frontiers in
ROBOTICS AND AI

SPECIALTY GRAND CHALLENGE ARTICLE
published: 06 October 2014
doi: 10.3389/frobt.2014.00007

Machine perception and learning grand challenge: situational intelligence using cross-sensory fusion

Shashi Phoha*

The Pennsylvania State University, University Park, PA, USA
**Correspondence: sxp26@psu.edu*

Edited and reviewed by:
Liyi Dai, U.S. Army Research Office, USA

Keywords: sensor fusion, situational intelligence, context learning, context-aware pattern classification, action dynamics

Humans possess only five senses and very effectively coordinate their cross-sensory perceptions to situate themselves in uncertain operational environments for extracting context relevant actionable intelligence. Machines, on the other hand, may be embodied with a wider variety of electronic sensing devices but lack such situational intelligence in interpreting the sensed information. Despite significant advances in sensing technologies, machine perception remains primitive when compared to human perception. Lack of situational intelligence results in processing of large amounts of irrelevant information, leading to the often cited "curse of dimensionality" and computational explosions. These, in turn, limit the power of data-driven abstract reasoning and problem-solving algorithms, cause a lack of focus for drawing upon relevant past knowledge, and inhibit situational learning. As a consequence, autonomous systems cannot be trusted to adapt their behavior to unanticipated operational conditions. Current behavior-based modeling approaches address these issues by developing world models that modularly decompose the problem space. This requires a very detailed and somewhat complete understanding of the operational environments as a prerequisite. Yet, such models invariably prove inadequate for real world operations due to the rigidness of the decompositions. Autonomous system designs, therefore, are not robust and machine learning methods remain brittle.

Situationally aware sensor fusion and machine perception present a new frontier in machine automation, which holds the promise of unprecedented levels of autonomy in executing complex tasks in dynamic operational environments. The goal of such automation is to accomplish these tasks with the perception and adaptation of humans, and often in collaboration with humans. Several technological challenges must be addressed to further the state-of-the-art toward this goal.

CONTEXT LEARNING AND *IN SITU* DECISION ADAPTATION

Faced with the challenge of data to action in a complex noisy world, research methods have emerged in diverse fields, over the past decade, for machines to extract current operational context from sensor data. These include physics based environmentally adaptive sensing models, innovations in image and scene processing, natural language processing, ubiquitous computing, and cognitive neuroscience. Current state-of-the-art research in these areas attempts to extract operational context for a specific sensing modality, like visual or auditory context, which is not relevant to other modalities. The notion of context itself is often incoherent and ill-defined across sensing modalities and applications: image processing research generally assumes only the visual scene to be the context for object detection; for human–machine interactions, context is often the linguistic semantics in which humans express the current instruction to autonomous systems; for ubiquitous or mobile computing, it is the computing environment, and in cognitive sciences, context is often modeled via attention and memory. In a multi-sensor operational environment, involving both hard and soft sensing modalities, a broad unified notion of context is needed. This notion should characterize situations in the physical, electronic, or tactical environments that affect the acquisition and interpretation of heterogeneous sensor data for machine perception and adaptation to specified goals. Furthermore, it is often necessary to iteratively sense the context automatically and treat it as an implicit dynamic input to the application for robust context-aware operations.

In their 2013 paper, Blasch et al. (2013) survey recent research efforts to accommodate the effects of context in information fusion for target tracking applications. Several of these approaches attempt to mitigate the effects of context on the feature space by designing statistical detection and classification algorithms that are invariant to context changes. However, feature extraction techniques often do not adapt well to the highly non-linear and non-stationary effects of the operational environment. An alternate approach to improving detection performance is to exploit differences in sensor behaviors across environments and treat them as a supplemental source for context-dependent-learning. This approach was recently proposed in Frigui et al. (2010) for learning regions of similar responses for each sensor. Formalizing this approach, a mathematical characterization of machine extractable context, applicable to all sensing modalities relevant to an application, was recently presented in Phoha et al. (2014), with the objective of enabling contextual decision-making in dynamic data-driven classification systems. Both intrinsic context, i.e., factors, which directly affect sensor measurements, as well as extrinsic context, i.e., factors that do not affect sensor measurements directly, but affect the interpretation of observed data, were analytically formulated. This analytical foundation can be used to characterize and represent situational intelligence for multi-sensor multi-target applications. Further work in integrating data-driven and model based methods for context learning, discovery of new

contexts that were not labeled during the training phase, and dynamic modeling of context drift, remain promising research areas for improving machine perception and machine learning via situational intelligence.

CROSS-SENSORY FUSION

Today's intelligent machines operate on a sensing infrastructure for measurement, communication, and computation with which they perceive the evolution of physical dynamic processes in their operational environment. Sensors require physical interaction with sensed phenomena to generate time series of measurements (temperature, pixel intensity, etc.) of the evolutionary dynamics caused by physical stimuli. Thus, they are subject to a number of noise factors. The multivariate *Information Space* generated by these time series represents an amorphous computing environment of high dimensionality with much redundancy. Furthermore, sensors of different modalities are subject to contextually variable performance in noisy environments. To extract reliable information from multi-modal sensors, soft or hard, it is necessary to more fully exploit their heterogeneity by fusing complementary information across modalities. Extant information fusion literature exploits this heterogeneity very minimally. Usually, decision-level fusion algorithms fuse the probability distributions generated independently by each sensor into a single decision. For humans, this is equivalent to fusing the perceptions of a blind person and a deaf person, instead of coordinating visual and aural sense perceptions of one individual. Causal information regarding feature level dependencies is lost in this process. Machine perception methods are needed, which more fully exploit sensor dependencies at the feature level. Contextual complementarity of heterogeneous sensors, for example, can be used to overcome sensing inaccuracies or data incompleteness. Just as humans can aptly coordinate their visual and auditory information to disambiguate scenes or sounds, automated algorithms are needed that exploit cross-sensor dependencies. These algorithms will exploit sensor-specific non-linear and non-stationary effects such as phase transitions caused by physical stimuli. Addressing the scientific and engineering challenges of deriving actionable intelligence from multiple sources of electronic inputs, with differing modality and contexts, is particularly important for adapting the behaviors of physical systems. Situationally intelligent synthesis of autonomous heterogeneous sensors will enable more robust and accurate perception of a physical process than what is possible from traditional methods of fusing perceptions of independent single-sensors.

