QuEST for Information Fusion in Multimedia Reports

Erik P. Blasch, Air Force Research Laboratory, Rome, NY, USA

Steven K. Rogers, Air Force Research Laboratory, Rome, NY, USA

Hillary Holloway, Systems &Technology Research, Woburn, MA, USA

Jorge Tierno, Barnstorm Research Corporation, Malden, MA, USA

Eric K. Jones, Systems &Technology Research, Woburn, MA, USA

Riad I. Hammoud, BAE Systems, Burlington, MA, USA

ABSTRACT

Qualia-based Exploitation of Sensing Technology (QuEST) is an approach to create a cognitive exoskeleton to improve human-machine decision quality. In this paper, the authors present QuEST-motivated man-machine information fusion system is presented for situation awareness. User-based situation awareness includes both elements of external sensory perception and internal cognitive explanation. The authors outline QuEST elements and are used as a reasoning approach to achieve human intelligence amplification (IA) in relation to data aggregation from machine artificial intelligence (AI). In a use case example for multimedia exploitation, the QuEST approach enhances enhanced understanding of the man (mind-body cognition) and the machine (sensor-based reasoning) by establishing a cohesive narrative of situational activities. QuEST tenets of structurally coherent, situated conceptualization, and simulated experience are utilized in organizing multimedia reports of Video Event Segmentation by Text (VEST).

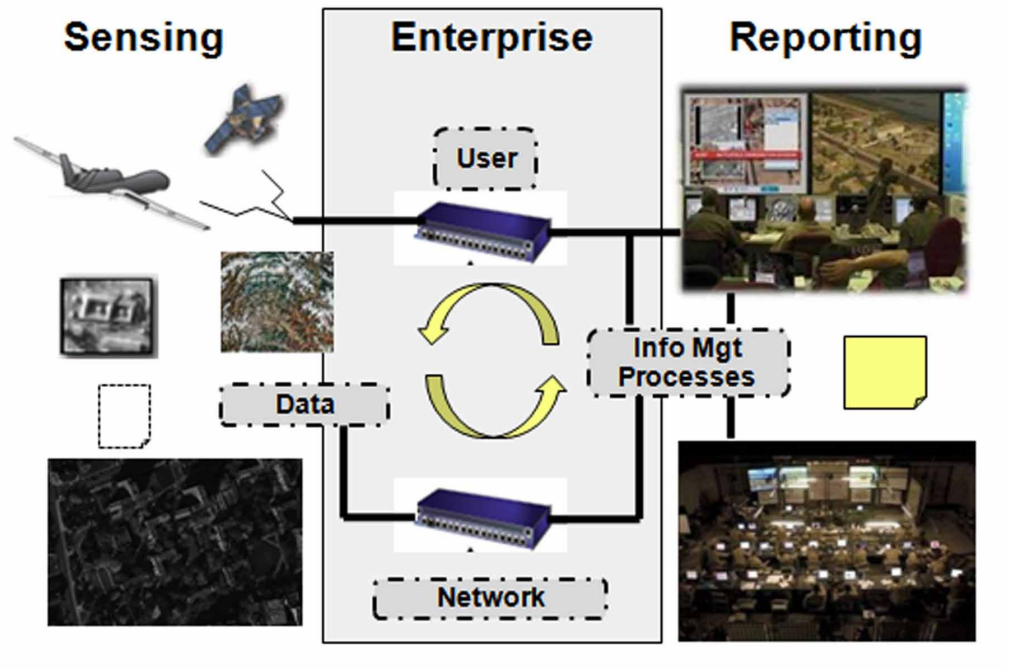
Keywords: *Automation and Autonomy, Cognition, Information Fusion, Intelligence Amplification, Qualia-based Exploitation of Sensing Technology (QuEST), Visualization*

1. INTRODUCTION

For many activities, a user-machine workflow is required for data analysis and mission effectiveness (Blasch, Bosse, Lambert, 2012). For example, integrated global Intelligence, Surveillance, and Reconnaissance (ISR) operations include a five-phase process: Planning and direction; Collection; Processing and Exploitation; Analysis and production; and Dissemination (PCPAD) (*Air Force Doctrine,* 2012). The PCPAD process is not linear or cyclical, but rather represents an multi-machine, multi-user enterprise of interrelated, simultaneous operations that can, at any given time, feed and be fed by other operations as illustrated in Figure

DOI: 10.4018/IJMSTR.2014070101

Figure 1. Information fusion in the enterprise



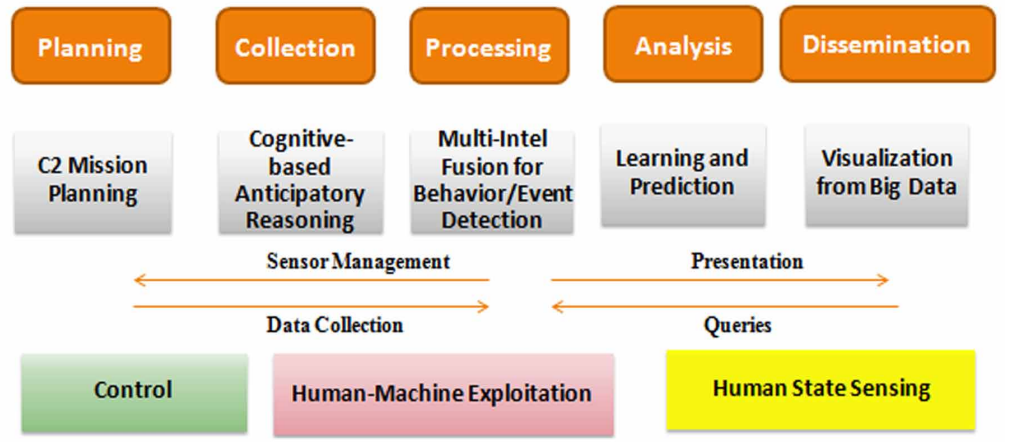
1. Key PCPAD process elements are machine tools to aid automation and decision making. For *processing*, it could be signals alignment, filtering and collection. For *exploitation,* it includes data correlation and association. For *dissemination*, visualization, interpretation and data reporting transfer knowledge. Exploitation and dissemination necessitate user refinement for situation analysis as well as data selection for reporting, picturing, and narration.

Qualia-based Exploitation of Sensing Technology (*QuEST*) is focused on the blending of traditional experience based interpretation with an artificially 'conscious' interpretation of the signals, data, and information. A driving motivation for QuEST is the theory that conscious has a role in robust decision making such as supporting a coherent narrative from sparse measurements.

In designing systems to augment user needs, it is desirable to provide intelligence amplification (IA) (Rogers, *et al*., 2003). *Qualia* provide a vocabulary for subjective analysis of stimuli. Qualia are the internal conscious perceptions of the basis set used to represent the stimuli and is a subjective aspect of the human's conscious perception of the stimuli. Qualia allow an agent to understand/evaluate relevant data in decision making. The more that a sensor/user understands and evaluates their Qualia (Rogers, *et al.,* 2008), the more self-confident they would be in decision making. Qualia encompass an important component to uncertainty reasoning associated with subjective beliefs, trust, and narratives in decision making. This includes the conscious representation of the subconscious processing and thus represents a sense of intuition. The goal of the user-machine processing is to select relevant data in forming a cohesive narrative that explains the *situation.*

Situation analysis includes three domains: (1) human factors for situation *awareness* (Endsley, 1995A), (2) information fusion for situation *assessment* (Blasch, *et al*., 2006), and (3) psychology for situation *representation* (Patterson, *et al.*, 2012). While all three domains

Figure 2. Planning to dissemination



have focused on situation understanding, the QuEST paradigm seeks to coordinate user cognitive processes with that of the sensed world as augmented by machines for situation narratives.

The PCPAD process highlights human-machine exploitation, shown in Figure 2 (Rogers, *et al*., 2014). *Sensing*-to-processing includes command and control (C2) and collection which incorporate cognitive determination for machine control. Processing-to-dissemination includes exploitation (or analysis) that focuses on cognitive interpretation of the data available, directly or indirectly, to reason about the *situation*. QuEST seeks to enhance processing with human-based sensing.

To afford an analyst access to large amounts of data, analytics, and federated access requires an enterprise information management and fusion process model. QuEST provides an approach to representation that suggests that one of the key functional purposes of consciousness is to provide a common framework for all sensing modalities for fusion. For example, aubible text experienced as qualia as elements from the visual channel are elements of internal thought. Qualia thus facilitate a fused cohesive interactive representation.

Key elements of user interaction with sensed data require:

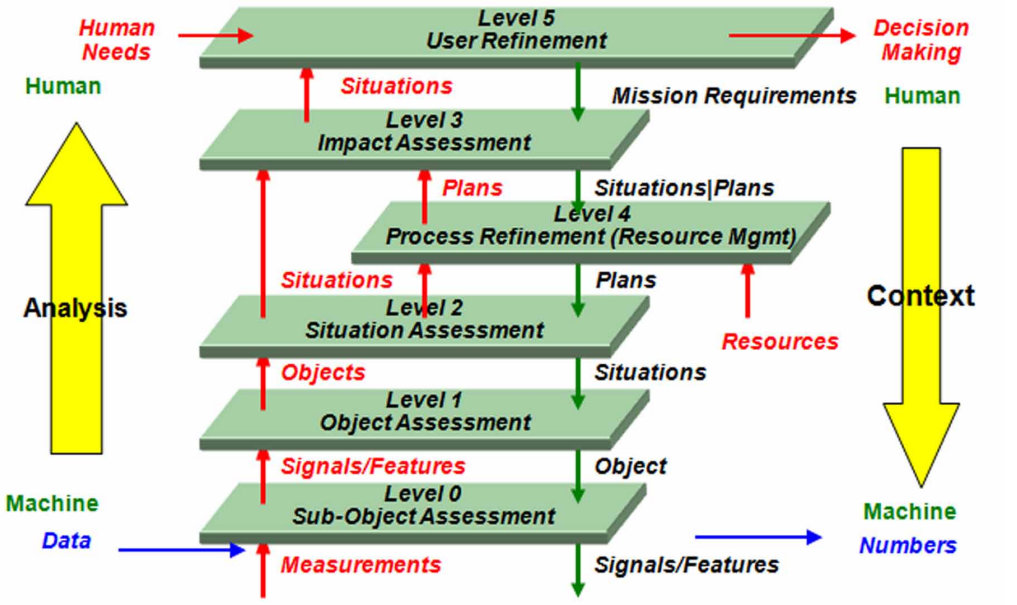
- Common framework: to ingest data, algorithms, and processing techniques;
- Complementary approaches: methods performing similar functions (e.g., image exploitation); and
- Interaction: allow machines and users to work with the various tools for analysis and subsequent information dissemination.

1.1. Information Fusion Based Situation Understanding

To combine user and machines for situation understanding, we use the information fusion paradigm as shown in Figure 3 to highlight QuEST motivations.

Figure 3 shows Level 1/2/3 (object/situation/impact) assessment and Level 4/5/6 (resource/user/mission) refinement. QuEST focuses on *sensing* (Level 0, 1 fusion) as well *exploitation* (Level 2, 4 fusion) from the machine *technology*. Qualia-based reasoning provides additional capabilities to the machine through the human. Using the information fusion levels; we contend that Level 5 (L5) fusion termed "*user refinement*" can be enhanced with QuEST which includes analysis relevant to the situation and mission context (L6). The QuEST paradigm implements artificial agents using a dual process model to afford 'wingman' solu-

Figure 3. Information fusion levels



tions to facilitate an alignment between humans and the decision aides.

1.2. QuEST-Based Situation Understanding

QuEST is focused on cohesive situation understanding:

A situation is any part of the subjective internal representation of an agent which can be comprehended as a whole by that agent through defining how it interacts with or is related to other parts of the representation in that agent. (Culbertson, et al., 2012)

QuEST for information fusion focuses on L5 fusion. L5 fusion includes operator collaboration with the machine (Blasch, Hanselman, 2000), situation awareness/assessment displays (Blasch, 2000), and trust (Blasch, *et al*., 2014). QuEST also requires human state sensing to afford the fusion between the computer wingman and human agents. QuEST however extends the notion of L5 fusion in that the traditional L5 fusion is based on the reasoning with the data collected from the enterprise, whereas QuEST includes the internal reasoning over the situation that is not always directly associated with the data available such as cognition of imagined, intuitive, and plausible representations of the situation using both the human and machine agents.

An emerging example includes *unstructured* text processing of a human that interprets information from documents as they form a narrative. For example, with data available on the web (e.g., twitter, documents), intelligent users need the capability to rapidly monitor and analyze event information over massive amounts of unstructured textual data (Panasyuk, *et al*., 2013). Text, from other human sources is subject to opinions, beliefs, and perceptions from the reader that interprets the information to form their own narrative (Fenstermacher, 2014).

