Collected Imagery Ontology: Semantics for Airborne Video

Alexander Mirzaoff

Geospatial Systems
ITT
Rochester, New York. August 2010
585-269-5700
Alexander.Mirzaoff@ITT.com

Abstract. A prototype Video Imagery Ontology has been developed to derive video imagery intelligence, VideoIMINT. The ontology includes the development of classes and properties to address video image content, and video collection metadata related to platforms, sensors and collection operations. Preliminary feature extraction of video imagery content classes was functionally utilized to identify important video segments in an integrated viewer. Integrated data storage systems and fusion processes are proposed and discussed.

Keywords: Ontology, semantic, imagery, video, intelligence.

1 Introduction

For decades, the increasing volume of imagery data has been a growing challenge for the military and intelligence communities, "*too much to look at…*" and "*most of the bits end up on the floor*". The coming of age of Video Intelligence Surveillance and Reconnaissance (VISR) has only exacerbated the problem by orders of magnitude. For areal coverage with multiple, high resolution cameras [1], operating at two hertz and greater frame rates, data volume is now calculated in yotta-bytes (10^24 bytes). Notwithstanding the computational, storage and networking problems associated with this amount of data, finding content via database searches through these many instances of data becomes very problematic. Lt. Gen. David A. Deptula remarked that the Air Force could soon be "swimming in sensors and drowning in data." [2]. The recognition in this comment of the sensor, as well as the data volumes, as part of the overwhelming information glut, is very important and telling as to how these systems are utilized.

Ontology structures, as a filter for domain information, and ontology enabled rules of organization, present many advantages to help navigate and automatically use such volumes of information. Ontologies can address apparent substantive conflicts of

detection when confronted with phenomena represented by different sensors (panchromatic, multispectral, infrared, RADAR…) on various platforms, collected under widely different circumstance in an automated, sensor to computer to human workflow.

Collected imagery data, and to a larger extent, the information represented, is an organizational, if not a metaphysical, challenge. Consider just two sensors, infrared (IR) and RADAR on the same aircraft. Does all the IR data go here to the IR data bin and all the RADAR data over there in the RADAR bin? Suppose we have both from the same area on different days, or perhaps one for 5 minutes and another data set for 5 hours? Specifically, how are such diverse collections correlated? Do we organize by spectra, by location, by time, or perhaps platform? Is intelligence driven or prioritized by location, time, content, or all these attributes and more? Obviously, these elements are all important, while to complicate matters, the importance varies from mission to mission.

Additionally, there are operational classes that impact domain organization; including aspects of, surveillance utility or operational reconnaissance. Elements of platform specification and platform performance, sensors and sensor performance, and products derived from mission data are also important. The ontological effort is to separate these concepts so that sensor performance, for instance, can be applied to any mission, describing sensor success in some qualitative and quantitative manner.

However, the most differentiating property of intelligence collections is data content: data defined features and objects extractable from a particular collection. While all other elements, or classes, of imagery collection, such as which aircraft, which sensor, provide a rich compilation of schematic information – subclasses and properties – it is the semantics of imagery content that moves this structure from the utility of databases to the world of ontologies. To understand this difference, consider the query "which *sensor* observed the IED explosion at *location x* during *time t*", as compared to, "were *individuals* observed prior to IED explosion at *location x* during *time t*". While building a database schema construct for object concept *sensor* is non-trivial, adding a class such as *individuals* which is, in fact, detected content of imagery, becomes a significant semantic encounter.

Thus, the initial effort has been to define, organize and build an ontology of the VISR domain, including imagery content classes, to enable automated data processing and domain query and management. Subsequent efforts will use this structure to develop the complex logic and relationships of this domain. Flexibility and change are driving principles so that the resulting ontology can be edited: modified as new knowledge is gained, particularly as imagery context is developed with more and more elements extracted from imagery data.

