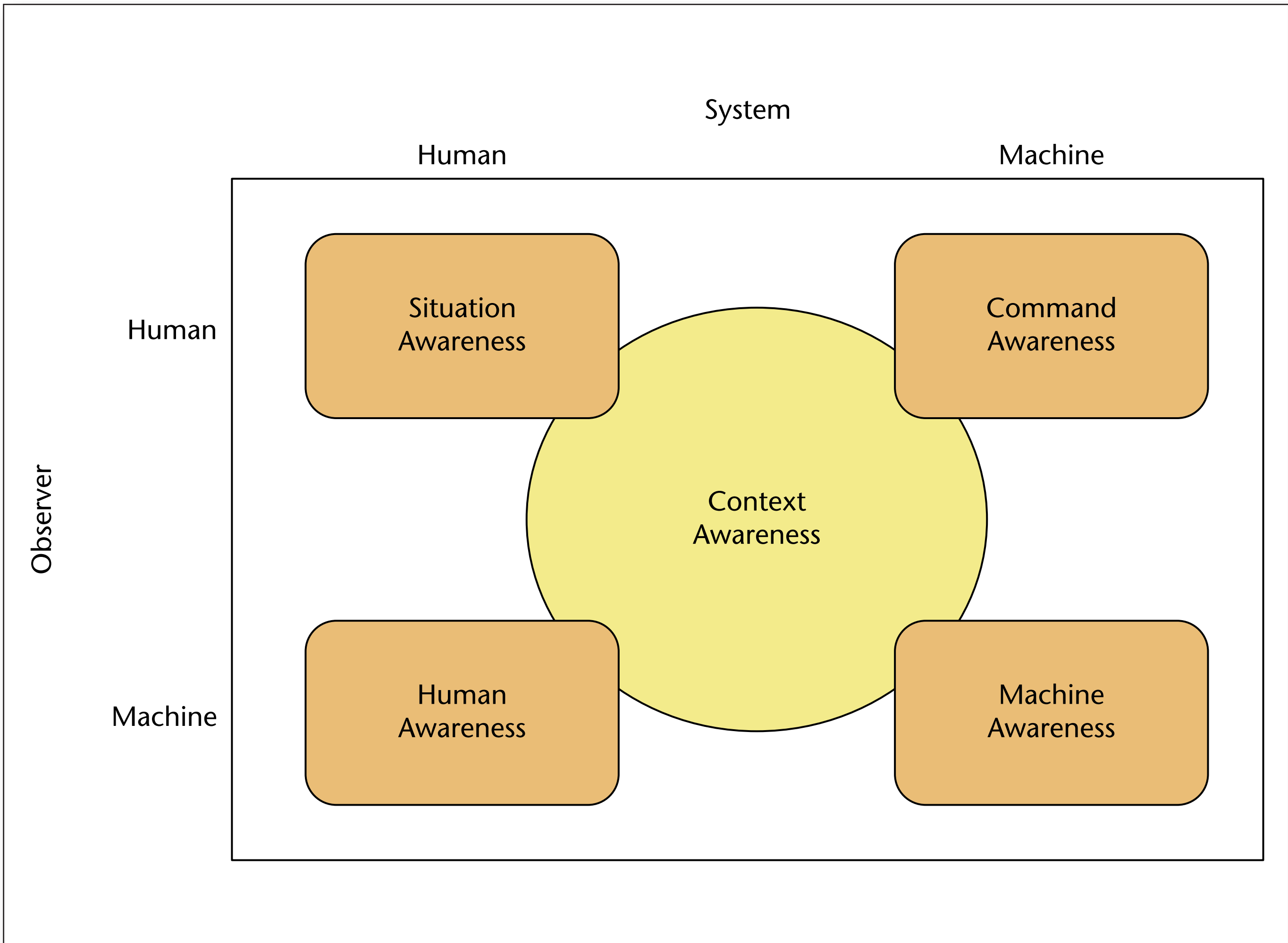

Figure 6. Human–Machine Context Awareness.

memory and incorporate AI agents; (6) Design wireless interactions — maintain communications, design access, secure, routing protocols; (7) Maximize integrity — provide performance under failures, quantify component criticality; (8) Ensure cyber security — analyze attacks, use systems theory for countermeasures; and (9) Establish cooperation — determine relations between qualitative semantics and quantitative control.