Computer sensed data is stochastic or deterministic from which we have to coordinate the agent information (Greene, *et al*., 2005). For example, with Gaussian observations, it is a stochastic probability analysis (e.g., Kalman

Filter); however, there is much *structural* information in the sensor models and sensitivities for a given state condition (which is a deterministic ontology) which could be used to improve the estimate of the situation such as multiple target tracking (Yang, *et al*., 2005; Rodriguez, *et al*., 2013). Thus, there is always a case for a combination of stochastic and deterministic decisions to deal with uncertainty elements in all modeling and system deployment.

A goal for QuEST is mission-responsive enterprise resource management that incorporates technology (sensing), human conscious thought (qualia), and combined user-machine interaction (exploitation) for robust decision making.

The rest of this paper includes Section 2 as an overview of information fusion including activity-based intelligence, automation, and autonomy. Section 3 discusses qualia. Section 4 highlights QuEST processes and tenets where Section 5 details cognitive models important to QuEST. Section 6 provides a QuEST model for information fusion. Section 7 provides a narrative analysis for video and text fusion with results in Section 8. Finally, Section 9 draws conclusions.

2. INFORMATION FUSION

Recent information fusion techniques include big data analytics (Blasch, Steinberg, *et al*., 2013). Multimedia analytics includes multi-intelligence fusion from which we seek sensor planning (DiBona, *et al*., 2006), resource management for situation assessment (Blasch, *et al*., 2008), geo-intelligence (Blasch, Deignan, *et al*., 2011), and analyst support (Blasch, Lambert, *et al*., 2012). With a user work-domain environment, a QuEST application is situational activity-based intelligence (ABI) (Blasch, Banas, *et al*., 2012). QuEST views humans using 'qualia' as the vocabulary for conscious thought, that could be an 'activity' or an 'object,' that connects ABI and reporting. Qualia provide a hypothetical representation to allow an agent to do 'prescriptive' analytics (causality versus correlation) on big data, thus generating a script for interactions to confirm or refute hypotheses between objects and activities.

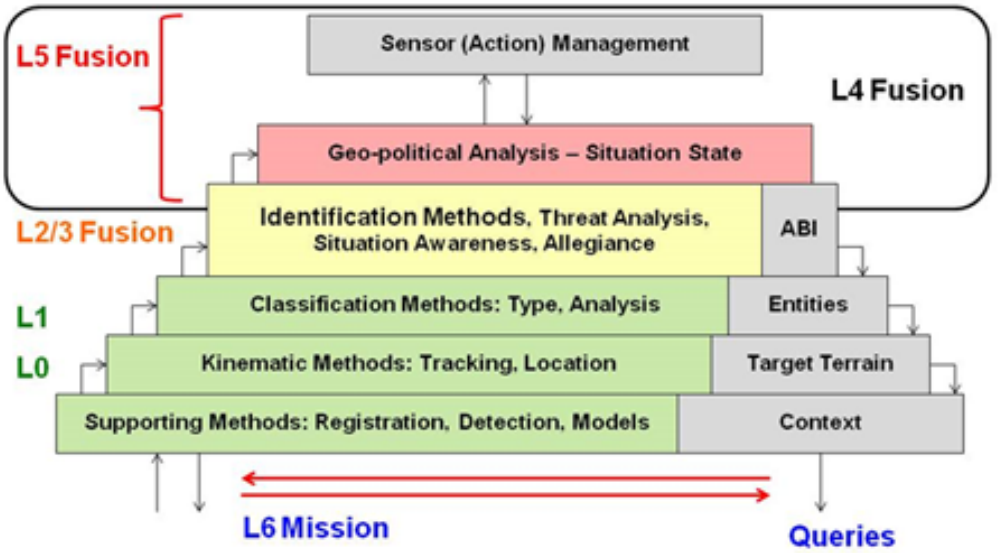
2.1. Event-Based Intelligence

To access dynamic data (data in motion), there is a need for modeling, scalable information architectures, and tools for pattern analysis (Blasch, Russell, *et al*., 2011). Technology should augment a human analyst work domain objectives as shown in Figure 4. Low-level information fusion (LLIF) of Level 0/1 object assessment tracking and classification (Blasch, Wang, *et al*., 2013; Hammoud, *et al*., 2014) can be determined from multimedia data exploitation.

Over the past decade, there have been many efforts to determine human activity from video. However, what is developed is machine extraction of video content such as events and activities semantically described, termed "labeling". In this paper, we seek to extend these concepts with user augmentation for which the content is described in text. One example is user analysis of video for semantic-based indexing of activities and truthing a video data set. To survey the entire community would be incomplete, so we present some of the more promising results that provided a survey on the need for user-defined semantic labels from audio and text fused with video content for video segmentation. Table 1 presents the results showing a progression of the video activity recognition community to extract, label, and multi-modal fusion activity analysis.

High-level information fusion (HLIF) includes analysis beyond Level 1 fusion and requires user-driven coordination. Many challenges exist for HLIF such as semantic analysis, evaluation, and systems design (Blasch, Bosse, Lambert, 2012). For example, effective coordination of estimation and management functions for sensing requires assessment based on mission needs (Blasch, Nagy, *et al*., 2014). Other challenges include ontologies supporting web services (Czajkowski, *et al*., 2004), machine translation (Czajkowski, *et al*., 2006), and evaluation (Costa, *et al*., 2012). These mission needs are sent to analysts who must *forage* for

Figure 4. Information fusion for activity-based analysis



data to answer queries, information needs, and mission perspectives. The analysis must perform *sensemaking* over the foraged data which requires pragmatic interfaces, visualizations, and analytics for users.

The analyst must observe emerging situation developments from observations. A combination of visualization techniques serves to coordinate the work flow between the machine and user. Human state sensing, not just visualization, needs bidirectional information flow between humans and computer agents. Using the power of the machine for autonomy and improving user interaction through automation is a current challenge.

2.2. Automation and Autonomy

Typical multimedia processing techniques assume limited human involvement; however most technology employs a combination of autonomy and automation:

- **Automation:** The system functions with no/little human operator involvement, however the system performance is limited to the specific actions it has been designed to do. Typically these are well-defined tasks that have predetermined responses, i.e. rule-based responses. (Note: we rather suggest Data Driven AI) (*Department of Defense Science Board*, 2012).
- **Autonomy:** Systems which have a set of intelligence-based capabilities that allow it to respond to situations that were not pre-programmed or anticipated in the design (i.e., decision-based responses). Autonomous systems have a degree of self-government, self-directed behavior (with the human's proxy for decisions) (*Department of Defense Science Board*, 2012).

For *automation*, we leverage HLIF methods such as users semantically describing time-stamps of situation-based event relationships. Event-recognition is dependent on analyst involvement as the enormous amounts of data are not exhaustibly defined by a machine (Hammoud, Sahin, *et al*., 2014). *Autonomy* includes control functions such as data-base management, storage, and indexing. The use of

*Table 1. Key techniques in human activity recognition using Video, Text, and Audio (VTA)**

	Video		Text			Fusion
	Tracking	**Labels**	**Extraction**	**Relation**	**Audio**	**V-T-A**
Hoogs, *et al*., 2001	√	√				
Hoogs, *et al*., 2003	√	√				
Denis, *et al*., 2003	√	√				
Chan, *et al*., 2003	√	√	√			
Basharat, *et al*., 2008	√	√				
Swears, *et al*., 2008	√	√				
Turaga, *et al*., 2008 (Survey)	√	√				
Reddy, *et al*., 2011	√	√				
Oh, *et al*., 2011		√	√			
Vondrick, *et al*., 2012		√	√			
Yuan, *et al*., 2013	√	√	√			
Tsitsoulis *et al*., 2013 (Survey)	√	√				
Wu, *et al*., 2005			√	□		
Zhang, *et al*., 2012	√	√		√		
Graham, *et al*., 2011		√		√		√*
Antony, *et al*., 2013		√		√		√*
Sandell, *et al*., 2013		√		√		√*
Fisher, *et al*., 2012	√	√		√		√*
Blasch, *et al*., 2013	√	√	√		√	√
Chellappa, *et al*., 2014 (Survey)	√	√	√	√	√	
Hammoud, *et al*., 2014	√	√	√	√	√	√

Notation: √ - Yes, □ – used for text-activity only in this study, √* - fusion with simulation

* Note: we did not include video and audio analysis as the assumption in the paper is that the audio is converted to text and is asynchronous with video collection.

automation/autonomy affords decisions to data (Blasch, 2014) over real world sensor, environment, and target (SET) operating conditions (Kahler, *et al*., 2008).

The elements of humans and machines requires integrated approaches that links situation awareness (human) with situation assessment (machine) for situation understanding as shown in Figure 5. Using the perception of multimedia information, both the human and machine can plan for coverage gaps of the surveillance needs using automation/autonomy.

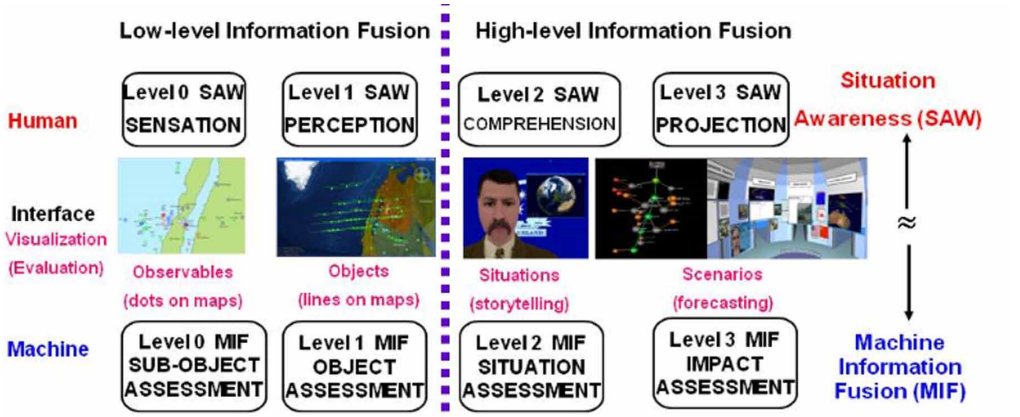
To bridge the human-machine interaction; elements of cues, context, and channels support the common representation of information as shown in Figure 6 (Blasch, 2013A).

To go from sensing to dissemination from cues, context, and channels; we highlight QuEST attributes by assessing qualia.

3. QUALIA

'Qualia' is a philosophical term referring to individual instances of a subjective conscious experience that refers to "what it is like" for a sensation (Cowell, 2001). Sensations include audio, visual, proprioception, smell, and taste.

Figure 5. Human-machine interaction through situations



QuEST also includes emotion and internal thoughts – anything you can introspect over. Combinations of sensations such as heat and sight can infer meaning such as 'the what' and 'the why' of pain. Qualia emanates from the mind-body problem (Harman, 1993). A common example is that perceived color is a derived sense as different individuals could report different colors for the same object.

If qualia were universal, assuming all people report the same things for the same stimuli, then a narrative of the color of a moving object as described by one person would be understood by another. QuEST suggests that although our individual qualia could be different as per the fact that we use similar processes to generate them, we can align our articulation of the experiences. Thus, we can have an understanding without requiring the experiences (i.e., qualia) be identical for a given stimulus. For example, a narrative includes knowledge of experiences by the person describing the color that affects the description of the object. To understand the role of qualia in exploitation, we need to look at some fundamental properties of qualia.

Qualia have been discussed in terms of codes, cues, clues, and affordances. Three laws of qualia are based in the physiological and cognitive perceptions from sensation to action that include (Ramachandran, Hirstein, 1997):

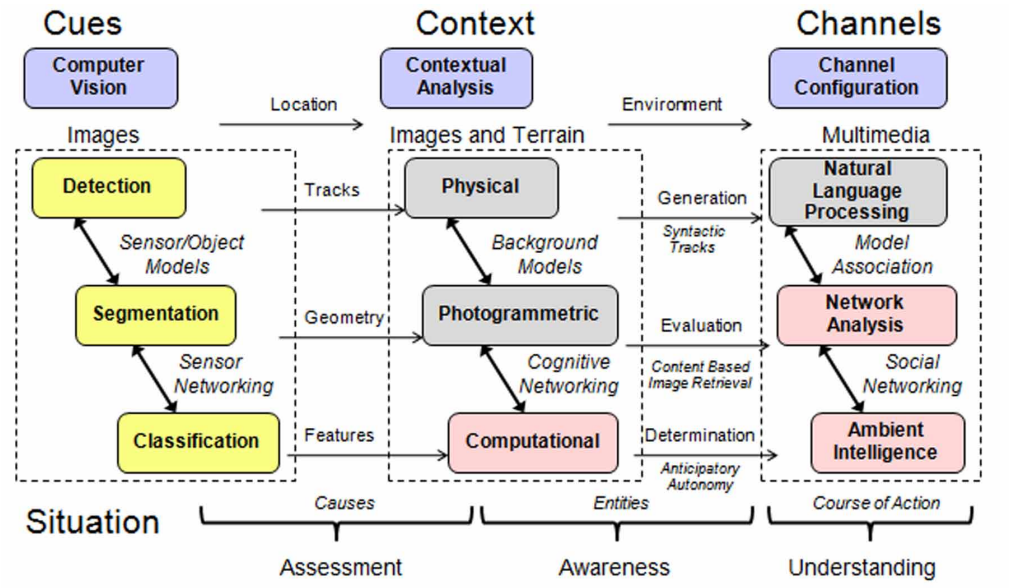
- Qualia are irrevocable and indubitable,
- Once the representation is created, what can be done with it is open-ended.
- The input invariably creates a representation that persists in short-term memory—long enough to allow time for choice of output.