CHARACTERIZING ACTION DYNAMICS

Information in a signal is physically encoded as patterns of organization (Stonier, 1990). Patterns in raw data are noise perturbed manifestations of causal structural relationships – spatial, temporal, or informational. There is compelling evidence that both perception of action (particularly visualization of events) and action itself are composed of certain invariant primitives that are performed with a certain structure (Verfaillie and Daems, 2002). Machine perception of multi-object action dynamics is the next challenge in this area. Fundamental innovations in signal representation methods are needed to *discover* action primitives in data streams as mathematical objects, and their organizational structure formulated as a *generative grammar* (Jerne, 1993) to synthesize them into higher level concepts. Treating these primitives as *words* formed from a symbolic representation of the Information Space, and their organizational structure as the generating grammar for event synthesis, a complex multi-object interaction in the real world can itself be described as a trajectory of words in a structured formal language. The formal language will consist of all possible trajectories in the Information Space. In the existing scientific literature, such models are heuristically abstracted from human understanding, and are almost always inadequate machine representations of the causal dynamics that generate the system trajectories. General purpose approaches are needed for rigorous constructive methods to *discover* the formal computational language (words and grammar) embedded in the observed sensor data as the most likely scientific mechanism that would generate system trajectories that preserve the statistics of the observed dataset. Such a scientific mechanism endows a probabilistic computational language to autonomous systems to characterize and explore the causal structure of the vast Information Space. Such characterizations can be used to address today's fundamental limitations of machine perception, characterize multi-object interactions, enable cross-sensory disambiguation, and improve machine learning and contextual decision making. Mathematically rigorous concepts of *symbolization, syntactic abstraction,* and *similarity* are needed to construct the combined algebraic and topological structure of the Information Space. These fundamental advances will enable the modeling of causal dependencies between emerging primitives of action captured in data streams and development of algorithms for prediction and inference. Initial research in this direction is presented in Wen et al. (2013). Thus, a general probabilistic theory for characterizing the multi-modal information dynamics is needed to simultaneously support *in situ* data compression and situation awareness for the entire spectrum of information analyses from data collection to actionable intelligence.

OTHER TECHNOLOGY CHALLENGES

Another technology challenge that will accelerate and promote the progress in machine perception and cognition of sensed information is that of perceptual user interfaces that add human understandable rendering of complex data sources and facilitate human–machine interactions.

These and other innovations in machine perception are essential for harnessing the potential of a dynamic data-rich world through multi-sensor, multi-level, data-to-decision approaches. They will enable unprecedented levels of dependable autonomy for traditional applications such as surveillance, object classification, target tracking, pattern discovery, machine learning, and data mining. In addition, they will enable new developments in cyber physical systems that will improve our quality of life in fields such as remote health care, emergency response, traffic flow management, power generation and delivery, machinery condition monitoring and diagnostics, geo-spatial analyses, social networking, economy, and humanities.

ACKNOWLEDGMENTS

The work was supported in part by the Air Force Office of Scientific Research (AFOSR) under Grant No. FA9550-12-1-0270 and by the Office of Naval Research (ONR) under Grant No N00014-11-1-0893.

REFERENCES

Blasch, E., Herrero, J. G., Snidaro, L., Linas, J., Seetharaman, G., and Palaniappan, K. (2013). Overview of contextual tracking approaches in information fusion. *Proc. SPIE* 87470B, 1–11. doi:10.1117/12.2016312

Frigui, H., Zhang, L., and Gader, P. (2010). Context-dependent multisensor fusion and its application to land mine detection. *IEEE Trans. Geosci. Remote Sens.* 48, 2528–2543. doi:10.1109/TGRS.2009.2039936

Jerne, N. (1993). The generative grammar of the immune system. *Scand J Immunol.* 38(1):2–8. doi:10.1111/j.1365-3083.1993.tb01687.x

Phoha, S., Virani, N., Chattopadhyay, P., Sarkar, S., Smith, B., and Ray, A. (2014). Context-aware dynamic data-driven pattern classification*. *Proc Comput. Sci.* 29, 1324–1333. doi:10.1016/j.procs.2014.05.119

Stonier, T. (1990). *Information and the Internal Structure of the Universe.* (Chicago: Springer-Verlag).

Verfaillie, K., and Daems, A. (2002). Representing and anticipating human actions in vision. *Vis. Cogn.* 9, 217–232. doi:10.1080/13506280143000403

Wen, Y., Ray, A., and Phoha, S. (2013). Hilbert space formulation of symbolic systems for signal representation and analysis. *Signal Process.* 93, 2594–2611. doi:10.1016/j.sigpro.2013.02.002

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 29 August 2014; accepted: 16 September 2014; published online: 06 October 2014.

Citation: Phoha S (2014) Machine perception and learning grand challenge: situational intelligence using cross-sensory fusion. Front. Robot. AI ***1****:7. doi: 10.3389/frobt.2014.00007*

This article was submitted to Sensor Fusion and Machine Perception, a section of the journal Frontiers in Robotics and AI.