Most of these challenges are associated with the data before employing AI methods. These challenges foster different design techniques to coordinate humans and machines. One solution for a CGS design utilizes three types of agents including (1) information fusion (IF), (2) operator infusion (OI), and (3) control diffusion (CD; Cruise et al. 2018). The agents provide different data for context awareness as the IF-agent processes physical data, the OI-agent provides social knowledge, and the CD-agent conducts cooperative action. The organization of these agents includes different architectures (figure 7) such as distributed agents, coordinated users, and actions that are *command-guided*.

For distributed agents, all systems possess the three agents (IF, OI, CD) as operator nodes through a cloud that synchronize plans and updates (figure 7, top). For coordinated users, the commander manages other human agents (OI) as swarm pilots, who then interact with machine agents (IF, CD) as fog nodes sending organized results (figure 7, bottom). For command-guided actions, the commander directs human–machine agents (OI/CD) that collect data (IF) to form decisions through a data refinement loop as user nodes (figure 7, left), or the commander provides goals to the machine–human agents (IF/OI) from which decisions determine the sensor control loop of what data to collect (CD) at edge nodes (figure 7, right).

Hence, a CGS system takes advantage of centralized command with distributed execution by expressing goals and having the contextual agents develop the sensing and action strategy. Of the many ideas presented in this article, the next section provides an example of using contemporary deep-multimodal image fusion with contextual analysis for command-guided collections.

Observer	System	
	Human	Machine
Human	Situation-aware AI (human analysis of human)	Command-aware AI human analysis of machine
Machine	Human-aware AI (machine analysis of human)	Machine-aware AI (machine analysis of machine)

Table 7. Context-Aware AI systems.