Qualia have association with awareness including aspects of emotion (experiences), unity (cohesion), body (context), convictions (beliefs), and free-will (decision making). Once a sensation is created, it cannot be recalled; however the interpretation persisting in conscious thought affords additional understanding as related to the cerebral processes. Fundamental to the qualia laws is that *attention*, as user-based sensation, initiates interpretations.

Dennett (1988) associates four properties to qualia:

- Ineffable; only communicated through direct experience.
- Intrinsic; independent to other experiences.
- Private; known to self without interpersonal comparisons.
- Directly or immediately apprehensible in consciousness; a complete quale experience is known instantly.

Figure 6. Situation modeling from cues, context, and channels



Other philosophers have pondered over the meaning of qualia as a discussion of concepts not easily understood or incorporated into human decision-making models. The metaphysical qualia concept is not correlated with brain states, processing, and representations; but the non-physical elements of qualia are important for user-based exploitation and reporting as evidenced in natural language narratives. QuEST takes the position that brain states are the source of qualia, but we can never fully understand qualia from neural correlates, which requires a theoretical model to understand the brain states.

4. QUEST

QuEST seeks a new set of processes that will be implemented in a computer agent (or set of computer agents) to improve decision quality of a human agent (or set of human agents).

- **Assumption 1:** Fundamental units of conscious cognition are situations.
- **Assumption 2:** Decision quality is dominated by the appropriate level of situational awareness (situation awareness is the perception of environmental elements with respect to time and/or space and/or logical connection, the comprehension of their meaning, and the projection of their status after some variable has changed, such as time.)
- **Assumption 3:** Cohesive narratives are the reported products of information fusion systems.

4.1. QuEST Processes

QuEST processes could be considered a new approach to situational assessment (processes that are used to achieve situational awareness) and situation understanding (comprehension of the meaning of the information as integrated with each other and in terms of the individual's goals). It is the "*so what*" of the data, or sensemaking ('a motivated, continuous effort to understand connections which can be among people, places, and events in order to anticipate their trajectories and act effectively,' Klein, *et al*., 2006) for decision quality.

Box 1.
Three processes defined in QuEST:
• *QuEST agent processes* implement blended dual process cognitive models (have both artificial conscious and artificial subconscious/intuition processes) for situational assessment
• *QuEST artificially conscious processes* all are constrained by the fundamental laws of the QUEST Theory of Consciousness (structural coherence, situation based, simulation / cognitively decoupled).
• *QuEST subconscious/intuition processes* do not use working memory and thus considered autonomous (do not require consciousness to act). Current approaches to data-driven artificial intelligence provide a wide range of options for implementing instantiations of capturing experiential knowledge used by these processes.

QuEST is developing a 'Theory of Knowledge' – to provide the foundations to understand what an agent or group of agents can know which fundamentally changes machine learning and human-computer decision making from an empirical effort to a scientific effort.

4.2. QuEST Tenets

As per the laws of qualia, QuEST seeks principles on which a belief or theory is based.

Three related ideas from the tenets include compression, exformation, and events. *Compression* encodes an infinite number of stimuli into a single quale (e.g., low bandwidth 50 bits/sec) for interaction. *Events* in situations may be communicated to another agent as an event potential akin to an evoked potential (i.e., brain response to a cognitive stimulus). Finally, *exformation* (pattern completion inferring mechanism) affords a conscious representation.

The conscious representation is blended with data-driven processing, exploitation and dissemination. The deliberation using that representation complements the conventional data-based representation with the ability to incorporate context. Context can either be stored or inferred by situating the hypothetical representation via simulation to generate a cohesive narrative as a meaning of the sensed data.

Together, the situated coherent experience extends processing, exploitation, and dissemination of information. For processing, it is the formation of structured and coherent understanding of collected data. For exploitation, it is the conceptualization of the situation. Finally, for dissemination, the reported results are appended by experiences with the pre-experienced or by imagined interpretations.

QuEST is exploring other modeling approaches for sensing-based situation reasoning, exploitation-based decision making, and technology-based information reporting.

5. QUEST MODELING

We review three cognitive modeling approaches that influence the QuEST modeling including the dual-process model, the ACT-R model, and the situation awareness model.

5.1. Dual-Process Model

The Dual Process (DP) model includes System 1 and System 2 attributes of cognition (Chaiken, Trope, 1999; Smith, DeCoster, 2000). The dual-processes are presented in Figure 7, which include pattern recognition and consciousness in working memory going from the stimulus (sensation) to the response (exploitation).

Patterson (Patterson, *et al*., 2010) has used the DP model for natural decision making that links implicit/tacit (System 1) with explicit

Box 2.

The three QuEST tenets are:

- *Structurally Coherent* - the conscious representation has to have enough mutual information with physical reality to facilitate **interaction** with the world in a stable, consistent and useful manner (e.g., learned predictable explanations, links, and outcomes).
- *Situated Conceptualization* - the fundamental units of conscious deliberation are **situation entities** (e.g., context-based gists, time/space/multi-modality representations, and plausible narratives).
- *Cognitively Decoupled* - the conscious representation is a hypothetical explanation of the present, past or imagined future, it is a simulation which is not a posting of sensor data (e.g., **exformation** and conceptual combination to generate a new meaning).

(System 2) analysis. Information fusion by a machine has also related explicit (LLIF) with that tacit (HLIF) data processing. Clearly, tacit situation assessment/awareness is affected by the user experiences, knowledge, and attention for situation interpretation. Furthermore, as per QuEST, the PCPAD cycle can be addressed by simulated (real or imagined) interpretations of the situation (Patterson, *et al*., 2013). A summary of the dual processing models are shown in Table 2 (Evans, Stanovich, 2013).

5.2. ACT-R

The Adaptive Control of Thought—Rational (ACT-R) model is a cognitive architecture for simulating and understanding human cognition. The ACT-R focus is on how people organize knowledge and produce intelligent behavior including perception, thoughts, and actions on the world. The QUEST team has used ACT-R as a basis for Qualia analysis (Vaughan, *et al*., 2014) including robustness in decision making (Walsh, 2013; Walsh, *et al.*, 2013). Building

Figure 7. Dual processing model

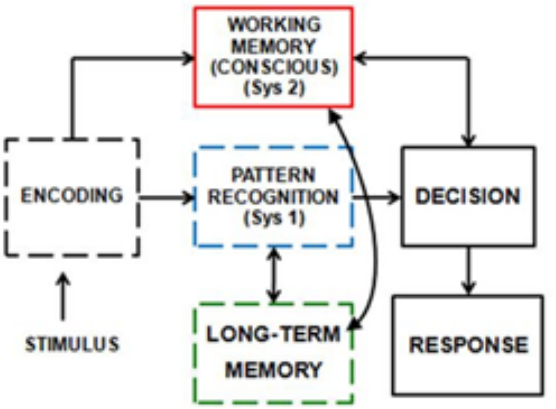


Table 2. Summary dual processing modes

Intuitive Processing	Reflective Processing
SYS 1	SYS 2
Autonomous (does not require working memory)	Cognitive (requires working memory)
Fast, parallel, high-capacity	Slow, serial, capacity-limited
Associative, contextualized, experienced decision making	Rule-based, abstract, consequential decision making
Animal-based, basic emotions	Human-based, complex emotions
Implicit knowledge	Explicit knowledge
Non-conscious	Conscious

It is noted that these processing mode descriptions follow the themes of autonomous, associative, and cognitive phases of learning (Fitts, Posner, 1962) and skills, rules, and knowledge in decision making (Rasmussen, 1983). Between sensation (autonomous or skills) and exploitation (cognition and knowledge), associative and rule-based processing is done.

on (Kennedy, Patterson, 2013), four levels of the cognitive architecture shown in Figure 8 include:

Cognitive framework Layer [L1] (autonomous mind, algorithmic mind, reflexive mind) where the reflective mind includes beliefs, goals, knowledge to produce narratives,
Qualia Processing Layer [L2] seeks to resolve the relationship between external stimuli and properties representing internally generated (evoked) qualia for a modally cohesive narrative,
Technical Implementation Layer [L3] is the generation, computational, retrieval and processing of information, and
External Interface Layer [L4] includes the semantic, visualization, and methods of user interaction with machines.

Three agents (minds) include:

1. The autonomous mind – reactive to stimulus
2. The algorithmic mind – strategies for control
3. The reflexive mind – rational deliberative processing

Building on the dual process model and ACT-R, we revisit human decision making for situation awareness and information fusion.

5.3. Situation Awareness (SAW) Model

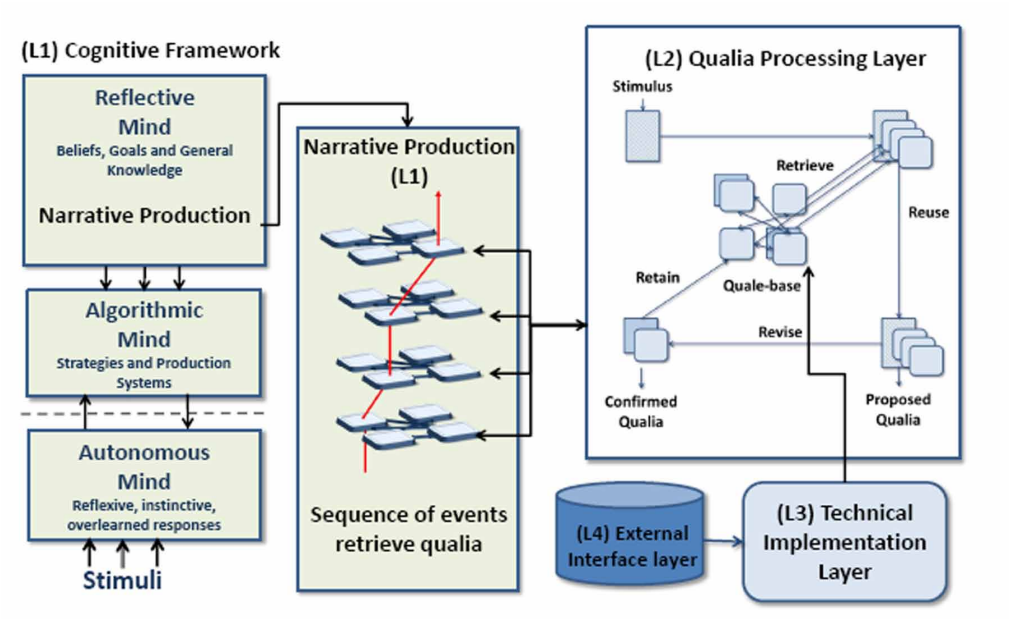
The Situation Awareness (SAW) model (Endsley, 1995A) includes perception, comprehension, and projection with extensions for workload, working memory, and attention (see Figure 9).

The perception layer is related to the Sys 2 cognitive processing or working memory. The QuEST 'qualia-based' (introspectively available) step of the ACT-R could incorporate user experiences for information processing. As the SAW model is descriptive, many overlaps exist. For example, there is no perception of experiences out of context – since all qualia are situated. Comprehension and projection could include data recollection, inference information, cultural models (Blasch, Salerno, *et al*., 2013), imagined futures, as well as course of actions.

5.4. Information Fusion Model

The Data Fusion Information Group (DFIG) model (*Blasch et al.,* 2005) relates to the DP model for user refinement, the CogArch model for man-machine interactions, and the SAW model for comprehension and projection. A

Figure 8. Cognitive Architecture (CogArch) model



DFIG element, Figure 10, is situation/threat understanding knowledge of us and them, as others have goals/beliefs that require qualia to determine the social narrative.

6. QUEST FOR INFORMATION FUSION

Using the tenets of QuEST and related cognitive models, we seek to bring together these ideas in a QuEST analysis.

6.1. Analyst-based QuEST Model

The QuEST discussion seeks to differentiate a QuEST agent versus an atomic agent. The atomic agent consists of a representor (mapping between sensor and representation) and an exploiter (mapping between representation and stimuli). An atomic agent has single process while the QuEST agent has dual processes (see Table 2). Two questions are: what is quality of data and what is the value of information (Blasch, Valin, *et al*., 2010)?

