Staged Exploitation of Video Moving Target Detections, with Adherence to the Motion Imagery Standards Profile

Shannon Fehlmann, David Booth, Merrilyn Fiebig, Stephen Craig, Michael Royce
ISR Division
Defence Science and Technology Organisation
Edinburgh, Australia
{Shannon.Fehlmann, David.Booth, Merrilyn.Fiebig, Stephen.Craig, Michael.Royce}@dsto.defence.gov.au

***Abstract*—The endurance, pervasiveness and sophistication of airborne surveillance platforms and attached sensor payloads provides an ever increasing source of valuable, high quality imagery. However, as demands on analysts increase, automated cueing of features of interest, such as moving objects, becomes desirable. We describe our implementation of Video Moving Target Indication (VMTI) using a configuration of The Analysts' Detection Support System (ADSS). During the VMTI process, metadata is populated with target attributes. To promote interoperability, the metadata is compliant with international standards, namely The Motion Imagery Standards Board Profile, and in particular Engineering Guideline (EG) 0903[1].**

We introduce an interfacing design and a first implementation of the EG 0903. Using EG 0903, ADSS VMTI was interfaced to an airborne Sensor Management System (SMS) for the realtime transmission of VMTI metadata to analyst mapping tools and for realtime inclusion in an MPEG-2 Transport Stream (MPEG-2 TS). The MPEG-2 TS with EG 0903 was decoded by video players attached to the airborne network and was also saved to persistent storage. Detected moving targets in the video provided by our airborne imaging sensor were geographically and time associated and overlaid on a situational awareness and mapping tool. The MPEG-2 TS with EG 0903 was successfully decoded by standard compliant video player software, and is available for subsequent offline exploitation.

This paper describes a leading edge capability, architecture and constituent video exploitation algorithms. The development and utility of the architecture has influenced and driven international standards, namely MISB EG 0903, the use of which is demonstrated and reported for the first time. This work facilitates improved situational awareness, enriched metadata for video players and supports video content aware software such as video search.

Keywords-video metadata; airborne surveillance; video moving target indication; tracking

I. INTRODUCTION

Video imagery collected from a loitering airborne platform provides a powerful means of enhancing the tactical ground picture and improving situational awareness in general; an important reason being the capacity for motion imagery to capture scene dynamics. One way of filtering (or assessing the value of) particular sections of video is on the basis of vehicular or human activity within it. Video analysts continually search for signs of movement but over extended periods they can become fatigued, leading to missed targets, particularly when attention is focused on the more obvious movers. As the deployment of UAVs (in particular) increases, the demand for, and on, Imagery Analysts will increase. A relative shortage of analysts will potentially lead to some imagery not being considered at all, and will also necessitate a greater reliance on automated processing. The detection of moving targets in video has potential for automation, and the resulting tools can aid analysts at numerous points in the exploitation chain provided that the detection attributes can be stored in the accompanying video metadata.

The image processing engine used for this paper is a Video Moving Target Indication (VMTI) system first described by Jones et al. [1]. It operates by detecting and then tracking image pixels that are inconsistent with a gradually adapting background model. As previously, the VMTI system is implemented within the Analysts' Detection Support System (ADSS) which provides the architecture for interconnecting component algorithms to solve particular detection problems [2].

Metadata is considered vital for effective exploitation of ISR (Intelligence, Surveillance and Reconnaissance) related imagery because it provides context [3], particularly in a geographical sense. Until recently, the lack of effective geo-location and pointing information associated with video imagery limited its exploitation potential, particularly when associating with other types of still imagery and intelligence. Recently however, digital metadata has come into widespread use. It is increasingly compliant with NATO STANAG (Standardisation Agreement) 4609, and Motion Imagery Standards Board (MISB) standard 0601 [4] has become the accepted local data set within Defence for the transmission of metadata elements within motion imagery.