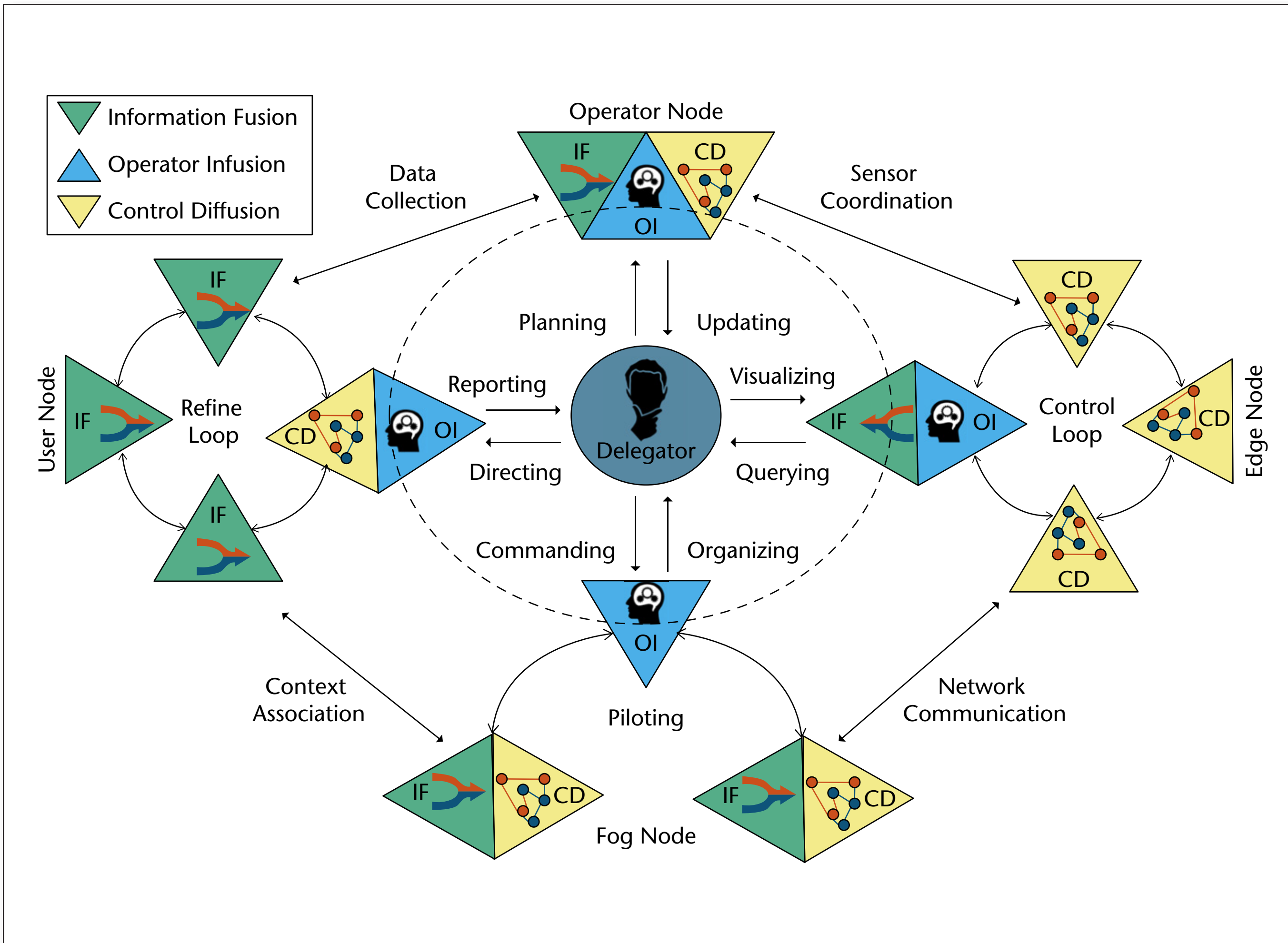

Figure 7. Conceptual CGS Architectures.

Example: Coordination DL Image Fusion

Assume that a commander desires multiperspectives and multimodal observations of an object. In such a case, there are available UAVs each with different types of sensors. The goal is for the swarm to evolve such that the positions for the viewpoints of the multimodal sensors (for example, visual and infrared) observe an object from different perspectives and distances. The maximum approach (Yang et al. 2013) is to position the swarm sensors at orthogonal viewpoints (90 degrees). However, a competing analysis is sensor fusion in which the view of the object should be the same (0 degrees) to maximize data registration. The overlap and orthogonal viewpoints are competing constraints that support performance, but require coordinated control. For the analysis, three things are leveraged: models — theoretical performance of task success based on the range to an object; methods — empirical performance of deep-multimodal image fusion; and