Qualia characterize the data with knowledge to provide conscious meaning of the data *context*. Context aids in the encoding and representation of the data. Using the representation, exploitation can be done for decision making which is then organized for reporting as new stimuli. Using the context from qualia helps to determine the relevancy (e.g., value) of information along with some situation understanding of the data quality. Figure 11 shows the integration of the various models as cognitive agents (information fusion nodes) to access sensory data, provide that to Sys1 and Sys2 processing, and output narratives. The key attribute here is that QuEST identifies the imagined present to complement Sys1 representations and imaged futures for Sys2 processing. Fusing imagined present/futures results in better decision making as planned narratives for reporting.

A QuEST example is video, text, and mission analytics diagnosed by a machine and a human. The real-world is sensed as *data* from humans and machines. QuEST *cognitive/computational agents* provide the central analytics between data and Sys1/Sys2 human reasoning.

Figure 9. Situation Awareness (SAW) model

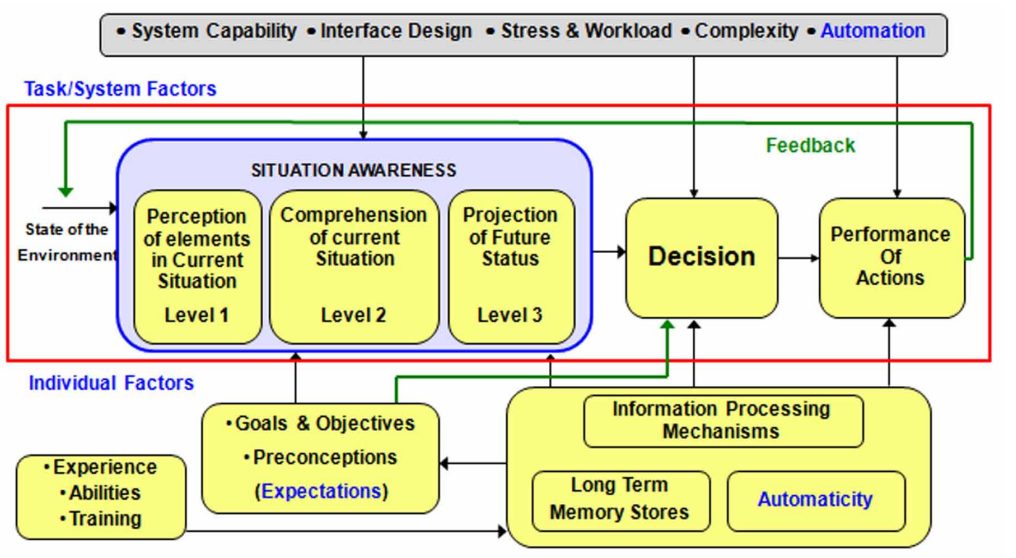


Figure 10. Data fusion information group model

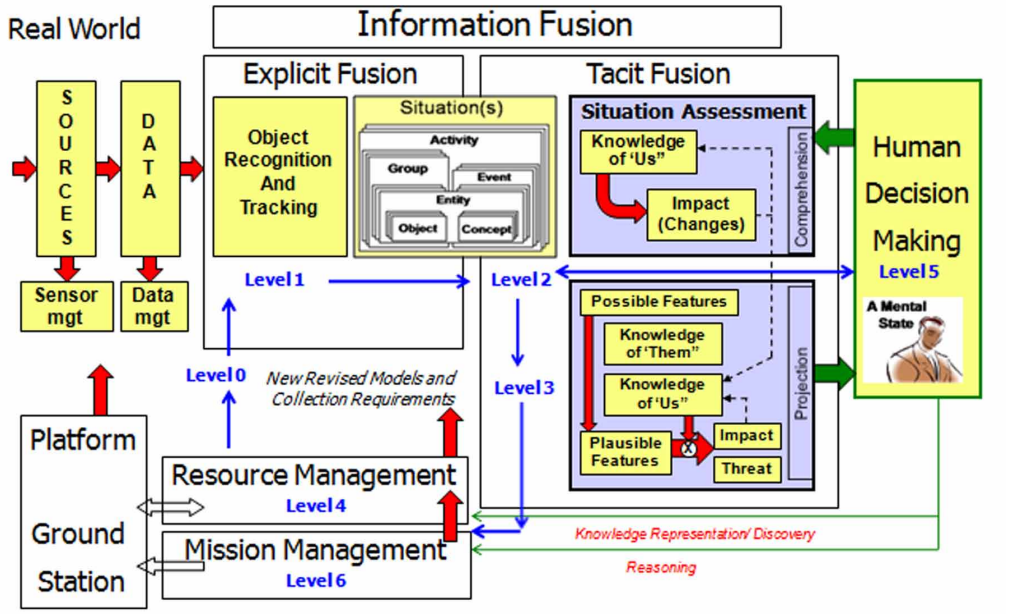
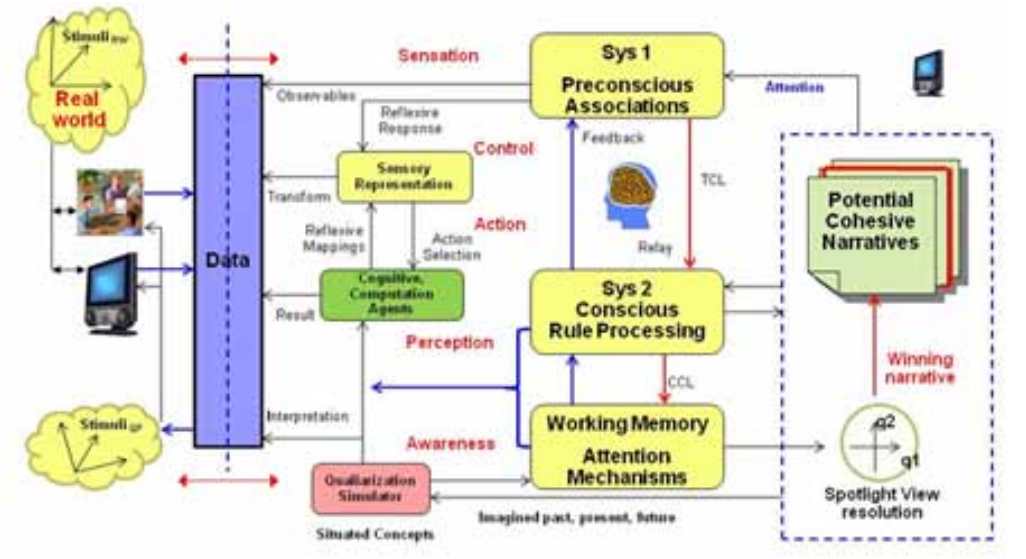


Figure 11. QuEST model for information fusion



Note that information fusion has typically only addressed computational agents (Figure 11 top) for sensation/control/action. However, QuEST agents add perception/awareness from Sys2 and working memory. Figure 11 (bottom) highlights the situated concepts from QuEST (from imagined past/present/futures) as narratives that conscious thought attends to the sensed information. QuEST acknowledges that awareness is not decoupled from the narrative; whereas information fusion assumes that the user attends only to the display perceptions for situation assessment.

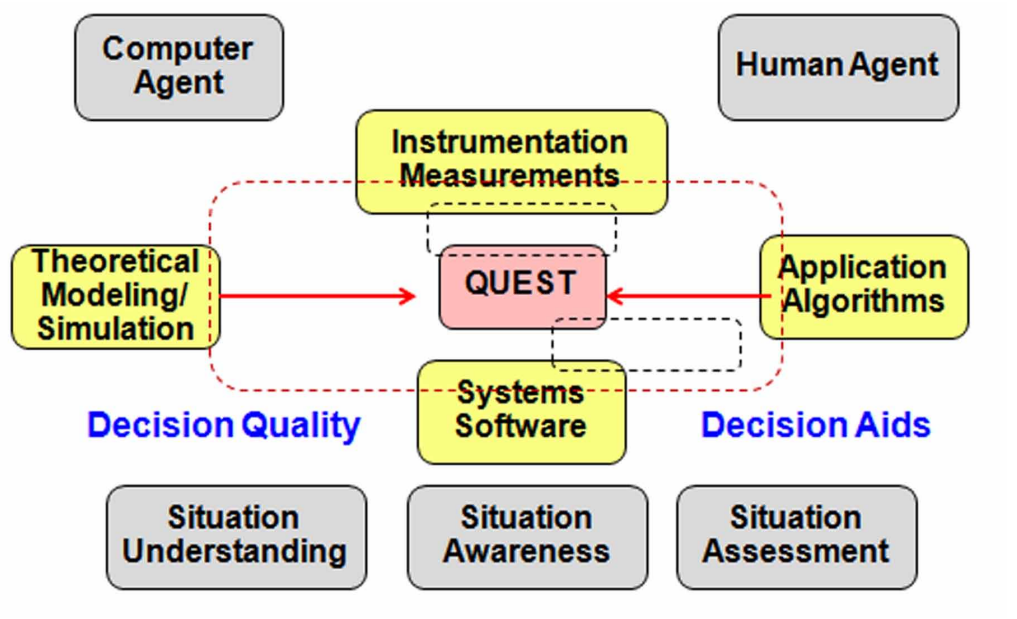
6.2. Machine-based QuEST Agents

Knowledge representation relates to natural language processing such as semantics in decision making. QuEST agents will work with machines for representation and data validity that enables the automatic reporting of cohesive narratives. Key enablers include cloud technologies (Liu, *et al*., 2014) for future information fusion developments as well as User-Defined Operating Pictures (UDOP) (Blasch, 2013B) for visualization. One method that brings together big data, sensing and exploitation is the Dynamic Data-Driven Applications Systems (DDDAS) paradigm (Blasch, Seetharaman, *et al*., 2013). As shown in Figure 12, there are human, computer, and QuEST agents that support situation analysis (understanding, awareness, and assessment). Of the four DDDAS methods, (theoretical modeling, engineering measurements, processing algorithms, and systems software), QuEST overlaps with measurements *sensing* and software applications for *exploitation*. Enhancing decision aids would lead to improved decision quality as the user provides introspective measurements.

7. QUEST ANALYSIS

QuEST improves human-machine decision quality by assisting the analysis with event-based processing for segmenting key activities. Situation awareness (Endsley, 1988) includes perception, comprehension, and projection. We focus on sensor, user, and mission (SUM) data management (Blasch, 2006). The goal is utilize methods of evaluation from both the user (Fitts, 1954) as well as the multimdeia data (Nghiem, *et al*., 2007). The novel idea is to utilize the user

Figure 12. Big data interactions with the QuEST model



semantics in conjunction with the video data (Ahanger, Little, 2001) and the scene context (Yang, *et al*., 2009) to segment the video stream into clips supporting a narrative.

We designed a segmentation capability to determine how messages and video data are grouped into "events" for gestalt-based presentation (i.e., proximally grouping relevant information). The grouping is derived from video segment boundaries, which are defined as the endpoints of a time interval, where an activity is considered to have occurred within and across the time interval.

The QuEST algorithm proceeds in three steps and follows the work of (Cooper, *et al.,* 2005), which includes computing similarity matrices for the streaming, deriving a novelty function for each matrix, and detecting event edges by finding the maxima of a combined novelty function.

7.1. Video Similarity Function

For a pair of time indices, t_1 and t_2, we measure how similar the streaming data at t_1 is to the data at t_2. We expect the similarity function to have high values when t_1 and t_2 are both within a homogeneous segment of the video, and low values otherwise.

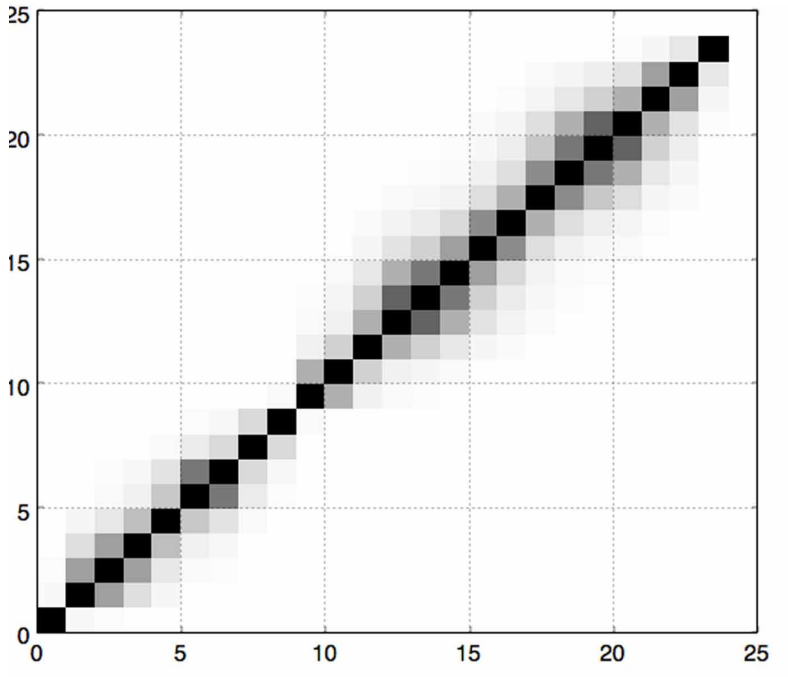
Consider a buffer of length l of a given data stream. Entries in this buffer are pairs (t_i, v_i) where t_i is the i^{th} time index and v_i is the corresponding value of the data stream. We order the events in each data stream according to their time indexes.