2 Initial Classes

The VISR classes that were initially proposed include the following:

Domain Classes	Subclasses
1. ***Platform Sensor and Sensor Operation***	a. Platforms b. Sensors c. Operational Parameters d. Calibration and Quality Metrics
2. ***Collection and Collection Performance***	a. Collection Variables b. Collection Operational Parameters c. Collection Performance Metrics
3. ***Mission and Targets***	a. Mission Description b. Detection and Characterization
4. ***Imagery and Exploitation Products***	a. VideoIMINTHierarchy b. Product Descriptions c. Product Utility d. Data Assurance Metrics
5. ***Integrated Ontology***	a. Relationship and Rule Algorithms

Table 1. Initial Organizational Construct

Due to programmatic limitation, only classes 1, 2, and 4d were developed. The ***Integrated Ontology*** concept was dropped because the major classes covered the domain rather completely for this application, requiring no further integration, and relationships were an outcome of structure, even at the lowest levels of, for example, sensor calibration and product utility.

The AAF Profile for Aerial Surveillance and Photogrammetry Applications (ASPA) specification [3] provides an excellent starting structure to begin differentiating such concepts as performance and metrics in this ontology. This metadata specification is an XML type structured document that lends itself well to transition into a Resource Description Framework (RDF) for use in a hierarchical ontology.

The ASPA specification covers a great deal of video support information, including where and when it was collected as well as sensor data and platform data, so that the consequent instances of a particular mission, reflected in the video metadata, easily populate the ontology classes of platform and sensor. Such information, semantically consistent, and further constrained by the ontology structure, can form the basis for subsequent queries that reveal much richer content than at first apparent. In fact, structuring the ontology in this manner sets up the entire domain in a logical and computationally complete structure. To further enhance the subsequent utility, the ontology is written in the Ontology Web Language (OWL) standardized by the World Wide Web Consortium (W3C) in the Descriptive Logic (DL) version.

The VISR ontology design was based upon an upper level ontology utilized by the National Center for Ontology Research (NCOR). In this approach, Entities and Events constitute the two main component classes of the upper level, with an entity comprised of two main branches, the Dependent and Independent Continuants.

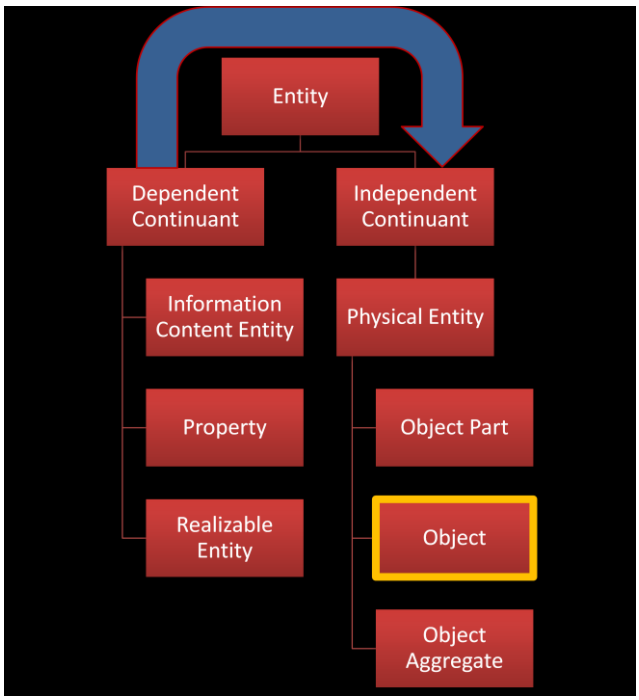

Figure 1. Upper Classes of the NCOR ontology.

Working down through *Independent Continuant* branch to the class of *Object*, we find that this area of the ontology includes subclasses for *Information Bearing Entity*, *Image Bearing Entity*, and both *VideoImage Bearing Entity* and *StillImage Bearing Entity*. Including these as subclasses of *Information Bearing Entity* allows for the later expansion of the class to include other sensor data such as from SIGINT or MASINT collection systems.

On the *Dependent Continuant* side of the ontology structure, we find the *Information Content Entity* from which is derived a *Descriptive Information Content Entity*, and subsequently the class *Image* and a subclass *Video Image*, an image that contains a moving (or extended temporal) representation of some Entity or Event, or *Still Image*, an image that contains a non-moving (or limited temporal) representation of some Entity or Event. These classes are what we would normally think of as the image or the video, while the *Image Bearing Entity*, including both *VideoImage Bearing Entity* and *StillImage Bearing Entity* are bearers of some *Video Image* or some *Still Image* found in the *Dependent Continuant* side of the ontology.