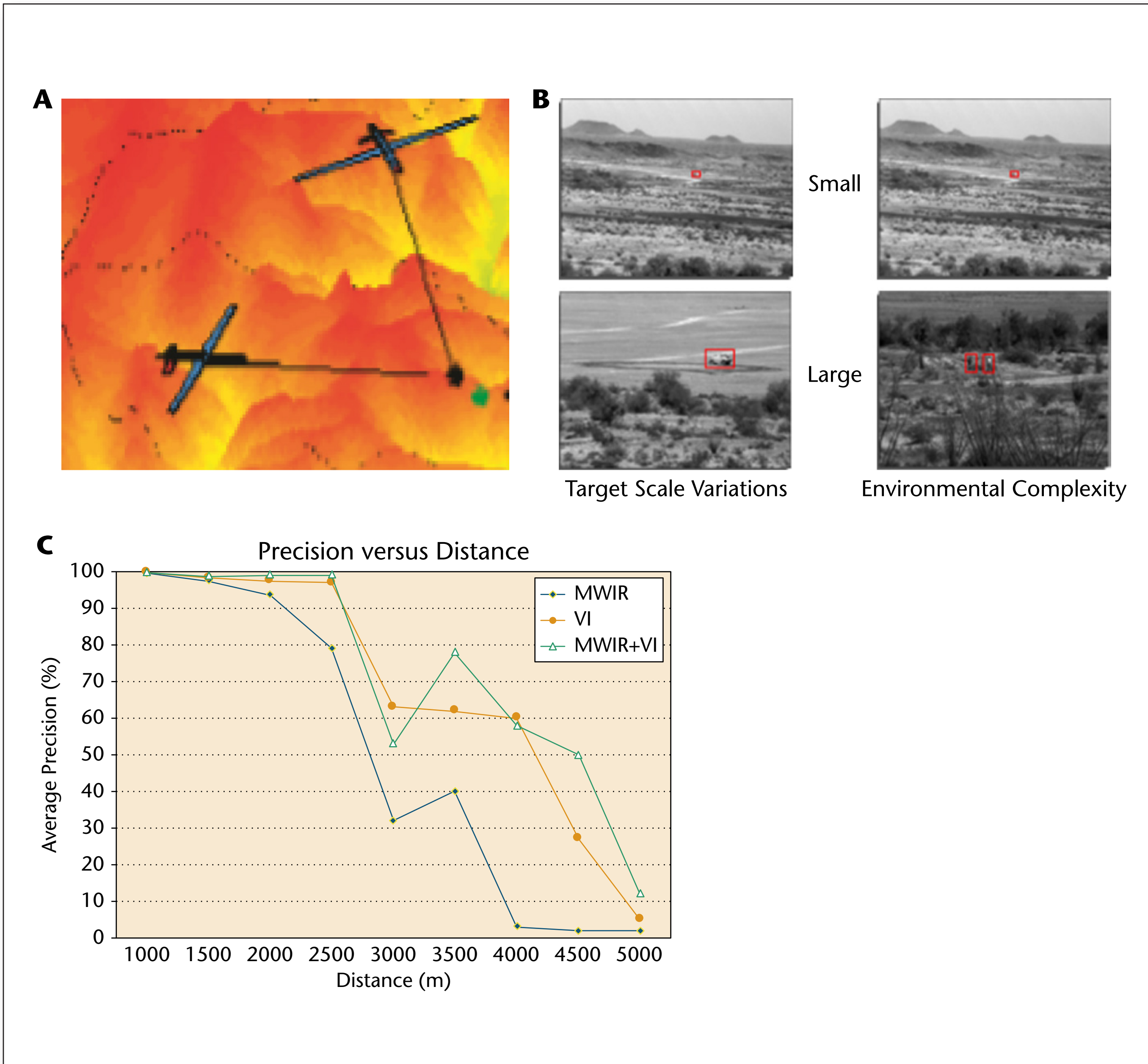

Figure 8. Deep Multimodal Image Fusion Analysis.

(A) Swarm of sensors. (B) Imagery showing number of pixels for object recognition. (C) Increased precision using DL methods with fused imagery versus single modalities.

control — perspective performance from coordinated UAV positioning.

Theoretical models relate the object distance to the image resolution available for visual electro-optical and infrared object detection and classification (Kahler and Blasch 2010; Blasch et al. 2018). When the sensor-to-object distance gets smaller, the resolution increases and the classification improves. Hence, it is advantageous to have the command-guided UAV move closer to the object (in range). To investigate multimodal image fusion using DL methods, a theoretical analysis of the task success compares a single modality (for example, electro-optical) versus multimodality (for example, electro-optical/infrared), which indicates that it is generally desired to get within 3 kilometers of the object (Blasch and Kahler 2015) based on the sampling distance.

With the theoretical performance analysis showing advantages in image fusion, an interest in AI methods for contextual analysis was developed. Context from the scenario includes lighting conditions (whether to use visual, infrared, or visual + infrared) imagery results, and the position of the sensors as a function of range. Deep-multimodal image fusion was performed using the Defense Systems Information Analysis Center Automation Target Recognition Algorithm Development Image Database package (DSIAC 2018). Inherently, the multiview swarm sensing advantage (figure 8a) must be balanced with the action complexity (figure 8b). Figure 8b highlights the outlined objects in red boxes to demonstrate the complex contextual variations of the sensor (near or far) and environment (open or cluttered). The Defense Systems Information Analysis Center Automation Target Recognition database provides many images to investigate variations of scenarios over models and methods. The combination of the theoretical models (that is, data at rest) and DL experimental results (that is, data in collect) enhance performance. Figure 8c shows the results highlighting that, for greater than 3 kilometers, DL image fusion should be used but, as the sensor-to-object gets closer, visual imagery should be selected.

For CGS multimodal sensing and action (that is, data in motion), the user commands a desired result and the context-based performance includes the mathematical models, DL image analysis, and UAV routing as a human–machine agent design (command-guided architecture in figure 7). In a CGS scenario, it is assumed that a user selects the desired result such as verified object recognition from two types of sensors. In a noncooperative sensing scenario, both UAVs pursue the target without knowledge of the other UAV. For the cooperative sensing case, the human–machine agents guide the systems to different positions, taking advantage of the context information. Examples include understanding multiperspective coordination, environmental conditions, and object behavior.