As well as being used to describe mission, platform, sensor and collection details, the metadata can store image content descriptors which, when populated and interrogated, offers the potential to cue analysts, reduce bandwidth requirements (using locally adaptive compression), and facilitate staged (or collaborative) exploitation. Metadata is particularly useful for

[1] This document refers to EG 0903 in general and in practice implements EG 0903.0. Subsequent to this implementation EG 0903 was elevated to a MISB Recommended Practice (RP) 0903.2.

recording events that are of little interest at that instant but which may become so later. It provides a means of accessing key portions of the video imagery with little compute effort, following which other forms of analysis (by human or machine) can be instigated. Booth et al. [5] used the free text field defined in Row 578 of the "SMPTE Metadata Dictionary" V4 Sept. 2002 to store descriptions of image content in a Lisp-like format. Since then, the Motion Imagery Standards Board (MISB) Engineering Guideline 0903 (EG) [6] that defines a Local Data Set (LDS) for the delivery of VMTI metadata has been defined and promulgated. The benefit of delivering this metadata in a standards compliant interfacing format is the assurance that the interfacing is best practice and that any compliant system will be compatible.

This paper presents the relevant imagery systems background, overviews our image processing systems, the standards compliant interfacing approach and the experimentation undertaken. It also summarises some of the analysis potential that this type of work presents, and in particular the opportunities it opens for video exploitation in a staged, or collaborative fashion, thereby promoting multi-source, multi-INT fusion.

II. Overview

The document is structured as follows. Section III describes the software systems, namely, the ADSS, the VMTI algorithms, and the Sensor Management System (SMS) software interfaces. Section IV describes the computers and the ISR sensors. Section V describes the metadata; its origins, and its flow through the system. Section VI outlines the imaging trial used to exercise the system, with the sections which follow outlining the findings and conclusions.

III. Software Systems

The software subsystems used for this work are: the Analysts' Detection Support System (ADSS) configured with an ADSS Video Moving Target Indication (VMTI) image processing pipeline and the Defence Experimentation Airborne Platform (DEAP) Sensor Management System (SMS). Both systems have been developed by Intelligence, Surveillance and Reconnaissance Division (ISRD) of the Defence Science and Technology Organisation (DSTO) of Australia.

A. ADSS

ADSS is a novel software system that assists Image Analysts by identifying targets in wide area surveillance imagery. ADSS is composed of a suite of specialist image processing algorithms and a software framework. The framework abstracts programming challenges not of direct concern to image processing algorithm developers such as image formats, hardware interfaces, compute parallelisation, user interfaces, and software build and debug tools [7].

In practice, to solve a given image processing problem, an ADSS developer constructs a pipeline of several specialist image processing algorithms (ADSS modules) [7]. ADSS was originally developed for automated processing of Synthetic Aperture Radar (SAR) imagery [2] but has evolved to a more general image and data processing framework.

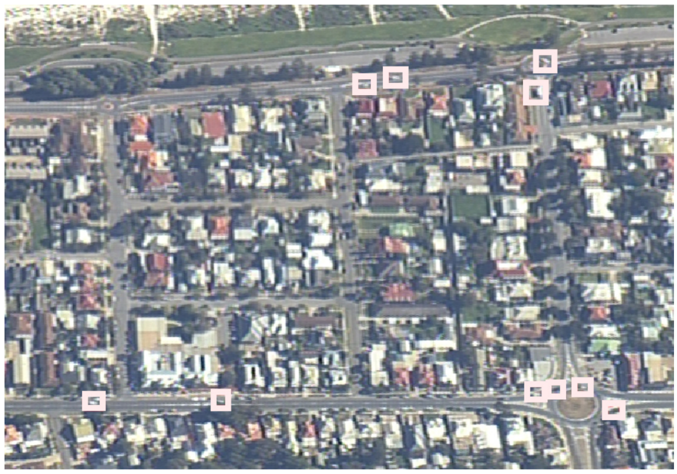

Figure 1. A screen capture of the visualisation provided by the ADSS VMTI Plotter module. The plotted moving targets in this case are vehicles. A highlight has been hand drawn over the plotted boxes for legibility.

B. The ADSS VMTI Image Processing Pipeline

An ADSS VMTI pipeline was described in previous work [8]. An overview is shown in Fig. 6 (the Fundamental ADSS Pipeline). To summarise, inter frame registration in the video sequence is achieved, and camera motion removed, by detecting and tracking features between frames [9]. After the inter frame affine transform has been calculated, a background model can then be developed by maintaining a block of recent frames in a moving temporal window and calculating an average intensity for each pixel [1]. Movement can then be calculated by comparing pixel values in the current frame against the calculated average background. To eliminate environmental movement and establish movement tracks persisting over time a straightforward temporal tracker was used to disregard movement against the background that was not persistent.