This differentiation provides for the description and definition of additional object classes such as *Pixel* and *Geospatial Region* as an *Independent Continuant* of the pictures that may be subsequently created. Additionally, the ontology can describe classes of *Object* such as *Facility*, *Vehicle* and *Sensor* independently of any particular *Facility*, *Vehicle*, and *Sensor*, again providing a means to specify facilities that are then imaged with particular attribute subclasses such as *Airport* or *Aircraft*.
There is another type of *Physical Entity* class called *Object Aggregate*, of which a subclass is a *Platform*. This *Platform* has properties denoted as `has_part,` such as **ImageSensor** and another `has_part`, **Aircraft**. In this manner, we can now construct a complex object, a **UAV**, as shown in Figure 2.
So with such a construct, we have the ability to present an image, describe its content (through some content extraction algorithm, such as feature extraction or automated target recognition) and relate that content to associated collection parameters (e.g. sensor, frame location, time, altitude…) as well as quality metrics of sensor performance that would be reflected in pixel characterizations, for example.

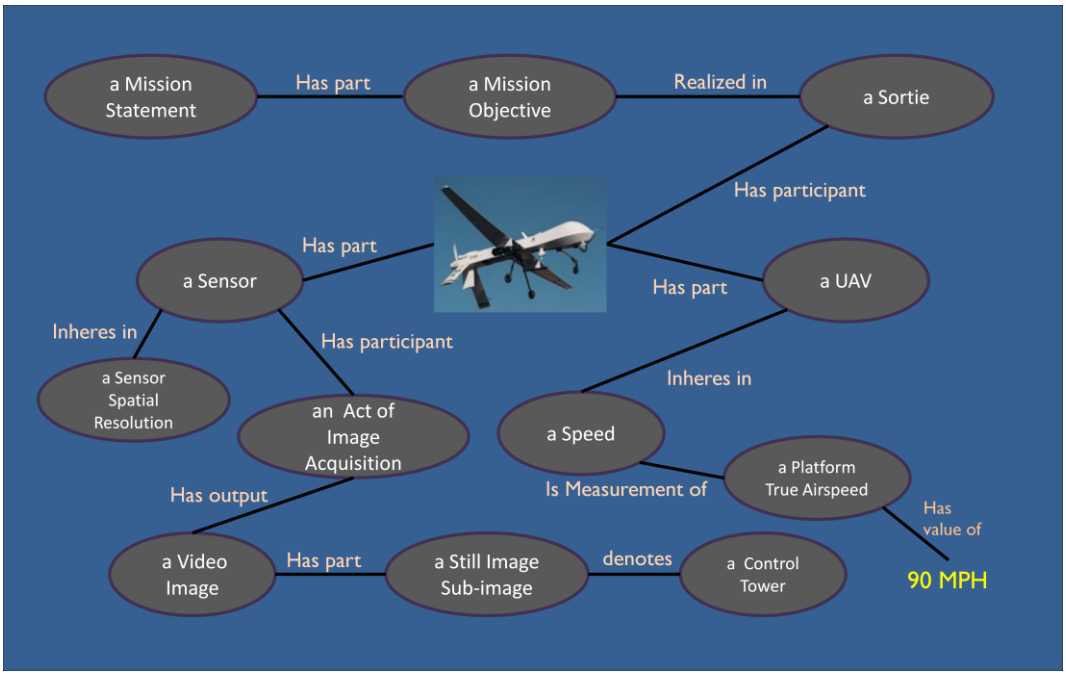

Figure 2. Real World Object as multi-class constructs

Note also that the instances of aircraft properties such as Speed, Direction and Location can be found in the ASPA metadata that is passed with the Predator UAV Datalink Local Metadata Set data elements (*i.e.* video metadata) [4]. Furthermore, since this information is dynamic, it can be updated and associated with any frame of the video collection.

3 Data as Image Content

A key aspect of making this VideoIMINT ontology useful is the ability to extract content from image data. That is, to be able to identify objects (e.g. vehicles, people, weapons), preferably in an automatic manner, from the collected data. There are two aspects to this problem: first, the image content itself – the targets of interest, and second, the support data provided by the sensor and sensor platform as well as from other opportunistic sources. First, we will review the challenges associated with discovering imagery content.