The scenario presents three ideas for future context-based AI, including scenario autonomy, environmental reasoning, and situation understanding. The first is that the human–machine teaming leverages context-by commands for system autonomy. If the user makes a command too early, the system does not yet have the learned techniques to make the appropriate decision based on data alone. However, if the user provides a general command to maximize object recognition, the CGS can optimize data collection in support of the mission. Context supports achieving the goal from the system agents for cooperative intelligence.

The second attribute includes context-of information (Snidaro et al. 2016). A high object speed indicates that the object is likely on a road network. The CGS uses context-of estimates for the likely directions of the moving target and can maneuver platforms to optimize recognition. Other context-of methods include lighting conditions and range. Hence, inclusion of the physical intelligence can improve performance robustness.

The third attribute is context-for assessment (Snidaro et al. 2016). Referenced entities or data support understanding, such as the hypothesized object intent behavior. The most probable location provides a context-for assessment based on the object intent from which agents can estimate the appropriate actions. The context-for approach provides social intelligence to provide situation understanding.

Future AI systems can leverage DL techniques and contextual knowledge to build models incorporating physical intelligence, to use historical data as social intelligence, and to afford human–machine dynamic data for robust situation cooperative intelligence. The scenario analysis presents a simple example, providing a strategy leveraging multimodal sensing and action for complex scenarios.

Conclusions

Many AI ML approaches are based on acquiring a large set of labeled data. DL and statistical relational leaning demonstrate context-based analysis for physical intelligence. However, DL contextual understanding is limited as the statistical model is built on labeled data, but does not know details external to the data; that is, a contextual label drift available from social intelligence. Incomplete, partial, or ambiguous data require information fusion as cooperative intelligence to resolve dynamic situations. This article presents context-based AI for human–machine shared context by addressing data management concerns (that is, at rest, in collect, in transit, in motion, in use). A CGS example is presented that leverages theoretical knowledge and experimental deep multimodal image fusion results, where performance increases with context information. Improvements come from context-by goals for flexible cooperative control, context-of environment information for physical sensing, and context-for data from social knowledge for situation understanding.

Future opportunities being pursued for context-based AI include generative adversarial networks that act as agents based on information from real or augmented, collected or modeled, and analyzed or simulated data coordinated with IF, OI, or CD agents. The AI context-agent approaches expand with human inputs accounting for interface design considerations and complex scenarios. As many AI systems are designed for only one context, distributed methods utilizing transfer learning could support representations across agent contexts to support unknown situations and improve robustness.

References

Adomavicius, G.; Mobasher, B.; Ricci, F.; and Tuzhilin, A. 2011. Context-Aware Recommender Systems. *AI Magazine* 32(3): 67–80. doi.org/10.1609/aimag.v32i3.2364.

Amershi, S.; Cakmak, M.; Knox, W. B.; and Kulesz, T. 2014. Power to the People: The Role of Humans in Interactive Machine Learning. *AI Magazine* 35(4): 105–20. doi.org/10.1609/aimag.v35i4.2513.

Blasch, E. 2013. Enhanced Air Operations Using JView for an Air-Ground Fused Situation Awareness UDOP. In *2013 IEEE/AIAA 32nd Digital Avionics Systems Conference (DASC)*. Piscataway, NJ: Institute of Electrical and Electronics Engineers. doi.org/10.1109/DASC.2013.6712597.

Blasch, E.; Bosse, E.; and Lambert, D. 2012. *High-Level Information Fusion Management and Systems Design*. Norwood, MA: Artech House.

Blasch, E.; Chen, H-M.; Irvine, J. M.; Wang, Z.; Chen, G.; Nagy, J.; and Scott, S. 2018. Prediction of Compression Induced Image Interpretability Degradation. *Optical Engineering* 57(4): 043108. doi.org/10.1117/1.OE.57.4.043108.

Blasch, E., and Kahler, B. 2015. V-NIIRS Fusion Modeling for EO/IR Systems. In *2015 National Aerospace and Electronics Conference (NAECON)*. Piscataway, NJ: Institute of Electrical and Electronics Engineers. doi.org/10.1109/NAECON.2015.7443033.