A Plotter module was used to display vehicle and pedestrian track histories on the ADSS VMTI (Host) Computer. The plotter was used to verify the quality of the imagery received by the ADSS VMTI Computer and for qualitative comparison of targets generated by the ADSS VMTI Computer against targets displayed as end product in the tested analyst imaging tools: FalconView [10] and MissionMonitor [11]. Fig. 1 depicts a part of the target indication visualisation from the Plotter.

C. The Video Audio Metadata Processor and Integrated Recording Environment (VAMPIRE) and Sensor Management System (SMS)

The SMS is software that interfaces to the imaging sensor, in this case a WESCAM MX-20 [12] and ADSS. The VAMPIRE encodes the video outputs from each of the three imaging sensors in the MX-20 and multiplexes the imagery with the metadata stream generated by the SMS. Motion imagery from the three sensors are simultaneously generated

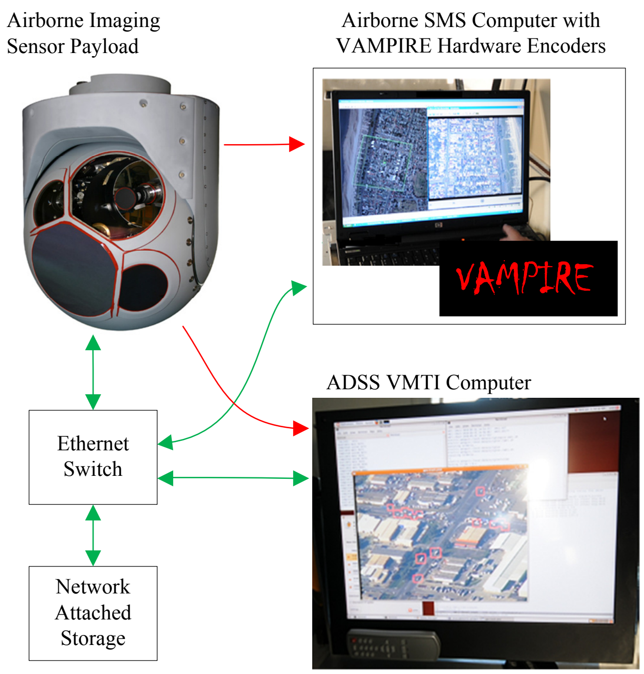

Figure 2. An interconnection diagram showing the physical appearance of the sensor payload (WESCAM MX-20 [12]), the VAMPIRE and SMS displaying analyst tools and the ADSS VMTI Computer with a realtime target plot.

and output in STANAG 4609 compliant MPEG-2 Transport Streams (MPEG-2 TS). This motion imagery is recorded to disk and multicast to provide a realtime interface to analyst tools such as FalconView and MissionMonitor. A DSTO developed software plug-in written for FalconView decodes the metadata to display the geolocation of the aircraft and ingest situational awareness data from the sensors. The SMS allows the user to target and cross-cue the MX-20 via geographically referenced map overlays.

IV. HARDWARE

Physical interconnection of the hardware was straightforward. Video inputs are taken from the MX-20 by the ADSS VMTI Computer and the VAMPIRE. Inter-systems communication, metadata and encoded video were transmitted over a switched Gigabit Ethernet network. Hardware systems interconnection is outlined in Fig. 2.

A. *Computers*

The ADSS VMTI computer is a Linux computer with two, four core processors and two Nvidia GeForce 200 Series GPUs. A typical hardware configuration for an ADSS processing pipeline is to allow one core per ADSS compute module. The frame registration, background modelling and plotter modules are GPU accelerated using the Cg shader language. The attractive compute density offered by GPU accelerated computing [13] is important for realtime image processing where space might be constrained or transportability is important.

The SMS consists of a laptop with Windows XP, the analyst imaging tools under test, the SMS software and the VAMPIRE is a rack mounted computer containing hardware video encoders and removable storage.