For live message events, the value of the event will be equal for any value of time index. The similarity matrix will be an l by l matrix where the $(i, j)^{\text{th}}$ entry is computed according to the equation:

$$S(i,j) = e^{-\left[\frac{(v_i + v_j)^2}{r}\right]} \quad (1)$$

where r is a scaling factor that is chosen for each data stream independently. Each entry in the similarity matrix will range between 0,

Figure 13. Similarity matrix



when the corresponding values are very different relative to the scaling factor, to 1, when the two values are equal. The similarity matrix is symmetric and all the values in the diagonal are 1. Figure 13 shows a typical similarity function where values close to 1 are shaded dark and values close to zero are white. When events are present, the area around the diagonal will be a sequence of dark squares of different sizes, each corresponding to a homogeneous segment in the video. An example of a clear event is at (10,10) because there is no confusion in the similarity matrix.

7.2. Event Edge Detection

The next step detects boundaries between the homogeneous regions in the similarity function. Define a function of the time index i (that we denote $N(i)$ for novelty function) with the following property: when i is within a homogeneous region of the similarity function, $N(i)$ is small; and when i is at the boundary between two homogeneous functions, $N(i)$ is large. Event boundaries can then be found where $N(i)$ has a local maximum.

To formalize the novelty function, we define first a tapered, checkerboard kernel of size k, G_k as a $2k$ by $2k$ matrix where the $(i, j)^{th}$ entry is

$$G_k(i,j) \triangleq \begin{cases} Ae^{-\left[\frac{i^2+j^2}{s}\right]} & \text{if } i > k; j > k \\ & \text{or if } i \le k; j \le k \\ -Ae^{-\left[\frac{i^2+j^2}{s}\right]} & \text{otherwise} \end{cases} \quad (2)$$

where s is a scale used to taper the kernel towards the edges, and A is a kernel normalization constant.

For each data source, we compute a novelty function N by convolving its similarity func-

tion with the checkerboard kernel as show in Equation (3):

$$N(i) = \sum_{\ell;m=k}^{k} S(i+k, m+k) G_k(i,m) \tag{3}$$

The last step is to combine the novelty functions for each multi-modal data stream into a single novelty function, interpolating all $N(i)$ to a common time index and then computing their weighted sum. Each local maximum of the combined novelty function corresponds to an event boundary.

7.3. Normalized Rand Index

The performance is compared against analyst truth: a *valuable* video segmentation generated by an analyst which is user specific and not unique, but 'valuable'. The performance metric must fairly compare event boundaries with different number of segments. It is also desirable that the metric makes accommodation for refinement – when segmentation may agree at one level, but the algorithm may not - in cases where the true activities may be more finely segmented into event time boundaries. For these reasons, and following the suggestion of (Unnikrishnan *et al.*, 2005), we selected a *normalized Rand index* for our engineering metric.

The *standard Rand Index* (Rand, 1971), (*RI*) is computed by counting all the pairs of video frames that were either: a) assigned to the same segment by both the analyst and machine, or b) assigned to different segments by both the analyst and machine segmentation. In other words it gives credit for keeping together things that are related and for separating things that are not related.

The *normalized Rand Index* (*nRI*) compares the *RI* value with the expected value of a random selection of segment edges, and divides it with the maximum value,

$$nRI = \frac{RI - E[RI]}{RI_{\max} - E[RI]} \tag{4}$$

where *RI* compares the computed and analyst-truth, $E[RI]$ is the expected value using randomly distributed segment boundaries, and RI_{max} is the maximum possible value of the index. If $nRI = 1$, the two segmentations match exactly, and if $nRI = 0$ if the computed segmentation performs as well as a random one. If $nRI \leq 0$, the computed segmentation is worse than what would be expected from a random one.

8. QUEST EXAMPLES

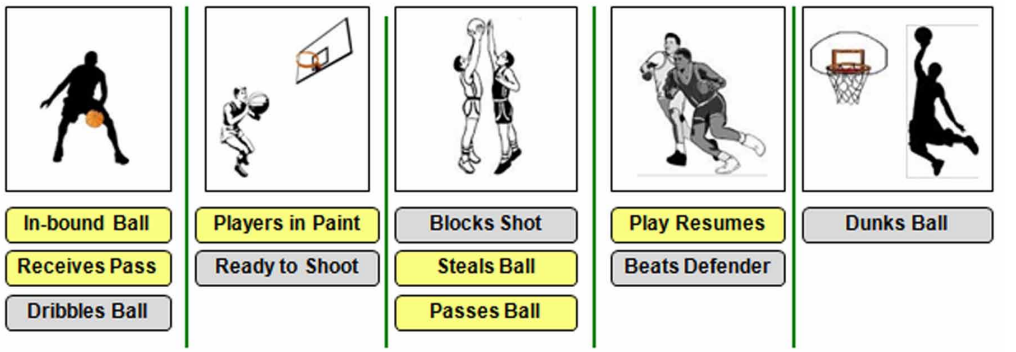
8.1. Example 1: QuEST Narrative

Using QuEST, we present an example of its use in a PCPAD cycle. The example consists of three steps: (1) multimodal activity segmentation, (2) graphical fusion of relevant information, and (3) QuEST-based cohesive narrative reporting.

Step 1. **Multimodal Activity Segmentation:** Information fusion agents are tasked with collecting and analyzing video data with user call-out text data. The event determination (point in time) forms boundaries of activities (perceptions with durations). For example, a sports announcer is describing the facts as well as ancillary information about the situation. Figure 14 highlights an example for a sports (basketball) game. Individual video clips are exploited for multiple mover tracking, player classification, and team relationships. The text information helps in the segmenting of important activities as surrounded by events boundaries (Holloway, *et al.*, 2014). Note that the same stimulus could be used by a coach trying to instruct lessons learned to his/her players (offense and defense) as well as game highlights.

Step 2. **Graphical Fusion of Relevant Information:** A graphical information fusion system links the text data with the video exploitation products (Blasch, Levchuk, *et al.*, 2014). The nodes of relevant information are analyzed

Figure 14. Activities separated by events



for the plausible narratives as determined by the graph in Figure 15. The machine-derived information presented to the analyst is a series of exploited video-to-text products that need to be combined for a collective report. Hence the storage of information is not tracks, but that of activities. The graph at the right of Figure 15 highlights a cohesive narrative from the fused exploited video and extracted text.

Step 3. **QuEST-Based Cohesive Narrative Reporting:** The segmented activities determined from the graphical information fusion are linked for a cohesive narrative story that is disseminated (as available) to other users and machines. Figure 16 highlights various activities of a sporting event combined with external narratives that results in the cohesive narrative. The data alone, without the context, could produce a different narrative.

8.2. Example 2: QuEST Event Tracking

QuEST event tracking does not deal with the raw video pixels, but video exploitation data such as live commentary and metadata. QuEST brings together heterogeneous features (derived from video and text) into a common fusion algorithm to detect the event boundaries (e.g., start, stop) of an activity.

8.2.1. Live Commentary

A user generates textual live commentary (e.g., a sports analyst maintaining a social media presence). Live commentary provides truth-like analysis of salient activities, but is inherently delayed and not always synchronized with the video. Examples of live commentary are subdivided into three types:

- **Announcer:** Person watching the video and calling out actions such as "that was a foul."
- **Commentator:** Internal discussion about the video with announcer such as, "did you see that?"
- **External:** Other's discussions of interesting activities, but not directly seen in the video such as a referee discussing a call, "number 19 holding."

Live commentary messages contain descriptions of what the analysts (announcer, commentator, and referee) see relevant to the video. Some of the language in the messages is constrained (dictated by protocol – e.g. "holding") but not all messages need to adhere to conventional standards (e.g., "hacking"). For example, commentary may contain casual communication unrelated to the game, or may reference observations about video that is in the past. According to our robustness directive, we will not parse and interpret most messages. Instead, for the analysis presented here, we use only the

Figure 15. Graphical processing linking nodes and links

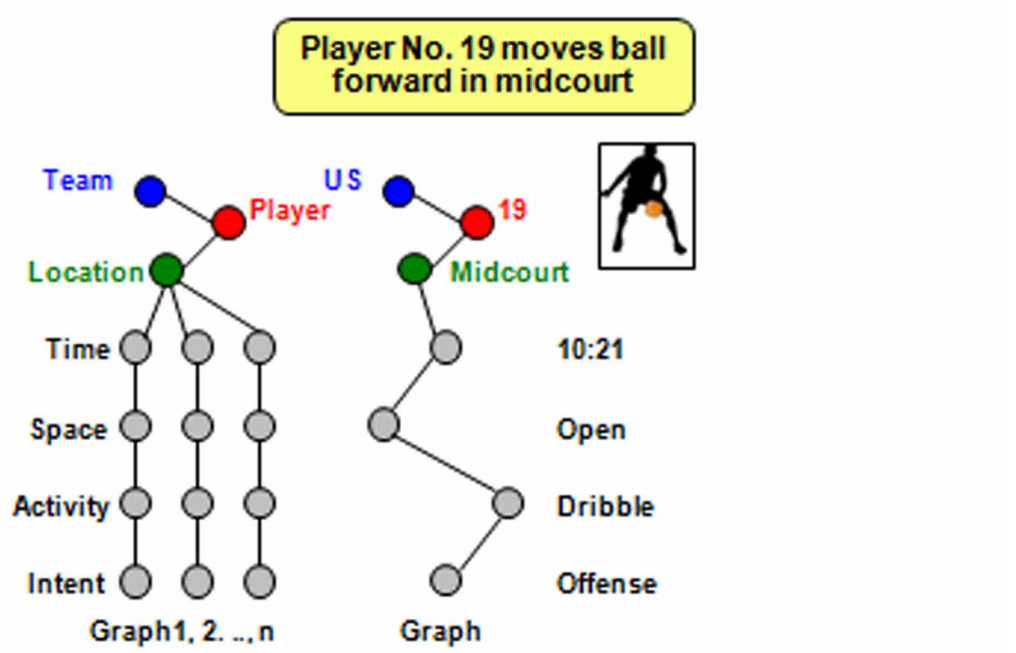
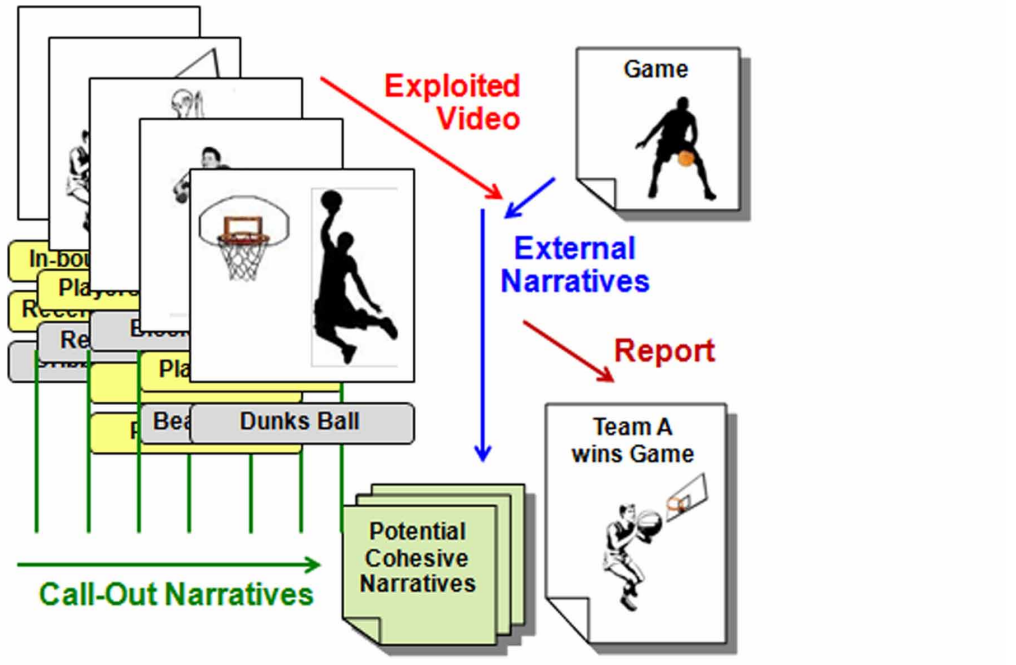


Figure 16. Cohesive narrative from activity segmentation fused with exploited sensed data and extracted semantic data



timestamps of the live commentary messages as clues to the significance of the events on the screen. Furthermore, we can utilize the amount of discussion as a simple cue of importance, where the more dense the message traffic over a fixed interval, the more relevant something is in determining an emerging event. Increased accuracy can be achieved by using the content of at least some of the more structured messages to determine meaningful events, as well as the tone and speed of the discussion.