Blasch, E.; Liu, Z.; Liu, S.; and Zheng, Y. 2018. Deep Learning Measures of Effectiveness. Paper presented at the IEEE National Aerospace & Electronics Conference, Dayton, OH, July 23–26.

Blasch, E.; Nagy, J.; Aved, A.; Jones, E. K.; Pottenger, W. M.; Basharat, A.; Hoogs, A.; Schneider, M.; Hammoud, R.; Chen, G.; Shen, D.; and Ling, H. 2014. Context Aided Video-to-Text Information Fusion. In *International Conference on Information Fusion*. Piscataway, NJ: Institute of Electrical and Electronics Engineers.

Blasch, E.; S. Ravela; and A. Aved, editors. 2018. *Handbook of Dynamic Data Driven Applications Systems*. Berlin: Springer. doi.org/10.1007/978-3-319-95504-9.

Blasch, E.; Salerno, J. J.; Kadar, I.; Yang, S. J.; Fenstermacher, L. H.; Endsley, M.; and Grewe, L. L. 2013. Summary of Human Social, Cultural, Behavioral (HSCB) Modeling for Information Fusion Panel discussion. In *Proceedings SPIE 8745, Signal Processing, Sensor Fusion, and Target Recognition XXII*. Bellingham, WA: Society for Optics and Photonics Technology. doi.org/10.1117/12.2016305.

Bredeweg, B.; Liem, J.; Beek, W.; Linnebank, F.; Gracia, J.; Lozano, E.; and Wißner, M. 2013. DynaLearn —An Intelligent Learning Environment for Learning Conceptual Knowledge. *AI Magazine* 34(4): 46–65. doi.org/10.1609/aimag.v34i4.2489.

Cruise, R.; Blasch, E; S. Natarajan; and Raz, A. 2018. Cyber-Physical Command-Guided Swarm. Journal of Defense Systems Information Analysis Center. 5(2). www.dsiac.org/resources/journals/dsiac/spring-2018-volume-5-number-2/cyber-physical-command-guided-swarm.

Domingo, P. 2015. *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World*. New York: Basic Books.

DSIAC. 2018. ATR (Automated Target Recognition) Algorithm Development Image Database (online). Belcamp, MD: Defense Systems Information Analysis Center (DSIAC). www.dsiac.org/resources/research-materials/cds-dvds-databases/atr-algorithm-development-image-database. Accessed June 1, 2018.

Dunik, J.; Straka, O.; Simandl, M.; and Blasch, E. 2015. Random-Point-Based Filters: Analysis and Comparison in Target R Tracking. *IEEE Transactions on Aerospace and Electronic Systems* 51(2): 1403–21. doi.org/10.1109/TAES.2014.130136

Fogg, A. 2017. A History of Deep Learning. www. import.io/post/history-of-deep-learning/. Accessed August 2018.

Hintze, A. 2016. Understanding the Four Types of AI, From Reactive Robots to Self-Aware Beings. The Conversation. theconversation.com/understanding-the-four-types-of-Ai-from-reactive-robots-to-self-Aware-beings-67616. Accessed August 15, 2018.

Kahler, B., and Blasch, E. 2010. Predicted Radar/Optical Feature Fusion Gains for Target Identification. In *Proceedings of the IEEE National Aerospace Electronics Conference (NAECON), 2010*. Piscataway, NJ: Institute of Electrical and Electronics Engineers. doi.org/10.1109/NAECON.2010.5712986.

Kashoob, S.; Caverlee, J.; and Ding, Y. 2009. A Categorical Model for Discovering Latent Structure in Social Annotations. In Proceedings of the Third International AAAI Conference on Weblogs and Social Media. Palo Alto, CA: Association for the Advancement of Artificial Intelligence Press.

Kim, M.; Han, D. K.; and Ko, H. 2016. Joint Patch Clustering-Based Dictionary Learning for Multimodal Image Fusion. *Information Fusion* 27: 198–214. doi.org/10.1016/j.inffus.2015.03.003.

Li, C.-S., and Darema, F. 2017. Distributed Behavior Model Orchestration in Cognitive Internet of Things Solution. *Enterprise Information Systems* 12(4): 414–34. doi.org/10.1080/17517575.2017.1355984.