B. *Airborne Imaging Sensor*

The Airborne Imaging Sensor Payload was an L-3 Wescam MX-20 turret [12] (see Fig. 2). It contains three gyrostabilised imaging sensors: an electro-optic (colour) wide field of view (FOV) camera; an electro-optic (colour) narrow FOV camera and an infrared (3-5 micron) camera.

Timestamped estimates of the gimbal’s location, orientation, and the internal parameters of the EO sensors are input into a sensor model on the SMS. This sensor model is then projected onto a digital elevation model (Digital Terrain Elevation Data Level II). This provides estimates for the slant range, as well as the geographical location corresponding to the frame centre and corner points of the EO sensors.

V. SOFTWARE INTERFACING

First, we present the developed ADSS VMTI interfacing and then discuss how the generated metadata is propagated through the interconnected airborne systems. Fig. 3 summarises the interfacing.

In brief, the integration is achieved by the ADSS VMTI Computer ingesting system time from the SMS. Timecoded moving targets (as pixel locations) are then provided back to the SMS. The SMS receives timecoded geolocation metadata describing aircraft position and velocity, the imaging sensor orientation and internal parameters. This allows the SMS to fuse the VMTI targets with geolocation metadata, first by timecoding and second, by referencing a digital elevation model to geolocate targets. Geolocated targets are plotted on the Realtime Visualisation Tools. Fig. 3 also outlines data transmitted over software interfaces, the relevant standards and the video capture hardware interconnects.

A. *Metadata Standards*

Standards compliant metadata provides an agreed upon best practice for intercommunication of heterogeneous systems. In the airborne VMTI domain the use of the Key Length Value (KLV) format is particularly important to ensure optimum use of comparatively low bandwidth and potentially unreliable data channels [14]. A typical scenario where a communications link with these characteristics may exist would be an aircraft communicating with a ground station.

The KLV standards used for systems interconnection in our experimentation are outlined in Fig. 3. The Motion Imagery Standards Board (MISB) Standard (STD) 0601.4 [15] for transmission of Defence digital motion imagery systems data was used to convey geo and time coding. The MISB EG 0903.0 VMTI metadata standard [6] was used to convey timecoding to ADSS VMTI, used to convey timecoded targets to the SMS and embedded in the SMS generated STD 0601.4 which was then multiplexed with the encoded video into an MPEG-2 Transport Stream.

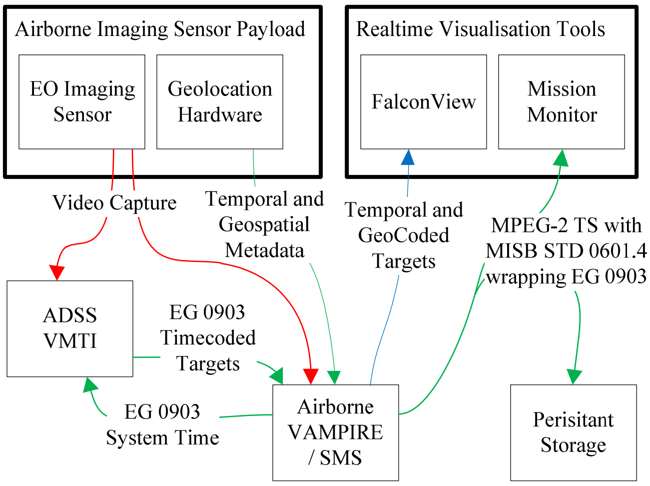

Figure 3. The logical interconnectivity of the system. Video capture interconnects are analog, the SMS to FalconView interface is via an API on a shared host computer and all other interfacing is over the switched ethernet network.

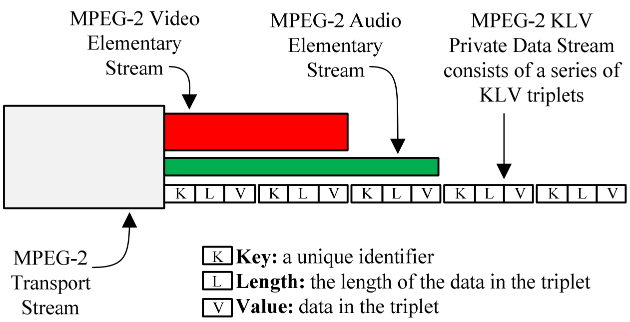

Figure 4. MPEG-2 transport of video, audio and KLV metadata.

```
00: 06 0E 2B 34 02 0B 01 01 0E 01 03 03 06 00 00 00
10: 20 02 07 04 A5 3E 3B 57 AE 40 05 02 00 01 06 02
20: 00 01 65 09 08 01 01 05 00 00 07 8B D1 04 02 00
30: 00 06 0E 2B 34 02 0B 01 01 0E 01 03 03 06 00 00
```

Figure 5. A representative, timecoded MISB EG 0903 compliant KLV hexadecimal byte stream produced by the ADSS VMTI computer.