Ontological classes of content at first appear to be straight forward – vehicles, facilities, infrastructure, people… yet extracting these target object instances to populate these classes is a complex and elusive process to undertake in an automated manner. Manual tagging is an option that will be used for the foreseeable future, and facilitating this functionality in an efficient, icon driven manner is an additional objective of the VideoIMINT ontology effort, as is editing the classes of the ontology to be able to add additional target classes.

Automated feature –object extraction from imagery, and in this case video, continues to be an evolving and complex process. Much of the early efforts in understanding and classifying data from overhead remote sensors were in the area of Geographical (or more recently, Geospatial) Information Systems (GIS). For earth observing systems, in order to classify sensor data and build an ontology, Camara *et al* [5] originally argued for a concept of objects as a subset of geospatial fields while acknowledging the overly generic boundaries of this idea. With this approach, everything in the world is a field or an object in the field. This bodes well for constructing a subsequent ontological model since the separation of objects is axiomatic. The problem with such an approach, is deciding, from a sensor viewpoint, rather than a geographer's, which is field and which is object. From a purely GIS perspective the field/object solution is more semantic than image data content oriented; transcriptions of know objects in the world: mountains, rivers, roads…rather an *a priori* method of knowledge recording, provide a framework for ontology constructs: everyone knows a river, and there it is. However, the limitations of this world view were understood when, for example, one would try to decide where the very dynamic river object began and the river bank ended. This was difficult enough to ascertain during a ground survey, much less from overhead sensors looking at terrain during different times of year.

For modern intelligence gathering systems, finding and identifying a road can be accomplished, for the most part, automatically. However, finding a road that is more earth than road can be difficult, requiring perhaps special sensor configurations as well as special data processing. This is a case of the "object" merging with the "field". In fact, the entire problem of object recognition in sensed data can be reduced to first detecting the object,, that is, separating it from the background, and then recognizing what the object is and subsequently characterizing the target object [6]. Furthermore, tracking, or maintaining a view of the detection, a key capability for a video surveillance system, presupposes that 1) an object of interest has been detected and 2) the same object is being recognized in subsequent temporal increments: that is, being tracked.

While detection and tracking of objects in motion came into formal study during World War Two with the invention of RADAR, and the technical evolution of tracking since that time has of course been significant, yet the fundamental problems are the same. The issues have centered on state estimators, probability, statistics, and linear system analysis, all somewhat outside the scope of this paper. Yaakov Bar-Shalom portrays the problem as "…estimation of states of targets in surveillance systems operating in a multitatarget-multisensor environment. This problem is characterized by measurement origin uncertainty."[7]. However, once a system dominates uncertainty, target classification and population of ontological entity objects may proceed. Ontology refinement becomes a function of simply combining the extracted target objects with the collection associated metadata so that a vehicle image in one collect is differentiated from a vehicle image in a different collect. The fuzzy boundaries of the river-bank object can be quantified by a metadata structure with metrics appropriate to the target, or qualified by a time of year tag. Multiple target ambiguity is reduced in a sequence by noting position based on platform geoposition and camera pointing: information carried in the metadata stream [3].
In our preprocessing to populate ontology classes representing image content, we were able to successfully employ multiphase image decomposition and shape recognition algorithms [8] to extract target objects from video scenes. Local contextual information combined with statistical boosting was part of this image analysis process. Learning object representation is also an important part of the analysis and compatible with multiframe video so that subsequent collects of similar objects will enhance recognition success.