Figure 17 shows the distribution of message timestamps for a simulated commentary session based on a sample surveillance video. The commentary messages in this session include both descriptive timely messages, as well as side conversations. Simple inspection of the message timestamp distributions suggests that it is an information rich source. Two salient characteristics of this source are the lengths of the commentary strings and the commentary density in each string. Both vary significantly for this 5-minute sample.

For example:

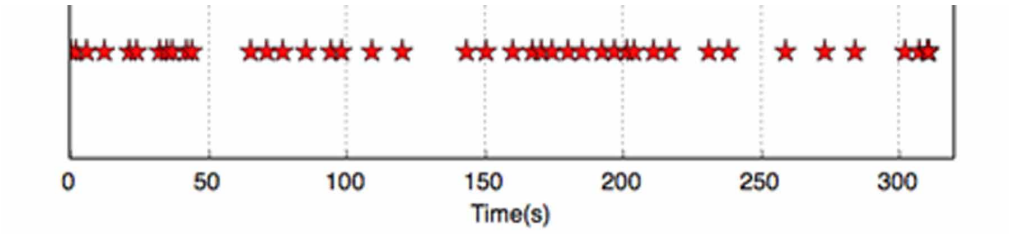
- Interval [0-45s] has consistent commentary with a rise in density that ends abruptly; this pattern is associated with a discussion to a culmination of activity (e.g., "passing, moving up court, shoot");
- Interval [140-210s] is longer in duration with higher density in the middle (e.g., "moving up the court, passing, foul, did you see that, goes to the line");
- Interval [300-310s] is dense with no duration (e.g., "shoots, scores").

Commentary analysis shown in Figure 17 is determined by context and cultural factors. For a sports commentary, much of the analysis is reasonable; however, in other video analyses, such as in disaster relief, more is needed to deal with the complexity of the situation. To supplement the commentary, QuEST uses video metadata as described in the next section.

8.2.2. MetaData

Although the commentary message timestamps carry much information of the scene activity, they are an imperfect and incomplete source. First, we have observed that increases in chat message frequency tend to be a lagging indicator. Second, without attempting to parse the messages, it is difficult to separate the relevant from the auxiliary or irrelevant messages. Parsing the messages is difficult because they do not always follow a defined structure. While communications from aircrews follow rigid procedures for safety and efficiency reasons, and semi-structured sports commentary could apply to known situations, complex situations give rise to unstructured communication results. Encumbering the analyst with new constraints, terminology, or procedures would limit their commentary. We thus complement the commentary timestamp information with two additional sources including centerpoint ground speed and number of moving entities.

Figure 17. Distribution of chat timestamps for a simulated session based on a 5-minute video. Overlapping stars denotes increases in message density



8.2.3. CenterPoint Ground Speed

Although technically the centerpoint speed is mechanically derived, it contains human input: a skilled camera operator will pan and zoom the sensor to follow the action. The reaction of the camera operator is faster than that of the user commenting on the video.

We derive ground speed by differentiating and low pass filtering the geo-registered latitude and longitude of the frame center point. Since the framerate is much higher than the resolution we need for video segmentation, and to minimize computation, we only update centerpoint speed once a second. Figure 18 shows the normalized center point speed for the test case video. Variations in speed during the 5-minute video are significantly above noise levels produced by camera jitter and geo-registration errors. The speed variations exhibit structure suggesting that centerpoint speed is also an information rich data source.

For example by simple inspection of Figure 18, we can identify the following intervals:

- [0-45s] has decreased camera velocity and abruptly ends which then moves at 50s, which could be a following an activity (e.g., panning to one side of the field);
- [160-210s] has a high speed movement, but consistent (e.g., moving up the court and focusing on moving players); and
- [300-310s] is dense with no duration, (e.g., zooming in on a ball going to the basket).

8.2.4. Number of Moving Entities

The second metadata stream we will use is the mean number of moving entities per frame. Although in cluttered areas this signal can be expected to be noisy, in many relevant applications there will be few moving entities in the frame or can be resolved through image subtraction. The absolute number of moving entities is not necessarily significant, but a sudden change can be correlated with an activity of interest. For example, inspection of Figure 19 reveals the following intervals

- [0-45s] focuses in on a few number of movers, which could be a following an activity (e.g., following one player);
- [140-210s] has a consistent mover density (e.g., one player moving up the court); and
- [300-310s] has a single mover, (e.g., zooming in on a ball going to the basket).

We note that the variation in moving entities is not necessarily related to an activity of interest. An example is [80-90s], which has a high number of movers, such as a camera panning out on the audience. This may have nothing to do with the game or commentary (e.g., during a time out when the commentator is waiting for the next play and the camera operator is doing something not in coordination with the play).

8.3. Example 3: QuEST Quantitative Fusion Results

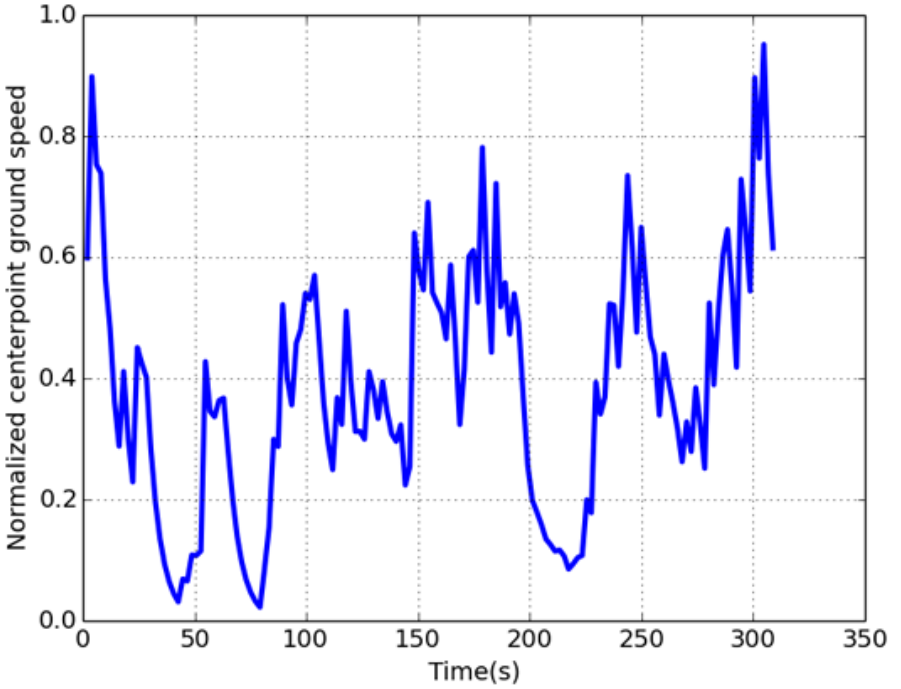
8.3.1. Sum of Fusion Results

We have tested the QuEST algorithm on a data set based on synthetic but relevant video stream from a sporting event. The associated metadata and simulated commentary reflects multiple activities over a single session (e.g., portion of a basketball game). For this data set, we computed the segment edges corresponding to the message stream alone, the metadata alone, and the combination of the two. The results are shown in Figure 20.

The analyst truth was based solely on the video, and without having access to the outcomes of the algorithm, nor the timestamps or content of the live commentary messages. As noted above, key intervals {0-45s, 140-210s, 300-310s} were derived after the fact and organized with the corresponding analysis (Figures 17-19). We optimized the performance by varying the metadata and message timestamp coefficients.

To determine the estimated combinations of activities, we use a fusion as a sum of Gaussians. As shown in Figure 20, we have a series of point estimates $\{x_v: v = 1, \ldots V\}$ for the estimate video points as well as estimates for

Figure 18. Normalized center point speed produced from georegistered frame data



the call-out data $\{x_c: c= 1, \ldots C\}$. To combine the results over all points, we have $x_a = x_v + x_c$. The point estimates are assumed to be the mean of the event such that $A = V + C$ wherein $\mu_A = \mu_V + \mu_C$. Likewise, the uncertainty associated with the combined variance is $\sigma_A^2 = \sigma_V^2 + \sigma_C^2$. Thus, for all estimates, a combined distribution is processed as:

$$P(u,\mu_A,\sigma_A) = \frac{1}{\sqrt{2\pi\left(\sigma_V^2+\sigma_C^2\right)}} e\left(\frac{\left[u\left(\mu_V+\mu_C\right)\right]^2}{\left[2\left(\sigma_V^2+\sigma_C^2\right)\right]}\right) \quad (5)$$

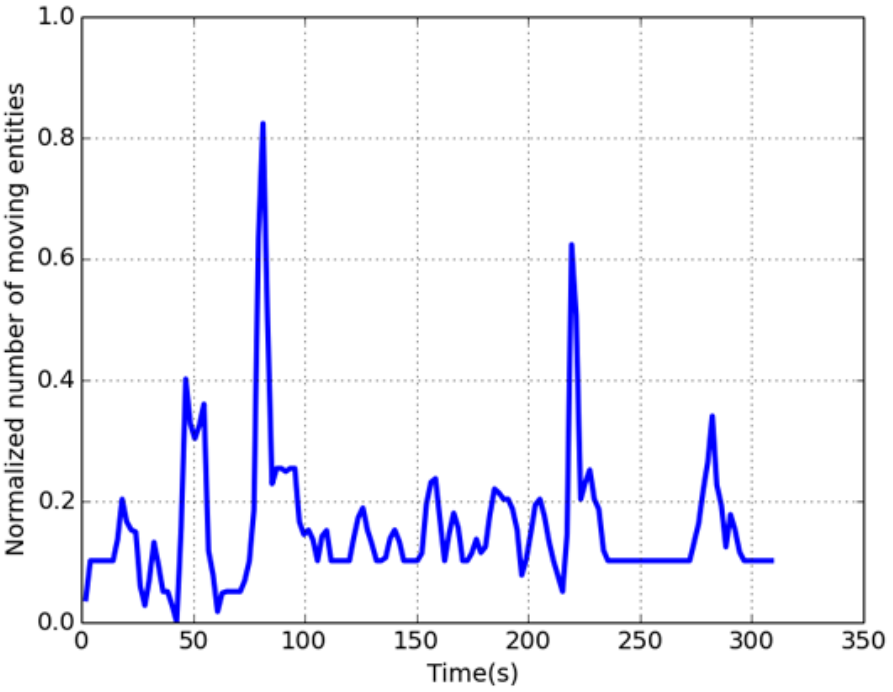
Using the sum of the Gaussians of the measurements, we can fuse the results to get a measure of frequency of activities as shown in Figure 21. From Figure 20, we see that there are instances with multiple corresponding activities of interest as shown between 50-80 and 140-180 seconds. These intervals fall within the suggested analyst-truth event times (shown by black lines in Figure 21). The peaks of the fused Gaussians give a measure of performance that can be used to determine the interval of the activities of interest.

8.3.2. Analysis of Fusion Results for Multimedia Segmentation

Some notes on the results obtained.

- In the first 75 seconds of the video, analysis of the commentary stream of people chatting alone provides a good approximation to the analyst-truth event edges, although, as anticipated, the predicted events lag the real events. Combining commentary and metadata in this same period adds an additional detection (at about 47 seconds).
- In the segment between 75 and 200 seconds, commentary alone is a poor indicator. Event boundaries are detected but they are substantially delayed from the actual

Figure 19. Normalized number of moving entities per frame



events. Introduction of metadata, allows the detector to predict the edges at 120 and 200 seconds.

- In the segment after 200 seconds, commentary timestamps do not provide sufficient information to segment the video. In this case most of the information is derived from the metadata streams. QuEST is able to correctly predict the edges at 210, 255, and 300 seconds.
- The nRI used to tune QuEST, intentionally penalizes sub divisions of clusters less heavily than straddling ones. This explains in part why the analyst-truth events with boundaries 75-120, 120-200, 210-250, and 250 to 300 are correctly identified by QuEST, but also further divided into sub-events.