B. The EG 0903 Standard for VMTI Systems Integration

In complying with EG 0903 for this systems design and experimentation, a benefit is that the system interconnects become pluggable. The ADSS VMTI implementation becomes interchangeable with any other compliant VMTI implementation for the system described by Fig. 3. The corollary is, of course that ADSS VMTI can be plugged into any compliant EG 0903 surveillance system. In fact, the standard and interfaces allow for this pluggable reconfiguration to occur during system operation.

C. ADSS VMTI with MISB EG 0903 KLV

The MISB EG 0903 KLV was transmitted as binary datagrams over a User Datagram Protocol (UDP) socket using the airborne network. KLV consists of a series of Key Length Value triplets. Fig. 4 describes how the KLV format is embedded with MPEG, KLV can also be transmitted as a standalone metadata stream.

Fig. 5 is an indicative snapshot of a datagram transmitted from the ADSS VMTI to the SMS. These datagrams are however, interpreted as a KLV byte stream at the application layer. The first 16 bytes uniquely identify the KLV triplet, bytes 17 to 25 are the timecoding and bytes 34 to 44 are the pixel location of the moving target.

D. Augmenting the ADSS VMTI Pipeline

The EG 0903 interfacing necessary for the ADSS VMTI pipeline is outlined in Fig. 6. The fundamental ADSS VMTI pipeline [8] ingests video via the Video Capture Service offered by the ADSS framework. The Video Capture Service decodes imagery once and presents it as an uncompressed buffer to any requesting ADSS module.

ADSS pipelines are selections of ADSS modules connected with communication links in directed acyclic graphs. Inter module communication is unidirectional. That is, ADSS modules read from their predecessors and write to their successors. For our experimentation the Fundamental ADSS VMTI Pipeline was augmented with EG 0903 interface modules.

The EG 0903 System Time ingestion module reads the current SMS System Time as KLV over UDP. The system time is then propagated along the ADSS pipeline using the ADSS framework Command and Data Language (CDL). The CDL timestamp progressed through video frame registration, background modelling, tracking and plotting modules, and was read by the EG 0903 Timecoded Target Output module and used to maintain a copy of the current time.

On receiving notification from the tracker (by way of a CDL message) that a moving target has been identified in the imagery, the EG 0903 Timecoded Target Output module generates a timecoded KLV UDP datagram and transmits it to the SMS.

It is worth noting that the ADSS EG 0903 interface could be elaborated to output additional information about the target such as colour, a mask to describe the target's shape, an image chip containing a picture of the target or the target's track

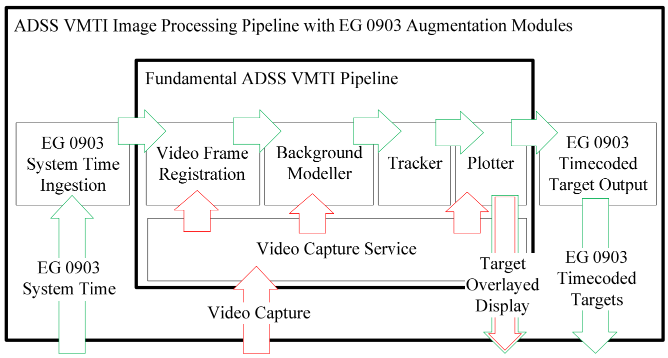

Figure 6. A representation of the ADSS VMTI pipeline used for this experimentation. Metadata propagates through the complete pipeline, captured imagery is ingested by modules as needed.