Li, H.; Sudusinghe, K.; Liu, Y.; Yoon, J.; Van Der Schaar, M.; Blasch, E.; and Bhattacharyya, S. S. 2017. Dynamic, Data-Driven Processing of Multispectral Video Streams. *IEEE Aerospace and Electronic Systems Magazine* 32(7): 50–7. doi.org/10.1109/MAES.2017.160132.

Liu, S.; John, V.; Blasch, E.; Liu, Z.; and Huang, Y. 2018. IR2VI: Enhanced Night Environmental Perception by Unsupervised Thermal Image Translation. Paper presented at the IEEE CVPR Workshop. openaccess.thecvf.com/content_cvpr_2018_workshops/papers/w21/Liu_IR2VI_Enhanced_Night_CVPR_2018_paper.pdf

Liu, Z.; Blasch, E.; Bhatnagar, G.; John, V.; Wu, W.; and Blum, R. S. 2018. Fusing Synergistic Information From MultiModal Images: An Overview from Implementation to Performance Assessment. *Information Fusion* 42: 127–45. doi.org/10.1016/j.inffus.2017.10.010.

Minsky, M. L., and Papert, S. A. 1969. *Perceptrons: An Introduction to Computational Geometry*. Cambridge, MA: The MIT Press.

Ngiam, J.; Koshla, A.; Kim, M.; Nam, J.; Lee, N.; and Ng, A. 2011. Multimodal Deep Learning. In *ICML'11 Proceedings of the 28th International Conference on Machine Learning*. 689–96. Bellevue, WA: Association for Computing Machinery.

Peterson, C., and Paley, D. A. 2011. Multivehicle Coordination in an Estimated Time-Varying Flowfield. Journal of Guidance, Control, and Dynamics 34(1): 177–91. doi.org/10.2514/1.50036.

Rajkumar, R.; De Niz, D.; and Klein, M. 2017. *Cyber-Physical Systems*. Boston, MA: Addison-Wesley Professional.

Scerri, S.; Debattista, J.; Attard, J.; and Rivera, I. 2015. A Semantic Infrastructure for Personalizable Context-Aware Environments. *AI Magazine* 36(2): 73–84. doi.org/10.1609/aimag.v36i2.2586.

Seetharaman, G.; Lakhotia, A.; and Blasch, E. 2006. Unmanned Vehicles Come of Age: The DARPA Grand Challenge. *IEEE Computer Society Magazine* 39(12): 26–9. doi.org/10.1109/MC.2006.447.

Shen, D.; Blasch, E.; Zulch, P.; Distasio, M.; Niu, R.; Lu, J; Wang, Z.; and Chen, G. 2018. A Joint Manifold Leaning-Based Framework for Heterogeneous Upstream Data Fusion. Journal of Algorithms and Computational Technology. doi.org/10.1177/1748301818791507.

Shishika, E., and Paley, D. A. 2017. Mosquito-Inspired Swarming for Decentralized Pursuit with Autonomous Vehicles. In *2017 American Controls Conference (ACC)*, 923–9.

Piscataway, NJ: Institute of Electrical and Electronics Engineers. doi.org/10.23919/ACC.2017.7963071.

Snidaro, L.; J. Garcia; J. Llinas; and E. Blasch, editors. 2016. *Context-Enhanced Information Fusion: Boosting Real-World Performance with Domain Knowledge*. Berlin: Springer. doi.org/10.1007/978-3-319-28971-7.

Urmson, C.; Baker, C.; Dolan, J.; Rybski, P.; Salesky, B.; Whittaker, W.; Ferguson, D.; and Darms, M. 2009. Autonomous Driving in Traffic: Boss and the Urban Challenge. *AI Magazine* 30(2): 17–28. doi.org/10.1609/aimag.v30i2.2238.

Vyas, N.; Farringdon, J.; Andre, D.; and Stivoric, J. 2012. Machine Learning and Sensor Fusion for Estimating Continuous Energy Expenditure. *AI Magazine* 33(2): 55–66. doi.org/10.1609/aimag.v33i2.2408.