The normalized Rand Index, *nRI*, values obtained (see Table 3) support a qualitative analysis. An *nRI* value of 0 indicates that the segmentation is not better than a random assignment. (The *nRI* can also be negative, indicating that the algorithm made more assignment errors than would be expected with random assignment). An *nRI* value of λ indicates that QuEST makes a fraction (1-λ) of the errors that a random assignment would make. Results in Table 3 indicate that the optimal blend makes half the errors that a random assignment would make and only 60% of the errors made using commentary information alone. It is important to note that a fraction of these errors are explained by subdividing an analyst-truth segment into smaller sub-segments.

8.3.3 Product of Image Segementation

As per Figure 16, we are building a story board for a narrative summary. A sporting event has a natural duration for reporting that we are emulating. Figure 22 shows how a clip from a

Figure 20. Video Segment (or event) Boundaries for activities are computed using commentary times, centerpoint speed, and number of moving entities

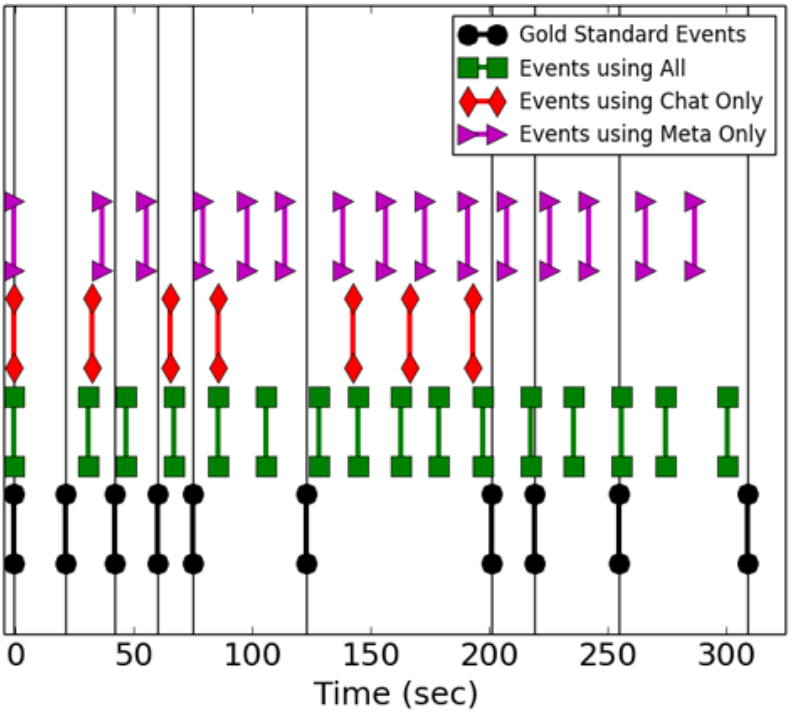


Figure 21. Sum of text and video gaussians for edge detection

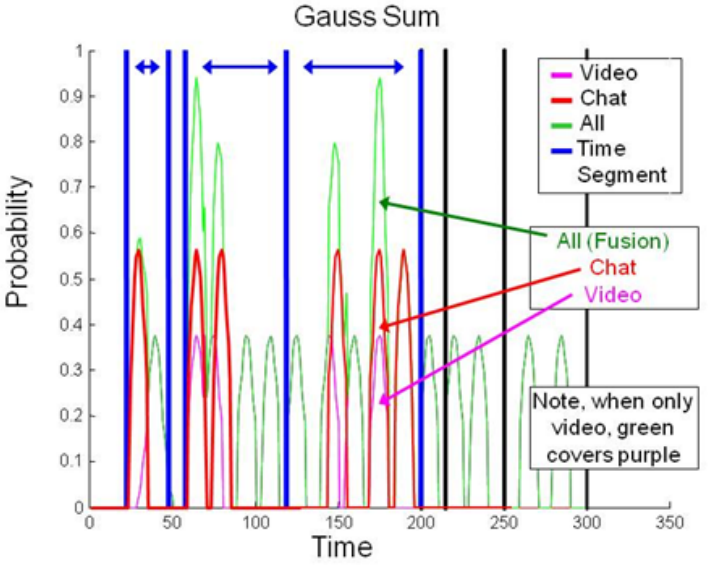
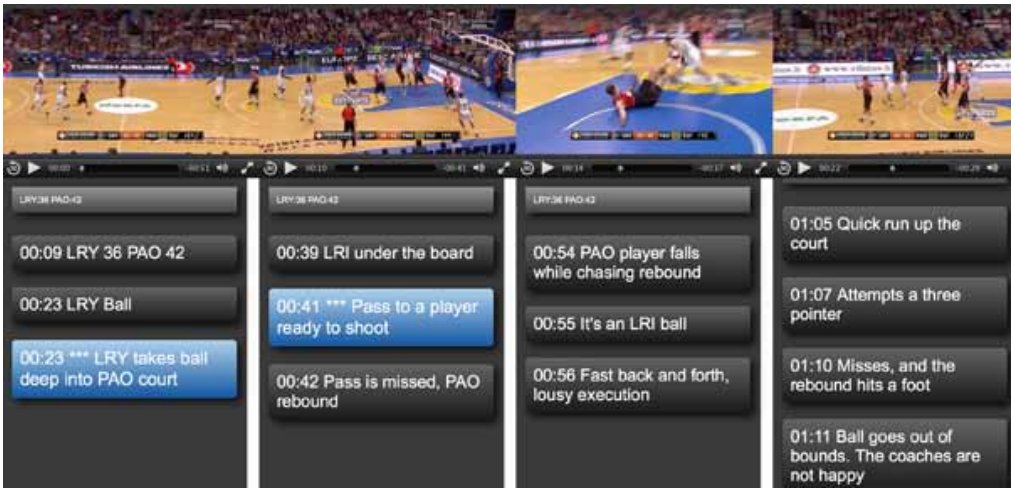


Figure 22. Multimedia video/text segmentation for narratives (Data use: 2013.10.17 Lietuvos Rytas vs. Panathinaikos, available at: https://mega.co.nz/#!0FImyZzC!YzW82m7_MMyZpKE0r3l5XyeP0MpwdqJ_8JnH7ZrXwIE



basketball game may be separated into distinct plays by identifying boundaries based on commentary and metadata (The plays are: Offense by LRY, quick rebound by PAO that ends in a foul, a second offense by LRY which ends in a missed three pointer and the ball being deflected out of bounds).

9. CONCLUSION

In this paper, we highlighted QuEST for information fusion and demonstrated examples with multimedia data. We reviewed qualia as a basis of user experiences and imaged futures that enhanced exploited data sensed sources and included in PCPAD reporting. Looking at cognitive models and information fusion models, we utilized the tenets of QuEST modeling for information fusion that include *structurally coherent, situated conceptualization,* and *simulated experience* for analysis of narratives for situation understanding (awareness, assessment, and reasoning). Video event segmentation by text demonstrates how the QuEST cognitive experience can be used to interpret multimedia data to provide a situated coherent narrative of the key activities extracted from the video and chat data.

ACKNOWLEDGMENT

The authors thank the AFRL QuEST team for the insightful discussions, literature review, and challenges to the thought of consciousness. Erik Blasch was supported under an AFOSR grant in Dynamic Data-Driven Applications

Table 3. All-source blended segment edges better than those obtained individually

Data Sources Fused	Normalized Rand Index
Commentary Only	0.1776
Metadata Only	0.4488
Optimal Blend	0.4966

Systems and support from AFRL. Other discussions provided by: Laurie Fenstermacher, Robert E. Patterson, Jared Culbertson, and Andres Rodriguez of AFRL. Support for the video analysis was also provided by Haibin Ling. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of Air Force Research Laboratory, or the U.S. Government.

REFERENCES

Ahanger, G., & Little, T. D. C. (2001). Data Semantics for Improving Retrieval Performance of Digital News Video Systems. *IEEE Transactions on Knowledge and Data Engineering*, *13*(3), 352–360. doi:10.1109/69.929894

Antony, R. T., & Karakowski, J. A. (2010). Homeland security application of the Army Soft Target Exploitation and Fusion (STEF) system, *Proc. SPIE*, Vol. 7666. doi:10.1117/12.852508

Basharat, A., Gritai, A., & Shah, M. (2008). *Learning object motion patterns for anomaly detection and improved object detection*. IEEE CVPR. doi:10.1109/CVPR.2008.4587510

Blasch, E. (2006). Sensor, User, Mission (SUM) Resource Management and their interaction with Level 2/3 fusion, *Int. Conf. on Info Fusion*. doi:10.1109/ICIF.2006.301791

Blasch, E. (2013A). Book Review: 3C Vision: Cues, Context, and Channels, IEEE Aerospace and Elec. Sys. Magazine, Vol. 28, No. 2, Feb.

Blasch, E. (2013B). Enhanced Air Operations Using JView for an Air-Ground Fused Situation Awareness UDOP, *AIAA/IEEE Digital Avionics Systems Conference*.

Blasch, E. (2014). Decisions-to-Data using Level 5 Information Fusion, *Proc. SPIE*, Vol. 9079.

Blasch, E. et al.. (2005). DFIG Level 5 (User Refinement) issues supporting Situational Assessment Reasoning, *Int. Conf. on Information Fusion*.

Blasch, E., & Banas, C. et al.. (2012). Pattern Activity Clustering and Evaluation (PACE), *Proc. SPIE*, Vol. 8402.

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

Blasch, E., Deignan, P. B. Jr, & Dockstader, S. L. et al.. (2011). Contemporary Concerns in Geographical/Geospatial Information Systems (GIS) Processing, *Proc. IEEE Nat. Aerospace Electronics Conf (NAECON)*. doi:10.1109/NAECON.2011.6183099

Blasch, E., & Hanselman, P. (2000). Information Fusion for Information Superiority, *IEEE Nat'l Aerospace and Electronics Conf.*, Blasch, E. (2000). Assembling a distributed fused Information-based Human-Computer Cognitive Decision Making Tool, *IEEE Aerospace and Electronic Sys. Mag.,* Vol. 15, No. 5, pp. 11-17, May.

Blasch, E., Jøsang, A., Dezert, J., et al. (2014). URREF Self-Confidence in Information Fusion Trust, Int'l. Conf. on Information Fusion, 2014.

Blasch, E., Kadar, I., & Hintz, K. et al.. (2008). Resource Management Coordination with Level 2/3 Fusion Issues and Challenges [Mar.]. *IEEE Aerospace and Electronic Systems Magazine*, *23*(3), 32–46. doi:10.1109/MAES.2008.4476103

Blasch, E., Kadar, I., & Salerno, J. et al.. (2006). Issues and Challenges in Situation Assessment (Level 2 Fusion) [Dec.]. *J. of Advances in Information Fusion*, *1*(2), 122–139.

Blasch, E., Lambert, D. A., Valin, P., Kokar, M. M., Llinas, J., Das, S., & Shahbazian, E. et al. (2012). High Level Information Fusion (HLIF) Survey of Models, Issues, and Grand Challenges. *IEEE Aerospace and Electronic Systems Magazine*, *27*(9), 4–20. doi:10.1109/MAES.2012.6366088

Blasch, E., Levchuk, G., & Staskevich, G. et al.. (2014). Visualization of Graphical Information Fusion Results. *Proceedings of the Society for Photo-Instrumentation Engineers*, *9091*, 2014.

Blasch, E., Nagy, J., & Aved, A. et al.. (2014). *Context aided Video-to-Text Information Fusion, Int'l*. Conf. on Information Fusion.

Blasch, E., Russell, S., & Seetharaman, G. (2011). Joint Data Management for MOVINT Data-to-Decision Making, *Int. Conf. on Info Fusion*.

Blasch, E., Salerno, J., Yang, S. J., Fenstermacher, L., Kadar, I., Endsley, M., & Grewe, L. (2013). Summary of Human, Social, Cultural, Behavioral (HCSB) Modeling for Information Fusion, *Proc. SPIE*, Vol. 8745.

Blasch, E., & Seetharaman, G. et al.. (2013). Dynamic Data Driven Applications Systems (DDDAS) modeling for Automatic Target Recognition, *Proc. SPIE*, Vol. 8744. doi:10.1117/12.2016338

Blasch, E., Steinberg, A., Das, S., Llinas, J., Chong, C.-Y., Kessler, O., & Waltz, E., &, White, F. (2013). Revisiting the JDL model for information Exploitation, *Int'l Conf. on Info Fusion*.

Blasch, E., Valin, P., & Bossé, E. (2010). Measures of Effectiveness for High-Level Fusion, *Int'l Conf. on Info Fusion*.

Blasch, E., Wang, Z., Ling, H., Palaniappan, K., Chen, G., Shen, D., & Seetharaman, G. et al. (2013). Video-Based Activity Analysis Using the L1 tracker on VIRAT data, *IEEE Applied Imagery Pattern Rec. Workshop*. doi:10.1109/AIPR.2013.6749311

Chaiken, S., & Trope, Y. (1999). *Dual-process Theories in Social Psychology*. Guilford Press.