4 Integration of Content

The pivotal classes to be developed in the VideoIMINT ontology are the classes of imagery content and targets. Since both methods of extracting such imagery features: manual and automatic are utilized, an important aspect of the development effort was to define these classes and properties so as not preclude one or the other method while remaining consistent with other class and property descriptions.
In the ontology design, we have already constructed a class of *Property*, a *Dependent Continuant* of *Entity* class. The elements of *Property* include physical properties such as Location, Direction, Distance, Height, Length and other physical features. While these properties can apply to aircraft, sensors or other *Independent Continuant*, *Physical Entity* Classes, they can also apply to imagery content classes such as an Airfield *Control Tower*. Thus, new classes can be added to capture the concepts of imagery content – targets, and the existing property classes can be used to define them, dimensions and location.
The human analyst can efficiently recognize and tag content in videos, however, as was posited early in this paper, there is just too much to view. Therefore, automated extraction processing is an important mechanism for populating instances of the ontology while recognition of content detail and differentiation is not necessarily important. As was demonstrated in the development program, simple recognition of a "runway", a "control tower" facility and "aircraft" was sufficient to locate specific

video segments to vastly reduce the amount of data a human needed to review. The recognition of aircraft type was unnecessary, only the fact of presence of aircraft in the video made an enormous difference in video volume requiring human review. Figure 3 shows how segments of the video were highlighted by the ontology reasoner "knowing" that the class of *Aircraft* had been populated by the recognition engine. Those segments of identified video also reference the associated geospatial location, time, sensor and other details regarding the collection.

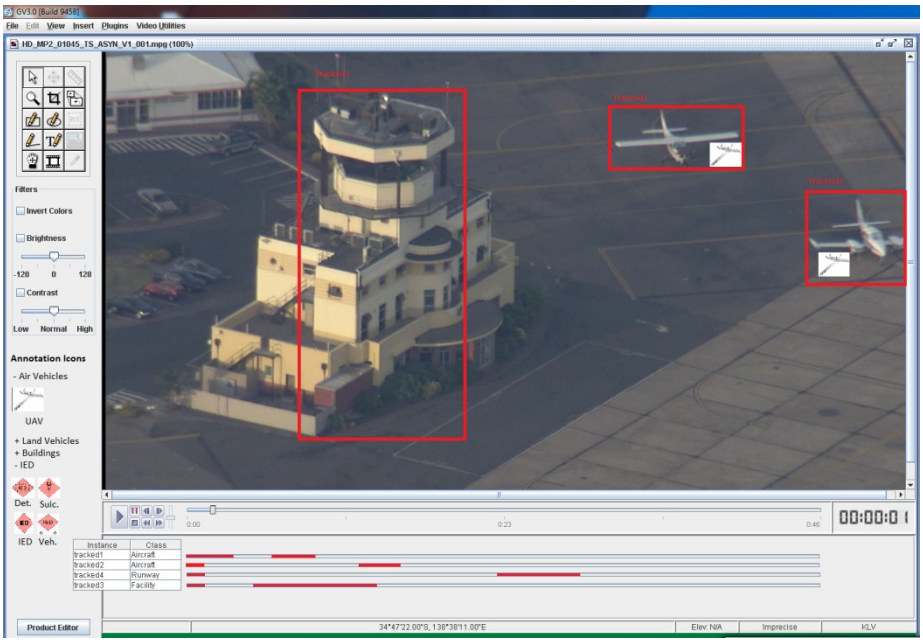

Figure 3. Video viewer showing highlighted segments of recognized content. The analyst has only to skip to that segment to find aircraft – and perhaps add his own tag of type identification.

5 Intelligence Assurance

A practical aspect of all this metadata information, along with the imagery (*InformationContentEntity*) is the inherent ability to determine quality of collection at any time, and conversely, the ability to predict collection performance *a priori*, in order to manage missions in terms of platform/sensor and operations to complete mission requirements and fulfill Essential Elements of Information (EEI) needs. That is to say, if the mission is to image an SA-6 Integrated Air Defense System (IADS) as opposed to determine whether individuals in an urban area are carrying Man Portable Air Defense systems (MANPADS), the proper combination of aircraft/sensor/altitude can be determined prior to mission execution: essentially a dynamic National Imagery Interpretability Rating Scale (NIIRS) for video collection to drive tasking.

Furthermore, imagery can be subject to valuation for quality metrics, such as consistent General Image Quality Equation (GIQE) [9] performance in regards to factors such as spatial resolution in terms of ground sample distance (GSD), relative edge response (RER) and overall system modulation transfer function (MTF) [10], after the fact, to determine system performance efficacy. All of these factors can be calculated, in many cases dynamically, but certainly as simple reasoned queries into the ontology. The true value is that the semantics of system performance are enforced by the ontology such that the variables of formula are consistent, yielding comparative and useful results. It is then possible to understand how one platform/sensor combination will perform, or is performing, relative to another under varying conditions for various missions.