Figure 7. A screen capture depicting a moving target overlay on a FalconView map. Moving targets in the scene, the boresight ray and approximate sensor footprint are also depicted.

history information. In addition, as data ingested from the SMS is propagated through each VMTI module, commands might be issued from the SMS to (for example) configure the tracker for a particular target.

E. *Exploitation with the SMS*

Upon receiving an EG 0903 VMTI timecoded target the SMS generates STD 0601.4 KLV triples to describe the geospatial location of an object. Each EG 0903 triplet is nested in the STD 0601.4 and included in the MPEG-2 Transport Stream. Additionally, the SMS makes the geographical location and time associated targets available in realtime to FalconView via a plug-in API. A realtime moving target overlay can be presented to the FalconView user, see Fig. 7 for this visualisation.

The MPEG-2 Transport Stream is used in two ways. First, the stream is read in realtime by MissionMonitor. The stream could be broadcast over the onboard network to service an arbitrary number of video viewers. MissionMonitor is able to interpret the KLV embedded in the MPEG-2 TS. The user is able to toggle a MissionMonitor overlay to highlight moving targets on the live motion imagery stream or during playback. Second, the stream is saved to network attached storage for later review.

VI. EXPERIMENTATION

It was important to evaluate the performance of the systems, interfaces and metadata standards in a representative operational environment.

A. *Overview*

The hardware: the ADSS VMTI Computer, Airborne VAMPIRE and SMS, the Airborne Imaging Sensor Payload and network infrastructure (see Fig. 2) were installed on an airborne platform. The performance of the system was observed over a number of trial flights. For this work the experimentation team were interested in exploring the performance of the:

- systems
- systems interfaces
- EG 0903 standard for systems interconnection
- automated time and geo association to detected pixel targets and automated inclusion in the MPEG-2 TS
- effectiveness of visualisations on situational awareness tools

B. *Troubleshooting*

Unexpected challenges associated with the airborne deployment included:

- debugging configuration issues arising from the airborne network environment - packet sniffing was required
- optimising ADSS VMTI algorithm performance for the scenarios – sensor and platform characteristics, target size, atmospheric conditions
- electromagnetic noise in the airborne environment meant additional shielding on analog cabling

C. *Results*

The intention was to perform a qualitative assessment of the relevant aspects of the systems performance listed previously and to evaluate the benefit arising from the interconnection of these airborne systems. Discussion of the performance follows.

The ADSS VMTI Computer was able to provide target indications from the live imagery feed in realtime to the SMS and the SMS was able provide the geo associated targets to FalconView in realtime. Measured latency for FalconView targets and MPEG-2 TS production was approximately 500ms. This was deemed suitable for realtime operator situational awareness.

Propagation of the time and geocoding through the interconnected systems by STD 0601.4 and otherwise was effective. Propagation of time coding and target pixel locations through the interconnected systems by EG 0903 was effective. Fig. 7 is an indicative example of the accuracy of the geo data for use as a map overlay. The circles depicted are vehicles travelling along the adjacent suburban roadways. While, to human eyes the geospatial accuracy of the target indication is not perfect, we feel this type of visualisation is still of value for realtime target cueing.

The accuracy of the geospatial location data fit varied with the collection geometry of the scene. It was generally best when the MX-20 turret and aircraft were at constant velocity with minimal roll/slew/pitch and where wind buffeting of the aircraft caused by air turbulence was minimal.

During the experimentation, the ADSS VMTI Computer and the SMS were powered off and back on, and disconnected numerous times. The systems interfaces re-established the metadata flows with no special intervention from the experimentation team.

VII. Conclusion

This paper has described a leading edge capability, architecture and constituent video exploitation algorithms. The development and utility of the architecture has influenced and driven international standards, namely MISB EG 0903, effectiveness has been the use of which has been demonstrated. More specifically, we have demonstrated effective, robust systems interoperability by using MISB EG 0903 compliant metadata to communicate moving target information in an operations like scenario. The MISB EG 0903 standard was effective at communicating system time to the ADSS VMTI Computer, timecoded pixel targets to the SMS and for inclusion in the generated MPEG-2 Transport Stream.

The metadata allowed us to produce realtime target visualisations as target markers on a mapping tool – FalconView and to display the realtime MPEG-2 TS with target annotations in a MISB EG 0903 compliant video viewer – MissionMonitor. Subsequent analysis of the metadata enriched video will support video content search, and promote multi-source and multi-INT fusion.