Chan, M. T., Hoogs, A., Schmiederer, J., & Petersen, M. (2004). Detecting Rare Events in Video Using Semantic Primitives with HMM, *International Conference on Pattern Recognition*. doi:10.1109/ICPR.2004.1333726

Chellappa, R. (2014). *Frontiers in Image and Video Analysis NSF/FBI/DARPA Workshop Report*. Accessed at http://www.umiacs.umd.edu/~rama/NSF_report.pdf, Dec. 2014.

Cooper, M., Foote, J., Girgensohn, A., & Wilcox, L. (2005). Temporal event clustering for digital photo collections. *ACM Transactions on Multimedia Computing, Communications, and Applications*, *1*(3), 269–288. doi:10.1145/1083314.1083317

Costa, P. C. G., & Laskey, K. B. et al.. (2012). Towards Unbiased Evaluation of Uncertainty Reasoning: The URREF Ontology, *Int. Conf. on Information Fusion*.

Cowell, C. W. (2001). Minds, Machines and Qualia: A Theory of Consciousness, PhD Dissertation, UC Berkeley.

Culbertson, J., Sturtz, K., Oxley, M., & Rogers, S. K. (2012). Probabilistic Situations for Reasoning, IEEE Cognitive Methods in Situation Awareness and Decision Support (CogSIMA).

Czajkowski, M., Ashpole, B., Hughes, T., & Le, T. (2006). Bridging Semantic eGovernment Applications using Ontology-to-Ontology Message Translation, *AAAI Spring Symposium*.

Czajkowski, M., Buczak, A. L., & Hofmann, M. O. (2004). Dynamic Agent Composition from Semantic Web Services, *International Workshop on Semantic Web and Databases*.

Denis, N., & Jones, E. (2003). Spatio-Temporal Pattern Detection Using Dynamic Bayesian Networks, *IEEE Conf. on Decision and Control*. doi:10.1109/CDC.2003.1272263

Dennett, D. C. (1988). Quining Qualia. In A. J. Marcel & E. Bislach (Eds.), *Consciousness in Contemporary Science*. Oxford University Press.

DiBona, P., Belov, N., & Pawlowski, A. (2006). *Plan-Driven Fusion: Shaping the Situation Awareness Process using Empirical Plan Data, Int'l. Conf. on Information Fusion*. on Information Fusion.

Endsley, M. R. (1988). Design and evaluation for situation awareness enhancement, *Proceedings of the Human Factors Society 32nd Annual Meeting*, Santa Monica, CA: Human Factors and Ergonomics Society, pp. 97-101, 1988.

Endsley, M. R. (1995A). Toward a Theory of Situation Awareness. *Human Factors*, *37*(1), 32–64. doi:10.1518/001872095779049543

Endsley, M. R. (1995B). Measurement of situation awareness in dynamic systems. *Human Factors*, *37*(1), 65–84. doi:10.1518/001872095779049499

Evans, J. St. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, *8*(3), 223–241. doi:10.1177/1745691612460685

Fenstermacher, L. (2014). Information fusion: telling the story (or threat narrative)," *Proc. SPIE*, Vol. 9091.

Fisher, Y., & Beyerer, J. (2012). Defining Dynamic Bayesian Networks for Probabilistic Situation Assessment, *International Conference on Information Fusion*.

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement [June]. *Journal of Experimental Psychology*, *47*(6), 381–391. doi:10.1037/h0055392 PMID:13174710

Fitts, P. M. and Posner, M. I. (1962). Human Performance. Brooks/ Cole. Department of Defense Science Board (2012). The role of autonomy in DoD systems, July.

Global Integrated Intelligence, Surveillance, and Reconnaissance Operations, Air Force Doctrine Document 2-0, 6 Jan. (2012).

Graham, J. L., Hall, D. L., & Rimlan, J. (2011). A synthetic dataset for evaluating soft and hard fusion algorithms. *Proceedings of the Society for Photo-Instrumentation Engineers*, *Vol.* 8062.

Greene, K., Cooper, D. G., Buczak, A. L., Czajkowski, M., Vagle, J. L., & Hofmann, M. O. (2005). *Cognitive Agents for Sense and Respond Logistics, Autonomous Agents and Multi-Agents System conference*. AAMAS.

Hammoud, R. I., & Sahin, C. S. et al.. (2014). *Multi-Source Multi-Modal Activity Recognition in Aerial Video Surveillance, Int. Computer Vision and Pattern Recognition*. ICVPR.

Hammoud, R. I., Sahin, C. S., Blasch, E., Rhodes, B., & Wang, T. (2014). Automatic Association of Chats and Video Tracks for Activity Learning and Recognition in Aerial Video Surveillance. *Sensors (Basel, Switzerland)*, *14*(10), 19843–19860. doi:10.3390/s141019843 PMID:25340453

Harman, G. (1993). Some philosophical issues in cognitive science: qualia, intentionality, and the mind-body problem. In *Foundations of Cognitive Neuroscience* (pp. 831–848). Cambridge, MA: MIT Press.

Holloway, H., Jones, E. K., & Kaluzniacki, A. et al.. (2014). Activity Recognition using Video Event Segmentation with Text (VEST), *Proc. SPIE*, Vol. 9091.

Hoogs, A., Mundy, J., & Cross, G. (2001). "Multi-Modal Fusion for Video Understanding," *IEEE Applied Imagery Pattern Rec. Workshop*.

Hoogs, A., Rittscher, J., Stein, G., & Schmiederer, J. (2003). *Video Content Annotation Using Visual Analysis and a Large Semantic Knowledgebase*. IEEE CVPR. doi:10.1109/CVPR.2003.1211487

Kahler, B., & Blasch, E. (2008). Sensor Management Fusion Using Operating Conditions, *Proc. IEEE Nat. Aerospace Electronics Conf. (NAECON)*.

Kennedy, W. G., & Patterson, R. E. (2012). Modeling Intuitive Decision Making in ACT-R, *Proc. of International Conf. on Cognitive Modeling*.

Klein, G., Moon, B., & Hoffman, R. (2006). Making sense of sensemaking 1: Alternative perspectives. *IEEE Intelligent Systems*, *21*(4), 70–73. doi:10.1109/MIS.2006.75

Liu, B., Blasch, E., & Chen, Y. et al.. (2014). Information Fusion in a Cloud Computing Era: A Systems-Level Perspective [Oct.]. *IEEE AES Magazine*, *29*(10), 16–24. doi:10.1109/MAES.2014.130115

Nghiem, A. T., Thonnat, B. M., & Ma, R. (2007). *A new Evaluation Approach for Video Processing Algorithms*. WWVC. doi:10.1109/WMVC.2007.2

Oh, S., Hoogs, A., Perera, A., Cuntoor, N., & Chen, C.-C. et al.. (2011). *A Large-scale Benchmark Dataset for Event Recognition in Surveillance Video*. IEEE CVPR. doi:10.1109/CVPR.2011.5995586

Panasyuk, A. et al.. (2013). *Extraction of Semantic Activities from Twitter Data*. Semantic Tech. for Intelligence, Defense, and Security.

Patterson, R., Fournier, L., Williams, L., Amann, R., Tripp, L., & Pierce, B.P. (2012). System dynamics modeling of sensory-driven decision priming, *Journal of Cognitive Eng. and Decision Making*.

Patterson, R., Pierce, B., Boydstun, A., Park, L., Shannon, J., Tripp, L., & Bell, H. (2013). Training intuitive decision Making in a simulated real-world environment. *Human Factors*, Apr; 55(2): 333-45. PMID:23691829

Patterson, R., Pierce, B. J., Bell, H. H., & Klein, G. (2010). Implicit learning, tacit knowledge, expertise development, and naturalistic decision making. *Journal of Cognitive Engineering and Decision Making*, *4*, 289–303.

Ramachandran, V. S., & Hirstein, W. (1997). Three laws of qualia: What neurology tells us about the biological functions of consciousness, *Journal of Consciousness*.

Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, *66*(336), 845–850. doi:10.1080/01621459.1971.10482356

Rasmussen, J. (1983). Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models," *IEEE T. Systems, Man, and Cybernetics*, *13*(3), 257–266. doi:10.1109/TSMC.1983.6313160

Reddy, K. K., & Cuntoor, N. et al.. (2012). "Human Action Recognition in Large-Scale Datasets Using Histogram of Spatiotemporal Gradients, *IEEE Int'l Conf. on Adv. Video and Signal-Based Surveillance*. doi:10.1109/AVSS.2012.40

Rodriguez, A., Boddeti, V. N., Kumar, B. V. K. V., & Mahalanobis, A. (2013). Maximum Margin Correlation Filter: A New Approach for Localization and Classification. *IEEE Transactions on Image Processing*, *22*(3), 631–664. doi:10.1109/TIP.2012.2220151 PMID:23014751

Rogers, S. K. et al.. (2008). The life and death of ATR/sensor fusion and the hope for resurrection, *Proc. of SPIE*, Vol. 6967.

Rogers, S. K. et al.. (2014). *Sensing as a Service*. AFRL.

Rogers, S. K., Kabrisky, M., Bauer, K., & Oxley, M. (2003). Computing Machinery and Intelligence Amplification. In D. B. Fogel & C. J. Robinson (Eds.), *Computational Intelligence: The Experts Speak* (pp. 25–44). Piscataway, NJ: IEEE Press.

Sandell, N. F., Savell, R., Twardowski, D., & Cybenko, G. (2009). HBML: A Representation Language for Quantitative Behavioral Modeling in the Human Terrain, *Social Computing and Behavioral Modeling*, Liu, H., Salerno, J. J., Young, M. J. (eds.), Springer.

Smith, E. R., & DeCoster, J. (2000). Dual-Process Models in Social and Cognitive Psychology: Conceptual Integration and Links to Underlying Memory Systems. *Personality and Social Psychology Review*, *4*(2), 108–131. doi:10.1207/S15327957PSPR0402_01

Swears, E., Hoogs, A., & Perera, A. G. A. (2008). *Learning Motion Patterns in Surveillance Video using HMM Clustering*. IEEE WMCV. doi:10.1109/WMVC.2008.4544063

Tsitsoulis, A., & Bourbakis, N. (2013). A first stage comparative survey on vision-based human activity recognition, *Int. Journal on AI Tools*, vol. 24, no.6.

Turaga, P., Chellappa, R., Subrahmanian, V. S., & Udrea, O. (2008). Machine Recognition of Human Activities: A Survey. *IEEE Transactions on Circuits and Systems for Video Technology*, *18*(11), 1473–1488. doi:10.1109/TCSVT.2008.2005594

Unnikrishnan, R., Pantofaru, C., & Hebert, M. (2005). A measure for objective evaluation of image segmentation algorithms, *Robotics Institute*, Paper 366.

Vaughan, S. L., Mills, R. F., Grimaila, M. R., Peterson, G. L., & Rogers, S. K. (2014). Narratives as a Fundamental Component of Consciousness, *Comp. Models of Narrative Workshop*.

Vondrick, C., Patterson, D., & Ramanan, D. (2012). Efficiently Scaling up Crowdsourced Video Annotation: A Set of Best Practices for High Quality, Economical Video Labeling. *International Journal of Computer Vision*.

Walsh, M. M. (2013). Robustness as a guiding principle for the design of cognitive and behavior systems, *Conference on Behavior Representation in Modeling and Simulation (BRiMS)*.

Walsh, M. M., Einstein, E. H., & Gluck, K. A. (2013). A quantification of robustness. *Journal of Applied Research in Memory and Cognition*, *2*(3), 137–148. doi:10.1016/j.jarmac.2013.07.002

Wu, T., & Pottenger, W. M. (2005). A Semi-Supervised Active Learning Algorithm for Information Extraction from Textual Data. *Journal of the American Society for Information Science and Technology*, *56*(3), 258–271. doi:10.1002/asi.20119

Yang, C., & Blasch, E. (2009). Kalman Filtering with Nonlinear State Constraints [Jan.]. *IEEE Transactions on Aerospace and Electronic Systems*, *45*(1), 70–84. doi:10.1109/TAES.2009.4805264

Yang, Y., Liu, J., & Shah, M. (2009). *Video Scene Understanding Using Multi-scale Analysis*. ICCV. doi:10.1109/ICCV.2009.5459376

Yuan, C., Li, X., Hu, W., Ling, H., & Maybank, S. (2013). 3D R Transform on Spatio-Temporal Interest Points for Action Recognition, *IEEE Conf. on Computer Vision and Pattern Recognition* (CVPR). doi:10.1109/CVPR.2013.99

Zhang, X., Li, W., Hu, W., & Ling, H. et al.. (2013). Block covariance based L1 tracker with a subtle template dictionary. *Pattern Recognition, Vol. 46, Issue 7, 1750–1761*.