6 Data Storage and Fusion

An integrated approach to video collection systems that includes processing, ontology mapping, storage and fusion would certainly enhance the overall utility and value of this intelligence source. Integration of an ontology with a tightly coupled storage system can yield value in the same manner as designing a data schema will for any data storage system. In fact, there are many similarities between a database schema and an ontology. However, one of the major differences is that a schema is essentially a static construct and does not support logical inferences in the way an ontology does. [11, 12] For example, a query into ontology might ask if a particular imaged runway can support a large cargo aircraft. The ontology can explore data rules regarding classes of runways, aircrafts and their properties, one of which may be a relationship between aircraft type with a property of landing *Distance (length)*, and *Weight* (*Load*) while the classes and subclasses of *Object*←*Facility*←*Airport*←***Runway*** will have a similar property of *Length* and another of *Load*. Thus, if a runway image falls into a particular runway ontology class, then the inferred condition that it will support certain aircraft is straight forward. The Database, on the other hand, has the explicit requirement of a schema entry to identify that runway has a certain characteristics as part of a data storage tuple, without inferring a particular aircraft can use that runway.

6.1 Storage Approaches

While in theory, the ontology for VideoIMINT could operate on any data video that was known to the ontology (*i.e.* standard video products); a tightly coupled storage system is more efficient. The ability to reference the storage system upon which the ontology operates is a great advantage. Short-term storage will make searches more efficient and rapid while longer term retrieval, the forensic search, can be enhanced as a class in the ontology with rules guiding which data is stored for what periods of time. Temporal redundancy, similar to information redundancy, can guide the "compression" of video for longer term, more efficient, storage if the storage rules operating on this data are clearly defined (semantically consistent). For example, a vehicle "track" can include the content of the tracked video as only a segment vector of the video frame through time. Utilizing a common method of video compression,

the “background” can be intermittent frames (I-frames, B-frames or P-frames of the MPEG specification) that maintain the slower changes in the surrounding scene. I t would be unnecessary to retain all the traditional I, B, or P-frames but rather only those frames useful to understand the context of the tracked target. Further compression is achieved by rendering these frames as wavelet compressed data according to the JPEG2000 compression schemes [13]. The track itself can be stored as a separate class of wavelet, type *Track*, with useful subclasses and properties. Regardless of scheme employed to store data from video, control of a short-term storage of data will enhance the operation of the ontology.

6.2 Data Fusion

When building an ontology of imagery content and associated metadata, these classes become the inputs for stipulated data fusion processing, at least for lower levels of the Joint Directors of Laboratories (JDL) Data Fusion Model (1998 revision) [14] that include Object Detection and Assessment and Object Refinement.
As targets are detected and assessed, declarations of object are made which in turn enables the population of ontological object classes (*e.g.* vehicle). The thresholds and rules governing this instancing are the same thresholds and rules that will (or will not) satisfy subsequent fusion processing of these detected objects. Associations of metadata, related to these instances will allow further Object Refinement in the sense of positioning, sizing and characterizing the ontological object thus enhancing fusion processes with associated metrics. Such qualifications will enable overall correctness of initial assessments in terms of accuracy, precision, and error within the fusion process.
Consider fusing two different collects of video data, from different sensor types, at different times covering a similar geographical location. The imagery must be collected, located, registered spatially and temporally, while the characteristics of the sensor, the look angle and altitude (for resolution purposes) all need to be considered to just begin the fusion process. However, the classes and properties that have been described previously in this paper do just that. Utilizing the metadata alone, almost all sensors and platforms provide this information, and it is rendered by the ontology into appropriate classes with properties. That information which is not collected, for example, pixel image resolution, can be readily calculated from sensor specification, sensor pointing data, and platform performance data, all readily available. The only other fusion requirement is that the ontology enforces semantic consistency of units and metrics. The fusion processes can now be built into the data processing chain with sensor selection tasking “switches” to choose appropriate sensors for a particular mission and appropriate systems operations. The data preconditioning for fusion is completed: leaving specific, mission related fusion processes, with inputs necessary for predictable, consistent sensor data fusion.
The construction of the ontology must however, consider such subsequent processing in the design of classes and properties. While the necessary metadata and class descriptions can be built, they may not be consistently populated from one sensor to the next of one collect to the next. We may provide the facility for the subsequent operation, which does not, however, guarantee fusion.