Wu, R.; Liu, B.; Chen, Y.; Blasch, E.; Ling, H.; and Chen, G. 2016. A Container-Based Elastic Cloud Architecture for Pseudo Real-Time Exploitation of Wide Area Motion Imagery (WAMI) Stream. *Journal of Signal Processing Systems for Signal, Image, and Video Technology* 88(2): 219–31. doi.org/10.1007/s11265-016-1206-6.

Yang, C., and Blasch, E. 2008. Fusion of Tracks with Road Constraints. *Journal of Advances in Information Fusion* 3(1): 14–32.

Yang, C.; Kaplan, L. M.; and Blasch, E. 2012. Performance Measures of Covariance and Information Matrices in Resource Management for Target State Estimation. *IEEE Transactions on Aerospace and Electronic Systems* 48(3): 2594–613. doi.org/10.1109/TAES.2012.6237611.

Yang, C.; Kaplan, L. M.; Blasch, E.; and Bakich, M. 2013. Optimal Placement of Heterogeneous Sensors for Targets with Gaussian Priors. *IEEE Transactions on Aerospace and Electronic Systems* 49(3): 1637–53. doi.org/10.1109/TAES.2013.6558009.

Zheng, Y.; Blasch, E.; and Liu, Z. 2018. *Multispectral Image Fusion and Colorization*. Bellingham, WA: Society for Optics and Photonics Technology.

Zuo, H.; Fan, H.; Blasch, E.; and Ling, H. 2017. Combining Convolutional and Recurrent Neural Networks for Human Skin Detection. *IEEE Signal Processing Letters* 24(3): 289–93. doi.org/10.1109/LSP.2017.2654803.

Erik Blasch is a program officer at the US Air Force Office of Scientific Research. He received a BS in mechanical engineering from the Massachusetts Institute of Technology; master's degrees in mechanical, health science, and industrial engineering from the Georgia Institute of Technology; master's degrees in electronics, economics, and business from Wright State University; and a PhD in electrical engineering from Wright State University. He has compiled more than 800 papers, 27 patents, and 5 books focusing on robotics, information fusion, and man–machine systems, starting with the 1994 AAAI Robot Competition Winner (*AI Magazine*, 16(2) 1995). He is an International Society of Information Fusion member, an American Institute of Aeronautics and Astronautics Associate Fellow, an International Society for Photonics and Optics Fellow, and an Institute of Electrical and Electronics Engineers Fellow.

Robert Cruise is a research scientist at the US Naval Surface Warfare Center, Crane Division. He received his bachelor's degrees in physics and electrical engineering from Notre Dame, an MBA degree from the University of Chicago, and a PhD degree in theoretical nuclear physics from Indiana University with minors in applied mathematics and computer science. His research focuses on IF, CD, and the application of ML and control technologies to systems and robotics design. Previously, Cruise served as an engineer and scientist researching software-defined radios, ground-based operational surveillance systems, and fire-control systems at the US Naval Surface Warfare Center.

Alex J. Aved is a senior researcher with the US Air Force Research Laboratory, Information Directorate, Rome, New York. He received a BA degree in computer science and mathematics in 1999 from Anderson University, an MS degree in computer science from Ball State University, and a PhD degree in computer science in 2013 from the University of Central Florida, Orlando. His research interests include multimedia databases, stream processing (via CPU, GPU, or coprocessor), and dynamically executing models with feedback loops incorporating measurement and error data to improve the accuracy of the model.

Uttam K. Majumder is a senior electronics engineer at the US Air Force Research Laboratory. He received a BS degree in computer science from The City College of New York, graduating summa cum laude. He earned an MS degree from the US Air Force Institute of Technology, an MBA degree from Wright State University, and a PhD in electrical engineering from Purdue University. He was also an adjunct faculty member at Wright State University, teaching signal and image processing. He received an US Air Force Distinguished Civilian Award, a US Air Force Research Laboratory Science and Technology achievement award for radar systems development, and is a senior member of the Institute of Electrical and Electronics Engineers.

Todd V. Rovito is a research computer scientist with the US Air Force Research Laboratory, Sensors Directorate, Dayton, Ohio. He received a BS degree in information technology in 2001 from the University of Cincinnati, and a MS in computer science in 2005 from Wright State University. Rovito's research interest includes 3D reconstruction and reasoning applications for commercial space systems. He aggressively uses open source software to develop exploitation systems quickly, generate the best value, and deliver high performance products to the customers.