VIII. Discussion

These systems have been improved since this initial work. ADSS GPU accelerated modules now use the Nvidia CUDA interface and the Plotter displays its visualisations using X Window System calls. This means the ADSS VMTI Computer can now be deployed as a rack mounted system without a host display. The Plotter visualisations can now be displayed on a laptop hosting an X server over an ethernet connection. SMS development since this work has focused on improving integration with high definition sensors, improved performance of geolocation algorithms, and software refactoring in the form of componentisation.

The VMTI EG 0903 is now available as MISB Recommended Practice (RP) 0903.2 [16]. Like others [17], we see content based video search as an important tool for limiting the time analysts spend reviewing video. Associating RP 0903.2 KLV means standards based video search can traverse all compliant video.

With respect to improving the quality of the metadata during rapid changes in aircraft and sensor velocity, we feel the application of statistical techniques to smooth and interpolate metadata may also prove useful.

Acknowledgement

The imaging, metadata and trials lead was Merrilyn Fiebig, the SMS integration and development lead was Stephen Craig (DSTO Contractor), the ADSS integration lead was Shannon Fehlmann with imaging research oversight by Dr David Booth. Michael Royce led the development of VMTI EG 0903. Our thanks are also extended to Warwick Holen, Peter Perry, Dr Nick Redding and Rod Smith and the ADSS development team for their support.

References

[1] R. Jones, B. Ristic, N. J. Redding, and D. M. Booth, "Moving Target Indication and Tracking from Moving Sensors", in Proc. DICTA'05, page 46, 2005.

[2] N. J. Redding, D. I. Kettler, G. Blucher, and P. Perry, "The Analysts' Detection Support System for deploying a network of target detection and recognition algorithms in SAR exploitation", in Proc. Int. Conf. on RADAR, pp. 448–453, 2003.

[3] D. L. Young, "Motion imagery metadata standards assist in object and activity classification" AIPR'10, 2010, pp.1–4, Oct. 2010.

[4] UAS Datalink Local Metadata Set, MISB STD 0601.4, Mar. 2010.

[5] D. M. Booth, N. J. Redding, R. Jones, M. Smith, I. J. Lucas, and K. L. Jones, "Federated Exploitation of All-Source Imagery", in Proc. DICTA'05, pp. 243–250, 2005.

[6] Video Moving Target Indicator Local Data Set, MISB EG 0903.0, Sep 2009.

[7] N. Redding, "Design of the Analysts' Detection Support System for Broad Area Aerial Surveillance", DSTO, Australia, DSTO-TR-0746, 1998.

[8] R. Jones, D. Booth, and N. Redding, "Video Moving Target Indication in the Analysts' Detection Support System", DSTO, Australia, DSTO-RR-0306, 2006.

[9] Shi, and C. Tomasi, "Good features to track," Proc. CVPR'00, pp.593–600, June 1994

[10] (2011) The FalconView™ webpage. [Online]. Available: http://www.falconview.org/trac/FalconView

[11] (2011) The Mediaware MissionMonitor™ webpage. [Online]. Available: http://mediaware.com.au/Products/Product.aspx?ProductId=22

[12] (2011) The WESCAM™ MX-20 webpage. [Online]. Available: http://www.l-3com.com/wescam/products/products_services_1h.asp

[13] D. Owens, D. Luebke, N. Govindaraju, M. Harris, J. Krüger, A.E. Lefohn, and T. J. Purcell, "A Survey of General-Purpose Computation on Graphics Hardware", Computer Graphics Forum, pp. 80–113, 2007.

[14] MISB, Key-Length-value (KLV) User's Guide, Draft, unpublished, Jul 2010

[15] UAS Datalink Local Metadata Set, MISB STD 0601.4, Mar. 2010.

[16] MISB, Video Moving Target Indicator Local Data Set, MISB RP 0903.2, Feb. 2011.

[17] S. Chang, W. Chen, H. J. Meng, H. Sundaram and D. Zhong, "A fully automated content-based video search engine supporting spatiotemporal queries" Circuits and Systems for Video Technology, IEEE Transactions on, vol. 8, no.5, pp. 602–615, Sep 1998.