7 Summary

This prototype ontology construct for Video Imagery Intelligence collection demonstrated the value of integrating video metadata along with specification information and imagery content in an organized, semantically consistent structure based on standards. Additionally, direct logical queries into the ontology were able to identify video segments with tagged and extracted features and mark those segments for review by an analyst. The ontology structures appear to be a valuable and useful tool to bring under control the growing volumes of data that is being collected by Unmanned Aerial Vehicles in various mission circumstances. The ontologies, if developed correctly, can also be used as both a mission planning system and a dynamic control system based on proven approaches such as NIIRS guided tasking. Overall performance quality can be monitored in real time to ensure the efficient and effective operation of intelligence collection platforms.
Finally, the use of ontologies enforces a semantic consistency as well as maintenance of performance information that forms the basis of sensor data fusion. Using the information collected and categorized by the ontology promises to facilitate building new fusion processes based on simple class relationships such as location, dimensional information and sensor operational performance.

Acknowledgements

Mr. Ron Rudnicki of CUBRC and the National Center for Ontological Research for his guidance and help in developing the OWL ontology used in this project.
Mr. Todd Howlett of the Air Force Research Lab at Rome, NY for sponsoring this activity as part of AFRL research into MultiINT information systems understanding.

References

[1] A single, 16 Mpixel camera with twenty-four bytes per pixel (color) and a two hertz frame rate would generate about 455 Gbytes of data in ten minutes. Six such cameras on a platform would push well into the terabyte range in 10 minutes.

[2] http://www.nytimes.com/2010/01/11/business/11drone.html

[3] *Advanced Authoring Format Profile for Aerial Surveillance and Photogrammetry Applications*, Version 1.0, National Geospatial-Intelligence Agency Motion Imagery Standards Board, Washington D.C. January 8, 2006.

[4] *Ibid, UAV Datalink Local Metadata Set.*

[5] Câmara, G., Egenhofer, M., Fonseca, F., and Monteiro, A. M. V. *What's in an Image?* in: Montello, D. R., (Ed.), Spatial Information Theory—A Theoretical Basis for GIS, International Conference COSIT '01, Santa Barbara, CA.

[6] Waldman, G., Wootton, J., Electro Optics Systems Performance Modeling (pp190-192). Artech House, Norwood, MA. 1993.

[7] Bar-Shalom, Y., Li, Xiao-Rong, Multitarget-Multisensor Tracking: Principles and Techniques. Storrs, Connecticut, 1995.

[8] Corso, J.J., New York State University at Buffalo, Department of Computer Science and Engineering. 2010.

[9] Leachtenauer, J.C., Malila, W., Irvina, J., Colburn, L., Salvaggio, N. *General Image Quality Equation: GIQE*. Applied Optics, Vol.36, No. 32. November 1997.

[10] Granger, E.M., Cupery, K.N. *An optical merit function (SQF), which correlates with subjective image judgments*. Photographic Science and Engineering, Vol. 16, No. 3, May-June 1972.

[11] Horrocks, I., Ontologies and Databases, a W3C Presentation. Oxford University, 2010.

[12] Motik, B., University of Manchester; Ian Horrocks, Oxford University; Ulrike Sattler, University of Manchester, *Bridging the Gap Between OWL and Relational Databases*, World Wide Web Conference Committee, May 2007.

[13] *ISO/IEC 15444-1:2004 | ITU-T Rec. T.800* defines a set of lossless (bit-preserving) and lossy compression methods for coding bi-level, continuous-tone grey-scale, palletized colour, or continuous-tone colour digital still images.http://www.iso.org/iso/catalogue_detail.htm?csnumber=37674

[14] Hall, D., Llinas, J., Steinberg, A. N., Bowman, C.L., *Handbook of Multisensor Data Fusion*. pp **1**-7,8, **2**-5, ed. David Hall, James Llinas, CRC Press, LLC Boca Raton, Florida, 2001.